Probabilidad y variables aleatorias para ML con R y Python

Ricardo Alberich, Juan Gabriel Gomila y Arnau Mir

2019-11-26

Contents

Consulta el curso completo creado por Ricardo Alberich, Juan Gabriel Gomila y Arnau Mir solamente en Udemy

Puedes consultar todas las transparecias del curso en formato HTML desde nuestro Gihub.io

Asienta las bases para convertirte en el Data Scientist del futuro con todo el contenido de estadística descriptiva del curso. En particular verás los mismos contenidos que explicamos en primero de carrera a matemáticos, ingenieros, economistas, biólogos, médicos o informáticos.

1. Probabilidad
2. Variables aleatorias
3. Distribuciones notables
4. Complementos avanzados
5. Variables bidimnsionales
6. Variables multidimensionales
7. Convergencia y Teorema Central del límite

Y todo con más de 30 horas de vídeo a demanda, cientos de ejercicios, tareas, talleres y trucos de los profesores para que te conviertas en un experto de la materia.

Pre requisitos: Teoría de conjuntos y combinatoria

0.1 Antes de empezar

0.1.1 Consideraciones

Para aprender cálculo de probabilidades son necesarios conocimientos de:

1. Cálculo: Derivadas, integrales, límites, sumas de series…
2. Geometría básica y álgebra lineal : rectas, hiperplanos, volúmenes… Matrices, valores propios…
3. Teoría de conjuntos y combinatoria…..

Por experiencia sabemos que la mayoría de estudiantes tienen más conocimientos de cálculo, geometría y matrices.

Pero muchos tienen una falta de conocimientos en teoría básica de conjuntos y combinatoria (matemática discreta).

0.1.2 Teoría de conjuntos

Definición de conjunto

La definición de conjunto es una idea o noción primitiva. Es decir es una idea básica del pensamiento humano: un conjunto es una colección de objetos: números, imágenes… cualquier cosa, jugadores de fútbol, palabras, colores ….

La teoría de conjuntos básicas es simple y natural y es la que necesitamos para este curso.

La teoría de conjuntos matemática es más complejas y presenta varias paradojas como la paradoja de Russell.

La idea o noción práctica de conjunto es la de una colección de objetos de un cierto tipo.

Estas colecciones o conjuntos se pueden definir por:

- Compresión: reuniendo los objetos que cumplen una propiedad p
- Extensión: dando una lista exhaustiva de los miembros del conjunto

0.1.3 Conjuntos básicos

Los conjuntos suelen tener un conjunto madre como por ejemplo

- $\mathbb{N} = \{0, 1, 2, \ldots\}$
- $\mathbb{Z} = \{\ldots, -2, -1, 0, 1, 2, \ldots\}$
- $\mathbb{Q} = \left\{\frac{p}{q} \quad \middle| \quad p, q \in \mathbb{Z} \text{ y } q \neq 0.\right\}$
- $\mathbb{R} = \{\text{Todos los puntos de una recta.}\}$
- $\mathbb{C} = \left\{a + b \cdot i \quad \middle| \quad a, b \in \mathbb{R}\right\}$ Los números complejos $\quad a + b \cdot i$.
- Alfabeto $= \{a, b, c, \ldots, A, B, C, \ldots\}$.
- Palabras $= \{paz, guerra, amor, probabilidad, \ldots\}$.

Recordemos que i es la unidad imaginaria que cumple que $\sqrt{i} = -1$.

0.1.4 Características y propiedades básicas de los conjuntos

- Si a cada objeto x de Ω le llamaremos **elemento del conjunto** Ω y diremos que x pertenece a Ω. Lo denotaremos por $x \in \Omega$.
- Un **conjunto de un elemento**, por ejemplo $\{1\}$ recibe el nombre de **conjunto elemental** (o **singleton** del inglés).
- Sea A otro conjunto diremos que A **es igual a** B si todos los elementos A están en B y todos los elementos de B están en A. Por ejemplo $A = \{1, 2, 3\}$ es igual a $B = \{3, 1, 2\}$.
- Si A es otro conjunto tal que si $x \in A$ entonces $x \in B$ diremos que A es un subconjunto de o que está contenido en B. Lo denotaremos por $A \subseteq B$.
- El conjunto que no tiene elementos se denomina conjunto vacío y se denota por el símbolo $\emptyset$.
- Dado A un conjunto cualquiera obviamente $\emptyset \subseteq A$.

Tomemos como conjunto base $\Omega = \{1, 2, 3\}$

- Ω es un conjunto de cardinal 3, se denota por $\#(\Omega) = 3$ o por $|\Omega| = 3$
- El conjunto Ω tiene $2^3 = 8$ subconjuntos.
- el vacio $\emptyset$ y los elementales $\{1\}, \{3\}, \{3\}$
- los subconjuntos de dos elementos: $\{1, 2\}, \{1, 3\}, \{2, 3\}$
- El conjunto total de tres elementos $\Omega = \{1, 2, 3\}$.

Dado un conjunto Ω podemos construir el **conjunto de todas sus partes** (todos sus subconjuntos) al que denotamos por $\mathcal{P}(\Omega)$. También se denomina de forma directa partes de Ω.

Cardinal de las partes de un conjunto

El cardinal de la partes de un conjunto es $\#(\mathcal{P}(\Omega)) = 2^{\#(\Omega)}$.

Por ejemplo $\#\left(\mathcal{P}(\{1,2,3\})\right) = 2^{\#(\{1,2,3\})} = 2^3 = 8$.

Efectivamente

$$\mathcal{P}(\{1,2,3\}) = \{\emptyset, \{1\}, \{2\}, \{3\}, \{1,2\}, \{1,3\}, \{2,3\}, \{1,2,3\}\}.$$

Dado un subconjunto A de Ω podemos construir la función característica de A

$$\chi_A : \Omega \to \{0,1\}$$

dado un $\omega \in \Omega$

$$\chi_A(\omega) = \begin{cases} 1 & \text{si } \omega \in A \\ 0 & \text{si } \omega \notin A \end{cases}$$

0.1.5 Operaciones con conjuntos

0.1.5.1 Intersección.

Sea Ω un conjunto y A y B dos subconjuntos de Ω.

El conjunto **intersección** de A B es el formado por todos los elementos que perteneces a A Y B, se denota por $A \cap B$.

Más formalmente

$$A \cap B = \left\{x \in \Omega \middle| x \in A \text{ y } x \in B\right\}.$$

0.1.5.2 Unión.

El conjunto **unión** de A y B es el formado por todos los elementos que perteneces a A O que pertenecen a B, se denota por $A \cup B$.

Más formalmente

$$A \cup B = \left\{x \in \Omega \middle| x \in A \text{ o } x \in B\right\}.$$

0.1.5.3 Diferencia.

El conjunto **diferencia** de A y B es el formado por todos los elementos que perteneces a A Y NO pertenecen a B, se denota por $A-B=A-(A\cap B)$.

Más formalmente

$$A-B=\left\{x\in\Omega\middle|x\in A \text{ y } x\notin B\right\}.$$

0.1.5.4 Complementario

El **complementario** de un subconjunto A de Ω es $\Omega-A$ Y se denota por A^c o $\overline{A}$.

Más formalmente

$$A^c=\left\{x\in\Omega\middle|x\notin A\right\}.$$

0.1.6 Más propiedades y definiciones

Sea Ω un conjunto y A,B,C tres subconjuntos de Ω

- Se dice que dos conjuntos A y B **son disjuntos** si $A\cap B=\emptyset$.
- $\Omega^c=\emptyset$.
- $\emptyset^c=\Omega$.
- $A\cup B=B\cup A$, $A\cap B=B\cap A$ conmutativas
- $(A\cup B)\cup C=A\cup(B\cup C)$, $(A\cap B)\cap C=A\cap(B\cap C)$ asociativas
- $A\cup(B\cap C)=(A\cup B)\cap(A\cup C)$, $A\cap(B\cup C)=(A\cap B)\cup(A\cap C)$ distributivas
- $(A^c)^c=A$ doble complementario
- $(A\cup B)^c=A^c\cap B^c$, $(A\cap B)^c=A^c\cup B^c$ leyes de De Morgan

0.1.7 Con R, ejemplos.

Con R los conjuntos de pueden definir como vectores

```
(Omega=c(1,2,3,4,5,6,7,8,9,10))
```

```
##  [1]  1  2  3  4  5  6  7  8  9 10
```

```
(A=c(1,2,3,4,5))
```

```
## [1] 1 2 3 4 5
```

```
(B=c(1,4,5))
```

```
## [1] 1 4 5
```

```r
(C=c(4,6,7,8))
```

```
## [1] 4 6 7 8
```

$A \cap B$

```r
intersect(A,B)
```

```
## [1] 1 4 5
```

$A \cup B$

```r
union(A,B)
```

```
## [1] 1 2 3 4 5
```

$B - C$

```r
setdiff(B,C)
```

```
## [1] 1 5
```

$A^c = \Omega - A$

```r
setdiff(Omega,A)
```

```
## [1]  6  7  8  9 10
```

0.1.8 Con python

```python
Omega=set([1,2,3,4,5,6,7,8,9,10])
Omega
```

```
## set([1, 2, 3, 4, 5, 6, 7, 8, 9, 10])
```

```python
A=set([1,2,3,4,5])
A
```

```
## set([1, 2, 3, 4, 5])
```

```python
B=set([1,4,5])
B
```

```
## set([1, 4, 5])
```

```python
C=set([4,6,7,8])
C
```

```
## set([8, 4, 6, 7])
```

```python
A & B   ## intersección (&: and/y)
```

```
## set([1, 4, 5])
```

```
A | B   ## unión (|: or/o)
```

```
## set([1, 2, 3, 4, 5])
```

```
A - C   ## diferencia
```

```
## set([1, 2, 3, 5])
```

```
Omega-C ## complementario.
```

```
## set([1, 2, 3, 5, 9, 10])
```

0.2 Combinatoria

La combinatoria es una rama de la matemática discreta que entre otras cosas cuenta distintas configuraciones de objetos de un conjunto.

Por ejemplo si tenemos un equipo de baloncesto con 7 jugadores ¿cuántos equipos de 5 jugadores distintos podemos formar?

0.2.1 Número binomial.

Número combinatorio o número binomial

Nos da el número de subconjuntos de tamaño k de un conjunto de tamaño n. Este número es

$$\binom{n}{k} = \frac{n!}{k! \cdot (n-k)!}.$$

Recordemos que

$$n! = 1 \cdot 2 \cdot 3 \cdots n.$$

En nuestro caso con 7 jugadores $n = 7$ el número de equipos distintos de $k = 5$ es

$$\binom{7}{5} = \frac{7!}{5! \cdot (7-5)!} = \frac{7!}{5! \cdot 2!} = \frac{1 \cdot 2 \cdot 3 \cdot 4 \cdot 5 \cdot 6 \cdot 7}{1 \cdot 2 \cdot 3 \cdot 4 \cdot 5 \cdot 1 \cdot 2} = \frac{6 \cdot 7}{2} = \frac{42}{2} = 21.$$

Puedo formar 21 equipos distintos.

Ejercicio

Carga el paquete `gtools` de R y investiga la función `combinations(n, r, v, set, repeats.allowed)` para calcular todas las combinaciones anteriores.

0.2.2 Variaciones.

Variaciones

Con los número $\{1, 2, 3\}$ ¿cuántos números de dos cifras distintas podemos formar sin repetir ninguna cifra?

La podemos escribir

$$12, 13, 21, 23, 31, 32$$

Luego hay seis casos.

Denotaremos las variaciones (sin repetición) de k elementos (de orden k) de un conjunto de n elementos por V_k^n su valor es

$$V_k^n = \frac{n!}{(n-k)!} = (n-k+1)\cdot(n-k+2)\cdots n.$$

En nuestro ejemplo con $n = 3$ dígitos podemos escribir las siguientes variaciones de orden $k = 2$

$$V_{k=2}^{n=3} = \frac{3!}{(3-2)!} = \frac{1\cdot 2\cdot 3}{1} = 6.$$

Ejercicio

Carga el paquete `gtools` de R y investiga la función `permutations(n, r, v, set, repeats.allowed)` para calcular todas las variaciones anteriores.

0.2.3 Variaciones con repetición.

Variaciones con repetición

¿Y repitiendo algún dígito?

$$VR_k^n = n^k$$

Efectivamente en nuestro caso

$$11, 12, 13, 21, 22, 23, 31, 32, 33$$

$$VR_{k=2}^{n=3} = n^k.$$

0.2.4 Permutaciones

Las permutaciones de un conjunto de cardinal n son todas las variaciones de orden máximo n. Las denotamos y valen:

$$P_n = VR_n^n = n!$$

Por ejemplo todos los números que se pueden escribir ordenando todos los dígitos $\{1, 2, 3\}$ sin repetir ninguno

```
library(combinat)
for(permutacion in permn(3)) print(permutacion)
```

```
## [1] 1 2 3
## [1] 1 3 2
## [1] 3 1 2
## [1] 3 2 1
## [1] 2 3 1
## [1] 2 1 3
```

Efectivamente

$$P_3 = 3! = 1 \cdot 2 \cdot 3.$$

Ejercicio

Carga el paquete `combinat` de R e investiga la funcion `permn` para calcular todas las permutaciones anteriores.

Ejercicio

Investiga el paquete `itertools` y la función `comb` de `scipy.misc` de Python e investiga sus funciones para todas las formas de contar que hemos visto en este tema.

Ejercicio

La función gamma de Euler, cobrará mucha importancia en el curso de estadística. Comprueba que la función `gamma(x+1)` da el mismo valor que la función `factorial(x)` en `R` para todo $x = \{1, 2, 3 \cdots, 10\}$.

0.3 Para acabar

0.3.1 Otros asuntos básicos

Evidentemente nos hemos dejado muchas otras propiedades básicas de teoría de conjuntos y de combinatoria como:

- Números multinomiales.

- Combinaciones con repetición
- Propiedades de los números combinatorios.
- Binomio de Newton.
- Multinomio de Newton.

Si nos son necesarias las volveremos a repetir a lo largo del curso o bien daremos enlaces para que las podáis estudiar en paralelo.

Nota

Puedes repasar todos esos conceptos con ejercicios y más en el Curso de estadística descriptiva con `R` y `Python` con M. Santos y J.G. Gomila.

Chapter 1

Probabilidad

1.1 Probabilidades Básicas

Experimento aleatorio: experimento que repetido en las mismas condiciones puede dar resultados diferentes, pero que a largo plazo son predecibles

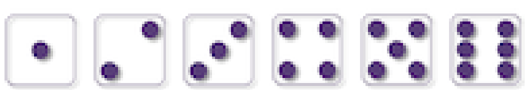
Ejemplo

Tirar un dado de 6 caras y anotar el número de puntos de la cara superior.

Suceso elemental: cada uno de los posibles resultados del experimento aleatorio

Ejemplo

Los sucesos elementales del ejemplo anterior serían:

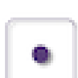
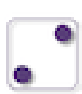
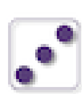
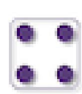
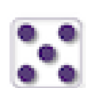
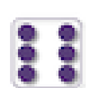


1.1.1 Definiciones básicas

Espacio muestral: el conjunto Ω formado por todos los sucesos elementales del experimento aleatorio

Ejemplo

El espacio muestral del ejemplo anterior del dado es

$\Omega = \{$  ,  ,  ,  ,  ,  $\}$

pero por comodidad, a partir de ahora pondremos

$$\Omega = \{1, 2, 3, 4, 5, 6\}$$

1.1.2 Definiciones básicas

Suceso : Cualquier subconjunto del espacio muestral.

Alguno sucesos notables que merece la pena nombrar son:

- Suceso seguro o cierto: Ω
- Suceso imposible o vacio: $\emptyset$
- Partes de un conjunto: $\mathcal{P}(\Omega)$: conjunto de todos los sucesos del experimento aleatorio (es decir, el conjunto de todos los subconjuntos de Ω)

Ejercicio

¿Cuantos elementos contiene el conjunto de partes de Ω del experimento anterior?

1.1.3 Ejemplo n-grama

Se define un n-grama de una palabra como el conjunto de n letras consecutivas de la misma (contando los blancos de inicio y final de palabra que marcamos como "_").

Ejemplo

Consideremos el experimento aleatorio que consiste en escoger al azar un 3-grama de la palabra "_Baleares_". Vamos a escribir el espacio muestral y algunos sucesos elementales del mismo.

En este caso, si consideramos la palabra "_Baleares_", el espacio muestral del experimento sería:

$$\Omega = \{\_Ba, Bal, ale, lea, ear, are, res, es\_\}$$

Algunos sucesos serían:

- 3-gramas que empiezan por a: $\{ale, are\}$
- 3-gramas de inicio y final de palabra: $\{\_Ba, es\_\}$
- 3-gramas que contengan una l: $\{Bal, ale, lea\}$

1.1.4 Operaciones con sucesos

Si tenemos dos sucesos $A, B \subseteq \Omega$, podemos definir:

- Ω: *suceso* total o *seguro*
- $\emptyset$:suceso *vacío* o *imposible*
- $A \cup B$: suceso *unión*; el que ocurre si sucede A o B
- $A \cap B$: suceso *intersección*; el que ocurre si sucede A y B
- A^c: suceso *complementario* el que sucede si NO sucede A.
- $A - B = A \cap B^c$: suceso *diferencia*, que acontece si sucede A y NO sucede B.

Sucesos incompatibles: A y B son *incompatibles* (o *disjuntos*) cuando $A \cap B = \emptyset$.

1.1.5 Ejemplo género

Ejemplo

Supongamos que el sexo se divide entre Mujeres y Hombres. Vamos a definir el espacio muestral, los sucesos elementales y a realizar algunas operaciones entre ellos.

- Estudiantes de esta clase: Ω
- Mujeres de esta clase: A
- Estudiantes que son zurdos B

Algunas operaciones entre los conjuntos:

- $A \cup B$: Est. que son mujeres o que son zurdos.
- $A \cap B$: Mujeres de esta clase que son zurdas.
- A^c: Hombres de esta clase.
- $A - B$: Mujeres de la clases que NO son zurdas.
- $B - A$: Hombres de la clase que son zurdos
- ¡Cuidado! No son incompatibles

1.1.6 Propiedades

Conmutativas:

$$A \cup B = B \cup A, \quad A \cap B = B \cap A$$

Asociativas:

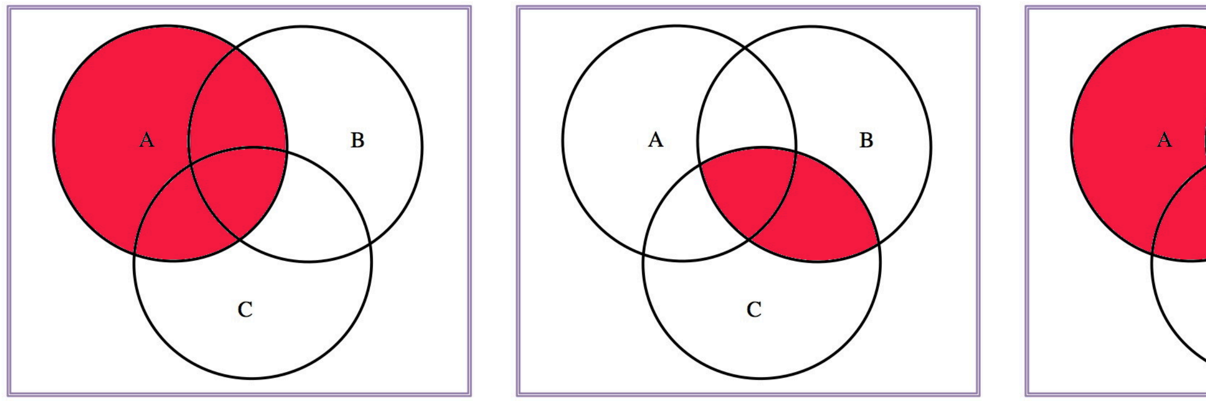
$$A \cup (B \cup C) = (A \cup B) \cup C, \quad A \cap (B \cap C) = (A \cap B) \cap C$$

Distributivas:

$$A \cap (B \cup C) = (A \cap B) \cup (A \cap C), \quad A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$$

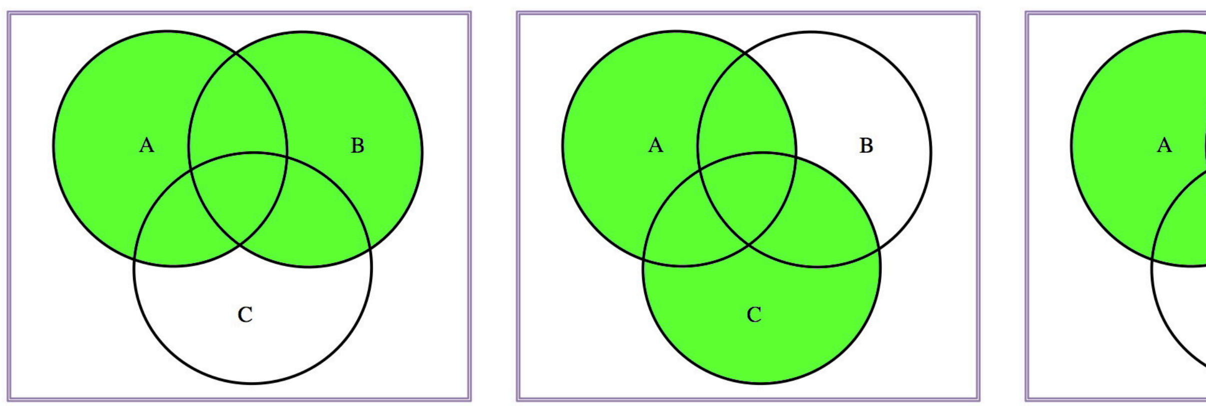
1.1.7 Propiedades

A	$B \cap C$	$A \cup (B \cap C)$



1.1.8 Propiedades

$A \cup B$	$A \cup C$	$(A \cup B) \cap (A \cup C)$

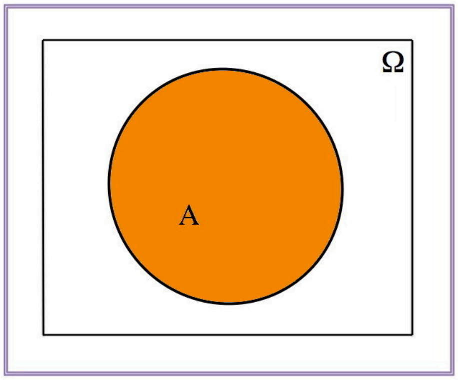


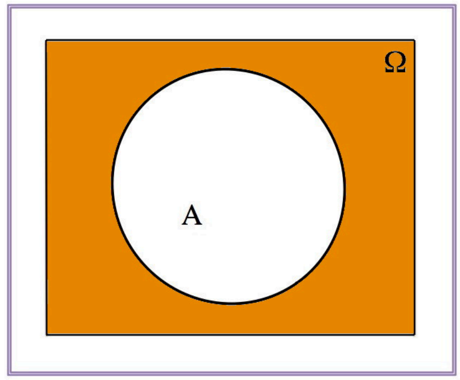
1.1.9 Propiedades

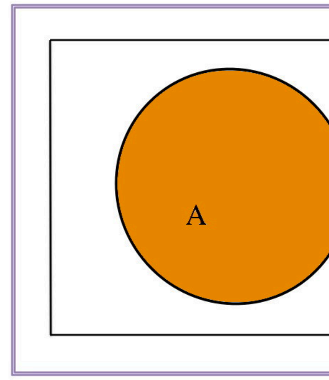
Complementario del complementario

$$(A^c)^c = A$$

A	A^c	$(A^c)^c$





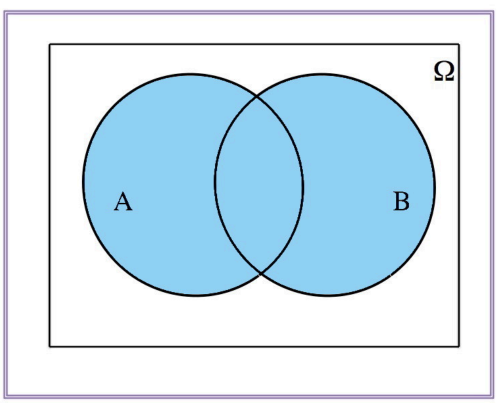


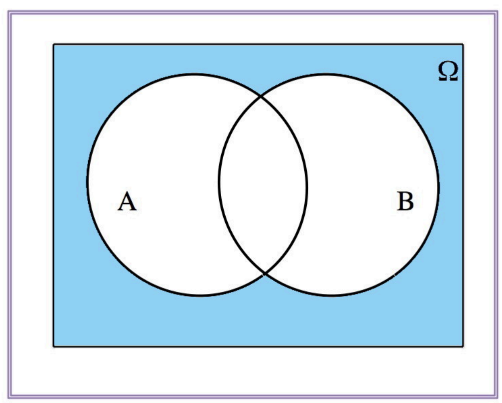
1.1.10 Propiedades

Leyes de De Morgan

$$(A \cup B)^c = A^c \cap B^c$$

$A \cup B$	$(A \cup B)^c$

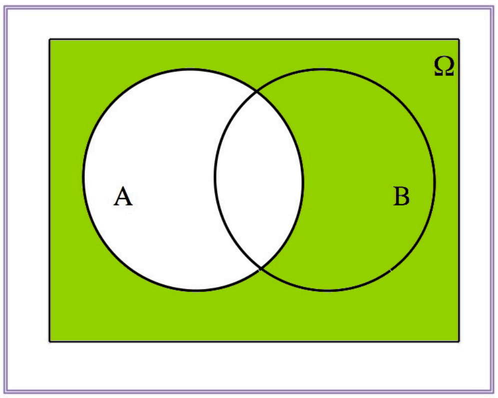
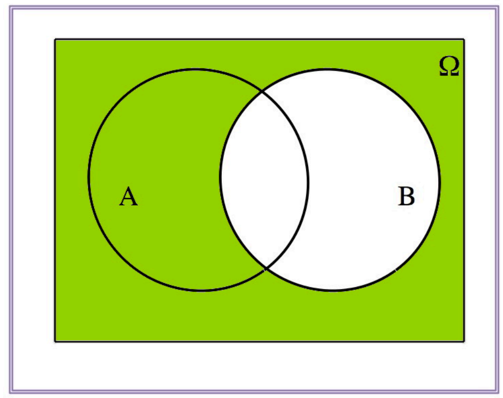




1.1.11 Propiedades

Leyes de De Morgan

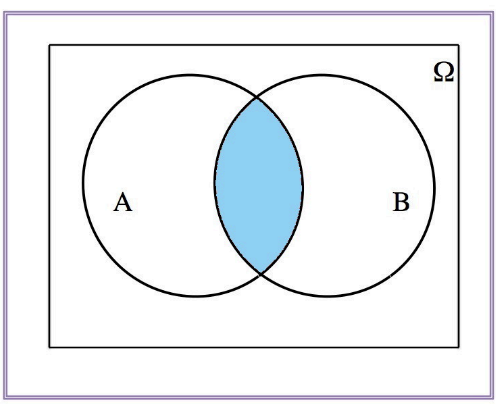
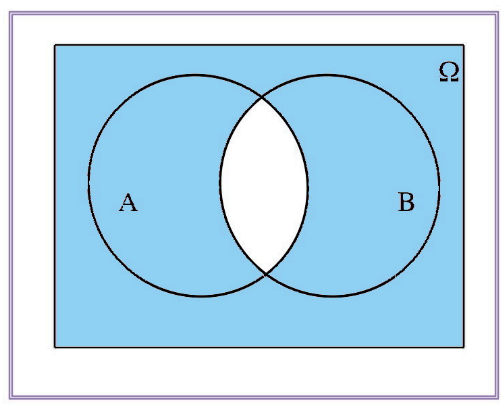
$$(A \cup B)^c = A^c \cap B^c$$

A^c	B^c	$A^c \cap B^c$
		

1.1.12 Propiedades

Leyes de De Morgan

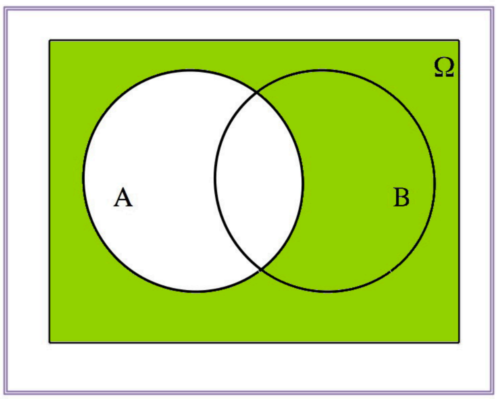
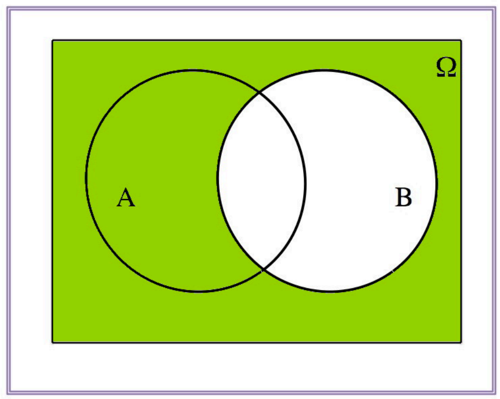
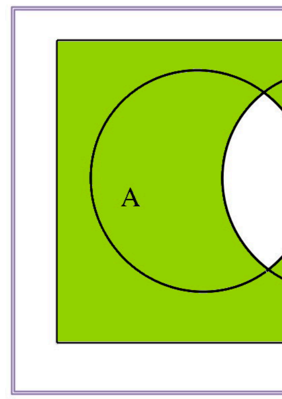
$$(A \cap B)^c = A^c \cup B^c$$

$A \cap B$	$(A \cap B)^c$
	

1.1.13 Propiedades

Leyes de De Morgan

$$(A \cap B)^c = A^c \cup B^c$$

A^c	B^c	$A^c \cup B^c$
		

1.1.14 Definición de probabilidad

La probabilidad de un suceso es una puntuación (*score*) numérico entre 0 y 1 que mide la verosimilitud de que este evento se produzca.

Esta verosimilitud puede estar justificada por

- Estimación personal
- Estimación de expertos
- La frecuencia con la que se da
- Cálculo formal

1.1.15 Definición de probabilidad

Sea Ω el espacio muestral de un experimento aleatorio. Supongamos que el número de posibles resultados, por el momento, es finito.

Una probabilidad sobre Ω es una aplicación $P : \mathcal{P}(\Omega) \to [0, 1]$ con las siguientes propiedades:

1. $0 \leq P(A) \leq 1$, para todo suceso A

2. $P(\Omega) = 1$
3. Si $\{A_1, A_2, \ldots, A_n\}$ son sucesos disjuntos dos a dos, entonces

$$P(A_1 \cup A_2 \cup \cdots \cup A_n) = P(A_1) + P(A_2) + \cdots + P(A_n)$$

Si $a \in \Omega$ es un suceso elemental cometeremos el abuso de notación de poner $P(a)$ en lugar de $P(\{a\})$

1.1.16 Ejemplo: grupos sangíneos

Ejemplo

En la página de la Fundación Banco de Sangre y Tejidos de las Islas Baleares podemos encontrar información sobre los porcentajes de tipos de sangre de los donantes de las Islas Baleares:

$$A: 46\%; B: 7.5\%; AB: 3.5\%; O: 43\%$$

¿Cuál es la probabilidad de que un balear donante de sangre no sea del tipo 0?

Experimento aleatorio: tipo de sangre de un paciente humano

$$\Omega = \{\text{A,B,AB,O}\}$$

Probabilidad de un suceso: se asimila al porcentaje observado de individuos

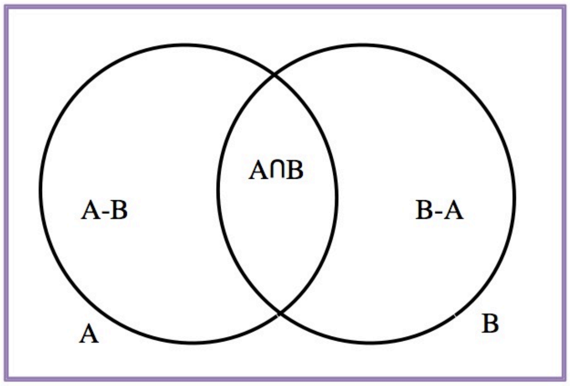
Suceso: $\{\text{O}\}^c = \{\text{A,B,AB}\}$

$$P(\{\text{O}\}^c) = P(\{\text{A,B,AB}\}) = P(\text{A}) + P(\text{B}) + P(\text{AB}) = 0.57$$

1.1.17 Propiedades

Propiedades básicas de la probabilidad

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

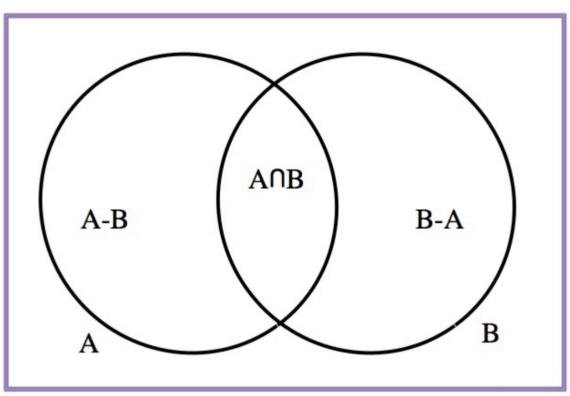


- Si $B \subseteq A$, entonces $0 \leq P(B) \leq P(A)$
- $P(A^c) = 1 - P(A)$

1.1.18 Propiedades

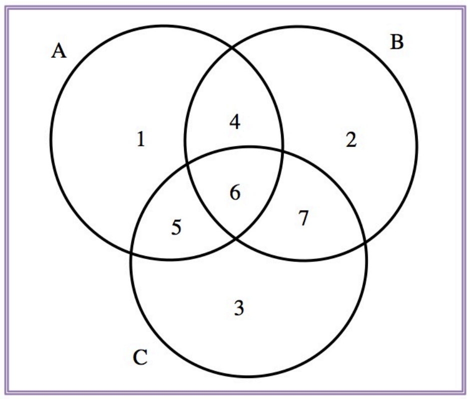
- $P(A \cup B) = P(A) + P(B) - P(A \cap B)$ porque

$$\begin{aligned}
P(A) + P(B) - P(A \cap B) &= P(A - B) + P(A \cap B) + \\
& P(B - A) + P(A \cap B) - P(A \cap B) \\
&= P(A - B) + P(A \cap B) + P(B - A) \\
&= P(A \cup B)
\end{aligned}$$



1.1.19 Propiedades

- $$\begin{aligned} P(A \cup B \cup C) &= P(A) + P(B) + P(C) \\ &\quad - P(A \cap B) - P(A \cap C) - P(B \cap C) + P(A \cap B \cap C) \end{aligned}$$



$$P(A \cup B \cup C) = P(1) + P(2) + P(3) + P(4) + P(5) + P(6) + P(7)$$

1.1.20 Propiedades

- Si $A = \{a_1, a_2, \ldots, a_k\}$, entonces
$$P(A) = P(a_1) + P(a_2) + \cdots + P(a_k)$$

- Si todos los sucesos elementales tienen la misma probabilidad,
$$P(A) = \frac{|A|}{|\Omega|} \Big(= \frac{\text{casos favorables}}{\text{casos posibles}} \Big)$$

1.1.21 Ejemplo: Frecuencia de vocales

Ejemplo

Los porcentajes de vocales de un determinado idioma (de alfabeto latino) según la Wikipedia son:

$$A : 18.7\%; E : 26.1\%; I : 25.7\%; O : 24.4\% U : 5.1\%$$

¿Cuál es la probabilidad que una vocal escogida al azar de este idioma sea una E o una O?

El espacio muestral del experimento es $\Omega = \{A, E, I, O, U\}$.

El suceso que deseamos analizar es $\{E, 0\}$.

Y su probabilidad es

$$P(\{E, O\}) = P(E) + P(O) = 0.261 + 0.244 = 0.505.$$

1.1.22 Ejemplo: Consumo de drogas

Segun un árticulo de El País, en un control especial de la policía el 0.1% de todos los conductores analizados en un control de tráfico dan positivo en un el test en cocaína, y el 1% da positivo en cannabis. Un 1.05% da positivo en alguno de los dos test.

¿Cuál es la probabilidad que un individuo analizado en el control de drogas escogido al azar no de positivo en ninguno de lo dos test?

Los sucesos elementales del enunciado del problema son:

- A: dar positivo en cocaína; $P(A) = 0.001$
- B: dar positivo en cannabis; $P(B) = 0.01$

En este caso nos interesa estudiar los sucesos:

- $A \cup B$: dar positivo en alguno de los dos test; $P(A \cup B) = 0.0105$
- $(A \cup B)^c$: no dar positivo en ninguno de los test

de donde, por tanto:

$$P((A \cup B)^c) = 1 - P(A \cup B) = 1 - 0.0105 = 0.9895.$$

1.1.23 Ejemplos: Consumo de drogas

Ejemplo

En un control especial de la policía el 0.1% de todos los conductores analizados en un control de tráfico dan positivo en un el test en cocaína, y el 1% da positivo en cannabis. Un 1.05% da positivo en alguno de los dos test.

¿Cuál es la probabilidad que un analizado al azar de positivo en los dos test en cocaína y cannabis?

Los sucesos elementales son:

- A: dar positivo en cocaína; $P(A) = 0.001$
- B: dar positivo en cannabis; $P(B) = 0.01$

En este caso nos interesa estudiar los sucesos:

- $A \cup B$: dar positivo en algún de los dos test; $P(A \cup B) = 0.0105$

- $A \cap B$: dar positivo en los dos test

de donde, por tanto:

$$\begin{aligned} P(A \cap B) &= P(A) + P(B) - P(A \cup B) \\ &= 0.001 + 0.01 - 0.0105 = 0.0005 \end{aligned}$$

1.1.24 Ejemplo: Control de drogas

Ejemplo

En un control especial de la policía el 0.1% de todos los conductores analizados en un control de tráfico dan positivo en un el test en cocaína, y el 1% da positivo en cannabis. Un 1.05% da positivo en alguno de los dos test.

¿Cuál es la probabilidad de que un conductor analizado de positivo en cocaína pero no en cannabis?

Los sucesos elementales son:

- A: dar positivo en cocaína; $P(A) = 0.001$
- B: dar positivo en cannabis; $P(B) = 0.01$

En este caso nos interesa estudiar los sucesos:

- $A \cap B$: dar positivo en los dos test; $P(A \cap B) = 0.0005$
- $B - A$: dar positivo en cocaína pero no en cannabis

de donde, por tanto:

$$P(B - A) = P(B) - P(A \cap B) = 0.01 - 0.0005 = 0.0095$$

1.2 Probabilidad condicionada

1.2.1 Probabilidad condicionada

Probabilidad condicionada: Dados dos sucesos A y B, con $P(A) > 0$, la probabilidad $P(B|A)$ de B condicionado a A es la probabilidad

- de que suceda B suponiendo que pasa A
- de que si pasa A, entonces suceda B
- de que un resultado de A también pertenezca a B

Se calcula a través de la definición:

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

1.2.2 Ejemplo: frecuencia género y gafas

Ejemplo

En una clase de 20 hombres y 30 mujeres, 15 hombres y 18 mujeres llevan gafas. Contestemos las siguientes preguntas:

- ¿Cuál es la probabilidad de que un alumno lleve gafas?

$$\frac{33}{50}$$

- ¿Cuál es la probabilidad de que un alumno sea mujer y lleve gafas?

$$\frac{18}{50}$$

1.2.3 Ejemplo: sexo y gafas

Ejemplo

En una clase de 20 hombres y 30 mujeres, 15 hombres y 18 mujeres llevan gafas. Contestemos las siguientes preguntas:

- ¿Cuál es la probabilidad de que un chica lleve gafas?

$$\frac{18}{30} = \frac{18/50}{30/50} = \frac{P(\text{mujer y gafas})}{P(\text{mujer})}$$

- Si escogemos un estudiante al azar ¿Cuál es la probabilidad que si es mujer, entonces lleve gafas?

$$\frac{18}{30}$$

1.2.4 Ejemplo

Ejemplo

En una clase de 20 hombres y 30 mujeres, 15 hombres y 18 mujeres llevan gafas. Contestemos las siguientes preguntas:

- ¿Cuál es la probabilidad de que un alumno que lleve gafas sea mujer?

$$\frac{18}{33} = \frac{18/50}{33/50} = \frac{P(\text{mujer y gafas})}{P(\text{gafas})}$$

- Si escogemos un estudiante al azar ¿Cuál es la probabilidad de que si lleva gafas, entonces sea mujer?

$$\frac{18}{33}$$

1.2.5 ¡Atención!

Hay que distinguir bien entre

- $P(A \cap B)$: probabilidad de A y B

Probabilidad de que sea mujer y lleve gafas

- $P(A|B)$: probabilidad de que si pasa B, entonces pase A.

Probabilidad de que, si es mujer, lleve gafas

Cuando utilizamos probabilidad condicional $P(A|B)$ estamos restringiendo el espacio muestral a B

1.2.6 Probabilidad condicionada. Propiedades

La probabilidad condicionada es una probabilidad

Proposición

Sea $A \subseteq \Omega$ un suceso tal que $P(A) > 0$. entonces

$$\begin{array}{rccc} P(-|A): & \mathcal{P}(\Omega) & \to & [0,1] \\ & B & \mapsto & P(B|A) \end{array}$$

satisface las propiedades de las probabilidades, como por ejemplo:

$$\begin{array}{l} P(B^c|A) = 1 - P(B|A) \\ P(B_1 \cup B_2|A) = P(B_1|A) + P(B_2|A) - P(B_1 \cap B_2|A) \end{array}$$

Ejercicio

Escribid el resto de propiedades que cumpliría una probabilidad condicionada al evento A.

1.2.7 Ejemplos

Ejemplo

Un 15% de los adultos son hipertensos, un 25% de los adultos creen que son hipertensos, y un 9% de los adultos son hipertensos y creen que lo son.

Si un adulto cree que es hipertenso, ¿cuál es la probabilidad que lo sea?

Sean los sucesos

- A: ser hipertenso; $P(A) = 0.15$
- B: creer ser hipertenso; $P(B) = 0.25$

entonces podemos definir el suceso:

- $A \cap B$: ser hipertenso y creerlo; $P(A \cap B) = 0.09$

de donde, la probabilidad condicionada de ser hipertenso creyéndonos que lo somos es:

$$P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{0.09}{0.25} = 0.36$$

1.2.8 Ejemplo

Ejemplo

Un 15% de los adultos son hipertensos, un 25% de los adultos creen que son hipertensos, y un 9% de los adultos son hipertensos y creen que lo son.

Si un adulto es hipertenso, ¿cuál es la probabilidad que crea que lo es?

Si tenemos los sucesos:

- A: ser hipertenso;
- B: creer ser hipertenso

entonces buscamos la probabilidad $P(B|A)$:

$$P(B|A) \quad = \frac{P(A \cap B)}{P(A)} = \frac{0.09}{0.15} = 0.6$$

1.2.9 Ejemplos: dígitos de control

Ejemplo

Un dígito de control de error toma el valor 0 en el 99% de los casos en que hay un error. Si la probabilidad de error en un mensaje es del 0.5%. \blue{¿cuál es la probabilidad de que el mensaje sea erróneo y el código de error tenga valor 0?

- B: mensaje con error; $P(B) = 0.005$
- A: código de error vale 0;
- $P(A|B) = 0.99$

entonces:

$$P(A \cap B) = P(B) \cdot P(A|B) = 0.005 \cdot 0.99 = 0.00495$$

1.2.10 Ejemplos

Ejemplo

Un 50% de correos recibidos en un servidor llevan adjuntos y un 65% son publicidad no deseada (SPAM). Sólo un 15% de estos correos no llevan adjuntos y no son SPAM.

- ¿Cuál es la probabilidad que un correo lleve adjunto si es SPAM?
- ¿Cuál es la probabilidad que un correo **no** tenga adjuntos si **no** es SPAM?

1.2.11 Ejemplos

Ejemplo

Un 50% de correos recibidos en un servidor llevan adjuntos y un 65% son publicidad no deseada (SPAM). Sólo un 15% de estos correos no llevan adjuntos y no son SPAM.

- ¿Cuál es la probabilidad que un correo lleve adjunto si es SPAM?

- A: llevar adjuntos; $P(A) = 0.5$
- S: SPAM; $P(S) = 0.65$
- $A^c \cap S^c = (A \cup S)^c$: no llevar adjunto y no ser SPAM; $P((A \cup S)^c) = 0.15$

$$P(A|S) = \frac{P(A \cap S)}{P(S)} =?$$

1.2.12 Ejemplos

Ejemplo

Un 50% de correos recibidos en un servidor llevan adjuntos y un 65% son publicidad no deseada (SPAM). Sólo un 15% de estos correos no llevan adjuntos y no son SPAM.

- ¿Cuál es la probabilidad que un correo lleve adjunto si es SPAM?

- $P(A) = 0.5, P(S) = 0.65, P(A^c \cap S^c) = P((A \cup S)^c) = 0.15$

- $P(A \cup S) = 1 - P((A \cup S)^c) = 0.85$

- $P(A \cap S) = P(A) + P(S) - P(A \cup S) = 0.3$

$$P(A|S) = \frac{P(A \cap S)}{P(S)} = \frac{0.3}{0.65} \approx 0.46$$

1.2.13 Ejemplos SPAM continuación

Ejemplo

Un 50% de correos recibidos en un servidor llevan adjuntos y un 65% son publicidad no deseada (SPAM). Sólo un 15% de estos correos no llevan adjuntos y no son SPAM.

- ¿Cuál es la probabilidad de que un correo no lleve adjuntos si no es SPAM?
- $P(A) = 0.5, P(S) = 0.65, P(A^c \cap S^c) = P((A \cup S)^c) = 0.15$

$$P(A^c|S^c) = \frac{P(A^c \cap S^c)}{P(S^c)} = \frac{P(A^c \cap S^c)}{1 - P(S)} = \frac{0.15}{0.35} \approx 0.43$$

1.2.14 Teorema de la probabilidad total

Teorema de la probabilidad total

Dados dos sucesos A y B se tiene que

$$\begin{aligned} P(B) &= P(B \cap A) + P(B \cap A^c) \\ &= P(A) \cdot P(B|A) + P(A^c) \cdot P(B|A^c) \end{aligned}$$

1.2.15 Teorema de la probabilidad total

Partición del espacio espacio muestral

Los sucesos $A_1, A_2, \ldots, A_n$ son una **partición** del espacio muestral Ω de un determinado experimento aleatorio, si cumplen las condiciones siguientes:

1. $A_1 \cup A_2 \cup \ldots \cup A_n = \Omega$
2. $A_1, A_2, \ldots, A_n$ son incompatibles dos a dos ($A_i \cap A_j = \emptyset$)

Teorema de la probabilidad total

Sea $A_1, A_2, \ldots, A_n$ una partición de Ω. Sea B un suceso cualquiera. Entonces

$$\begin{aligned} P(B) &= P(B \cap A_1) + \cdots + P(B \cap A_n) \\ &= P(A_1) \cdot P(B|A_1) + \ldots + P(A_n) \cdot P(B|A_n) \end{aligned}$$

1.2.16 Ejemplos

Ejemplo

Un dígito de control de error toma el valor 0 en un 99% de los casos en que hay un error y en un 5% de los mensajes sin error. La probabilidad de error en un mensaje es del 0.5%

¿Cuál es la probabilidad de que un mensaje escogido al azar tenga el dígito de control a 0?

Sean los sucesos del enunciado:

- B: mensaje con error; $P(B) = 0.005$
- A: código de error vale 0;

entonces obtenemos las probabilidades a partir del enunciado:

- $P(A|B) = 0.99$
- $P(A|B^c) = 0.05$

y por tanto,

$$\begin{aligned} P(A) &= P(B) \cdot P(A|B) + P(B^c) \cdot P(A|B^c) \\ &= 0.005 \cdot 0.99 + 0.995 \cdot 0.05 = 0.0547 \end{aligned}$$

1.2.17 Clasificación o Diagnósticos

Consideremos alguna de las siguientes situaciones:

- Un algoritmo detecta si una transacción con tarjeta de crédito es fraude o no.
- Un algoritmo detecta si tiene o no que mostrar un anuncio en una web.
- Un prueba de embarazo.
- Una prueba médica para una enfermedad concreta.

Nos ceñiremos a la casuística más elemental el algoritmo de clasificación o la diagnosis solo da dos resultado **Positivo** (sí tienes la enfermedad, sí es un fraude) o **Negativo** (en caso contrario).

1.2.18 Clasificación o Diagnósticos

En todas estas situaciones podemos calcular lo que se llama **matriz de confusión** que representa todas las situaciones posibles. En el caso de estudiar una condición de tipo binario,

	El Test da Positivo	El Test da Negativo
Condición Positiva	Correcto	Error
Condición Negativa	Error	Correcto

1.2.19 Clasificación o Diagnósticos

En general los modelos y algoritmos de clasificación suelen aportar puntuaciones (*scores*) que determinan el grado de pertenencia a una clase, o que miden si dos objetos están en la misma clase.

Así el resultado del clasificador o del diagnóstico puede ser:

- **un número real**, en cuyo caso debe clasificador entre cada clase debe determinarse por un valor umbral (*threshold*) por ejemplo para determinar si una persona está estresado podemos dar un *scores* entre 0 y 1 (1 máximo estrés 0 estrés nulo):
- **un resultado discreto** que indica directamente una de las clases (esto es necesario si es un algoritmo que debe decidir qué hacer con el objeto.

1.2.20 Clasificación o Diagnósticos

1.2.21 Clasificación o Diagnósticos

Positivos y Negativos en Clasificación

Consideremos un problema de predicción de clases binario, en la que los resultados se etiquetan positivos (P) o negativos (N). Hay cuatro posibles resultados a partir de un clasificador binario como el propuesto.

- Si el resultado de una exploración es P y el valor dado es también P, entonces se conoce como un Verdadero Positivo (VP).
- Sin embargo si el valor real es N entonces se conoce como un Falso Positivo (FP).
- De igual modo, tenemos un Verdadero Negativo (VN) cuando tanto la exploración como el valor dado son N.
- Un Falso Negativo (FN) cuando el resultado de la predicción es N pero el valor real es P.

1.2.22 Clasificación o Diagnósticos

Un ejemplo aproximado de un problema real es el siguiente: consideremos una prueba diagnóstica que persiga determinar si una persona tiene una cierta enfermedad.

- Un falso positivo en este caso ocurre cuando la prueba predice que el resultado es positivo, cuando la persona no tiene realmente la enfermedad.
- Un falso negativo, por el contrario, ocurre cuando el resultado de la prueba es negativo, sugiriendo que no tiene la enfermedad cuando realmente sí la tiene.

1.2.23 Clasificación o Diagnósticos

En un diagnósticos de una cierta condición (por ejemplo, test embarazo, test de enfermedad), tenemos dos tipos de sucesos:

- T: el test da positivo
- M: el sujeto satisface la condición

Falsos Positivos y Falsos Negativos

- **Falsos positivos** $T \cap M^c$: El test da positivo, pero la condición no es da
- **Coeficiente de falsos positivos** $P(T|M^c)$
- **Falsos negativos** $T^c \cap M$: El test da negativo, pero la condición sí que se da
- **Coeficiente de falsos negativos**: $P(T^c|M)$

1.2.24 Clasificación o Diagnósticos

Ejemplo

Un test diseñado para diagnosticar una determinada enfermedad tiene un coeficiente de falsos negativos de 0.06, y un coeficiente de falsos positivos de 0.04. En un estudio masivo se observa que un 15% de la población da positivo al test.

¿Cuál es la probabilidad que una persona escogida aleatoriamente tenga esta enfermedad?

Los datos del problema son:

- T: dar positivo al test; $P(T) = 0.15$
- M: tener la enfermedad
- $P(T) = 0.15$, $P(T^c|M) = 0.06$, $P(T|M^c) = 0.04$
- ¿$P(M)$?

1.2.25 Ejemplos

- $P(T) = 0.15$, $P(T^c|M) = 0.06$, $P(T|M^c) = 0.04$

$$P(T) = P(M) \cdot P(T|M) + P(M^c) \cdot P(T|M^c)$$

donde

$$\begin{aligned} P(T|M) &= 1 - P(T^c|M) = 0.94 \\ P(M^c) &= 1 - P(M) \end{aligned}$$

Por lo tanto

$$\begin{aligned}
0.15 &= P(M)\cdot 0.94 + (1-P(M))\cdot 0.04 \\
&= 0.04 + 0.9P(M) \\
P(M) &= \frac{0.11}{0.9} \approx 0.1222.
\end{aligned}$$

1.3 Bayes

1.3.1 Fórmula de Bayes

Teorema de Bayes

Sean A y B dos sucesos. Si $P(B) > 0$, entonces

$$\begin{aligned}
P(A|B) &= \frac{P(A)\cdot P(B\big|A)}{P(B)} \\
&= \frac{P(A)\cdot P(B\big|A)}{P(A)\cdot P(B\big|A) + P(A^c)\cdot P(B\big|A^c)}
\end{aligned}$$

Ejercicio

Demostrar el teorema de Bayes utilizando que

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

1.3.2 Fórmula de Bayes

Teorema de Bayes

Sea $A_1, A_2, \ldots, A_n$ una partición de Ω. Sea B un suceso tal que $P(B) > 0$. entonces(para cualquier $i = 1, 2, \ldots, n$):

$$\begin{aligned}
P(A_i|B) &= \frac{P(A_i)\cdot P(B|A_i)}{P(B)} \\
&= \frac{P(A_i)\cdot P(B|A_i)}{P(A_1)\cdot P(B|A_1) + \cdots + P(A_n)\cdot P(B|A_n)}
\end{aligned}$$

Ejercicio

Demostrar el teorema de Bayes utilizando que

$$P(A_i|B) = \frac{P(A_i \cap B)}{P(B)} = \cdots$$

1.3.3 Ejemplos

Ejemplo

Un test para detección de VIH da positivo un 99% de los casos en los que está presente y en un 5% de los casos en los que el virus está ausente. En una población con un 0.5% de infectados por VIH, ¿cuál es la probabilidad que un individuo que haya dado positivo en el test esté infectado?

Los sucesos del ejemplo son:

- A: individuo infectado
- B: el test da positivo

de donde podemos calcular:

$$\begin{aligned} P(A|B) &= \frac{P(B|A) \cdot P(A)}{P(B|A) \cdot P(A) + P(B|A^c) \cdot P(A^c)} \\ &= \frac{0.99 \cdot 0.005}{0.005 \cdot 0.99 + 0.995 \cdot 0.05} = 0.09 \end{aligned}$$

1.3.4 Ejemplos

Ejemplo

Un test para detección de VIH da positivo un 99% de los casos en los que está presente y en un 5% de los casos en los que el virus está ausente. En una población con un 0.5% de infectados por VIH, ¿cuál es la probabilidad de que un individuo que haya dado **negativo** en el test **no** esté infectado?

Los sucesos del ejemplo son:

- A: individuo infectado
- B: el test da positivo

de donde podemos calcular:

$$\begin{aligned} P(A^c|B^c) &= \frac{P(B^c|A^c) \cdot P(A^c)}{P(B^c|A) \cdot P(A) + P(B^c|A^c) \cdot P(A^c)} \\ &= \frac{0.95 \cdot 0.995}{0.01 \cdot 0.005 + 0.95 \cdot 0.995} = 0.999947 \end{aligned}$$

1.3.5 Ejemplos

Ejercicio

Se ha observado que los cientes de una empresa de ventas por internet son de tres tipos, A, B y C, disjuntos dos a dos. La probabilidad que ser de cualquiera de cada uno de los tipos es 1/3, pero la probabilidad de compra de cada tipo es diferente: si es de tipo A compra un 50% de las veces, si de tipo B, un 75% de las veces, y de tipo C, un 60%.

Supongamos que llega un cliente ¿cuál es la probabilidad de que si ha comprado sea del tipo B?

- Los sucesos del ejercicio son A: el cliente es de tipo A; B: el cliente es de tipo B; C: el cliente es de tipo C;

$$P(A) = P(B) = P(C) = 1/3$$

Buscamos estrudiar el suceso E: el cliente compra

$$P(E|A) = 0.5, P(E|B) = 0.75, P(E|C) = 0.6$$

$$P(B|E) = \frac{P(E|B) \cdot P(B)}{P(E|A) \cdot P(A) + P(E|B) \cdot P(B) + P(E|C) \cdot P(C)} = \ldots$$

1.3.6 Ejemplos

Ejercicio

Un test de detección precoz de abandono de clientes de una empresa de telefonía da positivo el 97.5% de las ocasiones en las que, posteriormente, el cliente se da de baja, y un 12% de las veces en que no se dio de baja. La probabilidad que un cliente escogido al azar se dé de baja es de un 2%.

- ¿Cuál es la probabilidad que un individuo escogido al azar de positivo en el test?
- ¿Cuál es la probabilidad que un individuo escogido al azar se de de baja y dé positivo en el test?
- ¿Cuál es la probabilidad que un individuo que dé negativo en el test se dé de baja?

Definimos los sucesos y datos del ejercicio:

- T: Dar positivo al test
- B: darse de baja; $P(B) = 0.02$
- $P(T|B) = 0.975, P(T|B^c) = 0.12$

1.3.7 Ejemplos

$$P(B) = 0.02, P(T|B) = 0.975, P(T|B^c) = 0.12$$

- ¿Cuál es la probabilidad que un individuo escogido al azar de positivo en el test?

$$\begin{aligned} P(T) &= P(B) \cdot P(T|B) + P(B^c) \cdot P(T|B^c) \\ &= 0.02 \cdot 0.975 + 0.98 \cdot 0.12 = 0.1371 \end{aligned}$$

- ¿Cuál es la probabilidad que un individuo escogido al azar se de de baja y dé positivo en el test?

$$P(B \cap T) = P(B) \cdot P(T|B) = 0.02 \cdot 0.975 = 0.0195$$

1.3.8 Ejemplos

$$P(B) = 0.02, P(T|B) = 0.975, P(T|B^c) = 0.12$$

- ¿Cuál es la probabilidad que un individuo que dé negativo en el test se dé de baja?

$$\begin{aligned} P(B|T^c) &= \frac{P(B \cap T^c)}{P(T^c)} = \frac{P(B) - P(B \cap T)}{1 - P(T)} \\ &= \frac{0.02 - 0.0195}{1 - 0.1371} \approx 0.00058 \end{aligned}$$

- O también se obtiene así

$$P(B|T^c) = \frac{P(T^c|B) \cdot P(B)}{P(T^c|B) \cdot P(B) + P(T^c|B^c) \cdot P(B^c)}$$

donde

$$\begin{aligned} P(T^c|B) &= 1 - P(T|B) = 0.025, \\ P(T^c|B^c) &= 1 - P(T|B^c) = 0.88 \end{aligned}$$

1.4 Independencia de sucesos

1.4.1 Sucesos independientes

Sucesos Independientes

Diremos que los sucesos A y B son **independientes** si $P(A \cap B) = P(A) \cdot P(B)$

$A_1, \ldots, A_n$ son sucesos **independientes** cuando, para toda subfamilia $A_{i_1}, \ldots, A_{i_k}$,

$$P(A_{i_1} \cap \cdots \cap A_{i_k}) = P(A_{i_1}) \cdots P(A_{i_k})$$

Proposición Dados dos sucesos A y B con $P(A), P(B) > 0$, las siguientes afirmaciones son equivalentes:

1. A y B son independientes
2. $P(A|B) = P(A)$
3. $P(B|A) = P(B)$

1.4.2 Sucesos independientes

Proposición

Las siguientes afirmaciones son equivalentes:

1. A y B son independientes.
2. A^c y B son independientes.
3. A y B^c son independientes.
4. A^c y B^c son independientes.

1.4.3 Ejemplo billete avión

Ejemplo

En la web de viajes WEBTravel, el 55% de los clientes compra billete de avión, el 20% alojamiento en hotel, y el 60% billete de avión o alojamiento en hotel. ¿Son los sucesos comprar billete de avión y comprar alojamiento en hotel independientes?

Los sucesos y datos del ejemplo son:

- A: comprar billete de avión; $P(A) = 0.55$
- B: comprar alojamiento; $P(B) = 0.2$

por tanto, podemos calcular las probabilidades siguientes

$$\begin{array}{rcl} P(A \cap B) & = & P(A) + P(B) - P(A \cup B) \\ & = & 0.55 + 0.2 - 0.6 = 0.15 \\ P(A) \cdot P(B) & = & 0.55 \cdot 0.2 = 0.11 \end{array}$$

Por tanto, concluimos que son dependientes, ya que $P(A \cap B) \neq P(A) \cdot P(B)$.

1.4.4 Sucesos independientes vs disjuntos

Ejercicio

1. Dos sucesos A y B disjuntos, ¿son necesariamente independientes?
2. Dos sucesos A y B independientes, ¿son necesariamente disjuntos?
3. $\emptyset$ y un suceso cualquiera A, ¿son necesariamente independientes?
4. Ω y un suceso cualquiera A, ¿son necesariamente independientes?
5. ¿Qué condiciones se tienen que dar para que un suceso A sea independiente de si mismo?

Chapter 2

Variables Aleatorias

2.1 Introducción a las variables aleatorias

- Hasta ahora nuestros sucesos han sido de varios tipos: $\{C, +\}$ en la moneda, nombres de periódicos, ángulos en una ruleta, número de veces que sale cara en el lanzamiento de una moneda etc….
- Necesitamos estandarizar de alguna manera todos estos sucesos. Una solución es asignar a cada suceso un cierto conjunto de números reales, es decir, convertir todos los sucesos en *sucesos de números reales* para trabajar con ellos de forma unificada.
- Para conseguirlo utilizaremos unas funciones que transformen los elementos del espacio muestral en números; esta funciones son las variables aleatorias.

2.1.1 Definición de variable aleatoria

Comenzaremos dando una definición poco rigurosa, pero suficiente, de variable aleatoria.

Variable Aleatoria (definición práctica)

Una variable aleatoria (v.a.) es una aplicación que toma valores numéricos determinados por el resultado de un experimento aleatorio

Notación:

- Normalmente representaremos las v.a. por letras mayúsculas X, Y, Z…
- Los valores que "*toman*" las v.a. los representaremos por letras minúsculas (las mismas en principio) $x, y, z \ldots$

2.1.2 Ejemplo

Ejemplo

Lanzamos un dado convencional de parchís el espacio muestral del experimento es

$$\Omega = \{1, 2, 3, 4, 5, 6\}$$

Una v.a $X : \Omega \to \mathbb{R}$ sobre este espacio queda definida por

$$\begin{aligned} X(1) = 1, X(2) = 2, X(3) = 3, \\ X(4) = 4, X(5) = 5, X(6) = 6. \end{aligned}$$

- Ahora el suceso $A = \{2, 4, 6\}$, es decir "salir número par", es equivalente a $\{X = 2, X = 4, X = 6\}$.
- El suceso $B = \{1, 2, 3\}$, es decir "salir un número inferior o igual a 3" es en términos de la v.a. $\{X = 1, X = 2, X = 3\}$ o también $\{X \leq 3\}$.

2.1.3 Ejemplo

Ejemplo

Consideremos el experimento lanzar una anilla al cuello de una botella. Si acertamos a ensartar la anilla en la botella el resultado del experimento es *éxito* y *fracaso* en caso contrario.

El espacio muestral asociado a este experimento será $\Omega = \{\text{éxito, fracaso}\}$. Construyamos la siguiente variable aleatoria:

$$X : \{\text{éxito, fracaso}\} \to \mathbb{R}$$

definida por

$$X(\text{éxito}) = 1 \text{ y } X(\text{fracaso}) = 0.$$

2.1.4 Tipos de variables aleatorias

Hay dos tipos fundamentales de variables aleatorias, las discretas y las continuas.

Damos a continuación una definición informal.

Variables Aleatórias Discretas y Continuas

- Una variable aleatoria es **discreta** si sólo puede tomar una cantidad numerable de valores con probabilidad positiva.
- Las variables aleatorias **continuas** toman valores en intervalos.
- También existen las variables aleatorias **mixtas**; con una parte discreta y otra continua.

2.1.5 Ejemplo

Ejemplo

Son variables *aleatorias discretas*:

- Número de artículos defectuosos en un cargamento.
- Número de clientes que llegan a una ventanilla de un banco en una hora.
- Número de errores detectados en las cuentas de una compañía.
- Número de reclamaciones de una póliza de un seguro médico.

Son variables *aleatorias continuas*:

- Renta anual de una familia.
- Cantidad de petróleo importado por un país
- Variación del precio de las acciones de una compañía de telecomunicaciones.
- Porcentaje de impurezas en un lote de productos químicos.

2.2 Variables aleatorias discretas

2.2.1 Distribuciones de probabilidad para v.a. discretas.

- Pasamos ahora a describir el comportamiento de la v.a. Para ello utilizaremos distintas funciones que nos darán algunas probabilidades de la variable aleatoria.
- En el caso discreto estas funciones son la de probabilidad, y la función de distribución o de probabilidad acumulada.
- En el caso discreto la función de probabilidad es la que nos da las probabilidades de los sucesos elementales de la v.a. que definimos a continuación.

2.2.2 Función de probabilidad para variables discretas

Función de Probabilidad La *función de probabilidad* (*probability mass function* o incluso abusando de notación *probability density function*)

de una variable aleatoria discreta X a la que denotaremos por $P_X(x)$ está definida por

$$P_X(x) = P(X = x)$$

es decir la probabilidad de que X tome el valor x.

Si X no asume ese valor x, entonces $P_X(x) = 0$.

2.2.3 Funcion probabilidad discretas

Dominio de una variable aleatoria discreta

El conjunto

$$D_X = \{x \in \mathbb{R} \mid P_X(x) > 0\}$$

recibe el nombre de *dominio* de la v.a. y son los valores posibles de esta variable.

En el caso discreto lo más habitual es que $X(\Omega) = D_X$.

2.2.4 Ejemplo

Ejemplo parchís

Lanzamos un dado de parchís una vez, en esta ocasión representaremos los sucesos elementales por el número de puntos de la cara obtenida, tenemos que

$$\Omega = \{\text{1-puntos,2-puntos,3-puntos,4-puntos,5-puntos,6-puntos}\}$$

y la variables aleatoria $X : \Omega \to \mathbb{R}$ viene definida por

$$X(\text{i-puntos}) = i \text{ para } i = 1, 2, 3, 4, 5, 6.$$

Supongamos que el dado está bien balanceado. Entonces

$$P_X(1) = P_X(2) = P_X(3) = P_X(4) = P_X(5) = P_X(6) = \frac{1}{6}$$

Concretamente:

$$P_X(x) = \begin{cases} \frac{1}{6} & \text{si } x = 1, 2, 3, 4, 5, 6 \\ 0 & \text{en otro caso} \end{cases}$$

Su dominio es

$$D_X = \{1, 2, 3, 4, 5, 6\}.$$

2.2.5 Ejemplo

Ejemplo lanzamiento moneda

Sea X la v.a. asociada al lanzamiento de una moneda. Su espacio muestral es $\Omega = \{c, +\}$, la v.a. queda definida por:

$$X(\omega) = \begin{cases} 1 & \text{si } \omega = c \\ 0 & \text{si } \omega = + \end{cases}$$

Su función de probabilidad es:

$$P_X(x) = P(X = x) = \begin{cases} \frac{1}{2} & \text{si } x = 0, 1 \\ 0 & \text{en otro caso} \end{cases}$$

Finalmente su dominio es $D_X = \{0, 1\}$.

2.2.6 Ejemplo

Ejemplo urna con bolas

Tenemos una urna con tres bolas rojas, una negra y dos blancas. Realizamos una extracción y observamos el color de la bola entonces un espacio muestral es

$$\Omega = \{roja, blanca, negra\}.$$

Una variable aleatoria asociada al experimento es:

$$X(\omega) = \begin{cases} 1 & \text{si } \omega = roja \\ 2 & \text{si } \omega = negra \\ 3 & \text{si } \omega = blanca \end{cases}$$

2.2.7 Ejemplo

La función de probabilidad es

$$P_X(x) = \begin{cases} \frac{3}{6} & \text{si } x = 1 \\ \frac{1}{6} & \text{si } x = 2 \\ \frac{2}{6} & \text{si } x = 3 \\ 0 & \text{en otro caso} \end{cases}$$

El dominio de la v.a. X es $D_X = \{1, 2, 3\}$.

2.2.8 Propiedades de la función de probabilidad.

Propiedades básicas de la función de porbabilidad

Sea X una v.a. discreta $X:\Omega :\to \mathbb{R}$ con dominio D_X. Su función de probabilidad P_X verifica las siguientes propiedades:

- $0\leq P_X(x)\leq 1$ para todo $x\in\mathbb{R}$
- $\sum\limits_{x\in D_X} P_X(x)=1$

2.2.9 Ejemplo

Ejemplo: Lanzamiento moneda

Lanzamos al aire tres veces, de forma independiente, una moneda perfecta. El espacio muestral de este experimento es

$$\Omega=\{ccc, cc+, c+c, +cc, c++, +c+, ++c, +++\}$$

(expresados en orden de aparición).

Este espacio tiene todos los sucesos elementales equiprobables.

Consideremos la variable aleatoria asociada a este experimento:

$$X = \text{ número de caras en los tres lanzamientos.}$$

Su función de probabilidad es:

$$\begin{array}{l}
P(X=0)=P(\{+++\})=\frac{1}{8}\\
P(X=1)=P(\{c++, +c+, ++c\})=\frac{3}{8}\\
P(X=2)=P(\{cc+, c+c, +cc\})=\frac{3}{8}\\
P(X=3)=P(\{ccc\})=\frac{1}{8}
\end{array}$$

2.2.10 Ejemplo

Podemos reescribir la función de probabilidad de X de forma simplificada:

$$P_X(x)=\begin{cases}\frac{1}{8} & \text{si } x=0,3\\ \frac{3}{8} & \text{si } x=1,2\\ 0 & \text{en otro caso}\end{cases}$$

Efectivamente los valores de la función de distribución suman 1

$$\sum_{x=0}^{3} P_X(x)=\frac{1}{8}+\frac{3}{8}+\frac{3}{8}+\frac{1}{8}=1.$$

2.2.11 Función de distribución de variables aleatorias

Distribución de Probabilidad

La función de *distribución de probabilidad* (acumulada) de la v.a. X (de cualquier tipo; discreta o continua) $F_X(x)$ representa la probabilidad de que X tome un menor o igual que x es decir

$$F_X(x) = P(X \leq x)$$

Esta función también se denomina función de *distribución de probabilidad o simplemente función de distribución* de una v.a., y en inglés *cumulative distribution function* por lo que se abrevia con el acrónimo `cdf`.

2.2.12 Propiedades

Propiedades de la Función de Distribución

Sea X una v.a. y F_X su función de distribución:

1. $P(X > x) = 1 - P(X \leq x) = 1 - F_X(x)$
2. Sea a y b tales que $a < b$, $P(a < X \leq b) = P(X \leq b) - P(X \leq a) = F_X(b) - F_X(a)$

2.2.13 Propiedades

Demostración:

Tenemos que el complementario de X mayor que x es: $\overline{\{X > x\}} = \{X > x\}^c = \{X \leq x\}$. Además

$$P(X > x) = 1 - P(\overline{\{X > x\}}) = 1 - P(X \leq x) = 1 - F_X(x)$$

lo que demuestra la primera propiedad

Por otro lado, que X se encuentre entre dos valores a y b es $\{a < X \leq b\} = \{X \leq b\} - \{X \leq a\}$

ahora podemos hacer

$$\begin{aligned} P(a < X \leq b) &= P(\{X \leq b\} - \{X \leq a\}) \\ &= P(\{X \leq b\}) - P(\{X \leq a\}) \\ &= F_X(b) - F_X(a). \end{aligned}$$

2.2.14 Propiedades

Propiedades de la Función de Distribución

Sea F_X la función de distribución de una v.a. X entonces:

- $0 \leq F_X(x) \leq 1$.
- La función F_X es no decreciente.
- La función F_X es continua por la derecha.
- Si denotamos por $F_X(x_0^-) = \lim_{x \to x_0^-} F(x)$, entonces se cumple que $P(X < x_0) = F_X(x_0^-)$ y que $P(X = x_0) = F_X(x_0) - F_X(x_0^-)$.

2.2.15 Propiedades

Propiedades de la Función de Distribución

- Se cumple que $\lim_{x \to \infty} F_X(x) = 1$; $\lim_{x \to -\infty} F_X(x) = 0$.
- Toda función F verificando las propiedades anteriores es función de distribución de alguna v.a. X.
- $P(X > x) = 1 - F_X(x)$
- Dados $a, b \in \mathbb{R}$ con $a < b$

$$P(a < X \leq b) = F_X(b) - F_X(a).$$

2.2.16 Advertencia desigualdades estrictas

En las propiedades anteriores no se pueden cambiar en general las desigualdades de estrictas o no estrictas.

Veamos que propiedades tenemos cuando se cambian estas desigualdades.

Dada una F_X una función de distribución de la v.a. X y denotamos por

$$F_X(x_0^-) = \lim_{x \to x_0^-} F_X(x),$$

,

entonces se cumplen las siguientes igualdades...

2.2.17 Advertencia desigualdades estrictas

Propiedades

- $P(X = x) = F_X(x) - F_X(x^-)$
- $P(a < X < b) = F_X(b^-) - F_X(a)$
- $P(a \leq X < b) = F_X(b^-) - F_X(a^-)$

- $P(X < a) = F_X(a^-)$
- $P(a \leq X \leq b) = F_X(b) - F_X(a^-)$
- $P(X \geq a) = 1 - F_X(a^-)$

2.2.18 Propiedades

Más propiedades de la función de distribución

- Si F_X es continua en x se tiene que $P(X = x) = 0$. Así que si la v.a. es continua $P(X \leq a) = P(X < a) + P(X = a) = P(X < a)$ y propiedades similares.
- Sea X una variable aleatoria discreta que con dominio D_X y que tiene por función de probabilidad $P_X(x)$ entonces su función de distribución $F_X(x_0)$ es
$$F_X(x_0) = \sum_{x \leq x_0} P_X(x)$$

donde $\sum_{x \leq x_0}$ indica que sumamos todos los $x \in D_X$ tales que $x \leq x_0$

2.2.19 Propiedades

Demostración:

Si X es continua

$$P(X = a) = F(a) - F(a^-) = F(a) - F(a) = 0$$

por lo tanto

$$P(X \leq a) = P(X < a) + P(X = a) = P(X < a) + 0 = P(X < a).$$

lo que demuestra la primera propiedad.

Para demostrar la segunda basta hacer

$$F_X(x_0) = P(X \leq x_0) = P\left(\bigcup_{x \leq x_0; x \in D_X} \{x\}\right) = \sum_{x \leq x_0} P(X = x) = \sum_{x \leq x_0} P_X(x).$$

2.2.20 Ejemplo

Ejemplo

En el experimento del dado se tiene que:

$$P_X(x) = \begin{cases} \frac{1}{6} & \text{si } x = 1,2,3,4,5,6 \\ 0 & \text{en el resto de casos} \end{cases},$$

por lo tanto

$$F_X(x) = P(X \leq x) = \begin{cases} 0 & \text{si } x < 1 \\ \frac{1}{6} & \text{si } 1 \leq x < 2 \\ \frac{2}{6} & \text{si } 2 \leq x < 3 \\ \frac{3}{6} & \text{si } 3 \leq x < 4 \\ \frac{4}{6} & \text{si } 4 \leq x < 5 \\ \frac{5}{6} & \text{si } 5 \leq x < 6 \\ 1 & \text{si } 6 \leq x \end{cases}$$

2.2.21 Ejemplo

Calculemos más detalladamente algún valor de F_X, por ejemplo:

$$\begin{aligned} F_X(3.5) &= P(X \leq 3.5) = P(\{X=1\} \cup \{X=2\} \cup \{X=3\}) \\ &= P(\{X=1\}) + P(\{X=2\}) + P(\{X=3\}) \\ &= \frac{1}{6} + \frac{1}{6} + \frac{1}{6} = \frac{3}{6} = \frac{1}{2}, \end{aligned}$$

o de otra forma

$$\begin{aligned} F_X(3.5) &= \sum_{x \leq 3.5} P_X(x) = \sum_{x=1}^{3} P(X=x) \\ &= \sum_{x=1}^{3} \frac{1}{6} = 3 \cdot \frac{1}{6} = \frac{1}{2}. \end{aligned}$$

2.2.22 Propiedades de la función de distribución

Propiedad

Sea X una variable con función de distribución F_X entonces:

- $0 \leq F_X(x) \leq 1$ para todo x
- Si $x < x'$ entonces
$$F_X(x) \leq F_X(x').$$
Es una función creciente, no necesariamente estrictamente creciente.

- $\lim\limits_{x\to -\infty} F_X(x) = 0$ y $\lim\limits_{x\to +\infty} F_X(x) = 1$
- Es continua por la derecha $\lim\limits_{x\to x_0^+} F_X(x) = F_X(x_0)$.

2.3 Valores esperados o esperanza

2.3.1 Momentos de variables aleatorias discretas

- Al igual que en la estadística descriptiva se utilizan distintas medidas para resumir los valores centrales y para medir la dispersión de una muestra, podemos definir las correspondiente medidas para variables aleatorias.
- A estas medidas se les suele añadir el adjetivo *poblacionales* mientras que a las que provienen de la muestra se las adjetiva como *muestrales*.

Por ejemplo podemos buscar un valor que resuma toda la variable. Este valor es el que "*esperamos*" que se resuma la v.a. o esperamos que las realizaciones de la v.a. queden cerca de él. Veamos su definición formal.

2.3.2 Esperanza de un variable aleatoria discreta

Esperanza de una variable aleatoria discreta

El valor *esperado o esperanza* (*expected value* en inglés) $E(X)$ de una v.a. discreta X, se define como

$$E(X) = \sum_{x\in X(\Omega)} x P_X(x)$$

En ocasiones se le domina *media* (*mean* en inglés *mitjana* en catalán) poblacional o simplemente media y muy frecuentemente se la denota $\mu_X = E(X)$ o simplemente $\mu = E(X)$.

2.3.3 Interpretación de la media aritmética para v.a. discretas

Ejemplo

Supongamos que lanzamos un dado n veces y obtenemos unas frecuencias absolutas n_i para el resultado i con $i = 1, \ldots, 6$. Sea X la v.a. que nos representa el valor de una tirada del dado.

Calculemos la media aritmética (o media muestral) de los datos

$$\overline{x} = \frac{1 \cdot n_1 + 2 \cdot n_2 + 3 \cdot n_3 + 4 \cdot n_4 + 5 \cdot n_5 + 6 \cdot n_6}{n} = \sum_{x=1}^{6} x \frac{n_x}{n}.$$

Si $n \to \infty$ se tiene que $\lim\limits_{n\to\infty} \frac{n_x}{n} = P_X(x)$.

Por lo tanto $E(X) = \lim\limits_{n\to\infty} \sum_{x=1}^{6} x \frac{n_x}{n}$.

Entonces el valor esperado en una v.a. discreta puede entenderse como el valor promedio que tomaría una v.a. en un número grande de repeticiones.

2.3.4 Ejemplo

Ejemplo: Erratas en un texto

Sea X= número de erratas en una página de un texto con dominio $D_X = \{0, 1, 2\}$.

Resulta que

$$P(X = 0) = 0.42, P(X = 1) = 0.4, P(X = 2) = 0.18.$$

entonces

$$E(X) = 0 \cdot 0.42 + 1 \cdot 0.4 + 2 \cdot 0.18 = 0.76.$$

Elegida una página del texto al azar esperamos encontrar 0.76 errores por página.

Supongamos que en el editor nos paga 2 euros por cada página que encontremos con 1 error y 3 euros por cada página con dos errores (y nada por las páginas correctas) ¿Cuánto esperamos cobrar si analizamos una página?

$$0 \cdot 0.42 + 2 \cdot 0.4 + 3 \cdot 0.18 = 1.34$$

2.3.5 Esperanzas de funciones de variables aleatorias discretas

Esperanzas de funciones de variables aleatorias discretas

Sea X una v.a. discreta con función de probabilidad P_X y de distribución F_X. Entonces el *valor esperado de una función* $g(x)$ es :

$$E(g(X)) = \sum_{x} g(x) P_X(x).$$

2.3.6 Propiedades de los valores esperados

Propiedades

- $E(k) = k$ para cualquier constante k.
- Si $a \leq X \leq b$ entonces $a \leq E(X) \leq b$.
- Si X es una v.a. discreta que toma valores enteros no negativos entonces $E(X) = \sum_{x=0}^{+\infty}(1 - F_X(x))$.

Ejercicio

La demostración de las propiedades anteriores se deja como ejercicio.

2.3.7 Ejemplo

Ejemplo: paleta de colores aleatoria

Supongamos que estamos sentados delante de nuestro ordenador con un amigo y le decimos que en dos minutos podemos programar una paleta para poner colores a unos gráficos.

Queremos la que paleta tenga dos botones con las opciones color rojo y color azul. Como hemos programado a gran velocidad resulta que el programa tiene un error; cada vez que se abre la paleta los colores se colocan al azar (con igual probabilidad) en cada botón, así que no sabemos en que color hemos de pinchar.

Además, como nos sobraron 15 segundos para hacer el programa y pensando en la comodidad del usuario, la paleta se cierra después de haber seleccionado un color y hay que volverla a abrir de nuevo.

La pregunta es ¿cuál es el valor esperado del número de veces que hemos pinchar el botón de color azul antes de obtener este color?

2.3.8 Ejemplo

Llamemos X al número de veces que pinchamos en el botón azul (y nos sale rojo) hasta obtener el primer azul. La variable X toma valores en los enteros no negativos. Su función de probabilidad queda determinada por

$$P_X(x) = P(X = x) = P(\overbrace{rojo, rojo, \ldots, rojo}^{x\text{veces}}, azul) = \left(\frac{1}{2}\right)^{x+1}.$$

2.3.9 Propiedades de las series geométricas

Series geométricas

- Una *progresión geométrica* de razón r es una sucesión de la forma

$$r^0, r^1, \ldots, r^n, \ldots.$$

- La serie geométrica es la suma de todos los valores de la progresión geométrica $\displaystyle\sum_{k=0}^{+\infty} r^k$.
- Las sumas parciales desde el término n_0 al n de una progresión geométrica valen

$$\sum_{k=n_0}^{n} r^k = \frac{r^{n_0} - r^n r}{1-r}.$$

2.3.10 Propiedades de las series geométricas

Propiedades

- Si $|r| < 1$ la serie geométrica es convergente y

$$\sum_{k=0}^{+\infty} r^k = \frac{1}{1-r}$$

.

- En el caso en que se comience en n_0 se tiene que

$$\sum_{k=n_0}^{+\infty} r^k = \frac{r^{n_0}}{1-r}.$$

2.3.11 Propiedades de las series geométricas

Propiedades

- Si $|r| < 1$ también son convergentes las derivadas, respecto de r, de la serie geométrica y convergen a la derivada correspondiente. Así tenemos que

$$\begin{aligned}
\left(\sum_{k=0}^{+\infty} r^k\right)' &= \sum_{k=1}^{+\infty} k r^{k-1} \left(\frac{1}{1-r}\right)' = \frac{1}{(1-r)^2} \\
\left(\sum_{k=0}^{+\infty} r^k\right)'' &= \sum_{k=2}^{+\infty} k(k-1) r^{k-2} \left(\frac{1}{1-r}\right)'' = \frac{2}{(1-r)^3}
\end{aligned}$$

.

2.3.12 Ejemplo

Ejemplo (cont)

Si seguimos con el ejemplo de la paleta de colores, su esperanza es:

$$\begin{aligned} E(X) &= \sum_{x=0}^{+\infty} xP(X=x) = \sum_{x=0}^{+\infty} x\left(\frac{1}{2}\right)^{x+1} \\ &= \left(\frac{1}{2}\right)^2 \sum_{x=1}^{+\infty} x\left(\frac{1}{2}\right)^{x-1} = \left(\frac{1}{2}\right)^2 \frac{1}{\left(1-\frac{1}{2}\right)^2} = 1. \end{aligned}$$

Ahora calculemos su función de distribución

$$\begin{aligned} F_X(x) &= P(X \leq x) = \sum_{k=0}^{x} P(X=k) = \sum_{k=0}^{x} \left(\frac{1}{2}\right)^{k+1} \\ &= \frac{\frac{1}{2} - \frac{1}{2}^{x+1}\frac{1}{2}}{1-\frac{1}{2}} = 1 - \left(\frac{1}{2}\right)^{x+1} \end{aligned}$$

.

2.3.13 Ejemplo

Como la variable toma valores enteros positivos, podemos calcular su valor esperado de esta otra manera

$$E(X) = \sum_{x=0}^{+\infty} (1 - F_X(x)) = \sum_{x=0}^{+\infty} (\frac{1}{2})^{x+1} = \frac{1}{2}\frac{1}{1-\frac{1}{2}} = 1.$$

Ejercicio

Calculad el valor esperado de la variable

$$Y = \text{número de intentos para conseguir el color azul.}$$

2.3.14 Momentos de una variable aleatoria

Momentos de orden m

Llamaremos *momento de orden* m respecto al punto C a

$$E\left((X-C)^m\right)$$

- Cuando $C = 0$ los momentos reciben el nombre de *momentos respecto al origen*.
- Cuando $C = E(X)$ reciben el nombre de *momentos centrales o respecto de la media*. Luego la esperanza es el momento de orden 1 respecto al origen. Estos momentos son la versión poblacional de los momentos que vimos en el curso de estadística descriptiva, recibiendo estos último el nombre de momentos muestrales.

2.3.15 Resumen conceptos

- Hemos descrito el comportamiento aleatorio de una v.a. discreta mediante sus funciones de probabilidad P_X y de distribución F_X.
- También tenemos un valor central; el valor esperado $E(X)$.
- Como medida básica nos queda definir una medida de lo lejos que están los datos del valor central $E(X)$ una de estas medidas es la varianza de X.

2.4 Medidas de la variabilidad

2.4.1 Medidas de la variabilidad: la varianza

Varianza

Sea X una v.a. Llamaremos *varianza* de X a

$$Var(X) = E((X - E(X))^2)$$

Por lo tanto la varianza es el momento central de orden 2.

De forma frecuente se utiliza la notación

$$\sigma_X^2 = Var(X).$$

A la raíz cuadrada positiva de la varianza

$$\sigma_X = \sqrt{Var(X)}$$

se la denomina desviación típica o estándar de X.

2.4.2 Propiedades de la varianza

Propiedad

- Si X es una v.a. discreta con función de probabilidad P_X su varianza es

$$\sigma_X^2 = Var(X) = E((X - E(X))^2) = \sum_x (x - E(X))^2 P_X(x).$$

- Sea X una v.a.

$$Var(X) = E(X^2) - (E(X))^2 = \sum_x x^2 P_X(X) - (E(X))^2$$

2.4.3 Demostración

Demostración de b)

$$\begin{aligned} Var(X) &= \sum_x (x - E(X))^2 P_X(x) = \sum_x (x^2 - 2xE(X) + (E(X)^2) P_X(x) \\ &= \sum_x x^2 P_X(x) - E(X) \sum_x 2x P_X(x) + (E(X)^2) \sum_x P_X(x) \\ &= E(X^2) - 2E(X)E(X) + (E(X))^2 = E(X^2) - (E(X))^2. \end{aligned}$$

2.4.4 Ejemplo

Ejemplo

Calculemos en el ejemplo anterior la varianza del número de errores.

Recordemos que:

$$P(X = 0) = 0.42, \quad P(X = 1) = 0.4, \quad P(X = 2) = 0.18$$

y que

$$E(X) = 0.76$$

Entonces:

$$Var(X) = E(X^2) - (E(X))^2 = E(X^2) - (0.76)^2.$$

2.4.5 Ejemplo

Ahora necesitamos calcular

$$E(X^2) = 0^2(0.41) + 1^2(0.4) + 2^2(0.18) = 0.4 + 0.72 = 1.12$$

y por lo tanto

$$Var(X) = E(X^2) - (0.76)^2 = 1.12 - 0.5776 = 0.542$$

y

$$\sqrt{Var(X)} = \sqrt{0.542}$$

En resumen $\sigma_X^2 = 0.542$ y $\sigma_X = \sqrt{0.542}$

2.4.6 Propiedades de la varianza

Propiedades de la varianza

- $Var(X) \geq 0$
- $Var(cte) = E(cte^2) - (E(cte))^2 = cte^2 - cte^2 = 0$
- El mínimo de $E((X - C)^2)$ se alcanza cuando $C = E(X)$ y es $Var(X)$. Esta propiedad es una de las que hace útil a la varianza como medida de dispersión.

Ejercicio

Se deja como ejercicio la demostración de estas propiedades.

2.5 Esperanza y varianza de transformaciones lineales.

2.5.1 Transformaciones lineales.

Transformación lineal

Un **cambio de variable lineal** o **transformación lineal** de una v.a. X es otra v.a. $Y = a + bX$ donde $a, b \in \mathbb{R}$.

Esperanza de una transformación lineal

Sea X una v.a. con $E(X) = \mu_X$ y $Var(X) = \sigma_X^2$ y $a, b \in \mathbb{R}$. Entonces si $Y = a + bX$:

- $E(Y) = E(a + bX) = a + bE(X) = a + b\mu_X$.
- $Var(Y) = Var(a + bX) = b^2 Var(X) = b^2\sigma_X^2$

- $\sigma_Y = \sqrt{Var(Y)} = \sqrt{b^2 Var(X)} = |b|\sigma_X$

2.5.2 Demostración

Demostración:

$$\begin{aligned}
E(Y) &= E(a+bX) = \sum_x (a + b\cdot X)\cdot P_X(x) \\
&= a\sum_x P_X(x) + b\sum_x x\cdot P_X(x) \\
&= a + b\cdot E(X) = a + b\mu_X
\end{aligned}$$

Ejercicio

Las demostración de las demás propiedades se dejan como ejercicio.

2.6 Variables aleatorias continuas

2.6.1 Variables aleatorias continuas definición.

- Como ya hemos dicho las variables aleatorias continuas toman valores en intervalos o áreas.
- Lo más habitual es que estas variables tengan función de distribución continua y derivable (salvo a los más en una cantidad finita o numerable de puntos:-)).
- En lo que sigue supondremos que la función de distribución de variables aleatorias continuas cumplen estas propiedades.
- Notemos que si X es una v.a. con función de distribución continua se tiene que $P(X = x_0) = F_X(x_0) - F(x_0^-) = 0$. Por lo que no tiene sentido definir *función de probabilidad*.

2.6.2 Variables aleatorias continuas

- En general tendremos que $P(X < x_0) = P(X \leq x_0)$.
- Por otra parte podemos utilizar una regla parecida del cociente entre casos favorables y casos posibles de Laplace pero en este caso el conteo se hace por la *medida* de los casos posibles partida por la *medida* de los casos favorables.
- Veamos un ejemplo de v.a. continua, que ampliaremos en el tema siguiente, en el que se utilizan todos estos conceptos.

2.6.3 Ejemplo: Distribución uniforme en $[0, 1]$.

Ejemplo: distancia el dardo centro

Supongamos que lanzamos un dardo a una diana de radio 1, de forma que sea *equiprobable* cualquier distancia al centro (¡Cuidado! esto no es equivalente a que cualquier punto de la diana sea *equiprobable*).

Consideremos la v.a. continua X = distancia al centro.

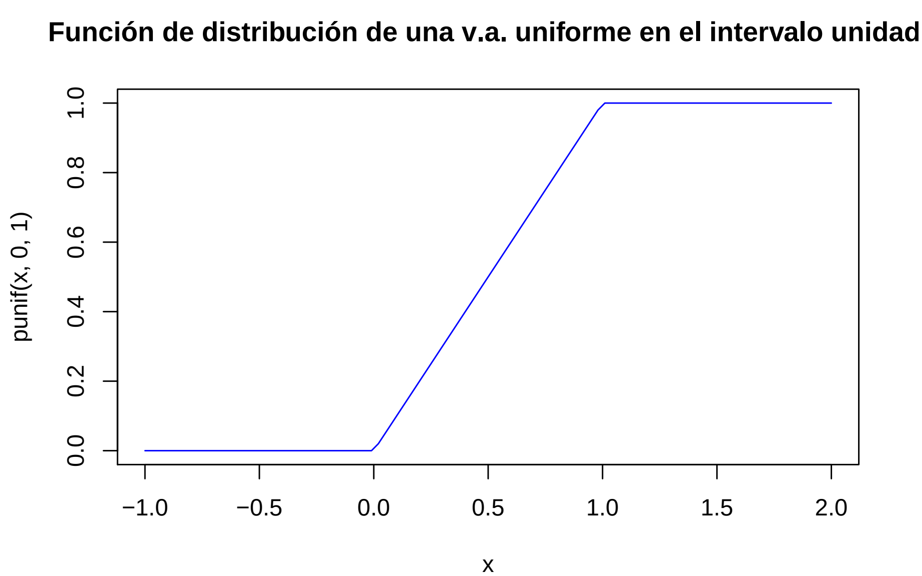
Su función de distribución es

$$F_X(x) = \begin{cases} 0 & \text{si } x \leq 0 \\ x & \text{si } 0 < x < 1 \\ 1 & \text{si } x \geq 1 \end{cases} .$$

Ya que

- C.F. *longitud favorable* es $x - 0$.
- C.P. *longitud posible* es $1 - 0$.
- Luego
$$P(X \leq x) = \frac{C.F.}{C.P.} = \frac{x-0}{1-0} = x$$

2.6.4 Gráfica de la función de distribución uniforme



2.6.5 Propiedades

En las variables continuas los sucesos del tipo $\{X \leq x\}$ y $\{X < x\}$ tendrán la misma probabilidad, y otros tipos de sucesos similares también, algunas de estas propiedades se explicitan en la siguiente proposición.

Propiedades

Dada una v.a. continua X se tiene que:

- $P(X \leq b) = P(X < b)$
- $P(X < b) = P(X < a) + P(a < X < b)$
- $P(a < X < b) = P(X < b) - P(X < a)$

2.6.6 Demostración

Demostración:

La primera es evidente $P(X \leq b) = P(X < b) + P(X = b) = P(X < b)$

Para demostrar la segunda, tenemos

$$\{X < a\} \cap \{a < X < b\} = \emptyset$$
$$\{X < a\} \cup \{a < X < b\} = \{X < b\}$$

entonces

$$\begin{aligned} P(X \leq b) = & \quad P(\{X < a\} \cup \{a < X < b\}) \\ & = P(X < a) + P(a < X < b) \end{aligned}$$

Ejercicio

La demostración de la tercera propiedad es similar a la segunda pero aplicando la primera. Queda de ejercicio.

2.6.7 Propiedades de la función de distribución

Las propiedades anteriores y combinaciones de ellas se pueden escribir utilizando la función de distribución de X:

Propiedades de la Función de Distribución

Dada una variable aleatoria continua se tiene que:

- $F_X(b) = F_X(a) + P(a < X < b)$
- $P(a < X < b) = F_X(b) - F_X(a)$

- $P(a \leq X \leq b) = F_X(b) - F_X(a)$

Ejercicio

Se deja la demostración como ejercicio

2.6.8 Ejemplo

Ejemplo: lanzamiento de dardos

En los dardos:

$$P(0.25 < X < 0.3) = F_X(0.3) - F_X(0.25) =$$

$$= 0.3 - 0.25 = 0.05$$

2.6.9 Función de densidad

Función de densidad

Una función $f : \mathbb{R} \to \mathbb{R}$ es una función de densidad sobre $\mathbb{R}$ si cumple que

- $f_X(x) \geq 0$ para todo $x \in \mathbb{R}$.
- f es continua salvo a lo más en una cantidad finita de puntos sobre cada intervalo acotado de $\mathbb{R}$.
- $\displaystyle\int_{-\infty}^{+\infty} f_X(x)dx = 1.$

2.6.10 Función de densidad de una variable aleatoria.

Función de densidad de una variable aleatoria

Sea X una v.a. con función de distribución F_X. Sea $f : \mathbb{R} \to \mathbb{R}$ una función de densidad tal que

$$F_X(x) = \int_{-\infty}^{x} f_X(t)dt. \text{ para todo } x \in \mathbb{R}.$$

Entonces X es una variable aleatoria continua y f_X es la densidad de la v.a. X.

2.6.11 Dominio de una variable aleatoria continua

El conjunto $D_X = \{x \in \mathbb{R} | f_x(x) > 0\}$ recibe el nombre de soporte o dominio de la variable aleatoria continua y se interpreta su conjunto de resultados posibles.

Ejemplo

En nuestra diana, la función f es una densidad

$$f_X(x) = \begin{cases} 0 & \text{si } x \leq 0 \\ 1 & \text{si } 0 < x < 1 \\ 0 & \text{si } 1 \leq x \end{cases}$$

que es la densidad de X, en efecto:

2.6.12 Densidad diana

$$f_X(x) = \begin{cases} 0 & \text{si } x \leq 0 \\ 1 & \text{si } 0 < x < 1 \\ 0 & \text{si } 1 \leq x \end{cases}$$

Si $x \leq 0$ entonces $\displaystyle\int_{-\infty}^{x} f_X(t)dt = 0$.

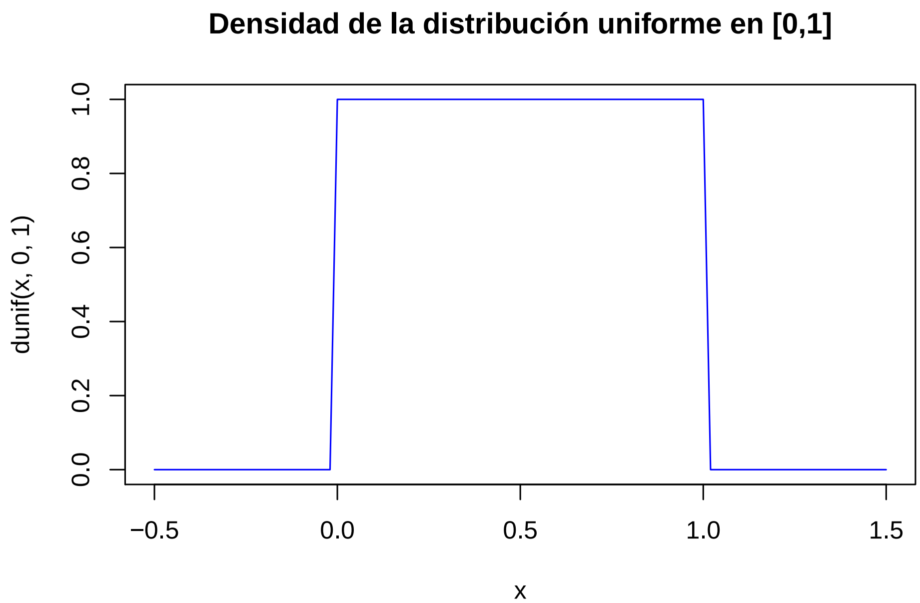
Si $0 \leq x \leq 1$ entonces $\displaystyle\int_{-\infty}^{x} f_X(t)dt = \int_0^x 1dt = x$.

Si $x \geq 1$ entonces $\displaystyle\int_{-\infty}^{x} f_X(t)dt = \int_0^1 1dt = 1$.

Por lo tanto $F_X(x) = \displaystyle\int_{-\infty}^{x} f_X(t)dt$ para todo $x \in \mathbb{R}$.

2.6.13 Densidad diana

```
curve(dunif(x,0,1),xlim=c(-0.5,1.5),col="blue",
      main="Densidad de la distribución uniforme en [0,1]")
```



2.6.14 Utilidad de la función de densidad

La función de densidad nos permite calcular diversas probabilidades.

Propiedades de la función de densidad

Sea X una v.a. continua con función de distribución F_X y de densidad f_X, entonces

-
$$\begin{aligned} P(a < X < b) &= P(a < X \leq b) = P(a \leq X < b) = \\ &P(a \leq X \leq b) = \int_a^b f_X(x)dx. \end{aligned}$$

- Si A es un conjunto adecuado de $\mathbb{R}$ entonces

$$P(X \in A) = \int_A f(x)dx = \int_{A \cap D_X} f(x)dx.$$

2.6.15 Utilidad de la función de densidad

Propiedades de la función de densidad

Sea X una v.a. continua con función de distribución F_X y de densidad f_X, entonces:

- Si f_x es continua en un punto x, F_X es derivable en ese punto y $F'_X(x) = f_X(x)$.
- $P(X = x) = 0$ para todo $x \in \mathbb{R}$.

Ejercicio

Comprobar estas propiedades en el ejemplo de la diana.

2.6.16 Ejemplo tiempo ejecución de un proceso

Ejemplo: Tiempo ejecución de un proceso.

Sea X = tiempo de ejecución de un proceso. Se supone que X sigue una distribución uniforme en dos unidades de tiempo, si tarda más el proceso se cancela.

Calculemos la función de densidad y de distribución de la v.a X.

Entonces

$$F_X(x) = P(X \leq x) = \frac{CF}{CP} = \frac{x}{2}$$

Luego su función de distribución es:

$$F_X(x) = \begin{cases} 0 & \text{si } x \leq 0 \\ \frac{x}{2} & \text{si } 0 < x < 2 \\ 1 & \text{si } 2 \leq x \end{cases}$$

2.6.17 Ejemplo tiempo ejecución de un proceso

Su función de densidad por su lado es:

$$f_X(x) = F'_X(x) = \begin{cases} 0 & \text{si } x \leq 0 \\ \frac{1}{2} & \text{si } 0 < x \leq 2 \\ 0 & \text{si } 2 \leq x \end{cases}$$

Efectivamente

- $f_X(x) \geq 0$, y tiene un conjunto finito de discontinuidades.
- $F_X(x) = \int_{-\infty}^{x} f_X(t)dt$. para todo $x \in \mathbb{R}$ (Ejercicio, resolverlo gráficamente.)
- $\int\limits_{-\infty}^{+\infty} f_X(x)dx = \int\limits_{0}^{2} \frac{1}{2}dx = \frac{x}{2} \mid_0^2 == \frac{2}{2} - \frac{0}{2} = 1.$

2.6.18 Ejemplo tiempo ejecución de un proceso

Ejercicio: Tiempo de un proceso:

Calcular la probabilidad de que uno de nuestros procesos tarde más de una unidad de tiempo en ser procesado. Calcular también la probabilidad de que dure entre 0.5 y 1.5 unidades de tiempo.

2.7 Esperanza y varianza para variables aleatorias continuas

2.7.1 Esperanza y varianza para variables aleatorias continuas

Los mismos comentarios y definiciones que se dieron en la sección correspondiente del tema de estadística descriptiva son aplicables aquí.

Así que sólo daremos las definiciones, la forma de cálculo y algunos ejemplos.

En lo que sigue, salvo que diagamos lo contrario, X es una v.a. continua con función de densidad $f_X(x)$

2.7.2 Esperanza y varianza para variables aleatorias continuas

Esperanza v.a. continuas

- Su esperanza es:

$$E(X) = \int_{-\infty}^{+\infty} x f_X(x) dx.$$

- Si $g(x)$ es una función de la variable X entonces:

$$E(g(X)) = \int_{-\infty}^{+\infty} g(x) \cdot f_X(x) dx.$$

2.7.3 Esperanza y varianza para variables aleatorias continuas

Varianza v.a. continuas

- Su varianza es:

$$Var(X) = \sigma_X^2 = E((X - \mu_X)^2) = \int_{-\infty}^{+\infty} (x - \mu_X)^2 f_X(x) dx.$$

- Su desviación típica es:

$$\sigma_X = +\sqrt{\sigma_X^2}.$$

2.7.4 Propiedades

Propiedades

- $\sigma_X^2 \geq 0$
- $Var(cte) = E(cte^2) - (E(cte))^2 = cte^2 - cte^2 = 0$
- $Var(x) = E(X^2) - \mu_X^2 = \int_{-\infty}^{+\infty} x^2 f_X(x) dx - \mu_X^2.$
- El mínimo de $E((X - C)^2)$ se alcanza cuando $C = E(X)$ y es $Var(X)$.

2.7.5 Ejemplo

Ejemplo: Dardo

Calcular μ_X y σ_X^2 en el dardo.

Resultado

$$\mu_X = \frac{1}{2},$$
$$E(X^2) = \frac{1}{3},$$
$$Var(X) = \frac{1}{12}.$$

2.7.6 Esperanza de trasformaciones lineales de v.a. continuas

Proposición

Sea X una v.a. continua con $E(X) = \mu_X$ y $Var(X) = \sigma_X^2$ sea $Y = a+bX$, donde $a, b \in \mathbb{R}$, es una nueva v.a. continua obtenida mediante una transformación lineal de X. Se verifican las mismas propiedades que en el caso discreto:

- $E(Y) = E(a + bX) = a + bE(X)$
- $Var(Y) = Var(a + bX) = b^2 Var(X)$
- $\sigma_Y = |b|\sigma_X$
- $Z = \frac{X - \mu_X}{\sigma_X}$ es una transformación lineal de X de forma que

$$E(Z) = 0 \text{ y } Var(Z) = 1$$

2.7.7 Ejemplo

Ejemplo

En una empresa de venta de vinos por internet, sea X = número de litros de vino del país vendidos en un año. Supongamos que sabemos que $E(X) = 10000$ y que $Var(X) = 100$ Supongamos que los gastos fijos de distribución son 50.000 € y el beneficio por litro es de 10 € por botella. Definimos $T = 10X - 50000$ que será el beneficio después de gastos.

Entonces la esperanza del beneficio es

$$E(T) = 10E(X) - 50000 = 50000$$

y

$$Var(T) = 10^2 Var(X) = 10000$$

2.8 Transformaciones de variables aleatorias

2.8.1 Transformaciones de variables aleatorias

Muchas variables aleatorias son funciones de otras v.a. En lo que sigue resumiremos diversas técnicas para dada una v.a. X y una transformación $Y = h(X)$ encontrar F_Y a partir de F_X.

2.8.2 Propiedad

Tranformación de v.a. discretas

Sea X una v.a. discreta con $X(\Omega) = \{x_1, x_2, \ldots, x_n, ..\}$ y sea $h : \mathbb{R} \to \mathbb{R}$ una aplicación. Entonces $Y = h(X)$ es también una v.a. discreta. Además si P_X y F_X son las funciones de probabilidad y de distribución de X entonces

- $P_Y(y) = \displaystyle\sum_{x_i | h(x_i) = y} P_X(x_i).$
- $F_Y(y) = \displaystyle\sum_{x_i | h(x_i) \leq y} P_X(x_i).$

2.8.3 Propiedades

Desafortunadamente para variables no discetas el asunto no es tan sencillo como el anterior, pues la transformación de, por ejemplo, una v.a. continua puede ser continua, discreta, mixta…

Transformación de v.a. continuas en continuas

Sea X una v.a. continua cuya función de densidad es f_X. Sea $h : \mathbb{R} \to \mathbb{R}$ una aplicación estrictamente monótona y derivable, tal que $h'(x) \neq 0$ para todo $x \in \mathbb{R}$. Sea $Y = h(X)$ la transformación de X por h. Entonces Y es una v.a. continua con función de densidad

$$f_Y(y) = \left. \frac{f_X(x)}{|h'(x)|} \right|_{x=h^{-1}(y)}$$

2.8.4 Propiedades

Densidad de una transformación de una v.a. continua

Sea X una v.a. continua cuya función de densidad es f_X. Sea

$$h : \mathbb{R} \to \mathbb{R}$$

una aplicación, no necesariamente monótona tal que :

- sea derivable con derivada no nula
- la ecuación $h(x) = y$ tiene un número finito de soluciones $x_1, x_2, .., x_n$

entonces:

$$f_Y(y) = \sum_{k=1}^{n} \left. \frac{f_X(x)}{|h'(x)|} \right|_{x=x_k}.$$

2.8.5 Método general del transformación de v.a.

Cuando no podamos aplicar las propiedades anteriores intentaremos calcular primero la función de distribución de la transformación y luego su densidad.

Notemos que en general si $Y = g(X)$ es una v.a. transformación de la v.a. X entonces

$$F_Y(y) = P(Y \leq y) = P(g(X) \leq y).$$

2.8.6 Método general del transformación de variables aleatorias

Por ejemplo si g es estrictamente creciente y cont.

$$F_Y(y) = P(g(X) \leq y) = P(X \leq g^{-1}(y)) = F_X(g^{-1}(y)).$$

y si g es estrictamente decreciente y cont.

$$F_Y(y) = P(g(X) \leq y) = P(X \geq g^{-1}(y)) = 1 - F_X(g^{-1}(y)).$$

2.9 Desigualdades básicas: Markov y Chebychev

2.9.1 Desigualdades de Markov y de Chebychev

- En esta sección distintas desigualdades que acotan determinadas probabilidades de una variable aleatoria.
- Estas desigualdades sirven en algunos casos para acotar probabilidades de determinados sucesos.
- También son útiles desde el punto de vista teórico, por ejemplo para justificar que la varianza es una mediada de la dispersión de los datos.

2.9.2 Desigualdad de Markov

Desigualdad de Markov

Sea X una v.a. positiva con $E(X)$ finita. Entonces

$$P(X \geq a) \leq \frac{E(X)}{a} \text{ para todo } a > 0.$$

Demostración:

Si X es continua y solo toma valores positivos

$$\begin{aligned} E(X) &= \int_{-\infty}^{+\infty} x f_X(x) dx = \int_0^{+\infty} x f_X(x) dx = \int_0^a x f_X(x) dx + \int_a^{+\infty} x f_X(x) dx \\ &\geq \int_a^{+\infty} x f_X(x) dx \geq a \int_a^{+\infty} f_X(x) dx = a \cdot P(X \geq a) \end{aligned}$$

.

de donde se sigue que

$$P(X \geq a) \leq \frac{E(X)}{a}.$$

2.9.3 Desigualdad de Markov

Corolario

Sea X una v.a. con $E(X)$ finita entonces para todo $a > 0$

$$P(|X| \geq a) \leq \frac{E(|X|)}{a}.$$

Ejercicio

Demuestra el corolario anterior a partir de la desigualdad de Markov.

2.9.4 Desigualdad de Chebychev

La desigualdad de Chebychev también se denomina de Chebyshov y en inglés Chebyshev.

Desigualdad de Chebychev

Sea X una v.a.con $E(X) = \mu$ y $Var(X) = \sigma^2$ entonces para todo $a > 0$

$$P(|X - \mu| \geq a) \leq \frac{\sigma^2}{a^2}$$

2.9.5 Demostración desigualdad de Chebychev

Demostración

Apliquemos la consecuencia de la desigualdad de Markov a la v.a. no negativa

$$Y^2 = (X - \mu)^2$$

entonces

$$P(Y^2 \geq a^2) \leq \frac{E(Y^2)}{a^2} = \frac{E((X - \mu)^2)}{a^2} = \frac{Var(X)}{a^2} = \frac{\sigma^2}{a^2}.$$

Por otra parte

$$P(Y^2 \geq a^2) = P(|Y| \geq a) = P(|X - \mu| \geq a).$$

hecho que, junto con la desigualdad anterior, demuestra el resultado.

2.9.6 Uso de la desigualdad de Chebychev

Utilidad básica de la desigualdad de Chebychev

Supongamos que X es una v.a. con $Var(X) = 0$ entonces.

Aplicando la desigualdad anterior

$$P(|X - E(X)| \geq a) = 0 \text{ para todo } a > 0$$

lo que implica que

$$P(X = E(X)) = 1.$$

Por lo que probabilidad de que X sea constantemente $E(X)$ es 1.

Lo que nos confirma la utilidad de la varianza es una medida de la dispersión de los datos.

2.9.7 Ejemplo

Ejemplo

Se sabe que el tiempo de respuesta medio y la desviación típica de un sistema multiusuario son 15 y 3 u.t.respectivamente. Entonces:

$$P(|X - 15| \geq 5) \leq \frac{9}{25} = 0.36.$$

Si substituimos a por $a \cdot \sigma$ en la desigualdad de Chebychev, nos queda:

$$P(|X - \mu| \geq a\sigma) \leq \frac{\sigma^2}{(a\sigma)^2} = \frac{1}{a^2}.$$

Que es otra manera de expresar la desigualdad de Chebychev.

2.9.8 Más formas de la desgualdad de Chebychev

La desigualdad de Chebychev también se puede escribir de al menos dos maneras más:

$$P(\mu - a \leq X \leq \mu + a) \geq 1 - \frac{\sigma^2}{a^2}$$

y tomado como $a = k \cdot \sigma$

$$P(\mu - k \cdot \sigma \leq X \leq \mu + k \cdot \sigma) \geq 1 - \frac{1}{k^2}$$

2.9.9 La varianza como medida de dispersión

Tomando la segunda expresión que hemos visto para la desigualdad de Chebychev para distintos valores de $k > 0$ tenemos la siguiente tabla.

k	$P(\|X - E(X)\| \geq k \cdot \sigma)$
1	≤ 1
2	≤ 0.25
3	≤ 0.111
4	≤ 0.0025

2.9.10 Interpretación de la desigualdad

- Por ejemplo para $k = 2$ esta desigualdad se puede interpretar como que dada una v.a. X con cualquier distribución que tenga $E(X)$ y $Var(X)$ finitos *la probabilidad de que un valor se aleje de la media* μ *más de* $a = 2$ *desviaciones típicas es menor o igual que* 0.25.
- Es decir sólo el 25% de los valores estarán alejados de la media más de 2σ

¡*Sea cual sea la distribución de la v.a.*!

Chapter 3

Distribuciones Notables

- En este tema estudiaremos diversos tipos de experimentos que son muy frecuentes y algunas de las variables aleatorias asociadas a ellos.
- Estas variables reciben distintos nombres que aplicaremos sin distinción al tipo de población del experimento a la variable o a su función de probabilidad, densidad o distribución.
- Empezaremos con las variables aleatorias discretas que se presentan con frecuencia ya que están relacionadas con situaciones muy comunes como el número de caras en varios lanzamiento de una moneda, el número de veces que una maquina funciona hasta que se estropea, el numero de clientes en una cola,…

3.1 Distribución Bernoulli

Distribución Bernoulli

- Consideremos un experimento con dos resultados posibles éxito (E) y fracaso (F). El espacio de sucesos será $\Omega=\{E,F\}$.
- Supongamos que la probabilidad de éxito es $P(E)=p$, y naturalmente $P(F)=1-p=q$ con $0<p<1$.
- Consideremos la aplicación

$$X:\Omega=\{E,F\}\to\mathbb{R}$$

definida por

$$X(E)=1,\, X(F)=0.$$

Su función de probabilidad es

$$P_X(x) = \begin{cases} 1-p=q & \text{si } x=0 \\ p & \text{si } x=1 \\ 0 & \text{en cualquier otro caso} \end{cases}.$$

Su función de distribución es

$$F_X(x) = P(X \leq x) = \begin{cases} 0 & \text{si } x<0 \\ 1-p=q & \text{si } 0 \leq x < 1 \\ 1 & \text{si } 1 \leq x \end{cases}.$$

- Bajo estas condiciones diremos que X **es una v.a. Bernoulli** o que sigue una ley de **distribución de probabilidad Bernoulli** de parámetro p.
- Lo denotaremos por

$$X \equiv Ber(p) \text{ o también } X \equiv B(1,p).$$

- A este tipo de experimentos (éxito/fracaso)se les denomina experimentos Bernoulli.
- Fue su descubridor un científico suizo Jacob Bernoulli, uno más de la de la conocida familia de científicos suizos Bernoulli

3.1.1 Esperanza de una v.a. $X\ Ber(p)$

Su **valor esperado** es

$$E(X) = \sum_{x=0}^{1} x \cdot P(X=x) = 0 \cdot (1-p) + 1 \cdot p = p.$$

Calculemos también $E(X^2)$

$$E(X^2) = \sum_{x=0}^{1} x^2 \cdot P(X=x) = 0^2 \cdot (1-p) + 1^2 \cdot p = p.$$

3.1.2 Varianza de una v.a. $X\ Ber(p)$

Su **varianza** es

$$Var(X) = E(X^2) - (E(X))^2 = p - p^2 = p \cdot (1-p) = p \cdot q.$$

Su desviación típica es

$$\sqrt{Var(X)} = \sqrt{p \cdot (1-p)}.$$

3.1.3 Resumen v.a con distribución Bernoulli

X Bernoulli	$Ber(p)$
$D_X =$	$\{0,1\}$
$P_X(x) = P(X=x) =$	$\begin{cases} q & \text{si } x=0 \\ p & \text{si } x=1 \\ 0 & \text{en otro caso} \end{cases}$
$F_X(x) = P(X \leq X) =$	$\begin{cases} 0 & \text{si } x<0 \\ q & \text{si } 0 \leq x < 1 \\ 1 & \text{si } 1 \leq x \end{cases}$
$E(X) = p$	$Var(X) = p \cdot q$

3.1.4 Distribución Bernoulli. Ejemplo

Veamos los cálculos básicos $Ber(p=0.25)$ en `R`.

```
dbinom(0,size=1,prob=0.25)
```

```
## [1] 0.75
```

```
dbinom(1,size=1,prob=0.25)
```

```
## [1] 0.25
```

```
rbinom(n=20,size = 1,prob=0.25)
```

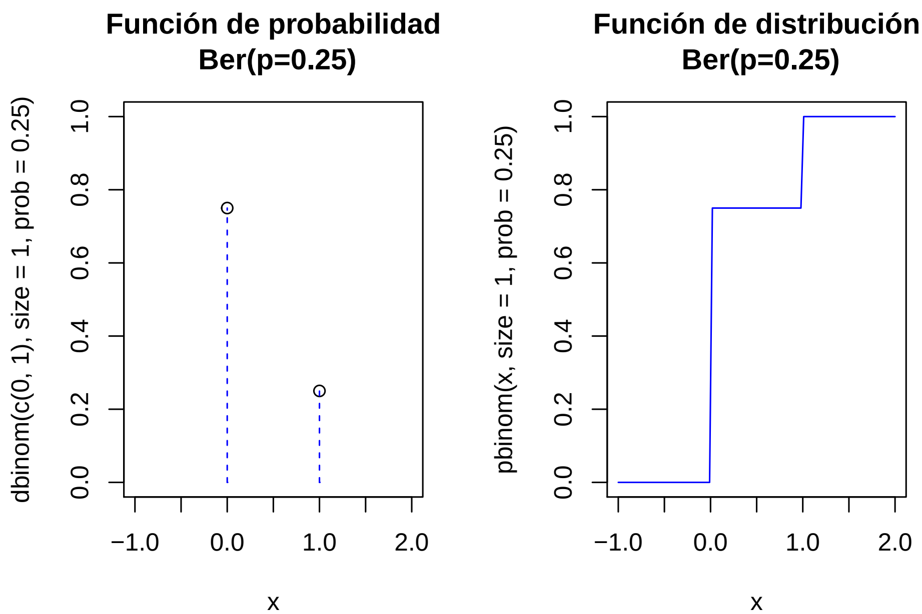
```
##  [1] 0 0 0 0 0 0 0 0 0 0 0 0 1 1 0 0 1 1 0 0
```

3.1.5 Distribución Bernoulli. Ejemplo

El siguiente código dibuja las función de probabilidad y la de distribución de una $Ber(p=0.25)$

```
par(mfrow=c(1,2))
plot(x=c(0,1),y=dbinom(c(0,1),size=1,prob=0.25),
     ylim=c(0,1),xlim=c(-1,2),xlab="x",
     main="Función de probabilidad\n Ber(p=0.25)")
lines(x=c(0,0,1,1),y=c(0,0.75,0,0.25), type = "h", lty = 2,col="blue")
curve(pbinom(x,size=1,prob=0.25),
      xlim=c(-1,2),col="blue",
      main="Función de distribución\n Ber(p=0.25)")
par(mfrow=c(1,1))
```

3.1.6 Distribución Bernoulli. Ejemplo



3.1.7 Gráficas interactivas $Ber(p)$

Para ejecutar el siguiente gráfico interactivo, solamente tienes que cargar el paquete `shiny` en tu ordenador y luego copiar/pegar las siguientes instrucciones. De este modo podrás observar los cambios en las distribuciones variando los parámetros.

```
sliderInput("p_ber", label = "Probabilidad éxito p:",
              min = 0.01, max = 0.99, value = 0.25, step = 0.01)

renderPlot({
par(mfrow=c(1,2))
  p=input$p_ber
plot(x=c(0,1),y=dbinom(c(0,1),size=1,prob=p),
     ylim=c(0,1),xlim=c(-0.5,2),xlab="x",pch=21,
     main=paste0(c("Función de probabilidad\n
                   Ber(p=",p,")"),collapse=""),bg="black")
segments(x0=0,y0=0,x1=0,y1=1-p, col = "blue", lty =2)
segments(x0=1,y0=0,x1=1,y1=p, col = "blue", lty =2)
segments(x0=-1,y0=1-p,x1=0,y1=1-p, col = "blue", lty =2)
```