

Probabilidad y variables aleatorias para ML con R y Python

Ricardo Alberich, Juan Gabriel Gomila y Arnau Mir

2020-04-09

Probabilidad y variables aleatorias para ML con R y Python



R.Alberich, J.G.Gomila y Arnau Mir
Curso online completo
Disponible en Udemy
<https://tinyurl.com/ujr92xv>

Índice general

	5
Pre requisitos: Teoría de conjuntos y combinatoria	7
0.1. Teoría de conjuntos	7
0.2. Combinatoria	11
0.3. Para acabar	15
0.4. Otros aspectos a tener en cuenta	15
1. Probabilidad	17
1.1. Probabilidades Básicas	17
1.2. Probabilidad condicionada	26
1.3. Bayes	32
1.4. Independencia de sucesos	34
2. Variables Aleatorias	37
2.1. Introducción a las variables aleatorias	37
2.2. Variables aleatorias discretas	39
2.3. Variables aleatorias continuas	50
2.4. Transformaciones de variables aleatorias	56
2.5. Desigualdades de Markov y de Chebychev	58
2.6. Cuantiles de variables aleatorias	60
3. Distribuciones Notables	63
3.1. Distribuciones discretas	63
3.2. Cuantiles de distribuciones notables discretas	118
3.3. Distribuciones continuas	120
4. Variables Aleatorias. Complementos	139
4.1. Momentos de variables aleatorias	139
4.2. Asimetría de una variable aleatoria	142
4.3. Curtosis o apuntamiento de una variable aleatoria	145
4.4. Métodos de transformación	146
4.5. Fiabilidad	151
4.6. Generación de muestras de variables aleatorias por ordenador	152
4.7. Entropía	156

5. Vectores aleatorios bidimensionales	161
5.1. Dos variables aleatorias	161
5.2. Función de distribución conjunta	163
5.3. Variables aleatorias bidimensionales discretas	168
5.4. Variables aleatorias bidimensionales continuas	174
5.5. Independencia de variables aleatorias	184
5.6. Momentos conjuntos y valores esperados conjuntos	191
5.7. Variables aleatorias condicionales y valor esperado condicional	200
5.8. Variables aleatorias definidas como función de dos variables aleatorias conjuntas	208
6. Vectores aleatorios	219
6.1. Varias variables aleatorias	219
6.2. Función de distribución conjunta	221
6.3. Variables aleatorias n -dimensionales discretas	227
6.4. Variables aleatorias n -dimensionales continuas	235
6.5. Independencia de variables aleatorias	245
6.6. Momentos conjuntos y valores esperados conjuntos	249
7. Ley de los grandes números y Teorema Central del Límite	257
7.1. Muestras aleatorias simples	257
7.2. Convergencia de sucesiones de variables aleatorias	258
7.3. Leyes de los grandes números	263
7.4. Teorema Central del Límite	272

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

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

También tienes todas las distribuciones de probabilidad en nuestro Shiny interactivo.

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 bidimensionales
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

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

0.1. 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, jugadores de fútbol, palabras, colores...).

La teoría de conjuntos básica es simple y natural y es la que necesitamos para este curso. Por otro lado, la teoría de conjuntos matemática es más compleja 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

0.1.1. Conjuntos básicos

Los conjuntos suelen tener un conjunto madre como por ejemplo

- $\mathbb{N} = \{0, 1, 2, \dots\}$
- $\mathbb{Z} = \{\dots, -2, -1, 0, 1, 2, \dots\}$

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

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

0.1.2. Características y propiedades básicas de los conjuntos

- A cada objeto x de Ω lo 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 de 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 B 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 vacío \emptyset y los elementales $\{1\}, \{2\}, \{3\}$.
 - 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 las partes de un conjunto es $\#(\mathcal{P}(\Omega)) = 2^{\#(\Omega)}$.

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

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

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

donde dado un $\omega \in \Omega$

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

0.1.3. Operaciones con conjuntos

Intersección.

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

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

Más formalmente

$$A \cap B = \{x \in \Omega \mid x \in A \text{ y } x \in B\}.$$

Unión.

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

Más formalmente

$$A \cup B = \{x \in \Omega \mid x \in A \text{ o } x \in B\}.$$

Diferencia.

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

Más formalmente

$$A - B = \{x \in \Omega \mid x \in A \text{ y } x \notin B\}.$$

Complementario

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

Más formalmente

$$A^c = \{x \in \Omega \mid x \notin A\}.$$

0.1.4. 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 $A \cap B = \emptyset$.
- $\Omega^c = \emptyset$.
- $\emptyset^c = \Omega$.
- $A \cup B = B \cup A$, $A \cap B = B \cap A$ (comutativas).
- $(A \cup B) \cup C = A \cup (B \cup C)$, $(A \cap B) \cap C = A \cap (B \cap C)$ (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.

0.1.5. 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
```

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

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

$A \cap B$

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

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

$A \cup B$

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

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

$B - C$

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

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

$A^c = \Omega - A$

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

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

0.1.6. Con Python, ejemplos

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

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

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

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

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

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

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

## set([8, 4, 6, 7])
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. Combinaciones

Número combinatorio o número binomial

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

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

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

$$\begin{aligned} C_7^5 &= \binom{7}{5} = \frac{7!}{5!(7-5)!} = \frac{7!}{5!2!} \\ &= \frac{1 \cdot 2 \cdot 3 \cdot 4 \cdot 5 \cdot 6 \cdot 7}{1 \cdot 2 \cdot 3 \cdot 4 \cdot 5 \cdot 1 \cdot 2} = \frac{6 \cdot 7}{2} = \frac{42}{2} = 21. \end{aligned}$$

Podemos formar 21 equipos distintos.

Ejercicio

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

0.2.2. Combinaciones con repetición

En combinatoria, las combinaciones con repetición de un conjunto son las distintas formas en que se puede hacer una selección de elementos de un conjunto dado, permitiendo que las selecciones puedan repetirse.

El número CR_n^k de multiconjuntos con k elementos escogidos de un conjunto con n elementos satisface:

- Es igual al número de combinaciones con repetición de k elementos escogidos de un conjunto con n elementos.
- Es igual al número de formas de repartir k objetos en n grupos.

$$CR_n^k = \binom{n+k-1}{k} = \frac{(n+k-1)!}{k!(n-1)!}.$$

Ejemplo

Vamos a imaginar que vamos a repartir 12 caramelos entre Antonio, Beatriz, Carlos y Dionisio (que representaremos como A, B, C, D). Una posible forma de repartir los caramelos sería: dar 4 caramelos a Antonio, 3 a Beatriz, 2 a Carlos y 3 a Dionisio. Dado que no importa el orden en que se reparten, podemos representar esta selección como ‘AAAABBCCDDDD’.

Otra forma posible de repartir los caramelos podría ser: dar 1 caramelo a Antonio, ninguno a Beatriz y Carlos, los 11 restantes se los damos a Dionisio. Esta repartición la representamos como ‘ADDDDDDDDDDD’.

Recíprocamente, cualquier serie de 12 letras A, B, C, D se corresponde a una forma de repartir los caramelos. Por ejemplo, la serie ‘AAAABBBBBD’ corresponde a: Dar 4 caramelos a Antonio, 5 caramelos a Beatriz, ninguno a Carlos y 3 a Dionisio.

De esta forma, el número de formas de repartir los caramelos es:

$$CR_{12}^4 = \binom{12+4-1}{4}$$

0.2.3. Variaciones.

Variaciones

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

Los podemos escribir

12, 13, 21, 23, 31, 32

Luego hay seis casos

0.2.4. Variaciones (sin repetición).

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

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

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

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

Ejercicio

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

Variaciones con repetición

¿Y repitiendo algún dígito?

$$VR_n^k = n^k$$

Efectivamente, en nuestro caso

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

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

0.2.5. Permutaciones

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

$$P_n = V_n^n = n!$$

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 función `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 $x = \{1, 2, 3 \dots, 10\}$.

0.2.6. Números multinomiales. Permutaciones con repetición.

Consideremos un conjunto de elementos $\{a_1, a_2, \dots, a_k\}$.

Entonces, si cada uno de los objetos a_i de un conjunto, aparece repetido n_i veces para cada i desde 1 hasta k , entonces el número de permutaciones con elementos repetidos es:

$$PR_n^{n_1, n_2, \dots, n_k} = \frac{n!}{n_1! \cdot n_2! \cdot \dots \cdot n_k!},$$

donde $n = n_1 + n_2 + \dots + n_k$.

Ejemplo

¿Cuántas palabras diferentes se pueden formar con las letras de la palabra PROBABILIDAD?

El conjunto de letras de la palabra considerada es el siguiente: $\{A, B, D, I, L, O, P, R\}$ con las repeticiones siguientes: las letras A, B, D, e I, aparecen 2 veces cada una; y las letras L, O, P, R una vez cada una de ellas.

Por tanto, utilizando la fórmula anterior, tenemos que el número de palabras (permutaciones con elementos repetidos) que podemos formar es

$$PR_{12}^{2,2,2,2,1,1,1,1} = \frac{12!}{(2!)^4(1!)^4} = 29937600.$$

0.3. Para acabar

0.3.1. Principios básicos para contar cardinales de conjuntos

El principio de la suma

Sean A_1, A_2, \dots, A_n conjuntos disjuntos dos a dos, es decir $A_i \cap A_j = \emptyset$ para todo $i \neq j$, $i, j = 1, 2, \dots, n$. Entonces

$$\#(\cup_{i=1}^n A_i) = \sum_{i=1}^n \#(A_i).$$

Principio de unión exclusión

Consideremos dos conjuntos cualesquiera A_1, A_2 entonces el cardinal de su unión es

$$\#(A_1 \cup A_2) = \#(A_1) + \#(A_2) - \#(A_1 \cap A_2).$$

El principio del producto

Sean A_1, A_2, \dots, A_n

$$\begin{aligned} \#(A_1 \times A_2 \times \dots \times A_n) &= \#(\{(a_1, a_2, \dots, a_n) | a_i \in A_i, i = 1, 2, \dots, n\}) \\ &= \prod_{i=1}^n \#(A_i). \end{aligned}$$

0.4. Otros aspectos a tener en cuenta

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

- 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.

Capítulo 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:



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

Ejemplo

El espacio muestral del ejemplo anterior del dado es

$$\Omega = \left\{ \begin{array}{c} \text{dado 1} \\ \text{dado 2} \\ \text{dado 3} \\ \text{dado 4} \\ \text{dado 5} \\ \text{dado 6} \end{array} \right\}$$

pero por comodidad, a partir de ahora pondremos

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

Suceso : Cualquier subconjunto del espacio muestral.

Algunos sucesos notables que merece la pena nombrar son:

- Suceso seguro o cierto: Ω .
- Suceso imposible o vacío: \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

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

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 “_”).

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:

$$\Omega = \{_\underline{B}\underline{a}, \underline{B}\underline{a}\underline{l}, \underline{a}\underline{l}\underline{e}, \underline{l}\underline{e}\underline{a}, \underline{e}\underline{a}\underline{r}, \underline{a}\underline{r}\underline{e}, \underline{r}\underline{e}\underline{s}, \underline{e}\underline{s}_\}$$

Algunos sucesos serían:

- 3-gramas que empiezan por a : $\{ale, are\}$.
- 3-gramas de inicio y final de palabra: $\{\underline{B}a, es_\}$.
- 3-gramas que contengan una l : $\{Bal, ale, lea\}$.

1.1.1. 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*; el que acontece si sucede A y NO sucede B .

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

Ejemplo género

Supongamos que una clase 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:

- $A \cup B$: Estudiantes 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 clase que NO son zurdas.
- $B - A$: Hombres de la clase que son zurdos.
- ¡Cuidado! No son incompatibles.

1.1.2. Propiedades

Comutativas:

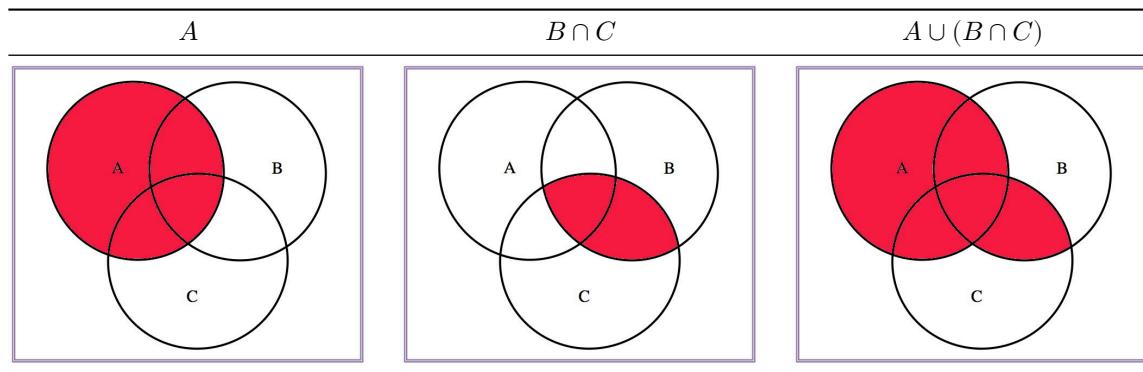
$$A \cup B = B \cup A, \quad A \cap B = B \cap A$$

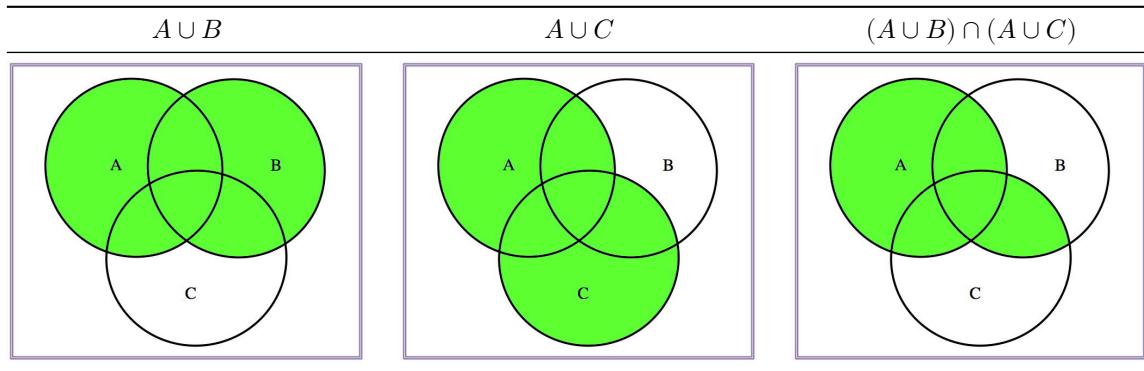
Asociativas:

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

Distributivas:

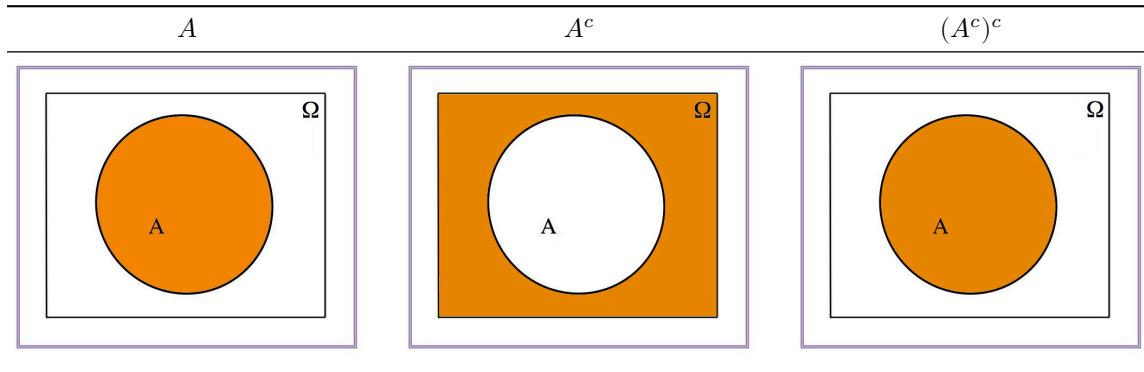
$$A \cap (B \cup C) = (A \cap B) \cup (A \cap C), \quad A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$$





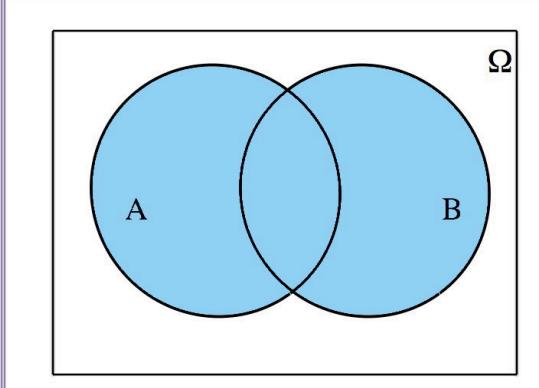
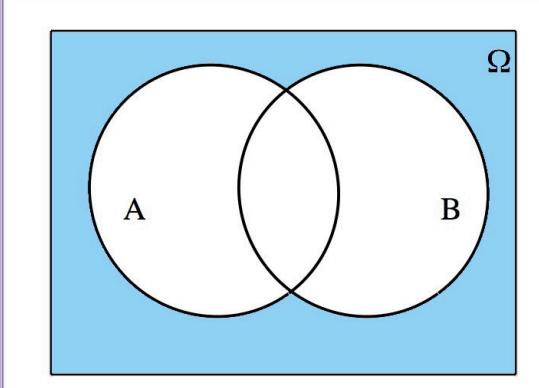
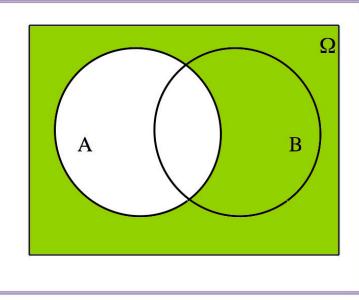
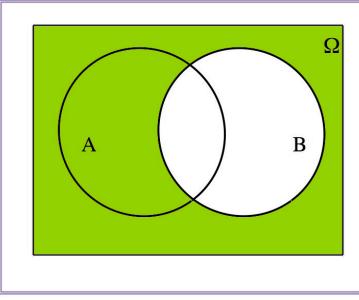
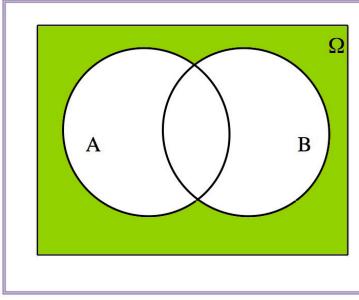
Complementario del complementario

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



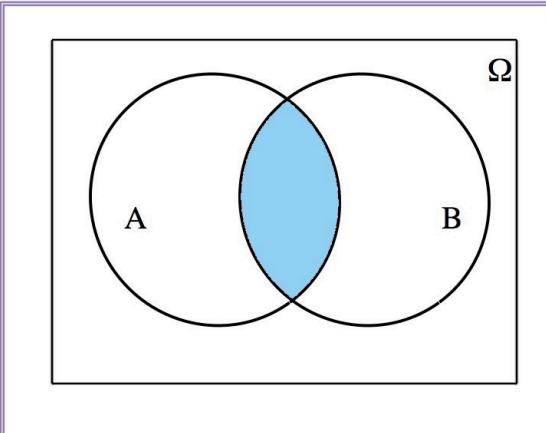
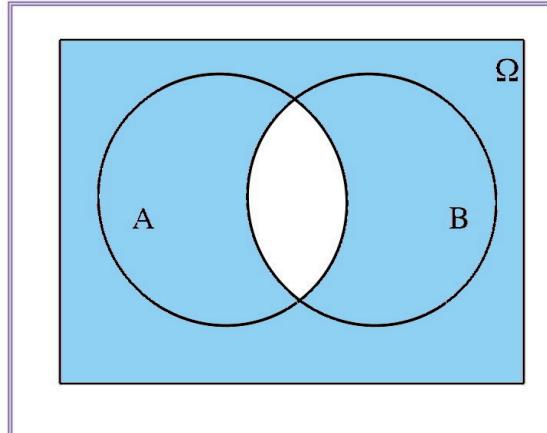
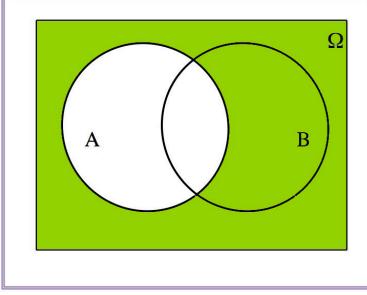
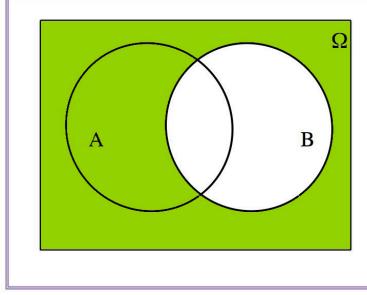
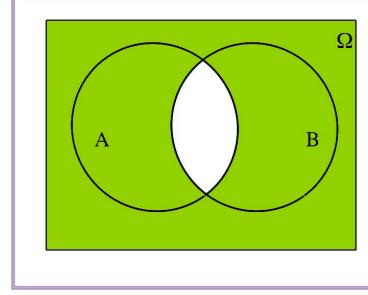
Leyes de De Morgan

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

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

Leyes de De Morgan

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

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

1.1.3. Definición de probabilidad

La probabilidad de un suceso es una puntuación (*score*) numérica 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

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 $P : \mathcal{P}(\Omega) \rightarrow [0, 1]$ con las siguientes propiedades:

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

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

$$P(A_1 \cup A_2 \cup \dots \cup A_n) = P(A_1) + P(A_2) + \dots + P(A_n).$$

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

Ejemplo: grupos sanguíneos

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:

$$A : 46\%; \quad B : 7,5\%; \quad AB : 3,5\%; \quad O : 43\%.$$

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

Experimento aleatorio: tipo de sangre de un paciente humano

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

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

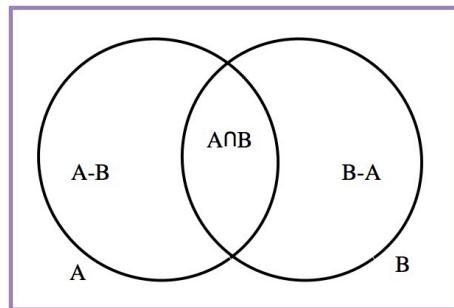
Suceso: $\{O\}^c = \{A, B, AB\}$

$$P(\{O\}^c) = P(\{A, B, AB\}) = P(A) + P(B) + P(AB) = 0,57.$$

1.1.4. Propiedades

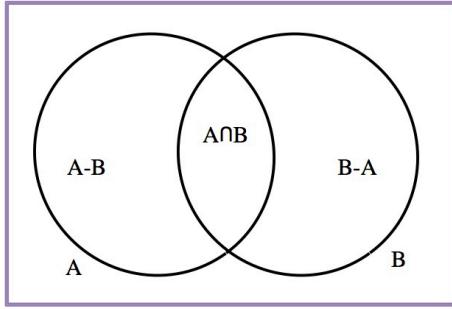
Propiedades básicas de la probabilidad

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



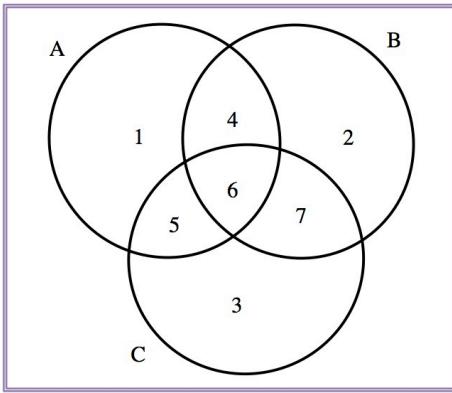
- Si $B \subseteq A$, entonces $0 \leq P(B) \leq P(A).$
- $P(A^c) = 1 - P(A).$
- $P(A \cup B) = P(A) + P(B) - P(A \cap B)$ porque

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



- $P(A \cup B \cup C)$:

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



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

- Si $A = \{a_1, a_2, \dots, a_k\}$, entonces

$$P(A) = P(a_1) + P(a_2) + \dots + P(a_k).$$

- Si todos los sucesos elementales tienen la misma probabilidad,

$$P(A) = \frac{|A|}{|\Omega|} \left(= \frac{\text{casos favorables}}{\text{casos posibles}} \right).$$

Ejemplo: Frecuencia de vocales

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 $\{E, O\}$.

Y su probabilidad es

$$P(\{E, O\}) = P(E) + P(O) = 0,261 + 0,244 = 0,505.$$

Ejemplo: Consumo de drogas

Según un artículo 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 dé positivo en ninguno de los 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:

$$P((A \cup B)^c) = 1 - P(A \cup B) = 1 - 0,0105 = 0,9895.$$

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 dé 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:

$$\begin{aligned} P(A \cap B) &= P(A) + P(B) - P(A \cup B) \\ &= 0,001 + 0,01 - 0,0105 = 0,0005. \end{aligned}$$

¿Cuál es la probabilidad de que un conductor analizado dé 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; $P(A \cap B) = 0,0005$.
- $A - B$: dar positivo en cocaína, pero no en cannabis,

de donde, por tanto:

$$P(A - B) = P(A) - P(A \cap B) = 0,001 - 0,0005 = 0,0005.$$

1.2. 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:

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

Ejemplo: frecuencia género y gafas

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?

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

- ¿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?

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

- ¿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?

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

1.2.1. ¡Atención!

Hay que distinguir bien entre

- $P(A \cap B)$: probabilidad de A **y** B , *Probabilidad de que lleve gafas y sea mujer*.
- $P(A|B)$: 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 .

1.2.2. Propiedades

La probabilidad condicionada es una probabilidad

Proposición

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

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

satisface las propiedades de las probabilidades, como por ejemplo:

$$\begin{aligned} P(B^c|A) &= 1 - P(B|A), \\ P(B_1 \cup B_2|A) &= P(B_1|A) + P(B_2|A) - P(B_1 \cap B_2|A). \end{aligned}$$

Ejercicio

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

Ejemplo: Hipertensos

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

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:

$$P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{0,09}{0,25} = 0,36.$$

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 $P(B|A)$:

$$P(B|A) = \frac{P(A \cap B)}{P(A)} = \frac{0,09}{0,15} = 0,6.$$

Ejemplo: dígitos de control

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 %. ¿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:

$$P(A \cap B) = P(B) \cdot P(A|B) = 0,005 \cdot 0,99 = 0,00495.$$

Ejemplo: SPAM

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?

¿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, $P((A \cup S)^c) = 0,15$.

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

- $P(A) = 0,5, P(S) = 0,65, P(A^c \cap S^c) = P((A \cup S)^c) = 0,15$,
- $P(A \cup S) = 1 - P((A \cup S)^c) = 0,85$,
- $P(A \cap S) = P(A) + P(S) - P(A \cup S) = 0,3$,

$$P(A|S) = \frac{P(A \cap S)}{P(S)} = \frac{0,3}{0,65} \approx 0,46.$$

¿Cuál es la probabilidad de que un correo no lleve adjuntos si no es SPAM?

$$P(A) = 0,5, P(S) = 0,65, P(A^c \cap S^c) = P((A \cup S)^c) = 0,15.$$

$$P(A^c|S^c) = \frac{P(A^c \cap S^c)}{P(S^c)} = \frac{P(A^c \cap S^c)}{1 - P(S)} = \frac{0,15}{0,35} \approx 0,43.$$

1.2.3. Teorema de la probabilidad total

Teorema de la probabilidad total

Dados dos sucesos A y B se tiene que

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

Partición del espacio espacio muestral

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

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

Teorema de la probabilidad total (generalización) Sea A_1, A_2, \dots, A_n una partición de Ω . Sea B un suceso cualquiera. Entonces

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

Ejemplo: Dígito de control de error

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,

$$\begin{aligned} P(A) &= P(B) \cdot P(A|B) + P(B^c) \cdot P(A|B^c) \\ &= 0,005 \cdot 0,99 + 0,995 \cdot 0,05 = 0,0547. \end{aligned}$$

1.2.4. 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.
- Una prueba de embarazo.
- Una prueba médica para una enfermedad concreta.

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

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

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 el clasificador entre cada clase debe determinarse por un valor umbral (*threshold*). Por ejemplo, para determinar si una persona está estresada 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).

Positivos y Negativos en Clasificación Consideremos un problema de predicción de clases binario, en el 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.

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.

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** $T \cap M^c$: El test da positivo, pero la condición no se 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**: $P(T^c|M)$.

Ejemplo: Prueba médica

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,
- $P(T) = 0,15$, $P(T^c|M) = 0,06$, $P(T|M^c) = 0,04$,
- ¿ $P(M)$?

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

donde

$$\begin{aligned} P(T|M) &= 1 - P(T^c|M) = 0,94, \\ P(M^c) &= 1 - P(M). \end{aligned}$$

Por lo tanto

$$\begin{aligned} 0,15 &= P(M) \cdot 0,94 + (1 - P(M)) \cdot 0,04 \\ &= 0,04 + 0,9 \cdot P(M), \\ P(M) &= \frac{0,11}{0,9} \approx 0,1222. \end{aligned}$$

1.3. Bayes

1.3.1. Fórmula de Bayes

Teorema de Bayes

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

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

Ejercicio

Demostrar el teorema de Bayes utilizando que

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

Teorema de Bayes generalizado

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

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

Ejercicio

Demostrar el teorema de Bayes utilizando que

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

Ejemplo: Test VIH

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?
- ¿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.

Ahora podemos calcular la probabilidad que un individuo que haya dado positivo en el test esté infectado:

$$\begin{aligned} P(A|B) &= \frac{P(B|A) \cdot P(A)}{P(B|A) \cdot P(A) + P(B|A^c) \cdot P(A^c)} \\ &= \frac{0,99 \cdot 0,005}{0,005 \cdot 0,99 + 0,995 \cdot 0,05} = 0,09. \end{aligned}$$

Para calcular la probabilidad de que un individuo que haya dado **negativo** en el test **no** esté infectado hacemos:

$$\begin{aligned} P(A^c|B^c) &= \frac{P(B^c|A^c) \cdot P(A^c)}{P(B^c|A) \cdot P(A) + P(B^c|A^c) \cdot P(A^c)} \\ &= \frac{0,95 \cdot 0,995}{0,01 \cdot 0,005 + 0,95 \cdot 0,995} = 0,999947. \end{aligned}$$

Ejercicio: Ventas por internet

Se ha observado que los clientes de una empresa de ventas por internet son de tres tipos, A, B y C, disjuntos dos a dos. La probabilidad de 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 estudiar el suceso E : el cliente compra, se tiene que:

$$P(E|A) = 0,5, P(E|B) = 0,75, P(E|C) = 0,6.$$

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

Ejercicio: Detección precoz de abandono de clientes

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 dé positivo en el test?
- ¿Cuál es la probabilidad que un individuo escogido al azar se dé 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; $P(B) = 0,02$,
- $P(T|B) = 0,975, P(T|B^c) = 0,12$.

$$P(B) = 0,02, P(T|B) = 0,975, P(T|B^c) = 0,12.$$

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

$$P(B \cap T) = P(B) \cdot P(T|B) = 0,02 \cdot 0,975 = 0,0195.$$

¿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, \dots, A_n son sucesos **independientes** cuando, para toda subfamilia A_{i_1}, \dots, A_{i_k} ,

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

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. $P(B|A) = P(B)$.

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.

Ejemplo: Compra billete avión

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

$$\begin{aligned} P(A \cap B) &= P(A) + P(B) - P(A \cup B) \\ &= 0,55 + 0,2 - 0,6 = 0,15, \\ P(A) \cdot P(B) &= 0,55 \cdot 0,2 = 0,11. \end{aligned}$$

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

1.4.2. 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 sí mismo?

Capítulo 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; estas 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 $X, Y, Z \dots$
- Los valores que “*toman*” las v.a. los representaremos por letras minúsculas (las mismas en principio) $x, y, z \dots$

Ejemplo: Dado seis caras

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

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

Una v.a $X : \Omega \rightarrow \mathbb{R}$ 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 número inferior o igual a 3” es en términos de la v.a. $\{X = 1, X = 2, X = 3\}$ o también $\{X \leq 3\}$.

Ejemplo: Juego lanzamiento anilla

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:

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

definida por

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

2.1.2. Tipos de variables aleatorias

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

Damos a continuación una definición informal.

Variables Aleatorias 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.

Ejemplo: Tipos de variables aleatorias

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.

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 $P_X(x) = 0$.

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 $X(\Omega) = D_X$.

Ejemplo: Juego del 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

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

y la variable aleatoria $X : \Omega \rightarrow \mathbb{R}$ viene definida por

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

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

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

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 $D_X = \{0, 1\}$.

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:

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

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 $D_X = \{1, 2, 3\}$.

Propiedades básicas de la función de probabilidad

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

- $0 \leq P_X(x) \leq 1$, para todo $x \in \mathbb{R}$.
- $\sum_{x \in D_X} P_X(x) = 1$.

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:

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

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:

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

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*.

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

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

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

Más 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 $F_X(x_0^-) = \lim_{x \rightarrow x_0^-} F_X(x)$, entonces se cumple que $P(X < x_0) = F_X(x_0^-)$ y que $P(X = x_0) = F_X(x_0) - F_X(x_0^-)$.
- Se cumple que $\lim_{x \rightarrow \infty} F_X(x) = 1$; $\lim_{x \rightarrow -\infty} F_X(x) = 0$.
- 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 $a < b$,

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

Advertencia desigualdades estrictas

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

Veamos qué propiedades tenemos cuando se cambian estas desigualdades.

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

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

entonces se cumplen las siguientes igualdades:

- $P(X = x) = F_X(x) - F_X(x^-)$.
- $P(a < X < b) = F_X(b^-) - F_X(a)$.
- $P(a \leq X < b) = F_X(b^-) - F_X(a^-)$.
- $P(X < a) = F_X(a^-)$.
- $P(a \leq X \leq b) = F_X(b) - F_X(a^-)$.
- $P(X \geq a) = 1 - F_X(a^-)$.
- 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 $x \leq x_0$.

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

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

Ejemplo: dado (continuación)

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,

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

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,

$$\begin{aligned} F_X(3,5) &= \sum_{x \leq 3,5} P_X(x) = \sum_{x=1}^3 P(X = x) \\ &= \sum_{x=1}^3 \frac{1}{6} = 3 \cdot \frac{1}{6} = \frac{1}{2}. \end{aligned}$$

Propiedades de la función de distribución

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 decir, es una función creciente, no necesariamente estrictamente creciente.
- $\lim_{x \rightarrow -\infty} F_X(x) = 0$ y $\lim_{x \rightarrow +\infty} F_X(x) = 1$.
- Es continua por la derecha: $\lim_{x \rightarrow x_0^+} F_X(x) = F_X(x_0)$.

2.2.2. Valores esperados o esperanza

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 correspondientes 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.

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 denomina **media** (*mean* en inglés) poblacional o simplemente media y muy frecuentemente se la denota $\mu_X = E(X)$ o simplemente $\mu = E(X)$.

Ejemplo: lanzamiento de un dado n veces

Supongamos que lanzamos un dado n veces y obtenemos unas frecuencias absolutas n_i para el resultado i con $i = 1, \dots, 6$. 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

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

Si $n \rightarrow \infty$ se tiene que $\lim_{n \rightarrow \infty} \frac{n_x}{n} = P_X(x)$.

Por lo tanto $E(X) = \lim_{n \rightarrow \infty} \sum_{x=1}^6 x \cdot \frac{n_x}{n}$.

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.

Ejemplo: Erratas en un texto

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

Resulta que

$$P(X = 0) = 0,42, \quad P(X = 1) = 0,4, \quad P(X = 2) = 0,18.$$

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 el editor nos paga 2 euros por cada página que encontramos 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?

$$0 \cdot 0,42 + 2 \cdot 0,4 + 3 \cdot 0,18 = 1,34.$$

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 :

$$E(g(X)) = \sum_x g(x) \cdot P_X(x).$$

Propiedades de los valores esperados

- $E(k) = k$ 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.

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 que la 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 qué 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?

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

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

Series geométricas

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

$$r^0, r^1, \dots, r^n, \dots$$

La serie geométrica es la suma de todos los valores de la progresión geométrica $\sum_{k=0}^{+\infty} r^k$.

Propiedades

- Las sumas parciales desde el término n_0 al n de una progresión geométrica valen

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

- 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

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

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

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

Ejemplo: paleta de colores (continuación)

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

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

Ahora calculemos su función de distribución

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

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

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

Ejercicio

Calculad el valor esperado de la variable

Y = número de intentos para conseguir el color azul.

Momentos de orden m

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

$$E((X - C)^m).$$

- 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.

Resumen de 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 .

2.2.3. Medidas de la variabilidad

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 .

Propiedad

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

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

- Sea X una v.a.

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

Demostración de b)

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

Ejemplo: número de errores (continuación)

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:

$$Var(X) = E(X^2) - (E(X))^2 = E(X^2) - (0,76)^2.$$

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 $\sigma_X = \sqrt{0,542}$.

Más 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.

2.2.4. 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 $Y = a + bX$:

- $E(Y) = E(a + bX) = a + b \cdot E(X) = a + b \cdot \mu_X$.
- $Var(Y) = Var(a + bX) = b^2 \cdot Var(X) = b^2 \cdot \sigma_X^2$.
- $\sigma_Y = \sqrt{Var(Y)} = \sqrt{b^2 \cdot Var(X)} = |b| \cdot \sigma_X$.

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.

2.3. Variables aleatorias continuas

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 mejor 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*.

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.

Ejemplo: distancia de un dardo al centro de la diana

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 $X = \text{distancia del dardo al centro de la diana}$.

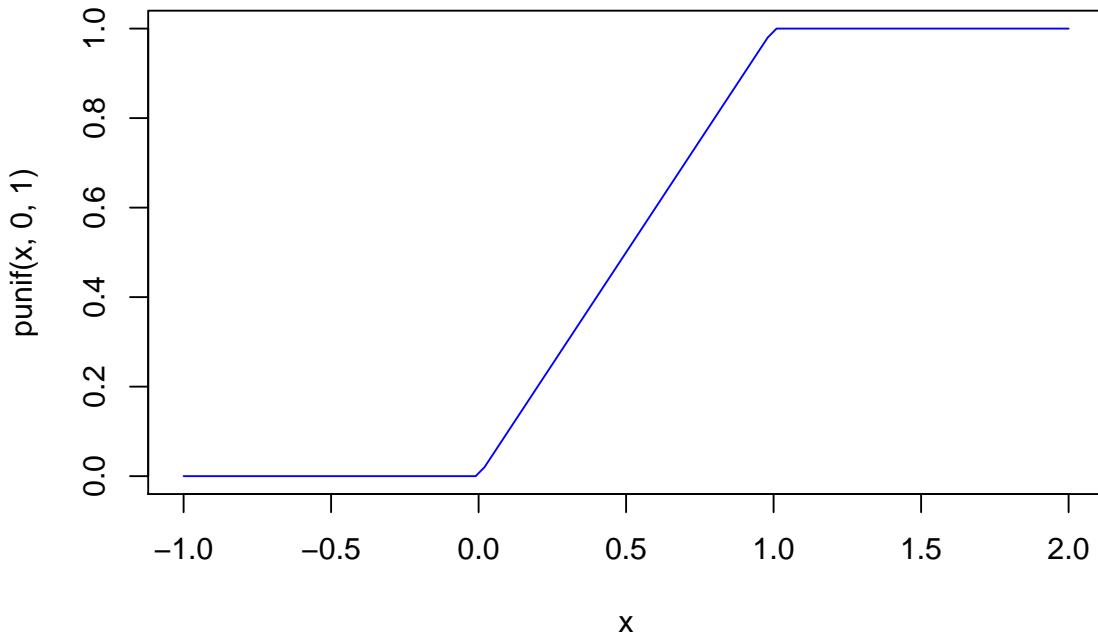
Su función de distribución es

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

- C.F. *longitud favorable* es $x - 0$.
- C.P. *longitud posible* es $1 - 0$.
- Luego

$$P(X \leq x) = \frac{\text{C.F.}}{\text{C.P.}} = \frac{x - 0}{1 - 0} = x.$$

Función de distribución de una v.a. uniforme en el intervalo unidad



2.3.1. 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:

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

Demostración:

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

Para demostrar la segunda, tenemos

$$\{X \leq a\} \cap \{a < X < b\} = \emptyset,$$

$$\{X \leq a\} \cup \{a < X < b\} = \{X < b\},$$

entonces

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

Ejercicio

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

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.

Ejemplo: diana (continuación)

En el ejemplo de la diana:

$$P(0,25 < X < 0,3) = F_X(0,3) - F_X(0,25) = 0,3 - 0,25 = 0,05.$$

2.3.2. Función de densidad

Función de densidad

Una función $f : \mathbb{R} \rightarrow \mathbb{R}$ 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 sumo en una cantidad finita de puntos sobre cada intervalo acotado de \mathbb{R} .
- $\int_{-\infty}^{+\infty} f_X(x) dx = 1$.

Función de distribución de una variable aleatoria

Sea X una v.a. con función de distribución F_X . Sea $f : \mathbb{R} \rightarrow \mathbb{R}$ 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 .

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 como su conjunto de resultados posibles.

Ejemplo: diana (continuación)

En nuestro ejemplo, la función f es una densidad

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

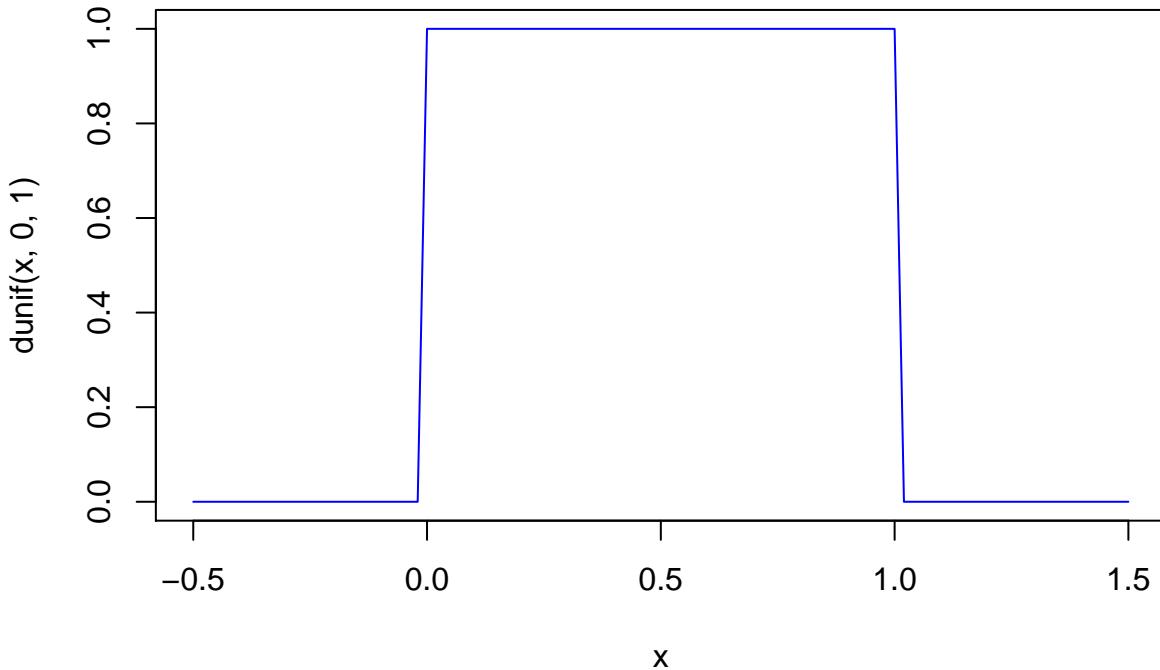
que es la densidad de X . En efecto:

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

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

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

Densidad de la distribución uniforme en [0,1]



2.3.3. 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 subconjunto de \mathbb{R} entonces

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

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.

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

$$F_X(x) = P(X \leq x) = \frac{\text{Casos Favorables}}{\text{Casos Posibles}} = \frac{x}{2}.$$

Luego su función de distribución es:

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

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: $\{0, 2\}$.
- $F_X(x) = \int_{-\infty}^x f_X(t)dt$, para todo $x \in \mathbb{R}$. (Ejercicio: resolverlo gráficamente.)
- $$\int_{-\infty}^{+\infty} f_X(x)dx = \int_0^2 \frac{1}{2} dx = \left[\frac{x}{2} \right]_0^2 = \frac{2}{2} - \frac{0}{2} = 1.$$

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.

2.3.4. 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 $f_X(x)$

Esperanza v.a. continuas

- Su esperanza es:

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

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

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

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:

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

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 $Var(X)$.

Ejemplo: diana (continuación)

Calcular μ_X y σ_X^2 en el ejemplo de la diana.

Resultado

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

Proposición

Sea X una v.a. continua con $E(X) = \mu_X$ y $Var(X) = \sigma_X^2$ sea $Y = a + b \cdot X$, 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:

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

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

Ejemplo: venta de vinos

En una empresa de venta de vinos por internet, sea X el 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

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

2.4. 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 .

Transformaciones de v.a. discretas

Sea X una v.a. discreta con $X(\Omega) = \{x_1, x_2, \dots, x_n, \dots\}$ y sea $h : \mathbb{R} \rightarrow \mathbb{R}$ 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

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

Desafortunadamente para variables no discretas, el resultado no es tan sencillo como la expresión 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 $h : \mathbb{R} \rightarrow \mathbb{R}$ una aplicación estrictamente monótona y derivable; por lo tanto, $h'(x) \neq 0$ 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

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

Densidad de una transformación de una v.a. continua

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

$$h : \mathbb{R} \rightarrow \mathbb{R},$$

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, \dots, x_n ,

entonces:

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

Método general de 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

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

Por ejemplo, si g es estrictamente creciente y continua,

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

y si g es estrictamente decreciente y continua,

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

2.5. Desigualdades de Markov y de Chebychev

En esta sección veremos 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 medida de la dispersión de los datos.

2.5.1. 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

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

de donde se sigue que

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

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.

2.5.2. Desigualdad de Chebychev

Desigualdad de Chebychev

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

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

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

Demostración

Aplicaremos 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.

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 la probabilidad de que X sea constantemente $E(X)$ es 1, hecho que nos confirma la utilidad de la varianza como una medida de la dispersión de los datos.

Ejemplo: tiempo de respuesta

Se sabe que el tiempo de respuesta medio y la desviación típica de un sistema multiusuario son 15 y 3 unidades de tiempo, 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.

Más formas de la desigualdad 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 $a = k \cdot \sigma$,

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

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	$\leq 0,25$
3	$\leq 0,111$
4	$\leq 0,0025$

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.!

2.6. Cuantiles de variables aleatorias

Si X es una v.a. con dominio D_X y $0 < p < 1$, llamaremos cuantil de orden p al menor valor perteneciente al dominio $x_p \in D_X$ tal que:

$$P(X \leq x_p) \geq p.$$

En \mathbb{R} , cada distribución X tiene la función $q_X(p, \dots)$ que devuelve precisamente el cuantil x_p tal que $P(X \leq x_p) \geq p$.

Dada una variable aleatoria X , si existe la inversa de la función de distribución de X , F_X^{-1} , el cuantil de orden p sería el valor que tiene la función F_X^{-1} en p : $x_p = F^{-1}(p)$.

En caso de no existir la inversa, dado p , definimos el conjunto A_p como:

$$A_p = \{x \in \mathbb{R}, \mid F_X(x) \geq p\}.$$

Entonces el cuantil p es el mínimo del conjunto A_p considerando sólo valores del dominio de la variable: $x_p = \min_{x \in D_X} (A_p)$. Este mínimo siempre existirá y nos da una fórmula explícita para calcular los cuantiles de cualquier variable aleatoria.

Ejemplo: variable aleatoria que nos da el resultado del lanzamiento de un dado

Sea X la variable aleatoria uniforme discreta que nos da el número de puntos obtenidos en el lanzamiento de un dado (seis caras numeradas del 1 al 6).

Su dominio es $D_X = \{1, 2, 3, 4, 5, 6\}$ y su función de probabilidad es

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

Su función de distribución es:

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

La función siguiente llamada `ddado` nos define la función de probabilidad de X para un dado de n caras:

```
ddado=function(x,n=6) {
  sapply(x,FUN=function(x) {
    if( x %in% c(1:n)){return(1/n)} else {return(0)}}
  )}
```

Por ejemplo, el valor de $P_X(0.5)$ sería:

```
ddado(0.5,n=6)
```

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

y los valores de $P_X(i)$ para $i = 1, \dots, 10$ sería:

```
ddado(1:10,n=6)
```

```
## [1] 0.1666667 0.1666667 0.1666667 0.1666667 0.1666667 0.1666667 0.1666667 0.0000000
## [8] 0.0000000 0.0000000 0.0000000
```

La función `pdado` nos da la función de distribución de X :

```
pdado=function(x,n=6)
{
  sapply(x,FUN=function(y){ if (y<1){ return(0)}else{if(y>=n){return(1)} else
  {return(sum(ddado(c(1:(floor(y))),n=n)))}}})
}
```

Los valores de $F_X(i)$ para $i = 0, \dots, 11$ serían:

```
pdado(0:11,6)
```

```
## [1] 0.0000000 0.1666667 0.3333333 0.5000000 0.6666667 0.8333333 1.0000000
## [8] 1.0000000 1.0000000 1.0000000 1.0000000 1.0000000
```

A continuación, construimos la función `qdado` que nos calcula el cuantil p , para $0 \leq p \leq 1$, de la variable X como el mínimo del conjunto A_p :

```
qdado=function(p,n=6){
  sapply(p,FUN=function(pp=p,nn=n)
  {
    if(pp<0 | pp>1) {return(NA)}
    else {
```

```
aux=pp>=pdado(1:n,nn)
aux
ifelse(all(!aux),return(1),return(max(which(pp>=pdado(1:n,nn)))))}}
```

Si $p = 1.5$ o $p = -1$, la función anterior nos devuelve NA ya que ni 1.5 ni -1 están entre 0 y 1:

```
qdado(1.5)
```

```
## [1] NA
```

```
qdado(-1)
```

```
## [1] NA
```

Los cuantiles $x_{0,1}$, $x_{0,5}$ $x_{0,6}$ y x_1 son:

```
qdado(c(0.1,0.5,0.6,1))
```

```
## [1] 1 3 3 6
```

Capítulo 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 número de clientes en una cola...

3.1. Distribuciones discretas

3.1.1. Distribución de 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

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

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 $X \equiv Ber(p)$ 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 conocida familia de científicos suizos Bernoulli.

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 $E(X^2)$:

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

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:

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

Resumen v.a con distribución Bernoulli

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

Ejemplo de Distribución Bernoulli

Veamos los cálculos básicos usando la distribución $Ber(p = 0,25)$ en R.

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

## [1] 0.75

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

## [1] 0.25

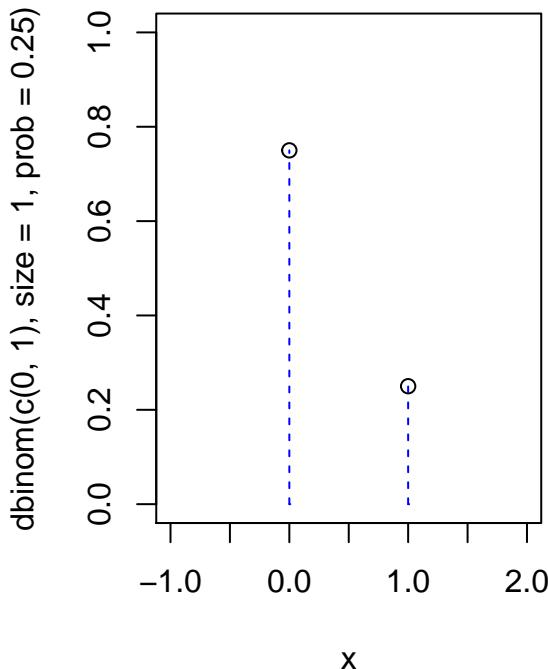
rbinom(n=20,size = 1,prob=0.25)

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

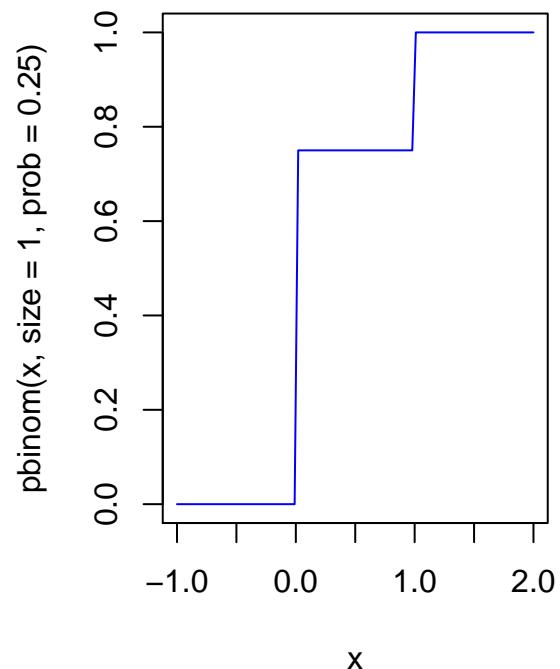
El siguiente código dibuja las función de probabilidad y la de distribución de una $Ber(p = 0,25)$

```
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))
```

**Función de probabilidad
 $Ber(p=0.25)$**



**Función de distribución
 $Ber(p=0.25)$**



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)
  segments(x0=-1,y0=p,x1=1,y1=p, col = "blue", lty =2)
  x=0:1
  y=pbis(x,size=1,prob=p)
  curve(pbinom(x,size=1,prob=p),
        xlim=c(-1,2),col="blue",
        main=paste0(c("Función de distribución\n Ber(p=",p,")"),collapse=""))
})

par(mfrow=c(1,1))
```

3.1.2. Distribución binomial

Si repetimos n veces de forma independiente un experimento Bernoulli de parámetro p , el espacio muestral Ω estará formado por cadenas de E 's y F 's de longitud n . Consideremos la v.a.:

$$X(\overbrace{EFFF \dots EEF}^n) = \text{número de éxitos en la cadena.}$$

A la variable aleatoria anterior se la conoce como distribución binomial de parámetros n y p , y lo denotaremos por $X \equiv B(n, p)$.

Función de probabilidad de una binomial

Su función de probabilidad es:

$$P_X(x) = \begin{cases} \binom{n}{x} \cdot p^x \cdot (1-p)^{n-x}, & \text{si } x = 0, 1, \dots, n, \\ 0, & \text{en otro caso.} \end{cases}$$

Función de distribución de una binomial

Su **función de distribución** no tiene una fórmula cerrada. Hay que acumular la función de probabilidad:

$$\begin{aligned} F_X(x) = P(X \leq x) &= \sum_{i=0}^x P_X(i) \\ &= \begin{cases} 0, & \text{si } x \leq 0, \\ \sum_{i=0}^k \binom{n}{i} \cdot p^i \cdot (1-p)^{n-i}, & \text{si } \begin{cases} k \leq x < k+1, \\ k = 0, 1, \dots, n, \end{cases} \\ 1, & \text{si } n \leq x. \end{cases} \end{aligned}$$

Números binomiales con R

Los números binomiales calculan el número de equipos de baloncesto distintos que ($k = 5$ jugadores) se pueden hacer con 6 jugadores ($n = 6$).

Es decir, cuántas maneras distintas hay para elegir (*choose*) 5 jugadores en un conjunto de 6 jugadores. Todo el mundo diría ¡¡¡6!!! Efectivamente con R es

```
choose(6,5)
```

```
## [1] 6
```

Con 10 jugadores, el número de equipos de 5 distintos es bastante más grande

```
choose(10,5)
```

```
## [1] 252
```

Y, por ejemplo, con un equipo de fútbol profesional que tiene en plantilla 22 jugadores (quitando los guardametas) se pueden formar ¡¡nada menos que!!

```
choose(22,10)
```

```
## [1] 646646
```

un bonito número capicúa que nos da el número de equipos distintos que se pueden formar.

Obviamente se tiene que una v.a. Bernoulli es una binomial con $n = 1$: $B(1, p) = Ber(p)$.

Ejercicio

Calculad las funciones de distribución de una binomial $B(n = 1, p = 0,3)$ y comprobad que coinciden con las distribuciones de una $Ber(p = 0,3)$.

Observaciones sobre la distribución binomial

- La probabilidad de fracaso se suele denotar con $q = 1 - p$, **sin ningún aviso adicional**, con el fin de acortar y agilizar la escritura de las fórmulas.
- Su **función de distribución no tienen una formula general**, hay que calcularla con una función de R o Python. En el siglo pasado se tabulaban en los libros de papel :-).
- En el material adicional os pondremos unas tablas de esta distribución para distintos valores de n y p para que disfrutéis de tan ancestral método de cálculo.
- Cualquier paquete estadístico u hoja de cálculo dispone de funciones para el cálculo de estas probabilidades, así que el **uso de las tablas** queda **totalmente anticuado**.

Esperanza de una $B(n, p)$

Su **esperanza** es:

$$E(X) = \sum_{k=0}^n k \cdot \binom{n}{k} \cdot p^k \cdot q^{n-k} = n \cdot p.$$

La esperanza de X^2 es:

$$E(X^2) = \sum_{k=0}^n k^2 \cdot \binom{n}{k} \cdot p^k \cdot q^{n-k} = n \cdot p \cdot q + (n \cdot p)^2.$$

Varianza de una $B(n, p)$

Su **varianza** es:

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

Su desviación típica es:

$$\sqrt{n \cdot p \cdot q} = \sqrt{n \cdot p \cdot (1 - p)}.$$

En temas posteriores veremos una forma sencilla del cálculo de la esperanza y varianza de una $B(n, p)$ como la suma de n v.a. $Ber(p)$ independientes.

Ejercicio

Justificar de forma intuitiva que si X_i con $i = 1, 2, \dots, n$ son v.a. $Ber(p)$ independientes entonces $X = \sum_{i=1}^n X_i$ sigue una distribución $B(n, p)$.

Resumen v.a con distribución binomial $B(n, p)$

X binomial	$B(n, p)$
$D_X =$	$\{0, 1, \dots, n\}$
$P_X(x) = P(X = x) =$	$\begin{cases} \binom{n}{x} p^x (1-p)^{n-x} & \text{si } x = 0, \dots, n \\ 0 & \text{en otro caso.} \end{cases}$
$F_X(x) = P(X \leq x) =$	no tiene fórmula (utilizad funciones de R o Python)
$E(X) = n \cdot p$	$Var(X) = n \cdot p \cdot (1 - p)$

Cálculos de la distribución Binomial con R

Veamos los cálculos básicos con funciones de R para una v.a X con distribución binomial $B(n = 10, p = 0,25)$.

Si queremos calcular con R algún valor de la función de distribución, como por ejemplo $F_X(0) = P(X \leq 0)$, tenemos que hacer:

```
binom(0, size=10, prob=0.25)
```

```
## [1] 0.05631351
```

y si queremos por ejemplo $F_X(4) = P(X \leq 4)$, tenemos que hacer:

```
pbinom(4,size=10,prob=0.25)
```

```
## [1] 0.9218731
```

Sin embargo, si queremos calcular algún valor de la función de probabilidad, como por ejemplo $P(X = 0)$, tenemos que hacer:

```
dbinom(0,size=10,prob=0.25)
```

```
## [1] 0.05631351
```

o, por ejemplo para $P(X = 4)$:

```
dbinom(4,size=10,prob=0.25)
```

```
## [1] 0.145998
```

Generación de muestras aleatorias con R

Generaremos una muestra aleatoria de 100 valores de una población con distribución $B(20, 0,5)$

```
set.seed(2019)
```

```
rbinom(100,size = 20,prob=0.5)
```

```
## [1] 12 11 9 11 6 6 12 5 7 11 12 11 8 8 11 11 7 11 9 10 9 10 14 8 8
## [26] 5 11 14 11 10 11 5 12 8 6 7 9 10 5 12 11 9 12 11 12 10 13 13 8 8
## [51] 9 7 6 9 10 9 16 13 6 6 8 8 11 9 12 15 9 7 12 11 9 8 9 8 11
## [76] 15 7 10 9 12 6 13 14 8 10 8 10 11 11 9 10 11 12 8 10 12 9 13 9 13
```

Ejemplo

El ejemplo anterior correspondería a repetir 100 veces el experimento de lanzar una moneda 20 veces y contar el número de caras.

Cálculos de la distribución Binomial con Python

Veamos los cálculos básicos con funciones de Python para una v.a X con distribución binomial $B(n = 10, p = 0,25)$.

Primero importamos la función `binom` de la librería `scipy.stats`:

```
from scipy.stats import binom
```

En general, en el paquete `scipy`, la función de probabilidad se invocará con el método `pmf`, la de distribución con el método `cdf`, mientras que una muestra aleatoria que siga esta distribución, con el método `rvs`. En todos ellos aparecerá siempre el parámetro `loc` que se utiliza para desplazar el dominio de la variable aleatoria. Por ejemplo, en este caso:

```
binom.pmf(k, n, p, loc) = binom.pmf(k - loc, n, p)
```

Para calcular los valores de la función de distribución como por ejemplo $F_X(0) = P(X \leq 0)$ y $F_X(4) = P(X \leq 4)$ utilizamos la función `cdf`:

```
binom.cdf(0,n=10,p=0.25)
```

```
## 0.056313514709472656
```

```
binom.cdf(4,n=10,p=0.25)
```

```
## 0.92187309265136719
```

Notemos que al no indicar el valor de `loc`, se le asume que toma el valor 0.

Para calcular los valores de la función de probabilidad $P(X = 0)$ y $P(X = 4)$ utilizamos la función `pmf`:

```
binom.pmf(0,n=10,p=0.25)
```

```
## 0.056313514709472684
```

```
binom.pmf(4,n=10,p=0.25)
```

```
## 0.14599800109863295
```

Notemos que al no indicar el valor de `loc`, se le asume que toma el valor 0.

Si queremos generar una muestras aleatorias que siga una distribución binomial, podemos usar la función `rvs`. En este caso, generaremos una muestra aleatoria de 100 valores de una población $B(20, 0,5)$

```
binom.rvs(n=20,p=0.25,size = 100)
```

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

Observación

Notemos que la secuencia aleatoria generada no es la misma que con R. De hecho, si volvemos a ejecutar esta función obtendremos una muestra aleatoria distinta.

```
binom.rvs(n=20,p=0.25,size = 100)
```

```
## array([ 6,  9,  4,  2,  6,  5,  4,  7,  4,  6,  1,  4,  6,  6,  6,  3,  4,
##        6,  4,  9,  4,  2,  7,  6,  6, 11,  3,  2,  1,  6,  5,  4,  7,  4,
##        0,  5,  3,  6,  5,  4,  5,  1,  4,  6,  5,  7,  4,  3,  4,  4,  3,
##        6,  2,  5,  6,  3,  7,  3,  3,  7,  0,  4,  4,  8,  5,  6,  6,  4,
##        0,  5,  6,  6,  3,  6,  7,  5,  0,  4,  1,  2,  4,  3,  4,  4,  2,
##        6,  4,  5,  2,  6,  4,  7,  4,  4,  7,  5,  6,  5,  4,  6])
```

Veamos algunos cálculos básicos con funciones de Python para la binomial $B(n = 10, p = 0,25)$.

```
binom.cdf(5,n=10,p=0.25)
```

```
## 0.98027229309082031
```

```
binom.pmf(1,n=10,p=0.25)

## 0.18771171569824247
binom.rvs(n=20,p=0.25,size=10)

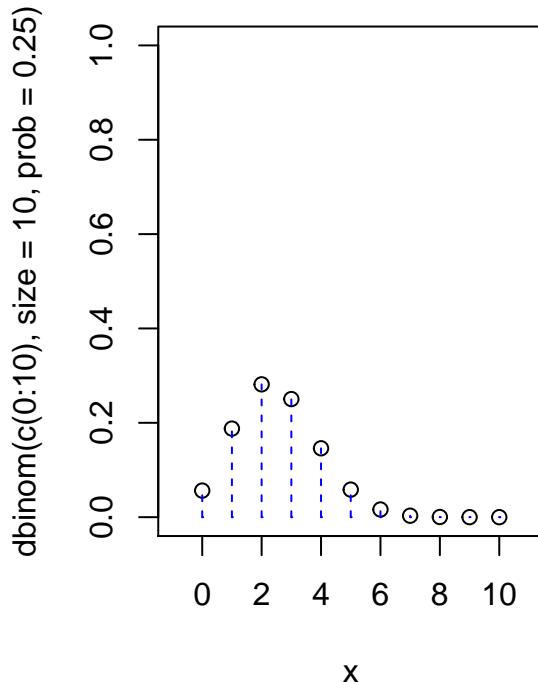
## array([4, 4, 5, 5, 3, 3, 4, 6, 7, 4])
```

Gráficas de la distribución Binomial con R

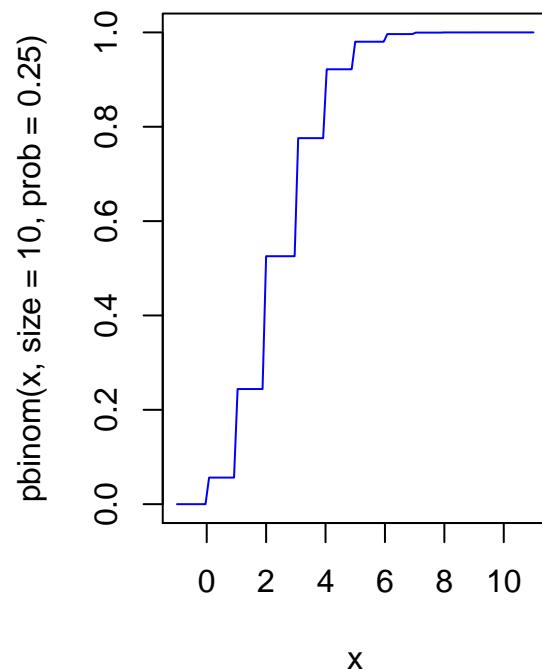
El siguiente código de R dibuja las función de probabilidad y la de distribución de una $B(n = 10, p = 0.25)$:

```
par(mfrow=c(1,2))
aux=rep(0,22)
aux[seq(2,22,2)]=dbinom(c(0:10),size=10,prob=0.25)
plot(x=c(0:10),y=dbinom(c(0:10),size=10,prob=0.25),
      ylim=c(0,1),xlim=c(-1,11),xlab="x",
      main="Función de probabilidad\n B(n=10,p=0.25)")
lines(x=rep(0:10,each=2),y=aux, type = "h", lty = 2,col="blue")
curve(pbinom(x,size=10,prob=0.25),
      xlim=c(-1,11),col="blue",
      main="Función de distribución\n B(n=10,p=0.25)")
par(mfrow=c(1,1))
```

**Función de probabilidad
 $B(n=10,p=0.25)$**



**Función de distribución
 $B(n=10,p=0.25)$**



Gráficas interactivas de la distribución Binomial

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.

```
fluidPage(
  fluidRow(
    column(6,
      sliderInput("n_binom", label = "Número de repeticiones n:",
                  min = 1, max = 50, value = 10, step = 1)),
    column(6,
      sliderInput("p_binom", label = "Probabilidad éxito p:",
                  min = 0.01, max = 0.99, value = 0.25, step = 0.01)
    )
  )
)

renderPlot({
  n=input$n_binom
  pr=input$p_binom

  par(mfrow=c(1,2))
  aux=rep(0,(n+1)*2)
  aux[seq(2,(n+1)*2,2)]=dbinom(c(0:n),size=n,prob=pr)
  plot(x=c(0:n),y=dbinom(c(0:n),size=n,prob=pr),
    ylim=c(0,1),xlim=c(-1,n+1),xlab="x",
    main=paste0(c("Función de probabilidad\n B(n=",n,",p=",pr,")"),collapse = ""))
  lines(x=rep(0:n,each=2),y=aux, type = "h", lty = 2,col="blue")
  curve(pbinom(x,size=n,p=pr),
    xlim=c(-1,n+1),col="blue",
    main=paste0(c("Función de distribución\n B(n=",n,",p=",pr,")"),
                collapse = ""))
  par(mfrow=c(1,1))
})
```

Gráficos de la distribución binomial con Python

Ejercicio

Buscad en la documentación de Python cómo se dibuja la función de probabilidad y de distribución de una binomial y recread los gráficos anteriores.

Pista: Necesitaremos investigar más librerías:

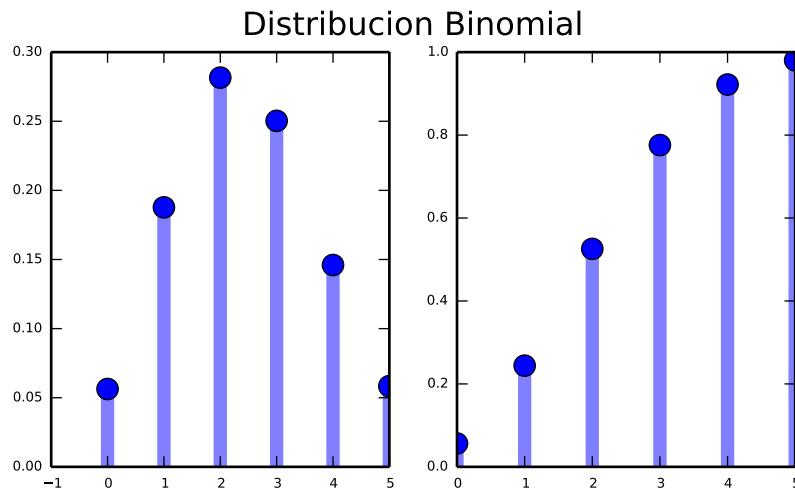
```
import numpy as np
import matplotlib.pyplot as plt

n, p = 10, 0.25
x = np.arange(binom.ppf(0.01, n, p), binom.ppf(0.99, n, p))
fig = plt.figure(figsize=(5, 2.7))
ax = fig.add_subplot(1, 2, 1)
```

```

ax.plot(x, binom.pmf(x, n, p), 'bo', ms=8, label='binom pmf')
ax.vlines(x, 0, binom.pmf(x, n, p), colors='b', lw=5, alpha=0.5)
for tick in ax.xaxis.get_major_ticks():
    tick.label.set_fontsize(5)
for tick in ax.yaxis.get_major_ticks():
    tick.label.set_fontsize(5)
ax = fig.add_subplot(1,2,2)
ax.plot(x, binom.cdf(x, n, p), 'bo', ms=8, label='binom pmf')
ax.vlines(x, 0, binom.cdf(x, n, p), colors='b', lw=5, alpha=0.5)
for tick in ax.xaxis.get_major_ticks():
    tick.label.set_fontsize(5)
for tick in ax.yaxis.get_major_ticks():
    tick.label.set_fontsize(5)
fig.suptitle('Distribucion Binomial')
plt.show()

```



Ejemplo: número de bolas rojas extraídas de una urna con reposición

Tenemos una urna con 100 bolas de las cuales 40 son rojas y 60 son blancas. Extraemos al azar una bola, anotamos su color y la devolvemos a (reponemos en) la urna.

Supongamos que repetimos este proceso $n = 10$ reponiendo en cada ocasión la bola extraída.

Consideremos la variable aleatoria X como el número de bolas rojas extraídas (con reposición) en $n = 10$ repeticiones del mismo experimento de Bernoulli.

Bajo estas condiciones repetimos $n = 10$ veces el mismo experimento de Bernoulli con probabilidad de éxito (sacar bola roja)

$$P(\text{Roja}) = P(x \geq 0) = p = \frac{40}{100} = 0.4.$$

Así que la variable X , que es el número de bolas rojas extraídas de la urna (con reposición) en $n = 10$ ocasiones, sigue una ley binomial $B(n = 10, p = 0.4)$.

Nos preguntamos:

1. ¿Cuál es la probabilidad de que saquemos exactamente 4 bolas rojas?
2. ¿Cuál es la probabilidad de que saquemos al menos 4 bolas rojas?
3. ¿Cuál es la probabilidad de que saquemos menos de 3 bolas rojas?
4. ¿Cuál es el valor esperado del número de bolas rojas?
5. ¿Cuál es la desviación típica del número de bolas rojas?

Solución 1. ¿Cuál es la probabilidad de que saquemos exactamente 4 rojas?

Utilizando la función de probabilidad, tenemos que:

$$\begin{aligned} P(X = 4) &= \binom{10}{4} \cdot 0,4^4 \cdot (1 - 0,4)^{10-4} = \frac{10!}{(10-4)! \cdot 4!} \cdot 0,4^4 \cdot 0,6^6 \\ &= \frac{7 \cdot 8 \cdot 9 \cdot 10}{1 \cdot 2 \cdot 3 \cdot 4} \cdot 0,4^4 \cdot 0,6^6 = 0,2508227. \end{aligned}$$

Con R:

```
dbinom(4, size=10, prob = 0.4)
```

```
## [1] 0.2508227
```

Solución 2. ¿Cuál es la probabilidad de que saquemos al menos 4 bolas rojas?

La probabilidad de sacar al menos 4 rojas se expresa como $P(X \geq 4) = 1 - P(X < 4) = 1 - P(X \leq 3)$:

$$\begin{aligned} P(X \leq 3) &= P(X = 0) + P(X = 1) + P(X = 2) + P(X = 3) \\ &= \binom{10}{0} \cdot 0,4^0 \cdot (1 - 0,4)^{10-0} + \binom{10}{1} \cdot 0,4^1 \cdot (1 - 0,4)^{10-1} \\ &\quad + \binom{10}{2} \cdot 0,4^2 \cdot (1 - 0,4)^{10-2} + \binom{10}{3} \cdot 0,4^3 \cdot (1 - 0,4)^{10-3} \\ &= 0,3822806. \end{aligned}$$

Con R:

```
pbinom(3, 10, 0.4)
```

```
## [1] 0.3822806
```

Así que

$$P(X \geq 4) = 1 - P(X < 4) = P(X \leq 3) = 1 - 0,3822806 = 0,6177194.$$

Otra manera usando R sería:

```
1-pbinom(3, 10, 0.4)
```

```
## [1] 0.6177194
```

Aunque en estos casos el parámetro `lower.tail = FALSE` es sin duda nuestra mejor opción:

```
pbinom(3, 10, 0.4, lower.tail = FALSE)
```

```
## [1] 0.6177194
```

Solución 3. ¿Cuál es la probabilidad de que saquemos menos de 3 bolas rojas?

$$\begin{aligned} P(X < 3) &= P(X \leq 2) = P(X = 0) + P(X = 1) + P(X = 2) \\ &= \binom{10}{0} \cdot 0,4^0 \cdot (1 - 0,4)^{10-0} + \binom{10}{1} \cdot 0,4^1 \cdot (1 - 0,4)^{10-1} \\ &\quad + \binom{10}{2} \cdot 0,4^2 \cdot (1 - 0,4)^{10-2} \\ &= 0,1672898. \end{aligned}$$

En R:

```
dbinom(0,10,0.4)+dbinom(1,10,0.4)+dbinom(2,10,0.4)
```

```
## [1] 0.1672898
```

```
pbinom(2,10,0.4)
```

```
## [1] 0.1672898
```

Solución 4. ¿Cuál es el valor esperado del número de bolas rojas?

Como X es una $B(n = 10, p = 0,4)$ sabemos que

$$E(X) = n \cdot p = 10 \cdot 0,4 = 4.$$

Aunque en Python tenemos la función `stats` que nos lo calcula directamente:

```
print("E(X) = {}".format(m=binom.stats(n = 10, p = 0.4, moments='m')))
```

```
## E(X) = 4.0
```

Solución 5. ¿Cuál es la desviación típica del número de bolas rojas?

La varianza es:

$$Var(X) = n \cdot p \cdot (1 - p) = 10 \cdot 0,4 \cdot 0,6 = 2,4.$$

Por lo tanto, la desviación típica es:

$$\sqrt{Var(X)} = \sqrt{2,4} = 1,5491933.$$

Aunque en Python tenemos la función `stats` que nos lo calcula directamente:

```
print("Var(X) = {}".format(v=binom.stats(n = 10, p = 0.4, moments='v')))
```

```
## Var(X) = 2.4
```

3.1.3. Distribución geométrica

Todos hemos jugado a, por ejemplo, tirar una moneda hasta que obtengamos la primera cara.

O también tirar una pelota a una canasta de baloncesto hasta obtener la primera canasta.

Desde otro punto de vista, también podemos intentar modelar el número de veces que accionamos una interruptor y la bombilla se ilumina hasta que falla.

O también, el número de veces que un cajero automático nos da dinero hasta que falla.

La **modelización de este tipo de problemas se consigue con la llamada distribución geométrica**.

Distribución geométrica

Repitamos un experimento Bernoulli, de parámetro p , de forma independiente hasta obtener el primer éxito.

Sea X la v.a. que cuenta el número de fracasos antes del primer éxito. Por ejemplo, que hayamos tenido x fracasos será una cadena de x fracasos culminada con un éxito. Más concretamente

$$P(\overbrace{FFF \dots F}^x E) = P(F)^x \cdot P(E) = (1-p)^x \cdot p = q^x \cdot p.$$

Su función de probabilidad es:

$$P_X(x) = P(X = x) = \begin{cases} (1-p)^x \cdot p, & \text{si } x = 0, 1, 2, \dots, \\ 0, & \text{en otro caso.} \end{cases}$$

La v.a. definida anteriormente diremos que sigue una distribución geométrica de parámetro p . La denotaremos por $Ge(p)$. Su dominio será: $D_X = \{0, 1, 2, \dots\}$.

Calculemos como ejemplo $P(X \leq 3)$. Por la propiedad de la probabilidad del suceso complementario tenemos que

$$P(X \leq 3) = 1 - P(X > 3) = 1 - P(X \geq 4)$$

Efectivamente, el complementario del evento $X \leq 3$ nos dice que hemos fracasado más de tres veces hasta conseguir el primer éxito, es decir, **hemos fracasado 4 o más veces**. Podemos simbolizar dicho evento de la forma siguiente:

$$\{X > 3\} = \{X \geq 4\} = \{FFFF\}$$

Ahora, al ser los intentos independientes, tenemos que:

$$\begin{aligned} P(X > 3) &= P(\{FFFF\}) = P(F) \cdot P(F) \cdot P(F) \cdot P(F) \\ &= (1-p) \cdot (1-p) \cdot (1-p) \cdot (1-p) = (1-p)^{3+1} = (1-p)^4. \end{aligned}$$

El valor de la función de distribución de X en $x = 3$ será, pues:

$$F_X(3) = P(X \leq 3) = 1 - P(X > 3) = 1 - (1-p)^{3+1}.$$

Generalizando el resultado anterior a cualquier entero positivo $k = 0, 1, 2, \dots$, tenemos:

$$F_X(k) = P(X \leq k) = 1 - (1-p)^{k+1}, \text{ si } k = 0, 1, 2, \dots$$

En general, tendremos que:

$$F_X(x) = P(X \leq x) = \begin{cases} 0, & \text{si } x < 0, \\ 1 - (1-p), & \text{si } k = 0 \leq x < 1, \\ 1 - (1-p)^2, & \text{si } k = 1 \leq x < 2, \\ 1 - (1-p)^3, & \text{si } k = 2 \leq x < 3, \\ 1 - (1-p)^{k+1}, & \text{si } \begin{cases} k \leq x < k+1, \\ \text{para } k = 0, 1, 2, \dots \end{cases} \end{cases}$$

De forma más compacta, tendremos que

$$F_X(x) = P(X \leq x) = \begin{cases} 0, & \text{si } x < 0, \\ 1 - (1-p)^{k+1}, & \text{si } \begin{cases} k \leq x < k+1, \\ \text{para } k = 0, 1, 2, \dots \end{cases} \end{cases}$$

Notemos que el límite de la función de distribución es:

$$\lim_{k \rightarrow +\infty} F_X(k) = \lim_{k \rightarrow +\infty} 1 - (1-p)^{k+1} = 1,$$

ya que $0 < 1-p < 1$.

Sumas derivadas series geométricas

Recordemos las propiedades siguientes del tema de variables aleatorias:

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

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

Esperanza de una v.a. $Ge(p)$

Recordemos que $P(X = x) = (1-p)^x \cdot p$ si $x = 0, 1, 2, \dots$ y aplicando la fórmula anterior con $r = 1-p$, tenemos:

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

Valor $E(X^2)$ de una v.a. $Ge(p)$

$$\begin{aligned} E(X^2) &= \sum_{x=0}^{+\infty} x^2 \cdot P_x(x) = \sum_{x=1}^{+\infty} x^2 \cdot (1-p)^x \cdot p \\ &= \sum_{x=1}^{+\infty} (x \cdot (x-1) + x) \cdot (1-p)^x \cdot p \\ &= \sum_{x=1}^{+\infty} x \cdot (x-1) \cdot (1-p)^x \cdot p + \sum_{x=1}^{+\infty} x \cdot (1-p)^x \cdot p \\ &= (1-p)^2 \cdot p \cdot \sum_{x=2}^{+\infty} x \cdot (x-1) \cdot (1-p)^{x-2} \\ &\quad + (1-p) \cdot p \sum_{x=1}^{+\infty} x \cdot (1-p)^{x-1} \\ &= (1-p)^2 \cdot p \cdot \sum_{x=2}^{+\infty} x \cdot (x-1) \cdot (1-p)^{x-2} \\ &\quad + (1-p) \cdot p \sum_{x=1}^{+\infty} x \cdot (1-p)^{x-1} \\ &= p \cdot (1-p)^2 \frac{2}{(1-(1-p))^3} + (1-p) \cdot p \frac{1}{(1-(1-p))^2} \\ &= p \cdot (1-p)^2 \frac{2}{p^3} + (1-p) \cdot p \frac{1}{p^2} \\ &= \frac{2 \cdot (1-p)^2}{p^2} + \frac{1-p}{p}. \end{aligned}$$

Varianza de una v.a. $Ge(p)$

$$\begin{aligned}
 Var(X) &= E(X^2) - E(X)^2 = \frac{2 \cdot (1-p)^2}{p^2} + \frac{1-p}{p} - \left(\frac{1-p}{p}\right)^2 \\
 &= \frac{2 \cdot (1-p)^2 + p \cdot (1-p) - (1-p)^2}{p^2} = \frac{(1-p)^2 + p \cdot (1-p)}{p^2} \\
 &= \frac{1-2 \cdot p + p^2 + p - p^2}{p^2} \\
 &= \frac{1-p}{p^2},
 \end{aligned}$$

y su desviación típica será

$$\sqrt{Var(X)} = \sqrt{\frac{1-p}{p^2}}.$$

Resumen $Ge(p)$ empezando en 0

$X = \text{Geométrica (empieza en 0)}$	número de fracasos para conseguir el primer éxito
$D_X =$	$\{0, 1, \dots n, \dots\}$
$P_X(x) = P(X = x) =$	$\begin{cases} (1-p)^x \cdot p, & \text{si } x = 0, 1, 2, \dots \\ 0, & \text{en otro caso.} \end{cases}$
$F_X(x) = P(X \leq x) =$	$\begin{cases} 0 & \text{si } x < 0 \\ 1 - (1-p)^{k+1}, & \text{si } \begin{cases} k \leq x < k+1, \\ \text{para } k = 0, 1, \dots \end{cases} \end{cases}$
$E(X) = \frac{1-p}{p}$	$Var(X) = \frac{1-p}{p^2}$

La variable geométrica que cuenta los intentos para obtener el primer éxito

Supongamos que sólo estamos interesados en el **número de intentos** para obtener el primer éxito.

Si definimos Y como número de intentos para obtener el primer éxito, entonces $Y = X + 1$, donde $X \equiv Ge(p)$.

Su dominio es $D_Y = \{1, 2, \dots\}$

La media se incrementa en un intento debido al éxito $E(Y) = E(X + 1) = E(X) + 1 = \frac{1-p}{p} + 1 = \frac{1}{p}$.

La varianza es la misma $Var(Y) = Var(X + 1) = Var(X) = \frac{1-p}{p^2}$.

Resumen $Ge(p)$ comenzando en 1

Y geométrica (que cuenta el éxito empieza en 1)	número de INTENTOS para OBTENER el primer éxito
$D_Y =$	$\{1, 2, \dots n, \dots\}$
$P_Y(y) = P(Y = y) =$	$\begin{cases} (1-p)^{y-1} \cdot p, & \text{si } y = 1, 2, 3, \dots \\ 0, & \text{en otro caso.} \end{cases}$
$F_Y(y) = P(Y \leq y) =$	$\begin{cases} 0, & \text{si } y < 1 \\ 1 - (1-p)^k, & \text{si } \begin{cases} k \leq y < k+1, \\ \text{para } k = 1, 2, 3, \dots \end{cases} \end{cases}$
$E(X) = \frac{1}{p}$	$Var(X) = \frac{1-p}{p^2}$

Propiedad de la falta de memoria

Sea X una v.a. discreta con dominio $D_X = \{0, 1, 2, \dots\}$, con $P(X = 0) = p$.

Entonces X sigue una ley $Ge(p)$ si, y sólo si,

$$P(X > k + j | X \geq j) = P(X > k)$$

para todo $k, j = 0, 1, 2, 3, \dots$

Demostración

Si X es geométrica, entonces el lado derecho de la igualdad es

$$P(X > k) = 1 - P(X \leq k) = 1 - (1 - (1 - p)^{k+1}) = (1 - p)^{k+1},$$

y el lado de izquierdo es:

$$\begin{aligned} P(X > k + j | X \geq j) &= \frac{P(\{X > k+j\} \cap \{X \geq j\})}{P(X \geq j)} = \frac{P(X > k+j)}{P(X \geq j)} = \frac{1 - P(X \leq k+j)}{1 - P(X \leq j-1)} \\ &= \frac{1 - (1 - (1-p)^{k+j+1})}{1 - (1 - (1-p)^{j-1+1})} = \frac{(1-p)^{k+j+1}}{(1-p)^j} = (1-p)^{k+1}, \end{aligned}$$

lo que demuestra la igualdad.

Para demostrar el recíproco, tomemos $j = 1$ y $k \geq 0$. Entonces, por la propiedad de la pérdida de memoria:

$$P(X > k + 1 | X \geq 1) = P(X > k)$$

Como $P(X = 0) = p$, tenemos que $P(X \geq 1) = 1 - P(X < 1) = 1 - P(X = 0) = 1 - p$.

Combinando las igualdades, tenemos que:

$$P(X > k + 1 | X \geq 1) = \frac{P(X > k + 1, X \geq 1)}{P(X \geq 1)} = \frac{P(X > k + 1)}{P(X \geq 1)} = P(X > k).$$

Así podemos poner que

$$\begin{aligned} P(X > k + 1) &= P(X \geq 1) \cdot P(X > k) = (1 - P(X < 1)) \cdot P(X > k) \\ &= (1 - P(X = 0)) \cdot P(X > k) = (1 - p) \cdot P(X > k). \end{aligned}$$

En general tenemos que:

$$P(X > k + 1) = (1 - p) \cdot P(X > k).$$

Del mismo modo para $j = 2$,

$$P(X > k + 2) = (1 - p) \cdot P(X > k + 1).$$

Restando la primera igualdad de la última obtenemos:

$$P(X > k + 1) - P(X > k + 2) = (1 - p) \cdot P(X > k) - (1 - p) \cdot P(X > k + 1),$$

de donde operando en cada lado de la igualdad obtenemos la recurrencia:

$$[1 - P(X \leq k + 1)] - [1 - P(X \leq k + 2)] = (1 - p) \cdot [P(X > k) - P(X > k + 1)]$$

Ahora operando,

$$\begin{aligned} P(X \leq k+2) - P(X \leq k+1) &= (1-p) \cdot [1 - P(X \leq k) - (1 - P(X \leq k+1))], \\ P(X = k+2) &= (1-p) \cdot [P(X \leq k+1) - P(X \leq k)], \\ P(X = k+2) &= (1-p) \cdot P(X = k+1). \end{aligned}$$

De forma similar, obtenemos

$$P(X = k+1) = (1-p) \cdot P(X = k).$$

Utilizando la recurrencia anterior, podemos calcular todas las probabilidades $P(X = k)$ a partir de la $P(X = 0) = p$:

$$\begin{aligned} P(X = 0) &= p, \\ P(X = 1) &= P(X = 0 + 1) = (1-p) \cdot P(X = 0) = (1-p) \cdot p, \\ P(X = 2) &= P(X = 1 + 1) = (1-p) \cdot P(X = 1) = (1-p) \cdot (1-p) \cdot p = (1-p)^2 \cdot p, \\ &\vdots \\ P(X = k) &= P(X = (k-1) + 1) = (1-p) \cdot P(X = k-1) = (1-p) \cdot (1-p)^{k-1} \cdot p \\ &= (1-p)^k \cdot p, \end{aligned}$$

lo que demuestra el recíproco, es decir, que X es $Geom(p)$.

Observación: Interpretación de la propiedad de la falta de memoria

La propiedad de la falta de memoria

$$P(X > k+j | X \geq j) = P(X > k),$$

significa que, aunque ya llevemos al menos j fracasos, la probabilidad de que fracasemos k veces más no disminuye, es la misma que era cuando empezamos el experimento.

A este efecto se le suele etiquetar con la frase **el experimento carece de memoria** o es un **experimento sin memoria** (*Memoryless Property*).

Ejemplo falta de memoria

Un ejemplo muy sencillo nos aclarará el alcance de esta propiedad es el ejercicio siguiente:

Ejercicio: la llave que abre la puerta

Tenemos un llavero con 10 llaves, sólo una de ellas abre una puerta. Cada vez que probamos una llave y falla olvidamos que llave hemos probado. ¿Cuál es la probabilidad de que si ya lo hemos intentado 5 veces necesitemos más de 4 intentos adicionales para abrir la puerta?

Tomemos $k = 4, j = 5$, aplicando la propiedad de la falta de memoria

$$P(X > 4 + 5 | X \geq 5) = P(X > 4)$$

Después de 5 fracasos no estamos “más cerca” de abrir la puerta. La propiedad de la falta de memoria nos dice que en **después de cada intento es como si empezásemos de nuevo a abrir la puerta**. Tras 5 fracasos la probabilidad de que fallemos más de 4 veces más es la misma que cuando lo intentamos la primera vez.

¿Cuál es el número esperado de fracasos hasta abrir la puerta?

$$E(X) = \frac{1-p}{p} = \frac{1 - \frac{1}{10}}{\frac{1}{10}} = \frac{\frac{9}{10}}{\frac{1}{10}} = 9.$$

La varianza es

$$Var(X) = \frac{1-p}{p^2} = \frac{1 - \frac{1}{10}}{\left(\frac{1}{10}\right)^2} = \frac{\frac{9}{10}}{\frac{1}{100}} = 90.$$

La desviación típica es $\sqrt{90} = 9,486833$.

Ejemplo: partidos hasta que el Barça gana al Madrid

Los partidos Real Madrid vs FC Barcelona de la liga española se suelen denominar **El Clásico**, sean en el Bernabeu (estadio del Real Madrid) o en el Camp Nou (estadio del Barça).

Sea X la variable que cuenta el número de veces consecutivas que en un partido de fútbol de la liga el Barça no gana al Madrid sea en el Camp Nou o el Bernabeu.

Nuestra amiga Aina es muy culé (hincha del Barça) y quiere averiguar cuántos partidos consecutivos de **El Clásico** tiene que ver hasta ver ganar al Barça por primera vez.

Le interesa estimar cuánto le va a costar este capricho. Tendrá que comprar las entradas y pagar los viajes de Barcelona a Madrid.

En datos históricos de **El clásico** en la wikipedia están los datos hasta el 3 de marzo de 2019: se han jugado en total 178 **Clásicos** donde el Real Madrid ganó en 72 ocasiones, el Barça en 72 y empataron 34 veces.

La pregunta es: ¿Cuántos partidos se tienen que jugar de media para ver ganar al Barça por primera vez?

Con los datos anteriores, podemos estimar que la probabilidad de que el Barça gane un clásico cualquiera es:

$$P(\text{Barça}) = \frac{72}{178} = 0,4045.$$

Por tanto, podemos modelar la variable X con una ley geométrica con probabilidad de éxito $p = P(\text{Barça}) = \frac{72}{178}$.

El número de partidos esperado para que el Barça gane por primera vez es:

$$E(X) = \frac{1-p}{p} = \frac{1 - 0,4045}{0,4045} = 1,4722,$$

con una varianza de:

$$Var(X) = \frac{1-p}{p^2} = \frac{1 - 0,4045}{0,4045^2} = 3,6397$$

y desviación típica:

$$\sqrt{3,6397} = 1,9078.$$

Cálculos con R

Veamos los cálculos básicos con R para la distribución geométrica $Ge(p = 0,25)$. R implementa la geométrica que cuenta el número de fracasos, $P(X = 0) = (1 - 0,25)^0 \cdot 0,25^1 = 0,25$:

```
dgeom(0,prob=0.25)
```

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

$$P(X \leq 0) = 1 - (1 - 0,25)^{0+1} = 1 - 0,75 = 0,25:$$

```
pgeom(0,prob=0.25)
```

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

$$P(X \leq 4) = 1 - (1 - 0,25)^{4+1} = 1 - 0,75 = 1 - 0,75^5 = 0,7626953:$$

```
pgeom(4,prob=0.25)
```

```
## [1] 0.7626953
```

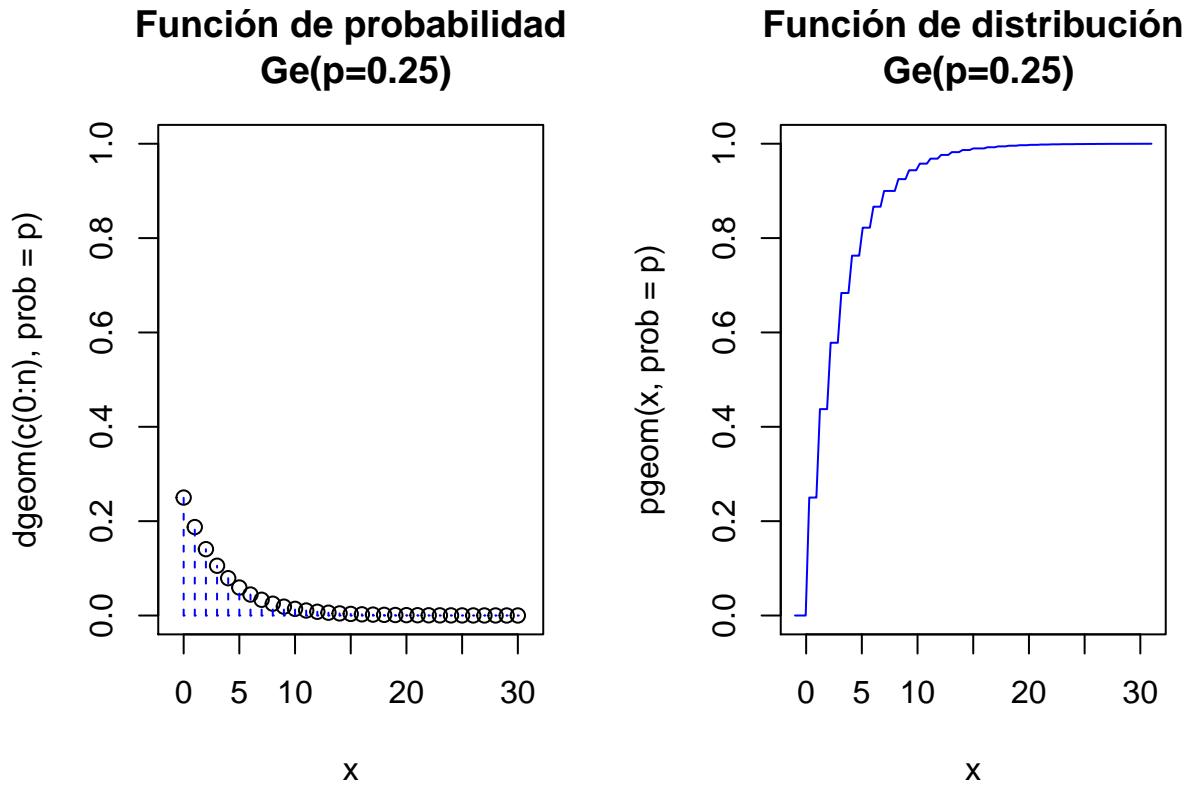
Una muestra aleatoria de tamaño 25 de una $Ge(0,25)$:

```
rgeom(n=25,prob=0.25)
```

```
## [1] 5 4 1 6 10 0 0 10 7 0 6 2 1 3 0 2 5 0 0 5 5 3 3 2 2
```

Gráficos con R

```
par(mfrow=c(1,2))
x=c(0:10)
plot(x=x,y=dgeom(x,prob=0.25),
      ylim=c(0,1),xlim=c(-1,11),xlab="x",
      main="Función de probabilidad\n Ge(p=0.25)")
lines(x=rep(0:10,each=2),y=aux, type = "h", lty = 2,col="blue")
aux0=dgeom(c(0:10),prob=0.25)
ceros=rep(0,21)
ceros
aux=ceros
aux[2*(c(1:11))]<-aux0
curve(pgeom(x,prob=0.25),
      xlim=c(-1,10),col="blue",
      main="Función de distribución\n Ge(p=0.25)")
par(mfrow=c(1,1))
```



Gráficas interactivas de la distribución Geométrica

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_geom", label = "Probabilidad de éxito:",
           min = 0.01, max = 0.99, value = 0.25, step = 0.01)
renderPlot({
  par(mfrow=c(1,2))
  p=input$p_geom
  n=30
  aux=rep(0,(n+1)*2)
  aux[seq(2,(n+1)*2,2)]=dgeom(c(0:n),prob=p)
  plot(x=c(0:n),y=dgeom(c(0:n),prob=p),
       ylim=c(0,1),xlim=c(-1,n+1),xlab="x",
       main=paste0(c("Función de probabilidad\n Ge(p=",p,")"),collapse = ""))
  lines(x=rep(0:n,each=2),y=aux, type = "h", lty = 2,col="blue")
  curve(pgeom(x,prob=p),
         xlim=c(-1,n+1),col="blue",
         main=paste0(c("Función de distribución\n Ge(p=",p,")"),collapse = ""))
  par(mfrow=c(1,1))
})
```

Cálculos con Python

Veamos los cálculos básicos con python para la distribución geométrica $Ge(p = 0,25)$. `scipy.stats` implementa la distribución geométrica que cuenta el número intentos, así que empieza en 1.

Cargamos la función de la librería

```
from scipy.stats import geom
```

La función de probabilidad es `geom.pmf(x,p,loc=0)=geom.pmf(x,p)`. Es una geométrica que cuenta el número de intentos para obtener el primer éxito y el valor por defecto del último parámetro es `loc=0`.

Si queremos la que cuenta el número de fracasos para obtener el primer éxito (la geométrica que empieza en 0) tenemos que usar `geom.pmf(x,p,loc=-1)`.

Es decir `geom.pmf(x,p,loc=-1)=geom.pmf(x-1,p,loc=0)`

Veamos pues los cálculos para la $Ge(p)$ que empieza en 0.

$$P(X = 0) = (1 - 0,25)^0 \cdot 0,25^1 = 0,25:$$

```
geom.pmf(0,p=0.25,loc=-1)
```

```
## 0.25
```

$$P(X \leq 0) = 1 - (1 - 0,25)^{0+1} = 1 - 0,75 = 0,25:$$

```
geom.cdf(0,p=0.25,loc=-1)
```

```
## 0.25
```

$$P(X \leq 4) = 1 - (1 - 0,25)^{4+1} = 1 - 0,75 = 1 - 0,75^5 = 0,7626953:$$

```
geom.cdf(4,p=0.25,loc=-1)
```

```
## 0.7626953125
```

Una muestra aleatoria de tamaño 25 de una $Ge(0,25)$:

```
geom.rvs(p=0.25, size=20, loc=-1)
```

```
## array([ 0,  5, 12,  4,  4,  0,  6,  1,  2,  4,  1,  0,  3,  2,  5,  5,  8,
##        3,  0, 13])
```

Ejercicio

¿Qué probabilidades son las que calcula el siguiente código y qué tipo de variables geométricas son?

```
geom.cdf(range(5),p=0.3,loc=0)
```

```
## array([ 0.      ,  0.3     ,  0.51    ,  0.657   ,  0.7599])
```

```
geom.cdf(range(5),p=0.3,loc=-1)
```

```
## array([ 0.3     ,  0.51    ,  0.657   ,  0.7599 ,  0.83193])
```

Cálculos con Python de la esperanza y varianza

Con Python también podemos calcular directamente algunos parámetros asociados a una función de distribución predefinida

```
geom.stats(p=0.25, loc=0, moments='mv')
```

```
## (array(4.0), array(12.0))
geom.stats(p=0.25, loc=-1, moments='mv')
```

```
## (array(3.0), array(12.0))
```

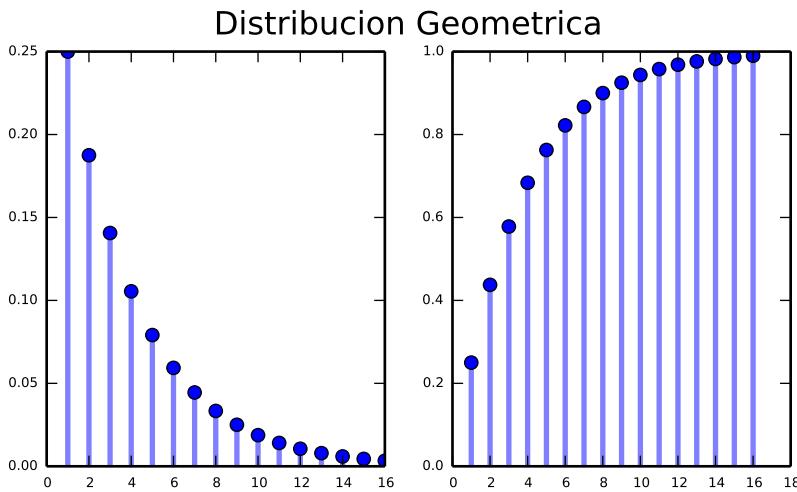
Ejercicio

Comprobad que las medias y las varianzas calculadas en el código anterior corresponden a una $Ge(p = 0,3)$ empezando en 1 a una $Ge(p = 0,3)$ empezando en 0.

¿Son las varianzas siempre iguales?

Gráficos con Python

```
p = 0.25
x = np.arange(geom.ppf(0.01, p), geom.ppf(0.99, p))
fig = plt.figure(figsize=(5, 2.7))
ax = fig.add_subplot(1,2,1)
ax.plot(x, geom.pmf(x, p), 'bo', ms=5, label='geom pmf')
ax.vlines(x, 0, geom.pmf(x, p), colors='b', lw=2, alpha=0.5)
for tick in ax.xaxis.get_major_ticks():
    tick.label.set_fontsize(5)
for tick in ax.yaxis.get_major_ticks():
    tick.label.set_fontsize(5)
ax = fig.add_subplot(1,2,2)
ax.plot(x, geom.cdf(x, p), 'bo', ms=5, label='geom pmf')
ax.vlines(x, 0, geom.cdf(x, p), colors='b', lw=2, alpha=0.5)
for tick in ax.xaxis.get_major_ticks():
    tick.label.set_fontsize(5)
for tick in ax.yaxis.get_major_ticks():
    tick.label.set_fontsize(5)
fig.suptitle('Distribucion Geometrica')
plt.show()
```



3.1.4. Distribución binomial negativa

El problema de la puerta con dos cerraduras

Supongamos que disponemos de 10 llaves distintas y tenemos que abrir una puerta con **dos cerraduras**.

Comenzamos por la primera cerradura, de tal forma que cada vez olvidamos qué llave hemos probado.

Una vez abierta la primera cerradura probamos de igual forma con la segunda hasta que también la abrimos.

Sea X la v.a. que cuenta el número de fracasos hasta abrir la puerta.

Acertar una llave de la puerta es un experimento Bernoulli con probabilidad de éxito $p = 0.1$. Lo repetiremos hasta obtener 2 éxitos.

En general, tendremos un experimento de Bernoulli con probabilidad de éxito $0 < p < 1$ tal que:

- Repetimos el experimento hasta obtener el n -ésimo éxito ¡¡abrir la maldita puerta!!.
- Sea X la v.a. que cuenta el número fallos hasta abrir la puerta, es decir, hasta conseguir el n -ésimo éxito. Notemos que no contamos los éxitos, solo contamos los fracasos.

Si representamos como es habitual un suceso como una cadena de F's y E's, para $n = 2$, algunos sucesos elementales serán:

$$\{EE, FEE, EFE, FFEE, FEFE, EFFE, FFFEE, FFEFE, FEFFE, EFFF\}.$$

Calculemos algunas probabilidades para $n = 2$:

$$\begin{aligned} P(X = 0) &= P(\{EE\}) = p^2, \\ P(X = 1) &= P(\{FEE, EFE\}) = 2 \cdot (1-p) \cdot p^2, \\ P(X = 2) &= P(\{FFEE, FEFE, EFFE\}) = 3 \cdot (1-p)^2 \cdot p^2, \\ P(X = 3) &= P(\{FFFEE, FFEFE, FEFFE, EFFF\}) = 4 \cdot (1-p)^3 \cdot p^2. \end{aligned}$$

En general, su función de probabilidad es

$$P_X(k) = P(X = k) = \begin{cases} \binom{k+n-1}{n-1} \cdot (1-p)^k \cdot p^n, & \text{si } k = 0, 1, \dots \\ 0, & \text{en otro caso.} \end{cases}$$

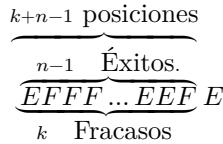
Una v.a. con este tipo de distribución recibe el nombre de **binomial negativa** y la denotaremos por $BN(n, p)$.

Notemos que $BN(1, p) = Ge(p)$.

Demostración

Justifiquemos el resultado. Sea X una $BN(n, p)$ y sea $k = 0, 1, 2, \dots$

$$P(X = k) = P(\text{Todas las cadenas de E's y F' con } k \text{ F, con } n \text{ E y acabadas en E})$$



De estas cadenas hay tantas como maneras de elegir de entre las $k + n - 1$ primeras posiciones $n - 1$ para colocar los éxitos. Esta cantidad es el número binomial $\binom{k+n-1}{n-1}$.

Números binomiales negativos

Dados dos enteros positivos n y k , se define el número binomial negativo como:

$$\binom{-n}{k} = \frac{(-n)(-n-1)\cdots(-n-k+1)}{k!}.$$

Los números binomiales negativos generalizan la fórmula de Newton para exponentes negativos:

$$(t+1)^{-n} = \sum_{k=0}^{+\infty} \binom{-n}{k} t^k.$$

R usa la función `choose` para calcular números binomiales, sean negativos o no. Veámoslo con un ejemplo:

$$\binom{-6}{4} = \frac{-6 \cdot (-6-1) \cdots (-6-2) \cdot (-6-3)}{4!} = \frac{-6 \cdot (-7) \cdot (-8) \cdot (-9)}{24} = \frac{3024}{24} = 126.$$

Si realizamos el cálculo con R obtenemos el mismo resultado:

```
choose(-6, 4)
```

```
## [1] 126
```

Esperanza de una $BN(n, p)$

Su **esperanza** es:

$$E(X) = \sum_{k=0}^{+\infty} k \cdot \binom{k+n-1}{n-1} \cdot (1-p)^k \cdot p^n = n \cdot \frac{1-p}{p}.$$

La **esperanza de X^2** es:

$$E(X^2) = \sum_{k=0}^{+\infty} k^2 \cdot \binom{k+n-1}{n-1} \cdot (1-p)^k \cdot p^n = n \cdot \frac{1-p}{p^2} + \left(n \cdot \frac{1-p}{p} \right)^2.$$

Varianza de una $BN(n, p)$

Por último, la **varianza** es:

$$Var(X) = E(X^2) - E(X)^2 = n \cdot \frac{1-p}{p^2} + \left(n \cdot \frac{1-p}{p} \right)^2 - \left(n \cdot \frac{1-p}{p} \right)^2 = n \cdot \frac{1-p}{p^2},$$

y la desviación típica es:

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

Resumen Binomial Negativa $BN(n, p)$

$X = BN(n, p)$	Número de fracasos antes de conseguir el n -ésimo éxito. Probabilidad de éxito p
$D_X =$	$\{0, 1, 2, 3, \dots\}$
$P_X(k) = P(X = k) =$	$\begin{cases} \binom{k+n-1}{n-1} \cdot (1-p)^k \cdot p^n, & \text{si } k = 0, 1, \dots \\ 0, & \text{en otro caso.} \end{cases}$
$F_X(x) = P(X \leq x) =$	$\begin{cases} 0, & \text{si } x < 0 \\ \sum_{i=0}^k P(X = i) & \text{si } \begin{cases} k \leq x < k+1, \\ k = 0, 1, 2, \dots \end{cases} \end{cases}$ Calcular la suma o utilizar funciones de R o Python.
$E(X) = n \cdot \frac{1-p}{p}$	$Var(X) = n \cdot \frac{1-p}{p^2}$

Ejercicio: Puerta con dos cerraduras

Recordemos nuestra puerta con dos cerraduras que se abren secuencialmente. Tenemos un manojo de 10 llaves casi idénticas de manera que cada vez que probamos una llave olvidamos qué llave hemos usado.

Sea X la v.a que nos da el número de intentos fallidos hasta abrir la puerta.

Estamos interesado en modelar este problema. La preguntas son:

1. ¿Cuál es la distribución de probabilidad de X la v.a que nos da el número fallos hasta abrir la puerta?
2. ¿Cuál es la función de probabilidad y de distribución de X ?
3. ¿Cuál es la probabilidad de fallar exactamente 5 veces antes de abrir la puerta?
4. ¿Cuál es la probabilidad de fallar más de 4?
5. ¿Cuál es el número esperado de fallos? ¿Y su desviación típica?

Solución 1. ¿Cuál es la distribución de probabilidad de X la v.a que nos da el número fallos hasta abrir la puerta?

Bajo estosas condiciones tenemos que la probabilidad de “éxito” de cada intento es $p = \frac{1}{10} = 0,1$. Como cada vez *olvidamos* qué llave hemos probado, cada intento será independiente del anterior.

Así que la variable X que queremos modelar cuenta el número de fallos de repeticiones sucesivas e independientes de un experimento $Ber(p = 0,1)$ hasta conseguir 2 éxitos en un experimento.

Por lo tanto podemos asegurar que X sigue un distribución $BN(n = 2, p = 0,1)$.

Solución 2. ¿Cuál es la función de probabilidad y de distribución de X ?

En general la función de probabilidad de una $BN(n, p)$ es

$$P_X(X = k) = \begin{cases} \binom{k+n-1}{n-1} \cdot (1-p)^k \cdot p^n, & \text{si } k = 0, 1, \dots \\ 0, & \text{en otro caso.} \end{cases}$$

Si aplicamos la expresión anterior para $n = 2$ y $p = 0,1$, obtenemos:

$$P_X(X = k) = \begin{cases} \binom{k+2-1}{2-1} \cdot 0,9^k \cdot 0,1^2, & \text{si } k = 0, 1, 2, \dots \\ 0, & \text{en otro caso.} \end{cases}$$

Simplificando,

$$P_X(X = k) = P(X = k) = \begin{cases} 0,01 \cdot (k+1) \cdot 0,9^k, & \text{si } k = 0, 1, 2, \dots \\ 0 & \text{en otro caso.} \end{cases}$$

La función de distribución en general es

$$F_X(x) = P(X \leq x) = \begin{cases} 0, & \text{si } x < 0, \\ \sum_{i=0}^k \binom{i+n-1}{n-1} \cdot (1-p)^{i+n-1} \cdot p^n, & \text{si } \begin{cases} k \leq x < k+1, \\ k = 0, 1, 2, \dots \end{cases} \end{cases}$$

Simplificando para $n = 2$, $p = 0,1$.

$$F_X(x) = P(X \leq x) = \begin{cases} 0, & \text{si } x < 0, \\ \sum_{i=0}^k 0,01 \cdot (i+1) \cdot 0,9^{i+1}, & \text{si } \begin{cases} k \leq x < k+1, \\ k = 0, 1, 2, \dots \end{cases} \end{cases}$$

Solución 3. ¿Cuál es la probabilidad de fallar exactamente 5 veces antes de abrir la puerta?

$$P(X = 5) = 0,01 \cdot (5+1) \cdot 0,9^5 = 0,06 \cdot 0,9^5 = 0,0354294.$$

Solución 4. ¿Cuál es la probabilidad de fallar más de 4?

Nos piden calcular $P(X > 4) = 1 - P(X \leq 4)$.

Calculemos primero $P(X \leq 4)$:

$$\begin{aligned} P(X \leq 4) &= \sum_{x=0}^4 P(X = x) \\ &= P(X = 0) + P(X = 1) + P(X = 2) + P(X = 3) + P(X = 4) \\ &= 0,01 \cdot (0+1) \cdot 0,9^0 + 0,01 \cdot (1+1) \cdot 0,9^1 + 0,01 \cdot (2+1) \cdot 0,9^2 \\ &\quad + 0,01 \cdot (3+1) \cdot 0,9^3 + 0,01 \cdot (4+1) \cdot 0,9^4 \\ &= 0,01 + 0,018 + 0,0243 + 0,02916 + 0,032805 = 0,114265. \end{aligned}$$

Por lo tanto

$$P(X > 4) = 1 - P(X \leq 4) = 1 - 0,114265 = 0,885735.$$

Solución 5. ¿Cuál es el número esperado de fallos? Y su desviación típica?

Como X sigue una ley $BN(n = 2, p = 0,1)$

$$E(X) = n \cdot \frac{1-p}{p} = 2 \cdot \frac{1-0,1}{0,1} = 18.$$

El número de fallos esperado es 18.

La varianza será:

$$Var(X) = n \cdot \frac{1-p}{p^2} = 2 \cdot \frac{1-0,1}{0,1^2} = 180.$$

La varianza de X es 180 y su desviación típica $\sqrt{180} = 13,41641$.

Cálculos con R

La función de R que calcula la función de probabilidad de la binomial negativa con sus parámetros básicos es:

```
dnbinom(x, size, prob,...)`
```

donde **size** (n) es el número de éxitos y **prob** (p), la probabilidad de éxito.

Así, en el ejemplo de la puerta con dos cerraduras, X es una $BN(n = size = 2, p = prob = 0,1)$. Por ejemplo, $P(X = 5)$ que hemos calculado en el ejemplo anterior, vale:

```
dnbinom(5,size=2,p=0.1)
```

```
## [1] 0.0354294
```

De forma similar calculamos $P(X \leq 4)$, $P(X > 4) = 1 - P(X \leq 4)$ y $P(X > 4)$.

```
pnbinom(4,size=2,p=0.1)
```

```
## [1] 0.114265
```

```
1-pnbinom(4,size=2,p=0.1)

## [1] 0.885735
pnbinom(4,size=2,p=0.1,lower.tail=FALSE)

## [1] 0.885735
```

Cálculos con Python

La función con Python es `nbinom.pmf(k, n, p, loc)`. Hay que cargarla desde `scipy.stats`

```
from scipy.stats import nbinom
```

Recordemos que de nuevo se cumple que

```
nbinom.pmf(k, n, p, loc) = nbinom.pmf(k-loc, n, p)

nbinom.pmf(k=5,n=2,p=0.1)

## 0.035429400000000041
nbinom.pmf(k=5,n=2,p=0.1,loc=0)

## 0.035429400000000041
nbinom.cdf(k=4,n=2,p=0.1)

## 0.11426500000000003
1-nbinom.cdf(k=4,n=2,p=0.1)

## 0.8857349999999994
```

Generemos 100 observaciones aleatorias de una $BN(n = 2, 0, 1)$. Es decir serán las veces que hemos fallado hasta abrir la puerta 100 veces.

```
nbinom.rvs(n=2, p=0.1, size=100)

## array([23, 31, 13, 7, 9, 28, 2, 19, 14, 5, 24, 10, 4, 55, 6, 20, 41,
##        21, 24, 17, 8, 15, 20, 19, 12, 35, 5, 36, 22, 25, 31, 22, 12, 16,
##        20, 9, 1, 4, 24, 27, 28, 6, 4, 2, 14, 22, 2, 28, 24, 22, 6,
##        20, 16, 22, 34, 16, 3, 45, 6, 6, 24, 3, 27, 24, 8, 7, 5, 11,
##        25, 8, 7, 3, 18, 19, 13, 26, 27, 9, 15, 32, 7, 10, 9, 8, 6,
##        11, 36, 9, 36, 2, 15, 2, 29, 15, 12, 5, 75, 6, 1, 25])
```

La **esperanza** y la **varianza** de una $BN(n = 2, 0, 1)$ valen:

```
n, p=2,0.1
params = nbinom.stats(n,p,moments='mv')
print("E(X)={m}".format(m=params[0]))

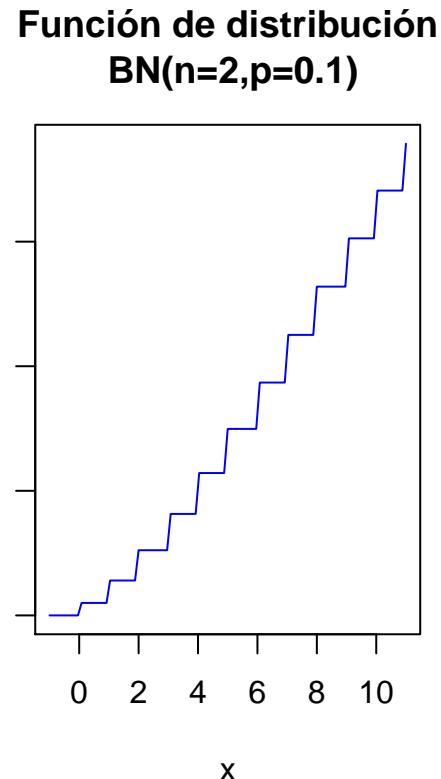
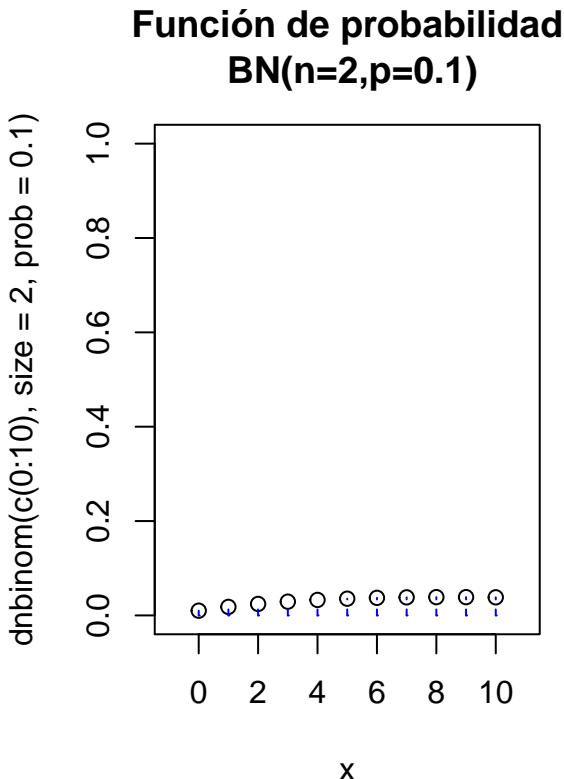
## E(X)=18.0
print("Var(X)={v}".format(v=params[1]))

## Var(X)=180.0
```

Gráficas de la binomial negativa con R

El siguiente código de R dibuja las función de probabilidad y la de distribución de una $BN(n = 2, p = 0,1)$

```
par(mfrow=c(1,2))
aux=rep(0,22)
aux[seq(2,22,2)]=dnbinom(c(0:10),size=2,prob=0.1)
plot(x=c(0:10),y=dnbinom(c(0:10),size=2,prob=0.1),
     ylim=c(0,1),xlim=c(-1,11),xlab="x",
     main="Función de probabilidad\n BN(n=2,p=0.1)")
lines(x=rep(0:10,each=2),y=aux, type = "h", lty = 2,col="blue")
curve(pnbinom(x,size=2,prob=0,1),
      xlim=c(-1,11),col="blue",
      main="Función de distribución\n BN(n=2,p=0.1)")
par(mfrow=c(1,1))
```



Gráficas interactivas binomial negativa

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.

```

fluidPage(
  fluidRow(
    column(6,
      sliderInput("n_nbinom", label = "Número de éxitos n:",
                  min = 1, max = 50, value = 20, step = 1)),
    column(6,
      sliderInput("p_nbinom", label = "Probabilidad de un éxito p:",
                  min = 0.01, max = 0.99, value = 0.8, step = 0.01)
    )
  )
)

renderPlot({
  n=input$n_nbinom
  pr=input$p_nbinom

  par(mfrow=c(1,2))
  aux=rep(0,(n+1)*2)
  aux[seq(2,(n+1)*2,2)]=dnbinom(c(0:n),size=n,prob=pr)
  plot(x=c(0:n),y=dnbinom(c(0:n),size=n,prob=pr),
        ylim=c(0,1),xlim=c(-1,n+1),xlab="x",
        main=paste0(c("Función de probabilidad\n BN(n=",n,",p=",pr,")"),
                    collapse = ""))
  lines(x=rep(0:n,each=2),y=aux, type = "h", lty = 2,col="blue")
  curve(pnbinom(x,size=n,p=pr),
         xlim=c(-1,n+1),col="blue",
         main=paste0(c("Función de distribución\n BN(n=",n,",p=",pr,")"),
                     collapse = ""))
  par(mfrow=c(1,1))
})

```

Ejercicio

Buscad en los manuales de Python cómo se dibuja la función de probabilidad y de distribución de una binomial negativa.

Necesitamos de nuevo más librerías

```

import numpy as np
from scipy.stats import nbinom
import matplotlib.pyplot as plt

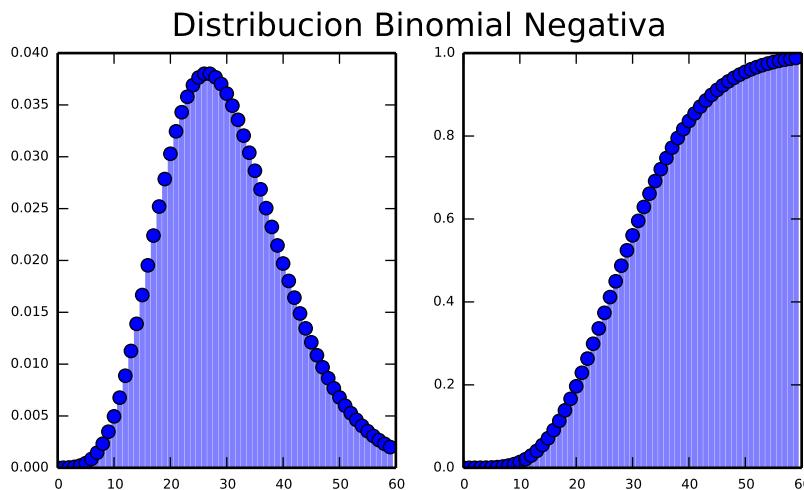
n, p = 10, 0.25
x = np.arange(0,nbinom.ppf(0.99, n, p))
fig = plt.figure(figsize=(5, 2.7))
ax = fig.add_subplot(1,2,1)
ax.plot(x, nbinom.pmf(x, n, p), 'bo', ms=5, label='nbinom pmf')
ax.vlines(x, 0, nbinom.pmf(x, n, p), colors='b', lw=2, alpha=0.5)
for tick in ax.xaxis.get_major_ticks():
  tick.label.set_fontsize(5)

```

```

for tick in ax.yaxis.get_major_ticks():
    tick.label.set_fontsize(5)
ax = fig.add_subplot(1,2,2)
ax.plot(x, nbinom.cdf(x, n, p), 'bo', ms=5, label='nbinom pmf')
ax.vlines(x, 0, nbinom.cdf(x, n, p), colors='b', lw=2, alpha=0.5)
for tick in ax.xaxis.get_major_ticks():
    tick.label.set_fontsize(5)
for tick in ax.yaxis.get_major_ticks():
    tick.label.set_fontsize(5)
fig.suptitle('Distribucion Binomial Negativa')
plt.show()

```



Ejercicio: acceso aleatorio a un sistema con triple clave

Supongamos que tenemos un sistema informático que tiene un programa de seguridad que genera accesos con claves de 3 dígitos 000, 001, ... 999. En total tenemos 1000 posibilidades.

Como una clave de tres dígitos es fácil de romper, proponemos considerar tres claves consecutivas de acceso al sistema, cada una de 3 dígitos.

Para acceder al sistema hay que dar las tres claves de forma consecutiva y por orden.

Es decir hasta que no averiguamos la primera clave no pasamos a la segunda clave.

Supongamos que cada vez que ponemos las dos claves olvidamos el resultado y seguimos poniendo claves al azar hasta adivinar la contraseña.

Así hasta conseguir entrar en el sistema.

Sea X la v.a que nos da el número de fallos antes de entrar en el sistema.

Estamos interesados en modelar este problema. La preguntas son:

1. ¿Cuál es la distribución de probabilidad de X , la v.a que nos da el número de fallos antes de acceder al sistema.

2. ¿Cuál es la función de probabilidad y de distribución del X ?
3. ¿Cuál es la probabilidad de fallar 150 veces antes de acceder en el sistema?
4. ¿Cuál es la probabilidad de fallar más de 150 veces antes de entrar en el sistema?
5. ¿Cuál es el número esperado de fallos antes de acceder al sistema? ¿Y su varianza?

Solución 1. ¿Cuál es la distribución de probabilidad de X , la v.a que nos da el número de fallos antes de acceder al sistema?

Bajo estas dos condiciones tenemos que la probabilidad de “éxito” de cada intento es $p = \frac{1}{1000} = 0,001$. Y como cada vez *olviamos* los dígitos cada intento será independiente del anterior.

Así que la variable X cuenta el número de fracasos independientes hasta conseguir 3 éxitos en un experimento $Ber(p = 0,001)$, por lo tanto X sigue un distribución $BN(n = 3, p = 0,001)$.

Solución 2. ¿Cuál es la función de probabilidad y de distribución del X

En general la función de probabilidad de una $BN(n, p)$ es:

$$P_X(X = x) = P(X = x) = \begin{cases} \binom{x+n-1}{n-1} \cdot (1-p)^x \cdot p^n, & \text{si } x = 0, 1, \dots \\ 0, & \text{en otro caso.} \end{cases}$$

En particular la función de probabilidad de una $BN(n = 3, p = 0,001)$ es

$$P_X(X = x) = P(X = x) = \begin{cases} \binom{x+2}{2} \cdot 0,999^x \cdot 0,001^3 & \text{si } x = 0, 1, 2, \dots \\ 0 & \text{en otro caso.} \end{cases}$$

Solución 3. ¿Cuál es la probabilidad de fallar 150 veces antes de acceder en el sistema?

Nos piden calcular la probabilidad siguiente:

$$P(X = 150) = \binom{152}{2} \cdot 0,999^{150} \cdot 0,001^3.$$

Realizaremos el cálculo anterior con ayuda de R:

```
choose(152,2)*0.999^150*0.001^3
```

```
## [1] 9.876743e-06
```

o, usando la función de R que nos calcula la función de probabilidad:

```
dbinom(150,size=3,p=0.001)
```

```
## [1] 9.876743e-06
```

Si queremos calcular la probabilidad anterior con Python, tenemos que hacer:

```
from scipy.special import binom
binom(152,2)*0.999**150*0.001**3
```

```
## 9.8767434596705257e-06
```

```
nbinom.pmf(150,n=3,p=0.001)
```

```
## 9.8767434596702174e-06
```

Vemos que es muy improbable fallar 150 veces antes de acceder al sistema.

Solución 4. ¿Cuál es la probabilidad de fallar más de 150 veces antes de entrar en el sistema?

Nos piden calcular la probabilidad siguiente:

$$P(X > 150) = 1 - P(X \leq 150).$$

Calculemos $P(X \leq 150)$

$$\begin{aligned} P(X \leq 150) &= P(X = 0) + P(X = 1) + P(X = 2) + \dots + P(X = 150) \\ &= \sum_{k=0}^{150} \binom{k+3-1}{3-1} \cdot (0,999)^k \cdot 0,001^3 = \dots = 5,2320035 \times 10^{-4} \end{aligned}$$

Si hacemos el cálculo con R, obtenemos:

```
pnbinom(150, 3, 0.001)
```

```
## [1] 0.0005232003
```

Si lo hacemos en Python, obtenemos el mismo resultado:

```
nbinom.cdf(150, n=3, p=0.001)
```

```
## 0.00052320034908240592
```

El valor pedido será pues:

$$P(X > 150) = 1 - P(X \leq 150) = 1 - 5,2320035 \times 10^{-4} = 0,9994768.$$

Vemos que es muy probable que fallemos más de 150 veces antes de entrar en el sistema.

Solución 5. ¿Cuál es el número esperado de fallos antes de acceder al sistema? ¿Y su varianza?

$$E(X) = n \cdot \frac{1-p}{p} = 3 \cdot \frac{1-0,001}{0,001} = 2997.$$

$$Var(X) = n \cdot \frac{1-p}{p^2} = 3 \cdot \frac{1-0,001^2}{0,001^2} = 2,997 \times 10^6.$$

Con Python:

```
params = nbinom.stats(n=3, p=0.001, moments='mv')
```

```
print("E(X) = {}".format(m=params[0]))
```

```
## E(X) = 2997.0
```

```
print("Var(X) = {}".format(v=params[1]))
```

```
## Var(X) = 2997000.0
```

Ejercicio: ¿Tres claves de tres dígitos o una de 9 dígitos?

Supongamos que ponemos una sola clave de 9 dígitos. Estudiemos en este caso la variable aleatoria que da el número de fallos antes de entrar en el sistema y comparemos los resultados.

Si seguimos suponiendo que cada vez ponemos la contraseña al azar, pero esta vez con una clave de 9 dígitos. La probabilidad de éxito será ahora $p = \frac{1}{10^9}$.

Si llamamos X_9 a la variable aleatoria que nos da el número de fallos antes de entrar en el sistema seguirá una distribución $Ge(p = \frac{1}{10^9} = 0,000000001)$.

Su valor esperado es

$$E(X_9) = \frac{1-p}{p} = \frac{1-0,000000001}{0,000000001} = 10 \times 10^8.$$

1000000000 son 1000 millones de fallos esperados hasta abrir la puerta.

Recordemos que con tres contraseñas de 3 dígitos el valor esperado de fallos era:

$$3 \cdot \frac{1-0,001}{0,001} = 2997.$$

Por lo tanto, desde el punto de vista de la seguridad, es mejor una clave larga de 9 dígitos que tres cortas si escribimos las contraseñas al azar.

3.1.5. Distribución de Poisson

Diremos que una v.a. discreta X con $X(\Omega) = \mathbb{N}$ tiene distribución de Poisson con parámetro $\lambda > 0$, y lo denotaremos por $Po(\lambda)$ si su función de probabilidad es:

$$P_X(x) = P(X = x) = \begin{cases} \frac{\lambda^x}{x!} e^{-\lambda}, & \text{si } x = 0, 1, \dots \\ 0, & \text{en otro caso.} \end{cases}$$

Usando que el desarrollo en serie de Taylor de la función exponencial es

$$e^\lambda = \sum_{x=0}^{+\infty} \frac{\lambda^x}{x!},$$

es fácil comprobar que la suma de la función de probabilidad en todos los valores del dominio de X , o sea, los enteros positivos, vale 1.

Además, recordemos que dado $x \in \mathbb{R} - \{0\}$ se tiene que

$$\lim_{n \rightarrow \infty} \left(1 + \frac{x}{n}\right)^n = e^x.$$

Usando la expresión anterior para $x = -\lambda$, tenemos:

$$\lim_{n \rightarrow \infty} \left(1 - \frac{\lambda}{n}\right)^n = \lim_{n \rightarrow \infty} \left(1 + \frac{-\lambda}{n}\right)^n = e^{-\lambda}.$$

La distribución de Poisson como “límite” de una binomial

La distribución de Poisson (Siméon Denis Poisson) aparece en el conteo de determinados eventos que se producen en un intervalo de tiempo o en el espacio.

Supongamos que nuestra variable de interés es X , el número de eventos en el intervalo de tiempo $(0, t]$, como por ejemplo el número de llamadas a un *call center* en una hora donde suponemos que se cumplen las siguientes condiciones:

1. El número promedio de eventos en el intervalo $(0, t]$ es $\lambda > 0$.
2. Es posible dividir el intervalo de tiempo en un gran número de subintervalos (denotemos por n al número de intervalos) de forma que:
 - La probabilidad de que se produzcan dos o más eventos en un subintervalo es despreciable.
 - El número de ocurrencias de eventos en un intervalo es independiente del número de ocurrencias en otro intervalo.
 - La probabilidad de que un evento ocurra en un subintervalo es $p_n = \frac{\lambda}{n}$.

Bajo estas condiciones, podemos considerar que el número de eventos en el intervalo $(0, t]$ será el número de “éxitos” en n repeticiones independientes de un proceso Bernoulli de parámetro p_n .

Entonces si $n \rightarrow \infty$ y $p_n \cdot n$ se mantiene igual a λ , resulta que la función de probabilidad de X_n se puede escribir como

$$\begin{aligned} P(X_n = k) &= \binom{n}{k} \cdot p_n^k \cdot (1 - p_n)^{n-k} \\ &= \binom{n}{k} \cdot \left(\frac{\lambda}{n}\right)^k \cdot \left(1 - \frac{\lambda}{n}\right)^{n-k} \\ &= \frac{\lambda^k}{k!} \cdot \frac{n!}{(n-k)! \cdot n^k} \cdot \left(1 - \frac{\lambda}{n}\right)^n \cdot \left(1 - \frac{\lambda}{n}\right)^{-k}. \end{aligned}$$

Si hacemos tender n hacia ∞ , obtenemos:

$$\lim_{n \rightarrow \infty} P(X_n = k) = \lim_{n \rightarrow \infty} \frac{\lambda^k}{k!} \cdot \frac{n!}{(n-k)! \cdot n^k} \cdot \left(1 - \frac{\lambda}{n}\right)^n \cdot \left(1 - \frac{\lambda}{n}\right)^{-k}.$$

Calculemos el límite de algunos de los factores de la expresión

$$\begin{aligned} \lim_{n \rightarrow \infty} \frac{n!}{(n-k)! \cdot n^k} &= \lim_{n \rightarrow \infty} \frac{n \cdot (n-1) \cdots (n-k+1)}{n^k} = \lim_{n \rightarrow \infty} \frac{n^k + \cdots}{n^k} = 1, \\ \lim_{n \rightarrow \infty} \left(1 - \frac{\lambda}{n}\right)^n &= e^{-\lambda}, \\ \lim_{n \rightarrow \infty} \left(1 - \frac{\lambda}{n}\right)^{-k} &= \lim_{n \rightarrow \infty} 1^{-k} = \lim_{n \rightarrow \infty} 1 = 1, \end{aligned}$$

donde en el último límite, hemos tenido en cuenta que k es constante.

Usando las expresiones halladas anteriormente, tenemos que el límite de la función de probabilidad de la variable X_n tiende a la función de probabilidad de la variable de Poisson de parámetro λ :

$$\lim_{n \rightarrow \infty} P(X_n = k) = \lim_{n \rightarrow \infty} \binom{n}{k} \cdot p_n^k \cdot (1 - p_n)^{n-k} = \frac{\lambda^k}{k!} \cdot 1 \cdot e^{-\lambda} \cdot 1 = \frac{\lambda^k}{k!} \cdot e^{-\lambda}.$$

Usando que las variables X_n tienen distribución $B(n, p_n = \frac{\lambda}{n})$, tenemos que el límite de binomiales de parámetros n y $p_n = \frac{\lambda}{n}$ es una distribución de Poisson de parámetro λ , $Po(\lambda)$.

Procesos de Poisson

Lo interesante de las variables Poisson es que podemos modificar (si el modelo lo permite) el intervalo de tiempo $(0, t]$ en el que contamos los eventos, siempre y cuando se cumplan las condiciones 1 y 2 enunciadas anteriormente en el nuevo intervalo de tiempo.

En general, podemos afirmar si la variable es Poisson en $(0, t]$, también lo será en cualquier subintervalo $(0, t']$ para todo t' tal que $0 < t' < t$.

De esta forma, podremos definir una serie de variables X_t de distribución $Po(\lambda \cdot t)$.

Definición de procesos de Poisson

Consideremos un experimento *Poisson* con λ igual al promedio de eventos en una unidad de tiempo (u.t.).

Si t es una cantidad de tiempo en u.t., la v.a. X_t definida como el número de eventos en el intervalo $(0, t]$ es una $Po(\lambda \cdot t)$.

El conjunto de variables $\{X_t\}_{t>0}$ recibe el nombre de **proceso de Poisson**.

Resumen de la distribución de Poisson $X \equiv Po(\lambda)$

X Poisson	λ
$D_X =$	$\{0, 1, \dots\}$
$P_X(x) = P(X = x) =$	$\begin{cases} \frac{\lambda^x}{x!} e^{-\lambda}, & \text{si } x = 0, 1, \dots \\ 0, & \text{en otro caso.} \end{cases}$
$F_X(x) = P(X \leq X) = \sum_{i=0}^k P(X = i) =$	$\begin{cases} 0, & \text{si } x < 0, \\ \sum_{i=0}^k \frac{\lambda^i}{i!} \cdot e^{-\lambda}, & \text{si } \begin{cases} k \leq x < k+1, \\ k = 0, 1, 2, \dots \end{cases} \end{cases}$ Calcular la suma o utilizar funciones de R o Python.
$E(X) = \lambda$	$Var(X) = \lambda$

Resumen proceso Poisson $X_t \equiv Po(\lambda \cdot t)$

X_t $Po(\lambda \cdot t)$	λ promedio por u.t.
$D_X =$	$\{0, 1, \dots\}$
$P_X(x) = P(X = x) =$	$\begin{cases} \frac{(\lambda \cdot t)^x}{x!} e^{-\lambda \cdot t} & \text{si } x = 0, 1, \dots \\ 0 & \text{en otro caso.} \end{cases}$
$F_X(x) = P(X \leq X) = \sum_{i=0}^k P(X = i) =$	$\begin{cases} 0, & \text{si } x < 0, \\ \sum_{i=0}^k \frac{(\lambda \cdot t)^i}{i!} \cdot e^{-\lambda \cdot t}, & \text{si } \begin{cases} k \leq x < k+1, \\ k = 0, 1, 2, \dots \end{cases} \end{cases}$ Calcular la suma o utilizar funciones de R o Python.
$E(X) = \lambda \cdot t$	$Var(X) = \lambda \cdot t$

Aproximación de la distribución binomial por la Poisson

Dada una variable aleatoria de distribución $B(n, p)$, si n es grande y p es pequeño podemos aproximar

la distribución anterior por una distribución Poisson de parámetro $\lambda = n \cdot p$, $Po(\lambda = n \cdot p)$.

Un criterio para decidir que la aproximación anterior es buena es que $n \geq 20$, o mejor, $n \geq 30$, $n \cdot p < 10$ y $p \leq 0,05$.

La aproximación de la función de probabilidad de una variable binomial a una variable de Poisson es óptima en los valores cercanos a $E(X) = \lambda$.

Gráficos de la aproximación binomial a la de Poisson

Suponemos que estamos en las condiciones anteriores: $n \geq 20$, $n \cdot p < 10$, $p \leq 0,05$.

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.

```
fluidPage(
  fluidRow(
    column(6,
      sliderInput("n_binomP", label = "Número de repeticiones n:",
                  min = 1, max = 100, value = 20, step = 1)),
    column(6,
      sliderInput("p_binomP", label = "Probabilidad éxito p:",
                  min = 0.001, max = 0.9, value = 0.05, step = 0.001)
    )
  )
)

renderPlot({
  n=input$n_binomP
  pr=input$p_binomP
  par(mfrow=c(1,2))
  aux=rep(0,(n+1)*2)
  aux[seq(2,(n+1)*2,2)]=dbinom(c(0:n),size=n,prob=pr)
  plot(x=c(0:n),y=dbinom(c(0:n),size=n,prob=pr),
        ylim=c(0,0.6),xlim=c(-1,n+1),xlab="x",
        main=paste0(c("Funciones de probabilidad\n B(n=",n,",p=",pr,"),",
                     "Po(lambda=",n*pr,")",collapse = ""))
  lines(x=rep(0:n,each=2),y=aux,pch=21, type = "h", lty = 2,col="blue")
  aux=rep(0,(n+1)*2)
  aux[seq(2,(n+1)*2,2)]=dpois(c(0:n),n*pr)
  points(x=c(0:n),y=dpois(c(0:n),n*pr),
         ylim=c(0,0.6),xlim=c(-1,n+1),xlab="x",pch=25,col="red")
  lines(x=rep(0:n,each=2),y=aux, type = "h", lty = 3,col="red")
  legend("topleft",legend=c("Binomial","Poisson"),col=c("blue","red"),
         pch=c(21,25),lty=c(2,3),byt = "n")
  curve(pbinom(x,size=n,p=pr),
         xlim=c(-1,n+1),col="blue",
         main=paste0(c("Funciones de distribución \n B(n=",n,",p=",pr,"),",
                      "Po(lambda=",n*pr,")",collapse = ""))
  curve(ppois(x,n*pr),
```

```

xlim=c(-1,n+1),col="red",add=TRUE)
if(all(c(n>=20,n*pr<10,pr<= 0.05))){aux_1="Condición\n TRUE"} else
  {aux_1="Condición\n FALSE"}
legend("topleft",legend=c(aux_1,paste0("n=",n),paste0("n*p=",n*pr),
  paste0("p=",pr)),bg="transparent",cex=0.8,bty = "n")
par(mfrow=c(1,1))
})

```

Ejemplo de una distribución de Poisson $Po(\lambda)$: trampa para insectos

La conocida lámpara antiinsectos o insecticida eléctrico atrae a los insectos voladores con una luz ultravioleta y los mata por electrocución.

Consideremos la v.a. X que cuenta el número de insectos caídos en la trampa en una hora. Supongamos que el número promedio de insectos que captura la trampa en una hora es $E(X) = 20$ y que podemos admitir que X sigue una ley de probabilidad $Po(\lambda = 20)$.

Nos piden

1. Comentar de forma breve si se cumplen intuitivamente las condiciones para tener una distribución Poisson.
2. Escribir de forma explícita la función de probabilidad y de distribución de X .
3. Calculad la probabilidad de que en una hora caigan en la trampa exactamente 21 insectos.
4. Calculad la probabilidad de que en una hora caigan en la trampa al menos 6 insectos.
5. ¿Cuál es el valor esperando, la varianza y la desviación típica de X ?

Solución 1. Comentar de forma breve si se cumplen intuitivamente las condiciones para tener una distribución Poisson.

1. El número promedio de eventos en el intervalo $(0, 1]$, una hora es $\lambda = 20 > 0$.
2. Es posible dividir el intervalo de tiempo de una hora en un gran número de subintervalos (denotemos por n al número de intervalos) de forma que:
 - La probabilidad de que se produzcan dos o más electrocuciones un subintervalo es despreciable. No es posible que dos mosquitos se electrocuten al mismo tiempo.
 - El número de ocurrencias, electrocuciones de insectos, en un intervalo es independiente del número de electrocuciones en otro intervalo.
 - La probabilidad de que un evento ocurra en un subintervalo es $p_n = \frac{\lambda}{n}$. Podemos dividir los 20 insectos promedio entre los n intervalos (trozo de hora) de forma que $p_n = \frac{\lambda}{n}$.
 - Por ejemplo si $n = 60$ tenemos que $p_n = \frac{20}{60} = \frac{1}{3}$. La probabilidad de que en un minuto la trampa chisporrotee es $\frac{1}{3}$.

Solución 2. Escribid de forma explícita la función de probabilidad y de distribución de X .

La distribución de probabilidad de un $Po(\lambda)$ es

$$P_X(x) = P(X = x) = \begin{cases} \frac{\lambda^x}{x!} e^{-\lambda}, & \text{si } x = 0, 1, \dots \\ 0, & \text{en otro caso.} \end{cases}$$

En nuestro caso, $\lambda = 20$:

$$P_X(x) = P(X = x) = \begin{cases} \frac{20^x}{x!} e^{-20}, & \text{si } x = 0, 1, \dots \\ 0, & \text{en otro caso.} \end{cases}$$

La función de distribución es:

$$F_X(x) = P(X \leq x) = \begin{cases} 0, & \text{si } x < 0, \\ \sum_{i=0}^k P(X = i) = \sum_{i=0}^k \frac{\lambda^i}{i!} \cdot e^{-\lambda}, & \text{si } \begin{cases} k \leq x < k+1, \\ k = 0, 1, 2, \dots \end{cases} \end{cases}$$

En nuestro caso:

$$F_X(x) = P(X \leq x) = \begin{cases} 0, & \text{si } x < 0, \\ \sum_{i=0}^k P(X = i) = \sum_{i=0}^k \frac{20^i}{i!} \cdot e^{-20}, & \text{si } \begin{cases} k \leq x < k+1, \\ k = 0, 1, 2, \dots \end{cases} \end{cases}$$

Solución 3. Calculad la probabilidad de que en una hora caigan en la trampa exactamente 21 insectos.

Nos piden la probabilidad siguiente:

$$P(X = 21) = \frac{20^{21}}{21!} e^{-20} = 0,0846051.$$

Para realizar el cálculo anterior, podemos usar R como calculadora o usar la función `dpois` que nos calcula la función de distribución de la variable de Poisson:

```
20^(21)/factorial(21)*exp(-20)
```

```
## [1] 0.08460506
```

```
dpois(21,lambda = 20)
```

```
## [1] 0.08460506
```

Solución 4. Calculad la probabilidad de que en una hora caigan en la trampa al menos 6 insectos.

Nos piden la probabilidad siguiente:

$$\begin{aligned} P(X \geq 6) &= 1 - P(X < 6) = 1 - P(X \leq 5) = 1 - F_X(5) = 1 - \sum_{x=0}^5 \frac{20^x}{x!} \cdot e^{-20} \\ &= 1 - \left(\frac{20^0}{0!} \cdot e^{-20} + \frac{20^1}{1!} \cdot e^{-20} + \frac{20^2}{2!} \cdot e^{-20} + \frac{20^3}{3!} \cdot e^{-20} + \frac{20^4}{4!} \cdot e^{-20} + \frac{20^5}{5!} \cdot e^{-20} \right) \\ &= 1 - e^{-20} \cdot \left(1 + 20 + \frac{400}{4} + \frac{8000}{6} + \frac{160000}{24} + \frac{3200000}{120} \right) \\ &= 1 - e^{-20} \cdot \left(\frac{1 \cdot 120 + 20 \cdot 120 + 400 \cdot 30 + 8000 \cdot 20 + 160000 \cdot 24 + 3200000 \cdot 1}{120} \right) \\ &= 1 - e^{-20} \cdot \left(\frac{4186520}{120} \right) = 1 - 7,1908841 \times 10^{-5} = 0,9999281. \end{aligned}$$

Solución 5. ¿Cuál es el valor esperado, la varianza y la desviación típica de X ?

El valor esperado del número de insectos caídos en la trampa en una hora es:

$$E(X) = \lambda = 20.$$

Su varianza es

$$Var(X) = \lambda = 20,$$

y su desviación típica vale:

$$\sqrt{Var(X)} = +\sqrt{\lambda} = +\sqrt{20} = 4,47214.$$

Cálculos con R

Consideremos por ejemplo una v.a. X con distribución $Po(\lambda = 3)$. Calculemos $P_X(0) = P(X = 0), P_X(1) = P(X = 1)$ con R:

```
dpois(0,lambda = 3)
```

```
## [1] 0.04978707
```

```
dpois(1,lambda = 3)
```

```
## [1] 0.1493612
```

Si quisiéramos hallar la función de distribución en los mismos valores anteriores, $F_X(0) = P(X \leq 0), F_X(1) = P(X \leq 1)$, haríamos lo siguiente:

```
ppois(0,lambda = 3)
```

```
## [1] 0.04978707
```

```
ppois(1,lambda = 3)
```

```
## [1] 0.1991483
```

```
dpois(0,lambda = 3)+dpois(1,lambda = 3) # es igual a ppois(1,lambda=3)
```

```
## [1] 0.1991483
```

A continuación, comprobemos que $F_X(10) = \sum_{x=0}^{10} P_X(x)$:

```
dpois(0:10,3)
```

```
## [1] 0.0497870684 0.1493612051 0.2240418077 0.2240418077 0.1680313557
```

```
## [6] 0.1008188134 0.0504094067 0.0216040315 0.0081015118 0.0027005039
```

```
## [11] 0.0008101512
```

```
sum(dpois(0:10,3))
```

```
## [1] 0.9997077
```

```
ppois(10,3)
```

```
## [1] 0.9997077
```

Si quisiéramos generar una secuencia de 100 observaciones para una distribución de Poisson de parámetro $\lambda = 3$, $Po(3)$, tendríamos que hacer:

```
rpois(n=100,lambda = 3)
```

```
## [1] 2 5 3 3 2 2 5 2 4 4 2 3 2 2 2 2 3 3 5 3 3 2 4 2 3 2 1 1 3 4 6 2 5 3 4 1
```

```
## [38] 1 6 3 4 1 4 3 4 3 0 2 1 4 3 0 2 4 2 3 5 2 1 3 3 4 2 5 0 3 1 1 4 6 4 5 0 4
```

```
## [75] 0 3 3 3 4 1 2 6 2 2 2 2 1 2 5 2 5 3 7 3 5 2 3 2 1 3
```

Ejercicio de la trampa para insectos (continuación)

En el ejercicio de la trampa para insectos teníamos que X es una $Po(20)$. Responded con R a la preguntas 3 y 4 de este ejercicio

Pregunta 3. Calculad la probabilidad de que en una hora caigan en la trampa exactamente 21 insectos.

Recordemos que la probabilidad pedida es $P(X = 21)$:

```
dpois(21,lambda=20) # P(X=21)
```

```
## [1] 0.08460506
```

Pregunta 4. Calculad la probabilidad de que en una hora caigan en la trampa al menos 6 insectos.

Recordemos que la probabilidad pedida es $P(X \geq 6) = 1 - P(X < 6) = 1 - P(X \leq 5)$:

```
ppois(5,lambda=20)
```

```
## [1] 7.190884e-05
```

```
1-ppois(5,lambda=20) # es 1-P(X<=5)=P(X>=6)
```

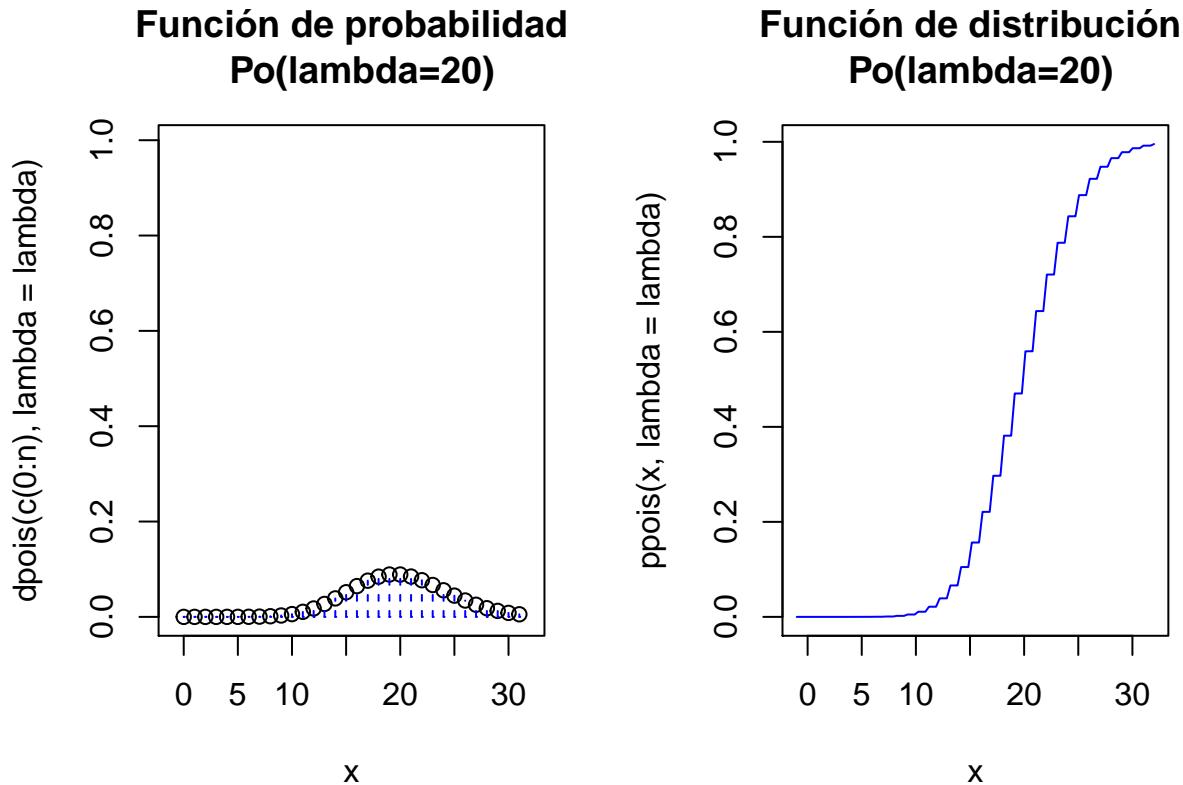
```
## [1] 0.9999281
```

```
# acumula hacia arriba P(X>5)=P(X>=6)=P(X=6)+P(X=7)+..  
ppois(5,lambda=20,lower.tail =FALSE )
```

```
## [1] 0.9999281
```

Gráficos de la distribución Poisson con R

```
lambda=20
par(mfrow=c(1,2))
n=qpois(0.99,lambda=lambda)
aux=rep(0,(n+1)*2)
aux[seq(2,(n+1)*2)]=dpois(c(0:n),lambda=lambda)
ymax=max(ppois(0:n,lambda=lambda))
plot(x=c(0:n),y=dpois(c(0:n),lambda=lambda),
      ylim=c(0,ymax),xlim=c(-1,n+1),xlab="x",
      main=paste0(c("Función de probabilidad\n Po(lambda=",lambda,")"),
                  collapse = ""))
lines(x=rep(0:n,each=2),y=aux,pch=21, type = "h", lty = 2,col="blue")
curve(ppois(x,lambda=lambda),
      xlim=c(-1,n+1),col="blue",
      main=paste0(c("Función de distribución \n Po(lambda=",lambda,")"),
                  collapse = ""))
par(mfrow=c(1,1))
```



Gráficos interactivos con R

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("lambda", label = "Promedio de eventos lambda",
            min = 1, max = 100, value = 20 , step = 1)
renderPlot({
  lambda=input$lambda
  par(mfrow=c(1,2))
  n=qpois(0.99,lambda=lambda)
  #n
  aux=rep(0,(n+1)*2)
  aux[seq(2,(n+1)*2,2)]=dpois(c(0:n),lambda=lambda)
  ymax=0.45
  plot(x=c(0:n),y=dpois(c(0:n),lambda=lambda),
    ylim=c(0,ymax),xlim=c(-1,n+1),xlab="x",
    main=paste0(c("Función de probabilidad\n Po(lambda=",lambda,")"),
    collapse = ""))
  lines(x=rep(0:n,each=2),y=aux,pch=21, type = "h", lty = 2,col="blue")
  curve(ppois(x,lambda=lambda),
    xlim=c(-1,n+1),col="blue",
    main=paste0(c("Función de distribución \n Po(lambda=",lambda,")"),
    collapse = ""))
})
```

```

collapse = ""))
par(mfrow=c(1,1))
}

```

Cálculos con Python

Realicemos los mismos cálculos realizados con R pero ahora usando Python. Recordemos que considerábamos una v.a. X con distribución $Po(\lambda = 3)$. Calculemos $P_X(0) = P(X = 0)$, $P_X(1) = P(X = 1)$ con Python

```
from scipy.stats import poisson
poisson.pmf(0, mu = 3)
```

```
## 0.049787068367863944
```

```
poisson.pmf(1, mu = 3)
```

```
## 0.14936120510359185
```

Si quisiéramos hallar las funciones de distribución en los mismos valores anteriores, $F_X(0) = P(X \leq 0)$, $F_X(1) = P(X \leq 1)$, tendríamos que hacer:

```
poisson.cdf(0, mu = 3)
```

```
## 0.049787068367863951
```

```
poisson.cdf(1, mu = 3)
```

```
## 0.19914827347145581
```

```
poisson.pmf(0, mu = 3) + poisson.pmf(1, mu = 3) # es igual a poisson.cdf(1, lambda=3)
```

```
## 0.19914827347145581
```

La comprobación de que $F_X(10) = \sum_0^{10} P_X(x)$ en Python se realiza de la forma siguiente:

```
range(0, 10)
```

```
## [0, 1, 2, 3, 4, 5, 6, 7, 8, 9]
```

```
poisson.pmf(range(0, 10), mu=3)
```

```
## array([ 0.04978707,  0.14936121,  0.22404181,  0.22404181,  0.16803136,
##        0.10081881,  0.05040941,  0.02160403,  0.00810151,  0.0027005])
```

```
sum(poisson.pmf(range(0, 10), mu=3))
```

```
## 0.99889751186988462
```

```
poisson.cdf(10, mu=3)
```

```
## 0.99970766304935266
```

Ejercicio de la trampa para insectos (continuación)

En el ejercicio de la trampa para insectos teníamos que X es una $Po(20)$. Responded con Python a la preguntas 3 y 4 de este ejercicio

Pregunta 3. Calculad la probabilidad de que en una hora caigan en la trampa exactamente 21 insectos.

Recordemos que la probabilidad pedida es $P(X = 21)$:

```
poisson.pmf(21, mu=20) # P(X=21)
```

```
## 0.084605064182937909
```

Pregunta 4. Calculad la probabilidad de que en una hora caigan en la trampa al menos 6 insectos.

La probabilidad pedida es $P(X \geq 6) = 1 - P(X \leq 5)$:

```
1-poisson.cdf(5, mu=20) # es 1-P(X<=5)=P(X>=6)
```

```
## 0.99992809115947157
```

Como ya hemos visto con `scipy.stats`, podemos pedir los momentos de una variable aleatoria $Po(3)$

```
poisson.stats(mu=3, moments='mv')
```

```
## (array(3.0), array(3.0))
```

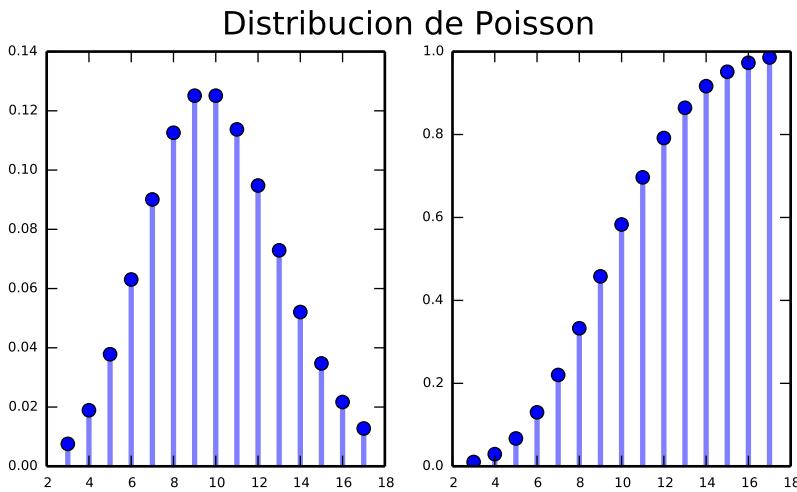
Y también generar secuencias de observaciones aleatorias de una población $Po(3)$:

```
poisson.rvs(mu=3, size=40)
```

```
## array([ 2,  1,  3,  1,  5,  2,  4,  0,  3,  3,  1,  3,  3,  4,  0,  3,  4,
##        5,  0,  2,  1,  1,  3,  1,  3,  2,  1,  4,  3,  2,  2,  4,  4,  3,
##        2,  2, 10,  4,  5,  3])
```

Gráficos con Python

```
from scipy.stats import poisson
mu = 10 ## mu = lambda
x = np.arange(poisson.ppf(0.01, mu), poisson.ppf(0.99, mu))
fig = plt.figure(figsize=(5, 2.7))
ax = fig.add_subplot(1, 2, 1)
ax.plot(x, poisson.pmf(x, mu), 'bo', ms=5, label='Poisson pmf')
ax.vlines(x, 0, poisson.pmf(x, mu), colors='b', lw=2, alpha=0.5)
for tick in ax.xaxis.get_major_ticks():
    tick.label.set_fontsize(5)
for tick in ax.yaxis.get_major_ticks():
    tick.label.set_fontsize(5)
ax = fig.add_subplot(1, 2, 2)
ax.plot(x, poisson.cdf(x, mu), 'bo', ms=5, label='Poisson cdf')
ax.vlines(x, 0, poisson.cdf(x, mu), colors='b', lw=2, alpha=0.5)
for tick in ax.xaxis.get_major_ticks():
    tick.label.set_fontsize(5)
for tick in ax.yaxis.get_major_ticks():
    tick.label.set_fontsize(5)
fig.suptitle('Distribucion de Poisson')
plt.show()
```



Gráficos interactivos para un proceso de Poisson $Po(\lambda \cdot t)$

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.

```

fluidPage(
  fluidRow(
    column(6,
      sliderInput("lambdapp", label="Promedio eventos por unidad de tiempo",
                  min = 0.1, max = 50, value =10 , step = 0.01)),
    column(6,sliderInput("t", label = "Intervalo de tiempo (0,t]",
                          min = 1, max = 120, value =1 , step = 0.5)))
  )
)

renderPlot({
  lambda1=input$lambdapp
  t=input$t
  lambda=lambda1*t ## es lambda* t
  par(mfrow=c(1,2))
  n=qpois(0.99,lambda=lambda)
  #n
  aux=rep(0,(n+1)*2)
  aux[seq(2,(n+1)*2,2)]=dpois(c(0:n),lambda=lambda)
  ymax=ppois(which.max(ppois(0:n,lambda))-1,lambda)*0.7
  plot(x=c(0:n),y=dpois(c(0:n),lambda=lambda),
        ylim=c(0,ymax),xlim=c(-1,n+1),xlab="x",
        main=paste0(c("Función de probabilidad\n Po(lambda=",lambda,")"),
                    collapse = ""))
  lines(x=rep(0:n,each=2),y=aux,pch=21, type = "h", lty = 2,col="blue")
}
)

```

```

curve(ppois(x,lambda=lambda),
      xlim=c(-1,n+1),col="blue",
      main=paste0(c("Función de distribución \n Po(lambda=",lambda,")"),
                  collapse = ""))
par(mfrow=c(1,1))
})

```

Ejemplo: Número de impactos de insectos en la visera de un casco

Un colega de trabajo, al que llamaremos JG, es muy aficionado a los grandes premios de velocidad tanto en coches como en motos.

Como es tan aficionado, está obsesionado con muchas de las más extravagantes estadísticas de estos deportes. En particular le propusimos que estudiara el número de insectos que chocan contra la visera de un casco de un motorista GP o de un conductor de fórmula 1 .

La idea es que el número de insectos está igualmente repartido por todo el circuito y de promedio impactan $\lambda > 0$ insectos por minuto. También es razonable suponer que:

- podemos dividir la superficie de la visera en cuadrados suficientemente pequeños de forma que la probabilidad de que caigan dos insectos en la misma zona es prácticamente 0.
- la probabilidad de que un insecto impacte en un cuadrado cualquiera de la visera es independiente de cualquier otro cuadrado.
- si hemos dividido la visera en n cuadrados la probabilidad p_n de impacto de un cuadrado vale $p_n = \frac{\lambda}{n}$.

Bajo estas condiciones, si denotamos por X_t como el número de insectos que ha impactado en la visera en el intervalo $(0, t]$ (en t minutos), podemos afirmar que X_t es un proceso de Poisson $Po(\lambda \cdot t)$.

Supongamos que nos dicen que $\lambda = 3$ insectos por minuto. Entonces el proceso de Poisson X_t seguirá una ley $Po(3 \cdot t)$.

Nos piden las probabilidades siguientes:

1. ¿Cuál es la probabilidad de que en 10 minutos impacten más de 25 insectos?
2. ¿Cuál es la probabilidad de que tengamos que esperar más de 2 minutos para observar el primer impacto?

Solución de 1. ¿Cuál es la probabilidad de que en 10 minutos impacten más de 25 insectos?

En este caso $t = 10$ y X_{10} es la variable aleatoria que nos da el número de insectos que impactan en 10 minutos o durante el intervalo $[0, 10]$. La distribución de X_{10} será de Poisson de parámetro $\lambda = 3 \cdot 10 = 30$, $Po(30)$.

Nos piden la probabilidad siguiente: $P(X > 25) = 1 - P(X \leq 25)$, que calculamos con ayuda de R:

```
1-dpois(25,lambda=3)
```

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

Solución de 2. ¿Cuál es la probabilidad de que tengamos que esperar más de 2 minutos para observar el primer impacto?

Nos piden la probabilidad siguiente $P(X_2 = 0)$ ya que la variable X_2 nos dice el número de impactos en dos minutos. La distribución de X_2 será de Poisson de parámetro $\lambda = 2 \cdot 3 = 6$, $Po(6)$. Si hemos de

esperar más de dos minutos para el primer impacto, significa que $X_2 = 0$:

$$P(X_2 = 0) = \frac{(6)^0}{0!} \cdot e^{-6} = e^{-6} = 0,002479.$$

Si usamos R, obtenemos:

```
6^0/factorial(0)*exp(-6)
```

```
## [1] 0.002478752
```

```
ppois(0,lambda=3*2)
```

```
## [1] 0.002478752
```

3.1.6. Distribución hipergeométrica

Supongamos que disponemos de una urna de sorteos que contiene m bolas blancas y n bolas rojas.

En total en esta urna hay $m + n$ bolas, m blancas y n rojas. Si extraemos dos bolas de la urna lo podemos hacer de dos formas:

- Extraer una, anotar su color y reponerla. Sacar otra y anotar su color. Hemos extraído la bola con reposición.
- Extraer simultáneamente dos bolas (sin reposición) y contar el número de bolas blancas.

Sea X es la v.a. que cuenta el número de bolas blancas extraídas.

- En el primer caso, X es una $B(n = 2, p = \frac{m}{m+n})$ ya que consiste en repetir dos veces el mismo experimento de Bernoulli.
- En el segundo caso, X sigue una distribución hipergeométrica que estudiaremos en esta sección.

Distribución hipergeométrica

Sean n , m y k tres números enteros positivos y tales que $k < m + n$.

Consideremos una urna que contiene $m + n$ bolas de las que m son blancas y las restantes n no (son no blancas).

El número total de bolas es $m + n$. Extraemos de forma aleatoria k bolas de la urna sin reemplazarlas.

Sea X la v.a. que cuenta el número de bolas blancas extraídas. Diremos que la distribución de X es hipergeométrica de parámetros m , n y k y la denotaremos por $H(m, n, k)$.

Su dominio es

$$D_X = \{x \in \mathbb{N} \mid \max\{0, k - n\} \leq x \leq \min\{m, k\}\}$$

Para explicarlo, veamos varios ejemplos:

- $H(m = 5, n = 2, k = 3)$. Tenemos $m = 5$ bolas blancas, $n = 2$ no blancas y sacamos $k = 3$ bolas sin reposición.
 - En este caso el mínimo de bolas blancas extraídas es $1 = k - n = 3 - 2$, ya que sólo hay dos no blancas.

- En cambio, el máximo sí es $k = 3$, ya que tenemos bolas blancas de “sobra”.
- $H(m = 2, n = 5, k = 3)$. Tenemos $m = 2$ bolas blancas, $n = 5$ no blancas y sacamos $k = 3$ bolas sin reposición.
 - En este caso el mínimo de bolas blancas es 0 ya que puedo sacar 3 no blancas.
 - En cambio, el máximo sí es $m = 2$, ya que aunque saquemos $k = 3$ bolas, al llegar a 2 ya hemos extraído todas las bolas blancas de la urna.
- $H(m = 10, n = 10, k = 3)$. Tenemos $m = 10$ bolas blancas, $n = 10$ no blancas y sacamos $k = 3$ bolas sin reposición.
 - En este caso podemos obtener desde 0 blancas hasta $k = 3$ blancas.

Su función de probabilidad es:

$$P_X(x) = \begin{cases} \frac{\binom{m}{x} \cdot \binom{n}{k-x}}{\binom{m+n}{k}}, & \text{si } \max\{0, k-n\} \leq x \leq \min\{m, k\}, \text{ para } x \in \mathbb{N}, \\ 0, & \text{en otro caso.} \end{cases}$$

Observación: otras parametrizaciones

En ocasiones se parametriza una v.a. hipergeométrica mediante $N = m + n$, número total de bolas, k , número de extracciones y p , probabilidad de extraer una bola blanca.

Así, podemos **parametrizar alternativamente** la distribución hipergeométrica como $H(N, k, p)$ donde $p = \frac{m}{N}$.

Resumen hipergeométrica $H(m, n, k)$

$X = H(m, n, k)$	Número de bolas blancas en k extracciones sin reposición de una urna con m bolas blancas y n no blancas
$D_X =$	$\{x \in \mathbb{N} \mid \max\{0, k-n\} \leq x \leq \min\{m, k\}\}$
$P_X(x) = P(X = x) =$	$\begin{cases} \frac{\binom{m}{x} \cdot \binom{n}{k-x}}{\binom{m+n}{k}}, & \text{si } x \in [\max\{0, k-n\}, \min\{m, k\}] \\ 0, & \text{en otro caso.} \end{cases}$
$F_X(x) = P(X \leq x)$	Hay que sumarla. Utilizad funciones de R o de Python.
$E(X) = \frac{k \cdot m}{m+n}$	$Var(X) = k \cdot \frac{m}{m+n} \cdot \left(1 - \frac{m}{m+n}\right) \cdot \frac{m+n-k}{m+n-1}$

Ejemplo: urna con $m = 15$ blancas, $n = 10$ rojas y $k = 3$ extracciones sin reposición

Tenemos una urna con 15 bolas blancas y 10 bolas rojas. Extraemos al azar tres bolas de la urna sin reposición. Sea X el número de bolas **blancas** extraídas. Bajo estas condiciones, la v.a. X sigue una ley de distribución $H(m = 15, n = 10, k = 3)$.

Nos piden:

1. Hallar la función de probabilidad de X .
2. Probabilidad de sacar dos bolas blancas.
3. Probabilidad de sacar más de una bola blanca.
4. Esperanza, varianza y desviación típica de X .

Solución de 1. Hallar la función de probabilidad de X

La función de probabilidad de X es:

$$P_X(x) = P(X = x) = \begin{cases} \frac{\binom{m}{x} \cdot \binom{n}{k-x}}{\binom{m+n}{k}}, & \text{si } \max\{0, k-n\} \leq x \leq \min\{m, k\}, \text{ para } x \in \mathbb{N}, \\ 0, & \text{en otro caso.} \end{cases}$$

Sustituyendo los parámetros m, n y k por $m = 15$, $n = 10$ y $k = 3$, obtenemos:

$$P_X(x) = P(X = x) = \begin{cases} \frac{\binom{15}{x} \cdot \binom{10}{3-x}}{\binom{25}{3}} = \frac{\binom{15}{x} \cdot \binom{10}{3-x}}{2300}, & \text{si } 0 \leq x \leq 3, \text{ para } x \in \mathbb{N}, \\ 0, & \text{en otro caso.} \end{cases}$$

Solución de 2. Probabilidad de sacar dos bolas blancas

La probabilidad de sacar 2 blancas será:

$$P(X = 2) = \frac{\binom{15}{2} \cdot \binom{10}{3-2}}{\binom{25}{3}}$$

Si calculamos con ayuda de R los números binomiales involucrados en la expresión anterior, obtenemos:

```
c(choose(15,2), choose(10,1), choose(25,3))
```

```
## [1] 105 10 2300
```

La probabilidad pedida, será, pues: $P(X = 2) = \frac{105 \cdot 10}{2300} = 0,4565217$.

Solución de 3. Probabilidad de sacar más de una bola blanca

La probabilidad de que saquemos más de 1 bola blanca es:

$$\begin{aligned} P(X > 1) &= 1 - P(X \leq 1) = 1 - (P(X = 0) + P(X = 1)) \\ &= 1 - \left(\frac{\binom{15}{0} \cdot \binom{10}{3}}{\binom{25}{3}} + \frac{\binom{15}{1} \cdot \binom{10}{2}}{\binom{25}{3}} \right) \\ &= 1 - \left(\frac{1 \cdot 120}{2300} + \frac{15 \cdot 45}{2300} \right) = 1 - \frac{120+15 \cdot 45}{2300} = 0,6543478. \end{aligned}$$

Solución de 4. Esperanza, varianza y desviación típica de X

El número esperado de bolas blancas extraídas para una v.a. X de distribución $H(m = 15, n = 10, k = 3)$ es:

$$E(X) = \frac{k \cdot m}{m + n} = \frac{3 \cdot 15}{15 + 10} = \frac{45}{35} = 1,285714.$$

La varianza vale:

$$\begin{aligned} Var(X) &= k \cdot \frac{m}{m+n} \cdot \left(1 - \frac{m}{m+n}\right) \cdot \frac{m+n-k}{m+n-1} \\ &= 3 \cdot \frac{15}{15+10} \cdot \left(1 - \frac{15}{15+10}\right) \cdot \frac{15+10-3}{15+10-1} \\ &= 3 \cdot \frac{15}{25} \cdot \left(1 - \frac{15}{25}\right) \cdot \frac{22}{24} = 3 \cdot \frac{15}{25} \cdot \frac{10}{25} \cdot \frac{22}{24} \\ &= 3 \cdot \frac{15}{25} \cdot \frac{10}{25} \cdot \frac{22}{24} = 0,66. \end{aligned}$$

Y por lo tanto, su desviación típica es:

$$+\sqrt{Var(X)} = +\sqrt{0,66} = 0,812404.$$

Cálculos con R

Sea X una v.a. $H(m, n, k)$. La función de R para calcular la función de probabilidad en un valor x , $P(X = x)$, es `dhyper(x, m, n, k)` y para calcular la función de distribución en un valor q , $P(X \leq q)$, es `phyper(q, m, n, k)`. Para generar una muestra de valores que siga la distribución $H(m, n, k)$, hay que usar la función `rhyper(nn, m, n, k)` donde `nn` es el número de observaciones aleatorias deseado de la muestra.

Por ejemplo, si X es una $H(m = 15, n = 10, k = 3)$, los valores de $P(X = 2)$ y que $P(X > 1) = 1 - P(X \leq 1)$ son:

```
dhyper(x=2, m=15, n=10, k=3)
```

```
## [1] 0.4565217
phyper(q=1, m=15, n=10, k=3) ## sí, le han puesto q ya veremos el porqué
## [1] 0.3456522
1-phyper(q=1, m=15, n=10, k=3)
## [1] 0.6543478
```

Una muestra aleatoria de este experimento de tamaño 200 sería:

```
rhyper(nn=200, m=15, n=10, k=3)
```

```
## [1] 2 3 1 3 1 2 2 3 2 2 1 2 1 2 2 3 3 1 1 1 1 0 2 3 2 1 3 2 2 2 2 3 2 3 3 3 2 0
## [38] 1 2 1 3 2 2 3 2 3 2 2 3 2 3 1 2 2 2 2 3 2 2 1 3 2 2 3 1 2 2 2 2 3 0 2 0
## [75] 3 2 2 2 1 2 2 3 1 1 1 2 2 2 2 1 1 3 2 2 3 2 2 1 1 1 3 3 2 2 2 1 3 2 2 2 1
## [112] 1 2 3 2 2 1 2 2 2 2 2 3 1 2 3 3 1 1 2 2 1 1 3 2 1 1 2 2 3 1 1 1 2 1 1 3
## [149] 1 2 2 3 3 2 3 1 2 1 2 2 2 1 2 3 1 3 3 2 2 1 3 3 1 1 2 2 2 2 2 3 2 1 2 1
## [186] 1 1 1 2 1 1 2 2 2 2 3 3 1 0 2
```

Gráficas con R

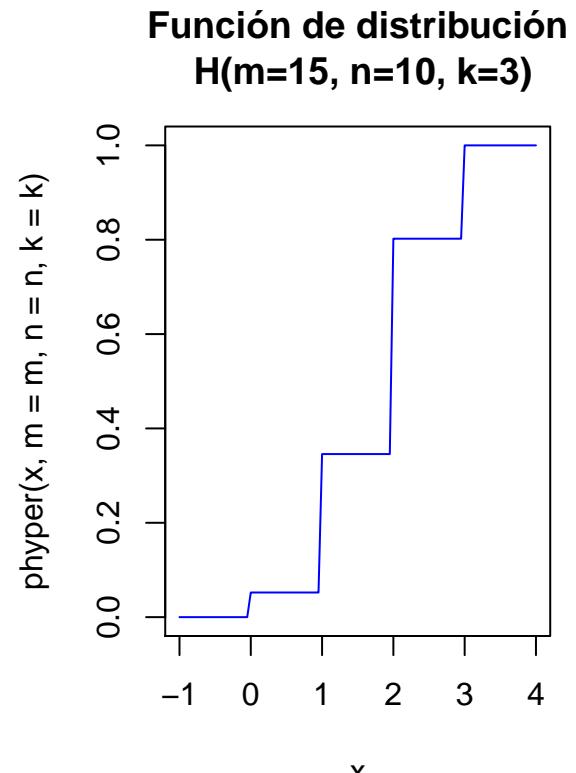
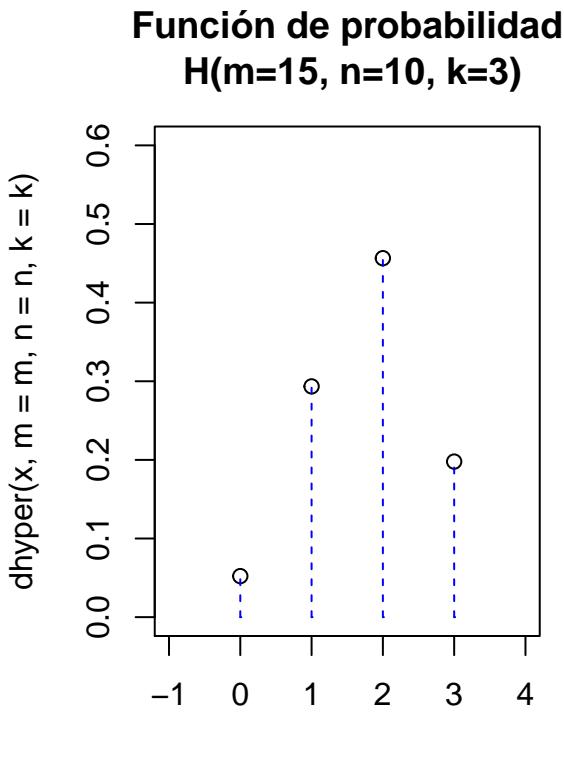
Los gráficos de la función de probabilidad y de la función de distribución en R se realizan de la forma siguiente:

```
par(mfrow=c(1,2))
m=15
n=10
k=3
a=max(c(0,k-n))
b=min(c(m,k))
l=b-a+1
aux=rep(0,2*l)
aux[seq(2,2*l,2)]=dhyper(c(a:b),m=m,n=n,k=k)
x=a:b
plot(x,y=dhyper(x,m=m,n=n,k=k),
ylim=c(0,0.6),xlim=c(a-1,b+1),xlab="x",
```

```

main=paste0("Función de probabilidad\n H(m=",m," , n=",n," , k=",k,"))")
lines(x=rep(a:b,each=2),y=aux, type = "h", lty = 2,col="blue")
curve(dhyper(x,m=m,n=n,k=k),
      xlim=c(a-1,b+1),col="blue",
      main=paste0("Función de distribución\n H(m=",m," , n=",n," , k=",k,"))")
par(mfrow=c(1,1))

```



Gráficos interactivos $H(m, n, k)$

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.

```

fluidPage(
  fluidRow(
    column(4,
      sliderInput("mh", label = "Número de bolas blancas m",
                 min = 1, max = 50, value =15, step = 1)),
    column(4,
      sliderInput("nh", label = "Número de bolas rojas n",
                 min = 1, max = 50, value =10 , step = 1)),
    column(4,
      sliderInput("kh", label = "Número bolas extraídas k",

```

```

        min = 1, max=25, value = 3, step = 1)
    )
}

renderPlot({
  m=input$mh
  n=input$nh
  k=input$kh
  #n=10
  #k=3
  #m=15
  par(mfrow=c(1,2))
  a=max(c(0,k-n))
  b=min(c(m,k))
  l=b-a+1
  aux=rep(0,times=2*l)
  aux[seq(2,2*l,2)]=dhyper(c(a:b),m=m,n=n,k=k)
  x=a:b
  plot(x,y=dhyper(x,m=m,n=n,k=k),
    ylim=c(0,0.6),xlim=c(a-1,b+1),xlab="x",
    main=paste0("Función de probabilidad\n H(m=",m,", n=",n,", k=",k,")"))
  lines(x=rep(a:b,each=2),y=aux, type = "h", lty = 2,col="blue")
  curve(pyper(x,m=m,n=n,k=k),
    xlim=c(a-1,b+1),col="blue",
    main=paste0("Función de distribución\n H(m=",m,", n=",n,", k=",k,")"))
  par(mfrow=c(1,1))
})

```

Comparación $H(m, n, k)$ y $B\left(k, \frac{m}{n+m}\right)$

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.

```

fluidPage(
  fluidRow(
    column(4,
      sliderInput("mh2", label = "Número de bolas blancas m",
                  min = 1, max = 50, value = 15, step = 1)),
    column(4,
      sliderInput("nh2", label = "Número de bolas rojas n",
                  min = 1, max = 50, value = 10 , step = 1)),
    column(4,
      sliderInput("kh2", label = "Número bolas extraídas k",
                  min = 1, max=25, value = 3, step = 1)
    )
  )
)

```

```

renderPlot({
  m=input$mh2
  n=input$nh2
  k=input$kh2
  #n=10
  #k=3
  #m=15
  pr=round(m/(n+m),4)
  a=max(c(0,k-n))
  b=min(c(m,k))
  l=b-a+1
  aux=rep(0,times=2*l)
  auxB=rep(0,times=2*(k+1))
  aux[seq(2,2*l,2)]=dhyper(c(a:b),m=m,n=n,k=k)
  x=a:b
  auxB[seq(2,2*(k+1),2)]=dbinom(0:k,k,pr)
  par(mfrow=c(1,2))
  plot(x=c(0:k),y=dbinom(c(0:k),size=k,prob=pr),
    ylim=c(0,0.6),xlim=c(-1,k+1),xlab="x",
    main=paste0("Funciones de probabilidad\n B(n=",n,"p=",pr,","
      H(m=",m,"n=", n,"k=",k,")"))
  lines(x=rep(0:k,each=2),y=aux,pch=21, type = "h", lty = 2,col="blue")
  #aux=rep(0,(n+1)*2)
  #aux[seq(2,(n+1)*2,2)]=dpois(c(0:n),n*pr)
  points(x=c(a:b),y=dhyper(c(a:b),m=m,n=n,k=k),
    ylim=c(0,0.6),xlim=c(-1,k+1),xlab="x",pch=25,col="red")
  lines(x=rep(0:(l-1),each=2),y=aux, type = "h", lty = 3,col="red")
  legend("topleft",legend=c("Binomial","Hipergeométrica"),col=c("blue","red"),
    pch=c(21,25),lty=c(2,3))
  curve(pbinom(x,size=k,p=pr),
    xlim=c(-1,k+1), col="blue",
    main=paste0("Funciones de distribución\n B(",k,",",pr,")
      H(m=",m,"n=", n,"k=",k,")"))
  curve(phyper(x,m=m,n=n,k=k),
    xlim=c(-1,k+1),col="red",add=TRUE)
  #if(all(c(n>=20,n*pr<10,pr<= 0.05))){aux_l="Condición VERDADERA"}
  else {aux_l="Condición FALSA"}
  #legend("topleft",legend=c(aux_l,paste0("n=",n),paste0("n*p=",n*pr),
  paste0("p=",pr)),bg="transparent",cex=0.5)
  par(mfrow=c(1,1))
})
}

```

Cálculos con Python

Sea X una $H(m, n, k)$. Las funciones de `scipy.stats` cambian los parámetros de la forma siguiente:

- M es el número total de bolas. Con nuestra parametrización $M = m + n$.
- n es el número de bolas blancas. Con nuestra parametrización $n = m$.

- N es el número de extracciones. Con nuestra parametrización $N = k$.

```
from scipy.stats import hypergeom
```

Los cálculos realizados anteriormente en R serían:

```
hypergeom.pmf(1,M=15+10,n=15,N=3)
```

```
## 0.29347826086956635
```

```
hypergeom.cdf(1,M=15+10,n=15,N=3)
```

```
## 0.34565217391304481
```

```
1-hypergeom.cdf(1,M=15+10,n=15,N=3)
```

```
## 0.65434782608695519
```

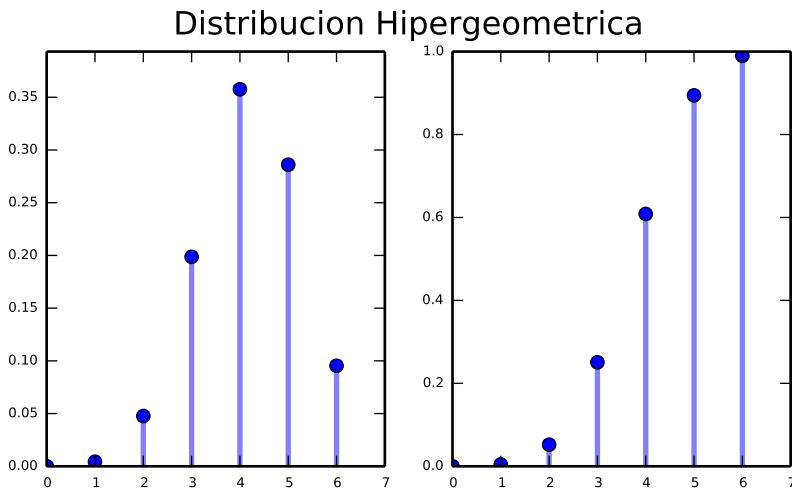
Una muestra aleatoria de este experimento sería:

```
hypergeom.rvs(M=15+10,n=15,N=3,size=100)
```

```
## array([1, 2, 2, 2, 2, 2, 2, 3, 2, 1, 1, 1, 1, 2, 2, 0, 2, 2, 3, 1, 0, 3,
##        1, 2, 2, 3, 2, 1, 2, 1, 0, 2, 2, 3, 1, 3, 2, 2, 2, 3, 1, 2, 3, 2, 3,
##        2, 2, 3, 1, 2, 1, 2, 3, 2, 2, 2, 2, 0, 1, 1, 1, 1, 3, 1, 2, 2, 3, 2,
##        2, 2, 2, 1, 2, 2, 1, 3, 2, 2, 1, 1, 3, 2, 3, 1, 3, 2, 3, 3, 1, 1, 3,
##        2, 0, 2, 2, 2, 2, 2])
```

Los gráficos de la función de probabilidad y de la función de distribución en python se realizan de la forma siguiente:

```
from scipy.stats import hypergeom
[M, n, N] = [20, 7, 12] ##20 elementos, 7 del tipo, extraemos 12
x = np.arange(max(0, N-M+n), min(n, N))
fig = plt.figure(figsize=(5, 2.7))
ax = fig.add_subplot(1, 2, 1)
ax.plot(x, hypergeom.pmf(x, M, n, N), 'bo', ms=5, label='hypergeom pmf')
ax.vlines(x, 0, hypergeom.pmf(x, M, n, N), colors='b', lw=2, alpha=0.5)
ax.set_ylim([0, max(hypergeom.pmf(x, M, n, N))*1.1])
for tick in ax.xaxis.get_major_ticks():
    tick.label.set_fontsize(5)
for tick in ax.yaxis.get_major_ticks():
    tick.label.set_fontsize(5)
ax = fig.add_subplot(1, 2, 2)
ax.plot(x, hypergeom.cdf(x, M, n, N), 'bo', ms=5, label='hypergeom cdf')
ax.vlines(x, 0, hypergeom.cdf(x, M, n, N), colors='b', lw=2, alpha=0.5)
for tick in ax.xaxis.get_major_ticks():
    tick.label.set_fontsize(5)
for tick in ax.yaxis.get_major_ticks():
    tick.label.set_fontsize(5)
fig.suptitle('Distribucion Hipergeometrica')
plt.show()
```



3.2. Cuantiles de distribuciones notables discretas

Ejemplo

Consideremos una v.a. X de distribución $B(5, 0,5)$.

Los cuantiles $x_{0,3}$, $x_{0,6}$ y $x_{0,8}$ son los siguientes:

```
qbinom(c(0.3,0.6,0.8),5,0.5)
```

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

Calculemos a mano, el valor $x_{0,3}$ y verifiquemos que da el mismo resultado que nos ha dado R.

La función de distribución de X es:

$$F_x(x) = P(X \leq x) = \begin{cases} 0, & x < 0, \\ 0,03125, & \text{si } 0 \leq x < 1, \\ 0,18750, & \text{si } 1 \leq x < 2, \\ 0,50000, & \text{si } 2 \leq x < 3, \\ 0,81250, & \text{si } 3 \leq x < 4, \\ 0,96875, & \text{si } 4 \leq x < 5, \\ 1,00000, & \text{si } 5 \leq x. \end{cases}$$

El cuantil $q = 0,3$ es el primer valor $x \in D_X$ tal que $F_X(x) = P(X \leq x_{0,3}) \geq 0,3$. Mirando la expresión anterior, comprobamos que $x_{0,3} = 2$ ya que $F_X(2) = P(X \leq 2) = 0,5 \geq 0,3$.

Ejercicio

Calcular los cuantiles de 0,6 y 0,8 de una $B(5, 0,5)$.

Gráfico interactivo que muestra los cuantiles de las distribuciones $B(n, p)$ y $Po(\lambda)$

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.

```

fluidPage(
  fluidRow(
    column(3,
      sliderInput("nq", label = "Par. n B(n,p)",
                  min = 1, max = 20, value = 10, step = 1)
    ),
    column(3,
      sliderInput("pq", label = "Par. p B(n,p)",
                  min = 0.01, max = 0.99, value = 0.5, step = 0.1)
    ),
    column(3,
      sliderInput("qq", label = "Cuantil q", value = 0.75, min = 0.01, max = 0.99,
                  step = 0.01)
    ),
    column(3,
      sliderInput("lq", label = "Par. lambda Po(lambda)", value = 5, min = 1, max = 20,
                  step = 1)
    )
  )

  renderPlot({
    n = input$nq
    p = input$pq
    q = input$qq
    lambda = input$lq
    par(mfrow = c(1, 2))
    #n=10;p=0.5;q=0.75;lambda=5
    #xx=c(seq(min(a,x),min(b,x),by=0.001))
    probsB = pbinom(0:n, n, p)
    curve(pbinom(x, n, p), xlim = c(0 - 0.25, n + 0.25), ylim = c(0, max(probsB + 0.05, 0.1)),
           col = "blue", main = paste0("Función distribución\n B(n=", n, ", p=", p, ")"),
           ylab = paste0("dbinom(x, n, , p, )"), yaxt = "n")
    segments(x0 = qbinom(q, n, p), y0 = 0, x1 = qbinom(q, n, p), y1 = q, lty = 2, col = "red")
    segments(x0 = qbinom(q, n, p), y0 = q, x1 = -0.25, y1 = q, lty = 2, col = "red")
    ytick = c(0.0, q, 1)
    axis(side = 2, at = ytick, labels = TRUE)
    axis(side = 1, at = qbinom(q, n, p), labels = TRUE)
    curve(ppois(x, lambda), xlim = c(0 - 0.25, 2.5 * lambda), ylim = c(0, 1 + 0.1),
           col = "blue", main = paste0("Función distribución \n Po(lambda=", lambda, ")"),
           ylab = paste0("dpois(x, lambda", lambda, ")"), yaxt = "n")
    segments(x0 = qpois(q, lambda), y0 = 0, x1 = qpois(q, lambda), y1 = q, lty = 2, col = "red")
    segments(x0 = qpois(q, lambda), y0 = q, x1 = -0.25, y1 = q, lty = 2, col = "red")
  })
)

```

```

  ytick=c(0.0,q,1)
  axis(side=2, at=ytick, labels = TRUE)
  axis(side=1, at=qpois(q,lambda), labels = TRUE)
  par(mfrow=c(1,1))
}

```

3.3. Distribuciones continuas

3.3.1. Distribución uniforme

Una v.a. continua X tiene una distribución uniforme sobre el intervalo real (a, b) , con $a < b$, si su función de densidad es

$$f_X(x) = \begin{cases} \frac{1}{b-a}, & \text{si } a < x < b, \\ 0, & \text{en cualquier otro caso.} \end{cases}$$

Ejercicio

Comprobar que el área comprendida entre f_X y la horizontal vale 1.

El área pedida vale:

$$\int_{-\infty}^{+\infty} f_x(x) \cdot dx = \int_a^b \frac{1}{b-a} \cdot dx = \left[\frac{x}{b-a} \right]_{x=a}^{x=b} = \frac{b}{b-a} - \frac{a}{b-a} = \frac{b-a}{b-a} = 1.$$

Su función de distribución es:

$$F_X(x) = \begin{cases} 0, & \text{si } x \leq a, \\ \frac{x-a}{b-a}, & \text{si } a < x < b, \\ 1, & \text{si } b \leq x. \end{cases}$$

Efectivamente:

- Si $x \leq a$, entonces

$$F_X(x) = \int_{-\infty}^x f(t) \cdot dt = \int_{-\infty}^x 0 \cdot dt.$$

- Si $a < x < b$ entonces ,

$$\begin{aligned} F_X(x) &= \int_{-\infty}^x f(t) \cdot dt = \int_{-\infty}^a 0 \cdot dt + \int_a^x \frac{1}{b-a} \cdot dt \\ &= 0 + \left[\frac{t}{b-a} \right]_{t=a}^{t=x} = \frac{x}{b-a} - \frac{a}{b-a} = \frac{x-a}{b-a}. \end{aligned}$$

- Por último si $x \geq b$ entonces,

$$\begin{aligned} F_X(x) &= \int_{-\infty}^x f(t) dt = \int_a^b \frac{1}{b-a} dt = \left[\frac{t}{b-a} \right]_{t=a}^{t=b} \\ &= \frac{b}{b-a} - \frac{a}{b-a} = \frac{b-a}{b-a} = 1. \end{aligned}$$

Denotaremos a la v.a. X uniforme en el intervalo (a, b) por $U(a, b)$.

Esperanza y varianza para una v.a. X $U(a, b)$

Calculemos la esperanza de X :

$$\begin{aligned} E(X) &= \int_{-\infty}^{+\infty} x \cdot f_X(x) dx = \int_a^b x \cdot \frac{1}{b-a} dx = \left[\frac{x^2}{2 \cdot (b-a)} \right]_{x=a}^{x=b} \\ &= \frac{b^2}{2 \cdot (b-a)} - \frac{a^2}{2 \cdot (b-a)} = \frac{b^2 - a^2}{2 \cdot (b-a)} = \frac{(b+a) \cdot (b-a)}{2 \cdot (b-a)} = \frac{b+a}{2}. \end{aligned}$$

De cara a calcular su varianza, calculemos primero la esperanza de X^2 :

$$\begin{aligned} E(X^2) &= \int_{-\infty}^{+\infty} x^2 f_X(x) dx = \int_a^b x^2 \frac{1}{b-a} dx = \left[\frac{x^3}{3 \cdot (b-a)} \right]_{x=a}^{x=b} \\ &= \frac{b^3 - a^3}{3 \cdot (b-a)} = \frac{b^2 + ab + a^2}{3}. \end{aligned}$$

Ejercicio

- Demostrad que la igualdad $b^3 - a^3 = (b-a) \cdot (b^2 + ab + a^2)$ es cierta.
- Utilizadla para el cálculo final del valor de $E(X^2)$.

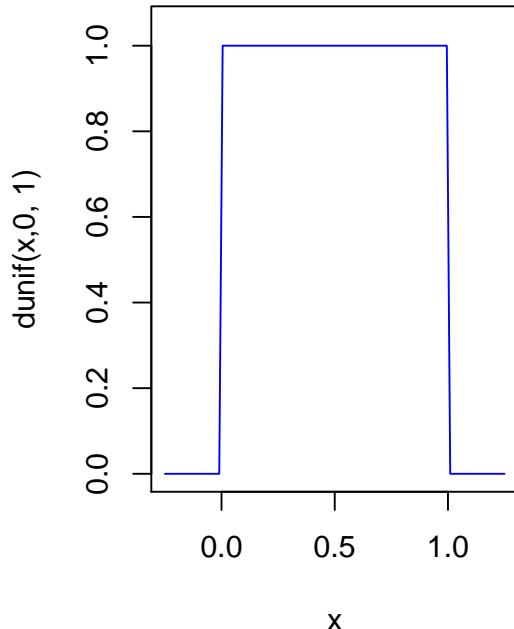
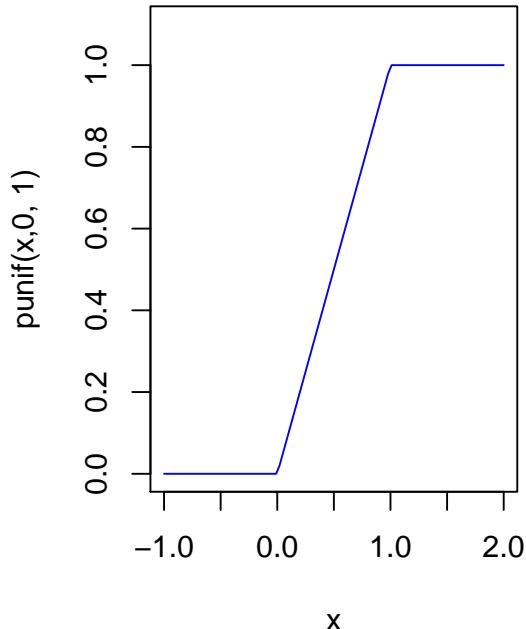
Calculemos $Var(X)$.

$$\begin{aligned} Var(X) &= E(X^2) - (E(X))^2 = \frac{b^2 + ab + a^2}{3} - \left(\frac{b+a}{2} \right)^2 \\ &= \frac{b^2 + ab + a^2}{3} - \frac{b^2 + 2ab + a^2}{4} \\ &= \frac{4 \cdot (b^2 + ab + a^2) - 3 \cdot (b^2 + 2ab + a^2)}{12} \\ &= \frac{b^2 - 2ab + a^2}{12} = \frac{4 \cdot 3}{12} \cdot \frac{(b-a)^2}{12}. \end{aligned}$$

Gráficas $U(0, 1)$

El código en R para dibujar la función de densidad y la función de distribución de una distribución $U(0, 1)$ es el siguiente:

```
par(mfrow=c(1, 2))
a=0;b=1
curve(dunif(x,a,b),xlim=c(a-0.25,b+0.25),ylim=c(0,max(1/(b-a)+0.05,0.1)),
       col="blue",main=paste0("Función densidad U(",a,",",b,")"),
       ylab=paste0("dunif(x,",a," , ",b,")"))
)
curve(punif(x,a,b),xlim=c(a-1,b+1),ylim=c(0,1.1),
       col="blue",main=paste0("Función de distribución U(",a,",",b,")"),
       ylab=paste0("punif(x,",a," , ",b,")",cex.axis=0.8)
)
par(mfrow=c(1,1))
```

Función densidad $U(0,1)$ **Función de distribución $U(0,1)$** **Gráficas interactivas $U(a, b)$**

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.

```
fluidPage(
  fluidRow(
    column(4,
      sliderInput("a1", label = "Parámetro a",
                  min = -5, max = 9, value = 0, step = 0.1)
    ),
    column(4,
      sliderInput("b1", label = "Parámetro b",
                  min = 10, max = 15, value = 5, step = 0.1)
    ),
    column(4,
      sliderInput("x1", label = "x", value = 9, min = -5, max = 15, step = 0.1)
    )
  )

  renderPlot({
    a = input$a1
    b = input$b1
    x = input$x1
    plot(dunif(x, a, b), type = "l")
  })
)
```

```

b=input$b1
x=input$x1
par(mfrow=c(1,2))
#a=0;b=1;x=0.25
xx=c(seq(min(a,x),min(b,x),by=0.001))
curve(dunif(x,a,b),xlim=c(a-0.25,b+0.25),ylim=c(0,max(1/(b-a)+0.05,0.1)),
       col="blue",main=paste0("Función densidad U(",a,",",b,")"),
       ylab=paste0("dunif(x,",a,",",b,")"),xaxt="n")
axis(side=1, at=c(a,x,b), labels = TRUE)
polygon(x=c(a,xx,min(x,b)),y=c(0,dunif(xx,a,b),0),
        density=20,col="skyblue")
curve(punif(x,a,b),xlim=c(a-1,b+1),ylim=c(0,1.1),col="blue",
       main=paste0("Función de distribución U(",a,",",b,")"),
       ylab=paste0("punif(x,",a,",",b,")"),xaxt="n",yaxt="n")
segments(x0=x,y0=0,x1=x,y1=punif(x,a,b),col="red",lty=2)
segments(x0=a-1.01,y0=punif(x,a,b),x1=x,y1=punif(x,a,b),col="red",lty=2)
axis(side=2, at=c(0,round(punif(x,a,b),1),1), labels = TRUE)
axis(side=1, at=c(a,x,b), labels = TRUE)
par(mfrow=c(1,1))
})

```

Transformación lineal de la v.a. uniforme

Si X sigue una distribución $U(a, b)$ entonces $Z = \frac{X-a}{b-a}$ sigue una distribución $U(0, 1)$.

Propiedad: Transformación lineal de la v.a. uniforme

Sea X una v.a $U(a, b)$.

Si $scale \neq 0$ y loc son dos constantes reales, entonces

- si $scale > 0$, $T = scale \cdot X + loc$ sigue una ley $U(scale \cdot a + loc, scale \cdot b + loc)$
- si $scale < 0$, $T = scale \cdot X + loc$ sigue una ley $U(scale \cdot b + loc, scale \cdot a + loc)$

Demostración

Supongamos que X sigue una ley $U(a, b)$, que $scale > 0$ y que $T = scale \cdot X + loc$. Dejamos el caso $scale < 0$ como ejercicio.

La función de distribución de X es:

$$F_X(x) = P(X \leq x) = \begin{cases} 0, & \text{si } x \leq a, \\ \frac{x-a}{b-a}, & \text{si } a \leq x \leq b, \\ 1, & \text{si } b \leq x. \end{cases}$$

Si T vale $T = scale \cdot X + loc$, su función de distribución será:

$$\begin{aligned}
 F_T(t) &= P(T \leq t) = P(scale \cdot X + loc \leq t) = P\left(X \leq \frac{t-loc}{scale}\right) = F_X\left(\frac{t-loc}{scale}\right) \\
 &= \begin{cases} 0, & \text{si } \frac{t-loc}{scale} \leq a \\ \frac{\frac{t-loc}{scale}-a}{b-a}, & \text{si } a \leq \frac{t-loc}{scale} \leq b, \\ 1, & \text{si } b \leq \frac{t-loc}{scale}, \end{cases} \\
 &= \begin{cases} 0, & \text{si } t \leq scale \cdot a + loc, \\ \frac{t-(scale \cdot a + loc)}{scale \cdot (b-a)}, & \text{si } scale \cdot a + loc \leq t \leq scale \cdot b + loc, \\ 1, & \text{si } scale \cdot b + loc \leq t, \end{cases} \\
 &= \begin{cases} 0, & \text{si } t \leq scale \cdot a + loc, \\ \frac{t-(scale \cdot a + loc)}{scale \cdot b + loc - (scale \cdot a + loc)}, & \text{si } scale \cdot a + loc \leq t \leq scale \cdot b + loc, \\ 1, & \text{si } scale \cdot b + loc \leq t, \end{cases}
 \end{aligned}$$

función que corresponde a la función de distribución de una v.a. $U(scale \cdot a + loc, scale \cdot b + loc)$, como queríamos demostrar.

Ejercicio

Sea X una variable $U(0, 1)$ y sea $T = scale \cdot X + loc$:

- Si T es $U(-5, 5)$ ¿qué valores toman $scale$ y loc ?
- Si $loc = -10$ y $scale = 10$ ¿qué distribución de probabilidad sigue T ?
- Si $loc = 0$ y $scale = -1$ ¿qué distribución de probabilidad sigue T ?

Resumen v.a con distribución uniforme, $U(a, b)$

Distribución uniforme	$U(a, b)$
Dominio	$D_X = (a, b)$
$f_X(x)$	$\begin{cases} \frac{1}{b-a}, & \text{si } a < x < b, \\ 0, & \text{en cualquier otro caso.} \end{cases}$
$F_X(x) = P(X \leq x) =$	$\begin{cases} 0, & \text{si } x \leq a \\ \frac{x-a}{b-a}, & \text{si } a \leq x \leq b, \\ 1, & \text{si } b \leq x. \end{cases}$
$E(X) = \frac{a+b}{2}$	$Var(X) = \frac{(b-a)^2}{12}$

Cálculos con R

Sea X una v.a. $U(a, b)$. Las funciones `dunif(x, a, b)` y `punif(x, a, b)` calculan la función de densidad y de distribución de X en el valor X . Por ejemplo, para $a = -1$, $b = 1$ y $x = 0,5$, los valores $f_X(x)$ y $F_X(x)$ valen:

```
dunif(x=0.5, min=-1,max=1)
```

```
## [1] 0.5
```

```
punif(q=0.5,min=-1,max=1)
```

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

La función `runif(n,a,b)` calcula una muestra de observaciones de tamaño n que sigan la distribución $U(a,b)$:

```
runif(n=5,min=-1,max=1)

## [1] -0.5502232 -0.4271506 -0.0515360  0.8052209  0.8983523
```

Por defecto, el valor de los parámetros `a` y `b` son 0 y 1, respectivamente:

```
dunif(x=0.5)
```

```
## [1] 1

punif(q=0.5)

## [1] 0.5

runif(n=5)

## [1] 0.382302539 0.009313886 0.351767001 0.294007361 0.071581515
```

Cálculos con Python

Sea X una v.a. $U(-1,1)$. Tomando como “base” la v.a. $U(0,1)$, los parámetros `loc` y `scale` valen: `loc=-1` y `scale=2`, ya que como hemos visto $X = 2 * U(0,1) - 1 = U(-1,1)$.

En Python, hay que usar dichos parámetros para calcular la función de densidad y de distribución:

```
from scipy.stats import uniform
uniform.pdf(0.5,loc=-1,scale=2)
```

```
## 0.5

uniform.ppf(0.5,loc=-1,scale=2)
```

```
## 0.0
```

Para generar una muestra de valores aleatorios, hay que usar la función `uniform.rvs`:

```
uniform.rvs(size=30,loc=-1,scale=2)
```

```
## array([ 0.48563877,  0.71491272, -0.51394451, -0.3376793 ,  0.87415556,
##        -0.27053797,  0.40585903,  0.72249594,  0.0342319 ,  0.31192043,
##        -0.09849789,  0.70697561, -0.64313489,  0.62142249, -0.71997377,
##        0.19841412,  0.37794679, -0.92802187,  0.6262023 ,  0.68311893,
##        -0.88836153,  0.2446337 ,  0.59574014, -0.89194793, -0.40223801,
##        0.47208819, -0.40267488, -0.53330863,  0.46773476, -0.52867392])
```

Los valores de los parámetros por defecto son `loc=0`, `scale=1`:

```
uniform.pdf(0.5)
```

```
## 1.0
```

```
uniform.ppf(0.5)
```

```
## 0.5
```

```
uniform.rvs(size=5)
## array([ 0.65116177,  0.00680686,  0.88286408,  0.46769238,  0.31648847])
```

3.3.2. Distribución exponencial

Distribución del tiempo entre dos eventos Poisson

Supongamos que tenemos un proceso Poisson con parámetro λ en una unidad de tiempo.

Dado un tiempo t , definimos N_t como el número de eventos en el intervalo de tiempo $(0, t]$. La distribución de $N(t)$ es una $Po(\lambda \cdot t)$. Consideremos la v.a. T como el tiempo transcurrido entre dos eventos Poisson consecutivos.

Sea $t > 0$, entonces

$$\begin{aligned} P(T > t) &= P(\text{Cero eventos en el intervalo}(0, t]) \\ &= P(N_t = 0) = \frac{(\lambda t)^0}{0!} e^{-\lambda t} = e^{-\lambda t}. \end{aligned}$$

Tomando complementarios, la función de distribución de T será:

$$F_T(t) = P(T \leq t) = 1 - P(T > t) = \begin{cases} 0, & \text{si } t \leq 0, \\ 1 - e^{-\lambda t}, & \text{si } t > 0, \end{cases}$$

Para hallar la función de densidad de T , basta derivar la expresión anterior:

$$f_T(t) = \begin{cases} \lambda \cdot e^{-\lambda t}, & \text{si } t > 0, \\ 0, & \text{si } t \leq 0. \end{cases}$$

Llamaremos a la variable T exponencial de parámetro λ y la denotaremos por $Exp(\lambda)$.

Propiedad de la falta de memoria

Sea X una v.a. $Exp(\lambda)$, entonces

$$P(X > s + t | X > s) = P(X > t) \text{ para todo } s, t \in \mathbb{R}$$

Demostración

Si X es una v.a. $Exp(\lambda)$ tenemos que $P(X > x) = 1 - P(X \leq x) = 1 - (1 - e^{-\lambda \cdot x}) = e^{-\lambda \cdot x}$ para todo $x > 0$.

Por tanto,

$$\begin{aligned} P(X > s + t | X > s) &= \frac{P(\{X > s+t\} \cap \{X > s\})}{P(X > s)} = \frac{P(X > s+t)}{P(X > s)} = \frac{e^{-\lambda \cdot (s+t)}}{e^{-\lambda \cdot s}} = \frac{e^{-\lambda \cdot s} \cdot e^{-\lambda \cdot t}}{e^{-\lambda \cdot s}} \\ &= e^{-\lambda \cdot t} = P(X > t). \end{aligned}$$

Ejemplo: el clásico problema del peluquero.

Una pequeña peluquería es regentada por un único peluquero. El peluquero está esperando al próximo cliente mientras lee el periódico.

Supongamos que la v.a. N_T , que representa el número de clientes que llegan en el intervalo $[0, t)$, es una $Po(\lambda \cdot t)$ entonces la variable T , tiempo entre dos clientes consecutivos, sigue una ley $Exp(\lambda)$.

Supongamos que t se mide en horas y que $\lambda = 4$ es el promedio de clientes por hora.

Se pide:

1. El tiempo esperado (en horas) y la varianza hasta el siguiente cliente.
2. ¿Cuál es la probabilidad de que nuestro peluquero esté sin clientes (leyendo el periódico) más de 30 minutos (0.5 horas)?

En este ejemplo, la propiedad de la pérdida de memoria significa que por ejemplo, si el peluquero lleva ya esperando más de $s > 0,25$ (un cuarto de hora), la probabilidad de que espere $t = 1/6$ de hora más (10 minutos) no cambia, sigue siendo $P(T > 0,25 + 1/6 | T > 0,25) = P(T > 1/6)$.

Solución de 1. El tiempo esperado (en horas) y la varianza hasta el siguiente cliente.

El tiempo esperado (en horas) hasta el siguiente cliente es

$$E(X) = \frac{1}{\lambda} = \frac{1}{4} = 0,25,$$

y la varianza es

$$Var(X) = \frac{1}{\lambda^2} = \frac{1}{4^2} = 0,0625.$$

Solución de 2. ¿Cuál es la probabilidad de que nuestro peluquero esté sin clientes (leyendo el periódico) más de 30 minutos (0.5 horas)?

La probabilidad pedida vale:

$$P(X > 0,5) = 1 - P(X \leq 0,5) = 1 - (1 - e^{-4 \cdot 0,5}) = e^{-2} = 0,1353353.$$

Usando R, la probabilidad anterior puede ser calculada de la forma siguiente:

```
pexp(0.5,rate=3)
```

```
## [1] 0.7768698
```

```
1-pexp(0.5,rate=3)
```

```
## [1] 0.2231302
```

```
pexp(0.5,rate=3,lower.tail = FALSE)
```

```
## [1] 0.2231302
```

Cálculos con R

Las funciones de densidad y de distribución de una variable exponencial de parámetro λ en un valor x se pueden obtener en R usando las funciones `dexp(x,lambda)` y `pexp(x,lambda)`, respectivamente. Para generar n valores aleatorios de una variable exponencial de parámetro λ , hay que usar la función `rexp(n,lambda)`. Veamos un ejemplo de aplicación de las tres funciones anteriores:

```
dexp(0.001,rate=3) # Alerta! No es una probabilidad, es una densidad y puede ser > 1
## [1] 2.991013
pexp(0.5,rate=3) # P(X<0.5)

## [1] 0.7768698
rexp(10,3) # Diez tiempos de una exponencial con lambda = 3
## [1] 0.32072389 0.57998323 0.26019628 0.48896233 0.31041354 0.11871455
## [7] 0.10798351 0.18699902 0.10023100 0.03619832
```

Cálculos con Python

Las funciones de densidad y de distribución de una variable exponencial de parámetro λ en un valor x se pueden obtener en Python usando las funciones `expon.pdf(x,scale=1/lambda)` y `expon.cdf(x,scale=1/lambda)`, respectivamente. Para generar n valores aleatorios de una variable exponencial de parámetro λ , hay que usar la función `expon.rvs(scale=1/lambda,size=n)`. Veamos un ejemplo de aplicación de las tres funciones anteriores:

```
from scipy.stats import expon
expon.pdf(0.0001,scale= 1./3)

## 2.9991001349865014
expon.cdf(0.5,scale= 1./3)

## 0.77686983985157021
expon.rvs(scale=1./3,size=10)

## array([ 0.29273001,  0.05024259,  0.09494715,  0.19175267,  0.10876354,
##        0.39534875,  0.16773158,  0.01956584,  0.41387611,  0.20937501])
```

Resumen v.a con distribución exponencial, $Exp(\lambda)$

X	$Exp(\lambda)$
$D_X =$	$(0, +\infty)$
$f_X(x) =$	$\begin{cases} \lambda e^{-\lambda x}, & \text{si } x > 0, \\ 0, & \text{si } x \leq 0. \end{cases}$
$F_X(x) = P(X \leq X) =$	$\begin{cases} 0, & \text{si } x \leq 0, \\ 1 - e^{-\lambda x}, & \text{si } x > 0. \end{cases}$
$E(X) = \frac{1}{\lambda}$	$Var(X) = \frac{1}{\lambda^2}$

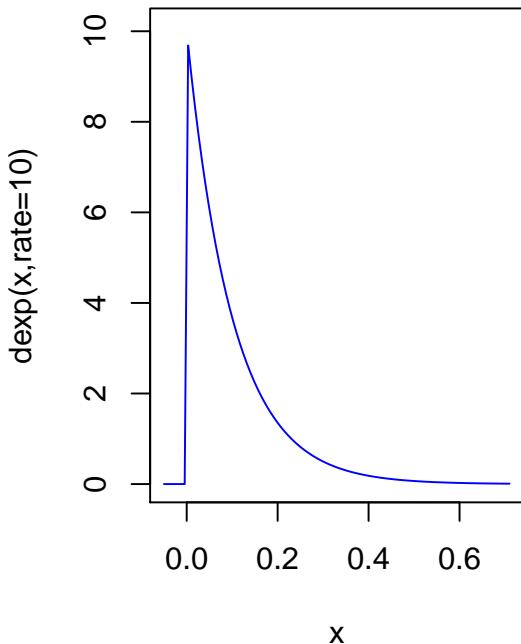
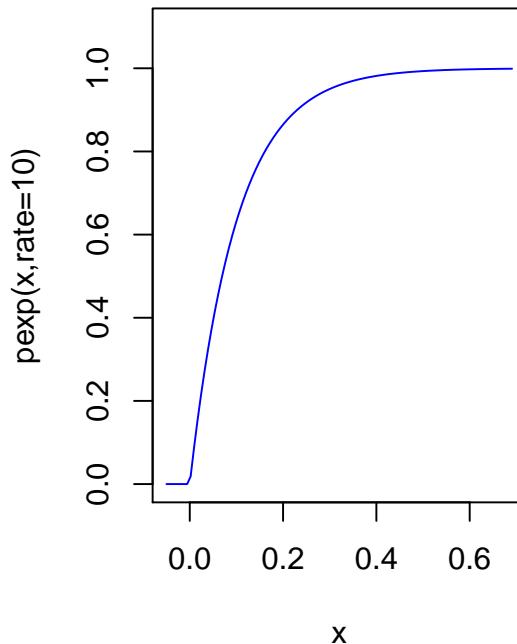
Gráficas de las funciones de densidad y de distribución de la variable aleatoria $Exp(\lambda = 10)$

```
lambda=10
par(mfrow=c(1,2))
curve(dexp(x,rate=lambda),xlim=c(-0.05,round(qexp(0.99,rate=lambda,2),2)+0.25),
      ylim=c(0,dexp(0,lambda)+0.1),col="blue",
      main=paste0("Función densidad Exp(",lambda,")")),
```

```

ylab=paste0("dexp(x,rate=",lambda,""))
curve(pexp(x,rate=lambda),xlim=c(-0.05,qexp(0.999,10)),ylim=c(0,1.1),col="blue",
      main=paste0("Función de distribución Exp(",lambda,""),
      ylab=paste0("pexp(x,rate=",lambda,""))
par(mfrow=c(1,1))

```

Función densidad Exp(10)**Función distribución Exp(10)****Ejercicio**

Consultad en el manual de Python `scipy.stats`.

Dibujad la función de densidad y de distribución de una $Exp(\lambda = 10)$.

Gráficas interactivas de una variable $Exp(\lambda)$

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.

```

fluidPage(
  fluidRow(
    column(4,
      sliderInput("le", label = "lambda",
                  min = 0.1, max = 3, value = 1, step = 0.1)
    ),
    column(4,
      sliderInput("xe", label = "X=x",

```

```

    min = 0, max = 5, value = 5, step = 0.1)
),
column(4,
  sliderInput("pe", label = "Cuantil p",
              min = 0.01, max = 1, value = 0.75, step = 0.01)
)
)

renderPlot({
  lambda=input$le
  p=input$pe
  x=input$xe
  #lambda=10;p=0.75;x=0.4
  xx=seq(0,x,by=0.001)
  par(mfrow=c(1,2))
  curve(dexp(x,rate=lambda),xlim=c(-0.05,round(qexp(0.999,rate=lambda),2)),
         ylim=c(0,dexp(0,lambda)+0.1),col="blue",
         main=paste0("Función densidad Exp(",lambda,")"),
  ylab=paste0("dexp(x,",lambda,")"),xaxt="n")
  axis(side=1, at=c(0,x,round(qexp(0.999,rate=lambda),2)),cex.axis=0.8)
  polygon(x=c(0,xx,max(x,xx)),y=c(0,dexp(xx, rate=lambda),0),
           density=20,col="skyblue")
  curve(pexp(x,rate=lambda),xlim=c(0.01,qexp(0.999,rate=lambda)+0.1),
         ylim=c(0,1.1),col="blue",
         main=paste0("Función de distribución Exp(",lambda,")"),
  ylab=paste0("pexp(x,",lambda,")"),xaxt="n",yaxt="n")
  segments(x0=qexp(p,lambda),x1=qexp(p,lambda),y0=0,y1=p,col="red",lty=2)
  segments(x0=0-0.05,y0=p,x1=qexp(p,lambda),y1=p,col="red",lty=2)
  axis(side=2, at=seq(0,1,0.1), labels = TRUE)
  axis(side=1, at=seq(0,round(qexp(0.999,rate=lambda),2),by=0.1), labels = TRUE)
  par(mfrow=c(1,1))
})
}

```

Ejercicio: las bombillas que no envejecen.

Supongamos que compramos una bombilla led que promete un **valor esperado** de duración de 10000 (1.14 años) horas de funcionamiento continuo. Además, nos aseguran que la distribución de X , el número de horas de funcionamiento continuo de una bombilla led, sigue una ley exponencial.

- Si X es $Exp(\lambda)$ ¿cuál es el valor del parámetro λ ?
 - ¿Cuál es la probabilidad de que una bombilla led ilumine más de 2 años?
 - Supongamos que ya tengo una bombilla led funcionando 1 año ¿Cuál es la probabilidad de que dure dos años más?
 - ¿Cuál es la varianza de la duración en horas de este tipo de bombillas?

3.3.3. Distribución normal o Gaussiana

Una de las variables aleatorias continua más populares es la llamada distribución normal o Gaussiana.

Distribución normal o de Gauss Diremos que una v.a. X sigue una ley normal de parámetros μ y σ y la denotaremos por $N(\mu, \sigma)$ si tiene por función de densidad:

$$f_X(x) = \frac{1}{\sqrt{2 \cdot \pi \cdot \sigma^2}} e^{-\frac{1}{2} \cdot \left(\frac{x-\mu}{\sigma}\right)^2},$$

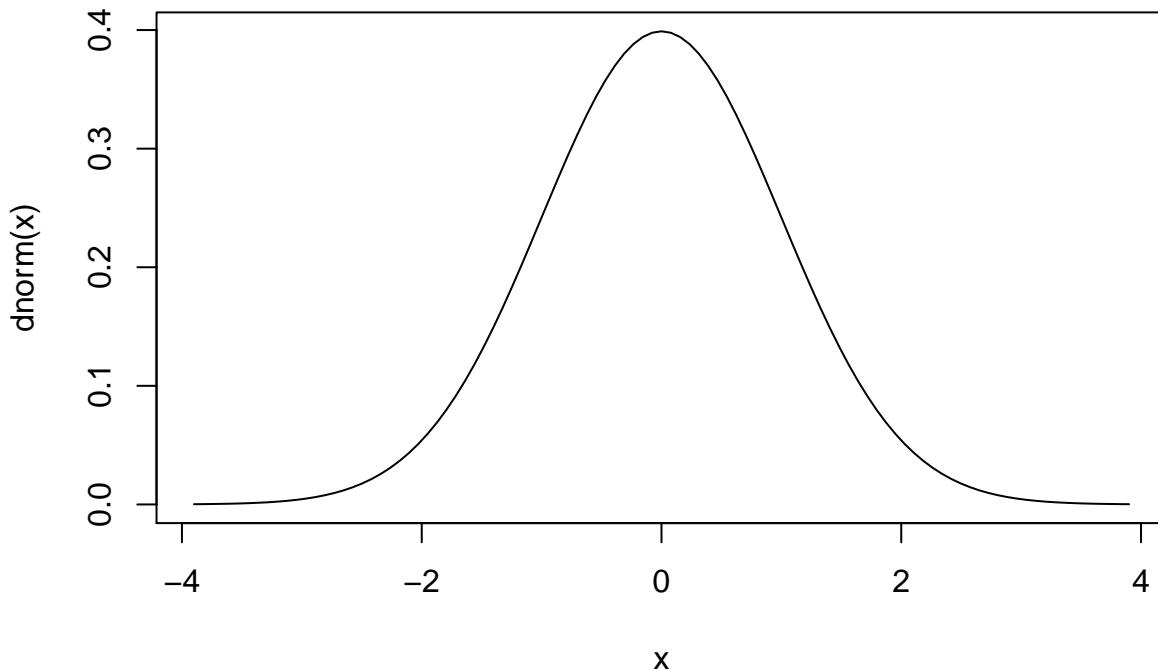
para todo $x \in \mathbb{R}$.

La gráfica de esta función de densidad es conocida como **campana de Gauss**.

La v.a. normal con $\mu = 0$ y $\sigma = 1$ recibe el nombre de normal estándar y se suele denotar por la letra $Z \sim N(0, 1)$.

```
curve(dnorm(x), main="Función de densidad de una normal estándar", xlim=c(-3.9,3.9))
```

Función de densidad de una normal estándar

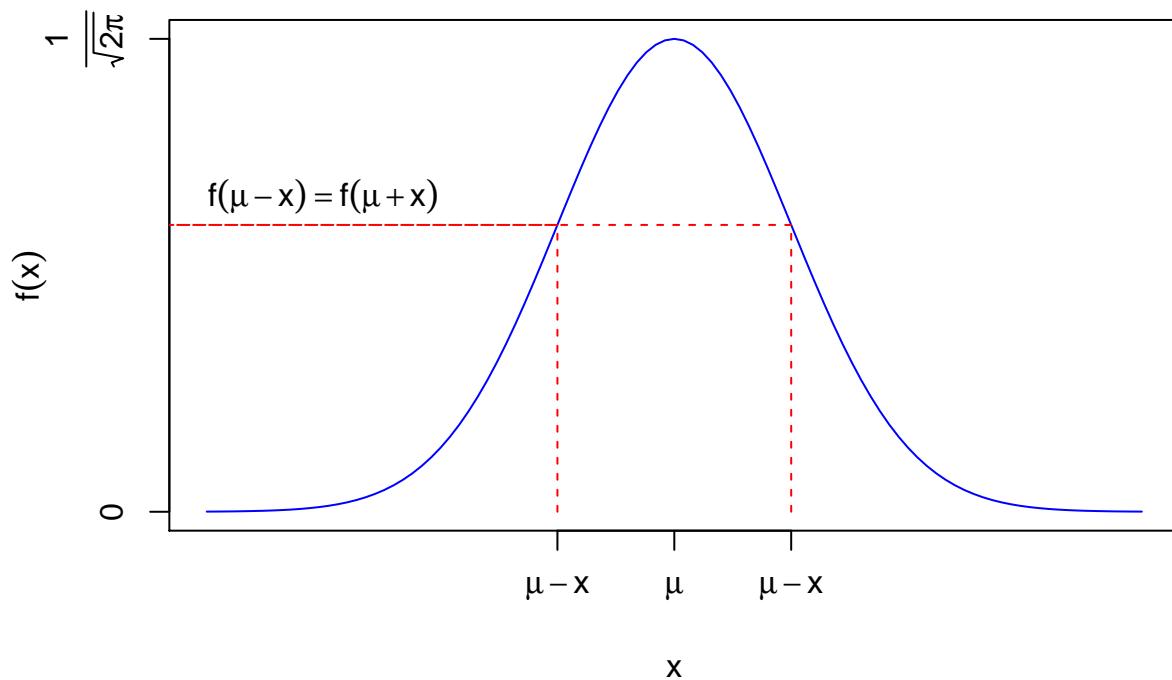


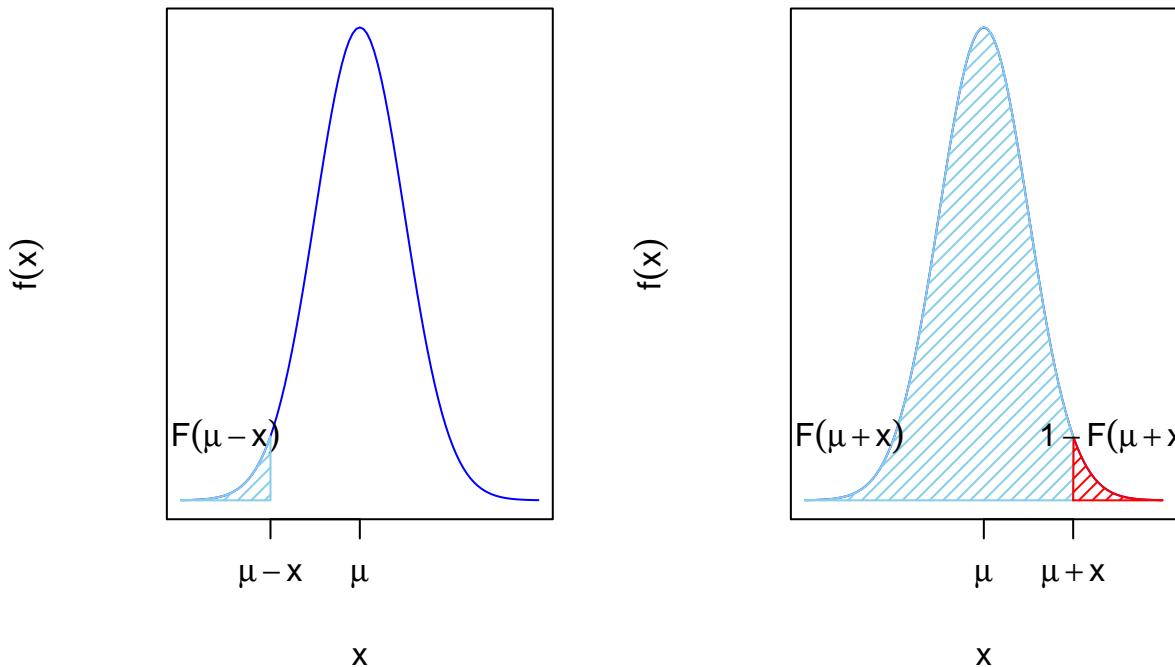
Propiedades de la función de densidad de la distribución normal

Sea X una v.a. $N(\mu, \sigma)$ y sean f_X su función de densidad y $F_X(x) = \int_{-\infty}^x f_X(t) dt$ su función de distribución. Entonces:

1. La función f_X verifica todas las propiedades de las funciones de densidad: $f_X(x) > 0$, para todo $x \in \mathbb{R}$ y $\int_{-\infty}^{\infty} f_X(x) dx = 1$.

2. La función $f_X(x)$ es simétrica respecto de la recta $x = \mu$: $f_X(\mu - x) = f_X(\mu + x)$, para todo $x \in \mathbb{R}$.
3. f_X tiene un único máximo absoluto en $x = \mu$ que vale $f_X(\mu) = \frac{1}{\sqrt{2\pi\sigma^2}}$.
4. Si F_X es la función de distribución de X , entonces $F_X(\mu + x) = 1 - F_X(\mu - x)$, para todo $x \in \mathbb{R}$.
5. En particular, si Z es una $N(0, 1)$, entonces $F_Z(-x) = 1 - F_Z(x)$ para todo $x \in \mathbb{R}$.
6. $Z = \frac{X-\mu}{\sigma}$ es una v.a. $N(0, 1)$ y $X = \sigma \cdot Z + \mu$ es una $N(\mu, \sigma)$ donde Z es la normal estándar.
7. La función f_X es continua.
8. $\lim_{x \rightarrow +\infty} f(x) = \lim_{x \rightarrow -\infty} f(x) = 0$ es decir tiene asíntota horizontal a derecha e izquierda.
9. f es estrictamente creciente si $x < \mu$ y decreciente si $x > \mu$.
10. Tiene dos puntos de inflexión en $x = \mu + \sigma$ y en $x = \mu - \sigma$.





Función de distribución de la $N(0,1)$

Su función de distribución es, como sabemos:

$$F(x) = \int_{-\infty}^x \frac{1}{\sqrt{2 \cdot \pi}} e^{-\frac{x^2}{2}} dt.$$

La función $F(x)$ no tiene ninguna expresión algebraica “decente”. Es por esta razón, y por comodidad, que esta función está tabulada o hay que calcularla usando un software estadístico.

Resumen v.a con distribución normal, $N(\mu, \sigma)$

X	$N(\mu, \sigma)$
$D_X =$	$\mathbb{R} = (-\infty, +\infty)$
$f_X(x) =$	$\frac{1}{\sqrt{2\pi}\sigma} \cdot e^{-\frac{(x-\mu)^2}{2\sigma^2}}$ para todo $x \in \mathbb{R}$.
$F_X(x) = P(X \leq x) =$	Utilizad la función de R <code>pnorm(x, mean=mu, sd=sigma)</code> o la función correspondiente en Python
$E(X) = \mu.$	$Var(X) = \sigma^2.$

Cálculos con R

Las funciones que calculan la función de densidad y de distribución de una variable $N(\mu, \sigma)$ en un valor x son `dnorm(x, mean=mu, sd=sigma)` y `pnorm(x, mean=mu, sd=sigma)`, respectivamente. Por ejemplo, para una variable $X \sim N(\mu = 1, \sigma = 2)$, la función de densidad $f_X(2)$ se puede calcular de la forma siguiente:

```
dnorm(2,mean=1,sd=2)
```

```
## [1] 0.1760327
```

y la función de distribución $F_X(2) = P(X \leq 2)$ de la forma siguiente:

```
pnorm(2,mean=1,sd=2)
```

```
## [1] 0.6914625
```

Si queremos calcular el cuantil x_q de una distribución normal $N(\mu, \sigma)$ que, recordemos es el valor que cumple que $P(X \leq x_q) = q$, tenemos que usar la función `qnorm(q,mean=mu,sd=sigma)`:

```
qnorm(0.95,mean=1,sd=2)
```

```
## [1] 4.289707
```

Para generar n valores aleatorios de una distribución normal $N(\mu, \sigma)$, hay que usar la función `rnorm(n,mean=mu,sd=sigma)`:

```
rnorm(n=5,mean=1,sd=2)
```

```
## [1] -1.7813382 0.7320555 2.7209162 2.9098518 2.5098235
```

Cálculos con Python

Para poder trabajar con una distribución normal $N(\mu, \sigma)$ en Python, tenemos que importar `norm` de `scipy.stats`. Los parámetros μ y σ son `loc` y `scale`, respectivamente.

```
from scipy.stats import norm
```

Por ejemplo, para una variable $X \sim N(\mu = 1, \sigma = 2)$, la función de densidad $f_X(2)$ se puede calcular de la forma siguiente:

```
norm.pdf(2,loc=1,scale=2)
```

```
## 0.17603266338214976
```

y la función de distribución $F_X(2) = P(X \leq 2)$, de la forma siguiente:

```
norm.cdf(2,loc=1,scale=2)
```

```
## 0.69146246127401312
```

Si queremos calcular el cuantil x_q de una distribución normal $N(\mu, \sigma)$, tenemos que usar la función `norm.ppf(q,loc=mu,scale=sigma)`:

```
norm.ppf(0.95,loc=1,scale=2)
```

```
## 4.2897072539029448
```

Para generar n valores aleatorios de una distribución normal $N(\mu, \sigma)$, hay que usar la función `norm.rvs(loc=mu,scale=sigma,size=n)`:

```
norm.rvs(loc=1,scale=2,size=5)
```

```
## array([ 1.16011924,  1.08431651,  1.67778047,  1.50212612, -0.22423355])
```

Ejercicio

Consultad SciPy.org para dibujar las funciones de densidad y de distribución con Python.

Gráficas interactivas usando los parámetros de la distribución normal

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.

```
fluidPage(
  fluidRow(
    column(3,
      sliderInput("m1", label = "mu1",
                  min = -10, max = 10, value = 0 , step = 0.05)
    ),
    column(3,
      sliderInput("s1", label = "sigma1",
                  min = 0.1, max = 5, value = 1, step = 0.1)
    ),
    column(3,
      sliderInput("m2", label = "mu2", value=4, min = -10, max = 10, step = 0.05)
    ),
    column(3,
      sliderInput("s2", label = "sigma2",
                  min = 0.1, max = 5, value = 1, step = 0.1)
    )
  )

  renderPlot({
    m1=input$m1
    m2=input$m2
    s1=input$s1
    s2=input$s2
    mins2=min(c(s1^2,s2^2))
    m=min(c(qnorm(0.01,m1,s1),qnorm(0.01,m2,s2)))
    M=max(c(qnorm(0.99,m1,s1),qnorm(0.99,m2,s2)))

    curve(dnorm(x,m1,s1),xlim=c(m,M),ylim=c(0,1/sqrt(2*pi*mins2)),col="red",lty=1)
    legend("topleft",legend=c(expression(N(mu[1],sigma[1])),
                               expression(N(mu[2],sigma[2]))),
           col=c("red","blue"),lty=c(1,2))
    curve(dnorm(x,m2,s2),add=TRUE,col="blue",lty=2)
  })
)
```

Transformaciones lineales de variables aleatorias normales

Propiedad: transformación lineal la distribución normal

Sea X una variable $N(\mu, \sigma)$ entonces la variable $Y = aX + b$ con $a \neq 0, b \in \mathbb{R}$ tiene distribución $N(a\mu + b, |a| \cdot \sigma)$

En el caso particular en que $a = \frac{1}{\sigma}$ y $b = \frac{-\mu}{\sigma}$ obtenemos la v.a. $Z = \frac{X - \mu}{\sigma}$, que se distribuye según una normal estándar $N(0, 1)$, es decir $E(Z) = 0$ y $Var(Z) = 1$. Dicha operación se denomina estandarización de la distribución normal.

La propiedad anterior nos permite calcular la función de distribución de cualquier v.a. $N(\mu, \sigma)$ a partir de la función de distribución de la distribución $Z = N(0, 1)$:

$$F_X(x) = P(X \leq x) = P\left(\frac{X - \mu}{\sigma} \leq \frac{x - \mu}{\sigma}\right) = P\left(Z \leq \frac{x - \mu}{\sigma}\right) = F_Z\left(\frac{x - \mu}{\sigma}\right).$$

Entonces, basta conocer la manera de calcular F_Z para poder calcular F_X , para $X = N(\mu, \sigma)$, para cualquier μ y cualquier $\sigma > 0$.

Propiedades de la distribución normal estándar

Sea Z una $N(0, 1)$.

En este caso, $\mu = 0$ y $\sigma = 1$. Podemos escribir algunas de las propiedades vistas para una distribución normal cualquiera de la forma siguiente:

- La propiedad $f_X(\mu - x) = f_X(\mu + x)$ se traduce a $f_Z(-x) = f_Z(x)$
- La propiedad $F_X(\mu - x) = 1 - F_X(\mu + x)$ se traduce a $F_Z(-x) = 1 - F(x)$.
- Dado $\delta > 0$,

$$P(-\delta \leq Z \leq \delta) = F_Z(\delta) - F_Z(-\delta) = F_Z(\delta) - (1 - F_Z(\delta)) = 2 \cdot F_Z(\delta) - 1.$$

Ejercicio: cálculos con la distribución normal estándar

Sea Z una distribución $N(0, 1)$. Calcular las siguientes probabilidades en función de F_Z :

- $P(-4 \leq Z \leq 4)$.
- $P(-2 \leq Z \leq 2)$.
- $P(Z \leq -2)$.
- $P(Z \leq 2)$.
- $P(Z \geq 2)$.
- $P(Z > 2)$.
- $P(Z = 2)$.
- $P(Z \geq -2)$.

Resolución:

- $P(-4 \leq Z \leq 4) = F_Z(4) - F_Z(-4) = 2 \cdot F_Z(4) - 1$.
- $P(-2 \leq Z \leq 2) = F_Z(2) - F_Z(-2) = 2 \cdot F_Z(2) - 1$.
- $P(Z \leq -2) = F_Z(-2) = 1 - F_Z(2)$.
- $P(Z \leq 2) = F_Z(2)$.
- $P(Z \geq 2) = 1 - P(Z < 2) = 1 - F_Z(2)$.
- $P(Z > 2) = 1 - P(Z \leq 2) = 1 - F_Z(2)$.
- $P(Z = 2) = 0$ ya que es una distribución continua.
- $P(Z \geq -2) = 1 - P(Z < -2) = 1 - F_Z(-2) = 1 - (1 - F_Z(2)) = F_Z(2)$.

Cálculo de probabilidades de la distribución normal $X = N(\mu, \sigma)$ en un intervalo usando la distribución $Z = N(0, 1)$

Para hallar la probabilidad de que X esté en un intervalo (a, b) cualquiera, podemos usar la función de distribución de Z de la siguiente manera:

$$\begin{aligned} P(a < X < b) &= P\left(\frac{a-\mu}{\sigma} < \frac{X-\mu}{\sigma} < \frac{b-\mu}{\sigma}\right) = \\ &= P\left(\frac{a-\mu}{\sigma} < Z < \frac{b-\mu}{\sigma}\right) = F_Z\left(\frac{b-\mu}{\sigma}\right) - F_Z\left(\frac{a-\mu}{\sigma}\right). \end{aligned}$$

Para el caso particular en que el intervalo esté centrado en la media μ , o sea existe un valor $\delta > 0$ tal que $(a, b) = (\mu - \delta, \mu + \delta)$, obtenemos:

$$P(\mu - \delta \leq X \leq \mu + \delta) = 2 \cdot F_Z\left(\frac{\delta}{\sigma}\right) - 1.$$

Ejercicio: cálculo de probabilidades de una distribución normal

Sea X una normal con media 2 y varianza 4. Calcular

- $P(1 < X < 2)$.
- $P(X > 3)$.

Solución

La primera probabilidad se calcula de la forma siguiente:

$$\begin{aligned} P(1 < X < 2) &= P\left(\frac{1-2}{2} < \frac{X-2}{2} < \frac{2-2}{2}\right) = P\left(-\frac{1}{2} < Z < 0\right) \\ &= F_Z(0) - F_Z(-0,5) = \frac{1}{2} - 1 + F_Z(0,5) = -\frac{1}{2} + F_Z(0,5). \end{aligned}$$

La segunda probabilidad se calcular de la forma siguiente:

$$P(X > 3) = P\left(\frac{X-2}{2} > \frac{3-2}{2}\right) = P(Z > 0,5) = 1 - F_Z(0,5).$$

Ejercicio

Sea X una normal con media 2 y varianza 4. Calcular usando R y con Python las probabilidades siguientes:

- $P(1 < X < 2)$.
- $P(X > 3)$.

Solución usando R

```
pnorm(2,mean=2,sd=2)-pnorm(1,mean=2,sd=2) #P(1< X< 2)
## [1] 0.1914625
pnorm(3,mean=2,sd=2,lower.tail =FALSE) #P(X>3)
## [1] 0.3085375
1-pnorm(3,mean=2,sd=2,lower.tail=TRUE) #P(X>3) = 1-P(X<=3)
## [1] 0.3085375
```

Solución usando Python

```
norm.cdf(2,loc=2,scale=2)-norm.cdf(1,loc=2,scale=2) #P(1< X< 2)
```

```
## 0.19146246127401312
```

```
1-norm.cdf(3,loc=2,scale=2) #P(X>3) = 1-P(X<=3)
```

```
## 0.30853753872598688
```

La distribución normal aproxima otras distribuciones

En los temas que siguen veremos como, bajo determinadas condiciones,

- la distribución normal puede aproximar la distribución binomial,
- la distribución normal puede aproximar la distribución Poisson
- la distribución normal es la distribución límite de la media aritmética de una muestra de variables aleatorias.

Capítulo 4

Variables Aleatorias. Complementos

4.1. Momentos de variables aleatorias

4.1.1. Momento de orden n

Definición. Sea X una variable aleatoria. Definimos el **momento de orden n** como $m_n = E(X^n)$.

Observación. El momento de orden 1 de una variable aleatoria es su valor medio o $E(X)$.

Los momentos de orden n caracterizan una variable X . O sea, que si conocemos todos los momentos de orden n , podemos deducir cuál es la distribución de X .

En general, el cálculo de los momentos de orden n para una variable X es bastante tedioso.

Ejemplos de momentos de orden n

Ejemplo: momento de orden n de una variable de Bernoulli de parámetro p

Sea X una variable de Bernoulli de parámetro p . Recordemos que su función de probabilidad es:

$$P_X(0) = q = 1 - p, \quad P_X(1) = p.$$

Su momento de orden n será:

$$m_n = E(X^n) = p \cdot 1^n + (1 - p) \cdot 0^n = p.$$

En este caso, todos los momentos de orden n valen p .

Ejemplo: momento de orden n de una variable exponencial de parámetro λ

Consideremos ahora una variable X exponencial de parámetro λ .

Recordemos que su función de densidad es: $f_X(x) = \lambda e^{-\lambda x}$, para $x \geq 0$ y 0, en caso contrario.

Su momento de orden n será:

$$m_n = E(X^n) = \int_0^{\infty} \lambda e^{-\lambda x} x^n dx = \frac{n!}{\lambda^n}.$$

La expresión anterior se puede obtener integrando por partes n veces y resolviendo los límites correspondientes. Dejámos al lector los cálculos correspondientes.

Fijémonos que los momentos de orden n tienden a infinito a medida que n crece: $\lim_{n \rightarrow \infty} m_n = \lim_{n \rightarrow \infty} \frac{n!}{\lambda^n} = \infty$.

Ejemplo: momento de orden n de una variable normal de parámetros $m = 0$ y $\sigma = 1$

Recordemos que su función de densidad es: $f_X(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$, para $x \in \mathbb{R}$.

Su momento de orden 1 será la esperanza de X : $m_1 = 0$ i su momento de orden 2 será: $m_2 = E(X^2) = \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} \cdot x^2 dx = 1$. La integral anterior se resuelve usando técnicas de integrales de dos variables. Dicho valor también se puede obtener usando que su varianza vale 1: $m_2 = \text{Var}(X) + E(X)^2 = \sigma^2 + 0^2 = 1$.

Los momentos de orden impar n serán cero ya que integramos una función impar: $m_n = E(X^n) = \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} \cdot x^n dx = 0$. O sea, si consideramos $g(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} \cdot x^n$, se verifica $g(-x) = -g(x)$, para todo $x \in \mathbb{R}$.

Si intentamos calcular el momento de orden 4, obtenemos: $m_4 = E(X^4) = \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} \cdot x^4 dx = 3$, usando técnicas de integración de dos variables otra vez.

4.1.2. Momento central de orden n

Definición. Sea X una variable aleatoria. Definimos el **momento central de orden n** como $\mu_n = E((X - \mu)^n)$, donde $\mu = E(X)$ es la media o la esperanza de la variable aleatoria X .

Observación. El momento central de orden 1 de una variable aleatoria es siempre 0:

$$\mu_1 = E((X - \mu)) = E(X) - E(\mu) = E(X) - E(X) = 0.$$

Observación. El momento central de orden 2 de una variable aleatoria es la varianza:

$$\mu_2 = E((X - \mu)^2) := \text{Var}(X).$$

Los momentos centrales de orden n caracterizan también una variable X . O sea, que si conocemos todos los momentos centrales de orden n , podemos deducir cuál es la distribución de X .

Proposición. La relación que hay entre los momentos centrales y los momentos de una variable aleatoria es la siguiente:

$$\mu_n = \sum_{k=0}^n (-1)^{n-k} \binom{n}{k} \mu^{n-k} m_k = \sum_{k=0}^n (-1)^k \binom{n}{k} \mu^k m_{n-k},$$

donde $\mu = E(X)$ recordemos que es la esperanza de la variable aleatoria X .

Demostración

Recordemos la definición de momento central de orden n y desarrollemos su expresión aplicando el **binomio de Newton**:

$$\mu_n = E((X - \mu)^n) = E \left(\sum_{k=0}^n (-1)^{n-k} \binom{n}{k} X^k \mu^{n-k} \right).$$

Aplicando la propiedad de la esperanza que la esperanza de la suma es la suma de esperanzas, obtenemos la expresión dada por la proposición:

$$\mu_n = \sum_{k=0}^n (-1)^{n-k} \binom{n}{k} \mu^{n-k} E(X^k) = \sum_{k=0}^n (-1)^{n-k} \binom{n}{k} \mu^{n-k} m_k.$$

Ejemplos de momentos centrales de orden n

Ejemplo: momento central de orden n de una variable de Bernoulli de parámetro p

Sea X una variable de Bernoulli de parámetro p . Recordemos que su función de probabilidad es:

$$P_X(0) = q = 1 - p, \quad P_X(1) = p.$$

Usando que $E(X) = p$, su momento central de orden n será:

$$\mu_n = E((X - p)^n) = p \cdot (1 - p)^n + (1 - p) \cdot (0 - p)^n = p(1 - p)^n + (-1)^n (1 - p)p^n.$$

Ejercicio

Demostrar que la expresión anterior corresponde a un polinomio de grado n .

Ejemplo: momento central de orden n de una variable exponencial de parámetro λ

Consideremos ahora una variable X exponencial de parámetro λ .

Recordemos que su función de densidad es: $f_X(x) = \lambda e^{-\lambda x}$, para $x \geq 0$.

Usando que $E(X) = \frac{1}{\lambda}$, su momento central de orden n será:

$$\mu_n = E\left(\left(X - \frac{1}{\lambda}\right)^n\right) = \int_0^\infty \lambda e^{-\lambda x} \left(x - \frac{1}{\lambda}\right)^n dx = \frac{a_n}{\lambda^n},$$

$$\text{donde } a_n = n! \sum_{k=0}^n \frac{(-1)^k}{k!}.$$

La expresión anterior fijado n se puede obtener integrando por partes n veces y resolviendo los límites correspondientes. Dejámos al lector los cálculos correspondientes. Sin embargo, la obtención de la fórmula general para n se sale del nivel del curso.

Fijémonos que los momentos centrales de orden n también tienden a infinito a medida que n crece:
 $\lim_{n \rightarrow \infty} \mu_n = \lim_{n \rightarrow \infty} \frac{a_n}{\lambda^n} = \infty$:

$$\lim_{n \rightarrow \infty} \mu_n = \lim_{n \rightarrow \infty} \frac{n! \sum_{k=0}^n \frac{(-1)^k}{k!}}{\lambda^n} = \lim_{n \rightarrow \infty} \sum_{k=0}^n \frac{(-1)^k}{k!} \cdot \lim_{n \rightarrow \infty} \frac{n!}{\lambda^n} = e^{-1} \cdot \infty = \infty.$$

Ejemplo: momento central de orden n de una variable normal de parámetros μ y σ

Recordemos que su función de densidad es: $f_X(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$, para $x \in \mathbb{R}$.

Su momento central de orden 2 será la varianza σ^2 : $\mu_2 = \sigma^2$.

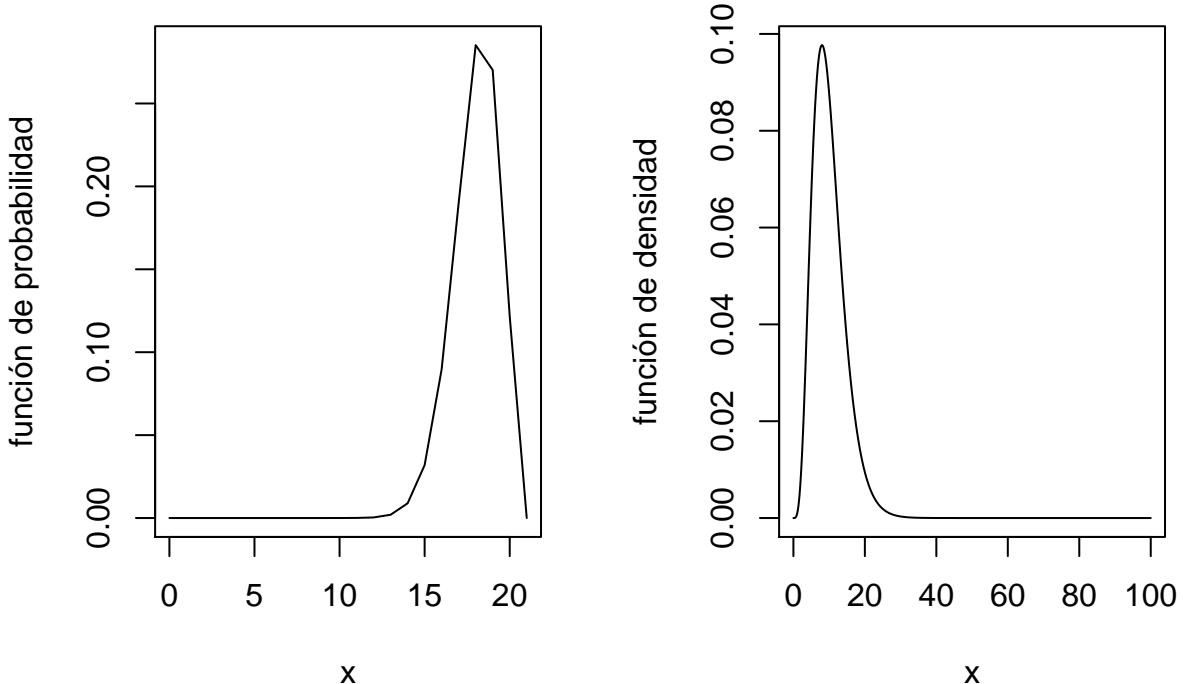
Los momentos centrales de orden impar n serán cero ya que integramos una función impar respecto $x = \mu$: $\mu_n = E((X - \mu)^n) = \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \cdot (x - \mu)^n dx = 0$. O sea, si consideramos $g(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \cdot (x - \mu)^n$, se verifica $g(\mu - x) = -g(\mu + x)$, para todo $x \in \mathbb{R}$.

Si intentamos calcular el momento central de orden 4, obtenemos: $\mu_4 = E((X - \mu)^4) = \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \cdot (x - \mu)^4 dx = 3\sigma^4$. La integral anterior puede resolverse con el cambio de variable $t = \frac{x-\mu}{\sigma}$ y usando que: $\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} \cdot t^4 dt = 3$.

4.2. Asimetría de una variable aleatoria

Una variable aleatoria tiene **asimetría positiva** si su función de densidad o de probabilidad presenta una cola a la **derecha** y **asimetría negativa**, si su función de densidad o de probabilidad presenta una cola a la **izquierda**.

Por ejemplo, en la figura siguiente, vemos la gráfica de la función de probabilidad de una variable aleatoria que presenta **asimetría negativa** a la izquierda y una función de densidad de una variable aleatoria que presenta **asimetría positiva** a la derecha:



¿Cómo calcular la asimetría de una variable aleatoria?

La asimetría de una variable aleatoria X se calcula a partir de sus momentos centrales de segundo y tercer orden:

$$\gamma_1 = E \left(\left(\frac{X - \mu}{\sigma} \right)^3 \right) = \frac{\mu_3}{\sigma^3},$$

donde $\mu = E(X)$ y $\sigma^2 = \text{Var}(X)$.

Dicho valor se denomina **coeficiente de asimetría de Pearson**.

Usando la relación ya vista entre los momentos centrales y los momentos, podemos expresar el **coeficiente de asimetría** en función de los momentos:

$$\gamma_1 = \frac{m_3 - 3\mu\sigma^2 - \mu^3}{\sigma^3}.$$

Dejamos al lector la comprobación de la expresión anterior.

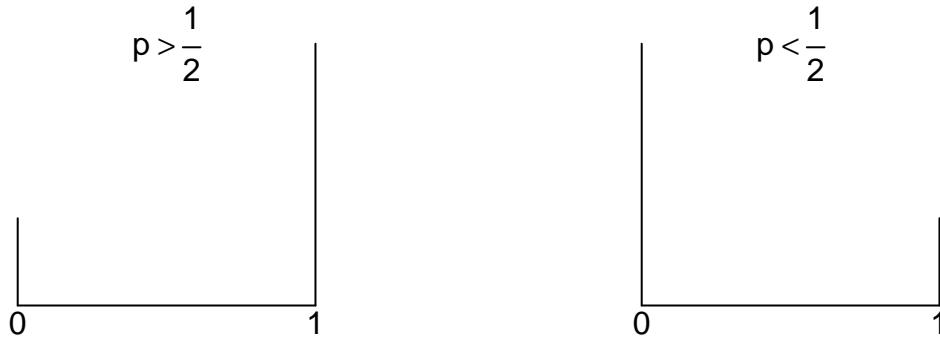
Por tanto, una variable aleatoria X tendrá simetría positiva o a la derecha si $\gamma_1 > 0$ y tendrá asimetría negativa o a la izquierda, si $\gamma_1 < 0$.

Ejemplo: cálculo del coeficiente de asimetría para una variable de Bernoulli de parámetro p

Sea X una variable de Bernoulli de parámetro p . Usando que $m_n = p$, para todo n y que $\mu_2 = \sigma^2 = p - p^2$, el coeficiente de asimetría γ_1 será:

$$\gamma_1 = \frac{p - 3p(p - p^2) - p^3}{\sqrt{(p - p^2)^3}} = \frac{p(1 - p)(1 - 2p)}{\sqrt{(p - p^2)^3}}.$$

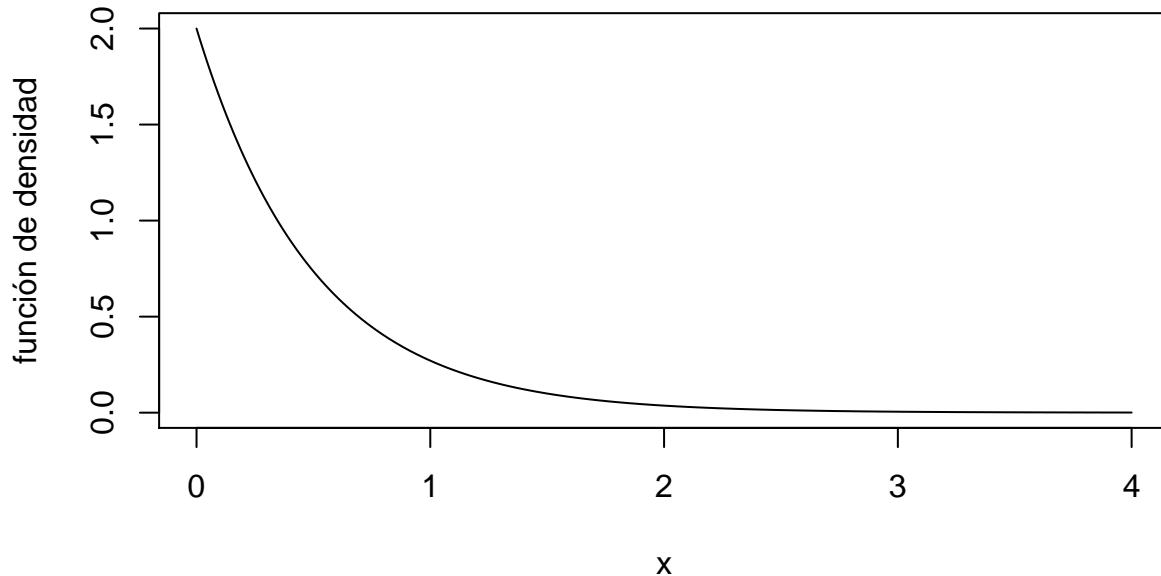
Por tanto, la variable de Bernoulli de parámetro p tendrá simetría negativa si $p > \frac{1}{2}$ y positiva, si $p < \frac{1}{2}$:



Ejemplo: cálculo del coeficiente de asimetría para una variable exponencial de parámetro λ

Sea X una variable exponencial de parámetro λ . Usando que $\sigma^2 = \frac{1}{\lambda^2}$ y $\mu_3 = \frac{a_3}{\lambda^3} = \frac{2}{\lambda^3}$, su coeficiente de asimetría de Pearson será: $\gamma_1 = \frac{\frac{2}{\lambda^3}}{\frac{1}{\lambda^2}} = 2$.

Entonces presenta asimetría positiva o a la derecha tal como se observa en su función de densidad:



Ejemplo: cálculo del coeficiente de asimetría para una variable normal de parámetros μ y σ

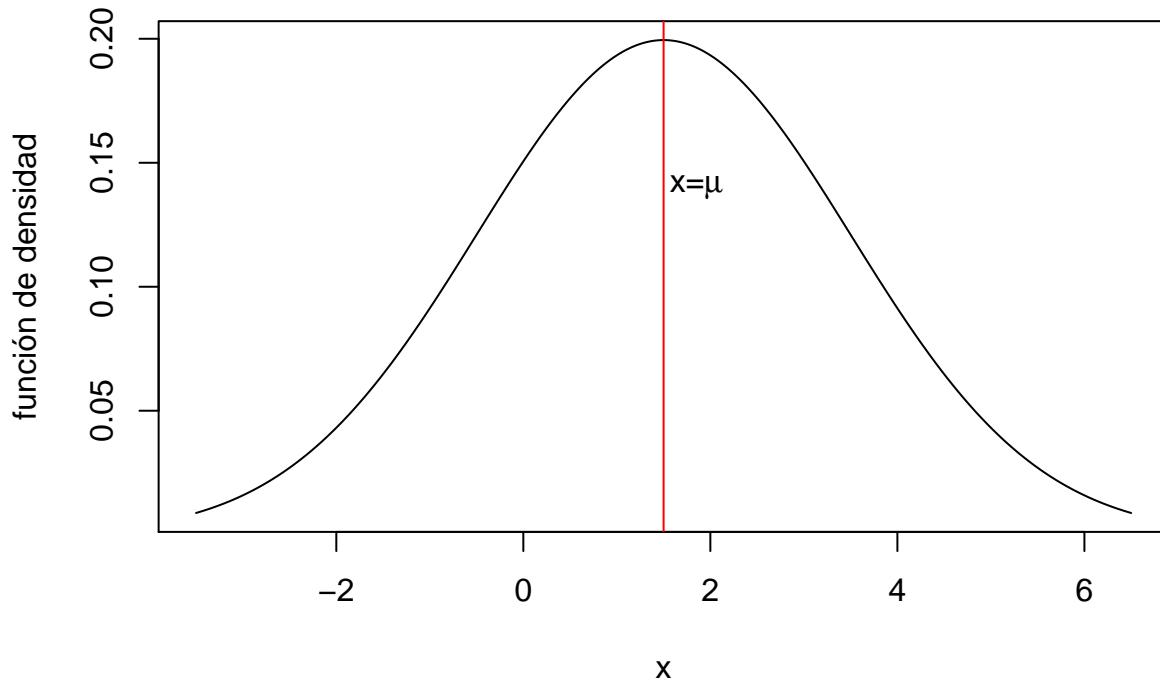
Sea X una variable aleatoria normal de parámetros μ y σ .

Tal como se ha indicado anteriormente, los momentos centrales de orden impar son nulos.

Por tanto, en este caso $\mu_3 = 0$ y, por tanto, $\gamma_1 = 0$.

Deducimos que la distribución normal es totalmente simétrica.

De hecho, usando que su función de densidad es $f_X(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$, para $x \in \mathbb{R}$, se puede comprobar que $f_X(\mu - x) = f_X(\mu + x)$, o sea, tiene el eje de simetría $x = \mu$:



4.3. Curtosis o apuntamiento de una variable aleatoria

La curtosis de una variable aleatoria X es una medida de cómo son las colas de su función de densidad.

Dicho en otras palabras, queremos medir de alguna manera la *tendencia* que tiene la variable aleatoria a tener valores atípicos o *outliers*.

La manera estándar de medir la curtosis de una variable aleatoria X es a partir de su **momento central de cuarto orden**:

$$\gamma_2 = E \left(\left(\frac{X - \mu}{\sigma} \right)^4 \right) = \frac{\mu_4}{\sigma^4},$$

donde recordemos que $\mu = E(X)$ y $\sigma^2 = \text{Var}(X)$.

A la expresión anterior se le denomina **medida de curtosis de Pearson**.

- Diremos que una variable aleatoria no tiene exceso de curtosis o **mesocúrtica** si $\gamma_2 \approx 3$.
- Diremos que una variable aleatoria tiene exceso positivo de curtosis o **leptocúrtica** si $\gamma_2 > 3$.
- Diremos que una variable aleatoria tiene exceso negativo de curtosis o **platicúrtica** si $\gamma_2 < 3$.

Ejemplo: cálculo del coeficiente de curtosis para una variable de Bernoulli de parámetro p

Sea X una variable aleatoria de parámetro p .

El momento central de cuarto orden de X será:

$$\mu_4 = p(1-p)^4 + (1-p)p^4 = p(1-p)(3p^2 - 3p + 1).$$

La medida de curtosis de Pearson será:

$$\gamma_2 = \frac{p(1-p)(3p^2 - 3p + 1)}{p^2(1-p)^2} = \frac{3p^2 - 3p + 1}{p(1-p)}.$$

Se puede comprobar (ejercicio para el lector) que si $p \in \left(\frac{3-\sqrt{3}}{6}, \frac{3+\sqrt{3}}{6}\right) \approx (0,211, 0,789)$, $\gamma_2 < 3$ y, por tanto X será platicúrtica y en caso contrario, si $p \in \left(0, \frac{3-\sqrt{3}}{6}\right) \cup \left(\frac{3+\sqrt{3}}{6}, 1\right)$, $\gamma_2 > 3$ y, por tanto, X será leptocúrtica.

Ejemplo: cálculo del coeficiente de curtosis para una variable exponencial de parámetro λ

Sea X una variable exponencial de parámetro λ . Usando que $\sigma^2 = \frac{1}{\lambda^2}$ y $\mu_4 = \frac{\alpha_4}{\lambda^3} = \frac{9}{\lambda^4}$, su coeficiente de asimetría de Pearson será: $\gamma_2 = \frac{\frac{9}{\lambda^4}}{\frac{1}{\lambda^2}} = 9$.

Como $\gamma_2 > 3$, se trataría de una distribución leptocúrtica.

Ejemplo: cálculo del coeficiente de curtosis para una variable normal de parámetros μ y σ

Sea X una variable aleatoria normal de parámetros μ y σ .

Tal como se ha indicado anteriormente, el momento central de orden 4 vale: $\mu_4 = 3\sigma^4$.

Su coeficiente de curtosis será:

$$\gamma_2 = \frac{\mu_4}{\sigma^4} = \frac{3\sigma^4}{\sigma^4} = 3.$$

Deducimos, por tanto, que toda distribución normal es mesocúrtica o no tiene exceso (ni positivo ni negativo) de curtosis.

4.4. Métodos de transformación

Hemos visto anteriormente que el cálculo de los **momentos** o los **momentos centrados** de una variable aleatoria X puede ser muy complicado y muy tedioso.

Por dicho motivo, vamos a introducir un conjunto de funciones que nos permitirán calcular los **momentos** de la variable X de forma relativamente sencilla.

4.4.1. Función generatriz de momentos

Definición de función generatriz de momentos: Sea X una variable aleatoria X con función de probabilidad P_X en el caso discreto o función de densidad f_X en el caso continuo.

Sea $t \in \mathbb{R}$ un valor real cualquiera.

Definimos la función generatriz de momentos $m_X(t)$ en el valor t como: $m_X(t) = E(e^{tX})$.

Ejemplo: cálculo de la función generatriz de momentos para una variable de Bernoulli de parámetro p

Sea X una variable aleatoria de Bernoulli de parámetro p . Recordemos que su función de probabilidad es:

$$P_X(0) = q = 1 - p, \quad P_X(1) = p.$$

Su función generatriz de momentos será:

$$m_X(t) = E(e^{tX}) = pe^{t \cdot 1} + (1-p)e^{t \cdot 0} = pe^t + (1-p) = 1 + p(e^t - 1).$$

Ejemplo: cálculo de la función generatriz de momentos para una variable exponencial de parámetro λ

Sea X una variable aleatoria exponencial de parámetro λ . Recordemos que su función de densidad es: $f_X(x) = \lambda e^{-\lambda x}$, para $x \geq 0$ y 0, en caso contrario.

Su función generatriz de momentos será:

$$m_X(t) = E(e^{tX}) = \int_0^\infty e^{tx} \lambda e^{-\lambda x} dx = \lambda \int_0^\infty e^{(t-\lambda)x} dx = \lambda \left[\frac{e^{(t-\lambda)x}}{t-\lambda} \right]_{x=0}^{x=\infty} = \frac{\lambda}{\lambda-t}, \text{ si } t < \lambda.$$

En este caso vemos que el dominio de la función generatriz de momentos m_X es $(-\infty, \lambda)$, ya que si $t \geq \lambda$, la integral anterior no es convergente.

Fijémonos por lo que vendrá más adelante que, como $\lambda > 0$, el valor 0 pertenece al dominio de m_X .

Ejemplo: cálculo de la función generatriz de momentos para una variable normal de parámetros μ y σ

Sea X una variable normal de parámetros μ y σ .

Recordemos que su función de densidad es: $f_X(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$, para $x \in \mathbb{R}$.

Su función generatriz de momentos será:

$$\begin{aligned} m_X(t) &= E(e^{tX}) = \int_{-\infty}^{\infty} e^{tx} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx = \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^{\infty} e^{tx - \frac{(x-\mu)^2}{2\sigma^2}} dx \\ &= \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^{\infty} e^{-\frac{1}{2\sigma^2}((x-(\sigma^2t+\mu))^2 - 2\sigma^2t\mu - \sigma^4t^2)} dx = \frac{1}{\sqrt{2\pi}\sigma} e^{\frac{1}{2}(2t\mu + \sigma^2t^2)} \int_{-\infty}^{\infty} e^{-\frac{1}{2\sigma^2}(x-(\sigma^2t+\mu))^2} dx \\ &= e^{\frac{1}{2}(2t\mu + \sigma^2t^2)} \left(\frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^{\infty} e^{-\frac{1}{2\sigma^2}(x-(\sigma^2t+\mu))^2} dx \right) = e^{t\mu + \frac{\sigma^2t^2}{2}}. \end{aligned}$$

La integral del último paréntesis se resuelve haciendo el cambio de variable $u = x - \sigma^2t$ y usando que la integral de la función de densidad de X sobre todo \mathbb{R} vale 1.

Relación entre la función generatriz de momentos y los momentos

La razón del nombre que lleva la **función generatriz de momentos** es que podemos obtener todos los momentos de la variable a partir de ella:

Proposición. Sean X una variable aleatoria con **función generatriz de momentos** $m_X(t)$. Entonces, el momento de orden n de X se puede obtener de la forma siguiente:

$$m_n = E(X^n) = \frac{d}{dt^n} m_X(t)|_{t=0} = m_X^{(n)}(0).$$

O sea, el momento de orden n de X es la derivada n -ésima de la función generatriz de momentos evaluada en $t = 0$.

Demostración

Recordemos la definición de la función generatriz de momentos: $m_X(t) = E(e^{tX})$.

La idea de la demostración es probar por inducción que $m_X^{(n)}(t) = E(e^{tX} \cdot X^n)$.

Veámoslo para $n = 1$: $m'_X(t) = E(e^{tX} \cdot X)$.

Seguidamente, apliquemos inducción sobre n . Supongamos que $m_X^{(n)}(t) = E(e^{tX} \cdot X^n)$ y veamos que $m_X^{(n+1)}(t) = E(e^{tX} \cdot X^{n+1})$: $m_X^{(n+1)}(t) = \frac{d}{dt}(m_X^{(n)}(t)) = \frac{d}{dt}E(e^{tX} \cdot X^n) = E(e^{tX} \cdot X^{n+1})$, tal como queríamos demostrar.

Ahora si aplicamos la expresión demostrada $m_X^{(n)}(t) = E(e^{tX} \cdot X^n)$ a $t = 0$, obtenemos: $m_X^{(n)}(0) = E(X^n) = m_n$, tal como dice la proposición.

Ejemplo: aplicación de la proposición en el caso en que X es una variable de Bernoulli de parámetro p

En este caso, recordemos que: $m_X(t) = 1 + p(e^t - 1)$.

Se puede comprobar que $m_X^{(n)}(t) = pe^t$. Por tanto:

$$m_n = m_X^{(n)}(0) = p,$$

tal como habíamos calculado anteriormente.

Ejemplo: aplicación de la proposición en el caso en que X es una variable exponencial de parámetro λ

En este caso, recordemos que: $m_X(t) = \frac{\lambda}{\lambda-t}$, para $t < \lambda$ pero como $\lambda > 0$, $t = 0$ cumple la expresión anterior.

Dejamos como ejercicio para el lector comprobar que: $m_X^{(n)}(t) = \frac{\lambda n!}{(\lambda-t)^{n+1}}$.

Por tanto:

$$m_n = m_X^{(n)}(0) = \frac{\lambda n!}{\lambda^{n+1}} = \frac{n!}{\lambda^n},$$

expresión que ya habíamos obtenido anteriormente.

Ejemplo: aplicación de la proposición en el caso en que X es una variable normal de parámetros μ y σ

En este caso, recordemos que: $m_X(t) = e^{t\mu + \frac{\sigma^2 t^2}{2}}$.

Aplicando la fórmula de los momentos para $n = 1$ obtenemos: $m'(t) = e^{t\mu + \frac{\sigma^2 t^2}{2}} (\mu + t\sigma^2)$, que en $t = 0$ vale: $m'(0) = \mu = E(X)$, tal como ya sabemos.

Si la aplicamos para $n = 2$, obtenemos: $m''(t) = e^{t\mu + \frac{\sigma^2 t^2}{2}} ((\mu + t\sigma^2)^2 + \sigma^2) = e^{t\mu + \frac{\sigma^2 t^2}{2}} (t^2\sigma^4 + \mu^2 + \sigma^2 + 2t\mu\sigma^2)$, que en $t = 0$ vale: $m''(0) = \mu^2 + \sigma^2 = E(X^2)$, tal como ya sabemos.

Para $n = 3$ obtenemos: $m'''(t) = e^{t\mu + \frac{\sigma^2 t^2}{2}} (\mu + \sigma^2 t) \left((\mu + \sigma^2 t)^2 + 3\sigma^2 \right)$, que en $t = 0$ vale: $m'''(0) = 3\sigma^2\mu = E(X^3)$, valor que correspondería al momento de tercer orden de X .

Por último, para $n = 4$, obtenemos: $m^{(iv)}(t) = e^{\mu t + \frac{\sigma^2 t^2}{2}} \left(6\sigma^2 (\mu + \sigma^2 t)^2 + (\mu + \sigma^2 t)^4 + 3\sigma^4 \right)$, que en $t = 0$ vale: $m^{(iv)}(0) = 6\sigma^2 \mu^2 + \mu^4 + 3\sigma^4 = E(X^4)$, valor que correspondería al momento de cuarto orden de X .

4.4.2. Función característica

Definición de función característica: Sea X una variable aleatoria X con función de probabilidad P_X en el caso discreto o función de densidad f_X en el caso continuo.

Sea $w \in \mathbb{R}$ un valor real cualquiera.

Definimos la función característica $\phi_X(w)$ en el valor w como: $\phi_X(w) = E(e^{iwX})$, donde i es el número complejo $i = \sqrt{-1}$.

Observación: Si X es una variable continua, la **función característica** $\phi_X(w)$ puede interpretarse como la **transformada de Fourier** de la **función de densidad** de X : $\phi(w) = \int_{-\infty}^{\infty} f_X(x)e^{ixw} dx$.

Por tanto, usando la fórmula de la **antitransformada de Fourier**, podemos escribir la **función de densidad** $f_X(x)$ como función de la **función característica** de X , $\phi(w)$: $f_X(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \phi_X(w)e^{-ixw} dw$.

Observación: En el caso discreto, o sea, Si X es una variable discreta, la **función característica** $\phi_X(w)$ se escribe como función de la **función de probabilidad** $P_X(x_k)$ con **Dominio** $D_X = \{x_k, k\}$ como: $\phi(w) = \sum_k P_X(x_k)e^{ixw_k}$.

En los casos en que los x_k sean enteros, $x_k = k$, que son la mayoría, la ecuación anterior es la **transformada de Fourier de la secuencia** $P_X(k)$. Dicha función es una *función periódica* en w de periodo 2π ya que $e^{i(w+2\pi)k} = e^{iwk}$.

Por tanto, usando la fórmula de **inversión**, podemos escribir la **función de probabilidad** $P_X(k)$ como función de la función característica de X , $\phi(w)$: $P_X(k) = \frac{1}{2\pi} \int_0^{2\pi} \phi_X(w)e^{-ikw} dw$.

Ejemplo: cálculo de la función característica para una variable de Bernoulli de parámetro p

Sea X una variable aleatoria de Bernoulli de parámetro p . Recordemos que su función de probabilidad es:

$$P_X(0) = q = 1 - p, \quad P_X(1) = p.$$

Su función característica será:

$$\phi_X(w) = E(e^{iwX}) = pe^{iw \cdot 1} + (1 - p)e^{iw \cdot 0} = pe^{iw} + (1 - p) = 1 + p(e^{iw} - 1).$$

Comprobemos la fórmula de la inversión:

$$\begin{aligned} P_X(1) &= \frac{1}{2\pi} \int_0^{2\pi} (1 + p(e^{iw} - 1)) e^{-iw \cdot 1} dw = \frac{1}{2\pi} \left(\int_0^{2\pi} (1 - p)e^{-iw} dw + \int_0^{2\pi} p dw \right) \\ &= \frac{1}{2\pi} \left((1 - p) \left[\frac{e^{-iw}}{-i} \right]_0^{2\pi} + 2\pi p \right) = \frac{1}{2\pi} ((1 - p) \cdot 0 + 2\pi p) = p, \\ P_X(0) &= \frac{1}{2\pi} \int_0^{2\pi} (1 + p(e^{iw} - 1)) e^{-iw \cdot 0} dw = \frac{1}{2\pi} \left(\int_0^{2\pi} (1 - p) dw + \int_0^{2\pi} p e^{iw} dw \right) \\ &= \frac{1}{2\pi} \left((1 - p) \cdot 2\pi + p \left[\frac{e^{iw}}{i} \right]_0^{2\pi} \right) = \frac{1}{2\pi} ((1 - p) \cdot 2\pi + p \cdot 0) = 1 - p. \end{aligned}$$

Ejemplo: cálculo de la función característica para una variable exponencial de parámetro λ

Sea X una variable aleatoria exponencial de parámetro λ . Recordemos que su función de densidad es: $f_X(x) = \lambda e^{-\lambda x}$, para $x \geq 0$ y 0, en caso contrario.

Su función característica será:

$$\phi_X(w) = E(e^{iwX}) = \int_0^\infty e^{iwx} \lambda e^{-\lambda x} dx = \lambda \int_0^\infty e^{(i w - \lambda)x} dx = \lambda \left[\frac{e^{(i w - \lambda)x}}{i w - \lambda} \right]_{x=0}^{x=\infty} = \frac{\lambda}{\lambda - i w}.$$

La expresión anterior es válida para todo $w \in \mathbb{R}$ ya que su valor sería: $\phi_X(w) = \frac{\lambda}{\lambda - i w} \cdot \frac{\lambda + i w}{\lambda + i w} = \frac{\lambda^2 + i \lambda w}{\lambda^2 + w^2} = \frac{\lambda^2}{\lambda^2 + w^2} + i \frac{\lambda w}{\lambda^2 + w^2}$. En la última expresión hemos separado la parte real de la imaginaria.

Calculemos la función de densidad a partir de la función característica:

$$f_X(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \frac{\lambda}{\lambda - i w} e^{-i w x} dw = a e^{-a x},$$

si $x > 0$ y 0 en caso contrario. El cálculo de la integral anterior debe realizarse usando el *Teorema de los Residuos*, Residue theorem y se sale de los objetivos de este curso.

Ejemplo: cálculo de la función característica para una variable normal de parámetros μ y σ

Sea X una variable normal de parámetros μ y σ .

Recordemos que su función de densidad es: $f_X(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$, para $x \in \mathbb{R}$.

Su función característica será:

$$\begin{aligned} \phi_X(w) &= E(e^{iwX}) = \int_{-\infty}^{\infty} e^{iwx} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx = \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^{\infty} e^{iwx - \frac{(x-\mu)^2}{2\sigma^2}} dx \\ &= \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^{\infty} e^{-\frac{1}{2\sigma^2}((x-(\sigma^2 iw + \mu))^2 - 2\sigma^2 iw\mu + \sigma^4 w^2)} dx = \frac{1}{\sqrt{2\pi}\sigma} e^{\frac{1}{2}(2iw\mu - \sigma^2 w^2)} \int_{-\infty}^{\infty} e^{-\frac{1}{2\sigma^2}(x-(\sigma^2 iw + \mu))^2} dx \\ &= e^{\frac{1}{2}(2iw\mu - \sigma^2 w^2)} \left(\frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^{\infty} e^{-\frac{1}{2\sigma^2}(x-(\sigma^2 iw + \mu))^2} dx \right) = e^{iw\mu - \frac{\sigma^2 w^2}{2}}. \end{aligned}$$

La integral del último paréntesis se resuelve haciendo el cambio de variable $u = x - \sigma^2 iw$ y usando que la integral de la función de densidad de X sobre todo \mathbb{R} vale 1.

Calculemos la función de densidad a partir de la función característica:

$$\begin{aligned} f_X(x) &= \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{iwx - \frac{\sigma^2 w^2}{2}} e^{-iwx} dw = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{\left(\frac{iw\sigma}{\sqrt{2}} + \frac{\mu-x}{\sigma\sqrt{2}}\right)^2 - \frac{(\mu-x)^2}{2\sigma^2}} dw = \frac{1}{2\pi} e^{-\frac{(\mu-x)^2}{2\sigma^2}} \int_{-\infty}^{\infty} e^{\left(\frac{iw\sigma}{\sqrt{2}} + \frac{\mu-x}{\sigma\sqrt{2}}\right)^2} dw \\ &= \frac{1}{2\pi} e^{-\frac{(\mu-x)^2}{2\sigma^2}} \int_{-\infty}^{\infty} e^{-\left(\frac{w\sigma}{\sqrt{2}} + \frac{\mu-x}{i\sigma\sqrt{2}}\right)^2} dw \quad \text{cambio de variable } u = \frac{w\sigma}{\sqrt{2}} + \frac{\mu-x}{i\sigma\sqrt{2}}, \quad \frac{1}{2\pi} e^{-\frac{(\mu-x)^2}{2\sigma^2}} \int_{-\infty}^{\infty} \frac{\sqrt{2}}{\sigma} e^{-u^2} du \\ &\stackrel{\int_{-\infty}^{\infty} e^{-u^2} du = \sqrt{\pi}}{=} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(\mu-x)^2}{2\sigma^2}} \sqrt{\pi} = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}, \end{aligned}$$

función que coincide con la densidad de la distribución $N(\mu, \sigma)$.

Relación entre la función característica y los momentos

La relación entre la **función característica** y los **momentos** es la siguiente:

Proposición. Sean X una variable aleatoria con **función característica** $\phi_X(w)$. Entonces, el momento de orden n de X se puede obtener de la forma siguiente:

$$m_n = E(X^n) = \frac{1}{i^n} \frac{d}{dw^n} \phi_X(w)|_{w=0} = \frac{1}{i^n} \phi_X^{(n)}(0).$$

O sea, el momento de orden n de X es la derivada n -ésima de la función característica evaluada en $w = 0$ dividido por i^n .

Ejercicio

La demostración se realiza de forma similar a la demostración de la proposición que relaciona la función generatriz de momentos y los momentos.

Se deja como ejercicio al lector.

Ejercicio

Realizar los mismos ejemplos que los realizados para la función generatriz de momentos. O sea:

- Si X es una variable de Bernoulli de parámetro p , demostrar usando la función característica que para todo n , $m_n = E(X^n) = p$.
- Si X es una variable exponencial de parámetro λ , demostrar usando la función característica que para todo n , $m_n = E(X^n) = \frac{n!}{\lambda^n}$.
- Si X es una variable normal de parámetros μ y σ , demostrar usando la función característica que $E(X) = \mu$, $E(X^2) = \mu^2 + \sigma^2$, $E(X^3) = 3\sigma^2\mu$ y $E(X^4) = 6\sigma^2\mu^2 + \mu^4 + 3\sigma^4$.

4.5. Fiabilidad

Sea $T \geq 0$ una variable aleatoria que nos da, por ejemplo, el tiempo de vida de cierto componente o dispositivo.

Vamos a definir medidas para estudiar la fiabilidad de este tipo de variables aleatorias.

Definición: Sea $T \geq 0$ una variable aleatoria. La **fiabilidad** de T en el tiempo t se define como la probabilidad que el sistema, componente o dispositivo funcione en el tiempo t : $R(t) = P(T > t)$.

Observación: Dada una variable $T \geq 0$, la relación existente entre la **fiabilidad** R y la **función de distribución** F_T es la siguiente:

$$R(t) = P(T > t) = 1 - P(T \leq t) = 1 - F_T(t)$$

4.5.1. Tiempo medio de vida

Observación: Dada una variable $T \geq 0$ continua, el **tiempo medio de vida** de la variable T sería $E(T)$. Entonces, este **tiempo medio de vida** se puede calcular como: $E(T) = \int_0^\infty R(t) dt$.

Veámoslo. Para ello basta ver que $E(T) = \int_0^\infty (1 - F_T(t)) dt$, donde $F_T(t)$ es la función de distribución de la variable T :

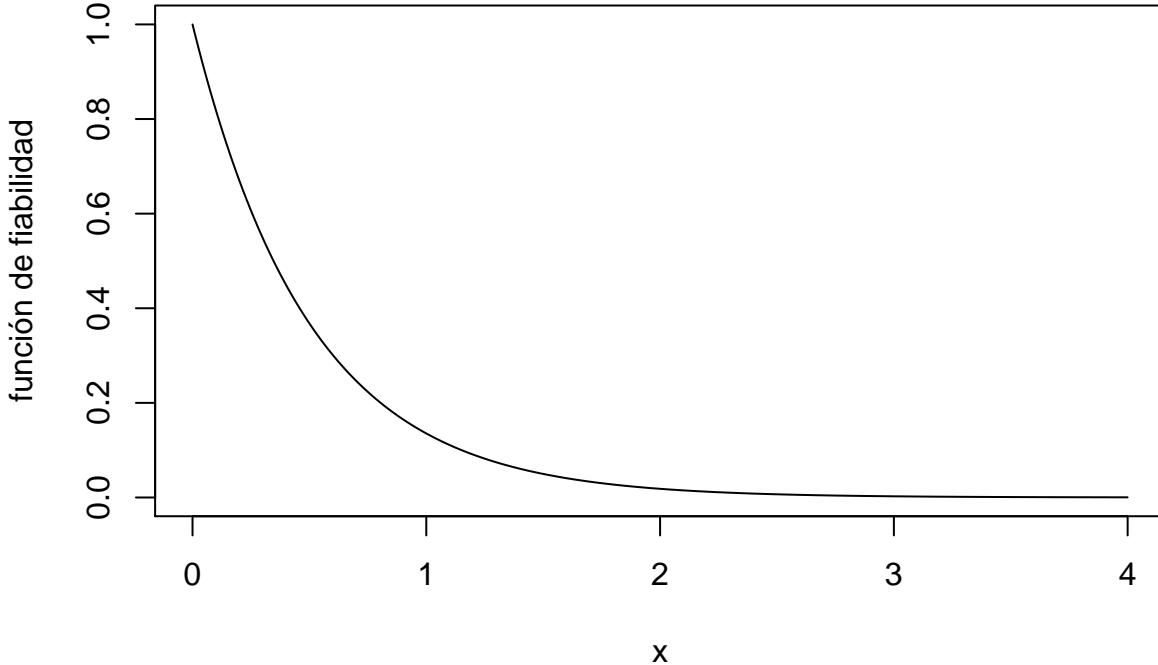
$$\begin{aligned} E(T) &= \int_{t=0}^{t=\infty} 1 - F_T(t) dt = \int_{t=0}^{t=\infty} \int_{u=t}^{u=\infty} f_T(u) du dt \\ &= \int_{u=0}^{u=\infty} f_T(u) \int_{t=0}^{t=u} dt du = \int_{u=0}^{u=\infty} f_T(u) \cdot u du = E(T), \end{aligned}$$

donde $f_T(u)$ sería la función de densidad de la variable T en el valor u .

Ejemplo

Sea T una variable aleatoria exponencial de parámetro λ .

La fiabilidad de T sería: $R(t) = P(T > t) = 1 - F_T(t) = e^{-\lambda t}$:



4.6. Generación de muestras de variables aleatorias por ordenador

La simulación por **computadora** de cualquier fenómeno aleatorio implica la **generación de variables aleatorias** con distribuciones prefijadas de antemano.

Por ejemplo, la simulación de un sistema de colas implica generar el tiempo entre las llegadas de los clientes, así como los tiempos de servicio de cada cliente.

Fijémonos que fijar la variable aleatoria X es equivalente a fijar la **función de distribución** $F_X(x)$ o la **función de densidad** $f_X(x)$ en el caso continuo o la **función de probabilidad** $P_X(x)$ en el caso discreto.

Todos los métodos que vamos a describir presuponen que podemos generar **números aleatorios** que se distribuyen **uniformemente** entre 0 y 1. En R se puede hacer usando la función `runif(n)`, donde `n` es la cantidad de números aleatorios entre 0 y 1 a generar.

4.6.1. Método de transformación

El **método de transformación** se basa en el resultado siguiente:

Proposición. Sea X una variable aleatoria con función de distribución $F_X(x)$. Supongamos que $F_X(x)$ es estrictamente creciente o que existe $F_X^{-1}(y)$, para todo $y \in [0, 1]$. Sea Y la variable aleatoria definida como: $Y = F_X(X)$. Entonces la distribución de Y es uniforme en el intervalo $[0, 1]$.

Demostración:

Claramente, por propia definición de Y , tenemos que el dominio de Y es $[0, 1]$ ya que el conjunto recorrido de la función de distribución de cualquier variable es el intervalo $[0, 1]$.

Para ver que la distribución de Y es $U[0, 1]$ basta comprobar que $F_Y(y) = y$, para todo $y \in [0, 1]$:

$$\begin{aligned} F_Y(y) &= P(Y \leq y) = P(F_X(X) \leq y) && \text{usando que } F_X \text{ es estrictamente creciente} \\ &= F_X(F_X^{-1}(y)) = y. \end{aligned}$$

Usando la proposición anterior, dada una variable X , como la distribución de la variable aleatoria $Y = F_X(X)$ es $U[0, 1]$, si hacemos $X = F_X^{-1}(Y)$, tendremos que si sabemos generar una muestra de Y , aplicándole a la muestra la función F_X^{-1} tendremos generada una muestra de X .

Ejemplo: generar una muestra de una variable exponencial de parámetro λ

Recordemos que si X es exponencial de parámetro λ , su función de distribución es: $F_X(x) = 1 - e^{-\lambda x}$.

Hallaremos a continuación F_X^{-1} :

$$y = 1 - e^{-\lambda x}, \Leftrightarrow 1 - y = e^{-\lambda x}, \Leftrightarrow \ln(1 - y) = -\lambda x, \Leftrightarrow x = -\frac{1}{\lambda} \ln(1 - y).$$

Por tanto, $F_X^{-1}(y) = -\frac{1}{\lambda} \ln(1 - y)$.

Generemos una muestra con R de 25 valores de una variable exponencial de parámetro $\lambda = 2$ usando el método anterior:

```
n=25
lambda=2
muestra.y = runif(n)
muestra.x = -(1/lambda)*log(1-muestra.y)
muestra.x

## [1] 0.35877893 0.25280668 0.11355513 0.71885498 1.19013394 0.49444154
## [7] 0.23443983 0.07049329 1.48215515 0.09206527 0.14783308 2.35604428
## [13] 0.41728297 0.43470135 0.02910931 0.36337987 0.08532398 0.27981457
## [19] 0.40904272 0.02482599 0.25061234 1.67562739 0.07212948 0.27227039
## [25] 0.23932819
```

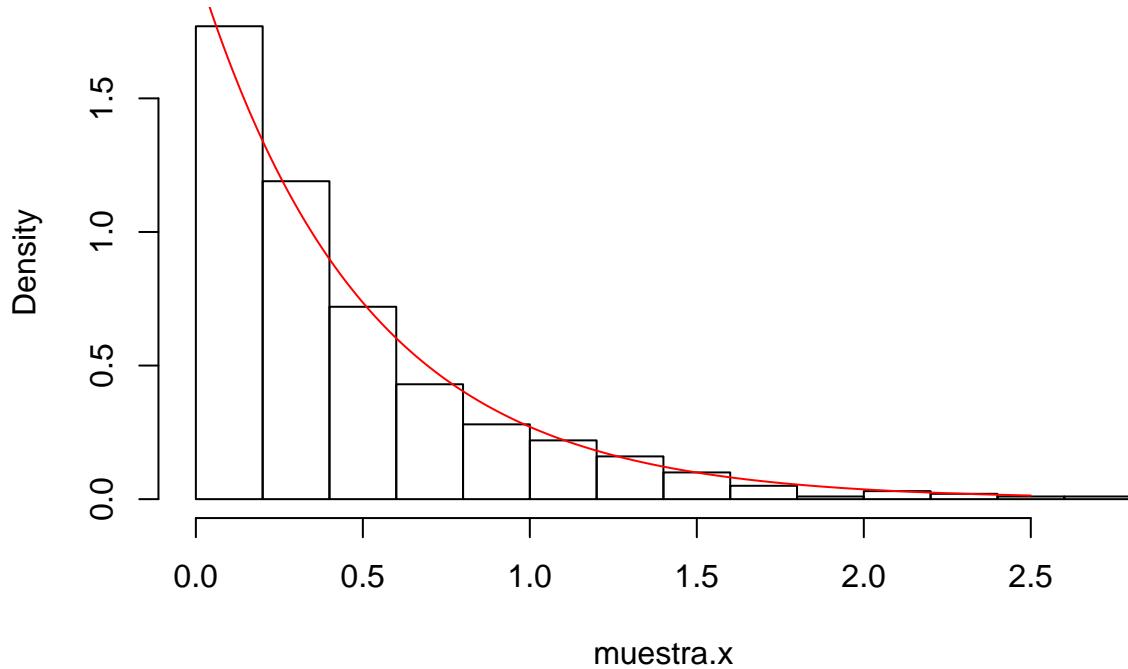
Vamos a testear si nuestro método funciona.

Para ello generaremos una muestra de 500 valores usando el método de transformación y dibujaremos su **histograma de frecuencias relativas**.

Seguidamente dibujaremos la función de densidad de la variable exponencial de parámetro λ y compararemos los resultados:

```
n=500
lambda=2
muestra.y = runif(n)
muestra.x = -(1/lambda)*log(1-muestra.y)
hist(muestra.x,freq=FALSE,main="Histograma de la muestra")
x2=seq(from=0,to=2.5,by=0.01)
lines(x2,dexp(x2,lambda),col="red")
```

Histograma de la muestra



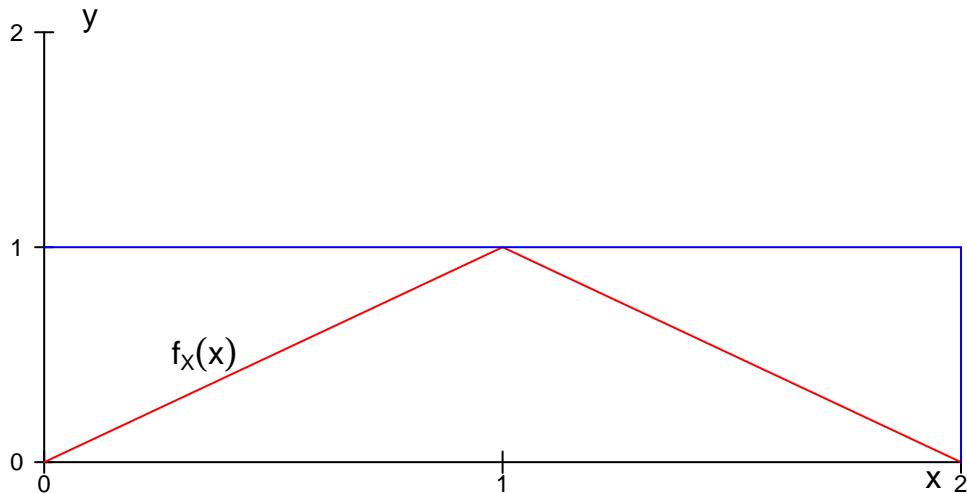
4.6.2. Método de rechazo

Sea X una variable aleatoria continua tal que su función de densidad verifica:

- Existen valores a y b tal que $f_X(x) = 0$ si $x \notin [a, b]$.
- Existen valores c y d tal que $f_X(x) \in [c, d]$, si $x \in [a, b]$.

En resumen, los puntos $(x, f(x))$ pertenecen al rectángulo $[a, b] \times [c, d]$ y en caso contrario $f_X(x) = 0$.

En el gráfico siguiente, $a = 0$, $b = 2$, $c = 0$ y $d = 1$.



Para generar una **muestra aleatoria** de la variable X , hacemos lo siguiente:

- 1) generamos un valor aleatorio x entre a y b .
- 2) generamos un valor aleatorio y entre c y d .
- 3) si $y \leq f_X(x)$, aceptamos x como valor de la muestra. En caso contrario, volvemos a empezar en 1.

Ejemplo

El gráfico de la figura anterior corresponde a la función de densidad siguiente:

$$f_X(x) = \begin{cases} x, & \text{si } 0 \leq x \leq 1, \\ 2 - x, & \text{si } 1 \leq x \leq 2, \\ 0, & \text{en caso contrario.} \end{cases}$$

Vamos a generar una muestra de 25 valores usando el **método del rechazo**:

```
a=0; b=2; c=0; d=1; n=25; i=1;
f = function(x){ifelse(x>=0 & x<=1,x,ifelse(x>=1&x<=2,2-x,0))}
muestra=c()
while(i <=n){
  x=runif(1,a,b)
  y=runif(1,c,d)
  if(y <= f(x)){muestra=c(muestra,x); i=i+1}
}
muestra

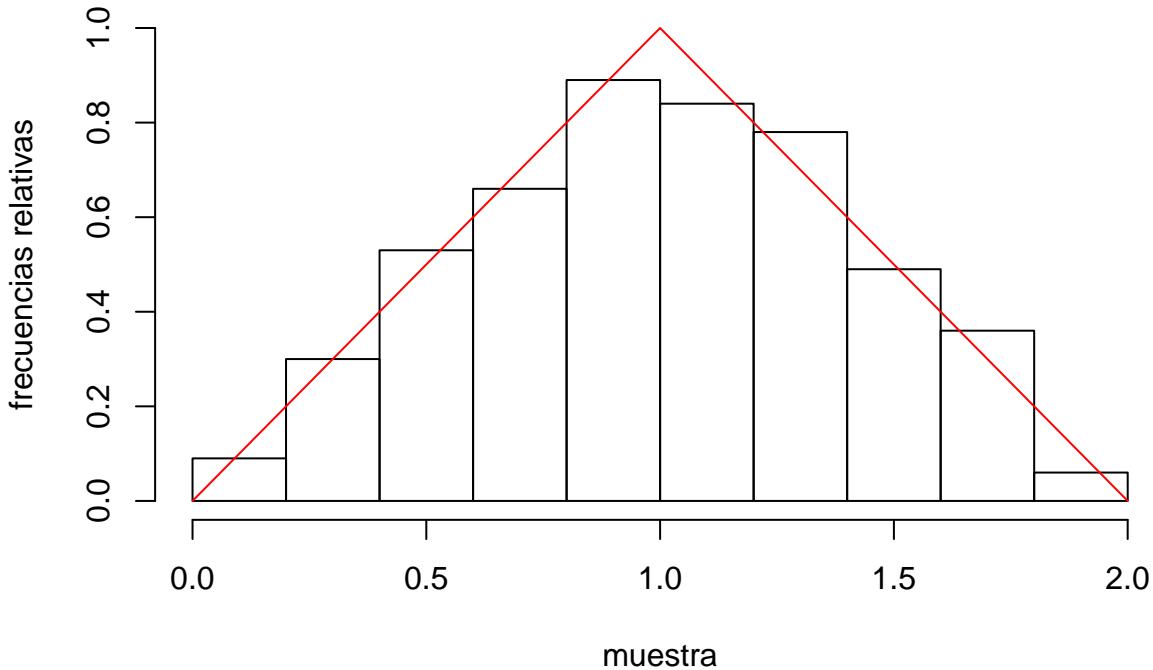
## [1] 0.8934245 1.0241151 0.7485943 1.1318743 1.1174486 0.8956145 1.6270601
## [8] 0.6737312 0.9584882 0.5762514 0.7390811 0.4276671 1.5210227 0.8506008
## [15] 0.4126257 0.9897587 1.2166050 1.6098125 0.8727734 0.8756381 1.4853939
## [22] 1.1539542 1.2353779 0.5157756 1.2873756
```

Como hicimos con el ejemplo del **método de transformación**, vamos a generar una muestra de 500 valores de la variable X , vamos a dibujar el **histograma de frecuencias relativas** junto con la

función de densidad para ver si ésta se aproxima a dicho histograma:

```
a=0; b=2; c=0; d=1; n=500; i=1;
f = function(x){ifelse(x>=0 & x<=1,x,ifelse(x>=1&x<=2,2-x,0))}
muestra=c()
while(i <=n){
  x=runif(1,a,b)
  y=runif(1,c,d)
  if(y <= f(x)){muestra=c(muestra,x); i=i+1}
}
hist(muestra,freq=FALSE,main="Histograma de la muestra")
x2=seq(from=0,to=2,by=0.01)
lines(x2,f(x2),col="red")
```

Histograma de la muestra



4.7. Entropía

La **entropía** es una medida de la **incertidumbre** en un experimento aleatorio.

Veremos cómo la **entropía** cuantifica la **incertidumbre** por la cantidad de **información** requerida para especificar el resultado de un experimento aleatorio.

4.7.1. Entropía de una variable aleatoria

Supongamos que tenemos una variable aleatoria X discreta con valores enteros: $D_X = \{1, 2, \dots, N\}$.

Sea $k \in D_X$ un valor de la variable. Estamos interesados en cuantificar la **incertidumbre** del suceso $A_k = \{X = k\}$.

O sea, cuánta **menos incertidumbre** tenga A_k , más **alta será su probabilidad**, y cuánta **más incertidumbre, menos probabilidad** de aparecer A_k .

Una medida que cumple las condiciones anteriores es la siguiente: $I(A_k) = I(\{X = k\}) = \ln\left(\frac{1}{P(X=k)}\right) = -\ln(P(X=k))$.

Por ejemplo, si $P(A_k) = 1$, o sea, A_k aparece “**seguro**”, entonces tiene incertidumbre **nula**, $I(A_k) = 0$, y si $P(A_k) = 0$, o sea, A_k no aparece “**nunca**”, tiene incertidumbre **máxima**, $I(A_k) = \infty$.

La motivación anterior hace que definamos la **entropía** de una variable aleatoria de la forma siguiente:

Definición: Sea X una variable aleatoria con función de densidad $f_X(x)$ en el caso continuo o función de probabilidad $P_X(x)$ en el caso discreto. Definimos **entropía de X** como: $H_X = E(-\ln(f_X)) = \int_{-\infty}^{\infty} -\ln(f_X(x))f_X(x) dx$, en el caso continuo y, $H_X = E(-\ln(P_X)) = \sum_{x_k \in D_X} -\ln(P_X(x_k))P_X(x_k)$, en el caso discreto.

Ejemplo: entropía de una variable de Bernoulli de parámetro p

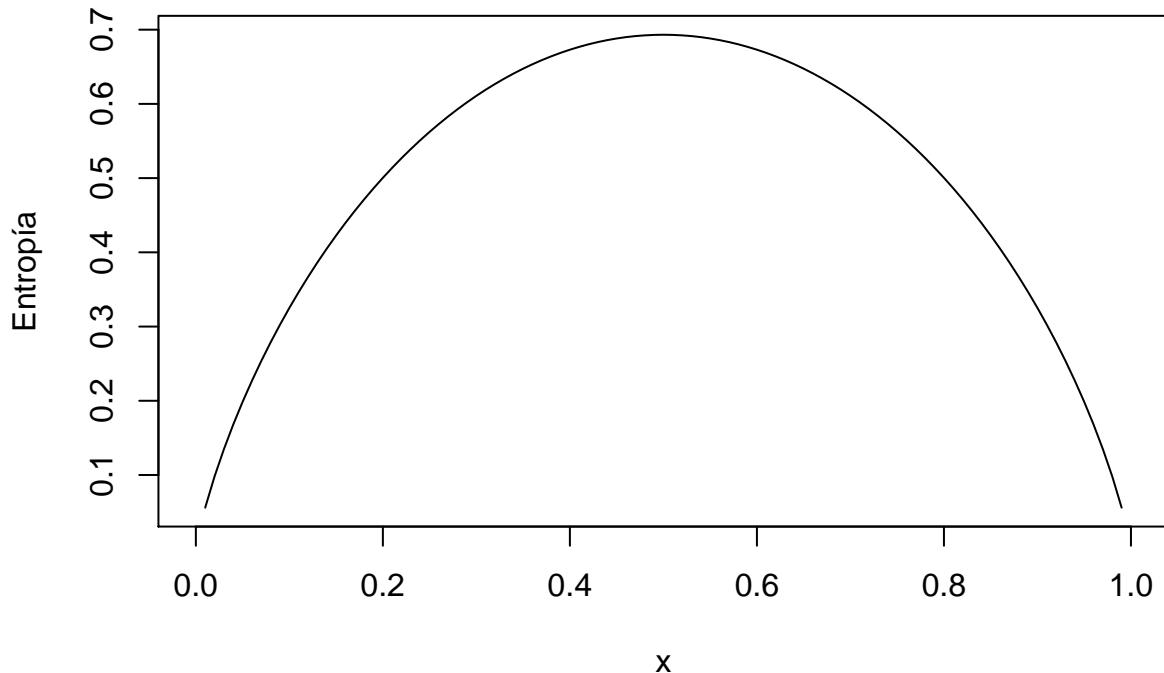
Sea X una variable de Bernoulli de parámetro p .

Recordemos que su función de probabilidad P_X es: $P_X(0) = 1 - p = q$, $P_X(1) = p$.

La entropía de X será:

$$H_X = E(-\ln(P_X)) = -(1-p) \cdot \ln(1-p) - p \cdot \ln p.$$

El gráfico de la entropía se puede observar en el gráfico siguiente donde X tiene entropía máxima cuando $p = \frac{1}{2}$ que sería cuando X tiene incertidumbre máxima al tratar de adivinar el resultado de X y X tiene entropía mínima cuando $p = 0$ o $p = 1$ ya que en estos casos el resultado de X sería siempre 0 o 1, respectivamente.



Ejemplo: Entropía de una variable aleatoria exponencial de parámetro λ

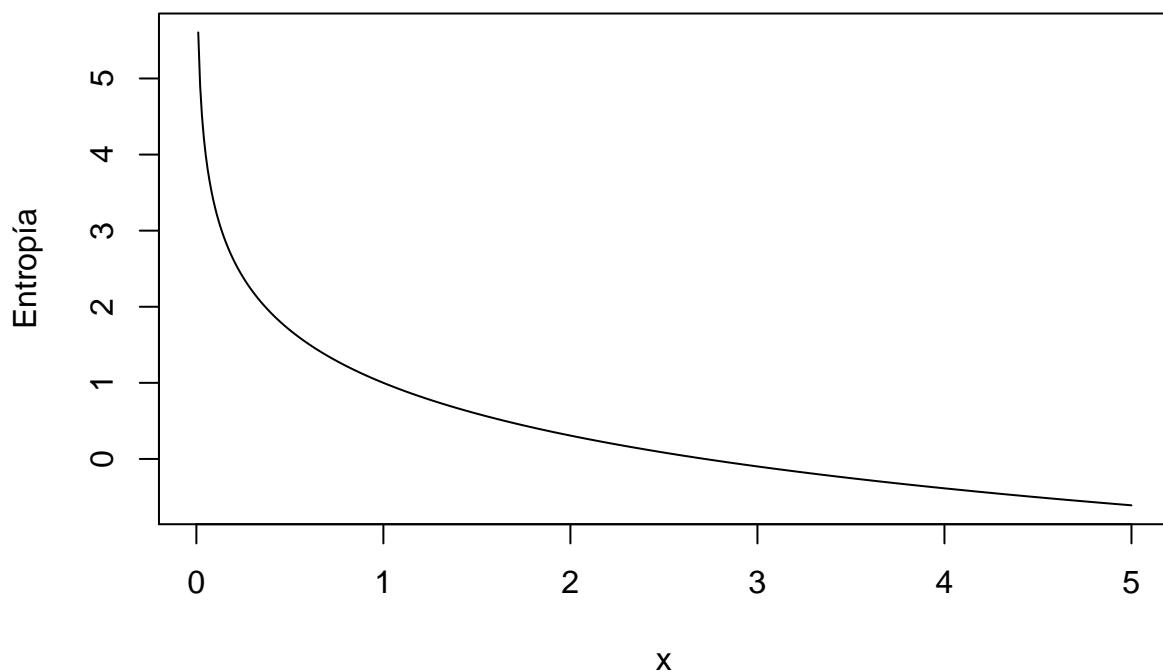
Sea X una variable aleatoria exponencial de parámetro λ .

Recordemos que su función de densidad es: $f_X(x) = \lambda e^{-\lambda x}$, si $x \geq 0$ y $f_X(x) = 0$, en caso contrario.

Su entropía será:

$$\begin{aligned}
 H_X &= E(-\ln(f_X)) = - \int_0^\infty \ln(\lambda e^{-\lambda x}) \lambda e^{-\lambda x} dx = -\lambda \int_0^\infty (\ln(\lambda) - \lambda x) e^{-\lambda x} dx \\
 &= -\ln(\lambda) \int_0^\infty \lambda e^{-\lambda x} dx + \lambda \int_0^\infty \lambda x e^{-\lambda x} dx = -\ln(\lambda) \int_0^\infty f_X(x) dx + \lambda E(X) \\
 &= -\ln(\lambda) + \lambda \frac{1}{\lambda} = 1 - \ln(\lambda).
 \end{aligned}$$

El gráfico de la entropía se puede observar en el gráfico siguiente donde X tiene entropía máxima cuando $\lambda = 0$ que sería cuando X tiene incertidumbre máxima al tratar de adivinar el resultado de X al tener media $E(X) = \frac{1}{\lambda} = \infty$ y X tiene entropía mínima cuando λ tiende a ∞ ya que su media $E(X) = \frac{1}{\lambda}$ tendería a 0.



Capítulo 5

Vectores aleatorios bidimensionales

5.1. Dos variables aleatorias

Muchos experimentos aleatorios involucran varias variables aleatorias.

Por ejemplo, dado un individuo de 30 años escogido al azar de una cierta población, medir su altura y su peso conjuntamente.

Otro ejemplo más complejo es la medición continuada de un *fenómeno aleatorio* que se repite en el tiempo, como sería medir la temperatura media un día determinado del año, por ejemplo el día 1 de enero en un cierto lugar.

La variable aleatoria que nos da la medición en 10 años es una variable aleatoria de varias variables que involucra 10 variables aleatorias supuestas independientes e idénticamente distribuidas, lo que en **estadística inferencial** se le llama una **muestra aleatoria simple**.

5.1.1. Definición

Recordemos que una **variable aleatoria** X es una aplicación que toma valores numéricos para cada resultado de un experimento aleatorio:

$$\begin{aligned} X : \Omega &\longrightarrow \mathbb{R} \\ w &\longrightarrow X(w). \end{aligned}$$

A partir de la definición anterior, generalizamos la noción de **variable aleatoria unidimensional** a **variable aleatoria bidimensional**:

Definición de variable aleatoria bidimensional: Dado un experimento aleatorio con **espacio muestral** Ω , definimos **variable aleatoria bidimensional** (X, Y) a toda aplicación

$$\begin{aligned} (X, Y) : \Omega &\longrightarrow \mathbb{R}^2 \\ w &\longrightarrow (X(w), Y(w)). \end{aligned}$$

Ejemplo: lanzamiento dos dados

Consideremos el experimento aleatorio de lanzar un dado no trucado dos veces.

Sea S la suma de los resultados obtenidos y P el producto de los mismos.

La variable aleatoria (S, P) que asigna a cada resultado $w = (x_1, x_2)$ donde x_1 es el resultado obtenido en el primer lanzamiento y x_2 , el resultado obtenido en el segundo, los valores: $S(w) = x_1 + x_2$ y $P(w) = x_1 \cdot x_2$ es una variable aleatoria bidimensional.

El suceso $\{2 \leq S \leq 4, 3 \leq P \leq 6\}$ seria:

$$\{2 \leq S \leq 4, 3 \leq P \leq 6\} = \{(1, 3), (3, 1), (2, 2)\}.$$

Ejemplo

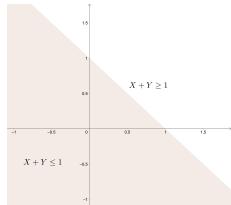
Consideremos el experimento aleatorio de elegir al azar un estudiante de primer curso de grado. Sea w el estudiante elegido. Consideremos la variable aleatoria (H, W) que asigna a dicho estudiante w , $H(w)$: la altura de dicho estudiante en cm. y $W(w)$: el peso de dicho estudiante en kg.

Estamos interesado en sucesos del tipo $A = \{H \leq 176, W \leq 85\}$, es decir, el conjunto de estudiantes que miden menos de 1.76 m. y que pesan menos de 85 kg.

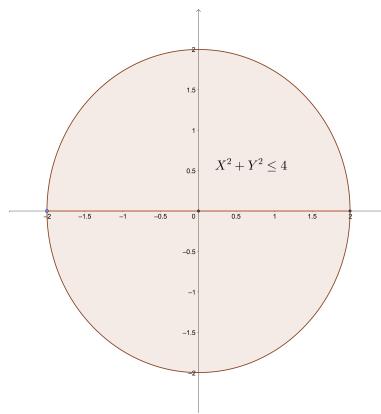
5.1.2. Representación del dominio de una variable aleatoria bidimensional

Los sucesos que se derivan de una **variable aleatoria bidimensional** estan especificados por regiones del plano. Veamos algunos ejemplos:

Suceso: $\{X + Y \leq 1\}$. Es la zona sombreada del gráfico siguiente:



Suceso: $\{X^2 + Y^2 \leq 4\}$. Es la zona sombreada del gráfico siguiente:



Suceso: $\{\max\{X, Y\} \geq 1\}$. Esta zona es la sombreada del gráfico siguiente:

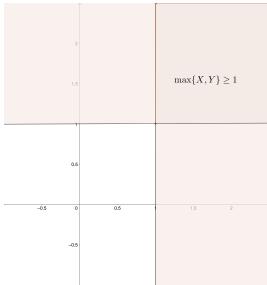


Figura 5.1: Representación suceso

La probabilidad de que la **variable bidimensional** pertenezca a una cierta **región del plano B** se define de la forma siguiente:

$$P((X, Y) \in B) = P\{w \in \Omega, | (X(w), Y(w)) \in B\},$$

es decir, la probabilidad anterior es la probabilidad del suceso formado por los elementos de $w \in \Omega$ que cumplen que su **imagen** por la **variable aleatoria bidimensional** (X, Y) esté en B .

Por ejemplo, si consideramos $B = \{X + Y \leq 1\}$, $P((X, Y) \in B)$ es la probabilidad del suceso formado por los elementos w de Ω tal que la suma de las imágenes por X e Y sea menor o igual que 1: $X(w) + Y(w) \leq 1$.

5.2. Función de distribución conjunta

5.2.1. Definición

Dada una **variable aleatoria bidimensional** (X, Y) , queremos estudiar cómo se distribuye la probabilidad de sucesos cualesquiera de la forma $\{(X, Y) \in B\}$, donde B es una región del plano.

Para ello, definimos la **función de distribución conjunta**:

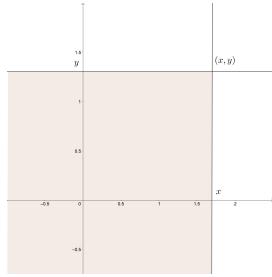
Definición de función de distribución conjunta: Dada una variable bidimensional (X, Y) , definimos su **función de distribución conjunta** F_{XY} a la función definida sobre \mathbb{R}^2 de la manera siguiente:

$$\begin{aligned} F_{XY} : \mathbb{R}^2 &\longrightarrow \mathbb{R} \\ (x, y) &\longrightarrow F_{XY}(x, y) = P(X \leq x, Y \leq y). \end{aligned}$$

Por lo tanto, dado un valor $(x, y) \in \mathbb{R}^2$, consideramos la región del plano $(-\infty, x] \times (-\infty, y]$:

Entonces la **función de distribución conjunta** en el valor (x, y) es la probabilidad del suceso formado por aquellos elementos tal que la imagen por la **variable aleatoria bidimensional** (X, Y) caen dentro de la región sombreada en el gráfico anterior:

$$\begin{aligned} F_{XY}(x, y) &= P\{w \in \Omega, | (X(w), Y(w)) \in (-\infty, x] \times (-\infty, y]\} \\ &= P\{w \in \Omega, | X(w) \leq x, Y(w) \leq y\}. \end{aligned}$$



5.2.2. Propiedades

Sea (X, Y) una variable bidimensional. y sea F_{XY} su **función de distribución conjunta**. Dicha función satisface las propiedades siguientes:

- La función de distribución conjunta es no decreciente en cada una de las variables:

$$\text{Si } x_1 \leq x_2, \text{ y } y_1 \leq y_2, \text{ entonces, } F_{XY}(x_1, y_1) \leq F_{XY}(x_2, y_2).$$

- $F_{XY}(x, -\infty) = F_{XY}(-\infty, y) = 0$, $F_{XY}(\infty, \infty) = 1$, para todo $x, y \in \mathbb{R}$.
- Las variables aleatorias X e Y se llaman **variables aleatorias marginales** y sus funciones de distribución F_X y F_Y pueden hallarse de la forma siguiente como función de la **función de distribución conjunta** F_{XY} :

$$F_X(x) = F_{XY}(x, \infty), \quad F_Y(y) = F_{XY}(\infty, y),$$

para todo $x, y \in \mathbb{R}$.

- La función de distribución conjunta es continua por el “norte” y por el “este”:

$$\begin{aligned} \lim_{x \rightarrow a^+} F_{XY}(x, y) &= \lim_{x \rightarrow a, x > a} F_{XY}(x, y) = F_{XY}(a, y), \\ \lim_{y \rightarrow b^+} F_{XY}(x, y) &= \lim_{y \rightarrow b, y > b} F_{XY}(x, y) = F_{XY}(x, b), \end{aligned}$$

para todo $a, b \in \mathbb{R}$. Ver la siguiente figura.

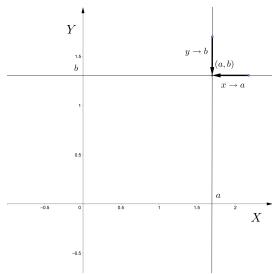
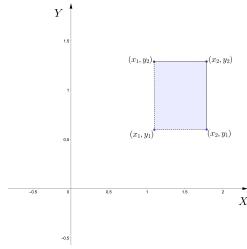


Figura 5.2: Función de distribución conjunta

- Dados $x_1 < x_2$ y $y_1 < y_2$, consideramos B el rectángulo de vértices (x_1, y_1) , (x_1, y_2) , (x_2, y_1) y (x_2, y_2) : $(x_1, x_2] \times (y_1, y_2]$. Entonces,

$$\begin{aligned} P((X, Y) \in B) &= F_{XY}(x_2, y_2) - F_{XY}(x_2, y_1) - F_{XY}(x_1, y_2) \\ &\quad + F_{XY}(x_1, y_1). \end{aligned}$$

**Ejemplo: Función de distribución uniforme**

Consideremos una variable aleatoria bidimensional (X, Y) con **función de distribución conjunta**:

$$F_{XY}(x, y) = \begin{cases} 0, & \text{si } x < 0, \text{ o } y < 0, \\ xy, & \text{si } 0 \leq x \leq 1, \text{ } 0 \leq y \leq 1, \\ x, & \text{si } 0 \leq x \leq 1, \text{ } y > 1, \\ y, & \text{si } 0 \leq y \leq 1, \text{ } x > 1, \\ 1, & \text{si } x \geq 1, \text{ } y \geq 1. \end{cases}$$

En la figura siguiente, hemos representado por zonas cómo está definida F_{XY} .

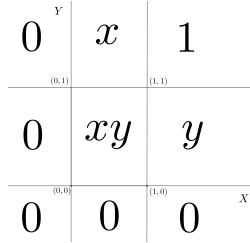


Figura 5.3: Función de distribución uniforme bidimensional.

Comprobemos algunas de las propiedades que hemos enunciado anteriormente:

- Claramente $F_{XY}(x, -\infty) = F_{XY}(-\infty, y) = 0$ ya que $F_{XY}(x, y) = 0$ si $x < 0$ o $y < 0$. Por tanto, si hacemos tender x o y hacia $-\infty$, obtendremos que $F_{XY}(x, -\infty) = F_{XY}(-\infty, y) = 0$.
- De la misma manera $F_{XY}(\infty, \infty) = 1$ ya que $F_{XY}(x, y) = 1$ para $x > 1$ e $y > 1$. Por tanto, si hacemos tender x e y hacia ∞ , obtendremos $F_{XY}(\infty, \infty) = 1$.
- Hallemos las marginales:

$$F_X(x) = F_{XY}(x, \infty) = \begin{cases} 0, & \text{si } x < 0, \\ x, & \text{si } 0 \leq x \leq 1, \\ 1, & \text{si } x > 1. \end{cases}$$

Para ver la expresión anterior basta trazar la recta vertical $X = x$ en el gráfico anterior y ver hacia dónde tiende a medida que la y se va hacia ∞ .

¿Habéis averiguado cuál es la distribución de X ?

¡Efectivamente!, X es la uniforme en el intervalo $(0, 1)$.

Dejamos como ejercicio hallar la distribución marginal para la variable Y .

- Comprobemos que F_{XY} es continua por el “norte” y el “este” en el punto $(1, 1)$ que es un punto problemático:

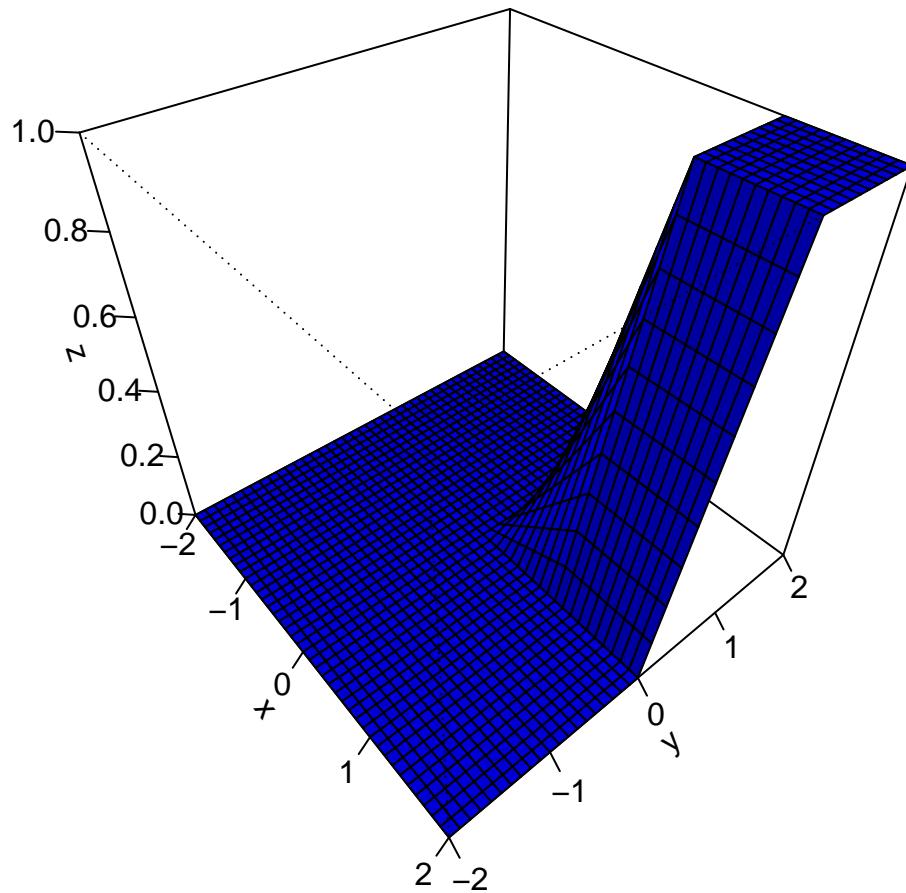
$$\lim_{x \rightarrow 1, x > 1} F_{XY}(x, 1) = \lim_{x \rightarrow 1, x > 1} 1 = F_{XY}(1, 1),$$

$$\lim_{y \rightarrow 1, y > 1} F_{XY}(1, y) = \lim_{y \rightarrow 1, y > 1} 1 = F_{XY}(1, 1).$$

Representemos en un gráfico tridimensional la **función de distribución conjunta** usando la función `persp` de R para x e y entre -2 y 2.

Primero definimos la **función** y luego la dibujamos:

```
f.dist.con = function(x,y){ifelse(x<0 | y<0,0,
                                 ifelse(x>=0 & x<=1 & y>=0 & y<=1,x*y,
                                 ifelse(x>=0 & x<=1 & y >1,x,ifelse(y>=0 & y<=1 & x>1,y,1))))}
x=seq(from=-2,to=2,by=0.1)
y=seq(from=-2,to=2,by=0.1)
z=outer(x,y,f.dist.con)
persp(x,y,z,theta=50,phi=40,col="blue",shade=0.25,ticktype="detailed")
```



Ejemplo: dos lanzamientos de un dado no trucado

Consideremos el experimento aleatorio de lanzar dos veces un dado no trucado.

Sea (S, P) la **variable aleatoria bidimensional** que nos da la suma y el producto de los resultados obtenidos, respectivamente.

La **función de distribución conjunta** en el valor $(3, 4)$ es:

$$F_{XY}(3, 4) = P(S \leq 3, P \leq 4) = P\{(1, 1), (1, 2), (2, 1)\} = \frac{3}{36} = \frac{1}{12} \approx 0,083,$$

ya que Ω tiene en total 36 resultados:

$$\Omega = \{(1, 1), (1, 2), \dots, (6, 6)\}.$$

y los únicos resultados en los que la suma es menor o igual que 3 y el producto menor o igual que 4 son $(1, 1)$ (suma 2 producto 1), $(1, 2)$ (suma 3 y producto 2) y $(2, 1)$ (suma 3 y producto 2).

Ejercicio

Hallar el valor de la **función de distribución conjunta** para la **variable aleatoria bidimensional** anterior (S, P) en los valores (i, j) siguientes: $(4, 5)$, $(4, 9)$, $(5, 9)$, $(6, 10)$.

5.3. Variables aleatorias bidimensionales discretas

Definición de variable aleatoria bidimensional discreta: Sea (X, Y) una **variable aleatoria bidimensional**. Diremos que es discreta cuando su conjunto de valores en \mathbb{R}^2 , $(X, Y)(\Omega)$ es un conjunto finito o numerable.

En la mayoría de los casos, dicho conjunto es un subconjunto de los enteros naturales.

Ejemplo: lanzamiento de dos dados (continuación)

La variable aleatoria bidimensional anterior que nos daba la suma y el producto de los resultados obtenidos por los dos lanzamientos, respectivamente es discreta ya que:

$$(S, P)(\Omega) = \{(2, 1), (3, 2), (4, 3), (4, 4), (5, 4), (5, 6), (6, 5), (6, 8), (6, 9), (7, 6), (7, 10), (7, 12), (8, 12), (8, 15), (8, 16), (9, 18), (9, 20), (10, 24), (10, 25), (11, 30), (12, 36)\}.$$

Comprobar que el conjunto $(S, P)(\Omega)$ dado por el ejemplo coincide con la expresión dada. O lo que es lo mismo, hallar el conjunto $(S, P)(\Omega)$:

$$\begin{aligned} (S, P) : \Omega &\longrightarrow \mathbb{R}^2 \\ (1, 1) &\longrightarrow (S(1, 1), P(1, 1)) = (2, 1), \\ (1, 2) &\longrightarrow (S(1, 2), P(1, 2)) = (3, 2), \\ \vdots &\quad \vdots \\ (6, 6) &\longrightarrow (S(6, 6), P(6, 6)) = (12, 36). \end{aligned}$$

5.3.1. Función de probabilidad conjunta

Definición de función de probabilidad conjunta: Dada una **variable aleatoria bidimensional discreta** (X, Y) con $(X, Y)(\Omega) = \{(x_i, y_j), i = 1, 2, \dots, j = 1, 2, \dots\}$, definimos la función de probabilidad discreta P_{XY} para un valor $(x, y) \in \mathbb{R}^2$ de la siguiente forma:

$$\begin{aligned} P_{XY} : \mathbb{R}^2 &\longrightarrow \mathbb{R} \\ (x, y) &\longrightarrow P_{XY}(x, y) = P(X = x, Y = y). \end{aligned}$$

Observación: Si $(x, y) \notin (X, Y)(\Omega)$, el valor de la **función de probabilidad conjunta** en (x, y) es nula: $P_{XY}(x, y) = 0$. El motivo es que, en este caso, el conjunto $\{w \in \Omega, |(X(w), Y(w)) = (x, y)\} = \emptyset$ es vacío pues $(x, y) \notin (X, Y)(\Omega)$.

Ejemplo: lanzamiento de dos dados (continuación) Por tanto, de cara a calcular P_{XY} basta conocer los valores de $P_{XY}(x_i, y_j)$ para $(x_i, y_j) \in (X, Y)(\Omega)$:

X/Y	y_1	y_2	...	y_N
x_1	$P_{XY}(x_1, y_1)$	$P_{XY}(x_1, y_2)$...	$P_{XY}(x_1, y_N)$
x_2	$P_{XY}(x_2, y_1)$	$P_{XY}(x_2, y_2)$...	$P_{XY}(x_2, y_N)$
\vdots	\vdots	\vdots	\vdots	\vdots
x_M	$P_{XY}(x_M, y_1)$	$P_{XY}(x_M, y_2)$...	$P_{XY}(x_M, y_N)$

La **función de probabilidad conjunta** es:

P/S	2	3	4	5	6	7	8	9	10	11	12
1	$\frac{1}{36}$	0	0	0	0	0	0	0	0	0	0
2	0	$\frac{2}{36}$	0	0	0	0	0	0	0	0	0
3	0	0	$\frac{2}{36}$	0	0	0	0	0	0	0	0
4	0	0	$\frac{1}{36}$	$\frac{2}{36}$	0	0	0	0	0	0	0
5	0	0	0	0	$\frac{1}{36}$	0	0	0	0	0	0
6	0	0	0	$\frac{1}{36}$	0	$\frac{2}{36}$	0	0	0	0	0
8	0	0	0	0	$\frac{1}{36}$	0	0	0	0	0	0
9	0	0	0	0	$\frac{1}{36}$	0	0	0	0	0	0
10	0	0	0	0	0	$\frac{2}{36}$	0	0	0	0	0
12	0	0	0	0	0	$\frac{2}{36}$	$\frac{2}{36}$	0	0	0	0
15	0	0	0	0	0	0	$\frac{1}{36}$	0	0	0	0
16	0	0	0	0	0	0	$\frac{1}{36}$	0	0	0	0
18	0	0	0	0	0	0	0	$\frac{2}{36}$	0	0	0
20	0	0	0	0	0	0	0	$\frac{2}{36}$	0	0	0
24	0	0	0	0	0	0	0	0	$\frac{2}{36}$	0	0
25	0	0	0	0	0	0	0	0	$\frac{1}{36}$	0	0
30	0	0	0	0	0	0	0	0	0	$\frac{2}{36}$	0
36	0	0	0	0	0	0	0	0	0	0	$\frac{1}{36}$

Vamos a definir unas funciones en R para calcular la **función de probabilidad conjunta**.

La función `pdado` devuelve la probabilidad de que salga la cara `x` en un dado de `n` caras donde por defecto `n = 6`:

```
pdado =function(x,n=6)  sapply(x,FUN=function(x)
  if( x %in% c(1:n)) {return(1/n)} else {return(0)})
```

Vamos a probarla. La probabilidad de que salga la cara 4 en un dado de 6 caras vale:

```
pdado(4,6)
```

```
## [1] 0.1666667
```

La función `pdado2` devuelve la probabilidad de que salgan las caras `x` e `y` cuando lanzamos un dado de `n` caras dos veces:

```
pdado2 =function(x,y,n=6) {pdado(x,n)*pdado(y,n)}
```

Por ejemplo la probabilidad de que salgan las caras 3 y 4 en un dado de 6 caras es:

```
pdado2(3,4,6)
```

```
## [1] 0.02777778
```

La función `psum_prod` nos da la **función de probabilidad conjunta** de la suma y el producto cuando lanzamos dos dados de `n` caras:

```

psum_prod=function(x,y,n=6){
  Dxy=data.frame(d1=rep(1:n,each=n),d2=rep(1:n,times=n))
  Dxy$suma=Dxy$d1+Dxy$d2
  Dxy$producto=Dxy$d1*Dxy$d2
  aux=Dxy[Dxy$suma==x & Dxy$producto==y,]
  sum(apply(aux[,1:2],FUN=function(x) {pdado2(x[1],x[2],n=n)},1))
}

```

Por ejemplo, sabemos que $P_{SP}(6, 8) = \frac{2}{36} = 0,0556$:

`psum_prod(6,8)`

```
## [1] 0.05555556
```

la tabla de la **función de probabilidad conjunta** para la variable (S, P) hacemos lo siguiente:

```

n=6
Dxy=data.frame(d1=rep(1:n,each=n),d2=rep(1:n,times=n))
Dxy$suma=Dxy$d1+Dxy$d2
Dxy$producto=Dxy$d1*Dxy$d2
tabla.func.prob.conjunta=prop.table(table(Dxy$suma,Dxy$producto))
knitr::kable(round(tabla.func.prob.conjunta[,1:9],3))

```

```
knitr::kable(round(tabla.func.prob.conjunta[,10:18],3))
```

lo cual finaliza nuestros cálculos en R.

5.3.1.1. Propiedades de la función de probabilidad conjunta

Sea (X, Y) una variable aleatoria bidimensional discreta con conjunto de valores $(X, Y)(\Omega) = \{(x_i, y_j) | i = 1, 2, \dots, j = 1, 2, \dots\}$. Entonces su función de probabilidad conjunta verifica las propiedades siguientes:

La suma de todos los valores de la **función de probabilidad conjunta** sobre el conjunto de valores siempre vale 1:

$$\sum_i \sum_j P_{XY}(x_i, y_j) = 1.$$

Sea B una región del plano. El valor de la probabilidad $P((X, Y) \in B)$ se puede calcular de la forma siguiente:

$$P((X, Y) \in B) = \sum_{(x_i, y_j) \in B} P_{XY}(x_i, y_j).$$

Es decir, la probabilidad de que la variable bidimensional tome valores en B es igual a la suma de todos aquellos valores de la función de probabilidad conjunta que están en B .

En particular, tenemos la siguiente propiedad que relaciona la **función de distribución conjunta** con la **función de probabilidad conjunta**:

$$F_{XY}(x, y) = \sum_{x_i \leq x, y_j \leq y} P_{XY}(x_i, y_j).$$

Dicha expresión se deduce de la expresión anterior considerando $B = (-\infty, x] \times (-\infty, y]$.

Ejercicio

En el ejemplo del lanzamiento de los dos dados. Comprobad usando la tabla de la función de probabilidad conjunta que la suma de todos sus valores suma 1.

Ejemplo: lanzamientos de dados (continuación)

Apliquemos la fórmula que relaciona la función de distribución conjunta con la función de probabilidad conjunta para $(x, y) = (5, 4)$.

Recordemos la tabla de la función de probabilidad conjunta hasta $S = 5$ y $P = 4$:

Observamos que los únicos valores $(x_i, y_j) \in (X, Y)(\Omega)$ que verifican $x_i \leq 5$ y $y_j \leq 4$ son $(2, 1)$, $(3, 2)$, $(4, 3)$, $(4, 4)$ y $(5, 4)$. Por tanto,

$$\begin{aligned} F_{SP}(5, 4) &= P_{SP}(2, 1) + P_{SP}(3, 2) + P_{SP}(4, 3) + P_{SP}(4, 4) + P_{SP}(5, 4) \\ &= \frac{1}{36} + \frac{2}{36} + \frac{2}{36} + \frac{1}{36} + \frac{2}{36} = \frac{8}{36} = \frac{2}{9}. \end{aligned}$$

Es decir, “a largo plazo”, de cada 9 ocasiones que lanzamos un dado dos veces; en 2 ocasiones obtenemos un resultado cuya suma es menor o igual que 5 y cuyo producto es menor o igual que 4.

Para definir la **función de distribución conjunta** definimos la función siguiente en R:

```
func.dist.conj = function(x,y,n=6){
  sum(tabla.func.prob.conjunta[as.integer(rownames(tabla.func.prob.conjunta))<=x,
                                as.integer(colnames(tabla.func.prob.conjunta)) <=y])
}
```

Comprobemos que $F_{SP}(5, 4) = \frac{2}{9} = 0,2222$:

```
func.dist.conj(5,4)
```

```
## [1] 0.2222222
```

5.3.2. Distribuciones marginales

Consideremos una variable aleatoria **bidimensional discreta** (X, Y) con **función de probabilidad conjunta** $P_{XY}(x_i, y_j)$, con $(x_i, y_j) \in (X, Y)(\Omega)$, $i = 1, 2, \dots$, $j = 1, 2, \dots$

La tabla de la **función de probabilidad conjunta** contiene suficiente información para obtener las **funciones de probabilidad** de las variables X e Y .

Dichas variables X e Y se denominan **distribuciones marginales** y sus correspondientes **funciones de probabilidad, funciones de probabilidad marginales** P_X de la variable X y P_Y de la variable Y .

Veamos cómo obtener P_X y P_Y a partir de la tabla P_{XY} .

Proposición. Expresión de las funciones de probabilidad marginales. Sea (X, Y) una variable aleatoria **bidimensional discreta** con **función de probabilidad conjunta** $P_{XY}(x_i, y_j)$, con $(x_i, y_j) \in (X, Y)(\Omega)$, $i = 1, 2, \dots$, $j = 1, 2, \dots$

Las **funciones de probabilidad marginales** $P_X(x_i)$ y $P_Y(y_j)$ se calculan usando las expresiones siguientes:

$$\begin{aligned} P_X(x_i) &= \sum_{j=1} P_{XY}(x_i, y_j), \quad i = 1, 2, \dots, \\ P_Y(y_j) &= \sum_{i=1} P_{XY}(x_i, y_j), \quad j = 1, 2, \dots \end{aligned}$$

Si consideramos la función P_{XY} como una tabla bidimensional en la que en la primera fila están los valores de la variable Y (y_1, y_2, \dots) y en la primera columna están los valores de la variable X (x_1, x_2, \dots). Para obtener la **función de probabilidad marginal** de la variable X en el valor x_i , $P_X(x_i)$, hay que sumar todos los valores de $P_{XY}(x_i, y_j)$ correspondientes a la fila i -ésima y para obtener la **función de probabilidad marginal** de la variable Y en el valor y_j , $P_Y(y_j)$, hay que sumar todos los valores de $P_{XY}(x_i, y_j)$ correspondientes a la columna j -ésima.

5.3.2.1. Ejemplo

Ejemplo de la suma y el producto de los resultados de dos lanzamientos de un dado

Hallemos la función de probabilidad marginal para la suma de los resultados S :

Usando la expresión de la probabilidad marginal tenemos que

$$\begin{aligned}
 P_S(2) &= P_{SP}(2, 1) = \frac{1}{36}, \\
 P_S(3) &= P_{SP}(3, 2) = \frac{2}{36}, \\
 P_S(4) &= P_{SP}(4, 3) + P_{SP}(4, 4) = \frac{2}{36} + \frac{1}{36} = \frac{3}{36} = \frac{1}{12}, \\
 P_S(5) &= P_{SP}(5, 4) + P_{SP}(5, 6) = \frac{2}{36} + \frac{2}{36} = \frac{4}{36} = \frac{1}{9}, \\
 P_S(6) &= P_{SP}(6, 5) + P_{SP}(6, 8) + P_{SP}(6, 9) = \frac{2}{36} + \frac{2}{36} + \frac{1}{36} = \frac{5}{36}, \\
 P_S(7) &= P_{SP}(7, 6) + P_{SP}(7, 10) + P_{SP}(7, 12) = \frac{2}{36} + \frac{2}{36} + \frac{2}{36} = \frac{6}{36} = \frac{1}{6}, \\
 P_S(8) &= P_{SP}(8, 12) + P_{SP}(8, 15) + P_{SP}(8, 16) = \frac{2}{36} + \frac{2}{36} + \frac{1}{36} = \frac{5}{36}, \\
 P_S(9) &= P_{SP}(9, 18) + P_{SP}(9, 20) = \frac{2}{36} + \frac{2}{36} = \frac{4}{36} = \frac{1}{9}, \\
 P_S(10) &= P_{SP}(10, 24) + P_{SP}(10, 25) = \frac{2}{36} + \frac{1}{36} = \frac{3}{36} = \frac{1}{12}, \\
 P_S(11) &= P_{SP}(11, 30) = \frac{2}{36}, \\
 P_S(12) &= P_{SP}(12, 36) = \frac{1}{36}.
 \end{aligned}$$

La **función de probabilidad marginal** de la suma S queda resumida en la tabla siguiente:

S	2	3	4	5	6	7	8	9	10	11	12
P_S	$\frac{1}{36}$	$\frac{2}{36}$	$\frac{3}{36}$	$\frac{4}{36}$	$\frac{5}{36}$	$\frac{6}{36}$	$\frac{5}{36}$	$\frac{4}{36}$	$\frac{3}{36}$	$\frac{2}{36}$	$\frac{1}{36}$

Para hallar la **función de probabilidad marginal** de la suma basta sumar las filas de la tabla que nos daba la **función de probabilidad conjunta**:

```
marginal.suma = apply(tabla.func.prob.conjunta, 1, sum)
marginal.suma
```

```
##      2      3      4      5      6      7      8
## 0.02777778 0.05555556 0.08333333 0.11111111 0.13888889 0.16666667 0.13888889
##      9     10     11     12
## 0.11111111 0.08333333 0.05555556 0.02777778
```

De la misma manera, para hallar la **función de probabilidad marginal** del producto basta sumar las columnas de la tabla anterior:

```
marginal.producto = apply(tabla.func.prob.conjunta, 2, sum)
marginal.producto
```

```
##      1      2      3      4      5      6      8
## 0.02777778 0.05555556 0.05555556 0.08333333 0.05555556 0.11111111 0.05555556
##      9     10     12     15     16     18     20
## 0.02777778 0.05555556 0.11111111 0.05555556 0.02777778 0.05555556 0.05555556
##     24     25     30     36
## 0.05555556 0.02777778 0.05555556 0.02777778
```

5.4. Variables aleatorias bidimensionales continuas

5.4.1. Definición

Recordemos la definición de **variable continua unidimensional**: X es continua si existe una función $f_X : \mathbb{R} \rightarrow \mathbb{R}$, llamada **función de densidad** no negativa $f_X(x) \geq 0$, para todo $x \in \mathbb{R}$ tal que para cualquier intervalo (a, b) , la probabilidad de que X esté en (a, b) se calcula de la forma siguiente:

$$P(X \in B) = P(a < X < b) = \int_B f_X(x) du = \int_a^b f_X(x) dx.$$

La generalización natural es, entonces:

Definición de variable aleatoria bidimensional continua. Sea (X, Y) una variable aleatoria bidimensional. Diremos que (X, Y) es continua si existe una función $f_{XY} : \mathbb{R}^2 \rightarrow \mathbb{R}$ llamada **función de densidad** no negativa $f_{XY}(x, y) \geq 0$ para todo $(x, y) \in \mathbb{R}^2$ tal que dado cualquier región B del plano, la probabilidad de que (X, Y) esté en B se calcula de la forma siguiente:

$$P((X, Y) \in B) = \int \int_B f_{XY}(x, y) dx dy.$$

5.4.1.1. Ejemplos

Ejemplo: cálculo probabilidad distribución bidimensional

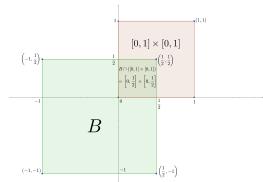
Consideremos la siguiente **función de densidad**:

$$f_{XY}(x, y) = \begin{cases} 1, & \text{si } 0 \leq x \leq 1, 0 \leq y \leq 1, \\ 0, & \text{en caso contrario.} \end{cases}$$

En este caso, si consideramos $B = [-1, \frac{1}{2}] \times [-1, \frac{1}{2}]$, la probabilidad de que (X, Y) esté en B se calcularía de la forma siguiente:

$$\begin{aligned} P((X, Y) \in B) &= \int \int_B f_{XY}(x, y) dx dy = \int_{-1}^{\frac{1}{2}} \int_{-1}^{\frac{1}{2}} f_{XY}(x, y) dx dy \\ &= \int_0^{\frac{1}{2}} \int_0^{\frac{1}{2}} 1 dx dy = \int_0^{\frac{1}{2}} 1 dx \int_0^{\frac{1}{2}} 1 dy = \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{4}. \end{aligned}$$

En la figura siguiente hemos dibujado en morado la región donde f_{XY} no es cero, es decir $[0, 1] \times [0, 1]$, la región B en verde y la región intersección de las dos anteriores que es donde tenemos que integrar la **función de densidad** dada.



5.4.2. Propiedades de la función de densidad

Sea (X, Y) una **variable aleatoria bidimensional continua** con **función de densidad** f_{XY} . Esta función verifica las propiedades siguientes:

- La integral de dicha función sobre todo el plano vale 1:

$$\int \int_{\mathbb{R}^2} f_{XY}(x, y) dx dy = 1.$$

Para ver dicha propiedad, basta considerar $B = \mathbb{R}^2$, tener en cuenta que el suceso $(X, Y) \in \mathbb{R}^2$ es el total Ω y aplicar la definición de f_{XY} :

$$P((X, Y) \in \mathbb{R}^2) = 1 = \int \int_{\mathbb{R}^2} f_{XY}(x, y) dx dy.$$

- La relación que hay entre la **función de distribución** F_{XY} y la **función de densidad** f_{XY} es la siguiente:

$$F_{XY}(x, y) = \int_{-\infty}^x \int_{-\infty}^y f_{XY}(u, v) du dv.$$

Para ver dicha propiedad, basta considerar $B = (-\infty, x] \times (-\infty, y]$ y aplicar la definición de **función de distribución**:

$$F_{XY}(x, y) = P((X, Y) \in (-\infty, x] \times (-\infty, y]) = \int_{-\infty}^x \int_{-\infty}^y f_{XY}(u, v) du dv.$$

- La relación que hay entre la **función de densidad** f_{XY} y la **función de distribución** F_{XY} es la siguiente:

$$f_{XY}(x, y) = \frac{\partial^2 F_{XY}(x, y)}{\partial x \partial y}.$$

Dicha propiedad se deduce de la anterior, derivando primero respecto a x y después respecto a y para eliminar las dos integrales.

- Las **funciones de densidad marginales** de las variables X e Y , $f_X(x)$ y $f_Y(y)$ respectivamente, se calculan de la forma siguiente:

$$f_X(x) = \int_{-\infty}^{\infty} f_{XY}(x, y) dy, \quad f_Y(y) = \int_{-\infty}^{\infty} f_{XY}(x, y) dx$$

5.4.2.1. Ejemplos

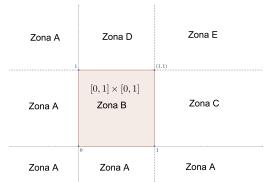
Ejemplo: continuación ejemplo anterior

Comprobemos las propiedades usando la **función de densidad** del ejemplo anterior: $f_{XY}(x, y) = \begin{cases} 1, & \text{si } 0 \leq x \leq 1, 0 \leq y \leq 1, \\ 0, & \text{en caso contrario.} \end{cases}$

- La integral de f_{XY} sobre todo el plano vale 1:

$$\begin{aligned} \int \int_{\mathbb{R}^2} f_{XY}(x, y) dx dy &= \int_0^1 \int_0^1 1 dx dy \\ &= \int_0^1 1 dx \int_0^1 1 dy = 1 \cdot 1 = 1. \end{aligned}$$

- Vamos a calcular la función de distribución F_{XY} . Para ello dividimos el plano en 5 zonas tal como muestra la figura siguiente:

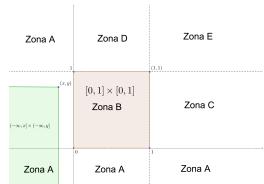


Sea (x, y) un punto cualquiera de \mathbb{R}^2 . De cara a calcular $F_{XY}(x, y)$ tenemos que averiguar el conjunto intersección siguiente: $([0, 1] \times [0, 1]) \cap ((-\infty, x] \times (-\infty, y])$ ya que el dominio donde f_{XY} es no nula es $[0, 1] \times [0, 1]$ y la función de distribución $F_{XY}(x, y)$ valdrá:

$$\begin{aligned} F_{XY}(x, y) &= \int_{-\infty}^x \int_{-\infty}^y f_{XY}(u, v) du dv \\ &= \int \int_{([0,1]\times[0,1]) \cap ((-\infty,x]\times(-\infty,y])} f_{XY}(u, v) du dv. \end{aligned}$$

- Caso $(x, y) \in \text{Zona A}$ o $x < 0$ o $y < 0$ En este caso: $([0, 1] \times [0, 1]) \cap ((-\infty, x] \times (-\infty, y)) = \emptyset$. Ver figura siguiente donde la zona morada ($[0, 1] \times [0, 1]$) no se interseca con la zona verde $((-\infty, x] \times (-\infty, y))$.

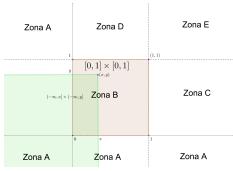
Por tanto en este caso, $F_{XY}(x, y) = 0$.



- Caso $(x, y) \in \text{Zona B}$, o $(x, y) \in [0, 1] \times [0, 1]$. En este caso: $([0, 1] \times [0, 1]) \cap ((-\infty, x] \times (-\infty, y)) = [0, x] \times [0, y]$. Ver figura siguiente.

Por tanto en este caso, 19

$$F_{XY}(x, y) = \int_0^x \int_0^y 1 du dv = \int_0^x 1 du \int_0^y 1 dy = x \cdot y.$$



Dejamos como ejercicio los otros casos. En resumen:

$$F_{XY}(x, y) = \begin{cases} 0, & \text{si } x < 0, \text{ o } y < 0, \\ xy, & \text{si } (x, y) \in [0, 1] \times [0, 1], \\ x, & \text{si } 0 \leq x \leq 1, \text{ } y > 1, \\ y, & \text{si } x > 1, \text{ } 0 \leq y \leq 1, \\ 1, & \text{si } x > 1, \text{ } y > 1. \end{cases}$$

¿Os suena?

Ver el primer ejemplo que pusimos del tema. Es la misma variable aleatoria bidimensional. Ahora sabemos que se trata de una **variable aleatoria bidimensional continua**.

Comprobemos ahora que si derivamos dos veces la expresión de F_{XY} , primero respecto x y después respecto y , obtendremos la **función de densidad** f_{XY} .

Si derivamos respecto x obtenemos:

$$\frac{\partial F_{XY}(x, y)}{\partial x} = \begin{cases} 0, & \text{si } x < 0, \text{ o } y < 0, \\ y, & \text{si } (x, y) \in [0, 1] \times [0, 1], \\ 1, & \text{si } 0 \leq x \leq 1, \text{ } y > 1, \\ 0, & \text{si } x > 1, \text{ } 0 \leq y \leq 1, \\ 0, & \text{si } x > 1, \text{ } y > 1. \end{cases}$$

Si ahora derivamos respecto y obtenemos:

$$\frac{\partial^2 F_{XY}(x, y)}{\partial y \partial x} = \begin{cases} 0, & \text{si } x < 0, \text{ o } y < 0, \\ 1, & \text{si } (x, y) \in [0, 1] \times [0, 1], \\ 0, & \text{si } 0 \leq x \leq 1, \text{ } y > 1, \\ 0, & \text{si } x > 1, \text{ } 0 \leq y \leq 1, \\ 0, & \text{si } x > 1, \text{ } y > 1, \end{cases}$$

expresión que coincide con la **función de densidad** $f_{XY}(x, y)$.

Hallaremos para finalizar las **funciones de densidad marginales**. Empecemos con $f_X(x)$:

$$f_X(x) = \int_{-\infty}^{\infty} f_{XY}(x, y) dy.$$

Recordemos que la región donde no se anulaba la **función de densidad conjunta** f_{XY} era el cuadrado $[0, 1] \times [0, 1]$. Por tanto, fijado x , el valor de $f_X(x)$ es no nulo si la recta vertical $X = x$ interseca dicho

cuadrado. Y esto ocurre siempre que $x \in (0, 1)$. Por tanto,

$$f_X(x) = \begin{cases} \int_0^1 f_{XY}(x, y) dy = \int_0^1 1 dy = 1, & \text{si } x \in (0, 1), \\ 0, & \text{en caso contrario.} \end{cases}$$

Por tanto la variable X sigue la distribución uniforme en el intervalo $[0, 1]$.

Dejamos como ejercicio comprobar que la variable Y también sigue la distribución uniforme en el mismo intervalo.

Ejemplo: otra función de densidad bidimensional

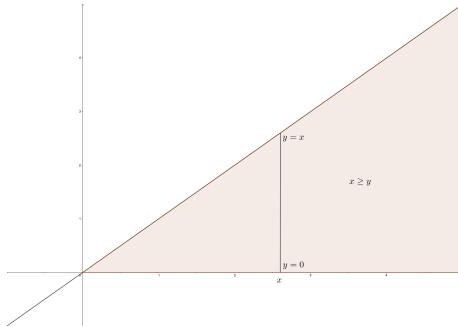
Consideremos la variable aleatoria bidimensional (X, Y) con **función de densidad**:

$$f_{XY}(x, y) = \begin{cases} c \cdot e^{-x} \cdot e^{-y}, & 0 \leq y \leq x < \infty, \\ 0, & \text{en caso contrario,} \end{cases}$$

donde c es un valor que se tiene que hallar para que f_{XY} sea función de densidad. Calcular c y comprobar todas las propiedades de la función de densidad.

Para hallar c , hemos de imponer que la integral de la función anterior debe ser 1 sobre todo el plano \mathbb{R}^2 .

Primero fijémonos en como es la región de integración (zona morada de la figura). Fijado un valor $x \geq 0$, el valor y va desde $y = 0$ hasta $y = x$. Por tanto, para calcular el valor de c , hay que hacer lo siguiente:



$$\begin{aligned} 1 &= \int_{\mathbb{R}^2} f_{XY}(x, y) dx dy = \int_{x=0}^{x=\infty} \int_{y=0}^{y=x} c \cdot e^{-x} \cdot e^{-y} dy dx \\ &= c \cdot \int_{x=0}^{x=\infty} e^{-x} \cdot \int_{y=0}^{y=x} e^{-y} dy dx = c \cdot \int_{x=0}^{x=\infty} e^{-x} \cdot [-e^{-y}]_{y=0}^{y=x} dx \\ &= c \cdot \int_{x=0}^{x=\infty} e^{-x} \cdot (1 - e^{-x}) dx = c \cdot \int_{x=0}^{x=\infty} (e^{-x} - e^{-2x}) dx \\ &= c \cdot [-e^{-x} + \frac{1}{2}e^{-2x}]_{x=0}^{x=\infty} = c \left(1 - \frac{1}{2}\right) = \frac{c}{2}. \end{aligned}$$

El valor de c es $c = 2$.

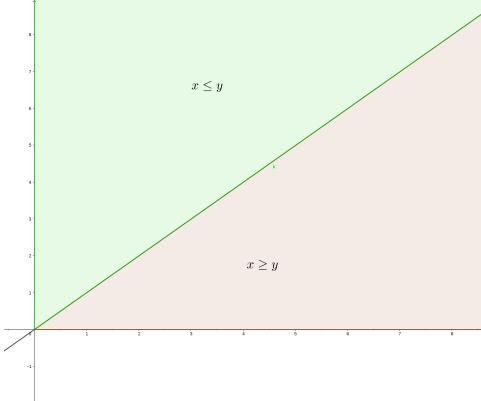
Vamos a calcular seguidamente su función de distribución.

Fijémonos que, en este caso, si $x < 0$ o $y < 0$, $F_{XY}(x, y) = 0$, ya que el dominio $B = (-\infty, x] \times (-\infty, y]$ no interseca la zona morada del gráfico anterior.

Suponemos entonces que $x \geq 0$ e $y \geq 0$.

Vamos a considerar dos casos:

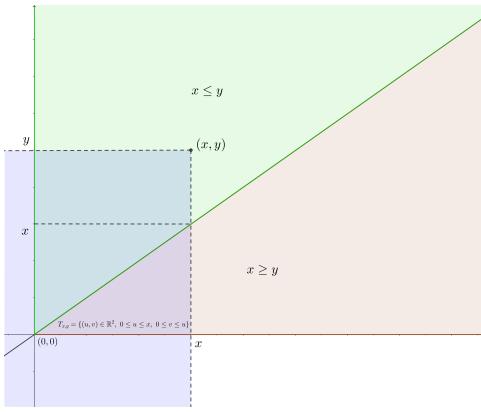
- $x \leq y$. Ver zona verde del gráfico siguiente.
- $x \geq y$. Ver zona morada del gráfico siguiente.



- Caso $x \leq y$ (zona verde de la figura adjunta). En este caso, si hacemos la intersección de la región $B = (-\infty, x] \times (-\infty, y]$ (zona azul) con la zona morada o región donde $f_{XY}(x, y) \neq 0$ obtenemos el triángulo $T_{x,y} = \{(u, v) \in \mathbb{R}^2, 0 \leq u \leq x, 0 \leq v \leq u\}$. Ver figura adjunta.

Por tanto,

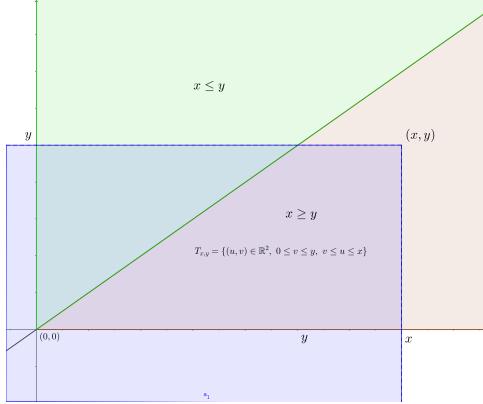
$$\begin{aligned} F_{XY}(x, y) &= \int_{u=0}^{u=x} \int_{v=0}^{v=u} f_{XY}(u, v) dv du = 2 \cdot \int_{u=0}^{u=x} e^{-u} \int_{v=0}^{v=u} e^{-v} dv du \\ &= 2 \cdot \int_{u=0}^{u=x} e^{-u} \cdot [-e^{-v}]_{v=0}^{v=u} du = 2 \cdot \int_{u=0}^{u=x} e^{-u} \cdot (1 - e^{-u}) du \\ &= 2 \int_{u=0}^{u=x} (e^{-u} - e^{-2u}) du = 2 \cdot [-e^{-u} + \frac{1}{2} \cdot e^{-2u}]_{u=0}^{u=x} \\ &= 2 \cdot \left(-e^{-x} + \frac{1}{2} \cdot e^{-2x} + 1 - \frac{1}{2} \right) = 1 - 2 \cdot e^{-x} + e^{-2x}. \end{aligned}$$



- Caso $x \geq y$ (zona morada de la figura adjunta). En este caso, si hacemos la intersección de la región $B = (-\infty, x] \times (-\infty, y]$ (zona azul) con la zona morada o región donde $f_{XY}(x, y) \neq 0$ obtenemos el trapecio $T_{x,y} = \{(u, v) \in \mathbb{R}^2, 0 \leq v \leq y, v \leq u \leq x\}$. Ver figura adjunta.

Por tanto,

$$\begin{aligned}
 F_{XY}(x, y) &= \int_{v=0}^{v=y} \int_{u=v}^{u=x} f_{XY}(u, v) dv du = 2 \cdot \int_{v=0}^{v=y} e^{-v} \int_{u=v}^{u=x} e^{-u} du dv \\
 &= 2 \cdot \int_{v=0}^{v=y} e^{-v} \cdot [-e^{-u}]_{u=v}^{u=x} dv = 2 \cdot \int_{v=0}^{v=y} e^{-v} \cdot (e^{-v} - e^{-x}) dv \\
 &= 2 \cdot \int_{v=0}^{v=y} (e^{-2v} - e^{-v-x}) dv = 2 \cdot \left[-\frac{1}{2}e^{-2v} + e^{-v-x} \right]_{v=0}^{v=y} \\
 &= 2 \cdot \left(-\frac{1}{2} \cdot e^{-2y} + e^{-x-y} + \frac{1}{2} - e^{-x} \right) \\
 &= 1 - 2 \cdot e^{-x} - e^{-2y} + 2 \cdot e^{-x-y}.
 \end{aligned}$$



En resumen:

$$F_{XY}(x, y) = \begin{cases} 1 - 2 \cdot e^{-x} + e^{-2x}, & \text{si } x \geq 0, y \geq 0, x \leq y, \\ 1 - 2 \cdot e^{-x} - e^{-2y} + 2 \cdot e^{-x-y}, & \text{si } x \geq 0, y \geq 0, x \geq y, \\ 0, & \text{en caso contrario.} \end{cases}$$

Comprobemos a continuación que si derivamos dos veces la expresión de F_{XY} , primero respecto x y después respecto y , obtendremos la **función de densidad** f_{XY} .

Si derivamos respecto x obtenemos:

$$\frac{\partial F_{XY}(x, y)}{\partial x} = \begin{cases} 2 \cdot e^{-x} - 2 \cdot e^{-2x}, & \text{si } x \geq 0, y \geq 0, x \leq y, \\ 2 \cdot e^{-x} - 2 \cdot e^{-x-y}, & \text{si } x \geq 0, y \geq 0, x \geq y, \\ 0, & \text{en caso contrario.} \end{cases}$$

Si ahora derivamos respecto y obtenemos:

$$\frac{\partial^2 F_{XY}(x, y)}{\partial y \partial x} = \begin{cases} 0, & \text{si } x \geq 0, y \geq 0, x \leq y, \\ 2 \cdot e^{-x-y}, & \text{si } x \geq 0, y \geq 0, x \geq y, \\ 0, & \text{en caso contrario.} \end{cases}$$

expresión que coincide con la **función de densidad** $f_{XY}(x, y)$.

Hallemos las **funciones de densidad marginales**. Fijémonos que basta tener en cuenta los casos en que $x \geq 0$ e $y \geq 0$ ya que en caso contrario tanto $f_X(x)$ como $f_Y(y)$ son nulas.

$$\begin{aligned} f_X(x) &= \int_{-\infty}^{\infty} f_{XY}(x, y) dy = \int_{y=0}^{y=x} 2 \cdot e^{-x-y} dy = 2 \cdot [-e^{-x-y}]_{y=0}^{y=x} \\ &= 2(e^{-x} - e^{-2x}), \text{ si } x \geq 0, \end{aligned}$$

$$\begin{aligned} f_Y(y) &= \int_{-\infty}^{\infty} f_{XY}(x, y) dx = \int_{x=y}^{x=\infty} 2 \cdot e^{-x-y} dx = 2 \cdot [-e^{-x-y}]_{x=y}^{x=\infty} \\ &= 2 \cdot e^{-2y}, \text{ si } y \geq 0. \end{aligned}$$

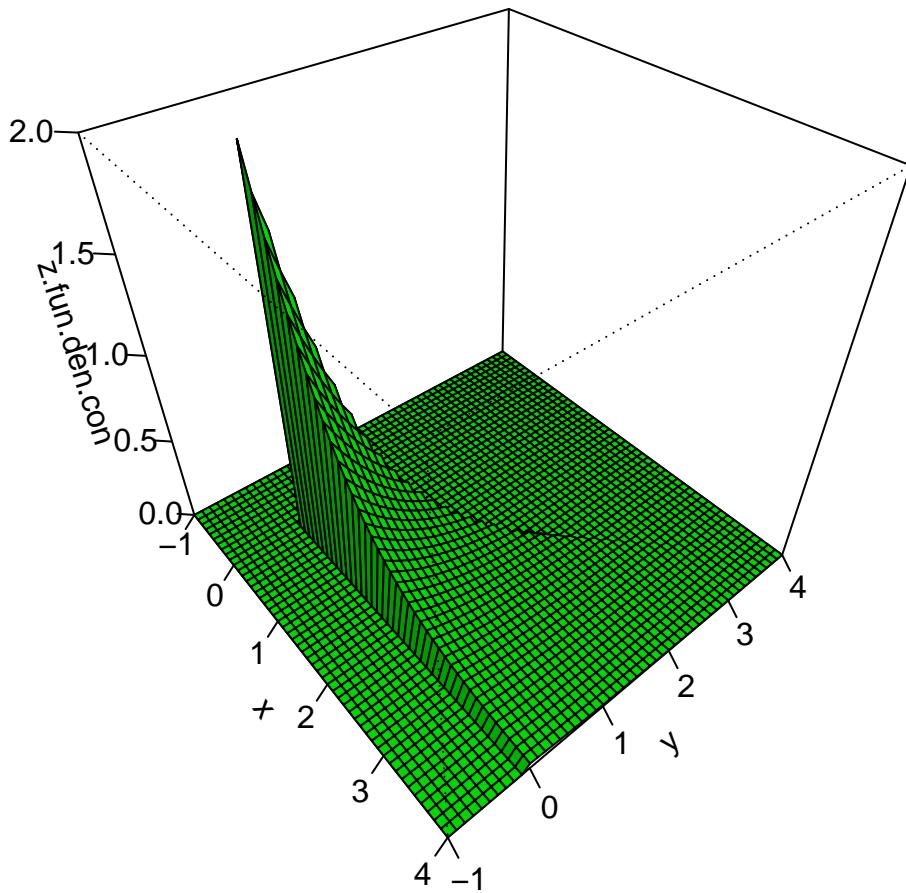
Vemos que la variable Y corresponde a una distribución exponencial de parámetro $\lambda = 2$.

Dibujemos la **función de densidad conjunta** y la **función de distribución conjunta** con R. Primero las definimos:

```
fun.den.con = function(x,y){ifelse(x>=0 & y>=0 & x>=y,
                                    2*exp(-x-y),0)}
fun.dist.con = function(x,y){ifelse(x>=0 & y>=0 & x<=y,
                                    1-2*exp(-x)+exp(-2*x),ifelse(x>=0 & y>=0 & x>=y,
                                    1-2*exp(-x)-exp(-2*y)+2*exp(-x-y),0))}
```

A continuación las dibujamos para x e y entre -1 y 4 :

```
x=seq(from=-1,to=4,by=0.1)
y=seq(from=-1,to=4,by=0.1)
z.fun.den.con=outer(x,y,fun.den.con)
z.fun.dist.con = outer(x,y,fun.dist.con)
persp(x,y,z.fun.den.con,theta=50,phi=40,col="green",shade=0.25,ticktype="detailed")
```



5.4.3. La distribución gaussiana bidimensional

Vamos a generalizar la distribución normal a dos dimensiones.

Definición de distribución gaussiana bidimensional. Diremos que la distribución de la variable aleatoria bidimensional (X, Y) es **gaussiana bidimensional** dependiendo del parámetro ρ si su **función de densidad conjunta** es:

$$f_{XY}(x, y) = \frac{1}{2 \cdot \pi \cdot \sqrt{1 - \rho^2}} e^{-\frac{(x^2 - 2 \cdot \rho \cdot x \cdot y + y^2)}{2 \cdot (1 - \rho^2)}}, \quad -\infty < x, y < \infty.$$

Propiedades de la **función de densidad de la variable gaussiana bidimensional**:

- Para cualquier punto $(x, y) \in \mathbb{R}^2$, la **función de densidad** es no nula: $f_{XY}(x, y) > 0$.
- La **función de densidad** tiene un único máximo absoluto en el punto $(0, 0)$ que vale $f_{XY}(0, 0) = \frac{1}{2\pi\sqrt{1-\rho^2}}$. Por tanto, para $\rho = 0$, dicho máximo alcanza el mínimo valor posible y si $\rho \rightarrow \pm 1$, dicho máximo tiende a ∞ .
- Las densidades marginales $f_X(x)$ y $f_Y(y)$ son normales $N(0, 1)$.

Veámoslo con $f_X(x)$. Por simetría, quedaría deducido para $f_Y(y)$:

$$\begin{aligned} f_X(x) &= \frac{1}{2\pi\sqrt{1-\rho^2}} \int_{-\infty}^{\infty} e^{-\frac{(x^2-2\rho x \cdot y+y^2)}{2(1-\rho^2)}} dy = \frac{1}{2\pi\sqrt{1-\rho^2}} e^{-\frac{x^2}{2(1-\rho^2)}} \int_{-\infty}^{\infty} e^{-\frac{(y-\rho x)^2}{2(1-\rho^2)}} dy \\ &= \frac{1}{2\pi\sqrt{1-\rho^2}} e^{-\frac{x^2}{2(1-\rho^2)}} \int_{-\infty}^{\infty} e^{-\frac{(y-\rho x)^2}{2(1-\rho^2)}} e^{\frac{\rho^2 \cdot y^2}{2(1-\rho^2)}} dy \\ &= \frac{1}{2\pi\sqrt{1-\rho^2}} e^{-\frac{x^2}{2}} \int_{-\infty}^{\infty} e^{-\frac{(y-\rho x)^2}{2(1-\rho^2)}} dy, \text{ hacemos el cambio } z = \frac{y-\rho x}{\sqrt{1-\rho^2}} \\ &= \frac{1}{2\pi\sqrt{1-\rho^2}} e^{-\frac{x^2}{2}} \int_{-\infty}^{\infty} e^{-\frac{z^2}{2}} \sqrt{1-\rho^2} dy = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}, \end{aligned}$$

función que coincide con la **función de densidad** de la variable $N(0, 1)$.

En el último paso hemos usado que

$$\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{z^2}{2}} dz = 1,$$

ya que correspondería al área de una **función de densidad** de una distribución $N(0, 1)$.

5.4.3.1. La distribución gaussiana bidimensional en R

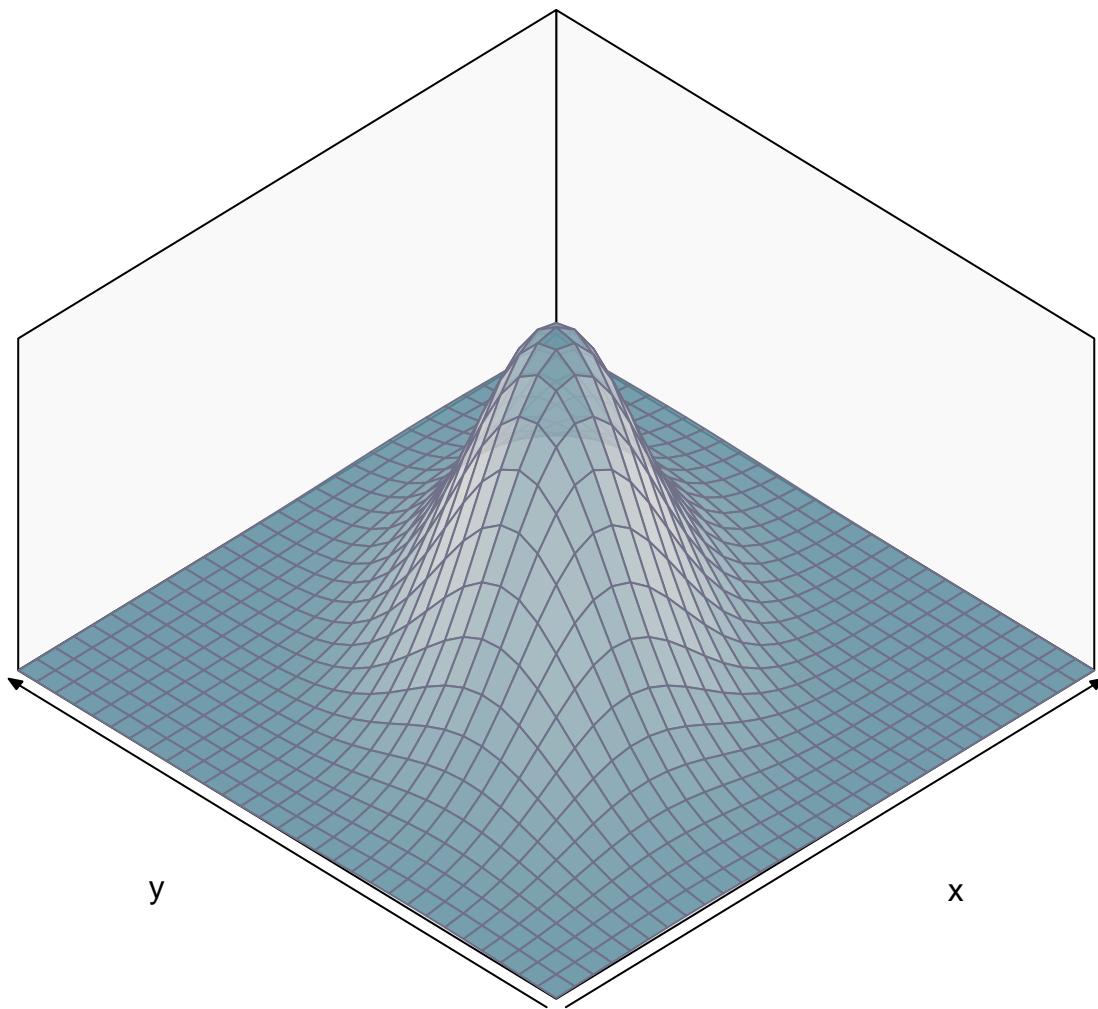
En R existe el paquete **bivariate** para trabajar con algunas distribuciones conjuntas; en particular, con la **distribución normal bidimensional**.

La función que nos la densidad de la **distribución normal bidimensional** es **nbvpdf** y tiene 5 parámetros: la **media** de X (μ_X), la **media** de Y (μ_Y), la **desviación típica** de X (σ_X), la **desviación típica** de Y (σ_Y) y un concepto que veremos más adelante, la **correlación** entre X e Y (ρ_{XY}).

En el ejemplo que estamos tratando, los valores de los parámetros anteriores son: $\mu_X = \mu_Y = 0$, $\sigma_X = \sigma_Y = 1$ y $\rho_{XY} = \rho$.

Vamos a hacer un gráfico de la **distribución normal bidimensional** para $\rho = \frac{1}{2}$.

```
library(bivariate)
f = nbvpdf (0, 0, 1, 1, 0.5)
plot(f,TRUE)
```



5.5. Independencia de variables aleatorias

5.5.1. Independencia de variables aleatorias discretas

Recordemos que dos sucesos A y B son independientes si $P(A \cap B) = P(A) \cdot P(B)$.

¿Cómo trasladar dicho concepto al caso de variables aleatorias?

En el caso de **variables aleatorias discretas bidimensionales** vimos que, dada una variable aleatoria bidimensional discreta (X, Y) con $(X, Y)(\Omega) = \{(x_i, y_j), i = 1, 2, \dots, j = 1, 2, \dots\}$, los sucesos de la forma $\{X = x_i, Y = y_j\}$ determinaban cómo se distribuían los valores de la variable (X, Y) . De ahí la definición siguiente:

Definición de independencia para variables aleatorias bidimensionales discretas. Sean (X, Y) una **variable aleatoria bidimensional discreta** con $(X, Y)(\Omega) = \{(x_i, y_j), i = 1, 2, \dots, j = 1, 2, \dots\}$ y

función de probabilidad P_{XY} y **funciones de probabilidad marginales** P_X y P_Y . Entonces X e Y son independientes si:

$$P_{XY}(x_i, y_j) = P_X(x_i) \cdot P_Y(y_j), \quad i = 1, 2, \dots, j = 1, 2, \dots$$

o dicho de otra forma:

$$P(X = x_i, Y = y_k) = P(X = x_i) \cdot P(Y = y_k), \quad i = 1, 2, \dots, j = 1, 2, \dots$$

Ejemplo: suma y el producto de los resultados de dos lanzamientos de un dado

Consideramos la variable aleatoria (S, P) donde S representa la suma de los valores obtenidos al lanzar dos veces un dado y P , su producto.

En este caso S y P no son independientes ya que recordemos que por ejemplo $P_{SP}(3, 2) = \frac{2}{36}$, $P_S(3) = \frac{2}{36}$ y $P_P(2) = \frac{2}{36}$, ya que en este último caso, sólo hay dos posibles resultados en los que el producto dé 2: el $(1, 2)$ y el $(2, 1)$.

Entonces no se cumple que $P_{SP}(3, 2) = P_S(3) \cdot P_P(2)$, ya que $\frac{2}{36} \neq \frac{2}{36} \cdot \frac{2}{36}$.

De ahí que no sean independientes ya que la condición anterior se debería cumplir para todos los valores x_i e y_k y hemos encontrado un contraejemplo en donde no se cumple.

Observación. Si la tabla de la **función de probabilidad conjunta** de (X, Y) contiene algún 0, X e Y no pueden ser independientes. ¿Podéis decir por qué?

Ejemplo: un caso de imdependencia

Veamos un caso de independencia. Consideramos el experimento aleatorio de lanzar un dado dos veces. Sea X el resultado del primer lanzamiento e Y , el resultado del segundo lanzamiento.

Veamos que, en este caso, X e Y son independientes.

El valor de $(X, Y)(\Omega) = \{(1, 1), (1, 2), \dots, (6, 6)\}$, en total 36 resultados.

La **función de probabilidad conjunta** en un valor cualquiera (i, j) con $i, j \in \{1, 2, 3, 4, 5, 6\}$ es: $P_{XY}(i, j) = \frac{1}{36}$ ya que la probabilidad que salga i en el primer lanzamiento es $\frac{1}{6}$ y la probabilidad de que salga j en el segundo lanzamiento, también. Por tanto, la probabilidad de que salga i en el primer lanzamiento y j en el segundo es: $\frac{1}{6} \cdot \frac{1}{6} = \frac{1}{36}$.

Las **funciones de densidad marginales** de X e Y son:

X o Y	1	2	3	4	5	6
P_X o P_Y	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$

Por tanto, para todo (i, j) con $i, j \in \{1, 2, 3, 4, 5, 6\}$ se cumplirá:

$$P_{XY}(i, j) = \frac{1}{36} = \frac{1}{6} \cdot \frac{1}{6} = P_X(i) \cdot P_Y(j).$$

Deduzcamos que son independientes.

Para comprobar si dos variables aleatorias X e Y son independientes o no en R en general, una vez calculada la tabla de la **función de probabilidad**, podemos calcular la tabla de **independencia**

teórica $P_T(x_i, y_j)$ y compararlas. Ésta segunda tabla se define de la forma siguiente:

$$P_T(x_i, y_j) = P_X(x_i) \cdot P_Y(y_j),$$

donde P_X y P_Y son las distribuciones marginales.

La tabla de **independencia teórica** en el caso de la suma y el producto se calcularían de la forma siguiente:

```
tabla.ind.teor = marginal.suma%*%t(marginal.producto)
tabla.ind.teor = as.data.frame(tabla.ind.teor)
rownames(tabla.ind.teor)=rownames(tabla.func.prob.conjunta)
colnames(tabla.ind.teor)=colnames(tabla.func.prob.conjunta)
```

Si comparamos los resultados de la tabla de **independencia teórica** mostrada a continuación con los resultados de la tabla de la **función de probabilidad conjunta**, veréis que no son iguales. Por tanto, S y P no son **independientes**.

	1	2	3	4	5	6	8	9	10
2	0.001	0.002	0.002	0.002	0.002	0.003	0.002	0.001	0.002
3	0.002	0.003	0.003	0.005	0.003	0.006	0.003	0.002	0.003
4	0.002	0.005	0.005	0.007	0.005	0.009	0.005	0.002	0.005
5	0.003	0.006	0.006	0.009	0.006	0.012	0.006	0.003	0.006
6	0.004	0.008	0.008	0.012	0.008	0.015	0.008	0.004	0.008
7	0.005	0.009	0.009	0.014	0.009	0.019	0.009	0.005	0.009
8	0.004	0.008	0.008	0.012	0.008	0.015	0.008	0.004	0.008
9	0.003	0.006	0.006	0.009	0.006	0.012	0.006	0.003	0.006
10	0.002	0.005	0.005	0.007	0.005	0.009	0.005	0.002	0.005
11	0.002	0.003	0.003	0.005	0.003	0.006	0.003	0.002	0.003
12	0.001	0.002	0.002	0.002	0.002	0.003	0.002	0.001	0.002

	12	15	16	18	20	24	25	30	36
2	0.003	0.002	0.001	0.002	0.002	0.002	0.001	0.002	0.001
3	0.006	0.003	0.002	0.003	0.003	0.003	0.002	0.003	0.002
4	0.009	0.005	0.002	0.005	0.005	0.005	0.002	0.005	0.002
5	0.012	0.006	0.003	0.006	0.006	0.006	0.003	0.006	0.003
6	0.015	0.008	0.004	0.008	0.008	0.008	0.004	0.008	0.004
7	0.019	0.009	0.005	0.009	0.009	0.009	0.005	0.009	0.005
8	0.015	0.008	0.004	0.008	0.008	0.008	0.004	0.008	0.004
9	0.012	0.006	0.003	0.006	0.006	0.006	0.003	0.006	0.003
10	0.009	0.005	0.002	0.005	0.005	0.005	0.002	0.005	0.002
11	0.006	0.003	0.002	0.003	0.003	0.003	0.002	0.003	0.002
12	0.003	0.002	0.001	0.002	0.002	0.002	0.001	0.002	0.001

lo cual finaliza nuestros cálculos en R.

5.5.2. Independencia de variables aleatorias continuas

La definición dada para **variables aleatorias discretas** se traslada de forma natural a las **variables aleatorias continuas**:

Definición de independencia para variables aleatorias bidimensionales continuas. Sean (X, Y) una **variable aleatoria bidimensional continua** con **función de densidad conjunta** f_{XY} y **funciones de densidad marginales** f_X y f_Y . Entonces X e Y son independientes si:

$$f_{XY}(x, y) = f_X(x) \cdot f_Y(y), \text{ para todo } x, y \in \mathbb{R}.$$

5.5.2.1. Ejemplos

Ejemplo: densidad uniforme en el cuadrado unidad

Recordemos el ejemplo siguiente visto donde teníamos una **variable aleatoria bidimensional continua** (X, Y) con **función de densidad conjunta**:

$$f_{XY}(x, y) = \begin{cases} 1, & \text{si } 0 \leq x \leq 1, 0 \leq y \leq 1, \\ 0, & \text{en caso contrario.} \end{cases}$$

y con densidad marginales:

$$f_X(x) = \begin{cases} 1, & \text{si } 0 \leq x \leq 1, \\ 0, & \text{en caso contrario.} \end{cases} \quad f_Y(y) = \begin{cases} 1, & \text{si } 0 \leq y \leq 1, \\ 0, & \text{en caso contrario.} \end{cases}$$

Veamos que son independientes.

Consideremos dos casos:

- $(x, y) \in [0, 1] \times [0, 1]$. En este caso:

$$f_{XY}(x, y) = 1 = 1 \cdot 1 = f_X(x) \cdot f_Y(y).$$

- $(x, y) \notin [0, 1] \times [0, 1]$. En este caso:

$$f_{XY}(x, y) = 0 = f_X(x) \cdot f_Y(y),$$

ya que si $(x, y) \notin [0, 1] \times [0, 1]$, o $x \notin [0, 1]$ o $y \notin [0, 1]$. Por tanto $f_X(x) = 0$ o $f_Y(y) = 0$. En cualquier caso, $f_X(x) \cdot f_Y(y) = 0$.

Ejemplo: otra función de densidad (continuación)

Recordemos el ejemplo siguiente visto donde teníamos una **variable aleatoria bidimensional continua** (X, Y) con **función de densidad conjunta**:

$$f_{XY}(x, y) = \begin{cases} 2 \cdot e^{-x} \cdot e^{-y}, & 0 \leq y \leq x < \infty, \\ 0, & \text{en caso contrario,} \end{cases}$$

y con densidad marginales:

$$f_X(x) = 2(e^{-x} - e^{-2x}), \text{ si } x \geq 0, \quad f_Y(y) = 2e^{-2y}, \text{ si } y \geq 0.$$

En este caso no son independientes ya que claramente $f_{XY}(x, y) \neq f_X(x) \cdot f_Y(y)$.

En este caso, recordemos que la **función de densidad conjunta** de (X, Y) es:

$$f_{XY}(x, y) = \frac{1}{2 \cdot \pi \cdot \sqrt{1 - \rho^2}} e^{-\frac{(x^2 - 2 \cdot \rho \cdot x \cdot y + y^2)}{2(1 - \rho^2)}}, \quad -\infty < x, y < \infty.$$

Las **funciones de densidad marginales** de X e Y correspondían a $N(0, 1)$:

$$\begin{aligned} f_X(x) &= \frac{1}{\sqrt{2\pi}} \cdot e^{-\frac{x^2}{2}}, \quad -\infty < x < \infty, \\ f_Y(y) &= \frac{1}{\sqrt{2\pi}} e^{-\frac{y^2}{2}}, \quad -\infty < y < \infty. \end{aligned}$$

¿Para qué valor(es) de ρ las variables normales estándar X e Y son independientes?

o, ¿para qué valor(es) de ρ se cumple?

$$f_X(x) \cdot f_Y(y) = \frac{1}{2 \cdot \pi} e^{-\frac{x^2+y^2}{2}} = \frac{1}{2 \cdot \pi \cdot \sqrt{1 - \rho^2}} e^{-\frac{(x^2 - 2 \cdot \rho \cdot x \cdot y + y^2)}{2(1 - \rho^2)}}.$$

La respuesta es claramente para $\rho = 0$.

Por tanto, ρ se puede interpretar como un parámetro de independencia, cuánto más cercano a cero esté, más cerca de la independencia estarán las variables X e Y .

5.5.3. Relación de la independencia y la función de distribución

El siguiente resultado nos da la relación entre la **independencia de variables aleatorias** y su **función de distribución conjunta**:

Teorema. Sea (X, Y) una variable aleatoria bidimensional. Entonces X e Y son independientes si, y sólo si, la **función de distribución conjunta** es el producto de las **funciones de distribución marginales** en todo valor $(x, y) \in \mathbb{R}^2$:

$$F_{XY}(x, y) = F_X(x) \cdot F_Y(y), \quad (x, y) \in \mathbb{R}^2.$$

Ejemplo

Consideramos el experimento aleatorio de lanzar un dado dos veces. Sea X el resultado del primer lanzamiento e Y , el resultado del segundo lanzamiento.

Recordemos que, en este caso, X e Y son independientes.

En primer lugar notemos que si $x < 1$ o $y < 1$, $F_{XY}(x, y) = 0$ ya que el suceso $\{X \leq x, Y \leq y\}$ es vacío.

De la misma forma como $x < 1$ o $y < 1$, o el suceso $\{X \leq x\}$ o el suceso $\{Y \leq y\}$ son vacíos. Por tanto, o $F_X(x) = 0$ o $F_Y(y) = 0$.

En cualquier caso, se cumple $F_{XY}(x, y) = 0 = F_X(x) \cdot F_Y(y)$.

Podemos suponer, por tanto, que $x \geq 1$ e $y \geq 1$.

Sea $(x, y) \in \mathbb{R}^2$ con $x \geq 1$ e $y \geq 1$. Podemos suponer tal que existen dos valores i y j en $\{1, 2, \dots\}$ con $i \leq x < i+1$ y $j \leq y < j+1$.

El valor de la **función de distribución conjunta** en (x, y) es:

$$F_{XY}(x, y) = \begin{cases} \frac{i \cdot j}{36}, & \text{si } i \leq 6, j \leq 6, \\ \frac{6 \cdot i}{36}, & \text{si } i \leq 6, j \geq 6, \\ \frac{6 \cdot j}{36}, & \text{si } i \geq 6, j \leq 6, \\ 1, & \text{si } i \geq 6, j \geq 6, \end{cases}$$

ya que:

$$\begin{aligned} F_{XY}(x, y) &= P(X \leq i, Y \leq j) = P(\{(k, l) \in \{1, 2, 3, 4, 5, 6\}^2, | k \leq i, l \leq j\}) \\ &= P(\{(1, 1), \dots, (1, j), \dots, (i, 1), \dots, (i, j)\}) \\ &= \begin{cases} \frac{i \cdot j}{36}, & \text{si } i \leq 6, j \leq 6, \\ \frac{6 \cdot i}{36}, & \text{si } i \leq 6, j \geq 6, \\ \frac{6 \cdot j}{36}, & \text{si } i \geq 6, j \leq 6, \\ 1, & \text{si } i \geq 6, j \geq 6, \end{cases} \end{aligned}$$

ya que claramente el cardinal del conjunto

$$\{(1, 1), \dots, (1, j), \dots, (i, 1), \dots, (i, j)\}$$

es

$$\begin{cases} i \cdot j, & \text{si } i \leq 6, j \leq 6, \\ \cdot i, & \text{si } i \leq 6, j \geq 6, \\ 6 \cdot j, & \text{si } i \geq 6, j \leq 6, \\ 36, & \text{si } i \geq 6, j \geq 6. \end{cases}$$

Hallaremos ahora la función de distribución de X e Y que consiste en el resultado del lanzamiento de un dado.

Dado $x \in \mathbb{R}$ con $x \geq 1$, existe un i con $i \in \{1, 2, \dots, 6\}$ con $i \leq x < i + 1$. En este caso, el valor de $F_X(x)$ es:

$$F_X(x) = \begin{cases} \frac{i}{6}, & \text{si } i \leq 6, \\ 1, & \text{si } i \geq 6, \end{cases}$$

ya que:

$$\begin{aligned} F_X(x) &= F_X(i) = P(X \leq i) = P(\{k \in \{1, 2, 3, 4, 5, 6\}, | k \leq i\}) \\ &= \begin{cases} \frac{i}{6}, & \text{si } i \leq 6, \\ 1, & \text{si } i \geq 6, \end{cases} \end{aligned}$$

\begin{ ya que el cardinal del conjunto $\{k \in \{1, 2, 3, 4, 5, 6\}, | k \leq i\}$ es $\begin{cases} i, & \text{si } i \leq 6, \\ 6, & \text{si } i \geq 6. \end{cases}$

La función de distribución de Y es de la misma forma.

Por último, comprobemos que se verifica que $F_{XY}(x, y) = F_X(x) \cdot F_Y(y)$, si $x \geq 1$ e $y \geq 1$.

Sea $(x, y) \in \mathbb{R}^2$ y sean los enteros i y j tales que $i \leq x < i+1$ y $j \leq y < j+1$. Consideremos 4 casos:

- $i \leq 6, j \leq 6$. En este caso:

$$F_{XY}(x, y) = \frac{i \cdot j}{36} = \frac{i}{6} \cdot \frac{j}{6} = F_X(x) \cdot F_Y(y).$$

- $i \leq 6, j \geq 6$. En este caso:

$$F_{XY}(x, y) = \frac{6i}{36} = \frac{i}{6} \cdot 1 = F_X(x) \cdot F_Y(y).$$

- $i \geq 6, j \leq 6$. En este caso:

$$F_{XY}(x, y) = \frac{6j}{36} = 1 \cdot \frac{j}{6} = F_X(x) \cdot F_Y(y).$$

- $i \geq 6, j \geq 6$. En este caso:

$$F_{XY}(x, y) = 1 = 1 \cdot 1 = F_X(x) \cdot F_Y(y).$$

En resumen, para todo $(x, y) \in \mathbb{R}^2$ se verifica que $F_{XY}(x, y) = F_X(x) \cdot F_Y(y)$, tal como queríamos ver.

Ejemplo

Recordemos la variable aleatoria bidimensional continua con **función de densidad conjunta**:

$$f_{XY}(x, y) = \begin{cases} 1, & \text{si } 0 \leq x \leq 1, 0 \leq y \leq 1, \\ 0, & \text{en caso contrario.} \end{cases}$$

Su **función de distribución conjunta** es:

$$F_{XY}(x, y) = \begin{cases} 0, & \text{si } x < 0, \text{ o } y < 0, \\ x \cdot y, & \text{si } 0 \leq x \leq 1, 0 \leq y \leq 1, \\ x, & \text{si } 0 \leq x \leq 1, y > 1, \\ y, & \text{si } 0 \leq y \leq 1, x > 1, \\ 1, & x \geq 1, y \geq 1. \end{cases}$$

Recordemos también que las **distribuciones marginales** de X e Y eran uniformes en el intervalo $[0, 1]$. Por tanto, las **funciones de distribución marginales** son:

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

Recordemos que X e Y son independientes. Verifiquemos que $F_{XY}(x, y) = F_X(x) \cdot F_Y(y)$.

Distinguiremos cinco casos:

- $x < 0$ o $y < 0$. En este caso, $F_{XY}(x, y) = 0$ y, o $F_X(x) = 0$, si $x < 0$, o $F_Y(y) = 0$, si $y < 0$. En cualquier caso, se cumple que $F_{XY}(x, y) = F_X(x) \cdot F_Y(y)$.
- $0 \leq x \leq 1$, $0 \leq y \leq 1$. En este caso, $F_{XY}(x, y) = xy$, $F_X(x) = x$ y $F_Y(y) = y$. Claramente, se cumple que $F_{XY}(x, y) = F_X(x) \cdot F_Y(y)$.
- $0 \leq x \leq 1$, $y > 1$. En este caso, $F_{XY}(x, y) = x$, $F_X(x) = x$ y $F_Y(y) = 1$. Claramente, se cumple que $F_{XY}(x, y) = F_X(x) \cdot F_Y(y)$.
- $x > 1$, $0 \leq y \leq 1$. En este caso, $F_{XY}(x, y) = y$, $F_X(x) = 1$ y $F_Y(y) = y$. Claramente, se cumple que $F_{XY}(x, y) = F_X(x) \cdot F_Y(y)$.
- $x \geq 1$, $y \geq 1$. En este caso, $F_{XY}(x, y) = 1$, $F_X(x) = 1$ y $F_Y(y) = 1$. Claramente, se cumple que $F_{XY}(x, y) = F_X(x) \cdot F_Y(y)$.

5.6. Momentos conjuntos y valores esperados conjuntos

El **valor esperado** de una variable aleatoria X se identifica con el *centro de masa de la distribución de X* .

La **varianza** proporciona una medida de la *extensión de la distribución*.

En el caso de dos variables aleatorias, estamos interesados en cómo X e Y varían juntas.

En particular, nos interesa saber si la variación de X e Y está correlacionada. Por ejemplo, si X aumenta, ¿ Y tiende a aumentar o disminuir?

Los momentos conjuntos de X e Y , que se definen como valores esperados de las funciones de X e Y , proporcionan esta información.

5.6.1. Valor esperado de una función de dos variables aleatorias

Sea (X, Y) una variable aleatoria bidimensional.

Sea P_{XY} su **función de probabilidad conjunta** en el caso en que (X, Y) sea **discreta** y f_{XY} su **función de densidad conjunta** en el caso en que (X, Y) sea **continua**.

Sea $Z = g(X, Y)$ una **variable aleatoria unidimensional** función de las variables X e Y . Por ejemplo:

- Suma de las dos variables $g(x, y) = x + y$: $Z = X + Y$.
- Producto de las dos variables $g(x, y) = x \cdot y$: $Z = X \cdot Y$.
- Suma de los cuadrados de las variables $g(x, y) = x^2 + y^2$: $Z = X^2 + Y^2$.

Hay que tener en cuenta que Z , como **variable aleatoria unidimensional** tiene una **función de probabilidad** P_Z en el caso en que (X, Y) sea discreta y una **función de densidad** f_Z en el caso en que (X, Y) sea continua.

El siguiente resultado nos dice cómo calcular el **valor esperado** de Z sin tener que calcular P_Z o f_Z , sólo usando la información de la **variable aleatoria conjunta** (X, Y) :

Proposición. El valor esperado de Z se puede hallar usando la expresión siguiente:

- en el caso en que (X, Y) sea discreta con $(X, Y)(\Omega) = \{(x_i, y_j), i = 1, 2, \dots, j = 1, 2, \dots\}$,

$$E(Z) = E(g(X, Y)) = \sum_{x_i} \sum_{y_j} g(x_i, y_j) \cdot P(x_i, y_j),$$

- en el caso en que (X, Y) sea continua:

$$E(Z) = E(g(X, Y)) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(x, y) \cdot f_{XY}(x, y) dx dy.$$

5.6.1.1. Ejemplos

Ejemplo: suma y producto de dos dados (continuación)

Consideremos el ejemplo de la variable (S, P) que nos daba la suma y el producto de los resultados cuando lanzábamos dos veces un dado.

Vamos a calcular $E(S + P)$.

Recordemos que ya hemos calculado P_{SP} . La expresión de $E(S + P)$ es:

$$\begin{aligned} E(S + P) &= (2+1) \cdot P_{SP}(2, 1) + (3+2) \cdot P_{SP}(3, 2) + (4+3) \cdot P_{SP}(4, 3) \\ &\quad + (4+4) \cdot P_{SP}(4, 4) + (5+4) \cdot P_{SP}(5, 4) + (5+6) \cdot P_{SP}(5, 6) \\ &\quad + (6+5) \cdot P_{SP}(6, 5) + (6+8) \cdot P_{SP}(6, 8) + (6+9) \cdot P_{SP}(6, 9) \\ &\quad + (7+6) \cdot P_{SP}(7, 6) + (7+10) \cdot P_{SP}(7, 10) + (7+12) \cdot P_{SP}(7, 12) \\ &\quad + (8+12) \cdot P_{SP}(8, 12) + (8+15) \cdot P_{SP}(8, 15) + (8+16) \cdot P_{SP}(8, 16) \\ &\quad + (9+18) \cdot P_{SP}(9, 18) + (9+20) \cdot P_{SP}(9, 20) \\ &\quad + (10+24) \cdot P_{SP}(10, 24) + (10+25) \cdot P_{SP}(10, 25) \\ &\quad + (11+30) \cdot P_{SP}(11, 30) + (12+36) \cdot P_{SP}(12, 36) \\ &= 3 \cdot \frac{1}{36} + 5 \cdot \frac{2}{36} + 7 \cdot \frac{2}{36} + 8 \cdot \frac{1}{36} + 9 \cdot \frac{2}{36} + 11 \cdot \frac{2}{36} + 11 \cdot \frac{2}{36} + 14 \cdot \frac{2}{36} \\ &\quad + 15 \cdot \frac{1}{36} + 13 \cdot \frac{2}{36} + 17 \cdot \frac{2}{36} + 19 \cdot \frac{2}{36} + 20 \cdot \frac{2}{36} + 23 \cdot \frac{2}{36} + 24 \cdot \frac{1}{36} \\ &\quad + 27 \cdot \frac{2}{36} + 29 \cdot \frac{2}{36} + 34 \cdot \frac{2}{36} + 35 \cdot \frac{1}{36} + 41 \cdot \frac{2}{36} + 48 \cdot \frac{1}{36} \\ &= \frac{693}{36} = 19,25. \end{aligned}$$

Hallar el valor esperado de la suma $E(S + P)$ una vez hallada la tabla de la **función de probabilidad conjunta**, en R es bastante sencillo usando la función `outer`:

```
valores.suma = as.integer(rownames(tabla.func.prob.conjunta))
valores.producto = as.integer(colnames(tabla.func.prob.conjunta))
suma.valores = outer(valores.suma, valores.producto, "+")
(valor.esperado.suma = sum(suma.valores*tabla.func.prob.conjunta))
```

```
## [1] 19.25
```

Observación: En R para hallar el valor esperado de una función $g(X, Y)$, $E(g(X, Y))$ de las variables aleatorias X e Y , basta sustituir el valor `+` en el script anterior por `FUN=g`, definiendo previamente la función `g`.

Ejemplo: otra densidad (continuación)

Recordemos el ejemplo donde (X, Y) era una variable aleatoria bidimensional continua con **función de densidad conjunta**:

$$f_{XY}(x, y) = \begin{cases} 2 \cdot e^{-x} \cdot e^{-y}, & 0 \leq y \leq x < \infty, \\ 0, & \text{en caso contrario.} \end{cases}$$

Calculemos $E(X \cdot Y)$:

$$\begin{aligned} E(X \cdot Y) &= \int_{x=0}^{x=\infty} \int_{y=0}^{y=x} 2 \cdot x \cdot y e^{-x} \cdot e^{-y} dy dx \\ &= 2 \cdot \int_{x=0}^{x=\infty} x \cdot e^{-x} \cdot \int_{y=0}^{y=x} y \cdot e^{-y} dy dx \\ &= 2 \int_{x=0}^{x=\infty} x \cdot e^{-x} [-e^{-y} \cdot (y+1)]_{y=0}^{y=x} dx \\ &= 2 \cdot \int_{x=0}^{x=\infty} x \cdot e^{-x} \cdot (1 - e^{-x}(x+1)) dx \\ &= 2 \cdot \int_{x=0}^{x=\infty} x \cdot (e^{-x} - e^{-2x}) - x^2 \cdot e^{-2x} dx \\ &= 2 \cdot \left[-e^{-x}(x+1) + \frac{1}{4} \cdot e^{-2x}(1+2x) + \frac{1}{4} \cdot e^{-2x}(2x^2+2x+1) \right]_{x=0}^{x=\infty} \\ &= 2 \cdot \left(1 - \frac{1}{4} - \frac{1}{4} \right) = 1. \end{aligned}$$

En el último cálculo hemos usado integración por partes para integrar $\int xe^{-x} dx$, $\int xe^{-2x} dx$ y $\int x^2 e^{-2x} dx$.

Ejercicio

Hallar $E(X + Y)$ para el ejemplo anterior.

5.6.2. Valor esperado de una función de dos variables aleatorias independientes

El siguiente resultado nos simplifica el cálculo del **valor esperado de una función de dos variables aleatorias** en el caso en que sean **independientes**:

Proposición: cálculo del valor esperado de una función de dos variables aleatorias en el caso de independencia. Sea (X, Y) una variable aleatoria bidimensional tal que X e Y son independientes. Sea $Z = g(X, Y)$ una variable aleatoria unidimensional función de X e Y en la que podemos “separar” las variables x e y en la función g . Es decir, existen dos funciones g_x y g_y tal que $g(x, y) = g_x(x) \cdot g_y(y)$ para todo valor $x, y \in \mathbb{R}$. En este caso, el valor esperado de Z se puede calcular como:

$$E(Z) = E(g(X, Y)) = E_X(g_x(X)) \cdot E_Y(g_y(Y)).$$

Es decir, el cálculo de $E(g(X, Y))$ que es una suma doble en el caso de que (X, Y) sea **discreta** o una integral doble en el caso en que (X, Y) sea continua se transforma en el producto de dos sumas simples (caso **discreto**) o el producto de dos integrales simples (caso **continuo**):

$$\begin{aligned} E(Z) &= E(g(X, Y)) = \left(\sum_{x_i} g_x(x_i) \cdot P_X(x_i) \right) \cdot \left(\sum_{y_j} g_y(y_j) \cdot P_Y(y_j) \right), \\ &\quad \text{caso discreto,} \\ E(Z) &= E(g(X, Y)) = \left(\int_{-\infty}^{\infty} g_x(x) \cdot f_X(x) dx \right) \cdot \left(\int_{-\infty}^{\infty} g_y(y) \cdot f_Y(y) dy \right), \\ &\quad \text{caso continuo.} \end{aligned}$$

Un caso particular de aplicación de la proposición anterior es el cálculo de $E(X \cdot Y)$ cuando X e Y son independientes. En este caso $g(x, y) = x \cdot y$, $g_x(x) = x$, y $g_y(y) = y$.

Podemos escribir, por tanto:

$$E(X \cdot Y) = E_X(X) \cdot E_Y(Y).$$

5.6.2.1. Ejemplos

Ejemplo: lanzar dos veces un dado (continuación)

Recordemos el experimento aleatorio que consiste en lanzar un dado dos veces. Sea X el resultado del primer lanzamiento e Y , el resultado del segundo lanzamiento.

Hemos visto que X e Y son independientes.

Las marginales de X e Y recordemos que son las siguientes:

X o Y	1	2	3	4	5	6
$P_X(i)$ o $P_Y(i)$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$

Calculemos $E(X \cdot Y)$ usando la proposición anterior:

$$E(X \cdot Y) = E_X(X) \cdot E_Y(Y) = \left(\sum_{i=1}^6 i \cdot \frac{1}{6} \right) \cdot \left(\sum_{i=1}^6 i \cdot \frac{1}{6} \right) = \left(\frac{21}{6} \right)^2 = 12,25.$$

Dejamos como ejercicio el cálculo de $E(X \cdot Y)$ usando la **función de probabilidad conjunta** P_{XY} y comprobar que da el mismo resultado.

Ejemplo: densidad uniforme cuadrado unidad

Recordemos la variable aleatoria bidimensional continua con **función de densidad conjunta**:

$$f_{XY}(x, y) = \begin{cases} 1, & \text{si } 0 \leq x \leq 1, 0 \leq y \leq 1, \\ 0, & \text{en caso contrario.} \end{cases}$$

donde vimos que X e Y eran independientes y de distribución uniforme en el intervalo $[0, 1]$.

Calculemos $E(X \cdot Y)$ usando la proposición:

$$\begin{aligned} E(X \cdot Y) &= E_X(X) \cdot E_Y(Y) = \int_0^1 x \cdot 1 \, dx \cdot \int_0^1 y \cdot 1 \, dy \\ &= \left[\frac{x^2}{2} \right]_{x=0}^{x=1} \cdot \left[\frac{y^2}{2} \right]_{y=0}^{y=1} = \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{4}. \end{aligned}$$

Dejamos como ejercicio el cálculo de $E(X \cdot Y)$ usando la **función de densidad conjunta** f_{XY} y comprobar que da el mismo resultado.

5.6.3. Momentos conjuntos

A continuación vamos a definir el momento de orden (k, l) para una variable aleatoria bidimensional (X, Y) para intentar obtener información de su comportamiento conjunto:

Definición de momento conjunto. Sean (X, Y) una variable aleatoria bidimensional con **función de probabilidad conjunta** P_{XY} en el caso discreto y **función de densidad conjunta** f_{XY} en el caso continuo. Dados k y l números enteros positivos, definimos el **momento conjunto de orden** (k, l) para la variable (X, Y) como:

$$E(X^k Y^l) = \begin{cases} \sum_{x_i} \sum_{y_j} x_i^k y_j^l P_{XY}(x_i, y_j), & \text{caso discreto,} \\ \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x^k y^l f_{XY}(x, y) \, dx \, dy. & \text{caso continuo.} \end{cases}$$

Observación. Si consideramos $l = 0$, los momentos conjuntos de orden $(k, 0)$ coinciden con los momentos de orden k de la variable aleatoria X .

De la misma forma, considerando $k = 0$, los momentos conjuntos de orden $(0, l)$ coinciden con los momentos de orden l de la variable aleatoria Y .

Para $l = 1$ y $k = 1$ obtenemos el momento de orden $(1, 1)$ ya visto anteriormente: $E(X \cdot Y)$, denominado **correlación entre las variables** X e Y . Si dicha correlación es cero, $E(X \cdot Y) = 0$, se dice que las variables X e Y son **ortogonales**.

5.6.4. Momentos conjuntos centrados en las medias

A continuación definimos los **momentos conjuntos centrados en las medias**:

Definición de momento conjunto. Sean (X, Y) una variable aleatoria bidimensional con **función de probabilidad conjunta** P_{XY} en el caso discreto y **función de densidad conjunta** f_{XY} en el caso continuo. Sean $\mu_X = E(X)$ y $\mu_Y = E(Y)$ los **valores esperados** de las variables X e Y , respectivamente. Dados k y l números enteros positivos, definimos el **momento conjunto de orden** (k, l) **centrado en las medias** para la variable (X, Y) como:

$$E((X - \mu_X)^k \cdot (Y - \mu_Y)^l) = \begin{cases} \sum_{x_i} \sum_{y_j} (x_i - \mu_X)^k \cdot (y_j - \mu_Y)^l \cdot P_{XY}(x_i, y_j), & \text{caso discreto,} \\ \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (x - \mu_X)^k \cdot (y - \mu_Y)^l \cdot f_{XY}(x, y) \, dx \, dy. & \text{caso continuo.} \end{cases}$$

5.6.5. Covariancia entre las variables

El **momento conjunto centrado en las medias para $k = 1$ y $l = 1$** se denomina **covariancia** entre las variables X e Y :

$$\text{Cov}(X, Y) = E((X - \mu_X) \cdot (Y - \mu_Y)).$$

La covariancia puede calcularse a partir de la **correlación** entre las variables:

$$\text{Cov}(X, Y) = E((X - \mu_X) \cdot (Y - \mu_Y)) = E(X \cdot Y) - \mu_X \cdot \mu_Y,$$

ya que, usando las propiedades de la esperanza, tenemos:

$$\begin{aligned} E((X - \mu_X) \cdot (Y - \mu_Y)) &= E(X \cdot Y - \mu_Y \cdot X - \mu_X \cdot Y + \mu_X \cdot \mu_Y) \\ &= E(X \cdot Y) - \mu_Y \cdot E(X) - \mu_X \cdot E(Y) + \mu_X \cdot \mu_Y \\ &= E(X \cdot Y) - \mu_Y \cdot \mu_X - \mu_X \cdot \mu_Y + \mu_X \cdot \mu_Y \\ &= E(X \cdot Y) - \mu_X \cdot \mu_Y. \end{aligned}$$

Observación. Si las variables X e Y son **independientes**, su **covarianza** es nula ya que vimos que $E(X \cdot Y) = \mu_X \cdot \mu_Y$.

La **covarianza** es una medida de lo relacionadas están las variables X e Y :

- Si cuando $X \geq \mu_X$, también ocurre que $Y \geq \mu_Y$ o viceversa, cuando $X \leq \mu_X$, también ocurre que $Y \leq \mu_Y$, el valor $(X - \mu_X)(Y - \mu_Y)$ será positivo y la **covarianza** será positiva.
- Si por el contrario, cuando $X \geq \mu_X$, también ocurre que $Y \leq \mu_Y$ o viceversa, cuando $X \leq \mu_X$, también ocurre que $Y \geq \mu_Y$, el valor $(X - \mu_X)(Y - \mu_Y)$ será negativo y la **covarianza** será negativa.
- En cambio, si a veces ocurre una cosa y a veces ocurre otra, la **covarianza** va cambiando de signo y puede tener un valor cercano a 0.

5.6.5.1. Propiedades de la covarianza

- Sea (X, Y) una variable aleatoria bidimensional. Entonces la **varianza de la suma/resta** se calcula usando la expresión siguiente:

$$\text{Var}(X \pm Y) = \text{Var}(X) + \text{Var}(Y) \pm 2\text{Cov}(X, Y).$$

Demostración

La varianza de la suma/resta de las variables es, usando la propiedad de la **varianza**:

$$\text{Var}(X \pm Y) = E((X \pm Y)^2) - (E(X \pm Y))^2.$$

Desarrollando las expresiones anteriores, obtenemos:

$$\begin{aligned} \text{Var}(X \pm Y) &= E(X^2 + Y^2 \pm 2XY) - (E(X) \pm E(Y))^2 \\ &= E(X^2) + E(Y^2) \pm 2 \cdot E(X \cdot Y) \\ &\quad - (E(X)^2 + E(Y)^2 \pm 2 \cdot E(X) \cdot E(Y)) \\ &= E(X^2) - E(X)^2 + E(Y^2) - E(Y)^2 \\ &\quad \pm 2 \cdot (E(X \cdot Y) - E(X) \cdot E(Y)) \\ &= \text{Var}(X) + \text{Var}(Y) \pm 2 \cdot \text{Cov}(X, Y), \end{aligned}$$

tal como queríamos ver.

Una consecuencia de la propiedad anterior es el resultado siguiente:

Proposición: si las variables son independientes, la varianza de la suma es la suma de varianzas. Sea (X, Y) una variable aleatoria bidimensional donde las variables X e Y son **independientes**. Entonces:

$$\text{Var}(X + Y) = \text{Var}(X) + \text{Var}(Y).$$

Demostración

La demostración es muy sencilla: basta aplicar la fórmula de la varianza de la suma y tener en cuenta que, como X e Y son independientes, su covarianza es cero: $\text{Cov}(X, Y) = 0$.

5.6.6. Coeficiente de correlación entre las variables

La **covarianza** depende de las unidades en las que están las variables X e Y ya que si $a > 0$ y $b > 0$, entonces:

$$\text{Cov}(a \cdot X, b \cdot Y) = a \cdot b \cdot \text{Cov}(X, Y).$$

Por tanto, si queremos “medir” la relación que existe entre las variables X e Y tendremos que “normalizar” la **covarianza** definiendo el **coeficiente de correlación** entre las variables X e Y :

Definición del coeficiente de correlación. Sea (X, Y) una variable aleatoria bidimensional. Se define el **coeficiente de correlación** entre las variables X e Y como:

$$\rho_{XY} = \frac{\text{Cov}(X, Y)}{\sqrt{\text{Var}(X)} \cdot \sqrt{\text{Var}(Y)}} = \frac{E(X \cdot Y) - \mu_X \cdot \mu_Y}{\sqrt{E(X^2) - \mu_X^2} \cdot \sqrt{E(Y^2) - \mu_Y^2}}.$$

Observación. Si las variables X e Y son **independientes**, su **coeficiente de correlación** $\rho_{XY} = 0$ es nulo ya que su **covarianza** lo es.

Notemos también que la **correlación** no tiene unidades y es invariante a cambios de escala.

Además, la **covarianza** de las **variables tipificadas** $\frac{X-\mu_X}{\sigma_X}$ y $\frac{Y-\mu_Y}{\sigma_Y}$ coincide con la **correlación** de X e Y .

El **coeficiente de correlación** es un valor normalizado ya que siempre está entre -1 y 1: $-1 \leq \rho_{XY} \leq 1$.

Para demostrar de este hecho, sean $\mu_X = E(X)$, $\mu_Y = E(Y)$, $\sigma_X = \sqrt{\text{Var}(X)}$ y $\sigma_Y = \sqrt{\text{Var}(Y)}$.

Consideremos la variable $Z = \left(\frac{X-\mu_X}{\sigma_X} \pm \frac{Y-\mu_Y}{\sigma_Y} \right)^2$. Como $Z \geq 0$, tenemos que $E(Z) \geq 0$. Desarrollemos el valor de $E(Z)$:

$$\begin{aligned} E(Z) &= E \left(\frac{X-\mu_X}{\sigma_X} \pm \frac{Y-\mu_Y}{\sigma_Y} \right)^2 = E \left(\left(\frac{X-\mu_X}{\sigma_X} \right)^2 + \left(\frac{Y-\mu_Y}{\sigma_Y} \right)^2 \pm 2 \left(\frac{X-\mu_X}{\sigma_X} \right) \cdot \left(\frac{Y-\mu_Y}{\sigma_Y} \right) \right) \\ &= E \left(\left(\frac{X-\mu_X}{\sigma_X} \right)^2 \right) + E \left(\left(\frac{Y-\mu_Y}{\sigma_Y} \right)^2 \right) \pm 2 \cdot E \left(\left(\frac{X-\mu_X}{\sigma_X} \right) \cdot \left(\frac{Y-\mu_Y}{\sigma_Y} \right) \right) \\ &= \frac{1}{\sigma_X^2} E \left((X - \mu_X)^2 \right) + \frac{1}{\sigma_Y^2} E \left((Y - \mu_Y)^2 \right) \pm \frac{2}{\sigma_X \cdot \sigma_Y} E \left((X - \mu_X)(Y - \mu_Y) \right) \\ &= \frac{1}{\sigma_X^2} \sigma_X^2 + \frac{1}{\sigma_Y^2} \sigma_Y^2 \pm \frac{2}{\sigma_X \cdot \sigma_Y} \text{Cov}(X, Y) = 1 + 1 \pm 2 \cdot \frac{\text{Cov}(X, Y)}{\sigma_X \cdot \sigma_Y} = 2 \cdot (1 \pm \rho_{XY}) \end{aligned}$$

Ahora, como $E(Z) \geq 0$, tenemos que $1 \pm \rho_{XY} \geq 0$, lo que significa que, por un lado $1 + \rho_{XY} \geq 0$ y, por otro, $1 - \rho_{XY} \geq 0$. De la primera inecuación, deducimos que $\rho_{XY} \geq -1$ y de la segunda, $\rho_{XY} \leq 1$.

En resumen, $-1 \leq \rho_{XY} \leq 1$, tal como queríamos ver.

5.6.6.1. Ejemplos

Ejemplo: otra densidad (continuación)

Hallemos el **coeficiente de correlación** para el ejemplo de la variable aleatoria bidimensional continua con **función de densidad conjunta**:

$$f_{XY}(x, y) = \begin{cases} 2 \cdot e^{-x} \cdot e^{-y}, & 0 \leq y \leq x < \infty, \\ 0, & \text{en caso contrario,} \end{cases}$$

Recordemos los cálculos realizados anteriormente:

- $E(X \cdot Y) = 1$.
- $f_X(x) = 2 \cdot (e^{-x} - e^{-2x})$, si $x \geq 0$. Su esperanza es:

$$\begin{aligned} E(X) &= \int_0^{\infty} x \cdot 2 \cdot (e^{-x} - e^{-2x}) \, dx = 2 \left[\frac{1}{4} e^{-2x} (2x + 1) - e^{-x} (x + 1) \right]_0^{\infty} \\ &= 2 \left(1 - \frac{1}{4} \right) = \frac{3}{2}. \end{aligned}$$

Calculemos a continuación su varianza: $\text{Var}(X) = E(X^2) - \mu_X^2$. El valor de $E(X^2)$ es:

$$\begin{aligned} E(X^2) &= \int_0^{\infty} x^2 \cdot 2 \cdot (e^{-x} - e^{-2x}) \, dx \\ &= 2 \cdot \left[\frac{1}{4} \cdot e^{-2x} \cdot (2 \cdot x^2 + 2 \cdot x + 1) - e^{-x} \cdot (x^2 + 2 \cdot x + 2) \right]_0^{\infty} \\ &= 2 \cdot \left(2 - \frac{1}{4} \right) = \frac{7}{2}. \end{aligned}$$

El valor de la varianza de X es: $\text{Var}(X) = \frac{7}{2} - \left(\frac{3}{2}\right)^2 = \frac{5}{4}$.

- La variable Y era exponencial de parámetro $\lambda = 2$. Por tanto, $E(Y) = \frac{1}{2}$, $\text{Var}(Y) = \frac{1}{4}$.

El **coeficiente de correlación** entre las variables X e Y es:

$$\rho_{XY} = \frac{E(X \cdot Y) - \mu_X \cdot \mu_Y}{\sqrt{\text{Var}(X)} \cdot \sqrt{\text{Var}(Y)}} = \frac{1 - \frac{3}{2} \cdot \frac{1}{2}}{\sqrt{\frac{5}{4}} \cdot \sqrt{\frac{1}{4}}} = \frac{\sqrt{5}}{5} \approx 0,447.$$

Vemos que la **correlación** entre las variables X e Y es positiva pero no demasiado ya que su valor no está cercano a 1.

Ejemplo: normal bidimensional

Recordemos que la **función de densidad** de la variable aleatoria **normal bidimensional** es:

$$f_{XY}(x, y) = \frac{1}{2\pi\sqrt{1-\rho^2}} e^{-\frac{(x^2-2\rho xy+y^2)}{2(1-\rho^2)}}, \quad -\infty < x, y < \infty.$$

Las **variables aleatorias marginales** son normales estándar o $N(0, 1)$.

Hallaremos el **coeficiente de correlación** ρ_{XY} en este caso.

Calculemos $E(X \cdot Y)$:

$$\begin{aligned}
 E(X \cdot Y) &= \int_{-\infty}^{\infty} xy \frac{1}{2\pi\sqrt{1-\rho^2}} e^{-\frac{(x^2-2\rho x \cdot y+y^2)}{2(1-\rho^2)}} dy dx \\
 &= \frac{1}{2\pi\sqrt{1-\rho^2}} \int_{x=-\infty}^{x=\infty} x \cdot e^{-\frac{x^2}{2(1-\rho^2)}} \int_{y=-\infty}^{y=\infty} ye^{-\frac{(-2\rho x \cdot y+y^2)}{2(1-\rho^2)}} dy dx \\
 &= \frac{1}{2\pi\sqrt{1-\rho^2}} \int_{x=-\infty}^{x=\infty} xe^{-\frac{x^2}{2(1-\rho^2)}} e^{\frac{\rho^2 \cdot x^2}{2(1-\rho^2)}} \int_{y=-\infty}^{y=\infty} ye^{-\frac{(y-\rho x)^2}{2(1-\rho^2)}} dy dx, \\
 &\quad \text{cambio de variable en la segunda integral } z = \frac{y-\rho x}{\sqrt{1-\rho^2}}, \\
 &= \frac{1}{2\pi\sqrt{1-\rho^2}} \int_{x=-\infty}^{x=\infty} xe^{-\frac{x^2}{2}} \int_{z=-\infty}^{z=\infty} (z\sqrt{1-\rho^2} + \rho x) \sqrt{1-\rho^2} e^{-\frac{z^2}{2}} dz \\
 &= \frac{1}{2\pi} \int_{x=-\infty}^{x=\infty} xe^{-\frac{x^2}{2}} \left(\sqrt{1-\rho^2} \int_{z=-\infty}^{z=\infty} z \cdot e^{-\frac{z^2}{2}} dz + \rho \cdot x \int_{z=-\infty}^{z=\infty} e^{-\frac{z^2}{2}} dz \right) dx
 \end{aligned}$$

Ahora, usando que el valor esperado de una variable $N(0, 1)$ es cero tenemos que: $\int_{z=-\infty}^{z=\infty} ze^{-\frac{z^2}{2}} dz = 0$, y usando que la integral de la **función de densidad** de la $N(0, 1)$ ($\frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}}$) sobre todo \mathbb{R} es 1, tenemos que: $\int_{z=-\infty}^{z=\infty} e^{-\frac{z^2}{2}} dz = \sqrt{2\pi}$.

Por tanto,

$$E(X \cdot Y) = \frac{\rho}{2\pi} \int_{x=-\infty}^{x=\infty} x^2 e^{-\frac{x^2}{2}} \sqrt{2\pi} dx = \frac{\rho}{\sqrt{2\pi}} \int_{x=-\infty}^{x=\infty} x^2 e^{-\frac{x^2}{2}} dx.$$

Por último, usando que la varianza de la distribución $Z = N(0, 1)$ es 1, tenemos que $\text{Var}(Z) = E(Z^2) - E(Z)^2$. Como $E(Z) = 0$, deducimos que $E(Z^2) = 1$:

$$\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} x^2 e^{-\frac{x^2}{2}} dx = 1, \Rightarrow \int_{-\infty}^{\infty} x^2 e^{-\frac{x^2}{2}} dx = \sqrt{2\pi}.$$

El valor de $E(X \cdot Y)$ es:

$$E(X \cdot Y) = \frac{\rho}{\sqrt{2\pi}} \sqrt{2\pi} = \rho.$$

La correlación entre las variables X e Y es precisamente ρ .

Ahora, usando que $\mu_X = \mu_Y = 0$ y $\sigma_X = \sigma_Y = 1$ ya que recordemos que las marginales son $N(0, 1)$, el **coeficiente de correlación** entre las variables X e Y es:

$$\rho_{XY} = \frac{E(X \cdot Y) - \mu_X \cdot \mu_Y}{\sqrt{\text{Var}(X)} \cdot \sqrt{\text{Var}(Y)}} = \frac{\rho - 0 \cdot 0}{1 \cdot 1} = \rho.$$

Por tanto, ρ es el **coeficiente de correlación** entre las variables X e Y y mide lo correlacionadas que están dichas variables.

5.6.7. Incorrelación e independencia

Hemos visto que si dos variables X e Y son **independientes**, entonces son **incorreladas**, es decir, la **covarianza** es 0 ($E(X \cdot Y) = E(X) \cdot E(Y)$).

El recíproco, sin embargo, es falso. Veamos un ejemplo de variables **incorreladas** que no son independientes.

Ejemplo de variables aleatorias incorreladas pero no independientes

Consideremos la variable aleatoria bidimensional continua con **función de densidad**:

$$f_{XY}(x, y) = \begin{cases} \frac{3}{8}(x^2 + y^2), & \text{si } (x, y) \in [-1, 1] \times [-1, 1], \\ 0, & \text{en caso contrario.} \end{cases}$$

Dejamos como ejercicio comprobar que es una **función de densidad**. Es decir, que es positiva y que la integral sobre todo el plano vale 1.

Calculemos las **densidades marginales**:

$$\begin{aligned} f_X(x) &= \int_{-1}^1 \frac{3}{8} \cdot (x^2 + y^2) dy = \frac{3}{8} \cdot \left[x^2 \cdot y + \frac{y^3}{3} \right]_{-1}^1 = \frac{3}{8} \cdot \left(2 \cdot x^2 + \frac{2}{3} \right) = \frac{3}{4} \cdot x^2 + \frac{1}{4}, \\ f_Y(y) &= \int_{-1}^1 \frac{3}{8} \cdot (x^2 + y^2) dx = \frac{3}{8} \cdot \left[\frac{x^3}{3} + y^2 x \right]_{-1}^1 = \frac{3}{8} \cdot \left(\frac{2}{3} + 2y^2 \right) = \frac{3}{4} \cdot y^2 + \frac{1}{4}. \end{aligned}$$

Los valores esperados de cada variable X e Y son:

$$\begin{aligned} E(X) &= \int_{-1}^1 x \cdot \left(\frac{3}{4} \cdot x^2 + \frac{1}{4} \right) dx = 0, \text{ al integrar una función impar,} \\ E(Y) &= \int_{-1}^1 x \left(\frac{3}{4} \cdot y^2 + \frac{1}{4} \right) dx = 0, \text{ al integrar una función impar.} \end{aligned}$$

El valor de la **correlación** entre X e Y es:

$$\begin{aligned} E(X \cdot Y) &= \int_{-1}^1 \int_{-1}^1 x \cdot y \cdot \frac{3}{8} \cdot (x^2 + y^2) dy dx \\ &= \frac{3}{8} \cdot \left(\int_{-1}^1 \int_{-1}^1 x^3 \cdot y dy dx + \int_{-1}^1 \int_{-1}^1 x \cdot y^3 dy dx \right) \\ &= \frac{3}{8} \left(\int_{x=-1}^{x=1} x^3 \left[\frac{y^2}{2} \right]_{y=-1}^{y=1} dx + \int_{y=-1}^{y=1} y^3 \left[\frac{x^2}{2} \right]_{x=-1}^{x=1} \right) = 0. \end{aligned}$$

El **coeficiente de correlación** entre X e Y es: $\rho_{XY} = E(X \cdot Y) - E(X) \cdot E(Y) = 0 - 0 \cdot 0 = 0$. Por tanto, son **incorreladas**.

En cambio no son **independientes** ya que claramente si $(x, y) \in [-1, 1] \times [-1, 1]$,

$$f_{XY}(x, y) = \frac{3}{8}(x^2 + y^2) \neq f_X(x) \cdot f_Y(y) = \left(\frac{3}{4}x^2 + \frac{1}{4} \right) \cdot \left(\frac{3}{4}y^2 + \frac{1}{4} \right).$$

5.7. Variables aleatorias condicionales y valor esperado condicional

Muchas **variables aleatorias bidimensionales** de interés práctico no son independientes.

Por ejemplo, la salida Y de un canal de comunicación debe depender de la entrada X para transmitir información.

En esta sección vamos a introducir variables aleatorias Y cuya distribución depende de otras X . Dichas variables se denominan **variables aleatorias condicionales**.

También nos interesa el valor esperado de la **variable condicional** Y suponiendo que conocemos $X = x$.

5.7.1. Variables aleatorias condicionales discretas

Sea (X, Y) una variable aleatoria bidimensional. Sea B un subconjunto de los números reales \mathbb{R} . Recordemos que la **probabilidad condicional** del suceso $\{Y \in B\}$ suponiendo que $X = x$ se definía de la forma siguiente:

$$P(Y \in B | X = x) = \frac{P(Y \in B, X = x)}{P(X = x)}, \text{ siempre que } P(X = x) > 0.$$

La definición anterior motiva la definición siguiente de **variable aleatoria condicional discreta**:

Definición de variable aleatoria condicional discreta. Sea (X, Y) una variable aleatoria bidimensional discreta con conjunto de valores $(X, Y)(\Omega) = \{(x_i, y_j) | i = 1, 2, \dots, j = 1, 2, \dots\}$ y **función de probabilidad conjunta** P_{XY} . Sean x_i un valor de $X(\Omega)$ con $P(X = x_i) > 0$. Entonces definimos la **función de probabilidad de la variable aleatoria condicional discreta** $Y|X = x_i$ como:

$$P_{Y|X=x_i}(y_j) = P(Y = y_j | X = x_i) = \frac{P(X = x_i, Y = y_j)}{P(X = x_i)} = \frac{P_{XY}(x_i, y_j)}{P_X(x_i)}.$$

Observación. La **función de probabilidad de la variable aleatoria condicional** $Y|X = x_i$ depende únicamente de la **función de probabilidad conjunta** de la variable aleatoria bidimensional (X, Y) .

Observación. Al ser $Y|X = x_i$ una variable aleatoria unidimensional, su **función de probabilidad** tiene que verificar que la suma de todos sus valores tiene que dar 1. Es decir:

$$\sum_{y_j} P(Y = y_j | X = x_i) = 1.$$

Veámoslo:

$$\begin{aligned} \sum_{y_j} P(Y = y_j | X = x_i) &= \sum_{y_j} \frac{P_{XY}(x_i, y_j)}{P_X(x_i)} = \frac{1}{P_X(x_i)} \sum_{y_j} P_{XY}(x_i, y_j) \\ &= \frac{1}{P_X(x_i)} \cdot P_X(x_i) = 1. \end{aligned}$$

Observación. Si X e Y son independientes, la distribución de $Y|X = x_i$ es la misma que la de Y , es decir, la **variable aleatoria condicional** $Y|X = x_i$ coincide con Y . Es decir, condicionar con $X = x_i$ no tiene ningún efecto sobre Y .

Efectivamente, veamos que $P_{Y|X=x_i}(y_j) = P_Y(y_j)$ para todo valor y_j de $Y(\Omega)$:

$$\begin{aligned} P_{Y|X=x_i}(y_j) &= \frac{P_{XY}(x_i, y_j)}{P_X(x_i)} \stackrel{\text{Por ser independientes}}{=} \frac{P_Y(y_j) \cdot P_X(x_i)}{P_X(x_i)} \\ &= P_Y(y_j). \end{aligned}$$

Observación. La definición de la **función de probabilidad** de la **variable aleatoria condicional** $X|Y = y_j$ se definiría de forma similar:

$$P_{X|Y=y_j}(x_i) = P(X = x_i | Y = y_j) = \frac{P(X = x_i, Y = y_j)}{P(Y = y_j)} = \frac{P_{XY}(x_i, y_j)}{P_Y(y_j)},$$

para todo $x_i \in X(\Omega)$.

Observación. Si tenemos la tabla de la **función de probabilidad conjunta** P_{XY} , para hallar la **función de distribución de la variable** $Y|X = x_i$ es equivalente a considerar la fila del valor x_i a la tabla y dividir todos los valores de la fila por la suma de los valores en dicha fila:

$Y X = x_i$	y_1	y_2	...	y_N
$P_{Y X=x_i}$	$\frac{P_{XY}(x_i, y_1)}{P_X(x_i)}$	$\frac{P_{XY}(x_i, y_2)}{P_X(x_i)}$...	$\frac{P_{XY}(x_i, y_N)}{P_X(x_i)}$

Observación. De la misma manera, si tenemos la tabla de la **función de probabilidad conjunta** P_{XY} , para hallar la **función de distribución de la variable** $X|Y = y_j$ es equivalente a considerar la columna del valor $y = y_j$ a la tabla y dividir todos los valores de la columna por la suma de los valores en dicha columna:

$X Y = y_j$	$P_{X Y=y_j}$
x_1	$\frac{P_{XY}(x_1, y_j)}{P_Y(y_j)}$
\vdots	\vdots
x_M	$\frac{P_{XY}(x_M, y_j)}{P_Y(y_j)}$

5.7.1.1. Ejemplos

Ejemplo de la suma y el producto de los resultados de dos lanzamientos de un dado

Vamos a hallar la **variable aleatoria condicional** $S|P = 12$.

Tenemos calculada la tabla de la **función de probabilidad conjunta** P_{SP} .

Si $P = 12$, los únicos valores x_i de $S(\Omega)$ para los que se verifica $P_{SP}(x_i, 12) \neq 0$ son 7 y 8.

Además si calculamos $P_P(12)$, obtenemos $P(P = 12) = \frac{4}{36}$ ya que hay 4 casos en que el producto da 12: (3, 4), (4, 3), (2, 6) y (6, 2).

Por tanto, la tabla de la **función de probabilidad condicional** de la variable $S|P = 12$ es:

$S P = 12$	$P_{S P=12}$
7	$\frac{\frac{2}{36}}{\frac{4}{36}} = \frac{1}{2}$
8	$\frac{\frac{2}{36}}{\frac{4}{36}} = \frac{1}{2}$

Ejemplo de la suma y el producto de los resultados de dos lanzamientos de un dado

Vamos a hallar la **variable aleatoria condicional** $P|S = 8$.

Si $S = 8$, los únicos valores y_j de $P(\Omega)$ para los que se verifica $P_{SP}(8, y_j) \neq 0$ son 12 y 15 y 16.

El valor de $P_S(8)$ recordemos que valía: $P_S(8) = \frac{5}{36}$.

Por tanto, la tabla de la **función de probabilidad condicional** de la variable $P|S = 8$ es:

$P S = 8$	12	15	16
$P_{P S=8}$	$\frac{\frac{2}{36}}{\frac{5}{36}} = \frac{2}{5}$	$\frac{\frac{2}{36}}{\frac{5}{36}} = \frac{2}{5}$	$\frac{\frac{1}{36}}{\frac{5}{36}} = \frac{1}{5}$

Para hallar la **variable aleatoria condicional** $S|P = 12$ hemos de condicionar por la columna $P = 12$ en la tabla de la **función de probabilidad conjunta**:

```
prob.cond.p12=tabla.func.prob.conjunta[,valores.producto==12]/
  sum(tabla.func.prob.conjunta[,valores.producto==12])
prob.cond.p12
```

```
## 2 3 4 5 6 7 8 9 10 11 12
## 0.0 0.0 0.0 0.0 0.0 0.5 0.5 0.0 0.0 0.0 0.0
```

El problema es que aparecen valores con **función de probabilidad marginal** nulos. Para eliminarlos hacemos lo siguiente:

```
prob.cond.p12.buena = prob.cond.p12[prob.cond.p12!=0]
prob.cond.p12.buena
```

```
## 7 8
## 0.5 0.5
```

Para hallar la **función de probabilidad marginal** $P|S = 8$, haríamos lo siguiente:

```
prob.cond.s8=tabla.func.prob.conjunta[valores.suma==8,]/
  sum(tabla.func.prob.conjunta[valores.suma==8,])
(prob.cond.s8.buena = prob.cond.s8[prob.cond.s8!=0])
```

```
## 12 15 16
## 0.4 0.4 0.2
```

5.7.2. Variables aleatorias condicionales continuas

La definición en el caso continua se hace cambiando la **función de probabilidad** por la **función de densidad**:

Definición de variable aleatoria condicional discreta. Sea (X, Y) una variable aleatoria bidimensional continua con **función de densidad conjunta** f_{XY} . Sean $x \in \mathbb{R}$ con $f_X(x) > 0$. Entonces definimos la **función de densidad** de la **variable aleatoria condicional continua** $Y|X = x$ como:

$$f_{Y|X=x}(y) = \frac{f_{XY}(x, y)}{f_X(x)}.$$

Observación. La **función de densidad** de la **variable aleatoria condicional continua** $Y|X$ depende únicamente de la **función de densidad conjunta** de la variable aleatoria bidimensional (X, Y) .

Observación. Al ser $Y|X = x$ una variable aleatoria unidimensional, su **función de densidad** tiene que verificar que la integral de dicha función sobre todo \mathbb{R} tiene que ser 1. Es decir:

$$\int_{-\infty}^{\infty} f_{Y|X=x}(y) dy = 1.$$

Veámoslo:

$$\begin{aligned} \int_{-\infty}^{\infty} f_{Y|X=x}(y) dy &= \int_{-\infty}^{\infty} \frac{f_{XY}(x,y)}{f_X(x)} dy = \frac{1}{f_X(x)} \int_{-\infty}^{\infty} f_{XY}(x,y) dy \\ &= \frac{1}{f_X(x)} \cdot f_X(x) = 1. \end{aligned}$$

Observación. Si X e Y son independientes, $Y|X = x = Y$, es decir, la **variable aleatoria condicional** $Y|X = x$ coincide con Y . Es decir, condicionar con $X = x$ no tiene ningún efecto sobre Y .

Efectivamente, veamos que $f_{Y|X=x}(y) = f_Y(y)$ para todo valor $y \in \mathbb{R}$.

$$f_{Y|X=x}(y) = \frac{f_{XY}(x,y)}{f_X(x)} \text{ Al ser independientes } \frac{f_Y(y) \cdot f_X(x)}{f_X(x)} = f_Y(y).$$

Observación. La definición de la **función de densidad** de la **variable aleatoria condicional** $X|Y = y$ se definiría de forma similar:

$$f_{X|Y=y}(x) = \frac{f_{XY}(x,y)}{f_Y(y)},$$

para todo $x \in \mathbb{R}$.

5.7.2.1. Ejemplos

Ejemplo: otra función de densidad (continuación)

Recordemos el ejemplo de la variable aleatoria bidimensional continua con **función de densidad**:

$$f_{XY}(x,y) = \begin{cases} 2 \cdot e^{-x} \cdot e^{-y}, & 0 \leq y \leq x < \infty, \\ 0, & \text{en caso contrario,} \end{cases}$$

Dado un valor $x_0 \geq 0$ cualquiera, vamos a hallar la **función de densidad** de la **variable aleatoria condicional** $Y|X = x_0$.

Fijémonos que, fijado un valor x_0 , los valores y para los cuales $f_{XY}(x_0, y) \neq 0$ cumplen $0 \leq y \leq x_0$. Por tanto,

$$f_{Y|X=x_0}(y) = \frac{f_{XY}(x_0, y)}{f_X(x_0)} = \frac{2 \cdot e^{-x_0} \cdot e^{-y}}{f_X(x_0)},$$

si $0 \leq y \leq x_0$, y $f_{Y|X=x_0}(y) = 0$, en caso contrario.

Recordemos que la **densidad marginal** de la variable X era: $f_X(x_0) = 2(e^{-x_0} - e^{-2x_0})$.

La **función de densidad marginal** de la variable $Y|X = x_0$ es:

$$f_{Y|X=x_0}(y) = \frac{2 \cdot e^{-x_0} \cdot e^{-y}}{2 \cdot (e^{-x_0} - e^{-2 \cdot x_0})} = \frac{e^{-y}}{1 - e^{-x_0}},$$

si $0 \leq y \leq x_0$, y $f_{Y|X=x_0}(y) = 0$, en caso contrario.

Sea ahora $y_0 > 0$. Calculemos ahora la **densidad marginal** de la variable $X|Y = y_0$.

Fijémonos que, fijado un valor y_0 , los valores x para los cuales $f_{XY}(x, y_0) \neq 0$ cumplen $y_0 \leq x \leq \infty$. Por tanto,

$$f_{X|Y=y_0}(x) = \frac{f_{XY}(x, y_0)}{f_Y(y_0)} = \frac{2 \cdot e^{-x} \cdot e^{-y_0}}{f_Y(y_0)},$$

si $y_0 \leq x \leq \infty$, y $f_{X|Y=y_0}(x) = 0$, en caso contrario.

Recordemos que la variable Y era exponencial de parámetro $\lambda = 2$. Por tanto, $f_Y(y_0) = 2e^{-2y_0}$.

La **función de densidad marginal** de la variable $X|Y = y_0$ es:

$$f_{X|Y=y_0}(x) = \frac{2 \cdot e^{-x} \cdot e^{-y_0}}{2 \cdot e^{-2y_0}} = \frac{e^{-x}}{e^{-y_0}},$$

si $y_0 \leq x \leq \infty$, y $f_{X|Y=y_0}(x) = 0$, en caso contrario.

Ejemplo: normal bidimensional

Sea (X, Y) una variable aleatoria bidimensional normal bidimensional con **densidad conjunta**:

$$f_{XY}(x, y) = \frac{1}{2\pi\sqrt{1-\rho^2}} e^{-\frac{(x^2-2\rho xy+y^2)}{2(1-\rho^2)}}, \quad -\infty < x, y < \infty.$$

Sea $x \in \mathbb{R}$. Hallemos la **función de densidad** de la **variable aleatoria condicionada** $Y|X = x$.

Recordemos que las **marginales** eran $N(0, 1)$. Por tanto, $f_X(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$.

La **función de densidad** de la variable condicional $Y|X = x$ es:

$$\begin{aligned} f_{Y|X=x}(y) &= \frac{f_{XY}(x, y)}{f_X(x)} = \frac{\frac{1}{2\pi\sqrt{1-\rho^2}} e^{-\frac{(x^2-2\rho xy+y^2)}{2(1-\rho^2)}}}{\frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}} \\ &= \frac{1}{\sqrt{2\pi \cdot (1-\rho^2)}} e^{-\frac{(y-\rho x)^2}{2(1-\rho^2)}}, \quad y \in \mathbb{R}. \end{aligned}$$

Concluimos que la **variable aleatoria condicional** $Y|X = x$ es una normal de parámetros $\mu_{Y|X=x} = \rho x$ y $\sigma_{Y|X=x}^2 = 1 - \rho^2$.

Observaciones

Tenemos dos observaciones con respecto al resultado obtenido:

- La varianza de la **variable aleatoria condicional** no depende de la x que se ha fijado. Sólo depende del parámetro ρ . La x sólo influye en la media de dicha variable.

- En el caso en que $\rho = 0$, que significa que X e Y son independientes, la distribución condicional de $Y|X = x$ es una $N(0, 1)$, distribución que coincide con la distribución de la **variable aleatoria marginal** Y .

5.7.3. Valores esperados condicionales

Definición de valor esperado condicional. Dada una variable aleatoria bidimensional (X, Y) , definimos el **valor esperado de la variable Y dado que $X = x$** como $E(Y|x)$, es decir, el valor esperado de la **variable aleatoria condicional** $Y|X = x$:

$$E(Y|x) = \begin{cases} \sum_{y_j} y_j \cdot P_{Y|X=x}(y_j), & \text{caso discreto,} \\ \int_{-\infty}^{\infty} y \cdot f_{Y|X=x}(y) dy, & \text{caso continuo.} \end{cases}$$

Tenemos el siguiente resultado relacionado con los valores esperados: el valor esperado respecto x del valor esperado de la **variable condicional** $Y|X = x$ coincide con el valor esperado de la variable Y :

Proposición. Sea (X, Y) una variable aleatoria bidimensional. Sean $E(Y|x)$ el **valor esperado condicional** de Y respecto x . Entonces el valor esperado de la *variable aleatoria* $E(Y|X)$ como función de la variable X es el valor esperado de la variable Y :

$$E_X(E(Y|X)) = E(Y).$$

Demostración

Haremos la demostración en el caso continuo. Dejamos como ejercicio la demostración para el caso discreto.

Sea f_{XY} la **función de densidad conjunta** y f_X y f_Y las **funciones de densidad marginales**.

El valor de $E_X(E(Y|X))$ es:

$$\begin{aligned} E_X(E(Y|X)) &= \int_{x=-\infty}^{x=\infty} E(Y|x) f_X(x) dx = \int_{x=-\infty}^{x=\infty} \int_{y=-\infty}^{y=\infty} y f_{Y|X=x}(y) dy f_X(x) dx \\ &= \int_{x=-\infty}^{x=\infty} \int_{y=-\infty}^{y=\infty} y \frac{f_{XY}(x,y)}{f_X(x)} f_X(x) dy dx = \int_{y=-\infty}^{y=\infty} y \int_{x=-\infty}^{x=\infty} f_{XY}(x,y) dx dy \\ &= \int_{y=-\infty}^{y=\infty} y f_Y(y) dy = E(Y), \end{aligned}$$

tal como queríamos ver.

5.7.4. Relación con el problema de la regresión general

El problema de la **regresión general** es el siguiente:

Sea (X, Y) una variable aleatoria bidimensional. Queremos hallar una función g tal que la variable $\hat{Y} = g(X)$ explique mejor la variable Y .

Dicho de forma más explícita, queremos hallar una función g tal que minimice el error cometido al aproximar Y por $\hat{Y} = g(X)$. Dicho error se define de forma natural como el valor esperado de la variable $(Y - g(X))^2$:

$$\min_g E((Y - g(X))^2).$$

El siguiente resultado nos dice cuál es la función g :

Proposición: La función g solución del problema de **regresión general** es la siguiente: $g(x) = E(Y|X = x)$.

Es decir, la función g asigna a cada valor x de la variable aleatoria X , el valor esperado de la **variable condicional** $Y|X = x$.

En resumen, la función $g(x) = E(Y|X = x)$ es la función que minimiza el error. A la curva $y = g(x)$ se la denomina **curva general de regresión de Y sobre X** .

5.7.5. Valores esperados condicionales. Caso general

Podemos generalizar los valores esperados condicionales en el sentido que en lugar de hallar $E(Y|X = x)$, hallar $E(g(Y)|X = x)$, donde g es una función de la variable aleatoria Y :

Definición de valor esperado condicional.

Dada una variable aleatoria bidimensional (X, Y) y una función g , definimos el **valor esperado de la variable $g(Y)$ dado que $X = x$** como $E(g(Y)|x)$, es decir, el valor esperado de la **variable aleatoria condicional** $g(Y)|X = x$:

$$E(g(Y)|x) = \begin{cases} \sum_{y_j} g(y_j)P_{Y|X=x}(y_j), & \text{caso discreto,} \\ \int_{-\infty}^{\infty} g(y)f_{Y|X=x}(y) dy, & \text{caso continuo.} \end{cases}$$

Observación: cuando $g(y) = y^k$, tenemos definidos los **momentos condicionados de orden k** de la variable $Y|X = x$.

5.7.5.1. Ejemplos

Ejemplo de la suma y el producto de los resultados de dos lanzamientos de un dado

Vamos a hallar el valor esperado de la **variable aleatoria condicional** $P|S = 8$.

Recordemos su **función de probabilidad**:

$P S = 8$	12	15	16
$P_{P S=8}$	$\frac{2}{36} = \frac{1}{18}$	$\frac{2}{36} = \frac{1}{18}$	$\frac{1}{36} = \frac{1}{36}$

Su valor esperado es:

$$E(P|S = 8) = 12 \cdot \frac{1}{18} + 15 \cdot \frac{1}{18} + 16 \cdot \frac{1}{36} = \frac{70}{18} = \frac{35}{9} = 3.888\overline{8}$$

El valor medio del producto de los resultados al lanzar un dado dos veces cuando la suma de dichos resultados es 8 vale 14.

El valor esperado de la variable $E(P|S = 8)$ es:

```
valores.cond.s8=as.integer(names(prob.cond.s8.buena))
sum(valores.cond.s8*prob.cond.s8.buena)
```

```
## [1] 14
```

Ejemplo: otra densidad (continuación)

Recordemos el ejemplo de la variable aleatoria bidimensional continua con **función de densidad**:

$$f_{XY}(x, y) = \begin{cases} 2e^{-x}e^{-y}, & 0 \leq y \leq x < \infty, \\ 0, & \text{en caso contrario,} \end{cases}$$

Ya sabemos que si fijamos $x_0 > 0$, la **función de densidad de la variable aleatoria condicionada** $Y|X = x_0$ es:

$$f_{Y|X=x_0}(y) = \begin{cases} \frac{e^{-y}}{1-e^{-x_0}}, & \text{si } 0 \leq y \leq x_0, \\ 0, & \text{en caso contrario.} \end{cases}$$

Hallemos su valor esperado:

$$E(Y|X = x_0) = \int_0^{x_0} y \frac{e^{-y}}{(1 - e^{-x_0})} dy = \frac{1}{(1 - e^{-x_0})} [-e^{-y}(y + 1)]_0^{x_0} = \frac{1 - e^{-x_0}(1 + x_0)}{1 - e^{-x_0}}.$$

Verifiquemos la propiedad vista anteriormente $E_X(E(Y|x)) = E(Y)$. Recordemos que la **función de densidad marginal** de la variable X era: $f_X(x) = 2(e^{-x} - e^{-2x})$, para $x > 0$:

$$\begin{aligned} E_X(E(Y|x)) &= \int_0^{\infty} E(Y|x) \cdot f_X(x) dx = \int_0^{\infty} \frac{1-e^{-x}(1+x)}{1-e^{-x}} \cdot 2(e^{-x} - e^{-2x}) dx \\ &= 2 \int_0^{\infty} \frac{1-e^{-x}(1+x)}{1-e^{-x}} e^{-x} (1 - e^{-x}) dx = 2 \int_0^{\infty} (e^{-x} - e^{-2x}(1+x)) dx \\ &= 2 \left[-e^{-x} + e^{-2x} \left(\frac{x}{2} + \frac{3}{4} \right) \right]_0^{\infty} = 2 \left(1 - \frac{3}{4} \right) = \frac{1}{2}. \end{aligned}$$

Recordemos que la variable Y era exponencial de parámetro $\lambda = 2$. Por tanto $E(Y) = \frac{1}{\lambda} = \frac{1}{2}$, valor que coincide con el hallado, tal como queríamos ver.

5.8. Variables aleatorias definidas como función de dos variables aleatorias conjuntas

Dado un experimento aleatorio, a veces estaremos interesados en una o más funciones de las variables asociadas con el experimento.

Por ejemplo, si consideramos el experimento aleatorio de lanzar un dado dos veces y definimos la **variable aleatoria bidimensional** (X_1, X_2) como la variable que nos da el resultado de cada lanzamiento, podemos expresar la suma y el producto como $S = X_1 + X_2$, $P = X_1 \cdot X_2$.

5.8. VARIABLES ALEATORIAS DEFINIDAS COMO FUNCIÓN DE DOS VARIABLES ALEATORIAS CONJUNTAS

Otros ejemplos podrían ser considerar el experimento aleatoria de realizar mediciones repetidas de la misma cantidad aleatoria. Entonces, podríamos estar interesados en el valor máximo y mínimo en el conjunto, así como la media muestral y la varianza muestral.

En esta sección presentamos métodos para determinar las probabilidades de eventos que involucran **funciones de dos variables aleatorias**.

Daremos métodos de cómo hallar la **función de distribución** y la **función de probabilidad** (caso discreto) o la **función de densidad** (caso continuo) de la variable aleatoria definida como función de la **variable aleatoria bidimensional**.

5.8.1. Variable aleatoria función de la variable aleatoria bidimensional

Proposición.

Sea (X, Y) una variable aleatoria bidimensional con **función de probabilidad** P_{XY} (caso discreto) o **función de densidad** (caso continuo). Sea g una función y definimos la **variable aleatoria unidimensional** Z como $Z = g(X, Y)$. Entonces la función de distribución de Z es:

$$\begin{aligned} F_Z(z) &= P(Z \leq z) = \sum_{(x_i, y_j), |g(x_i, y_j)| \leq z} P_{XY}(x_i, y_j), \quad z \in \mathbb{R}, \\ &\quad (\text{caso discreto}), \\ F_Z(z) &= P(Z \leq z) = \int \int_{(x, y) \in \mathbb{R}^2, |g(x, y)| \leq z} f_{XY}(x, y) dy dx, \quad z \in \mathbb{R}, \\ &\quad (\text{caso continuo}). \end{aligned}$$

Observación.

En el caso discreto, la variable aleatoria será discreta con valores $Z(\Omega) = \{z_{ij} = g(x_i, y_j), |(x_i, y_j) \in (X, Y)(\Omega)\}$. Hay que tener en cuenta que en dicho conjunto puede haber repeticiones, es decir, pueden existir dos parejas (i, j) y (i', j') tal que $z_{ij} = z_{i'j'}$.

La expresión de la **función de probabilidad** en el caso discreto se complica mucho debido a dichas repeticiones y es mejor hallarla en cada caso concreto.

La última observación se puede aplicar también en el caso continuo: la expresión de la **función de densidad** se halla en cada caso concreto.

Ejemplo del lanzamiento de un dado dos veces.

Consideremos el experimento aleatorio de lanzar dos veces un dado.

Sea (X, Y) la **variable aleatoria** bidimensional discreta ya estudiada anteriormente donde X nos da el resultado del primer lanzamiento e Y , el resultado del segundo lanzamiento.

Vimos que $(X, Y)(\Omega) = \{(i, j), i = 1, 2, 3, 4, 5, 6, j = 1, 2, 3, 4, 5, 6\}$ con **función de probabilidad conjunta** $P_{XY}(i, j) = \frac{1}{36}$, $i = 1, 2, 3, 4, 5, 6$, $j = 1, 2, 3, 4, 5, 6$.

Anteriormente hemos estudiado la suma S de los resultados. En este caso podemos interpretar $S = g(X, Y)$ donde $g(x, y) = x + y$.

Como la función S ya ha sido estudiada y el producto se ha dejado como ejercicio, estudiaremos la siguiente variable aleatoria función de X e Y : $Z = X^2 + Y^2$.

Realizaremos los cálculos con ayuda de R ya que hacerlos a mano es bastante tedioso.

Los valores de $Z(\Omega)$ son: $Z(\Omega) = \{z_{ij} = i^2 + j^2, i = 1, 2, 3, 4, 5, 6, j = 1, 2, 3, 4, 5, 6\}$. Observad que hay parejas (i, j) que dan lugar a los mismos valores, por ejemplo $1^2 + 2^2 = 2^2 + 1^2$, y, en general, si $i \neq j$, $z_{ij} = i^2 + j^2 = z_{ji} = j^2 + i^2$.

Para hallar el conjunto $Z(\Omega)$ usamos la función `outer` de R:

```
g=function(x,y){x^2+y^2} ## definimos la función g
sort(unique(as.vector(outer(1:6,1:6,g))))
```

```
## [1] 2 5 8 10 13 17 18 20 25 26 29 32 34 37 40 41 45 50 52 61 72
```

Vemos que hay 21 valores distintos de la variable Z .

Para hallar la **función de probabilidad** de Z hemos de calcular para cada valor z_k , las parejas (i, j) tal que $i^2 + j^2 = z_k$:

```
valores.variable.Z = sort(unique(as.vector(outer(1:6,1:6,g))))
matriz.valores = outer(1:6,1:6,g) ## aplicamos la función g a
## todas las parejas (i,j), i,j=1,2,3,4,5,6
frecuencias = c() ## vector donde guardaremos las frecuencias de los valores de Z
for (i in 1:length(valores.variable.Z)){
  z=valores.variable.Z[i]
  frecuencias=c(frecuencias,length(matriz.valores[matriz.valores==z]))
}
frecuencias

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

La **función de probabilidad** de Z es:

```
función.probabilidad.Z=data.frame(rbind(valores.variable.Z,round(frecuencias/36,3)))
rownames(función.probabilidad.Z)=c("Z","P_Z")
función.probabilidad.Z
```

```
##      X1     X2     X3     X4     X5     X6     X7     X8     X9     X10    X11
## Z  2.000 5.000 8.000 10.000 13.000 17.000 18.000 20.000 25.000 26.000 29.000
## P_Z 0.028 0.056 0.028 0.056 0.056 0.056 0.028 0.056 0.056 0.056 0.056 0.056
##      X12    X13    X14    X15    X16    X17    X18    X19    X20    X21
## Z  32.000 34.000 37.000 40.000 41.000 45.000 50.000 52.000 61.000 72.000
## P_Z 0.028 0.056 0.056 0.056 0.056 0.056 0.028 0.056 0.056 0.028
```

Ejemplo variables aleatorias continuas

Recordemos la variable aleatoria bidimensional (X, Y) con **función de densidad**:

$$f_{XY}(x, y) = \begin{cases} 2 \cdot e^{-x} \cdot e^{-y}, & 0 \leq y \leq x < \infty, \\ 0, & \text{en caso contrario,} \end{cases}$$

Consideremos la variable aleatoria $Z = X + Y$. Vamos a calcular la **función de densidad** de Z .

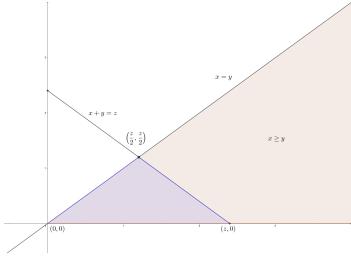
En primer lugar, los valores de Z para los que $f_Z(z) \neq 0$ cumplen $z \geq 0$ ya que $X \geq 0$ e $Y \geq 0$.

Calculemos la **función de distribución** de la variable Z . Sean $z \in \mathbb{R}$ con $z \geq 0$:

5.8. VARIABLES ALEATORIAS DEFINIDAS COMO FUNCIÓN DE DOS VARIABLES ALEATORIAS CONJUNTAS

$$\begin{aligned} F_Z(z) &= P(Z \leq z) = P(X + Y \leq z) \\ &= \int \int_{\{(x,y) \in \mathbb{R}^2, | x+y \leq z \} \cap \{(x,y) \in \mathbb{R}^2, | 0 \leq y \leq x < \infty \}} 2 \cdot e^{-x} \cdot e^{-y} dy dx \end{aligned}$$

El gráfico siguiente muestra en color violeta la región de integración para hallar $F_Z(z)$ dado un $z \geq 0$.



El valor de $F_Z(z)$ es: (fijémonos que primero fijamos la y y para cada y la x va desde la recta $x = y$ hasta la recta $x = z - y$)

$$\begin{aligned} F_Z(z) &= \int_{y=0}^{y=\frac{z}{2}} \int_{x=y}^{x=z-y} 2 \cdot e^{-x} \cdot e^{-y} dx dy = 2 \cdot \int_{y=0}^{y=\frac{z}{2}} e^{-y} \cdot [-e^{-x}]_{x=y}^{x=z-y} dy \\ &= 2 \cdot \int_{y=0}^{y=\frac{z}{2}} e^{-y} \cdot (e^{-y} - e^{y-z}) dy = 2 \cdot \int_{y=0}^{y=\frac{z}{2}} (e^{-2y} - e^{-z}) dy \\ &= 2 \cdot \left[-\frac{1}{2}e^{-2y} - e^{-z} \cdot y \right]_{y=0}^{y=\frac{z}{2}} = 2 \cdot \left(\frac{1}{2} - \frac{1}{2}e^{-z} - \frac{z}{2}e^{-z} \right) \\ &= 1 - e^{-z} \cdot (1 + z), \quad z \geq 0. \end{aligned}$$

La **función de densidad** de Z es:

$$f_Z(z) = F'_Z(z) = z \cdot e^{-z}, \quad z \geq 0,$$

y $f_Z(z) = 0$ en caso contrario.

Ejemplo de la suma de dos normales

Consideremos el caso en que la variable aleatoria (X, Y) tenga distribución **normal bidimensional**.

Recordemos que su **función de densidad conjunta** era:

$$f_{XY}(x, y) = \frac{1}{2 \cdot \pi \cdot \sqrt{1-\rho^2}} e^{-\frac{(x^2-2\rho xy+y^2)}{2(1-\rho^2)}}, \quad -\infty < x, y < \infty.$$

Consideremos $S = X + Y$. Estudiemos qué distribución tiene S .

Dado un valor $z \in \mathbb{R}$, la **función de distribución** de S en s es:

$$\begin{aligned}
F_S(s) &= P(S \leq s) = \int \int_{\{(x,y) \in \mathbb{R}^2, |x+y| \leq s\}} \frac{1}{2\pi\sqrt{1-\rho^2}} e^{-\frac{(x^2-2\rho xy+y^2)}{2(1-\rho^2)}} dy dx \\
&= \frac{1}{2\pi\sqrt{1-\rho^2}} \int_{x=-\infty}^{x=\infty} \int_{y=-\infty}^{y=s-x} e^{-\frac{(x^2-2\rho xy+y^2)}{2(1-\rho^2)}} dy dx \\
&= \frac{1}{2\pi\sqrt{1-\rho^2}} \int_{x=-\infty}^{x=\infty} e^{-\frac{x^2}{2(1-\rho^2)}} \int_{y=-\infty}^{y=s-x} e^{-\frac{(-2\rho xy+y^2)}{2(1-\rho^2)}} dy dx \\
&\quad \text{hacemos el siguiente cambio en la segunda integral } t = y + x \\
&= \frac{1}{2\pi\sqrt{1-\rho^2}} \int_{x=-\infty}^{x=\infty} e^{-\frac{x^2}{2(1-\rho^2)}} \int_{t=-\infty}^{t=s} e^{-\frac{(-2\rho x(t-x)+(t-x)^2)}{2(1-\rho^2)}} dt dx \\
F_S(s) &= \frac{1}{2\pi\sqrt{1-\rho^2}} \int_{x=-\infty}^{x=\infty} e^{-\frac{(1+\rho)x^2}{1-\rho^2}} \int_{t=-\infty}^{t=s} e^{-\frac{(t^2-2(1+\rho)t x)}{2(1-\rho^2)}} dt dx \\
&= \frac{1}{2\pi\sqrt{1-\rho^2}} \int_{x=-\infty}^{x=\infty} e^{-\frac{(1+\rho)x^2}{1-\rho^2}} \int_{t=-\infty}^{t=s} e^{-\frac{(t-(1+\rho)x)^2}{2(1-\rho^2)}} e^{\frac{(\rho+1)^2 x^2}{2(1-\rho^2)}} dt dx \\
&= \frac{1}{2\pi\sqrt{1-\rho^2}} \int_{x=-\infty}^{x=\infty} e^{-\frac{x^2}{2}} \cdot \sqrt{2\pi(1-\rho^2)} F_X(s) dx,
\end{aligned}$$

donde $F_X(s)$ es la función de distribución de una variable X normal de parámetros $\mu = (1+\rho)x$ y $\sigma^2 = 1-\rho^2$.

$$= \frac{1}{\sqrt{2\pi}} \int_{x=-\infty}^{x=\infty} e^{-\frac{x^2}{2}} \cdot F_X(s) dx.$$

Para calcular la **función de densidad** $f_S(s)$ aplicamos la expresión $f_S(s) = F'_S(s)$ y la derivación bajo el signo integral:

$$\begin{aligned}
f_S(s) &= \frac{1}{\sqrt{2\pi}} \int_{x=-\infty}^{x=\infty} e^{-\frac{x^2}{2}} \cdot f_X(s) dx = \frac{1}{\sqrt{2\pi}} \int_{x=-\infty}^{x=\infty} e^{-\frac{x^2}{2}} \cdot \frac{1}{\sqrt{2\pi(1-\rho^2)}} e^{-\frac{(s-(1+\rho)x)^2}{2(1-\rho^2)}} dx \\
&= \frac{e^{-\frac{s^2}{2(1-\rho^2)}}}{2\pi\sqrt{1-\rho^2}} \int_{x=-\infty}^{x=\infty} e^{-\frac{(2(1+\rho)x^2-2(1+\rho)sx)}{2(1-\rho^2)}} dx = \frac{e^{-\frac{s^2}{2(1-\rho^2)}}}{2\pi\sqrt{1-\rho^2}} \int_{x=-\infty}^{x=\infty} e^{-\frac{(x-\frac{s}{2})^2}{1-\rho}} e^{\frac{s^2}{4(1-\rho)}} dx
\end{aligned}$$

En la última integral hacemos el cambio $u = x - \frac{s}{2}$:

$$f_S(s) = \frac{e^{-\frac{s^2}{4(1+\rho)}}}{2\pi\sqrt{1-\rho^2}} \int_{u=-\infty}^{u=\infty} e^{-\frac{u^2}{1-\rho}} du.$$

A continuación usando que $f_Z(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}}$ es la función de densidad de la distribución $Z = N(0, 1)$, podemos escribir: $\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} dz = 1$, $\Rightarrow \int_{-\infty}^{\infty} e^{-\frac{u^2}{1-\rho}} du = \sqrt{2\pi}$.

Si en la última integral hacemos el cambio $v = \sqrt{\frac{2}{1-\rho}} u$, obtenemos:

$$\begin{aligned}
f_S(s) &= \frac{e^{-\frac{s^2}{4(1+\rho)}}}{2\pi\sqrt{1-\rho^2}} \int_{v=-\infty}^{v=\infty} e^{-\frac{v^2}{2}} \sqrt{\frac{1-\rho}{2}} dv = \frac{e^{-\frac{s^2}{4(1+\rho)}}}{2\pi\sqrt{2(1+\rho)}} \int_{v=-\infty}^{v=\infty} e^{-\frac{v^2}{2}} dv \\
&= \frac{e^{-\frac{s^2}{4(1+\rho)}}}{2\pi\sqrt{2(1+\rho)}} \sqrt{2\pi} = \frac{1}{\sqrt{2\pi 2(1+\rho)}} e^{-\frac{s^2}{4(1+\rho)}}, \quad s \in \mathbb{R}.
\end{aligned}$$

Dicha función de densidad corresponde a una distribución normal de parámetros $\mu = 0$ y $\sigma = \sqrt{2(1+\rho)}$.

5.8. VARIABLES ALEATORIAS DEFINIDAS COMO FUNCIÓN DE DOS VARIABLES ALEATORIAS CONJUNTAS

En resumen, la distribución de la suma de dos normales es una normal de parámetros $S = N(\mu = 0, \sigma = \sqrt{2(1 + \rho)})$.

5.8.2. Transformaciones lineales de variables aleatorias

Consideremos una variable aleatoria bidimensional continua (X, Y) con **función de densidad conjunta** f_{XY} .

Definimos la variable aleatoria bidimensional continua (U, V) a partir de una transformación lineal de la variable (X, Y) . Es decir, existe una matriz $\mathbf{M} = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$ y un vector $\mathbf{n} = \begin{pmatrix} \alpha \\ \beta \end{pmatrix}$ tal que:

$$\begin{aligned} \begin{pmatrix} U \\ V \end{pmatrix} &= \mathbf{M} \cdot \begin{pmatrix} X \\ Y \end{pmatrix} + \mathbf{n} = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \cdot \begin{pmatrix} X \\ Y \end{pmatrix} + \begin{pmatrix} \alpha \\ \beta \end{pmatrix}, \\ \Rightarrow \begin{cases} U = aX + bY + \alpha, \\ V = cX + dY + \beta. \end{cases} \end{aligned}$$

Para que (U, V) sea una variable aleatoria bidimensional, necesitamos que la matriz \mathbf{M} sea no singular, o $\det(\mathbf{M}) \neq 0$.

Nos preguntamos cuál es la relación entre la **función de densidad** de la variable (U, V) , f_{UV} y la **función de densidad** de la variable (X, Y) , f_{XY} . La expresión siguiente nos da dicha relación:

$$f_{UV}(u, v) = \frac{1}{|\det(\mathbf{M})|} f_{XY} \left(\mathbf{M}^{-1} \begin{pmatrix} u - \alpha \\ v - \beta \end{pmatrix} \right), \quad (u, v) \in \mathbb{R}^2.$$

Observación. Si la variable (X, Y) tiene una región D donde $f_{XY}(x, y) \neq 0$, para todo $(x, y) \in D$, antes de aplicar la expresión anterior para hallar la **función de densidad** de la variable (U, V) hemos de calcular cómo se transforma D con la matriz \mathbf{M} . Es decir, hay que hallar la región

$$D' = \mathbf{M}(D) = \{(u, v) \in \mathbb{R}^2, \text{ existe } (x, y) \in D \text{ con } (u, v) = \mathbf{M}(x, y) + \mathbf{n}\}.$$

Ejemplo: transformación lineal

Consideremos la variable (X, Y) continua con función de densidad:

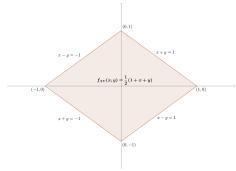
$$f_{XY}(x, y) = \begin{cases} \frac{1}{2}(1 + x + y), & \text{si } (x, y) \in R, \\ 0, & \text{en caso contrario.} \end{cases}$$

donde R es el rombo de vértices $(1, 0)$, $(0, 1)$, $(-1, 0)$ y $(0, -1)$, ver figura adjunta.

Otra forma de definir la función anterior es:

$$f_{XY}(x, y) = \begin{cases} \frac{1}{2}(1 + x + y), & -1 \leq x \leq 0, -1 - x \leq y \leq x + 1, \\ \frac{1}{2}(1 + x + y), & 0 \leq x \leq 1, x - 1 \leq y \leq 1 - x, \\ 0, & \text{en caso contrario.} \end{cases}$$

Dejamos como ejercicio al lector comprobar que la función anterior es una **función de densidad**.



Consideramos la variable aleatoria bidimensional (U, V) definida a partir de la variable (X, Y) :

$$\begin{pmatrix} U \\ V \end{pmatrix} = \begin{pmatrix} 1 & -1 \\ 1 & 1 \end{pmatrix} \cdot \begin{pmatrix} X \\ Y \end{pmatrix}, \Rightarrow \begin{cases} U = X - Y, \\ V = X + Y. \end{cases}$$

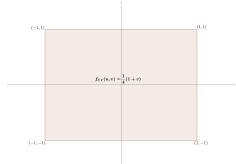
La región R se transforma en el cuadrado C de vértices $(1, 1)$, $(-1, 1)$, $(-1, -1)$ y $(1, -1)$ ya que si aplicamos la matriz a los vértices del rombo, obtenemos los vértices de cuadrado:

$$\begin{aligned} \begin{pmatrix} 1 & -1 \\ 1 & 1 \end{pmatrix} \cdot \begin{pmatrix} 1 \\ 0 \end{pmatrix} &= \begin{pmatrix} 1 \\ 1 \end{pmatrix}, & \begin{pmatrix} 1 & -1 \\ 1 & 1 \end{pmatrix} \cdot \begin{pmatrix} 0 \\ 1 \end{pmatrix} &= \begin{pmatrix} -1 \\ 1 \end{pmatrix}, \\ \begin{pmatrix} 1 & -1 \\ 1 & 1 \end{pmatrix} \cdot \begin{pmatrix} -1 \\ 0 \end{pmatrix} &= \begin{pmatrix} -1 \\ -1 \end{pmatrix}, & \begin{pmatrix} 1 & -1 \\ 1 & 1 \end{pmatrix} \cdot \begin{pmatrix} 0 \\ -1 \end{pmatrix} &= \begin{pmatrix} 1 \\ -1 \end{pmatrix}. \end{aligned}$$

Ver la figura adjunta.

Para hallar la **función de densidad** f_{UV} necesitamos escribir X e Y en función de U y V :

$$\begin{pmatrix} X \\ Y \end{pmatrix} = \begin{pmatrix} 1 & -1 \\ 1 & 1 \end{pmatrix}^{-1} \cdot \begin{pmatrix} U \\ V \end{pmatrix} = \begin{pmatrix} \frac{1}{2} & \frac{1}{2} \\ -\frac{1}{2} & \frac{1}{2} \end{pmatrix} \cdot \begin{pmatrix} U \\ V \end{pmatrix}, \Rightarrow \begin{cases} X = \frac{1}{2}(U + V), \\ Y = \frac{1}{2}(-U + V). \end{cases}$$



La **función de densidad** f_{UV} es,

$$\begin{aligned} f_{UV}(u, v) &= \frac{1}{\left| \det \begin{pmatrix} 1 & -1 \\ 1 & 1 \end{pmatrix} \right|} \cdot f_{XY} \left(\frac{1}{2}(u + v), \frac{1}{2}(-u + v) \right) \\ &= \frac{1}{2} \cdot \frac{1}{2} \cdot \left(1 + \frac{1}{2}(-u + v) + \frac{1}{2}(u + v) \right) = \frac{1}{4}(1 + v), \end{aligned}$$

para (u, v) perteneciente al cuadrado C de vértices $(1, 1)$, $(-1, 1)$, $(-1, -1)$ y $(1, -1)$, o si se quiere para $-1 \leq u \leq 1$, $-1 \leq v \leq 1$, y $f_{UV}(u, v) = 0$, en caso contrario.

Observamos que es más cómodo trabajar con las variables (u, v) en vez de trabajar con las variables (x, y) por dos razones:

- La región donde la **función de densidad** no es nula es más simple, ya que trabajar con un cuadrado simplifica mucho más los cálculos que trabajar con un rombo a la hora de hallar la **función de distribución**, **densidades marginales**, **densidades condicionadas**, **valores esperados**, etc.

5.8. VARIABLES ALEATORIAS DEFINIDAS COMO FUNCIÓN DE DOS VARIABLES ALEATORIAS CONJUNTAS

- La expresión de la **función de densidad** también es más simple, ya que sólo depende de la segunda variable v ; sin embargo, la **función de densidad** inicial f_{XY} dependía de las dos variables x e y .

Ejemplo

Consideremos el caso en que la variable aleatoria (X, Y) tenga distribución **normal bidimensional**.

Recordemos que su **función de densidad conjunta** era:

$$f_{XY}(x, y) = \frac{1}{2\pi\sqrt{1-\rho^2}} e^{-\frac{(x^2-2\rho xy+y^2)}{2(1-\rho^2)}}, \quad -\infty < x, y < \infty.$$

Recordemos que las **distribuciones marginales** son distribuciones $N(0, 1)$.

La idea es hallar la **función de densidad conjunta** de una distribución normal bidimensional para la que sus **distribuciones marginales** sean dos normales $N(\mu_1, \sigma_1)$ y $N(\mu_2, \sigma_2)$.

Recordemos que si $Z = N(0, 1)$, entonces $\sigma_1 \cdot Z + \mu_1 = N(\mu_1, \sigma_1)$. Este hecho, motiva que consideremos el cambio lineal siguiente a las variables X e Y :

$$\begin{pmatrix} U \\ V \end{pmatrix} = \begin{pmatrix} \sigma_1 & 0 \\ 0 & \sigma_2 \end{pmatrix} \cdot \begin{pmatrix} X \\ Y \end{pmatrix} + \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}, \Rightarrow \begin{cases} U = \sigma_1 \cdot X + \mu_1, \\ V = \sigma_2 \cdot Y + \mu_2. \end{cases}$$

La función de densidad conjunta f_{UV} es:

$$\begin{aligned} f_{UV}(u, v) &= \left| \begin{pmatrix} \sigma_1 & 0 \\ 0 & \sigma_2 \end{pmatrix} \right| f_{XY}\left(\frac{u-\mu_1}{\sigma_1}, \frac{v-\mu_2}{\sigma_2}\right) = \frac{1}{\sigma_1 \cdot \sigma_2} f_{XY}\left(\frac{u-\mu_1}{\sigma_1}, \frac{v-\mu_2}{\sigma_2}\right) \\ &= \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} e^{-\frac{\left(\left(\frac{u-\mu_1}{\sigma_1}\right)^2 - 2\rho\left(\frac{u-\mu_1}{\sigma_1}\right)\left(\frac{v-\mu_2}{\sigma_2}\right) + \left(\frac{v-\mu_2}{\sigma_2}\right)^2\right)}{2(1-\rho^2)}}, \end{aligned}$$

para $(u, v) \in \mathbb{R}^2$.

Si llamamos a la matriz $= \begin{pmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix}$, llamada **matriz de covarianzas** de la distribución normal (U, V) la **función de densidad** anterior puede escribirse como:

$$f_{UV}(u, v) = \frac{1}{2\pi\sqrt{|\Gamma|}} e^{-\frac{1}{2}(\mathbf{u}-\mathbf{u}_0^\top)^{-1}(\mathbf{u}-\mathbf{u}_0)}, \text{ donde } \mathbf{u} = \begin{pmatrix} u \\ v \end{pmatrix} \text{ y } \mathbf{u}_0 = \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}.$$

La variable aleatoria bidimensional (X, Y) es la **variable aleatoria tipificada** con respecto de la variable (U, V) .

5.8.3. Transformaciones generales de variables aleatorias

Consideremos una variable aleatoria bidimensional continua (X, Y) con **función de densidad conjunta** f_{XY} .

Definimos la variable aleatoria bidimensional continua (U, V) a partir de una transformación general de la variable (X, Y) . Es decir, existen dos funciones de dos variables g_1 y g_2 tal que:

$$U = g_1(X, Y), \quad V = g_2(X, Y).$$

Vamos a suponer que las funciones g_1 y g_2 son invertibles, es decir, dados (u, v) , podemos encontrar (x, y) tal que $x = h_1(u, v)$ e $y = h_2(u, v)$. Las funciones h_1 y h_2 son las inversas de las funciones g_1 y g_2 , respectivamente.

La **función de densidad conjunta** f_{UV} se puede expresar de la forma siguiente en función de la **función de densidad conjunta** f_{XY} :

$$\begin{aligned} f_{UV}(u, v) &= \left| \det \begin{pmatrix} \frac{\partial h_1}{\partial u} & \frac{\partial h_1}{\partial v} \\ \frac{\partial h_2}{\partial u} & \frac{\partial h_2}{\partial v} \end{pmatrix} \right| f_{XY}(h_1(u, v), h_2(u, v)) \\ &= \frac{1}{\left| \det \begin{pmatrix} \frac{\partial g_1}{\partial x} & \frac{\partial g_1}{\partial y} \\ \frac{\partial g_2}{\partial x} & \frac{\partial g_2}{\partial y} \end{pmatrix} \right|} f_{XY}(h_1(u, v), h_2(u, v)). \end{aligned}$$

A la matriz $\begin{pmatrix} \frac{\partial g_1}{\partial x} & \frac{\partial g_1}{\partial y} \\ \frac{\partial g_2}{\partial x} & \frac{\partial g_2}{\partial y} \end{pmatrix}$ se le llama **matriz jacobiana del cambio** y a la matriz $\begin{pmatrix} \frac{\partial h_1}{\partial u} & \frac{\partial h_1}{\partial v} \\ \frac{\partial h_2}{\partial u} & \frac{\partial h_2}{\partial v} \end{pmatrix}$, **matriz jacobiana del cambio inverso**.

Ejemplo: cambio a polares

Sea (X, Y) una variable aleatoria bidimensional cuya **función de densidad conjunta** es:

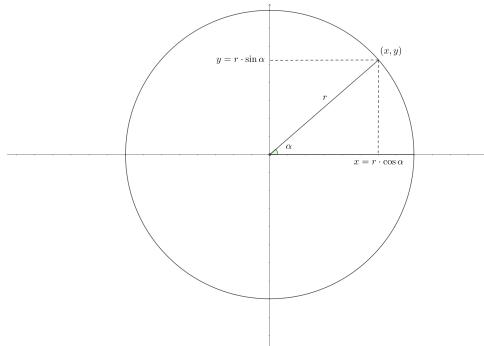
$$f_{XY}(x, y) = \begin{cases} \frac{2}{\pi} (x^2 + y^2), & \text{si } (x, y) \in D_1, \\ 0, & \text{en caso contrario,} \end{cases}$$

donde D_1 es el disco de radio 1:

$$D_1 = \{(x, y) \in \mathbb{R}^2, \mid x^2 + y^2 \leq 1\}.$$

El cambio a polares consiste en considerar las coordenadas polares (r, α) de un punto cualquiera (x, y) del plano, ver figura adjunta. El cambio que pasa de (r, α) a (x, y) (fijemos que es el cambio inverso, según nuestra notación o h_1 y h_2 , respectivamente) es:

$$x = h_1(r, \alpha) = r \cdot \cos \alpha, \quad y = h_2(r, \alpha) = r \cdot \sin \alpha.$$



Fijémonos que, con el cambio a polares, el disco unidad D_1 se transforma en el rectángulo $[0, 1] \times [0, 2\pi]$.

5.8. VARIABLES ALEATORIAS DEFINIDAS COMO FUNCIÓN DE DOS VARIABLES ALEATORIAS CONJUNTAS

Hallaremos el **jacobiano del cambio inverso**:

$$\det \begin{pmatrix} \frac{\partial h_1}{\partial u} & \frac{\partial h_1}{\partial v} \\ \frac{\partial h_2}{\partial u} & \frac{\partial h_2}{\partial v} \end{pmatrix} = \det \begin{pmatrix} \cos \alpha & -r \sin \alpha \\ \sin \alpha & r \cdot \cos \alpha \end{pmatrix} = r.$$

La **función de densidad conjunta** $f_{r\alpha}$ en las nuevas variables (polares) es:

$$f_{r\alpha}(r, \alpha) = r \cdot \frac{2}{\pi} ((r \cos \alpha)^2 + (r \sin \alpha)^2) = \frac{2}{\pi} \cdot r^3,$$

si $(r, \alpha) \in [0, 1] \times [0, 2\pi]$.

Podemos comentar que, gracias al cambio a polares, en este caso, es mucho más sencillo y cómodo trabajar con las variables (r, α) en vez de trabajar con las variables (x, y) por dos razones:

- La región donde la **función de densidad** no es nula es más simple, ya que trabajar con un rectángulo simplifica mucho más los cálculos que trabajar con un disco a la hora de hallar la **función de distribución**, **densidades marginales**, **densidades condicionadas**, **valores esperados**, etc.

Por ejemplo, comprobar que el área de la **función de densidad conjunta** $f_{r\alpha}$ da 1 es trivial:

$$\int_{r=0}^{r=1} \int_{\alpha=0}^{\alpha=2\pi} \frac{2}{\pi} r^3 d\alpha dr = \frac{2}{\pi} \cdot 2\pi \left[\frac{r^4}{4} \right]_{r=0}^{r=1} = 4 \cdot \frac{1}{4} = 1.$$

- La expresión de la **función de densidad** también es más simple, ya que sólo depende de la primera variable r ; sin embargo, la **función de densidad** inicial f_{XY} dependía de las dos variables x e y .

Capítulo 6

Vectores aleatorios

6.1. Varias variables aleatorias

En el capítulo anterior trabajamos con **variables aleatorias bidimensionales**

En este capítulo vamos a generalizar los conceptos introducidos para **variables aleatorias n -dimensionales**, con $n \geq 3$.

El ejemplo que comentamos en el capítulo de variables aleatorias de medir la temperatura media un día determinado del año durante 10 años es un ejemplo de variable aleatoria 10-dimensional.

6.1.1. Definición

La generalización de la noción de **variable aleatoria n -dimensional** a partir de la noción de **variable aleatoria bidimensional** es bastante obvia:

Definición de variable aleatoria n -dimensional: Dado un experimento aleatorio con **espacio muestral** Ω , definimos **variable aleatoria n -dimensional** $\mathbf{X} = (X_1, X_2, \dots, X_n)$ a toda aplicación

$$\begin{aligned}\mathbf{X} = (X_1, X_2, \dots, X_n) : \Omega &\longrightarrow \mathbb{R}^n \\ w &\longrightarrow \mathbf{X}(w) = (X_1(w), X_2(w), \dots, X_n(w)).\end{aligned}$$

Ejemplo: puertos entrada internet

Tenemos tres puertos de entrada de paquetes de internet.

Supongamos que cada milisegundo llega un paquete y el switch lo asigna a cada uno de los puertos con probabilidad $\frac{1}{3}$.

Estudiamos cómo se distribuyen los paquetes en 4 milisegundos.

Sea $\mathbf{X} = (X_1, X_2, X_3)$ la variable aleatoria 3-dimensional, donde X_i nos da el número de paquetes que ha recibido el puerto i -ésimo durante estos 4 milisegundos.

Por ejemplo, el suceso $\{X_1 \leq 1, X_2 \geq 3, X_3 \leq 1\}$ es $\{(0, 3, 0), (0, 3, 1), (0, 4, 0), (0, 4, 1), (1, 3, 0), (1, 3, 1), (1, 4, 0), (1, 4, 1)\}$.

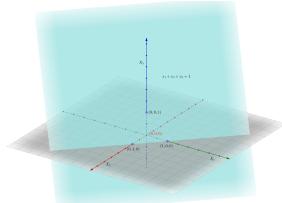
6.1.2. Representación del dominio de una variable aleatoria n -dimensional

Los sucesos que se derivan de una **variable aleatoria n -dimensional** están especificados por regiones del espacio n -dimensional.

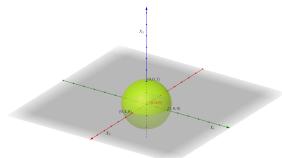
Veamos algunos ejemplos:

Suceso: $\{X_1 + X_2 + X_3 \leq 1\}$. En el gráfico siguiente, el plano $x_1 + x_2 + x_3 = 1$ separa el espacio en dos partes. Es la parte que corresponde al punto $(0, 0, 0)$.

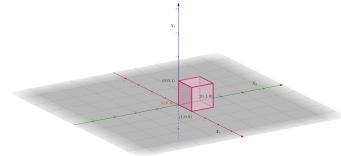
Si pensamos el plano anterior como un “espejo” es la parte de atrás del mismo.



Suceso: $\{X_1^2 + X_2^2 + X_3^2 \leq 1\}$. Es el interior de la esfera del gráfico siguiente:



Suceso: $\{0 \leq X_1 \leq 1, 0 \leq X_2 \leq 1, 0 \leq X_3 \leq 1\}$. Es el interior del cubo del gráfico siguiente:



La probabilidad de que la **variable n -dimensional** pertenezca a una cierta **región del n -espacio** $B \subset \mathbb{R}^n$ se define de la forma siguiente:

$$P((X_1, X_2, \dots, X_n) \in B) = P\{w \in \Omega, \mid (X_1(w), X_2(w), \dots, X_n(w)) \in B\},$$

dicho de otra manera, la probabilidad anterior es la probabilidad del suceso formado por los elementos de $w \in \Omega$ que cumplen que su **imagen** por la **variable aleatoria n -dimensional** (X_1, X_2, \dots, X_n) esté en B .

Por ejemplo, si consideramos $B = \{X_1 + X_2 + \dots + X_n \leq 1\}$, $P((X_1, X_2, \dots, X_n) \in B)$ es la probabilidad del suceso formado por los elementos w de Ω tal que la suma de las imágenes por X_i desde $i = 1$ hasta n sea menor o igual que 1: $X_1(w) + \dots + X_n \leq 1$.

6.2. Función de distribución conjunta

6.2.1. Definición

Dada una **variable aleatoria n -dimensional** (X_1, X_2, \dots, X_n) , queremos estudiar cómo se distribuye la probabilidad de sucesos cualesquiera de la forma $\{(X_1, X_2, \dots, X_n) \in B\}$, donde B es una región del espacio n -dimensional \mathbb{R}^n .

Para ello, definimos la **función de distribución conjunta**:

Definición de función de distribución conjunta: Dada una variable n -dimensional (X_1, X_2, \dots, X_n) , definimos su **función de distribución conjunta** $F_{X_1 \dots X_n}$ a la función definida sobre \mathbb{R}^n de la manera siguiente:

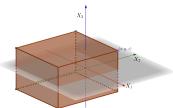
$$\begin{aligned} F_{X_1 \dots X_n} : \mathbb{R}^n &\longrightarrow \mathbb{R} \\ (x_1, \dots, x_n) &\longrightarrow F_{X_1 \dots X_n}(x_1, \dots, x_n) = P(X_1 \leq x_1, \dots, X_n \leq x_n). \end{aligned}$$

Es decir, dado un valor $(x_1, \dots, x_n) \in \mathbb{R}^n$, consideramos la región del espacio n -dimensional $(-\infty, x_1] \times \dots \times (-\infty, x_n]$.

Entonces la **función de distribución conjunta** en el valor (x_1, \dots, x_n) es la probabilidad del suceso formado por aquellos elementos tal que la imagen por la **variable aleatoria n -dimensional** (X_1, X_2, \dots, X_n) caen dentro de la región anterior:

$$\begin{aligned} F_{X_1 \dots X_n}(x_1, \dots, x_n) &= P\{w \in \Omega, \mid (X_1(w), \dots, X_n(w)) \\ &\quad \in (-\infty, x_1] \times \dots \times (-\infty, x_n]\} \\ &= P\{w \in \Omega, \mid X_1(w) \leq x_1, \dots, X_n(w) \leq x_n\}. \end{aligned}$$

El gráfico siguiente muestra el conjunto $(-\infty, x_1] \times (-\infty, x_2] \times (-\infty, x_3]$ en \mathbb{R}^3 para un valor (x, y, z) :



6.2.2. Propiedades

Sea (X_1, X_2, \dots, X_n) una variable n -dimensional. Sean $F_{X_1 \dots X_n}$ su **función de distribución conjunta**. Dicha función satisface las propiedades siguientes:

- La función de distribución conjunta es no decreciente en cada una de las variables:

Si $x_i \leq x'_i$, para todo i , entonces, $F_{X_1 \dots X_n}(x_1, \dots, x_n) \leq F_{X_1 \dots X_n}(x'_1, \dots, x'_n)$.

- $F_{X_1 \dots X_n}(x_1, \dots, x_{i-1}, -\infty, x_{i+1}, \dots, x_n) \stackrel{(i)}{=} 0$, para todo i y $F_{X_1 \dots X_n}(\infty, \dots, \infty) = 1$, para todo $x_1, \dots, x_n \in \mathbb{R}$.

- Las variables aleatorias X_1, \dots, X_n se llaman **variables aleatorias marginales** y sus funciones de distribución F_{X_1}, \dots, F_{X_n} pueden hallarse de la forma siguiente como función de la **función de distribución conjunta** $F_{X_1 \dots X_n}$:

$$F_{X_i}(x_i) = F_{X_1 \dots X_n}(\infty, \dots, \infty, \overset{(i)}{x_i}, \infty, \dots, \infty),$$

para todo $x_1, \dots, x_n \in \mathbb{R}$ y para todo $i = 1, \dots, n$.

- La función de distribución conjunta es continua por la derecha en todas las variables x_i :

$$\begin{aligned} & \lim_{x_i \rightarrow a^+} F_{X_1 \dots X_n}(x_1, \dots, x_{i-1}, \overset{(i)}{x_i}, x_{i+1}, \dots, x_n) \\ &= \lim_{x_i \rightarrow a, x_i > a} F_{X_1 \dots X_n}(x_1, \dots, x_{i-1}, \overset{(i)}{x_i}, x_{i+1}, \dots, x_n) \\ &= F_{X_1 \dots X_n}(x_1, \dots, x_{i-1}, \overset{(i)}{a}, x_{i+1}, \dots, x_n), \end{aligned}$$

para todo $a \in \mathbb{R}$ y para todo $i = 1, \dots, n$.

Ejemplo: densidad tridimensional

Consideremos una variable aleatoria 3-dimensional (X_1, X_2, X_3) con **función de distribución conjunta**:

$$F_{X_1 X_2 X_3}(x_1, x_2, x_3) = \begin{cases} 0, & \text{si } x_1 < 0, \text{ o } x_2 < 0, \text{ o } x_3 < 0 \\ x_1^2 \cdot x_2^2 \cdot x_3^2, & \text{si } 0 \leq x_1 \leq 1, 0 \leq x_2 \leq 1, 0 \leq x_3 \leq 1, \\ x_2^2 \cdot x_3^2, & \text{si } x_1 > 1, 0 \leq x_2 \leq 1, 0 \leq x_3 \leq 1, \\ x_1^2 \cdot x_3^2, & \text{si } 0 \leq x_1 \leq 1, x_2 > 1, 0 \leq x_3 \leq 1, \\ x_3^2, & \text{si } x_1 > 1, x_2 > 1, 0 \leq x_3 \leq 1, \\ x_1^2 \cdot x_2^2, & \text{si } 0 \leq x_1 \leq 1, 0 \leq x_2 \leq 1, x_3 > 1, \\ x_1^2, & \text{si } 0 \leq x_1 \leq 1, x_2 > 1, x_3 > 1, \\ x_2^2, & \text{si } x_1 > 1, 0 \leq x_2 \leq 1, x_3 > 1, \\ 1, & \text{si } x_1 \geq 1, x_2 \geq 1, x_3 \geq 1. \end{cases}$$

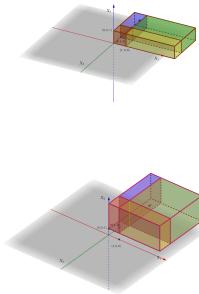
En las figuras siguientes, hemos representado por zonas cómo está definida $F_{X_1 X_2 X_3}$.

La primera figura muestra las zonas en la “planta baja” o para $0 \leq x_3 \leq 1$. En color marrón, está representada la región $0 \leq x_1 \leq 1, 0 \leq x_2 \leq 1, 0 \leq x_3 \leq 1$, en color amarillo, la región $x_1 > 1, 0 \leq x_2 \leq 1, 0 \leq x_3 \leq 1$, en color verde, la región $x_1 > 1, x_2 > 1, 0 \leq x_3 \leq 1$ y en color violeta, la región $0 \leq x_1 \leq 1, x_2 > 1, 0 \leq x_3 \leq 1$.

La segunda figura muestra las zonas del “primer piso” o para $x_3 > 1$. Los colores tienen un significado similar a los de la primera figura: en color marrón, está representada la región $0 \leq x_1 \leq 1, 0 \leq x_2 \leq 1, x_3 > 1$, en color amarillo u ocre, la región $x_1 > 1, 0 \leq x_2 \leq 1, x_3 > 1$, en color verde, la región $x_1 > 1, x_2 > 1, x_3 > 1$ y en color violeta, la región $0 \leq x_1 \leq 1, x_2 > 1, x_3 > 1$.

Comprobemos algunas de las propiedades que hemos enunciado anteriormente:

- Claramente $F_{X_1 X_2 X_3}(x_1, x_2, -\infty) = F_{X_1 X_2 X_3}(x_1, -\infty, x_3) = F_{X_1 X_2 X_3}(-\infty, x_2, x_3) = 0$ ya que $F_{X_1 X_2 X_3}(x_1, x_2, x_3) = 0$ si $x_1 < 0$ o $x_2 < 0$ o $x_3 < 0$. Por tanto, si hacemos tender



x_1 o x_2 o x_3 hacia $-\infty$, obtendremos que $F_{X_1 X_2 X_3}(x_1, x_2, -\infty) = F_{X_1 X_2 X_3}(x_1, -\infty, x_3) = F_{X_1 X_2 X_3}(-\infty, x_2, x_3) = 0$.

- De la misma manera $F_{X_1 X_2 X_3}(\infty, \infty, \infty) = 1$ ya que $F_{X_1 X_2 X_3}(x_1, x_2, x_3) = 1$ para $x_1 > 1, x_2 > 1$ y $x_3 > 1$. Por tanto, si hacemos tender x_1, x_2 y x_3 hacia ∞ , obtendremos $F_{X_1 X_2 X_3}(\infty, \infty, \infty) = 1$.
- Hallemos las marginales:

$$F_{X_1}(x_1) = F_{X_1 X_2 X_3}(x_1, \infty, \infty) = \begin{cases} 0, & \text{si } x_1 < 0, \\ x_1, & \text{si } 0 \leq x_1 \leq 1, \\ 1, & \text{si } x_1 > 1. \end{cases}$$

Para ver la expresión anterior basta trazar el plano $X_1 = x_1$ en el gráfico anterior y ver hacia dónde tiende a medida que las variables x_2 y x_3 se van hacia ∞ .

¿Habéis averiguado cuál es la distribución de X_1 ?

¡Efectivamente!, X_1 es la uniforme en el intervalo $(0, 1)$.

Dejamos como ejercicio hallar la distribución marginal para las variables X_2 e X_3 .

- Comprobemos que $F_{X_1 X_2 X_3}$ es continua por la derecha para las variables x_1, x_2 y x_3 en el punto $(1, 1, 1)$ que es un punto problemático:

$$\begin{aligned} \lim_{x_1 \rightarrow 1, x_1 > 1} F_{X_1 X_2 X_3}(x_1, 1, 1) &= \lim_{x_1 \rightarrow 1, x_1 > 1} 1 = F_{X_1 X_2 X_3}(1, 1, 1), \\ \lim_{x_2 \rightarrow 1, x_2 > 1} F_{X_1 X_2 X_3}(1, x_2, 1) &= \lim_{x_2 \rightarrow 1, x_2 > 1} 1 = F_{X_1 X_2 X_3}(1, 1, 1), \\ \lim_{x_3 \rightarrow 1, x_3 > 1} F_{X_1 X_2 X_3}(1, 1, x_3) &= \lim_{x_3 \rightarrow 1, x_3 > 1} 1 = F_{X_1 X_2 X_3}(1, 1, 1). \end{aligned}$$

Ejemplo: cálculo con R distribución conjunta

Realizar un gráfico 3D de la **función de distribución conjunta** no es posible ya que deberíamos pasar a \mathbb{R}^4 .

Lo que sí es posible es dibujar las curvas de nivel de dicha función para un valor de x_3 fijado.

En los gráficos siguientes dibujamos las curvas de nivel para $x_3 = 0, 0,5, 1$ i $x_3 = 1,5$.

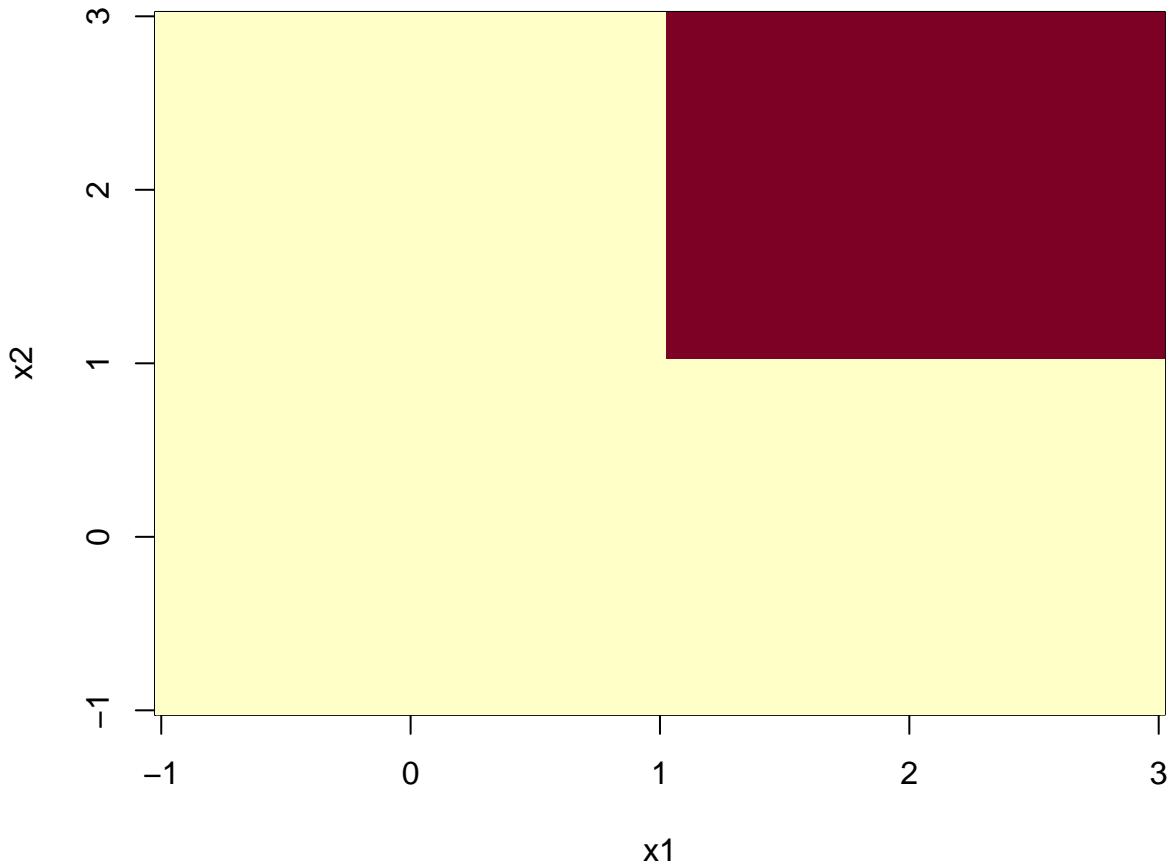
Primero definimos la **función** y luego la dibujamos para x_1 y x_2 entre -1 y 3 :

```

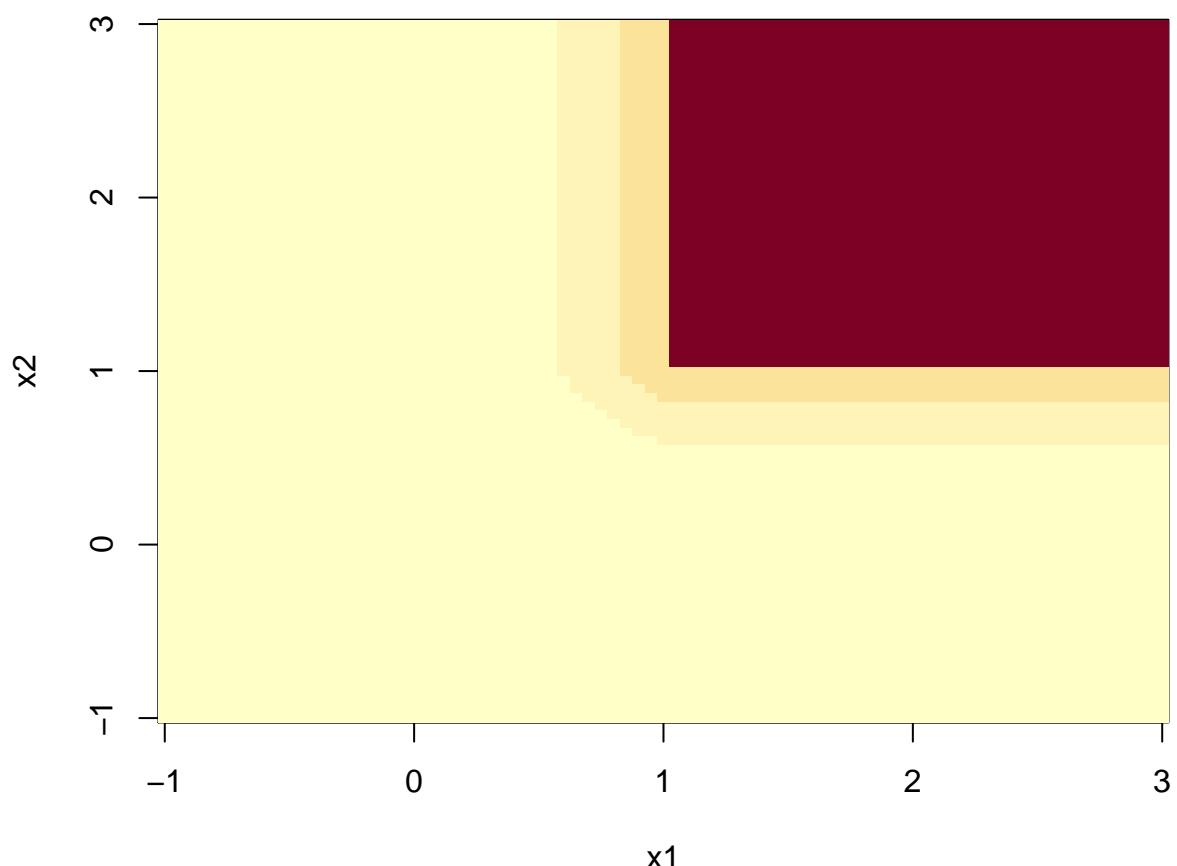
f.dist.con = function(x1,x2,x3){ifelse(x1<0 | x2<0 | x3 <0,0,
  ifelse(x1>=0 & x1<=1 & x2>=0 & x2<=1 & x3>=0 & x3<=1,
    x1^2*x2^2*x3^2,
  ifelse(x1>1 & x2>=0 & x2<=1 & x3>=0 & x3<=1,
    x2^2*x3^2,
  ifelse(x1>=0 & x1<=1 & x2>1 & x3>=0 & x3<=1,
    x1^2*x3^2,
  ifelse(x1>=0 & x1<=1 & x2>=0 & x2<=1 & x3>1,
    x1^2*x2^2,
  ifelse(x1>=0 & x1<=1 & x2 >1 & x3 >1,x1^2,
  ifelse(x1>1 & x2 >=0 & x2<=1 & x3 >1,x2^2,
  ifelse(x1>=0 & x1<=1 & x2>=0 & x2<=1 & x3 >1,
    x3^2,1))))))}

x1=seq(from=-1,to=3,by=0.05)
x2=seq(from=-1,to=3,by=0.05)
curva.nivel.0=outer(x1,x2,f.dist.con,x3=0)
curva.nivel.0.5=outer(x1,x2,f.dist.con,x3=0.5)
curva.nivel.1=outer(x1,x2,f.dist.con,x3=1)
curva.nivel.1.5=outer(x1,x2,f.dist.con,x3=1.5)
image(x1,x2,curva.nivel.0)

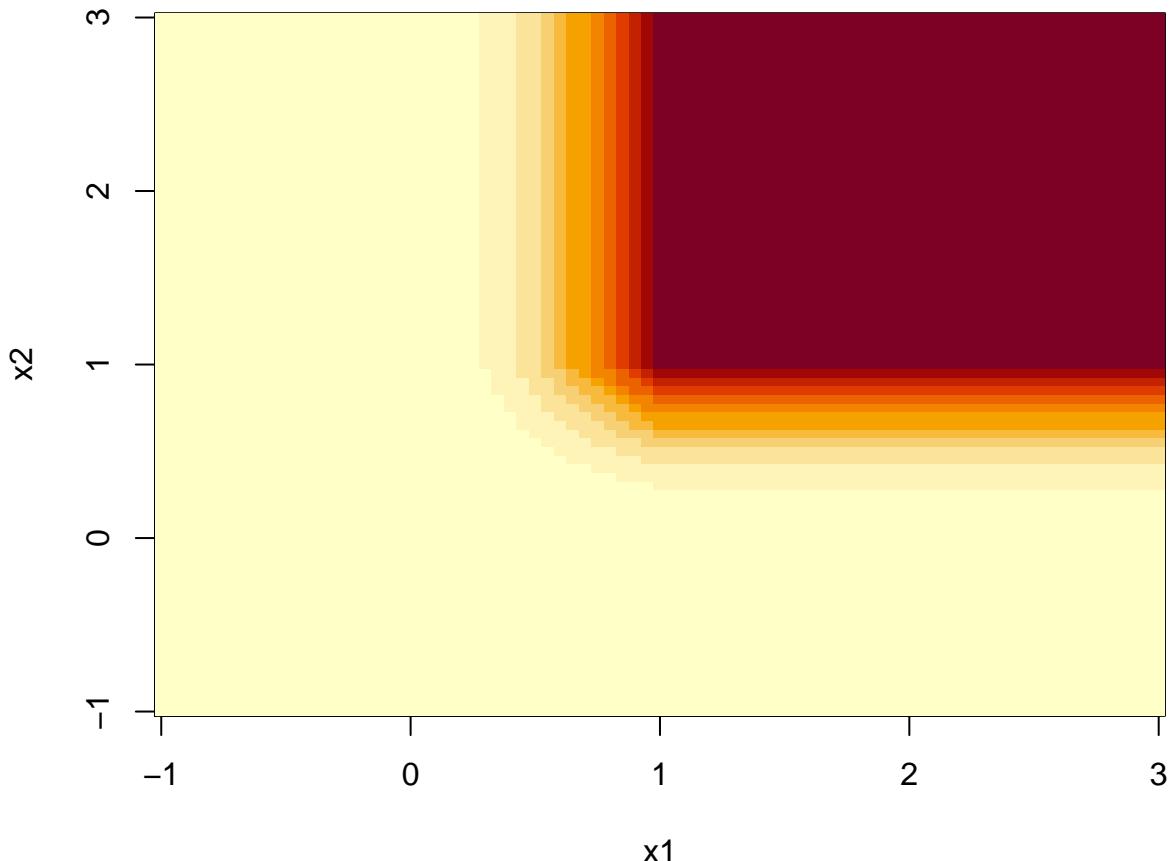
```



```
image(x1,x2,curva.nivel.0.5)
```



```
image(x1,x2,curva.nivel.1)
```



Ejemplo del lanzamiento de un dado tres veces

Consideremos el experimento aleatorio que consiste en lanzar un dado tres veces.

El espacio Ω de resultados es:

$$\Omega = \{(i, j, k), |i, j, k = 1, 2, 3, 4, 5, 6\}.$$

En total tendremos $6 \cdot 6 \cdot 6 = 6^3 = 216$ resultados posibles.

Consideremos la variable 3-dimensional $\mathbf{X} = (X_1, X_2, X_3)$, donde X_1 nos da el número de 1's obtenidos, X_2 , el número de 2's y X_3 , el número de 3's.

El conjunto $\mathbf{X}(\Omega)$ tiene en total 64 elementos ya que cada componente X_i puede tener en total 4 resultados: 0, 1, 2 o 3. Por tanto el conjunto total de resultados es: $4 \cdot 4 \cdot 4 = 4^3 = 64$.

El valor de función de distribución conjunta en el resultado $(0, 0, 0)$ es:

$$F_{X_1 X_2 X_3}(0, 0, 0) = p(X_1 \leq 0, X_2 \leq 0, X_3 \leq 0) = \frac{3^3}{6^3} = \left(\frac{1}{2}\right)^3 = 0,125,$$

ya que si $X_1 \leq 0$, $X_2 \leq 0$ y $X_3 \leq 0$, significa que no ha salido ni ningún 1, ni ningún 2 ni ningún 3. Sólo pueden salir 4's, 5's o 6's y existen $3 \cdot 3 \cdot 3 = 3^3 = 27$ posibilidades de que esto pase entre $6^3 = 216$ posibilidades posibles.

6.3. Variables aleatorias n -dimensionales discretas

Definición de variable aleatoria n -dimensional discreta: Sea (X_1, \dots, X_n) una **variable aleatoria n -dimensional**. Diremos que es discreta cuando su conjunto de valores en \mathbb{R}^n , $(X_1, \dots, X_n)(\Omega)$ es un conjunto finito o numerable.

En la mayoría de los casos, dicho conjunto es un subconjunto de los enteros naturales.

Ejemplo

La variable aleatoria 3-dimensional anterior que nos daba el número de 1's obtenidos, el número de 2's y el número de 3's es discreta ya que

$$\begin{aligned} \mathbf{X}(\Omega) = & \{(0, 0, 0), (1, 0, 0), (0, 1, 0), (0, 0, 1), (2, 0, 0), (0, 2, 0), \\ & (0, 0, 2), (3, 0, 0), (0, 3, 0), (0, 0, 3), (0, 1, 1), (1, 0, 1), \\ & (1, 1, 0), (0, 1, 2), (0, 2, 1), (1, 0, 2), (2, 0, 1), (1, 2, 0), \\ & (2, 1, 0), (0, 1, 3), (0, 3, 1), (1, 0, 3), (3, 0, 1), (1, 3, 0), \\ & (3, 1, 0), (0, 2, 2), (2, 0, 2), (2, 2, 0), (0, 2, 3), (3, 2, 0), \\ & (2, 0, 3), (3, 0, 2), (2, 3, 0), (3, 2, 0), (0, 3, 3), (3, 0, 3), \\ & (3, 3, 0), (1, 1, 1), (1, 1, 2), (1, 2, 1), (2, 1, 1), (1, 1, 3), \\ & (1, 3, 1), (3, 1, 1), (1, 2, 2), (2, 1, 2), (2, 2, 1), (1, 2, 3), \\ & (2, 1, 3), (1, 3, 2), (3, 1, 2), (2, 3, 1), (3, 2, 1), (1, 3, 3), \\ & (3, 1, 3), (3, 3, 1), (2, 2, 2), (2, 2, 3), (2, 3, 2), (3, 2, 2), \\ & (2, 3, 3), (3, 2, 3), (3, 3, 2), (3, 3, 3)\}. \end{aligned}$$

6.3.1. Función de probabilidad conjunta

Definición de función de probabilidad conjunta: Dada una **variable aleatoria n -dimensional discreta** (X_1, \dots, X_n) con $(X_1, \dots, X_n)(\Omega) = \{(x_{i_1}, x_{i_2}, \dots, x_{i_n}), i_1 = 1, 2, \dots, i_n = 1, 2, \dots\}$, definimos la función de probabilidad discreta $P_{X_1 \dots X_n}$ para un valor $(x_{i_1}, x_{i_2}, \dots, x_{i_n}) \in \mathbb{R}^n$ de la siguiente forma:

$$\begin{aligned} P_{X_1 \dots X_n} : \mathbb{R}^n &\longrightarrow \mathbb{R} \\ (x_{i_1}, x_{i_2}, \dots, x_{i_n}) &\longrightarrow P_{X_1 \dots X_n}(x_{i_1}, x_{i_2}, \dots, x_{i_n}) = P(X = x_{i_1}, \dots, X_n = x_{i_n}). \end{aligned}$$

Observación: Si $(x_{i_1}, x_{i_2}, \dots, x_{i_n}) \notin (X_1, \dots, X_n)(\Omega)$, el valor de la **función de probabilidad conjunta** en $(x_{i_1}, x_{i_2}, \dots, x_{i_n})$ es nulo: $P_{X_1 \dots X_n}(x_{i_1}, x_{i_2}, \dots, x_{i_n}) = 0$, ya que, en este caso, el conjunto $\{w \in \Omega, |(X_1(w), \dots, X_n(w)) = (x_{i_1}, x_{i_2}, \dots, x_{i_n})\} = \emptyset$ ya que recordemos $(x_{i_1}, x_{i_2}, \dots, x_{i_n}) \notin (X_1, \dots, X_n)(\Omega)$.

Por tanto, de cara a calcular $P_{X_1 \dots X_n}$ basta calcular $P_{X_1 \dots X_n}(x_{i_1}, \dots, x_{i_n})$ para $(x_{i_1}, \dots, x_{i_n}) \in (X_1, \dots, X_n)(\Omega)$.

Los valores de $P_{X_1 \dots X_n}(x_{i_1}, \dots, x_{i_n})$ estarían organizados en una tabla n -dimensional.

Ejemplo de la variable 3-dimensional que nos da el número de 1's, 2's y 3's en el lanzamiento de un dado tres veces

Para mostrar la **función de probabilidad conjunta** haremos una tabla bidimensional para cada valor de X_3 .

Como $X_3(\Omega) = \{0, 1, 2, 3\}$, en total mostraremos 4 tablas bidimensionales.

Tabla para $X_3 = 0$:

X_1/X_2	0	1	2	3
0	$\frac{1}{8}$	$\frac{1}{8}$	$\frac{1}{24}$	$\frac{1}{216}$
1	$\frac{1}{8}$	$\frac{1}{12}$	$\frac{1}{72}$	0
2	$\frac{1}{24}$	$\frac{1}{72}$	0	0
3	$\frac{1}{216}$	0	0	0

Tabla para $X_3 = 1$:

X_1/X_2	0	1	2	3
0	$\frac{1}{8}$	$\frac{1}{12}$	$\frac{1}{72}$	0
1	$\frac{1}{12}$	$\frac{1}{36}$	0	0
2	$\frac{1}{72}$	0	0	0
3	0	0	0	0

Tabla para $X_3 = 2$:

X_1/X_2	0	1	2	3
0	$\frac{1}{24}$	$\frac{1}{72}$	0	0
1	$\frac{1}{72}$	0	0	0
2	0	0	0	0
3	0	0	0	0

Tabla para $X_3 = 3$:

X_1/X_2	0	1	2	3
0	$\frac{1}{216}$	0	0	0
1	0	0	0	0
2	0	0	0	0
3	0	0	0	0

La función `fun.prod.con` nos da la **función de probabilidad conjunta** de la variable aleatoria \mathbf{X} cuando lanzamos un dado tres veces:

```
fun.prob.con=function(x1,x2,x3){
  n=6
  cuenta.1 =function(x){length(x[x==1])}
  cuenta.2 =function(x){length(x[x==2])}
  cuenta.3 =function(x){length(x[x==3])}
  Dxyz=data.frame(d1=rep(1:n,each=n),d2=rep(1:n,times=n),
                   d3=rep(1:n,each=n*n))
  X1=apply(Dxyz,1,cuenta.1)
  X2=apply(Dxyz,1,cuenta.2)
  X3=apply(Dxyz,1,cuenta.3)
  frecuencia = table(X1==x1 & X2==x2 & X3==x3)
  res=ifelse(length(frecuencia)==2,frecuencia[2],0)
  return(res/6^3)
}
```

Para construir la tabla de la **función de probabilidad conjunta** para la variable $\mathbf{X} = (X_1, X_2, X_3)$ con $X_3 = 0$ hacemos lo siguiente:

```
valores.variables=0:3
tabla.0 = c()
for (i in 1:length(valores.variables)){
  for (j in 1:length(valores.variables)){
    tabla.0=c(tabla.0,
              fun.prob.con(valores.variables[i],
                           valores.variables[j],0));
  }
}
tabla.0 = matrix(tabla.0,length(valores.variables),
                 length(valores.variables))
rownames(tabla.0)=valores.variables
colnames(tabla.0)=valores.variables
knitr::kable(tabla.0)
```

Con los demás valores de X_3 , lo haríamos de forma similar.

Tabla con $X_3 = 0$:

	0	1	2	3
0	0.1250000	0.1250000	0.0416667	0.0046296
1	0.1250000	0.0833333	0.0138889	0.0000000
2	0.0416667	0.0138889	0.0000000	0.0000000
3	0.0046296	0.0000000	0.0000000	0.0000000

Tabla con $X_3 = 1$:

	0	1	2	3
0	0.1250000	0.0833333	0.0138889	0
1	0.0833333	0.0277778	0.0000000	0
2	0.0138889	0.0000000	0.0000000	0
3	0.0000000	0.0000000	0.0000000	0

Tabla con $X_3 = 2$:

	0	1	2	3
0	0.0416667	0.0138889	0	0
1	0.0138889	0.0000000	0	0
2	0.0000000	0.0000000	0	0
3	0.0000000	0.0000000	0	0

Tabla con $X_3 = 3$:

	0	1	2	3
0	0.0046296	0	0	0
1	0.0000000	0	0	0
2	0.0000000	0	0	0
3	0.0000000	0	0	0

6.3.2. Propiedades de la función de probabilidad conjunta

Dada (X_1, \dots, X_n) una **variable aleatoria n -dimensional discreta** con conjunto de valores $(X_1, \dots, X_n)(\Omega) = \{(x_{i_1}, \dots, x_{i_n}) \mid i_1 = 1, 2, \dots, i_n = 1, 2, \dots\}$. Entonces su **función de probabilidad conjunta** verifica las propiedades siguientes:

La suma de todos los valores de la **función de probabilidad conjunta** sobre el conjunto de valores siempre vale 1:

$$\sum_{i_1} \cdots \sum_{i_n} P_{X_1 \dots X_n}(x_{i_1}, \dots, x_{i_n}) = 1.$$

Sea B una región del espacio \mathbb{R}^n . El valor de la probabilidad $P((X_1, \dots, X_n) \in B)$ se puede calcular de la forma siguiente:

$$P((X_1, \dots, X_n) \in B) = \sum_{(x_{i_1}, \dots, x_{i_n}) \in B} P_{X_1 \dots X_n}(x_{i_1}, \dots, x_{i_n}).$$

Es decir, la probabilidad de que la variable n -dimensional tome valores en B es igual a la suma de todos aquellos valores de la función de probabilidad conjunta que están en B .

En particular, tenemos la siguiente expresión relaciona la **función de distribución conjunta** con la **función de probabilidad conjunta**:

$$F_{X_1 \dots X_n}(x_1, \dots, x_n) = \sum_{x_{i_1} \leq x_1, \dots, x_{i_n} \leq x_n} P_{X_1 \dots X_n}(x_{i_1}, \dots, x_{i_n}).$$

Dicha expresión se deduce de la expresión anterior considerando $B = (-\infty, x_1] \times \cdots \times (-\infty, x_n]$.

Ejemplo anterior del lanzamiento de un dado tres veces

Ejercicio

Comprobad usando la tabla de la función de probabilidad conjunta que la suma de todos sus valores suma 1.

Apliquemos la fórmula que relaciona la función de distribución conjunta con la función de probabilidad conjunta para $(x_1, x_2, x_3) = (1, 2, 2)$.

Observamos que los únicos valores $(x_{i_1}, x_{i_2}, x_{i_3}) \in (X_1 X_2, X_3)(\Omega)$ que verifican $x_{i_1} \leq 1$, $x_{i_2} \leq 2$ y $x_{i_3} \leq 2$ son $(0, 0, 0)$, $(0, 0, 1)$, $(0, 0, 2)$, $(0, 1, 0)$, $(0, 1, 1)$, $(0, 1, 2)$, $(0, 2, 0)$, $(0, 2, 1)$, $(0, 2, 2)$, $(1, 0, 0)$, $(1, 0, 1)$, $(1, 0, 2)$, $(1, 1, 0)$, $(1, 1, 1)$, $(1, 1, 2)$, $(1, 2, 0)$, $(1, 2, 1)$ y $(1, 2, 2)$:

$$\begin{aligned}
 F_{X_1 X_2 X_3}(1, 2, 2) &= P_{X_1 X_2 X_3}(0, 0, 0) + P_{X_1 X_2 X_3}(0, 0, 1) + P_{X_1 X_2 X_3}(0, 0, 2) \\
 &\quad + P_{X_1 X_2 X_3}(0, 1, 0) + P_{X_1 X_2 X_3}(0, 1, 1) + P_{X_1 X_2 X_3}(0, 1, 2) \\
 &\quad + P_{X_1 X_2 X_3}(0, 2, 0) + P_{X_1 X_2 X_3}(0, 2, 1) + P_{X_1 X_2 X_3}(0, 2, 2) \\
 &\quad + P_{X_1 X_2 X_3}(1, 0, 0) + P_{X_1 X_2 X_3}(1, 0, 1) + P_{X_1 X_2 X_3}(1, 0, 2) \\
 &\quad + P_{X_1 X_2 X_3}(1, 1, 0) + P_{X_1 X_2 X_3}(1, 1, 1) + P_{X_1 X_2 X_3}(1, 1, 2) \\
 &\quad + P_{X_1 X_2 X_3}(1, 2, 0) + P_{X_1 X_2 X_3}(1, 2, 1) + P_{X_1 X_2 X_3}(1, 2, 2) \\
 &= \frac{1}{8} + \frac{1}{8} + \frac{1}{24} + \frac{1}{8} + \frac{1}{12} + \frac{1}{72} + \frac{1}{24} + \frac{1}{72} + 0 \\
 &\quad + \frac{1}{8} + \frac{1}{12} + \frac{1}{72} + \frac{1}{12} + \frac{1}{36} + 0 + \frac{1}{72} + 0 + 0 = \\
 &= \frac{11}{12} = 0,9167.
 \end{aligned}$$

La función de distribución conjunta es con código R:

```
fun.dis.con = function(x1, x2, x3){
  suma=0
  i1=0; i2=0; i3=0;
  while(i1 <=x1 & i1<=3){
    while(i2 <= x2 & i2<=3){
      while(i3<= x3 & i3 <=3){
        suma=suma+fun.prob.con(i1,i2,i3); i3=i3+1;
      }
      i3=0; i2=i2+1;
    }
    i2=0; i3=0; i1=i1+1;
  }
  return(suma)
}
```

Comprobemos que la **función de distribución conjunta** en el valor $(1, 2, 2)$ nos da el mismo resultado que vimos anteriormente:

```
fun.dis.con(1, 2, 2)
```

```
## [1] 0.9166667
```

6.3.3. Funciones de distribución marginales

Consideremos una variable aleatoria n -dimensional discreta $(X_1 \dots, X_n)$ con **función de probabilidad conjunta**

$P_{X_1 \dots, X_n}(x_{i_1}, \dots, x_{i_n})$, con $(x_{i_1}, \dots, x_{i_n}) \in (X_1 \dots, X_n)(\Omega)$, $i_1 = 1, 2, \dots$, $i_n = 1, 2, \dots$.

La tabla de la **función de probabilidad conjunta** contiene suficiente información para obtener las **funciones de probabilidad conjunta** de cualquier **variable aleatoria k -dimensional** $(X_{s_1}, \dots, X_{s_k})$ donde $\{s_1, \dots, s_k\}$ es un subconjunto de las componentes $\{1, \dots, n\}$ de la variable aleatoria n -dimensional.

Proposición. Expresión de las funciones de probabilidad marginales.

Sea $\mathbf{X} = (X_1, \dots, X_n)$ una variable aleatoria n -dimensional discreta con función de probabilidad conjunta $P_{X_1 \dots X_n}(x_{i_1}, \dots, x_{i_n})$, con $(x_{i_1}, \dots, x_{i_n}) \in (X_1, \dots, X_n)(\Omega)$, $i_1 = 1, 2, \dots, i_n = 1, 2, \dots$

Sea $\{s_1, \dots, s_k\}$ un subconjunto del conjunto de las componentes $\{1, \dots, n\}$ donde suponemos que $s_1 < s_2 < \dots < s_k$. Entonces la función de probabilidad conjunta $P_{X_{s_1} \dots X_{s_k}}$ de la variable aleatoria k -dimensional $(X_{s_1}, \dots, X_{s_k})$ se calcula usando la expresión siguiente:

$$P_{X_{s_1} \dots X_{s_k}}(x_{s_1}, \dots, x_{s_k}) = \sum_{x_t} P_{X_1 \dots X_n}(\mathbf{x}_s, \mathbf{x}_t),$$

donde $(\mathbf{x}_s, \mathbf{x}_t)$ es un valor de $\mathbf{X}(\Omega)$ tal que tiene como componente s_i el valor x_{s_i} para $i = 1, \dots, k$ y con \mathbf{x}_t queremos decir todas las demás componentes que no son las s_i . La suma tiene todos los sumandos \mathbf{x}_t para los que se cumpla que $(\mathbf{x}_s, \mathbf{x}_t) \in \mathbf{X}(\Omega)$.

En el caso particular de $n = 3$, podemos calcular las funciones de probabilidad conjunta de las variables unidimensionales X_1 , X_2 y X_3 y de las variables bidimensionales (X_1, X_2) , (X_1, X_3) y (X_2, X_3) .

Ejemplo de la variable 3-dimensional que nos da el número de 1's, 2's y 3's en el lanzamiento de un dado tres veces

Hallemos, en primer lugar, la función de probabilidad marginal de las variables X_1 , X_2 y X_3 . Empecemos con X_1 :

$$\begin{aligned} P_{X_1}(0) &= P_{X_1 X_2 X_3}(0, 0, 0) + P_{X_1 X_2 X_3}(0, 0, 1) + P_{X_1 X_2 X_3}(0, 0, 2) \\ &\quad + P_{X_1 X_2 X_3}(0, 0, 3) + P_{X_1 X_2 X_3}(0, 1, 0) + P_{X_1 X_2 X_3}(0, 1, 1) \\ &\quad + P_{X_1 X_2 X_3}(0, 1, 2) + P_{X_1 X_2 X_3}(0, 1, 3) + P_{X_1 X_2 X_3}(0, 2, 0) \\ &\quad + P_{X_1 X_2 X_3}(0, 2, 1) + P_{X_1 X_2 X_3}(0, 2, 2) + P_{X_1 X_2 X_3}(0, 2, 3) \\ &\quad + P_{X_1 X_2 X_3}(0, 3, 0) + P_{X_1 X_2 X_3}(0, 3, 1) + P_{X_1 X_2 X_3}(0, 3, 2) \\ &\quad + P_{X_1 X_2 X_3}(0, 3, 3) = 0,5787, \\ P_{X_1}(1) &= P_{X_1 X_2 X_3}(1, 0, 0) + P_{X_1 X_2 X_3}(1, 0, 1) + P_{X_1 X_2 X_3}(1, 0, 2) \\ &\quad + P_{X_1 X_2 X_3}(1, 0, 3) + P_{X_1 X_2 X_3}(1, 1, 0) + P_{X_1 X_2 X_3}(1, 1, 1) \\ &\quad + P_{X_1 X_2 X_3}(1, 1, 2) + P_{X_1 X_2 X_3}(1, 1, 3) + P_{X_1 X_2 X_3}(1, 2, 0) \\ &\quad + P_{X_1 X_2 X_3}(1, 2, 1) + P_{X_1 X_2 X_3}(1, 2, 2) + P_{X_1 X_2 X_3}(1, 2, 3) \\ &\quad + P_{X_1 X_2 X_3}(1, 3, 0) + P_{X_1 X_2 X_3}(1, 3, 1) + P_{X_1 X_2 X_3}(1, 3, 2) \\ &\quad + P_{X_1 X_2 X_3}(1, 3, 3) = 0,3472, \end{aligned}$$

$$\begin{aligned}
P_{X_1}(2) &= P_{X_1 X_2 X_3}(2, 0, 0) + P_{X_1 X_2 X_3}(2, 0, 1) + P_{X_1 X_2 X_3}(2, 0, 2) \\
&\quad + P_{X_1 X_2 X_3}(2, 0, 3) + P_{X_1 X_2 X_3}(2, 1, 0) + P_{X_1 X_2 X_3}(2, 1, 1) \\
&\quad + P_{X_1 X_2 X_3}(2, 1, 2) + P_{X_1 X_2 X_3}(2, 1, 3) + P_{X_1 X_2 X_3}(2, 2, 0) \\
&\quad + P_{X_1 X_2 X_3}(2, 2, 1) + P_{X_1 X_2 X_3}(2, 2, 2) + P_{X_1 X_2 X_3}(2, 2, 3) \\
&\quad + P_{X_1 X_2 X_3}(2, 3, 0) + P_{X_1 X_2 X_3}(2, 3, 1) + P_{X_1 X_2 X_3}(2, 3, 2) \\
&\quad + P_{X_1 X_2 X_3}(2, 3, 3) = 0,0694, \\
P_{X_1}(3) &= P_{X_1 X_2 X_3}(3, 0, 0) + P_{X_1 X_2 X_3}(3, 0, 1) + P_{X_1 X_2 X_3}(3, 0, 2) \\
&\quad + P_{X_1 X_2 X_3}(3, 0, 3) + P_{X_1 X_2 X_3}(3, 1, 0) + P_{X_1 X_2 X_3}(3, 1, 1) \\
&\quad + P_{X_1 X_2 X_3}(3, 1, 2) + P_{X_1 X_2 X_3}(3, 1, 3) + P_{X_1 X_2 X_3}(3, 2, 0) \\
&\quad + P_{X_1 X_2 X_3}(3, 2, 1) + P_{X_1 X_2 X_3}(3, 2, 2) + P_{X_1 X_2 X_3}(3, 2, 3) \\
&\quad + P_{X_1 X_2 X_3}(3, 3, 0) + P_{X_1 X_2 X_3}(3, 3, 1) + P_{X_1 X_2 X_3}(3, 3, 2) \\
&\quad + P_{X_1 X_2 X_3}(3, 3, 3) = 0,0046
\end{aligned}$$

La distribución marginal de la variable X_1 es la siguiente:

X_1	0	1	2	3
P_{X_1}	0,5787	0,3472	0,0694	0,0046

Las distribuciones marginales de las variables X_2 y X_3 coinciden con la distribución marginal de la variable X_1 . Lo dejamos como ejercicio.

A continuación, calculemos la **función de probabilidad marginal conjunta** de la variable (X_1, X_2) :

$$\begin{aligned}
P_{X_1 X_2}(0,0) &= P_{X_1 X_2 X_3}(0,0,0) + P_{X_1 X_2 X_3}(0,0,1) + P_{X_1 X_2 X_3}(0,0,2) \\
&\quad + P_{X_1 X_2 X_3}(0,0,3) = 0,2963, \\
P_{X_1 X_2}(0,1) &= P_{X_1 X_2 X_3}(0,1,0) + P_{X_1 X_2 X_3}(0,1,1) + P_{X_1 X_2 X_3}(0,1,2) \\
&\quad + P_{X_1 X_2 X_3}(0,1,3) = 0,2222, \\
P_{X_1 X_2}(0,2) &= P_{X_1 X_2 X_3}(0,2,0) + P_{X_1 X_2 X_3}(0,2,1) + P_{X_1 X_2 X_3}(0,2,2) \\
&\quad + P_{X_1 X_2 X_3}(0,2,3) = 0,0556, \\
P_{X_1 X_2}(0,3) &= P_{X_1 X_2 X_3}(0,3,0) + P_{X_1 X_2 X_3}(0,3,1) + P_{X_1 X_2 X_3}(0,3,2) \\
&\quad + P_{X_1 X_2 X_3}(0,3,3) = 0,0046, \\
P_{X_1 X_2}(1,0) &= P_{X_1 X_2 X_3}(1,0,0) + P_{X_1 X_2 X_3}(1,0,1) + P_{X_1 X_2 X_3}(1,0,2) \\
&\quad + P_{X_1 X_2 X_3}(1,0,3) = 0,2222, \\
P_{X_1 X_2}(1,1) &= P_{X_1 X_2 X_3}(1,1,0) + P_{X_1 X_2 X_3}(1,1,1) \\
&\quad + P_{X_1 X_2 X_3}(1,1,2) + P_{X_1 X_2 X_3}(1,1,3) = 0,1111, \\
P_{X_1 X_2}(1,2) &= P_{X_1 X_2 X_3}(1,2,0) + P_{X_1 X_2 X_3}(1,2,1) + P_{X_1 X_2 X_3}(1,2,2) \\
&\quad + P_{X_1 X_2 X_3}(1,2,3) = 0,0139, \\
P_{X_1 X_2}(1,3) &= P_{X_1 X_2 X_3}(1,3,0) + P_{X_1 X_2 X_3}(1,3,1) + P_{X_1 X_2 X_3}(1,3,2) \\
&\quad + P_{X_1 X_2 X_3}(1,3,3) = 0, \\
P_{X_1 X_2}(2,0) &= P_{X_1 X_2 X_3}(2,0,0) + P_{X_1 X_2 X_3}(2,0,1) + P_{X_1 X_2 X_3}(2,0,2) \\
&\quad + P_{X_1 X_2 X_3}(2,0,3) = 0,0556, \\
P_{X_1 X_2}(2,1) &= P_{X_1 X_2 X_3}(2,1,0) + P_{X_1 X_2 X_3}(2,1,1) + P_{X_1 X_2 X_3}(2,1,2) \\
&\quad + P_{X_1 X_2 X_3}(2,1,3) = 0,0139, \\
P_{X_1 X_2}(2,2) &= P_{X_1 X_2 X_3}(2,2,0) + P_{X_1 X_2 X_3}(2,2,1) + P_{X_1 X_2 X_3}(2,2,2) \\
&\quad + P_{X_1 X_2 X_3}(2,2,3) = 0, \\
P_{X_1 X_2}(2,3) &= P_{X_1 X_2 X_3}(2,3,0) + P_{X_1 X_2 X_3}(2,3,1) + P_{X_1 X_2 X_3}(2,3,2) \\
&\quad + P_{X_1 X_2 X_3}(2,3,3) = 0, \\
P_{X_1 X_2}(3,0) &= P_{X_1 X_2 X_3}(3,0,0) + P_{X_1 X_2 X_3}(3,0,1) + P_{X_1 X_2 X_3}(3,0,2) \\
&\quad + P_{X_1 X_2 X_3}(3,0,3) = 0,0046, \\
P_{X_1 X_2}(3,1) &= P_{X_1 X_2 X_3}(3,1,0) + P_{X_1 X_2 X_3}(3,1,1) \\
&\quad + P_{X_1 X_2 X_3}(3,1,2) + P_{X_1 X_2 X_3}(3,1,3) = 0, \\
P_{X_1 X_2}(3,2) &= P_{X_1 X_2 X_3}(3,2,0) + P_{X_1 X_2 X_3}(3,2,1) \\
&\quad + P_{X_1 X_2 X_3}(3,2,2) + P_{X_1 X_2 X_3}(3,2,3) = 0, \\
P_{X_1 X_2}(3,3) &= P_{X_1 X_2 X_3}(3,3,0) + P_{X_1 X_2 X_3}(3,3,1) \\
&\quad + P_{X_1 X_2 X_3}(3,3,2) + P_{X_1 X_2 X_3}(3,3,3) = 0,
\end{aligned}$$

La función de probabilidad marginal conjunta de la variable (X_1, X_2) queda resumida en la tabla siguiente:

X_1 / X_2	0	1	2	3
0	0,2963	0,2222	0,0556	0,0046
1	0,2222	0,1111	0,0139	0
2	0,0556	0,0139	0	0
3	0,0046	0	0	0

Las funciones de probabilidad marginales de las variables (X_1, X_3) y (X_2, X_3) dan el mismo resultado que la tabla anterior.

La **función de probabilidad marginal** de la variable X_1 en R se halla de la forma siguiente:

```
fun.marginal.X1 = function(x){
  suma=0;
  for (i in 0:3){for (j in 0:3){suma=suma+fun.prob.con(x,i,j)}}
  return(suma)
}
tabla.fun.marginal.X1=data.frame(fun.marginal.X1(0),fun.marginal.X1(1),
                                   fun.marginal.X1(2),fun.marginal.X1(3));
colnames(tabla.fun.marginal.X1)=0:3
knitr::kable(tabla.fun.marginal.X1)
```

	0	1	2	3
	0.5787037	0.3472222	0.0694444	0.0046296

La **función de probabilidad marginal** de la variable (X_1, X_2) en R se halla de la forma siguiente:

```
fun.marginal.X1.X2 = function(x,y){
  suma=0;
  for (i in 0:3){suma=suma+fun.prob.con(x,y,i)}
  return(suma)
}
tabla.fun.marginal.X1.X2=c()
for (i in 0:3){
  tabla.fun.marginal.X1.X2=cbind(
    tabla.fun.marginal.X1.X2,
    c(fun.marginal.X1.X2(i,0),
      )
  )
}
tabla.fun.marginal.X1.X2=as.data.frame(tabla.fun.marginal.X1.X2)
rownames(tabla.fun.marginal.X1.X2)=0:3
colnames(tabla.fun.marginal.X1.X2)=0:3
knitr::kable(tabla.fun.marginal.X1.X2)
```

	0	1	2	3
0	0.2962963	0.2222222	0.0555556	0.0046296
1	0.2222222	0.1111111	0.0138889	0.0000000
2	0.0555556	0.0138889	0.0000000	0.0000000
3	0.0046296	0.0000000	0.0000000	0.0000000

lo cual concluye nuestros cálculos con R.

6.4. Variables aleatorias n -dimensionales continuas

6.4.1. Definición

Recordemos la definición de **variable continua bidimensional**: (X, Y) es continua si existe una función $f_{XY} : \mathbb{R}^2 \rightarrow \mathbb{R}$, llamada **función de densidad conjunta** no negativa $f_{XY}(x, y) \geq 0$, para

todo $(x, y) \in \mathbb{R}^2$ tal que para cualquier región B del plano, la probabilidad de que (X, Y) esté en B se calcula de la forma siguiente:

$$P((X, Y) \in B) = \int \int_B f_{XY}(x, y) dx dy.$$

La generalización natural es, entonces:

Definición de variable aleatoria n -dimensional continua. Sea (X_1, \dots, X_n) una variable aleatoria n -dimensional. Diremos que (X_1, \dots, X_n) es continua si existe una función $f_{X_1, \dots, X_n} : \mathbb{R}^n \rightarrow \mathbb{R}$ llamada **función de densidad conjunta** no negativa $f_{X_1, \dots, X_n}(x_1, \dots, x_n) \geq 0$ para todo $(x_1, \dots, x_n) \in \mathbb{R}^n$ tal que dado cualquier región B del espacio n -dimensional, la probabilidad de que (X_1, \dots, X_n) esté en B se calcula de la forma siguiente:

$$P((X_1, \dots, X_n) \in B) = \int \cdots \int_B f_{X_1, \dots, X_n}(x_1, \dots, x_n) dx_1 \cdots dx_n.$$

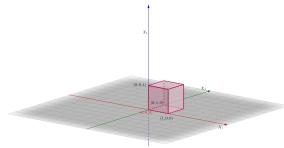
6.4.1.1. Ejemplos

Ejemplo

Consideremos una variable aleatoria 3-dimensional (X_1, X_2, X_3) con **función de densidad conjunta**:

$$f_{X_1, X_2, X_3}(x_1, x_2, x_3) = \begin{cases} 8x_1 \cdot x_2 \cdot x_3, & \text{si } 0 \leq x_1 \leq 1, 0 \leq x_2 \leq 1, 0 \leq x_3 \leq 1, \\ 0, & \text{en caso contrario.} \end{cases}$$

En la figura siguiente hemos dibujado en rosa la región donde f_{X_1, X_2, X_3} no es cero, es decir $[0, 1] \times [0, 1] \times [0, 1]$.



6.4.2. Propiedades de la función de densidad

Consideremos (X_1, \dots, X_n) una **variable aleatoria n -dimensional continua** con **función de densidad conjunta** f_{X_1, \dots, X_n} . Esta función verifica las propiedades siguientes:

- La integral de dicha función sobre todo el espacio n -dimensional vale 1:

$$\int \int_{\mathbb{R}^n} f_{X_1, \dots, X_n}(x_1, \dots, x_n) dx_1 \cdots dx_n = 1.$$

Para ver dicha propiedad, basta considerar $B = \mathbb{R}^n$, tener en cuenta que el suceso $(X_1, \dots, X_n) \in \mathbb{R}^n$ es el total Ω y aplicar la definición de f_{X_1, \dots, X_n} :

$$P((X_1, \dots, X_n) \in \mathbb{R}^n) = 1 = \int \cdots \int_{\mathbb{R}^n} f_{X_1, \dots, X_n}(x_1, \dots, x_n) dx_1 \cdots dx_n.$$

- La relación que hay entre la **función de distribución conjunta** F_{X_1, \dots, X_n} y la **función de densidad conjunta** f_{X_1, \dots, X_n} es la siguiente:

$$F_{X_1, \dots, X_n}(x_1, \dots, x_n) = \int_{-\infty}^{x_1} \cdots \int_{-\infty}^{x_n} f_{X_1, \dots, X_n}(u_1, \dots, u_n) du_1 \cdots du_n.$$

Para ver dicha propiedad, basta considerar $B = (-\infty, x_1] \times \cdots \times (-\infty, x_n]$ y aplicar la definición de **función de distribución conjunta**:

$$\begin{aligned} F_{X_1, \dots, X_n}(x_1, \dots, x_n) &= P((X_1, \dots, X_n) \in (-\infty, x_1] \times \cdots \times (-\infty, x_n]) \\ &= \int_{-\infty}^{x_1} \cdots \int_{-\infty}^{x_n} f_{X_1, \dots, X_n}(u_1, \dots, u_n) du_1 \cdots du_n. \end{aligned}$$

- La relación que hay entre la **función de densidad** F_{X_1, \dots, X_n} y la **función de distribución** f_{X_1, \dots, X_n} es la siguiente:

$$f_{X_1, \dots, X_n}(x_1, \dots, x_n) = \frac{\partial^n F_{X_1, \dots, X_n}(x_1, \dots, x_n)}{\partial x_1 \cdots \partial x_n}.$$

Dicha propiedad se deduce de la anterior, derivando primero respecto a x_1 , después respecto a x_2 y sucesivamente hasta llegar a x_n para eliminar las n integrales.

- La **función de densidad marginal** de la variable k dimensional $(X_{s_1}, \dots, X_{s_k})$ con $\{s_1, \dots, s_k\}$ un subconjunto de $\{1, \dots, n\}$, $f_{X_{s_1}, \dots, X_{s_k}}$ se calculan de la forma siguiente:

$$f_{X_{s_1}, \dots, X_{s_k}}(x_{s_1}, \dots, x_{s_k}) = \int_{x_{t_1}=-\infty}^{x_{t_1}=\infty} \cdots \int_{x_{t_{n-k}}=-\infty}^{x_{t_{n-k}}=\infty} f_{X_1, \dots, X_n}(x_1, \dots, x_n) dx_{t_1} \cdots dx_{t_{n-k}},$$

con $\{t_1, \dots, t_{n-k}\} = \{1, \dots, n\} \setminus \{s_1, \dots, s_k\}$. Es decir, las variables t 's son las que no aparecen en la definición de la variable aleatoria k dimensional $(X_{s_1}, \dots, X_{s_k})$.

6.4.2.1. Ejemplos

Ejemplo anterior

Comprobemos las propiedades usando la **función de densidad** del ejemplo anterior:

$$f_{X_1, X_2, X_3}(x_1, x_2, x_3) = \begin{cases} 8x_1 \cdot x_2 \cdot x_3, & \text{si } 0 \leq x_1 \leq 1, 0 \leq x_2 \leq 1, 0 \leq x_3 \leq 1, \\ 0, & \text{en caso contrario.} \end{cases}$$

La integral de f_{X_1, X_2, X_3} sobre todo el espacio 3D vale 1:

$$\begin{aligned} \iint_{\mathbb{R}^3} f_{X_1 X_2 X_3}(x_1, x_2, x_3) dx dy &= \int_0^1 \int_0^1 \int_0^1 8x_1 \cdot x_2 \cdot x_3 dx_1 dx_2 dx_3 \\ &= 8 \int_0^1 x_1 dx_1 \int_0^1 x_2 dx_2 \int_0^1 x_3 dx_3 = 8 \left[\frac{x_1^2}{2} \right]_0^1 \cdot \left[\frac{x_2^2}{2} \right]_0^1 \cdot \left[\frac{x_3^2}{2} \right]_0^1 \\ &= 8 \cdot \left(\frac{1}{2} \right)^3 = 1. \end{aligned}$$

Vamos a calcular la función de distribución $F_{X_1 X_2 X_3}$.

Recordemos que la expresión de la función de distribución en función de la función de densidad era:

$$F_{X_1 X_2 X_3}(x_1, x_2, x_3) = \int_{-\infty}^{x_1} \int_{-\infty}^{x_2} \int_{-\infty}^{x_3} f_{X_1 X_2 X_3}(u_1, u_2, u_3) du_1 du_2 du_3.$$

Como la región del espacio 3D donde $f_{X_1 X_2 X_3}(x_1, x_2, x_3)$ es no nula es el cubo unidad $[0, 1] \times [0, 1] \times [0, 1]$, fijado un punto del espacio (x_1, x_2, x_3) es fundamental calcular la intersección de dicho cubo unidad con la región $(-\infty, x_1] \times (-\infty, x_2] \times (-\infty, x_3]$.

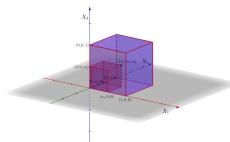
Dicha intersección $([0, 1] \times [0, 1] \times [0, 1]) \cap ((-\infty, x_1] \times (-\infty, x_2] \times (-\infty, x_3])$ es la región donde tendremos que integrar la función de densidad para hallar la función de distribución en el punto (x_1, x_2, x_3) .

Para hallar la región anterior, vamos a dividir el espacio 3D en tres “pisos”:

- “Sótano” o zona donde $x_3 < 0$.
- “Planta baja” o zona donde $0 \leq x_3 \leq 1$.
- “Primer piso” o zona donde $x_3 > 1$.
- Si (x_1, x_2, x_3) está en el “sótano” o $x_3 < 0$, claramente, $([0, 1] \times [0, 1] \times [0, 1]) \cap ((-\infty, x_1] \times (-\infty, x_2] \times (-\infty, x_3)) = \emptyset$. Por tanto, $F_{X_1 X_2 X_3}(x_1, x_2, x_3) = 0$.
- Si (x_1, x_2, x_3) está en la planta baja o $0 \leq x_3 \leq 1$, vamos a distinguir cuatro casos dependiendo de los valores de x_1 y x_2 :
 - $x_1 < 0$ o $x_2 < 0$. En este caso, $([0, 1] \times [0, 1] \times [0, 1]) \cap ((-\infty, x_1] \times (-\infty, x_2] \times (-\infty, x_3)) = \emptyset$. Por tanto, $F_{X_1 X_2 X_3}(x_1, x_2, x_3) = 0$.
 - $0 \leq x_1 \leq 1$ y $0 \leq x_2 \leq 1$. En este caso: $([0, 1] \times [0, 1] \times [0, 1]) \cap ((-\infty, x_1] \times (-\infty, x_2] \times (-\infty, x_3))$ que es igual a $[0, x_1] \times [0, x_2] \times [0, x_3]$, ver figura adjunta.

Por tanto,

$$\begin{aligned} F_{X_1 X_2 X_3}(x_1, x_2, x_3) &= \int_0^{x_1} \int_0^{x_2} \int_0^{x_3} 8x_1 x_2 x_3 dx_1 dx_2 dx_3 \\ &= 8 \left[\frac{x_1^2}{2} \right]_0^{x_1} \left[\frac{x_2^2}{2} \right]_0^{x_2} \left[\frac{x_3^2}{2} \right]_0^{x_3} = x_1^2 x_2^2 x_3^2. \end{aligned}$$



Seguimos en la planta “baja”,

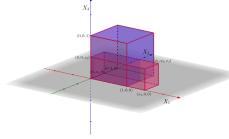
- Si $x_1 > 1$ y $0 \leq x_2 \leq 1$, $([0, 1] \times [0, 1] \times [0, 1]) \cap ((-\infty, x_1] \times (-\infty, x_2] \times (-\infty, x_3])$ que es igual a $[0, 1] \times [0, x_2] \times [0, x_3]$, ver figura adjunta. Hemos dibujado sólo la parte “positiva” de la región $(-\infty, x_1] \times (-\infty, x_2] \times (-\infty, x_3)$ ya que la parte “negativa” claramente no interseca con $[0, 1] \times [0, 1] \times [0, 1]$ para no complicar demasiado la figura.

En este caso,

$$\begin{aligned} F_{X_1 X_2 X_3}(x_1, x_2, x_3) &= \int_0^1 \int_0^{x_2} \int_0^{x_3} 8x_1 x_2 x_3 dx_1 dx_2 dx_3 \\ &= 8 \left[\frac{x_1^2}{2} \right]_0^1 \left[\frac{x_2^2}{2} \right]_0^{x_2} \left[\frac{x_3^2}{2} \right]_0^{x_3} = x_2^2 x_3^2. \end{aligned}$$

- Si $0 \leq x_1$ y $x_2 > 1$, es un caso parecido al caso anterior pero “cambiando los papeles” de x_1 y x_2 . Por tanto,

$$F_{X_1 X_2 X_3}(x_1, x_2, x_3) = x_1^2 x_3^2.$$

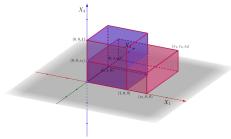


Seguimos en la planta “baja”,

- Si $x_1 > 1$ y $x_2 > 1$, entonces $([0, 1] \times [0, 1] \times [0, 1]) \cap ((-\infty, x_1] \times (-\infty, x_2] \times (-\infty, x_3))$ que es igual a $[0, 1] \times [0, 1] \times [0, x_3]$, ver figura adjunta. También hemos dibujado sólo la parte “positiva” de la región $(-\infty, x_1] \times (-\infty, x_2] \times (-\infty, x_3)$ ya que la parte “negativa” claramente no interseca con $[0, 1] \times [0, 1] \times [0, 1]$ para no complicar demasiado la figura.

En este caso,

$$F_{X_1 X_2 X_3}(x_1, x_2, x_3) = \int_0^1 \int_0^1 \int_0^{x_3} 8x_1 x_2 x_3 dx_1 dx_2 dx_3 = 8 \left[\frac{x_1^2}{2} \right]_0^1 \left[\frac{x_2^2}{2} \right]_0^1 \left[\frac{x_3^2}{2} \right]_0^{x_3} = x_3^2.$$



Supongamos ahora que (x_1, x_2, x_3) está en el “primer piso” o $x_3 > 1$. Aquí también vamos a distinguir 4 casos:

- $x_1 < 0$ o $x_2 < 0$. En este caso, $([0, 1] \times [0, 1] \times [0, 1]) \cap ((-\infty, x_1] \times (-\infty, x_2] \times (-\infty, x_3)) = \emptyset$. Por tanto, $F_{X_1 X_2 X_3}(x_1, x_2, x_3) = 0$.

- $0 \leq x_1 \leq 1$ y $0 \leq x_2 \leq 1$. En este caso: $([0, 1] \times [0, 1] \times [0, 1]) \cap ((-\infty, x_1] \times (-\infty, x_2] \times (-\infty, x_3]) = [0, x_1] \times [0, x_2] \times [0, 1]$, ver figura adjunta.

En este caso,

$$\begin{aligned} F_{X_1 X_2 X_3}(x_1, x_2, x_3) &= \int_0^{x_1} \int_0^{x_2} \int_0^1 8x_1 x_2 x_3 dx_1 dx_2 dx_3 \\ &= 8 \left[\frac{x_1^2}{2} \right]_0^{x_1} \left[\frac{x_2^2}{2} \right]_0^{x_2} \left[\frac{x_3^2}{2} \right]_0^1 = x_1^2 x_2^2. \end{aligned}$$

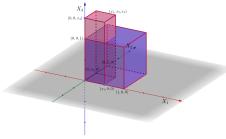


Figura 6.1: center

Seguimos en el “primer piso”,

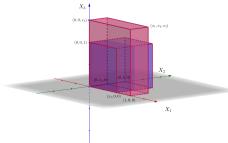
- Si $x_1 > 1$ y $0 \leq x_2 \leq 1$, $([0, 1] \times [0, 1] \times [0, 1]) \cap ((-\infty, x_1] \times (-\infty, x_2] \times (-\infty, x_3])$ que es igual a $[0, 1] \times [0, x_2] \times [0, 1]$, ver figura adjunta.

En este caso,

$$F_{X_1 X_2 X_3}(x_1, x_2, x_3) = \int_0^1 \int_0^{x_2} \int_0^1 8x_1 x_2 x_3 dx_1 dx_2 dx_3 = 8 \left[\frac{x_1^2}{2} \right]_0^1 \left[\frac{x_2^2}{2} \right]_0^{x_2} \left[\frac{x_3^2}{2} \right]_0^1 = x_2^2.$$

- Si $0 \leq x_1 \leq 1$ y $x_2 > 1$. Este caso es parecido al caso anterior pero “cambiando los papeles” de x_1 y x_2 . Por tanto,

$$F_{X_1 X_2 X_3}(x_1, x_2, x_3) = x_1^2.$$

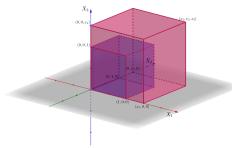


Seguimos en el “primer piso”,

- Si $x_1 > 1$ y $x_2 > 1$, $([0, 1] \times [0, 1] \times [0, 1]) \cap ((-\infty, x_1] \times (-\infty, x_2] \times (-\infty, x_3))$ que es iguala a $[0, 1] \times [0, 1] \times [0, 1]$, ver figura adjunta.

En este caso,

$$\begin{aligned} F_{X_1 X_2 X_3}(x_1, x_2, x_3) &= \int_0^1 \int_0^1 \int_0^1 8x_1 x_2 x_3 dx_1 dx_2 dx_3 \\ &= 8 \left[\frac{x_1^2}{2} \right]_0^1 \left[\frac{x_2^2}{2} \right]_0^1 \left[\frac{x_3^2}{2} \right]_0^1 = 1. \end{aligned}$$



En resumen:

$$F_{X_1 X_2 X_3}(x_1, x_2, x_3) = \begin{cases} 0, & \text{si } x_1 < 0, \text{ o } x_2 < 0, \text{ o } x_3 < 0 \\ x_1^2 \cdot x_2^2 \cdot x_3^2, & \text{si } 0 \leq x_1 \leq 1, 0 \leq x_2 \leq 1, 0 \leq x_3 \leq 1, \\ x_2^2 \cdot x_3^2, & \text{si } x_1 > 1, 0 \leq x_2 \leq 1, 0 \leq x_3 \leq 1, \\ x_1^2 \cdot x_3^2, & \text{si } 0 \leq x_1 \leq 1, x_2 > 1, 0 \leq x_3 \leq 1, \\ x_3^2, & \text{si } x_1 > 1, x_2 > 1, 0 \leq x_3 \leq 1, \\ x_1^2 \cdot x_2^2, & \text{si } 0 \leq x_1 \leq 1, 0 \leq x_2 \leq 1, x_3 > 1, \\ x_1^2, & \text{si } 0 \leq x_1 \leq 1, x_2 > 1, x_3 > 1, \\ x_2^2, & \text{si } x_1 > 1, 0 \leq x_2 \leq 1, x_3 > 1, \\ 1, & \text{si } x_1 \geq 1, x_2 \geq 1, x_3 \geq 1. \end{cases}$$

Dicha función era la función que nos sirvió como ejemplo a la hora de introducir las variables aleatorias n -dimensionales. Ahora sabemos que es continua y conocemos su función de densidad.

Comprobemos seguidamente que si derivamos tres veces la expresión de $F_{X_1 X_2 X_3}$, primero respecto x_1 , luego respecto x_2 y finalmente respecto x_3 , obtendremos la **función de densidad** $f_{X_1 X_2 X_3}$.

Si derivamos respecto x_1 obtenemos:

$$\frac{\partial F_{X_1 X_2 X_3}(x_1, x_2, x_3)}{\partial x_1} = \begin{cases} 0, & \text{si } x_1 < 0, \text{ o } x_2 < 0, \text{ o } x_3 < 0 \\ 2x_1 \cdot x_2^2 \cdot x_3^2, & \text{si } 0 \leq x_1 \leq 1, 0 \leq x_2 \leq 1, 0 \leq x_3 \leq 1, \\ 0, & \text{si } x_1 > 1, 0 \leq x_2 \leq 1, 0 \leq x_3 \leq 1, \\ 2x_1 \cdot x_3^2, & \text{si } 0 \leq x_1 \leq 1, x_2 > 1, 0 \leq x_3 \leq 1, \\ 0, & \text{si } x_1 > 1, x_2 > 1, 0 \leq x_3 \leq 1, \\ 2x_1 \cdot x_2^2, & \text{si } 0 \leq x_1 \leq 1, 0 \leq x_2 \leq 1, x_3 > 1, \\ 2x_1, & \text{si } 0 \leq x_1 \leq 1, x_2 > 1, x_3 > 1, \\ 0, & \text{si } x_1 > 1, 0 \leq x_2 \leq 1, x_3 > 1, \\ 0, & \text{si } x_1 \geq 1, x_2 \geq 1, x_3 \geq 1. \end{cases}$$

Si ahora derivamos respecto x_2 obtenemos:

$$\frac{\partial^2 F_{X_1 X_2 X_3}(x_1, x_2, x_3)}{\partial x_2 \partial x_1} = \begin{cases} 0, & \text{si } x_1 < 0, \text{ o } x_2 < 0, \text{ o } x_3 < 0 \\ 4x_1 \cdot x_2 \cdot x_3^2, & \text{si } 0 \leq x_1 \leq 1, 0 \leq x_2 \leq 1, 0 \leq x_3 \leq 1, \\ 0, & \text{si } x_1 > 1, 0 \leq x_2 \leq 1, 0 \leq x_3 \leq 1, \\ 0, & \text{si } 0 \leq x_1 \leq 1, x_2 > 1, 0 \leq x_3 \leq 1, \\ 0, & \text{si } x_1 > 1, x_2 > 1, 0 \leq x_3 \leq 1, \\ 4x_1 \cdot x_2, & \text{si } 0 \leq x_1 \leq 1, 0 \leq x_2 \leq 1, x_3 > 1, \\ 0, & \text{si } 0 \leq x_1 \leq 1, x_2 > 1, x_3 > 1, \\ 0, & \text{si } x_1 > 1, 0 \leq x_2 \leq 1, x_3 > 1, \\ 0, & \text{si } x_1 \geq 1, x_2 \geq 1, x_3 \geq 1. \end{cases}$$

Por último, si derivamos respecto x_3 , obtenemos:

$$\frac{\partial^3 F_{X_1 X_2 X_3}(x_1, x_2, x_3)}{\partial x_3 \partial x_2 \partial x_1} = \begin{cases} 0, & \text{si } x_1 < 0, \text{ o } x_2 < 0, \text{ o } x_3 < 0 \\ 8x_1 \cdot x_2 \cdot x_3, & \text{si } 0 \leq x_1 \leq 1, 0 \leq x_2 \leq 1, 0 \leq x_3 \leq 1, \\ 0, & \text{si } x_1 > 1, 0 \leq x_2 \leq 1, 0 \leq x_3 \leq 1, \\ 0, & \text{si } 0 \leq x_1 \leq 1, x_2 > 1, 0 \leq x_3 \leq 1, \\ 0, & \text{si } x_1 > 1, x_2 > 1, 0 \leq x_3 \leq 1, \\ 0, & \text{si } 0 \leq x_1 \leq 1, 0 \leq x_2 \leq 1, x_3 > 1, \\ 0, & \text{si } 0 \leq x_1 \leq 1, x_2 > 1, x_3 > 1, \\ 0, & \text{si } x_1 > 1, 0 \leq x_2 \leq 1, x_3 > 1, \\ 0, & \text{si } x_1 \geq 1, x_2 \geq 1, x_3 \geq 1, \end{cases}$$

expresión que coincide con la función de densidad $f_{X_1 X_2 X_3}(x_1, x_2, x_3)$.

Acabemos el ejemplo calculando las **funciones de densidad marginales** de las variables X_1, X_2, X_3 , $(X_1, X_2), (X_1, X_3), (X_2, X_3)$.

Debido a la simetría de la región donde $f_{X_1 X_2 X_3}(x_1, x_2, x_3)$ no se anula, es suficiente calcular la función de densidad marginal para las variables X_1 y (X_1, X_2) . Para ver las demás, basta cambiar los “papeles” de las variables correspondientes. Por ejemplo, la función de densidad de la variable X_2 es la misma que la de la variable X_1 cambiando x_1 por x_2 .

Para hallar la función de densidad de la variable X_1 , aplicamos la fórmula vista anteriormente:

$$f_{X_1}(x_1) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f_{X_1 X_2 X_3}(x_1, x_2, x_3) dx_2 dx_3.$$

Recordemos que la región donde no se anulaba la **función de densidad conjunta** $f_{X_1 X_2 X_3}$ era el cubo $[0, 1] \times [0, 1] \times [0, 1]$. Por tanto, fijado x_1 , el valor de $f_{X_1}(x_1)$ es no nulo si el plano “vertical” $X_1 = x_1$ interseca dicho cubo. Y esto ocurre siempre que $x_1 \in (0, 1)$. Por tanto,

$$f_{X_1}(x_1) = \begin{cases} \int_0^1 \int_0^1 8x_1 x_2 x_3 dx_2 dx_3 = 8x_1 \left[\frac{x_2^2}{2} \right]_0^1 \left[\frac{x_3^2}{2} \right]_0^1 = 2x_1, & \text{si } x_1 \in (0, 1), \\ 0, & \text{en caso contrario.} \end{cases}$$

Para hallar la función de densidad conjunta de la variable (X_1, X_2) , aplicamos la expresión siguiente:

$$f_{X_1 X_2}(x_1, x_2) = \int_{-\infty}^{\infty} f_{X_1 X_2 X_3}(x_1, x_2, x_3) dx_3.$$

En este caso, fijado x_1 y x_2 , tenemos que ver cuando la recta “vertical” $X_1 = x_1$, $X_2 = x_2$ intersecta el cubo $[0, 1] \times [0, 1] \times [0, 1]$ y esto ocurre siempre que $(x_1, x_2) \in [0, 1] \times [0, 1]$. Por tanto,

$$f_{X_1 X_2}(x_1, x_2) = \begin{cases} \int_0^1 8x_1 x_2 x_3 dx_3 \\ = 8x_1 x_2 \left[\frac{x_3^2}{2} \right]_0^1 = 4x_1 x_2 & , \quad \text{si } (x_1, x_2) \in [0, 1] \times [0, 1], \\ 0, & \text{en caso contrario.} \end{cases}$$

6.4.3. La distribución gaussiana n -dimensional

Vamos a generalizar la distribución normal a n dimensiones.

Definición de distribución gaussiana n -dimensional. Diremos que la distribución de la variable aleatoria n -dimensional (X_1, \dots, X_n) es **gaussiana n -dimensional** dependiendo del **vector de medias** y de la **matriz de covarianzas** Σ si su función de densidad conjunta es:

$$f_{X_1 \dots X_n}(x_1, \dots, x_n) = \frac{1}{(2\pi)^{\frac{n}{2}} \sqrt{|\Sigma|}} e^{-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})^\top \Sigma^{-1}(\mathbf{x}-\boldsymbol{\mu})},$$

$$-\infty < x_1, \dots, x_n < \infty,$$

Se denota $\mathbf{X} = (X_1, \dots, X_n)$ con $\mathbf{X} \sim N(\boldsymbol{\mu}, \Sigma)$, donde $\mathbf{x} = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}$, $\boldsymbol{\mu} = \begin{pmatrix} \mu_1 \\ \vdots \\ \mu_n \end{pmatrix}$ es el **vector de medias** de cada variable aleatoria X_1, \dots, X_n y Σ es la denominada **matriz de covarianzas** que nos sirve para estudiar la relación lineal entre las variables X_i , $i = 1, \dots, n$.

De hecho la componente (i, j) de la **matriz de covarianzas**, σ_{ij} es la **covarianza** entre las variables X_i y X_j .

Por tanto, los elementos de la diagonal de la **matriz de covarianzas**, σ_{ii} , es las varianzas de las variables X_i , $i = 1, \dots, n$.

Propiedades de la función de densidad de la variable gaussiana n -dimensional:

- Para cualquier punto $(x_1, \dots, x_n) \in \mathbb{R}^n$, la **función de densidad** es no nula: $f_{X_1 \dots X_n}(x_1, \dots, x_n) > 0$.
- La **función de densidad** tiene un único máximo absoluto en el punto que vale $f_{X_1 \dots X_n}(\boldsymbol{\mu}) = \frac{1}{(2\pi)^{\frac{n}{2}} \sqrt{|\Sigma|}}$.

Antes de estudiar cómo son las distribuciones de las **marginales** de una distribución **normal n -dimensional**, enunciemos el resultado siguiente:

Proposición (transformación afín de una **normal n -dimensional**) Sea $\mathbf{X} = N(\boldsymbol{\mu}, \Sigma)$ una distribución **normal n -dimensional**. Sea la variable k -dimensional \mathbf{Y} construida como $\mathbf{Y} = \mathbf{c} + \mathbf{C}\mathbf{X}$, con \mathbf{C} una matriz $k \times n$ y \mathbf{c} un vector k -dimensional. Entonces la variable \mathbf{Y} se distribuye como una variable **normal k -dimensional** de media $\mathbf{y} = \mathbf{c} + \mathbf{C}$ y matriz de covarianzas $\mathbf{y} = \mathbf{C}\mathbf{C}^\top$: $\mathbf{Y} \sim N(\mathbf{c} + \mathbf{C}\boldsymbol{\mu}, \mathbf{C}\Sigma\mathbf{C}^\top)$.

Usando la proposición anterior, podemos afirmar:

Proposición (distribución marginal de una variable normal n -dimensional) Sea $\mathbf{X} = N(\mu, \Sigma)$ una distribución **normal n -dimensional**. Sea $(X_{s_1}, \dots, X_{s_k})$ la variable k dimensional con las componentes s_1, \dots, s_k , con $s_i \in \{1, \dots, n\}$, entonces la variable $(X_{s_1}, \dots, X_{s_k})$ se distribuye según una **normal k -dimensional** de media $(\mu_{s_1}, \dots, \mu_{s_k})$ y matriz de covarianzas Σ' formada por las s_1, \dots, s_k filas y columnas de la matriz de covarianzas de la variable \mathbf{X} .

Para ver la proposición anterior a partir de la proposición de la transformación afín, hagamos un ejemplo concreto:

Ejemplo: normal multidimensional de dimensión 5

Consideremos una variable normal 5-dimensional de **vector de medias general** $\mu = (\mu_1, \mu_2, \mu_3, \mu_4, \mu_5)$ y **matriz de covarianzas**

$$= \begin{pmatrix} \sigma_{11} & \sigma_{12} & \sigma_{13} & \sigma_{14} & \sigma_{15} \\ \sigma_{21} & \sigma_{22} & \sigma_{23} & \sigma_{24} & \sigma_{25} \\ \sigma_{31} & \sigma_{32} & \sigma_{33} & \sigma_{34} & \sigma_{35} \\ \sigma_{41} & \sigma_{42} & \sigma_{43} & \sigma_{44} & \sigma_{45} \\ \sigma_{51} & \sigma_{52} & \sigma_{53} & \sigma_{54} & \sigma_{55} \end{pmatrix}$$

Queremos estudiar cuál es la distribución de la variable 3-dimensional (X_2, X_4, X_5) .

Para ello consideramos el vector $\mathbf{c} = 0$ y la matriz \mathbf{C} siguiente:

$$\mathbf{C} = \begin{pmatrix} 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix}$$

Esta matriz \mathbf{C} es de dimensiones 3×5 y vale 1 en los lugares (1, 2), (2, 4) y (3, 5). Fijémonos que las segundas componentes de los lugares anteriores son precisamente las componentes elegidas de la variable \mathbf{X} .

La matriz $\mathbf{Y} = \mathbf{C} \cdot \mathbf{X} = \begin{pmatrix} X_2 \\ X_4 \\ X_5 \end{pmatrix}$ vale precisamente la variable marginal que queremos estudiar.

Aplicando la proposición de la transformación afín, podemos afirmar que la distribución de la variable $\mathbf{Y} = \begin{pmatrix} X_2 \\ X_4 \\ X_5 \end{pmatrix}$ es una normal 3 dimensional de **vector de medias** $\mu_{\mathbf{Y}} = \mathbf{C} \mu = \begin{pmatrix} \mu_2 \\ \mu_4 \\ \mu_5 \end{pmatrix}$ y **vector de covarianzas**

$$\Sigma' = \mathbf{C} \mathbf{C}^{\top} = \begin{pmatrix} \sigma_{22} & \sigma_{24} & \sigma_{25} \\ \sigma_{42} & \sigma_{44} & \sigma_{45} \\ \sigma_{52} & \sigma_{54} & \sigma_{55} \end{pmatrix},$$

tal como indica la última proposición sobre distribuciones marginales.

6.5. Independencia de variables aleatorias

6.5.1. Independencia de variables aleatorias discretas

La generalización de **independencia** a variables aleatorias n -dimensionales es clara:

Definición de independencia para variables aleatorias multidimensionales discretas.

Sean (X_1, \dots, X_n) una **variable aleatoria n -dimensional discreta** con $(X_1, \dots, X_n)(\Omega) = \{(x_{i_1}, \dots, x_{i_n}), i_1 = 1, 2, \dots, i_n = 1, 2, \dots\}$ y **función de probabilidad conjunta** $P_{X_1 \dots X_n}$ y **funciones de probabilidad marginales** P_{X_1}, \dots, P_{X_n} . Entonces X_1, \dots, X_n son independientes si:

$$P_{X_1 \dots X_n}(x_{i_1}, \dots, x_{i_n}) = P_{X_1}(x_{i_1}) \cdots P_{X_n}(x_{i_n}), i_1 = 1, 2, \dots, i_n = 1, 2, \dots$$

o dicho de otra forma:

$$P(X_1 = x_{i_1}, X_n = x_{i_n}) = P(X_1 = x_{i_1}) \cdots P(X_n = x_{i_n}), i_1 = 1, 2, \dots, i_n = 1, 2, \dots$$

Ejemplo del lanzamiento de un dado tres veces

Consideremos el experimento aleatorio que consiste en lanzar un dado tres veces.

Recordemos que hemos estudiado la variable aleatoria (X_1, X_2, X_3) donde X_1 nos daba el número de 1's que han salido, X_2 , el número de 2's y X_3 , el número de 3's.

Las variables aleatorias anteriores no son independientes ya que, por ejemplo:

$$P_{X_1, X_2, X_3}(1, 1, 3) = 0 \neq P_{X_1}(1) \cdot P_{X_2}(1) \cdot P_{X_3}(3) = 0,3472 \cdot 0,3472 \cdot 0,0046 = 6 \times 10^{-4}.$$

Observación. Al igual que pasaba con las variables bidimensionales, si la tabla de la **función de probabilidad conjunta** de (X_1, \dots, X_n) contiene algún 0, las variables X_1, \dots, X_n no pueden ser independientes. ¿Podéis decir por qué?

Observación. Si X_1, \dots, X_n son variables aleatorias independientes, y consideramos una distribución marginal, por ejemplo $(X_{s_1}, \dots, X_{s_k})$, entonces las variables X_{s_1}, \dots, X_{s_k} también son independientes.

6.5.2. Independencia de variables aleatorias continuas

La definición dada para **variables aleatorias discretas** se traslada de forma natural a las **variables aleatorias continuas**:

Definición de independencia para variables aleatorias multidimensionales continuas.

Sean (X_1, \dots, X_n) una **variable aleatoria n -dimensional continua** con **función de densidad conjunta** $f_{X_1 \dots X_n}$ y **funciones de densidad marginales** f_{X_1}, \dots, f_{X_n} . Entonces X_1, \dots, X_n son independientes si:

$$f_{X_1 \dots X_n}(x_1, \dots, x_n) = f_{X_1}(x_1) \cdots f_{X_n}(x_n), \text{ para todo } x_1, \dots, x_n \in \mathbb{R}.$$

Ejemplo: variable aleatoria tridimensional continua

Recordemos el ejemplo siguiente visto donde teníamos una **variable aleatoria 3-dimensional continua** (X_1, X_2, X_3) con **función de densidad conjunta**:

$$f_{X_1 X_2 X_3}(x_1, x_2, x_3) = \begin{cases} 8x_1 \cdot x_2 \cdot x_3, & \text{si } 0 \leq x_1 \leq 1, 0 \leq x_2 \leq 1, 0 \leq x_3 \leq 1, \\ 0, & \text{en caso contrario.} \end{cases}$$

y con densidad marginales:

$$\begin{aligned} f_{X_1}(x_1) &= \begin{cases} 2x_1, & \text{si } 0 \leq x_1 \leq 1, \\ 0, & \text{en caso contrario.} \end{cases} & f_{X_2}(x_2) &= \begin{cases} 2x_2, & \text{si } 0 \leq x_2 \leq 1, \\ 0, & \text{en caso contrario.} \end{cases} \\ f_{X_3}(x_3) &= \begin{cases} 2x_3, & \text{si } 0 \leq x_3 \leq 1, \\ 0, & \text{en caso contrario.} \end{cases} \end{aligned}$$

Veamos que son independientes.

Consideremos dos casos:

- $(x_1, x_2, x_3) \in [0, 1] \times [0, 1] \times [0, 1]$. En este caso:

$$f_{X_1 X_2 X_3}(x_1, x_2, x_3) = 8x_1 x_2 x_3 = 2x_1 2x_2 2x_3 = f_{X_1}(x_1) \cdot f_{X_2}(x_2) \cdot f_{X_3}(x_3).$$

- $(x_1, x_2, x_3) \notin [0, 1] \times [0, 1] \times [0, 1]$. En este caso:

$$f_{X_1 X_2 X_3}(x_1, x_2, x_3) = 0 = f_{X_1}(x_1) \cdot f_{X_2}(x_2) \cdot f_{X_3}(x_3),$$

ya que si $(x_1, x_2, x_3) \notin [0, 1] \times [0, 1] \times [0, 1]$, o $x_1 \notin [0, 1]$ o $x_2 \notin [0, 1]$, o $x_3 \notin [0, 1]$. Por tanto $f_{X_1}(x_1) = 0$ o $f_{X_2}(x_2) = 0$ o $f_{X_3}(x_3) = 0$. En cualquier caso, $f_{X_1}(x_1) \cdot f_{X_2}(x_2) \cdot f_{X_3}(x_3) = 0$.

6.5.2.1. Ejemplo de la variable gaussiana n -dimensional

En este caso, recordemos que la **función de densidad conjunta** de (X_1, \dots, X_n) es:

$$f_{X_1 \dots X_n}(x_1, \dots, x_n) = \frac{1}{(2\pi)^{\frac{n}{2}} \sqrt{| \Sigma |}} e^{-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu})},$$

$$-\infty < x_1, \dots, x_n < \infty,$$

donde $\boldsymbol{\mu}$ es el vector de medias y $\boldsymbol{\Sigma}$, la matriz de covarianzas.

Recordemos también que las **funciones de densidad marginales** de X_1, \dots, X_n correspondían a $N(\mu_i, \sqrt{\sigma_{ii}})$, con $i = 1, \dots, n$:

$$f_{X_i}(x_i) = \frac{1}{\sqrt{2\pi\sigma_{ii}}} e^{-\frac{(x_i - \mu_i)^2}{2\sigma_{ii}}}, \quad -\infty < x_i < \infty, \quad i = 1, \dots, n.$$

¿Cómo tiene que ser la matriz de covarianzas para que las variables X_1, \dots, X_n sean independientes?

O, expresado matemáticamente,

$$\begin{aligned} f_{X_1}(x_1) \cdots f_{X_n}(x_n) &= \frac{1}{(2\pi)^{\frac{n}{2}} \sqrt{\sigma_{11} \cdots \sigma_{nn}}} e^{-\sum_{i=1}^n \frac{(x_i - \mu_i)^2}{2\sigma_{ii}}} = f_{X_1 \dots X_n}(x_1, \dots, x_n) \\ &= \frac{1}{(2\pi)^{\frac{n}{2}} \sqrt{|\Gamma|}} e^{-\frac{1}{2}(\mathbf{x} - \boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1} (\mathbf{x} - \boldsymbol{\mu})}. \end{aligned}$$

La respuesta es claramente cuando la matriz de covarianzas $\boldsymbol{\Sigma}$ es diagonal, es decir, si $\sigma_{ij} = 0$, para $i \neq j$, para todo $i, j = 1, \dots, n$.

En resumen, cuando la covarianza entre dos variables cualesquiera X_i y X_j distintas es cero, las variables normales X_1, \dots, X_n son independientes.

6.5.3. Relación de la independencia y la función de distribución

El siguiente resultado nos da la relación entre la **independencia de variables aleatorias** y su **función de distribución conjunta**:

Teorema. Sea (X_1, \dots, X_n) una variable aleatoria n -dimensional. Entonces X_1, \dots, X_n son independientes si, y sólo si, la **función de distribución conjunta** es el producto de las **funciones de distribución marginales** en todo valor $(x_1, \dots, x_n) \in \mathbb{R}^n$:

$$F_{X_1 \dots X_n}(x_1, \dots, x_n) = F_{X_1}(x_1) \cdots F_{X_n}(x_n), \quad (x_1, \dots, x_n) \in \mathbb{R}^n.$$

6.5.3.1. Ejemplos

Ejemplo: densidad tridimensional continuación

Recordemos la variable aleatoria 3-dimensional continua con **función de densidad conjunta**:

$$f_{X_1 X_2 X_3}(x_1, x_2, x_3) = \begin{cases} 8x_1 \cdot x_2 \cdot x_3, & \text{si } 0 \leq x_1 \leq 1, 0 \leq x_2 \leq 1, 0 \leq x_3 \leq 1, \\ 0, & \text{en caso contrario.} \end{cases}$$

Recordemos también **función de distribución conjunta** es:

$$F_{X_1 X_2 X_3}(x_1, x_2, x_3) = \begin{cases} 0, & \text{si } x_1 < 0, \text{ o } x_2 < 0, \text{ o } x_3 < 0 \\ x_1^2 \cdot x_2^2 \cdot x_3^2, & \text{si } 0 \leq x_1 \leq 1, 0 \leq x_2 \leq 1, 0 \leq x_3 \leq 1, \\ x_2^2 \cdot x_3^2, & \text{si } x_1 > 1, 0 \leq x_2 \leq 1, 0 \leq x_3 \leq 1, \\ x_1^2 \cdot x_3^2, & \text{si } 0 \leq x_1 \leq 1, x_2 > 1, 0 \leq x_3 \leq 1, \\ x_3^2, & \text{si } x_1 > 1, x_2 > 1, 0 \leq x_3 \leq 1, \\ x_1^2 \cdot x_2^2, & \text{si } 0 \leq x_1 \leq 1, 0 \leq x_2 \leq 1, x_3 > 1, \\ x_1^2, & \text{si } 0 \leq x_1 \leq 1, x_2 > 1, x_3 > 1, \\ x_2^2, & \text{si } x_1 > 1, 0 \leq x_2 \leq 1, x_3 > 1, \\ 1, & \text{si } x_1 \geq 1, x_2 \geq 1, x_3 \geq 1. \end{cases}$$

Recordemos las **funciones de densidad** de las **distribuciones marginales**:

$$\begin{aligned}f_{X_1}(x_1) &= \begin{cases} 2x_1, & \text{si } 0 \leq x_1 \leq 1, \\ 0, & \text{en caso contrario.} \end{cases} \\f_{X_2}(x_2) &= \begin{cases} 2x_2, & \text{si } 0 \leq x_2 \leq 1, \\ 0, & \text{en caso contrario.} \end{cases} \\f_{X_3}(x_3) &= \begin{cases} 2x_3, & \text{si } 0 \leq x_3 \leq 1, \\ 0, & \text{en caso contrario.} \end{cases}\end{aligned}$$

Dejamos como ejercicio verificar que las expresiones siguientes corresponden a las **funciones de distribución marginales**:

$$\begin{aligned}F_{X_1}(x_1) &= \begin{cases} 0, & \text{si } x_1 < 0, \\ x_1^2, & \text{si } 0 \leq x_1 \leq 1, \\ 1, & \text{si } x_1 > 1. \end{cases} \\F_{X_2}(x_2) &= \begin{cases} 0, & \text{si } x_2 < 0, \\ x_2^2, & \text{si } 0 \leq x_2 \leq 1, \\ 1, & \text{si } x_2 > 1. \end{cases} \\F_{X_3}(x_3) &= \begin{cases} 0, & \text{si } x_3 < 0, \\ x_3^2, & \text{si } 0 \leq x_3 \leq 1, \\ 1, & \text{si } x_3 > 1. \end{cases}\end{aligned}$$

Recordemos que X_1 , X_2 y X_3 son independientes. Por tanto verifiquemos que $F_{X_1 X_2 X_3}(x_1, x_2, x_3) = F_{X_1}(x_1) \cdot F_{X_2}(x_2) \cdot F_{X_3}(x_3)$ para todos los valores $x_1, x_2, x_3 \in \mathbb{R}$.

Distingamos los mismos casos que en la **función de distribución conjunta**:

- $x_1 < 0$, o $x_2 < 0$, o $x_3 < 0$. En este caso, o $F_{X_1}(x_1) = 0$, o $F_{X_2}(x_2) = 0$ o $F_{X_3}(x_3) = 0$. En cualquier caso:

$$F_{X_1 X_2 X_3}(x_1, x_2, x_3) = 0 = F_{X_1}(x_1) \cdot F_{X_2}(x_2) \cdot F_{X_3}(x_3).$$

- $0 \leq x_1 \leq 1$, y $0 \leq x_2 \leq 1$, y $0 \leq x_3 \leq 1$. En este caso:

$$F_{X_1 X_2 X_3}(x_1, x_2, x_3) = x_1^2 \cdot x_2^2 \cdot x_3^2 = F_{X_1}(x_1) \cdot F_{X_2}(x_2) \cdot F_{X_3}(x_3).$$

- $x_1 > 1$, y $0 \leq x_2 \leq 1$, y $0 \leq x_3 \leq 1$. En este caso:

$$F_{X_1 X_2 X_3}(x_1, x_2, x_3) = x_2^2 \cdot x_3^2 = 1 \cdot x_2^2 \cdot x_3^2 = F_{X_1}(x_1) \cdot F_{X_2}(x_2) \cdot F_{X_3}(x_3).$$

- $0 \leq x_1 \leq 1$, y $x_2 > 1$, y $0 \leq x_3 \leq 1$. En este caso:

$$F_{X_1 X_2 X_3}(x_1, x_2, x_3) = x_1^2 \cdot x_3^2 = x_1^2 \cdot 1 \cdot x_3^2 = F_{X_1}(x_1) \cdot F_{X_2}(x_2) \cdot F_{X_3}(x_3).$$

- $x_1 > 1$, y $x_2 > 1$, y $0 \leq x_3 \leq 1$. En este caso:

$$F_{X_1 X_2 X_3}(x_1, x_2, x_3) = x_3^2 = 1 \cdot 1 \cdot x_3^2 = F_{X_1}(x_1) \cdot F_{X_2}(x_2) \cdot F_{X_3}(x_3).$$

- $0 \leq x_1 \leq 1$, y $0 \leq x_2 \leq 1$, y $x_3 > 1$. En este caso:

$$F_{X_1 X_2 X_3}(x_1, x_2, x_3) = x_1^2 \cdot x_2^2 = x_1^2 \cdot x_2^2 \cdot 1 = F_{X_1}(x_1) \cdot F_{X_2}(x_2) \cdot F_{X_3}(x_3).$$

- $0 \leq x_1 \leq 1$, y $x_2 > 1$, y $x_3 > 1$. En este caso:

$$F_{X_1 X_2 X_3}(x_1, x_2, x_3) = x_1^2 = x_1^2 \cdot 1 \cdot 1 = F_{X_1}(x_1) \cdot F_{X_2}(x_2) \cdot F_{X_3}(x_3).$$

- $x_1 > 1$, y $0 \leq x_2 \leq 1$, y $x_3 > 1$. En este caso:

$$F_{X_1 X_2 X_3}(x_1, x_2, x_3) = x_2^2 = 1 \cdot x_2^2 \cdot 1 = F_{X_1}(x_1) \cdot F_{X_2}(x_2) \cdot F_{X_3}(x_3).$$

- $x_1 > 1$, y $x_2 > 1$, y $x_3 > 1$. En este caso:

$$F_{X_1 X_2 X_3}(x_1, x_2, x_3) = 1 = 1 \cdot 1 \cdot 1 = F_{X_1}(x_1) \cdot F_{X_2}(x_2) \cdot F_{X_3}(x_3).$$

6.6. Momentos conjuntos y valores esperados conjuntos

6.6.1. Valor esperado de una función de n variables aleatorias

Sea (X_1, \dots, X_n) una variable aleatoria n -dimensional.

Sea $P_{X_1 \dots X_n}$ su **función de probabilidad conjunta** en el caso en que (X_1, \dots, X_n) sea **discreta** y $f_{X_1 \dots X_n}$ su **función de densidad conjunta** en el caso en que (X_1, \dots, X_n) sea **continua**.

Sea $Z = g(X_1, \dots, X_n)$ una **variable aleatoria unidimensional** función de las variables X_1, \dots, X_n . Por ejemplo:

- Media aritmética de las n variables $g(x_1, \dots, x_n) = \frac{x_1 + \dots + x_n}{n}$: $Z = \frac{X_1 + \dots + X_n}{n}$.
- Media geométrica de las n variables $g(x_1, \dots, x_n) = \sqrt[n]{x_1 \cdots x_n}$: $Z = \sqrt[n]{X_1 \cdots X_n}$.
- Suma de los cuadrados de las variables $g(x_1, \dots, x_n) = x_1^2 + \dots + x_n^2$: $Z = X_1^2 + \dots + X_n^2$.

Hay que tener en cuenta que Z , como **variable aleatoria unidimensional** tiene una **función de probabilidad** P_Z en el caso en que (X_1, \dots, X_n) sea discreta y una **función de densidad** f_Z en el caso en que (X_1, \dots, X_n) sea continua.

El siguiente resultado nos dice cómo calcular el **valor esperado** de Z sin tener que calcular P_Z o f_Z , sólo usando la información de la **variable aleatoria conjunta** (X_1, \dots, X_n) :

Proposición. El valor esperado de Z se puede hallar usando la expresión siguiente:

- en el caso en que (X_1, \dots, X_n) sea discreta con $(X_1, \dots, X_n)(\Omega) = \{(x_{i_1}, \dots, x_{i_n}), i_1 = 1, 2, \dots, i_n = 1, 2, \dots\}$,

$$E(Z) = E(g(X_1, \dots, X_n)) = \sum_{x_{i_1}} \cdots \sum_{x_{i_n}} g(x_{i_1}, \dots, x_{i_n}) P(x_{i_1}, \dots, x_{i_n}),$$

- en el caso en que (X_1, \dots, X_n) sea continua:

$$\begin{aligned} E(Z) &= E(g(X_1, \dots, X_n)) \\ &= \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} g(x_1, \dots, x_n) f_{X_1 \dots X_n}(x_1, \dots, x_n) dx_1 \dots dx_n. \end{aligned}$$

6.6.2. Ejemplos

Ejemplo del lanzamiento de un dado tres veces

Consideremos el ejemplo de la variable (X_1, X_2, X_3) que nos daba el número de 1's, 2's y 3's en el lanzamiento de un dado tres veces.

Vamos a calcular $E(X_1 \cdot X_2 \cdot X_3)$.

El valor esperado anterior se calcula como sigue:

$$E(X_1 \cdot X_2 \cdot X_3) = \sum_{i_1=0}^3 \sum_{i_2=0}^3 \sum_{i_3=0}^3 i_1 \cdot i_2 \cdot i_3 \cdot P_{X_1 X_2 X_3}(i_1, i_2, i_3),$$

en total $4^3 = 64$ términos. Como el cálculo es tedioso, lo vamos a realizar con ayuda de R.

```
valor.esperado=0;
for (i1 in 0:3){
  for (i2 in 0:3){
    for (i3 in 0:3){
      valor.esperado=valor.esperado+i1*i2*i3*fun.prob.con(i1,i2,i3)
    }
  }
}
valor.esperado
```

```
## [1] 0.02777778
```

Ejemplo: densidad tridimensional (continuación)

Recordemos la variable aleatoria 3-dimensional (X_1, X_2, X_3) con **función de densidad conjunta**:

$$f_{X_1 X_2 X_3}(x_1, x_2, x_3) = \begin{cases} 8x_1 \cdot x_2 \cdot x_3, & \text{si } 0 \leq x_1 \leq 1, 0 \leq x_2 \leq 1, 0 \leq x_3 \leq 1, \\ 0, & \text{en caso contrario.} \end{cases}$$

Calculemos $E(X_1^2 + X_2^2 + X_3^2)$:

$$\begin{aligned} E(X_1^2 + X_2^2 + X_3^2) &= \int_{x_1=0}^{x_1=1} \int_{x_2=0}^{x_2=1} \int_{x_3=0}^{x_3=1} (x_1^2 + x_2^2 + x_3^2) 8 \cdot x_1 \cdot x_2 \cdot x_3 dx_1 dx_2 dx_3 \\ &= 8 \cdot \left(\int_{x_1=0}^{x_1=1} \int_{x_2=0}^{x_2=1} \int_{x_3=0}^{x_3=1} x_1^3 \cdot x_2 \cdot x_3 dx_1 dx_2 dx_3 + \int_{x_1=0}^{x_1=1} \int_{x_2=0}^{x_2=1} \int_{x_3=0}^{x_3=1} x_1 \cdot x_2^3 \cdot x_3 dx_1 dx_2 dx_3 \right. \\ &\quad \left. + \int_{x_1=0}^{x_1=1} \int_{x_2=0}^{x_2=1} \int_{x_3=0}^{x_3=1} x_1 \cdot x_2 \cdot x_3^3 dx_1 dx_2 dx_3 \right) \\ &= 8 \left(\left[\frac{x_1^4}{4} \right]_0^1 \left[\frac{x_2^2}{2} \right]_0^1 \left[\frac{x_3^2}{2} \right]_0^1 + \left[\frac{x_1^2}{2} \right]_0^1 \left[\frac{x_2^4}{4} \right]_0^1 \left[\frac{x_3^2}{2} \right]_0^1 + \left[\frac{x_1^2}{2} \right]_0^1 \left[\frac{x_2^2}{2} \right]_0^1 \left[\frac{x_3^4}{4} \right]_0^1 \right) \\ &= 8 \cdot 3 \cdot \frac{1}{16} = \frac{3}{2} = 1,5. \end{aligned}$$

6.6.3. Propiedad del valor esperado de la suma de variables

En estos momentos estamos en condiciones de demostrar el resultado siguiente:

Proposición. Valor esperado de la suma de variables aleatorias. Sea (X_1, \dots, X_n) una variable aleatoria n -dimensional. Entonces el valor esperado de la variable aleatoria suma de las variables es igual a la suma de los valores esperados de cada variable:

$$E(X_1 + \dots + X_n) = E(X_1) + \dots + E(X_n).$$

Demostración

Haremos la demostración para el caso continuo. Dejamos como ejercicio la demostración en el caso discreto.

Para calcular valor esperado de la suma de variables necesitamos la **función de densidad conjunta** $f_{X_1 \dots X_n}$:

$$\begin{aligned} E(X_1 + \dots + X_n) &= \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} (x_1 + \dots + x_n) f_{X_1 \dots X_n}(x_1, \dots, x_n) dx_1 \dots dx_n \\ &= \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} x_1 f_{X_1 \dots X_n}(x_1, \dots, x_n) dx_1 \dots dx_n + \dots \\ &\quad + \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} x_n f_{X_1 \dots X_n}(x_1, \dots, x_n) dx_1 \dots dx_n \\ &= E(X_1) + \dots + E(X_n). \end{aligned}$$

6.6.3.1. Ejemplos

Ejemplo del lanzamiento de un dado tres veces

Consideremos el ejemplo de la variable (X_1, X_2, X_3) que nos daba el número de 1's, 2's y 3's en el lanzamiento de un dado tres veces.

Comprobemos en este caso que $E(X_1 + X_2 + X_3) = E(X_1) + E(X_2) + E(X_3)$.

Calculemos $E(X_1 + X_2 + X_3)$ con R usando la misma técnica que en el ejemplo donde calculábamos $E(X_1 \cdot X_2 \cdot X_3)$:

```
valor.esperado=0;
for (i1 in 0:3){
  for (i2 in 0:3){
    for (i3 in 0:3){
      valor.esperado=valor.esperado+(i1+i2+i3)*fun.prob.con(i1,i2,i3)
    }
  }
}
valor.esperado

## [1] 1.5
```

La distribución marginal de cada una de las variables X_1 , X_2 y X_3 recordemos que es la siguiente:

X_1	0	1	2	3
P_{X_1}	0,5787	0,3472	0,0694	0,0046

El valor de $E(X_1) = E(X_2) = E(X_3)$ es:

```

esperanza.X1=0
for (i in 0:3){
  esperanza.X1=esperanza.X1+i*fun.marginal.X1(i)
}
esperanza.X1

## [1] 0.5

```

Claramente:

$$E(X_1 + X_2 + X_3) = 1,5 = E(X_1) + E(X_2) + E(X_3) = 0,5 + 0,5 + 0,5.$$

6.6.4. Valor esperado de una función de n variables aleatorias independientes

El siguiente resultado nos simplifica el cálculo del **valor esperado de una función de n variables aleatorias** en el caso en que sean **independientes**:

Proposición: cálculo del valor esperado de una función de n variables aleatorias en el caso de independencia. Sea (X_1, \dots, X_n) una variable aleatoria n -dimensional donde suponemos que X_1, \dots, X_n son independientes. Sea $Z = g(X_1, \dots, X_n)$ una variable aleatoria unidimensional función de X_1, \dots, X_n donde suponemos que podemos “separar” las variables x_1, \dots, x_n en la función g . O lo que es lo mismo, existen n funciones g_1, \dots, g_n tal que $g(x_1, \dots, x_n) = g_1(x_1) \cdots g_n(x_n)$ para todo valor $x_1, \dots, x_n \in \mathbb{R}$. En este caso, el valor esperado de Z se puede calcular como:

$$E(Z) = E(g(X_1, \dots, X_n)) = E_{X_1}(g_1(X_1)) \cdots E_{X_n}(g_n(X_n)).$$

El cálculo de $E(g(X_1, \dots, X_n))$ que es una suma múltiple en el caso de que (X_1, \dots, X_n) sea **discreta** o una integral múltiple en el caso en que (X_1, \dots, X_n) sea continua se transforma en el producto de n sumas simples (caso **discreto**) o el producto de n integrales simples (caso **continuo**):

$$\begin{aligned} E(Z) &= E(g(X_1, \dots, X_n)) \\ &= \left(\sum_{x_{i_1}} g_1(x_{i_1}) \cdot P_{X_1}(x_{i_1}) \right) \cdots \left(\sum_{x_{i_n}} g_n(x_{i_n}) \cdot P_{X_n}(x_{i_n}) \right), \text{ caso discreto,} \\ E(Z) &= E(g(X_1, \dots, X_n)) \\ &= \left(\int_{-\infty}^{\infty} g_1(x_1) \cdot f_{X_1}(x_1) dx_1 \right) \cdots \left(\int_{-\infty}^{\infty} g_n(x_n) \cdot f_{X_n}(x_n) dx_n \right), \text{ caso continuo.} \end{aligned}$$

Un caso particular de aplicación de la proposición anterior se produce cuando queramos calcular $E(X_1 \cdots X_n)$. En este caso $g(x_1, \dots, x_n) = x_1 \cdots x_n$, $g_1(x_1) = x_1$, y $g_n(x_n) = x_n$.

Podemos escribir, por tanto:

$$E(X_1 \cdots X_n) = E_{X_1}(X_1) \cdots E_{X_n}(X_n),$$

si X_1, \dots, X_n son independientes.

6.6.4.1. Ejemplos

Ejemplo: densidad tridimensional (continuación)

Recordemos el ejemplo con función de densidad conjunta:

$$f_{X_1 X_2 X_3}(x_1, x_2, x_3) = \begin{cases} 8x_1 \cdot x_2 \cdot x_3, & \text{si } 0 \leq x_1 \leq 1, 0 \leq x_2 \leq 1, 0 \leq x_3 \leq 1, \\ 0, & \text{en caso contrario.} \end{cases}$$

del que ya sabemos que las tres variables son independientes. Comprobemos que $E(X_1 \cdot X_2 \cdot X_3) = E(X_1) \cdot E(X_2) \cdot E(X_3)$.

Calculemos $E(X_1 \cdot X_2 \cdot X_3)$:

$$\begin{aligned} E(X_1 \cdot X_2 \cdot X_3) &= \int_0^1 \int_0^1 \int_0^1 x_1 \cdot x_2 \cdot x_3 \cdot 8x_1 \cdot x_2 \cdot x_3 dx_1 dx_2 dx_3 \\ &= 8 \int_0^1 \int_0^1 \int_0^1 x_1^2 \cdot x_2^2 \cdot x_3^2 dx_1 dx_2 dx_3 = 8 \left[\frac{x_1^3}{3} \right]_0^1 \left[\frac{x_2^3}{3} \right]_0^1 \left[\frac{x_3^3}{3} \right]_0^1 \\ &= 8 \cdot \frac{1}{3^3} = \frac{8}{27} = 0,2963. \end{aligned}$$

Recordemos que las **densidades marginales** son:

$$\begin{aligned} f_{X_1}(x_1) &= \begin{cases} 2x_1, & \text{si } 0 \leq x_1 \leq 1, \\ 0, & \text{en caso contrario.} \end{cases} & f_{X_2}(x_2) &= \begin{cases} 2x_2, & \text{si } 0 \leq x_2 \leq 1, \\ 0, & \text{en caso contrario.} \end{cases} \\ f_{X_3}(x_3) &= \begin{cases} 2x_3, & \text{si } 0 \leq x_3 \leq 1, \\ 0, & \text{en caso contrario.} \end{cases} \end{aligned}$$

Por tanto:

$$E(X_1) = E(X_2) = E(X_3) = \int_0^1 x \cdot 2x dx = 2 \int_0^1 x^2 dx = 2 \left[\frac{x^3}{3} \right]_0^1 = \frac{2}{3}.$$

Claramente, se verifica:

$$E(X_1 \cdot X_2 \cdot X_3) = \frac{8}{27} = E(X_1) \cdot E(X_2) \cdot E(X_3) = \left(\frac{2}{3} \right)^3.$$

6.6.5. Propiedades de la covarianza

Veamos cómo se calcula la **covarianza** de dos combinaciones lineales de variables aleatorias:

Proposición (covarianza de dos combinaciones lineales de variables aleatorias). Sean (X_1, \dots, X_n) e (Y_1, \dots, Y_n) dos variables aleatorias n -dimensionales. Sean a_1, \dots, a_n y b_1, \dots, b_n n parejas de valores reales. Sean U y V las variables aleatorias siguientes: $U = \sum_{i=1}^n a_i X_i$, $V = \sum_{i=1}^n b_i Y_i$. Entonces la **covarianza** de las variables U y V se calcula usando la expresión siguiente:

$$\text{Cov}(U, V) = \sum_{i=1}^n \sum_{j=1}^n a_i b_j \text{Cov}(X_i, Y_j).$$

Demostración

Usando la expresión de la **covarianza** podemos calcular la covarianza entre las variables U y V como:

$$\begin{aligned}\text{Cov}(U, V) &= E(UV) - E(U) \cdot E(V) = E\left(\sum_{i=1}^n a_i \cdot X_i \cdot \sum_{j=1}^n b_j \cdot Y_j\right) - E\left(\sum_{i=1}^n a_i \cdot X_i\right) E\left(\sum_{j=1}^n b_j \cdot Y_j\right) \\ &= \sum_{i=1}^n \sum_{j=1}^n a_i \cdot b_j \cdot E(X_i Y_j) - \sum_{i=1}^n a_i \cdot E(X_i) \sum_{j=1}^n b_j \cdot E(Y_j) \\ &= \sum_{i=1}^n \sum_{j=1}^n a_i \cdot b_j \cdot E(X_i Y_j) - \sum_{i=1}^n \sum_{j=1}^n a_i \cdot b_j \cdot E(X_i) \cdot E(Y_j) \\ &= \sum_{i=1}^n \sum_{j=1}^n a_i b_j (E(X_i Y_j) - E(X_i) \cdot E(Y_j)) = \sum_{i=1}^n \sum_{j=1}^n a_i \cdot b_j \cdot \text{Cov}(X_i, Y_j),\end{aligned}$$

tal como queríamos ver.

Si en la combinación lineal sólo hay las componentes de una sola variable aleatoria n -dimensional, la proposición anterior se convierte en la proposición siguiente:

Proposición (covarianza de una combinación lineal de variables aleatorias). Sean (X_1, \dots, X_n) una variable aleatoria n -dimensional. Sean a_1, \dots, a_n y b_1, \dots, b_n n parejas de valores reales. Sean U y V las variables aleatorias siguientes: $U = \sum_{i=1}^n a_i X_i$, $V = \sum_{i=1}^n b_i X_i$. Entonces la **covarianza** de las variables U y V se calcula usando la expresión siguiente:

$$\text{Cov}(U, V) = \sum_{i=1}^n a_i b_i \text{Var}(X_i) + \sum_{i=1}^n \sum_{j=1, j \neq i}^n a_i b_j \text{Cov}(X_i, X_j).$$

Demostración

Usando la misma técnica de demostración que en la proposición anterior, podemos calcular la covarianza entre las variables U y V como:

$$\begin{aligned}\text{Cov}(U, V) &= \sum_{i=1}^n \sum_{j=1}^n a_i \cdot b_j \cdot (E(X_i \cdot X_j) - E(X_i) \cdot E(X_j)) \\ &= \sum_{i=1}^n a_i \cdot b_i \cdot (E(X_i^2) - E(X_i)^2) + \sum_{i=1}^n \sum_{j=1, j \neq i}^n a_i \cdot b_j \cdot \text{Cov}(X_i, X_j) \\ &= \sum_{i=1}^n a_i \cdot b_i \cdot \text{Var}(X_i) + \sum_{i=1}^n \sum_{j=1, j \neq i}^n a_i b_j \text{Cov}(X_i, X_j),\end{aligned}$$

tal como queríamos ver.

Observación. Una forma equivalente de escribir la covarianza anterior es:

$$\text{Cov}(U, V) = \sum_{i=1}^n a_i \cdot b_i \cdot \text{Var}(X_i) + 2 \sum_{i=1}^n \sum_{j=1, j > i}^n a_i \cdot b_j \cdot \text{Cov}(X_i, X_j).$$

Como, en general $\text{Var}(U) = \text{Cov}(U, U)$, una consecuencia directa de la proposición anterior es la expresión de la varianza de una combinación lineal de variables aleatorias:

Proposición (varianza de una combinación lineal de variables aleatorias). Sean (X_1, \dots, X_n) una variable aleatoria n -dimensional. Sean a_1, \dots, a_n n valores reales. Sea U la variable aleatoria siguiente: $U =$

$\sum_{i=1}^n a_i X_i$. Entonces la **varianza** de la variable U se calcula usando la expresión siguiente:

$$\text{Var}(U) = \sum_{i=1}^n a_i^2 \text{Var}(X_i) + \sum_{i=1}^n \sum_{j=1, j \neq i}^n a_i a_j \text{Cov}(X_i, X_j).$$

Demostración

Para la demostración, basta tener en cuenta que $\text{Var}(U) = \text{Cov}(U, U)$ y aplicar la expresión de la covariancia entre U y V en la proposición que nos da la covarianza de una combinación lineal de variables aleatorias.

Observación. En este caso, también existe una forma equivalente de escribir la varianza anterior:

$$\text{Var}(U) = \sum_{i=1}^n a_i^2 \text{Var}(X_i) + 2 \sum_{i=1}^n \sum_{j=1, j > i}^n a_i a_j \text{Cov}(X_i, X_j).$$

Una consecuencia de la proposición anterior es que si las variables son **independientes**, la **varianza** de la suma es la suma de varianzas:

Proposición (varianza de la suma de variables aleatorias independientes). Sean (X_1, \dots, X_n) una variable aleatoria n -dimensional donde X_1, \dots, X_n son independientes.

Entonces la **varianza** de la variable $X_1 + \dots + X_n$ es la suma de las varianzas de cada variable aleatoria:

$$\text{Var}(X_1 + \dots + X_n) = \sum_{i=1}^n \text{Var}(X_i).$$

Demostración

Para la demostración, basta aplicar la proposición anterior de la varianza de una combinación lineal de variables aleatorias con $a_i = 1$, para todo $i = 1, \dots, n$ y tener en cuenta que como son independientes, $\text{Cov}(X_i, X_j) = 0$, para todo $i \neq j$.

Capítulo 7

Ley de los grandes números y Teorema Central del Límite

7.1. Muestras aleatorias simples

El pilar básico sobre el que se sustenta la **estadística inferencial** es el concepto de **muestra aleatoria simple**.

Una **muestra aleatoria simple**, desde el punto de vista de la probabilidad es una distribución n variables aleatorias, X_1, \dots, X_n todas independientes entre sí e idénticamente distribuidas ya que queremos simular la repetición de un experimento n veces de forma independiente.

Por tanto, estudiar una **muestra aleatoria simple** equivale a estudiar su distribución.

En muchos casos, nos bastará estudiar la distribución de una variable que “represente” a dicha **muestra aleatoria simple**: la media muestral definida como $\bar{X} = \frac{X_1 + \dots + X_n}{n}$.

Las **leyes de los grandes números** nos dicen que, de alguna manera (que concretaremos más adelante), la media muestral y la media poblacional se “parecen” a la larga o cuando el número de repeticiones n tiende a infinito.

El **Teorema Central del Límite** nos dice que la distribución de la media muestral tiende, sea cual sea la distribución de las variables X_i , a una normal. De ahí que la **distribución normal** sea la más importante en probabilidades y estadística.

7.1.1. La distribución de la media muestral

Vamos cómo se distribuye la media de un conjunto de variables normales e idénticamente distribuidas:

Proposición. Distribución de la media muestral de n variables normales independientes e idénticamente distribuidas. Sean X_1, \dots, X_n n variables normales de media μ y varianza σ^2 , todas normales e independientes. Consideraremos la variable $\bar{X} = \frac{X_1 + \dots + X_n}{n}$ la media muestral. Entonces la distribución de la variable aleatoria \bar{X} es normal de la misma media μ de las X_i y varianza $\frac{\sigma^2}{n}$.

Demostración

Consideramos la variable aleatoria n -dimensional $\mathbf{X} = (X_1, \dots, X_n)$. Dicha variable tendrá la distribución normal n -dimensional con vector de medias $\mu = (\mu_1, \dots, \mu_n)^\top$ y matriz de covarianzas diagonal ya que recordemos que las X_i son independientes y, por tanto, incorreladas o de covarianza nula:

$$= \begin{pmatrix} \sigma^2 & 0 & \dots & 0 \\ 0 & \sigma^2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \sigma^2 \end{pmatrix}.$$

Para hallar la variable \bar{X} , realizamos la transformación afín siguiente:

$$\bar{X} = \left(\frac{1}{n}, \dots, \frac{1}{n} \right) \cdot \begin{pmatrix} X_1 \\ X_2 \\ \vdots \\ X_n \end{pmatrix}.$$

Aplicando la proposición sobre la transformación afín sobre una variable normal n -dimensional que vimos en el capítulo de distribuciones n -dimensionales con matriz de cambio $\mathbf{C} = \left(\frac{1}{n}, \dots, \frac{1}{n} \right)$ y $\mathbf{c} = 0$, tenemos que la distribución de \bar{X} será normal de media $\mathbf{c} + \mathbf{C}\mu = \mu$ y varianza (o matriz de covarianzas 1×1):

$$\mathbf{C} \cdot \cdot \mathbf{C}^\top = \left(\frac{1}{n}, \dots, \frac{1}{n} \right) \cdot \begin{pmatrix} \sigma^2 & 0 & \dots & 0 \\ 0 & \sigma^2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \sigma^2 \end{pmatrix} \cdot \begin{pmatrix} \frac{1}{n} \\ \frac{1}{n} \\ \vdots \\ \frac{1}{n} \end{pmatrix} = \frac{\sigma^2}{n}.$$

7.2. Convergencia de sucesiones de variables aleatorias

En esta sección vamos a intentar concretar cómo la media muestral y la media poblacional de una **muestra aleatoria simple** se van pareciendo, así como la distribución de la **media muestral** se va “acercando” a la normalidad.

Para ello, necesitamos introducir un conjunto de conceptos relacionados con la convergencia de variables aleatorias.

En primer lugar, introduciremos el concepto de **sucesión de variables aleatorias**:

Definición de sucesión de variables aleatorias. Consideremos un experimento aleatorio sobre un **espacio muestral** Ω . Sea P una probabilidad definida sobre el conjunto de sucesos de Ω . Entonces, si $X_1, X_2, \dots, X_n, \dots$ son variables aleatorias definidas sobre Ω, P , diremos que forman una **sucesión de variables aleatorias** y lo denotaremos por $\{X_n\}_{n=1}^\infty$.

Ejemplo: lanzamiento de un dado

Consideremos el experimento aleatorio de ir lanzando un dado no trucado. Definimos la variable aleatoria X_n como el resultado del dado el lanzamiento n -ésimo.

Entonces, la sucesión de variables aleatorias X_1, \dots, X_n, \dots es la asociada al lanzamiento del dado.

¡Ojo! no confundir la sucesión de variables aleatorias X_1, \dots, X_n, \dots con la sucesión de resultados de dichas variables aleatorias x_1, \dots, x_n, \dots . Lo primero correspondería a variables aleatorias con su función de probabilidad, esperanza, varianza, etc., y lo segundo sería simplemente una sucesión numérica de valores enteros entre 1 y 6.

7.2.1. Convergencia casi segura

Definición de convergencia casi segura. Sea X_1, \dots, X_n, \dots una sucesión de variables aleatorias y sea X una variable aleatoria definida sobre el mismo espacio muestral Ω y con la misma probabilidad de sucesos. Diremos que la sucesión $\{X_n\}_{n=1}^{\infty}$ converge **casi seguramente** hacia X si

$$P\left(\{w \in \Omega \mid \lim_{n \rightarrow \infty} X_n(w) = X(w)\}\right) = 1.$$

Lo denotaremos por $X_n \xrightarrow{c.s.} X$.

Es decir, si el conjunto de elementos w del espacio muestral Ω que cumplen que el límite de la sucesión de números reales $(X_n(w))_n$ tiende a $X(w)$ tiene probabilidad 1.

De ahí viene el nombre de **casi segura**: el conjunto de valores w del espacio muestral tal que la sucesión numérica $(X_n(w))_n$ **no converge** a $X(w)$ tiene probabilidad 0.

Comprobar la **convergencia casi segura** a partir de la definición puede ser muy complicado. Por suerte, existe la proposición siguiente que nos hace la vida más fácil:

Proposición. Sea X_1, \dots, X_n, \dots una sucesión de variables aleatorias y sea X una variable aleatoria definida sobre el mismo espacio muestral Ω y con la misma probabilidad de sucesos. Entonces $X_n \xrightarrow{c.s.} X$ si, y sólo si, para todo valor $\epsilon > 0$, la serie siguiente

$$\sum_{n=1}^{\infty} P(|X_n - X| > \epsilon),$$

es convergente.

** Ejemplo: convergencia casi segura frecuencias de un dado**

Veamos si la sucesión $\{X_n\}_{n=1}^{\infty}$ tiene convergencia **casi segura** hacia la variable X cuya **función de probabilidad** es:

X	1	2	3	4	5	6
P_X	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$

En este caso el espacio muestral Ω es $\Omega = \{1, 2, 3, 4, 5, 6\}$ y la **función de probabilidad** de cada X_i corresponde con la tabla anterior.

Seguidamente, de cara a aplicar la proposición anterior, vamos a hallar la **función de probabilidad** de la variable $D_n = X_n - X$. Los valores de la variable anterior son: $D_n(\Omega) = \{-5, -4, -3, -2, -1, 0, 1, 2, 3, 4, 5\}$.

La **función de probabilidad** conjunta de la variable (X_n, X) será al ser X_n y X independientes:

$$P_{X_n X}(x_n, x) = P_{X_n}(x_n) \cdot P_X(x) = \frac{1}{6} \cdot \frac{1}{6} = \frac{1}{36},$$

para todo $x_n = 1, 2, 3, 4, 5, 6$ y para todo $x = 1, 2, 3, 4, 5, 6$.

La **función de probabilidad** de la variable D_n será:

$$\begin{aligned} P_{D_n}(-5) &= P_{X_n X}(1, 6) = \frac{1}{36}, \\ P_{D_n}(-4) &= P_{X_n X}(2, 6) + P_{X_n X}(1, 5) = \frac{2}{36}, \\ P_{D_n}(-3) &= P_{X_n X}(3, 6) + P_{X_n X}(2, 5) + P_{X_n X}(1, 4) = \frac{3}{36}, \\ P_{D_n}(-2) &= P_{X_n X}(4, 6) + P_{X_n X}(3, 5) + P_{X_n X}(2, 4) + P_{X_n X}(1, 3) = \frac{4}{36}, \\ P_{D_n}(-1) &= P_{X_n X}(5, 6) + P_{X_n X}(4, 5) + P_{X_n X}(3, 4) + P_{X_n X}(2, 3) + P_{X_n X}(1, 2) = \frac{5}{36}, \\ P_{D_n}(0) &= P_{X_n X}(6, 6) + P_{X_n X}(5, 5) + P_{X_n X}(4, 4) + P_{X_n X}(3, 3) + P_{X_n X}(2, 2) + P_{X_n X}(1, 1) = \frac{6}{36}, \\ P_{D_n}(1) &= P_{X_n X}(6, 5) + P_{X_n X}(5, 4) + P_{X_n X}(4, 3) + P_{X_n X}(3, 2) + P_{X_n X}(2, 1) = \frac{5}{36}, \\ P_{D_n}(2) &= P_{X_n X}(6, 4) + P_{X_n X}(5, 3) + P_{X_n X}(4, 2) + P_{X_n X}(3, 1) = \frac{4}{36}, \\ P_{D_n}(3) &= P_{X_n X}(6, 3) + P_{X_n X}(5, 2) + P_{X_n X}(4, 1) = \frac{3}{36}, \\ P_{D_n}(4) &= P_{X_n X}(6, 2) + P_{X_n X}(5, 1) = \frac{2}{36}, \\ P_{D_n}(5) &= P_{X_n X}(6, 1) = \frac{1}{36}. \end{aligned}$$

Sea ϵ un valor real entre 0 y 1: $0 < \epsilon < 1$. Entonces el suceso $\{|D_n| > \epsilon\}$ será el complementario del suceso $\{D_n = 0\}$ ya que el único valor entre -5 y 5 que no cumple $|D_n| > \epsilon$ es el valor $D_n = 0$. Por tanto:

$$P(|D_n| > \epsilon) = 1 - P(D_n = 0) = 1 - P_{D_n}(0) = 1 - \frac{1}{6} = \frac{5}{6}.$$

La serie $\sum_{n=1}^{\infty} \frac{5}{6}$ no es convergente de forma obvia. Por tanto, deducimos que la sucesión $\{X_n\}_{n=1}^{\infty}$ no converge **casi seguramente** hacia la variable X .

7.2.2. Convergencia en probabilidad

Definición de convergencia en probabilidad. Sea X_1, \dots, X_n, \dots una sucesión de variables aleatorias y sea X una variable aleatoria definida sobre el mismo espacio muestral Ω y con la misma probabilidad de sucesos. Diremos que la sucesión $\{X_n\}_{n=1}^{\infty}$ converge **en probabilidad** hacia X si para cualquier valor $\epsilon > 0$,

$$\lim_{n \rightarrow \infty} P(|X_n - X| > \epsilon) = 0.$$

Lo denotaremos por $X_n \xrightarrow{c.p.} X$.

El límite de la probabilidad de los sucesos formados por los $w \in \Omega$ tal que $|X_n(w) - X(w)| > \epsilon$ vale 0.

Observación. Una definición equivalente de **convergencia en probabilidad** es que para todo valor $\epsilon > 0$,

$$\lim_{n \rightarrow \infty} P(|X_n(w) - X(w)| \leq \epsilon) = 1.$$

Observación. La convergencia **casi segura** implica la convergencia **en probabilidad** ya que si la sucesión $\{X_n\}_{n=1}^{\infty}$ converge **casi seguramente** hacia X , la serie $\sum_{n=1}^{\infty} P(|X_n - X| > \epsilon)$ será convergente y, por tanto, el límite de su término $P(|X_n - X| > \epsilon)$ tenderá a cero, hecho que equivale a la convergencia **en probabilidad**.

El siguiente resultado nos puede ayudar algunas veces a comprobar la **convergencia en probabilidad**:

Proposición. Sea X_1, \dots, X_n, \dots una sucesión de variables aleatorias. Sea μ_n el valor medio de la variable X_n , $E(X_n) = \mu_n$ y σ_n^2 su varianza: $\text{Var}(X_n) = \sigma_n^2$. Supongamos que $\lim_{n \rightarrow \infty} \sigma_n^2 = 0$. Entonces,

$$X_n - \mu_n \xrightarrow{c.p.} 0.$$

Demostración

Usando la desigualdad de Chebyschev, podemos escribir:

$$P(|X_n - \mu_n| > \epsilon) \leq \frac{\sigma_n^2}{\epsilon^2}.$$

Tomando límite a cada parte de la desigualdad anterior tenemos:

$$0 \leq \lim_{n \rightarrow \infty} P(|X_n - \mu_n| > \epsilon) \leq \lim_{n \rightarrow \infty} \frac{\sigma_n^2}{\epsilon^2} = 0,$$

de donde deducimos que $\lim_{n \rightarrow \infty} P(|X_n - \mu_n| > \epsilon) = 0$, tal como queríamos ver.

Ejemplo del lanzamiento de un dado (continuación)

En el ejemplo anterior del lanzamiento de un dado, no hay convergencia en probabilidad ya que comprobamos que para $0 < \epsilon < 1$,

$$P(|X_n - \mu| > \epsilon) = \frac{5}{6}.$$

Por tanto, para $0 < \epsilon < 1$, $\lim_{n \rightarrow \infty} P(|X_n - \mu| > \epsilon) = \frac{5}{6} \neq 0$.

Ejemplo: convergencia en probabilidad distribuciones exponenciales

Consideremos las variables aleatorias X_n con función de densidad:

$$f_{X_n}(x) = \begin{cases} \lambda \cdot n \cdot e^{-\lambda \cdot n \cdot x}, & \text{si } x > 0, \\ 0, & \text{en caso contrario.} \end{cases}$$

Estas variables X_n tienen distribución exponencial de parámetro $\lambda \cdot n$.

Veamos que $\{X_n\}_{n=1}^{\infty} \xrightarrow{c.p.} 0$.

Dado $\epsilon > 0$, calculemos $P(|X_n| > \epsilon)$:

$$P(|X_n| > \epsilon) = \int_{\epsilon}^{\infty} \lambda \cdot n \cdot e^{-\lambda \cdot n \cdot x} dx = \lambda \cdot n \cdot \left[\frac{1}{-\lambda \cdot n} \cdot e^{-\lambda \cdot n \cdot x} \right]_{\epsilon}^{\infty} = e^{-\lambda \cdot n \cdot \epsilon} \xrightarrow{n \rightarrow \infty} 0,$$

tal como queríamos ver.

7.2.3. Convergencia en ley o en distribución

Definición de convergencia en ley o distribución. Sea X_1, \dots, X_n, \dots una sucesión de variables aleatorias y sea X una variable aleatoria definida sobre el mismo espacio muestral Ω y con la misma probabilidad

de sucesos. Sea F_{X_n} y F_X las funciones de distribución de la variable X_n y X , respectivamente. Diremos que la sucesión $\{X_n\}_{n=1}^{\infty}$ converge **en ley**, o **en distribución** hacia X si,

$$\lim_{n \rightarrow \infty} F_{X_n}(x) = F(x),$$

para todo valor $x \in \mathbb{R}$.

Lo denotaremos por $X_n \xrightarrow{L} X$.

El resultado siguiente simplifica algunas veces comprobar que la sucesión X_n converge en ley hacia X :

Proposición. Sea $\{X_n\}_{n=1}^{\infty}$ una sucesión de variables aleatorias y sea X una variable aleatoria definida sobre el mismo espacio muestral Ω y con la misma probabilidad de sucesos. Sean ϕ_{X_n} y ϕ_X las funciones características de X_n y X , respectivamente. Entonces, la sucesión converge **en ley** hacia X , $X_n \xrightarrow{L} X$, si, y sólo si,

$$\lim_n \phi_{X_n}(t) = \phi_X(t),$$

para cualquier número $t \in \mathbb{R}$.

Ejemplo de la distribución binomial $B(n, p)$

Veamos que si $X_n = B(n, p_n)$ tiene distribución binomial de parámetros n y p_n , con $p_n = \frac{\lambda}{n}$, con λ fijo,

$$B(n, p) \xrightarrow{L} Poiss(\lambda).$$

En el tema de distribuciones notables demostramos que para todo $k \in \{0, \dots, n\}$,

$$P(X_n = k) = \binom{n}{k} \cdot p_n^k \cdot (1 - p_n)^{n-k} \xrightarrow{n \rightarrow \infty} P(X = k) = \frac{\lambda^k}{k!} \cdot e^{-\lambda}.$$

Entonces tenemos que dado $x \in \mathbb{R}$, existe $k \in \{0, \dots, n\}$, tal que $k \leq x < k + 1$. Por tanto,

$$\begin{aligned} \lim_{n \rightarrow \infty} F_{X_n}(x) &= \lim_{n \rightarrow \infty} F_{X_n}(k) = \lim_{n \rightarrow \infty} P(X_n = 0) + \dots + P(X_n = k) \\ &= \lim_{n \rightarrow \infty} P(X_n = 0) + \dots + \lim_{n \rightarrow \infty} P(X_n = k) \\ &= P(X = 0) + \dots + P(X = k) \\ &= F_X(k) = F_X(x), \end{aligned}$$

tal como queríamos demostrar.

7.2.4. Relaciones entre las distintas convergencias

El resultado siguiente nos dice cuando un tipo de convergencia implica la otra:

Proposición. Sea $\{X_n\}_{n=1}^{\infty}$ una sucesión de variables aleatorias y sea X una variable aleatoria definida sobre el mismo espacio muestral Ω y con la misma probabilidad de sucesos. Entonces:

- Si $X_n \xrightarrow{c.s.} X$, entonces $X_n \xrightarrow{c.p.} X$.
- Si $X_n \xrightarrow{c.p.} X$, entonces $X_n \xrightarrow{L} X$.

En resumen, la convergencia más fuerte es la **casi segura**, luego vendría la convergencia **en probabilidad** y, por último, la convergencia **en ley**:

$$\text{Conv. casi segura} \Rightarrow \text{Conv. en probabilidad} \Rightarrow \text{Conv. en ley.}$$

7.3. Leyes de los grandes números

Como ya comentamos al principio del tema, las **leyes de los grandes números** estudian el comportamiento de la **media muestral** \bar{X}_n cuando la sucesión de variables aleatorias $\{X_n\}_{n=1}^{\infty}$ se va hacia infinito.

Más concretamente, diremos que una sucesión de variables aleatorias $\{X_n\}_{n=1}^{\infty}$ cumple una **ley de los grandes números** si existe un sucesión numérica $(a_n)_n$ tal que la sucesión de variables aleatorias $\{\bar{X}_n - a_n\}$ converge “de alguna manera” de las que hemos visto hacia 0.

Si este “algunas manera” es la convergencia más fuerte, o la **casi segura**, tendremos la **ley fuerte de los grandes números**.

En cambio, si la convergencia es **en probabilidad**, tendremos la **ley débil de los grandes números**.

7.3.1. Leyes débiles de los grandes números

Teorema. Ley débil de los grandes números. Sea $\{X_n\}_{n=1}^{\infty}$ una sucesión de variables aleatorias independientes dos a dos tal que sus varianzas existen y están acotadas por una constante independiente de n . Entonces,

$$\bar{X}_n - \frac{1}{n} \sum_{i=1}^n \mu_i \xrightarrow{c.p.} 0,$$

donde $\mu_i = E(X_i)$.

Dicho en otras palabras: en las condiciones de la proposición anterior, la diferencia entre la sucesión de **medias muestrales** como variables aleatorias y la sucesión numérica de las medias poblacionales de dichas variables aleatorias tiende **en probabilidad** hacia 0.

Demostración

Como las variables son independientes dos a dos la varianza de la suma es la suma de las varianzas:

$$\text{Var}(\bar{X}_n) = \frac{1}{n^2} \text{Var}\left(\sum_{i=1}^n X_i\right) = \frac{1}{n^2} \sum_{i=1}^n \sigma_i^2,$$

donde $\sigma_i^2 = \text{Var}(X_i)$.

Sabemos por hipótesis que existe una constante M tal que $\sigma_i^2 \leq M$ para todo i . Por tanto,

$$\text{Var}(\bar{X}_n) = \frac{1}{n^2} \sum_{i=1}^n \sigma_i^2 \leq \frac{1}{n^2} Mn = \frac{M}{n}.$$

El valor del valor medio de la media muestral será:

$$E(\bar{X}_n) = \frac{1}{n} \sum_{i=1}^n E(X_i) = \frac{1}{n} \sum_{i=1}^n \mu_i.$$

Usando la desigualdad de Chebyschev, deducimos, dado un $\epsilon > 0$:

$$P\left(\left|\bar{X}_n - \frac{1}{n} \sum_{i=1}^n \mu_i\right| > \epsilon\right) \leq \frac{\text{Var}(\bar{X}_n)}{\epsilon^2} \leq \frac{M}{n\epsilon^2}.$$

Por tanto, tomando límites en las dos partes de la desigualdad anterior, deducimos

$$\lim_{n \rightarrow \infty} P\left(\left|\bar{X}_n - \frac{1}{n} \sum_{i=1}^n \mu_i\right| > \epsilon\right) = 0,$$

tal como queríamos ver.

Del teorema anterior obtenemos las consecuencias siguientes:

Corolario. Sea $\{X_n\}_{n=1}^{\infty}$ una sucesión de variables aleatorias independientes dos a dos tal que todas tienen la misma esperanza μ y la misma varianza σ^2 . Entonces,

$$\bar{X}_n \xrightarrow{c.p.} \mu,$$

Demostración

En este caso tenemos que $\mu_i = \mu$ y, por tanto, $\frac{1}{n} \sum_{i=1}^n \mu_i = \frac{1}{n} \cdot n\mu = \mu$. Si aplicamos el teorema de la **ley débil de los grandes números** nos sale el resultado enunciado.

Corolario. Sea $\{X_n\}_{n=1}^{\infty}$ una sucesión de variables aleatorias independientes dos a dos e idénticamente distribuidas tal que todas tienen la misma esperanza μ . Entonces,

$$\bar{X}_n \xrightarrow{c.p.} \mu,$$

Demostración

Trivial a partir del Corolario anterior.

Ejemplo: lanzamiento de una moneda

Vamos a simular la **ley débil de los grandes números** en el caso en que el experimento aleatorio sea el lanzamiento de una moneda.

En este caso, tendremos que las variables aleatorias X_n tendrán distribución de Bernoulli de parámetro $p = \frac{1}{2}$.

La variable \bar{X}_n representa la proporción de caras ($X_n = 1$) en el lanzamiento de la moneda n veces. Nos preguntamos si dicha proporción de caras tiende al parámetro p en probabilidad.

Vamos a hallar una muestra para cada variable $\bar{X}_n = \frac{\sum_{i=1}^n X_i}{n}$.

Para ello, vamos a repetir el experimento de lanzar la moneda $N = 100$ veces y lo repetimos $k = 500$ ocasiones.

Los resultados estarán en una matriz $k \times N = 500 \times 100$ donde cada fila de la matriz simulará una repetición del experimento de lanzar la moneda $N = 100$ veces.

Dada la fila i -ésima, iremos calculando $\bar{X}_1^{(i)}, \bar{X}_2^{(i)}, \dots, \bar{X}_{N=100}^{(i)}$.

Luego, fijado un ϵ , para cada n , aproximaremos la probabilidad $P(|\bar{X}_n - \frac{1}{2}| > \epsilon)$ usando la fórmula de Laplace:

$$p_n = P\left(\left|\bar{X}_n - \frac{1}{2}\right| > \epsilon\right) \approx \frac{\#\left\{i \text{ tal que } \left|\bar{X}_n^{(i)} - \frac{1}{2}\right| > \epsilon\right\}}{k}.$$

Para comprobar dicha afirmación, la idea es hallar para cada valor n , una muestra para cada variable $\bar{X}_n = \frac{\sum_{i=1}^n X_i}{n}$.

Para hallar una muestra de cada variable \bar{X}_n , seguimos los pasos siguientes:

- En primer lugar, simulamos la repetición del experimento de lanzar la moneda $N = 100$ veces y lo repetimos $k = 500$ ocasiones. Los resultados estarán en una matriz $k \times N = 500 \times 100$ donde cada fila de la matriz simulará una repetición del experimento de lanzar la moneda $N = 100$ veces:

```
N=100
k=500
set.seed(2019)
## fijamos la semilla de aleatoriedad
## por reproducibilidad
valores.experimento=matrix(sample(c(0,1),N*k,replace=TRUE),k,N)
```

Los primeros resultados son:

```
##      [,1] [,2] [,3] [,4] [,5] [,6] [,7] [,8] [,9] [,10] [,11] [,12]
## [1,]     0     1     1     1     1     1     1     0     1     1     0     0
## [2,]     0     0     0     0     0     1     0     1     1     0     1     1
## [3,]     1     1     1     0     0     1     1     0     0     1     0     1
## [4,]     0     0     1     1     1     0     0     0     1     1     0     1
## [5,]     0     1     1     1     1     1     0     1     0     1     1     1
...
...
```

- En segundo lugar, dada la fila i -ésima de la matriz anterior, iremos calculando $\bar{X}_1^{(i)}, \bar{X}_2^{(i)}, \dots, \bar{X}_{N=100}^{(i)}$ guardando los resultados en una matriz de medias muestrales. Antes de nada, creamos la función que nos realizará la operación anterior dado un vector cualquiera x :

```
cálculo.xnbarra = function(x){
  return(cumsum(x)/(1:length(x)))
}
```

A partir de la matriz de los resultados, aplicamos la función anterior a cada fila y hallaremos una matriz con todas las $\bar{X}_n^{(i)}$:

```
matriz.medias.muestrales = t(apply(valores.experimento,
                                    1,cálculo.xnbarra))
```

La columna j -ésima de la matriz `matriz.medias.muestrales` contiene una muestra de $k = 500$ valores de la variable \bar{X}_j .

- En último lugar, fijado un ϵ , para cada n , aproximaremos la probabilidad $P\left(\left|\bar{X}_n - \frac{1}{2}\right| > \epsilon\right)$ usando la fórmula de Laplace:

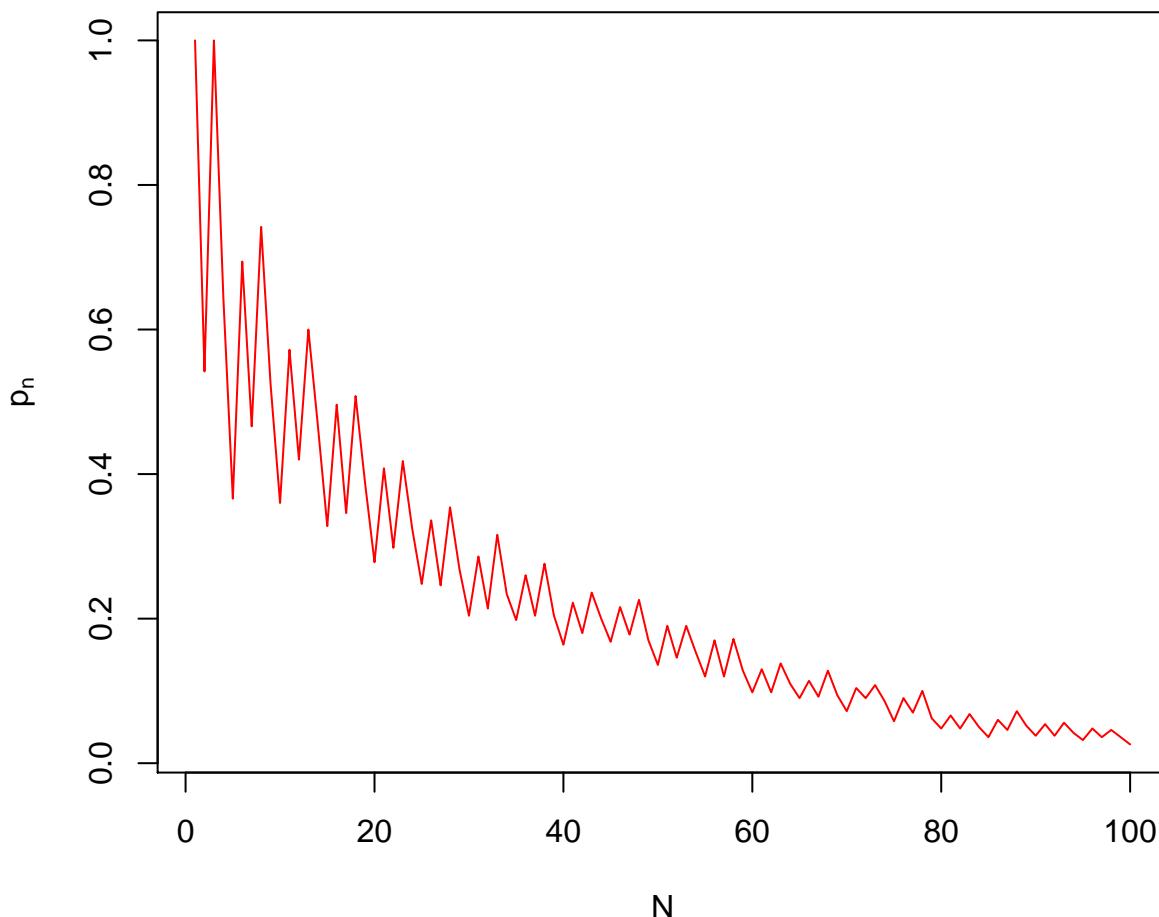
$$p_n = P\left(\left|\bar{X}_n - \frac{1}{2}\right| > \epsilon\right) \approx \frac{\#\left\{i \text{ tal que } \left|\bar{X}_n^{(i)} - \frac{1}{2}\right| > \epsilon\right\}}{k}.$$

La columna j -ésima de la matriz `matriz.medias.muestrales` es una muestra de la variable \bar{X}_j . Por tanto, para hallar la aproximación de p_n , miramos cuántos valores de la columna j -ésima de la matriz anterior verifican $\left|\bar{X}_j^l - \frac{1}{2}\right| > \epsilon$, para $l = 1, \dots, k$:

```
epsilon=0.1
probabilidades.pn= colSums(abs(matriz.medias.muestrales-0.5) > epsilon)/k
```

Para ver los resultados, dibujamos el gráfico n vs. p_n :

Simulación de la convergencia en probabilidad



Observamos que las probabilidades tienden a cero tal como nos dice el **Teorema de la ley débil de los grandes números**.

7.3.2. Convergencia de los momentos muestrales

‘Dada una sucesión de variables aleatorias, definimos los momentos muestrales de la forma siguiente:

Definición de los momentos muestrales de una sucesión de variables aleatorias. Sea $\{X_n\}_{n=1}^{\infty}$ una sucesión de variables aleatorias. Dado k valor entero positivo, definimos el **momento muestral** de orden k como la sucesión de variables aleatorias siguientes:

$$M_k^{(n)} = \frac{1}{n} \sum_{i=1}^n X_i^k.$$

Observación: el **momento muestral** de orden $k = 1$ es la **media muestral** \bar{X}_n .

Definición de los momentos muestrales centrados de una sucesión de variables aleatorias. Sea $\{X_n\}_{n=1}^{\infty}$ una sucesión de variables aleatorias. Dado k valor entero positivo, definimos el **momento muestral centrado en la media** de orden k como la sucesión de variables aleatorias siguientes:

$$MC_k^{(n)} = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X}_n)^k.$$

Observación: el **momento muestral centrado en la media** de orden $k = 2$ es la **varianza muestral** $S_{X_n}^2$.

Definición de la covarianza y el coeficiente de correlación muestral de una sucesión de variables aleatorias. Sea $\{(X_n, Y_n)\}_{n=1}^{\infty}$ una sucesión de variables aleatorias bidimensionales. Definimos la **covarianza muestral** como la sucesión de variables aleatorias siguientes:

$$S_{X_n, Y_n} = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X}_n)(Y_i - \bar{Y}_n),$$

y el **coeficiente de correlación muestral** como la sucesión de variables aleatorias siguientes:

$$R_{X_n, Y_n} = \frac{S_{X_n, Y_n}}{\sqrt{S_{X_n}^2 S_{Y_n}^2}}.$$

Dada $\{X_n\}_{n=1}^{\infty}$ una sucesión de variables aleatorias y k un valor entero positivo, en el tema de Complementos de variables aleatorias, definimos los momentos y los momentos centrales de orden k para cada de dichas variables como:

$$m_k^{(n)} = E(X_n^k), \quad \mu_k^{(n)} = E((X_n - \mu_n)^k),$$

donde μ_n es el valor medio de la variable X_n : $\mu_n = E(X_n)$.

Así mismo, dada $\{(X_n, Y_n)\}_{n=1}^{\infty}$ una sucesión de variables aleatorias bidimensionales, en el tema de variables aleatorias bidimensionales definimos para cada variable (X_n, Y_n) la covarianza $\sigma_{X_n Y_n}$ y el coeficiente de correlación $\rho_{X_n Y_n}$:

$$\sigma_{X_n Y_n} = E((X_n - \mu_{X_n})(Y_n - \mu_{Y_n})), \quad \rho_{X_n Y_n} = \frac{\sigma_{X_n Y_n}}{\sqrt{\sigma_{X_n}^2 \sigma_{Y_n}^2}}.$$

Dada una sucesión de variables aleatorias $\{X_n\}_{n=1}^{\infty}$, el resultado siguiente nos relaciona los **momentos muestrales** y los **momentos muestrales centrados en la media** con los **momentos** y los **momentos centrales** de cada variable:

Teorema. Convergencia de los momentos muestrales y los momentos muestrales centrados en la media. Sea $\{X_n\}_{n=1}^{\infty}$ una sucesión de variables aleatorias **independientes dos a dos e idénticamente distribuidas** y dado un entero positivo k , supongamos que para cada n , existe el **momento de orden** k , m_k y el **momento central de orden** k , μ_k , que no dependerán de n al ser idénticamente distribuidas. Entonces las sucesiones de variables aleatorias $\{M_n^{(k)}\}_{n=1}^{\infty}$ y $\{MC_n^{(k)}\}_{n=1}^{\infty}$ tienden a m_k y μ_k , respectivamente, en **probabilidad**

$$M_n^{(k)} \xrightarrow{c.p.} m_k, \quad MC_n^{(k)} \xrightarrow{c.p.} \mu_k.$$

Demostración

Consideremos la sucesión de variables aleatorias $\{X_n^k\}_{n=1}^\infty$. Como las variables aleatorias de la sucesión $\{X_n\}_{n=1}^\infty$ son independientes dos a dos e idénticamente distribuidas, las variables de la sucesión $\{X_n^k\}_{n=1}^\infty$ también lo serán.

La idea es aplicar la **ley débil de los grandes números** a la sucesión anterior.

El valor medio de cada variable de la sucesión $\{X_n^k\}_{n=1}^\infty$ será: $\tilde{\mu}_n^{(k)} = E(X_n^k) = m_k$ el momento de orden k .

Entonces, si hacemos $\frac{1}{n} \sum_{i=1}^n \tilde{\mu}_n^{(k)}$ obtenemos: $\frac{1}{n} n \cdot m_k = m_k$.

Aplicando la **ley débil de los grandes números** a la sucesión $\{X_n^k\}_{n=1}^\infty$, tendremos que

$$\overline{X^k}_n \xrightarrow{c.p.} m_k,$$

pero $\overline{X^k}_n$ vale:

$$\overline{X^k}_n = \frac{1}{n} \sum_{i=1}^n X_i^k,$$

variable aleatoria que coincide con el momento muestral de orden k , $M_n^{(k)}$, tal como queríamos demostrar.

Dejamos como ejercicio la demostración de los momentos centrales. Razonando de la misma manera, no tiene dificultad alguna.

Enunciemos ahora el resultado para las covarianzas y las correlaciones muestrales:

Teorema: convergencia de la covarianza y el coeficiente de correlación muestrales. Sea $\{(X_n, Y_n)\}_{n=1}^\infty$ una sucesión de variables aleatorias bidimensionales independientes dos a dos e idénticamente distribuidas. Sea σ_{XY} , ρ_{XY} la covarianza y el coeficiente de correlación de cada par de variables que, al ser idénticamente distribuidas, no dependen de n . Entonces las sucesiones de las covarianzas muestrales $\{S_{X_n, Y_n}\}_{n=1}^\infty$ y los coeficientes de correlación muestrales $\{R_{X_n, Y_n}\}_{n=1}^\infty$ tienden en probabilidad hacia σ_{XY} y ρ_{XY} , respectivamente:

$$S_{X_n, Y_n} \xrightarrow{c.p.} \sigma_{XY}, \quad R_{X_n, Y_n} \xrightarrow{c.p.} \rho_{XY}.$$

Demostración

Para la demostración basta aplicar la **ley débil de los grandes números** a las sucesiones $\{S_{X_n, Y_n}\}_{n=1}^\infty$ y $\{R_{X_n, Y_n}\}_{n=1}^\infty$. Dejamos los detalles como ejercicio.

7.3.3. Leyes fuertes de los grandes números

Vamos a dar una versión de la ley débil de los grandes números pero en lugar de tener convergencia en probabilidad, tendremos convergencia casi segura.

Teorema de Kolmogorov. Ley fuerte de los grandes números. Sea $\{X_n\}_{n=1}^{\infty}$ una sucesión de variables aleatorias independientes y con varianza σ_n^2 . Supongamos que la serie $\sum_{n=1}^{\infty} \frac{\sigma_n^2}{n^2}$, es convergente. Entonces la sucesión de las medias muestrales $\{\bar{X}_n\}_{n=1}^{\infty}$ cumplen la llamada **ley fuerte de los grandes números**:

$$\bar{X}_n - \frac{1}{n} \sum_{i=1}^n \mu_i \xrightarrow{c.s.} 0.$$

Asociados al resultado anterior tenemos los corolarios siguientes:

Corolario. Sea $\{X_n\}_{n=1}^{\infty}$ una sucesión de variables aleatorias independientes y con varianza σ_n^2 . Supongamos que existe una constante M tal que todas las varianzas están acotadas por M : $\sigma_n^2 \leq M$, para todo n . Entonces la sucesión de las medias muestrales $\{\bar{X}_n\}_{n=1}^{\infty}$ cumplen la llamada **ley fuerte de los grandes números**:

$$\bar{X}_n - \frac{1}{n} \sum_{i=1}^n \mu_i \xrightarrow{c.s.} 0.$$

Demostración

Si $\sigma_n^2 \leq M$ para todo n , la serie numérica $\sum_{n=1}^{\infty} \frac{\sigma_n^2}{n^2}$ será convergente ya que, por el criterio de acotación,

$$\sum_{n=1}^{\infty} \frac{\sigma_n^2}{n^2} \leq M \sum_{n=1}^{\infty} \frac{1}{n^2},$$

que es convergente.

Entonces aplicando el **Teorema de Kolmogorov** o la **ley fuerte de los grandes números**, tenemos el resultado.

Corolario. Sea $\{X_n\}_{n=1}^{\infty}$ una sucesión de variables aleatorias independientes de Bernoulli con el mismo parámetro p que es lo mismo que decir que son idénticamente distribuidas. Entonces la sucesión de las medias muestrales convergen **casi seguramente** hacia p :

$$\bar{X}_n \xrightarrow{c.s.} p.$$

Demostración

En este caso:

$$\frac{1}{n} \sum_{i=1}^n \mu_i = \frac{1}{n} \cdot n \cdot p = p.$$

También se verifica que $\sigma_n^2 = p(1-p)$. Por tanto, existe una constante M ($M = p(1-p)$) tal que $\sigma_n^2 \leq M$. Aplicando el Corolario anterior, obtenemos el resultado.

Ejemplo: lanzamiento moneda (continuación)

Vamos a repetir el ejemplo de las variables aleatorias de Bernoulli X_n , todas de parámetro $p = \frac{1}{2}$ y comprobar que las proporciones de caras cuando lanzamos la moneda n veces, es decir, las medias muestrales \bar{X}_n tienden **casi seguramente** hacia $p = \frac{1}{2}$.

La comprobación anterior es equivalente a ver que la serie:

$$\sum_{n=1}^{\infty} P(|\bar{X}_n - p| > \epsilon),$$

es convergente fijado $\epsilon > 0$.

Recordemos que en la variable `probabilidades.pn` calculábamos las probabilidades $P(|\bar{X}_n - p| > \epsilon)$ para un $\epsilon = 0,1$.

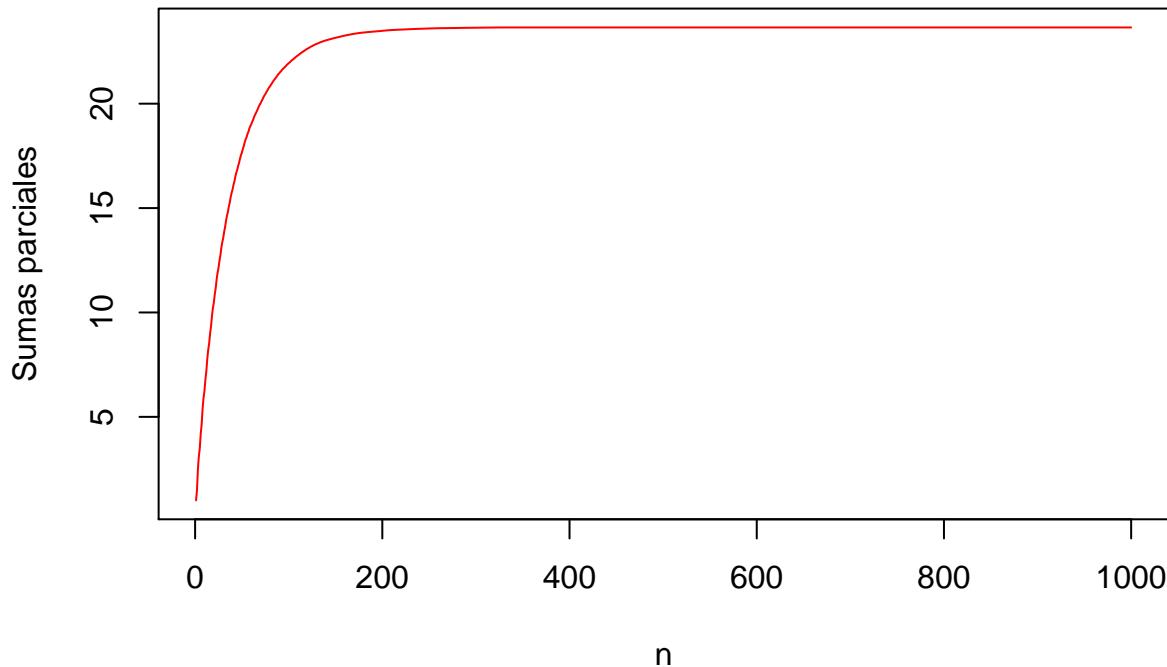
Comprobar que la serie anterior es convergente es equivalente a comprobar que las sumas parciales convergen:

```
cumsum(probabilidades.pn)
```

El problema es que la `n` y la `N` escogidas son demasiado pequeñas. Para realizar el experimento actual tenéis que considerar `n=1000` y `N=5000`. Id con cuidado que el programa os tardará un rato.

El gráfico de las sumas parciales se muestra a continuación:

```
N=1000
plot(1:N,cumsum(probabilidades.pn),xlab=expression(n),
      ylab="Sumas parciales",col='red', type='l')
```



Como se puede observar, la serie parece que converge.

7.4. Teorema Central del Límite

Sabemos que si una sucesión $\{X_n\}$ está formada por variables normales, la sucesión de medias muestrales $\left\{\bar{X}_n = \frac{\sum_{i=1}^n X_i}{n}\right\}_{n=1}^{\infty}$ también son normales ya que vimos en el tema de variables multidimensionales que si aplicamos una transformación afín (y, en particular, lineal) a una variable normal multidimensional, el resultado es una normal.

Para calcular la variable \bar{X}_n , es obvio que la transformación lineal es la siguiente:

$$\bar{X}_n = \left(\frac{1}{n}, \dots, \frac{1}{n} \right) \cdot \begin{pmatrix} X_1 \\ \vdots \\ X_n \end{pmatrix}.$$

Si además la sucesión de variables X_n son normales todas con media μ y varianza σ^2 , la sucesión $\{\bar{X}_n\}_{n=1}^{\infty}$ serán normales de media μ y varianza $\frac{\sigma^2}{n}$.

Estandarizando las variables anteriores, podemos concluir que las variables medias estandarizadas $Z_n = \left\{ \frac{\bar{X}_n - \mu}{\frac{\sigma}{\sqrt{n}}} \right\}_{n=1}^{\infty}$ todas son $N(0, 1)$.

El **Teorema Central del Límite** generaliza el resultado anterior en el sentido de que si las variables X_n no tienen por qué tener la distribución normal pero son independientes e idénticamente distribuidas, las variables Z_n correspondientes tienden **en ley** a una distribución normal estándar $N(0, 1)$.

En general, se dice que los valores medios de cualquier secuencia de números aproximadamente corresponden a una muestra de una normal.

7.4.1. Teorema Central del Límite

Teorema Central del Límite Sea $\{X_n\}_{n=1}^{\infty}$ una sucesión de variables aleatorias independientes e idénticamente distribuidas con $E(X_n) = \mu$ y $\text{Var}(X_n) = \sigma^2$ para todo n . Entonces:

$$\frac{\sum_{i=1}^n X_i - n\mu}{\sigma\sqrt{n}} \xrightarrow{L} N(0, 1).$$

Observación. Una condición equivalente a la tesis del **Teorema Central del Límite** es:

$$\frac{\bar{X}_n - \mu}{\frac{\sigma}{\sqrt{n}}} \xrightarrow{L} N(0, 1).$$

Basta dividir por n el numerador y el denominador de la tesis original del **Teorema Central del Límite**.

Para la demostración, usaremos dos propiedades de la función característica:

Proposición. Sea Y_1, \dots, Y_n n variables aleatorias independientes. Sea S_n la variable aleatoria suma de las variables anteriores, $S_n = \sum_{i=1}^n Y_i$. Entonces, para calcular ϕ_{S_n} , podemos usar la expresión siguiente::

$$\phi_{S_n}(w) = \phi_{Y_1}(w) \cdots \phi_{Y_n}(w),$$

donde w es cualquier valor real.

Demostración de la proposición

Por definición:

$$\begin{aligned}\phi_{S_n}(w) &= E(e^{iwS_n}) = E\left(e^{iw\sum_{i=1}^n Y_i}\right) = E(e^{iwY_1} \cdots e^{iwY_n}) \\ &\stackrel{Y_1, \dots, Y_n \text{ son independientes}}{=} E(e^{iwY_1}) \cdots E(e^{iwY_n}) \\ &= \phi_{Y_1}(w) \cdots \phi_{Y_n}(w).\end{aligned}$$

Proposición. Sea Y una variable aleatoria. Sea $U = kY$ la variable aleatoria Y multiplicada por un valor real k . Entonces, para calcular ϕ_U , podemos usar la expresión siguiente:

$$\phi_U(w) = \phi_Y(kw),$$

donde w es cualquier valor real.

Demostración de la proposición

Por definición:

$$\phi_U(w) = E(e^{iwU}) = E(e^{iwkY}) = \phi_Y(kw).$$

Demostración del Teorema Central del Límite

Usando la proposición que vimos al introducir la **convergencia en ley** que dice que una sucesión $\{X_n\}$ converge **en ley** hacia X si, y sólo si, $\lim_{t \rightarrow \infty} \phi_{X_n}(t) = \phi_X(t)$, donde ϕ representa la función característica y la condición anterior tiene que verificarse para todo valor $t \in \mathbb{R}$, basta demostrar que, si tomamos $Z_n = \frac{\sum_{i=1}^n X_i - n\mu}{\sigma\sqrt{n}}$,

$$\lim_{n \rightarrow \infty} \phi_{Z_n}(w) = \phi_Z(w),$$

para cualquier valor $w \in \mathbb{R}$, siendo $Z = N(0, 1)$.

Seguidamente, simplifiquemos la expresión $\phi_{Z_n}(w) = \phi_{\frac{\sum_{i=1}^n X_i - n\mu}{\sigma\sqrt{n}}}(w)$ usando las dos proposiciones anteriores. En primer lugar, teniendo en cuenta que las variables $\left\{ \frac{X_i - \mu}{\sigma\sqrt{n}} \right\}$ son independientes e idénticamente distribuidas, usando la primera proposición podemos escribir:

$$\phi_{Z_n}(w) = \left(\phi_{\frac{X-\mu}{\sigma\sqrt{n}}}(w) \right)^n,$$

donde X representa cualquiera de las variables X_i .

Usando la segunda proposición, podemos simplificar la expresión anterior aún más:

$$\phi_{Z_n}(w) = \left(\phi_{\frac{X-\mu}{\sigma\sqrt{n}}}(w) \right)^n = \left(\phi_{X-\mu} \left(\frac{w}{\sigma\sqrt{n}} \right) \right)^n.$$

Si desarrollamos por Taylor alrededor del valor $\hat{w} = 0$ la función característica $\phi_{X-\mu}(\hat{w})$ hasta segundo orden, obtenemos:

$$\phi_{X-\mu}(\hat{w}) = \phi_{X-\mu}(0) + \phi'_{X-\mu}(0)\hat{w} + \phi''_{X-\mu}(0)\frac{\hat{w}^2}{2} + O(\hat{w}^3),$$

donde $O(\hat{w}^3)$ simboliza los términos de orden \hat{w}^3 y superiores.

Los valores $\phi_{X-\mu}(0)$, $\phi'_{X-\mu}(0)$ y $\phi''_{X-\mu}(0)$ valen: (ver tema de Complementos de variables aleatorias)

$$\phi_{X-\mu}(0) = 1, \phi'_{X-\mu}(0) = \frac{1}{i}E(X-\mu) = 0, \phi''_{X-\mu}(0) = \frac{1}{i^2}E((X-\mu)^2) = -\sigma^2.$$

El desarrollo anterior será:

$$\phi_{X-\mu}(\hat{w}) = 1 - \frac{1}{2}\hat{w}^2\sigma^2 + O(\hat{w}^3),$$

Aplicando la expresión anterior para $\hat{w} = \frac{w}{\sigma\sqrt{n}}$, obtenemos:

$$\phi_{X-\mu}\left(\frac{w}{\sigma\sqrt{n}}\right) = 1 - \frac{1}{2}\left(\frac{w}{\sigma\sqrt{n}}\right)^2\sigma^2 + O\left(\frac{w}{\sigma\sqrt{n}}\right)^3 = 1 - \frac{w^2}{2n} + O\left(\frac{w^3}{n^{\frac{3}{2}}}\right),$$

La función característica de la variable Z_n será usando la expresión anterior:

$$\phi_{Z_n}(w) = \left(\phi_{X-\mu}\left(\frac{w}{\sigma\sqrt{n}}\right)\right)^n = \left(1 - \frac{w^2}{2n} + O\left(\frac{w^3}{n^{\frac{3}{2}}}\right)\right)^n$$

El objetivo es calcular el límite de la expresión anterior:

$$\lim_{n \rightarrow \infty} \phi_{Z_n}(w) = \lim_{n \rightarrow \infty} \left(1 - \frac{w^2}{2n} + O\left(\frac{w^3}{n^{\frac{3}{2}}}\right)\right)^n = \lim_{n \rightarrow \infty} e^{n \cdot \ln\left(1 - \frac{w^2}{2n} + O\left(\frac{w^3}{n^{\frac{3}{2}}}\right)\right)}.$$

Usando que para $z \approx 0$, $\ln(1-z) = z + O(z^2)$, el límite anterior será:

$$\lim_{n \rightarrow \infty} \phi_{Z_n}(w) = \lim_{n \rightarrow \infty} e^{n \cdot \left(-\frac{w^2}{2n} + O\left(\frac{w^4}{n^2}\right)\right)} = \lim_{n \rightarrow \infty} e^{\left(-\frac{w^2}{2} + O\left(\frac{w^4}{n}\right)\right)} = e^{-\frac{w^2}{2}},$$

y dicha expresión coincide con la función característica de la variable $N(0, 1)$, $\phi_Z(w)$.

Recordad que en el tema de Complementos de variables aleatorias vimos que si la variable U era $N(\mu, \sigma)$, $\phi_U(w) = e^{iw\mu - \frac{w^2\sigma^2}{2}}$. Aplicando la fórmula anterior para $\mu = 0$ y $\sigma = 1$, obtenemos $\phi_Z(w) = e^{-\frac{w^2}{2}}$.

7.4.2. Teorema Central del Límite en la práctica

El **Teorema Central del Límite** se aplica a la práctica en la forma siguiente:

Sea $\{X_n\}_{n=1}^{\infty}$ una sucesión de variables aleatorias independientes e idénticamente distribuidas con $E(X_i) = \mu$ y $\text{Var}(X_i) = \sigma^2$. Entonces, podemos aproximar para n grande ($n \geq 30$), la media muestral \bar{X}_n por:

$$\bar{X}_n = \frac{1}{n} \sum_{i=1}^n X_i \approx N\left(\mu, \frac{\sigma}{\sqrt{n}}\right),$$

o también:

$$\sum_{i=1}^n X_i \approx N(n\mu, \sigma\sqrt{n}),$$

Las aproximaciones anteriores se pueden obtener teniendo en cuenta que el **Teorema Central del Límite** nos dice que la variable $Z_n = \frac{\sum_{i=1}^n X_i - n\mu}{\sigma\sqrt{n}}$ es aproximadamente una $N(0, 1)$. Por tanto,

$$\frac{\sum_{i=1}^n X_i - n\mu}{\sigma\sqrt{n}} \approx N(0, 1), \Rightarrow \sum_{i=1}^n X_i \approx \sigma\sqrt{n} \cdot N(0, 1) + n\mu = N(n\mu, \sigma\sqrt{n}).$$

Dividiendo por n la aproximación anterior, obtenemos:

$$\bar{X}_n = \frac{1}{n} \sum_{i=1}^n X_i \approx \frac{1}{n} N(n\mu, \sigma\sqrt{n}) = N\left(\mu, \frac{\sigma}{\sqrt{n}}\right).$$

7.4.3. Teorema de Moivre-Laplace

Si aplicamos el **Teorema Central del Límite** en el caso en que las variables X_n son de Bernoulli de parámetro p , obtenemos el llamado **Teorema de Moivre-Laplace**:

Teorema de Moivre-Laplace Sea $\{X_n\}_{n=1}^{\infty}$ una sucesión de variables aleatorias independientes Bernoulli de parámetro p . La variable $\sum_{i=1}^n X_i$ será binomial de parámetros n y p , $B(n, p)$. Entonces:

$$\frac{B(n, p) - np}{\sqrt{n \cdot p \cdot (1-p)}} \xrightarrow{L} N(0, 1).$$

En la práctica, decimos que podemos aproximar una variable binomial de parámetros n y p por una distribución normal de parámetros $\mu = np$ y $\sigma = \sqrt{n \cdot p \cdot (1-p)}$:

$$B(n, p) \approx N(np, \sqrt{n \cdot p \cdot (1-p)}).$$

7.4.4. Aproximación de una suma de variables Poisson

Si aplicamos el **Teorema Central del Límite** en el caso en que las variables X_n son de Poisson de parámetro λ , obtenemos el resultado siguiente:

Proposición. Sea $\{X_n\}_{n=1}^{\infty}$ una sucesión de variables aleatorias independientes Poisson de parámetro λ . Entonces:

$$\frac{\sum_{i=1}^n X_i - n\lambda}{\sqrt{n \cdot \lambda}} \xrightarrow{L} N(0, 1).$$

Antes de ver la aplicación práctica del resultado anterior, veamos que suma de variables Poisson independientes de parámetro λ es una variable Poisson de parámetro $n\lambda$:

Proposición. Sea $\{X_n\}_{n=1}^{\infty}$ una sucesión de variables aleatorias independientes Poisson de parámetro λ . Entonces la variable $\sum_{i=1}^n X_i$ sigue la distribución de Poisson de parámetro $n\lambda$.

Demostración

En primer lugar, hallemos la función característica de la distribución de Poisson de parámetro λ . Sea $X = Poiss(\lambda)$. Su función característica en un valor w será:

$$\phi_X(w) = E(e^{iwX}) = \sum_{k=0}^{\infty} e^{iwk} \frac{\lambda^k}{k!} e^{-\lambda} = e^{-\lambda} \sum_{k=0}^{\infty} \frac{(\lambda e^{iw})^k}{k!} = e^{-\lambda} \cdot e^{\lambda e^{iw}} = e^{\lambda(e^{iw}-1)}.$$

Sea ahora la variable $S_n = \sum_{i=1}^n X_i$. Usando la proposición anterior que nos calcula la función característica de sumas de variables independientes, podemos escribir:

$$\phi_{S_n}(w) = \phi_{X_1}(w) \cdots \phi_{X_n}(w) = (e^{\lambda(e^{iw}-1)})^n = e^{n\lambda(e^{iw}-1)},$$

función característica que corresponde a una variable de Poisson de parámetro $n\lambda$, $Poiss(n\lambda)$.

Usando la proposición anterior, tenemos que la suma de variables Poisson independientes de parámetro λ sigue una distribución Poisson de parámetro $n\lambda$. Por tanto, podemos escribir usando el corolario del **Teorema Central del Límite** aplicado a variables Poisson:

$$Poiss(n\lambda) \approx N(n\lambda, \sqrt{n\lambda}).$$

Ejemplo: aplicación del Teorema de Moivre-Laplace

Sea X una distribución binomial de parámetros $n = 50$ y $p = \frac{1}{3}$.

Queremos conocer $P(X < 15)$ y $P(10 \leq X \leq 20)$.

Vamos a calcular las probabilidades anteriores usando el **Teorema de Moivre-Laplace**.

La variable X es aproximadamente una distribución normal X_N de parámetros $\mu = np = \frac{50}{3} = 16,6667$ y $\sigma = \sqrt{50 \cdot \frac{1}{3} \cdot \frac{2}{3}} = 3,3333$.

Por tanto:

$$\begin{aligned} P(X < 15) &= P(X \leq 14) \approx P(X_N \leq 14) = P\left(Z \leq \frac{14-16,6667}{3,3333}\right) \\ &= P(Z \leq -0,8) = 0,2119, \end{aligned}$$

$$\begin{aligned} P(10 \leq X \leq 20) &\approx P(10 \leq X_N \leq 20) = P\left(\frac{10-16,6667}{3,3333} \leq Z \leq \frac{20-16,6667}{3,3333}\right) \\ &= P(-2 \leq Z \leq 1) \\ &= P(Z \leq 1) - P(Z \leq -2) = 0,8413447 - 0,0227501 = 0,8186, \end{aligned}$$

donde $Z = N(0, 1)$.

Comparemos los valores aproximados anteriores con los valores “exactos” proporcionados por R:

```
pbinom(14, 50, 1/3)
```

```
## [1] 0.2612386
```

```
pbinom(20,50,1/3)-pbinom(9,50,1/3)
```

```
## [1] 0.8613685
```

Tenemos errores de 0.04938 y 0.04277, respectivamente.

Aunque n no es pequeño, $n = 50$, los errores anteriores no son demasiado pequeños.

Una razón por la que dichos errores no son pequeños es que aproximamos una distribución discreta (Binomial) cuyos valores van de 1 en 1 por una distribución normal, que es continua.

La corrección de continuidad de Fisher nos mejora la aproximación disminuyendo dichos errores.

7.4.5. Corrección de continuidad de Fisher

Cuando aplicamos el **Teorema Central del Límite** y aproximamos una distribución discreta que tiene valores enteros por una normal, hemos de aplicar lo que se llama **corrección de continuidad de Fisher**.

Sea X la variable discreta que queremos aproximar y X_N la variable normal que nos aparece cuando aplicamos el **Teorema Central del Límite**. Supongamos que queremos calcular $P(X \leq k)$, para un k entero. Entonces debemos hacer:

$$P(X \leq k) \approx P(X_N \leq k + 0,5).$$

Es decir, para tener en cuenta el valor k en la aproximación X_N hay que sumarle la mitad entre dos valores consecutivos (0.5 si los valores son enteros) de la variable X .

Id con cuidado, si queremos calcular $P(X < k)$, hay que hacer $P(X < k) = P(X \leq k - 1) \approx P(X_N \leq k - 1 + 0,5) = P(X_N \leq k - 0,5)$.

Ejemplo: continuación ejemplo anterior

Si aplicamos la continuidad de Fisher en el ejemplo anterior, obtenemos:

$$P(X < 15) = P(X \leq 14) \approx P(X_N \leq 14,5) = P\left(Z \leq \frac{14,5-16,6667}{3,3333}\right) = P(Z \leq -0,65) = 0,2578,$$

$$\begin{aligned} P(10 \leq X \leq 20) &= P(X \leq 20) - P(X \leq 9) \approx P(X_N \leq 20,5) - P(X_N \leq 9,5) \\ &= P\left(Z \leq \frac{20,5-16,6667}{3,3333}\right) - P\left(Z \leq \frac{9,5-16,6667}{3,3333}\right) = P(Z \leq 1,15) - P(Z \leq -2,15) \\ &= 0,8749281 - 0,0157776 = 0,8592, \end{aligned}$$

obteniendo unos errores de sólo 0.00339 y 0.00222, respectivamente.

7.4.6. Simulación del Teorema Central del Límite

Ejemplo de simulación de la aproximación de una variable binomial a una distribución normal

Para realizar la simulación anterior, consideremos una distribución binomial de parámetros $n = 100$ y $p = \frac{1}{2}$.

Según el **Teorema Central del Límite**, tenemos que

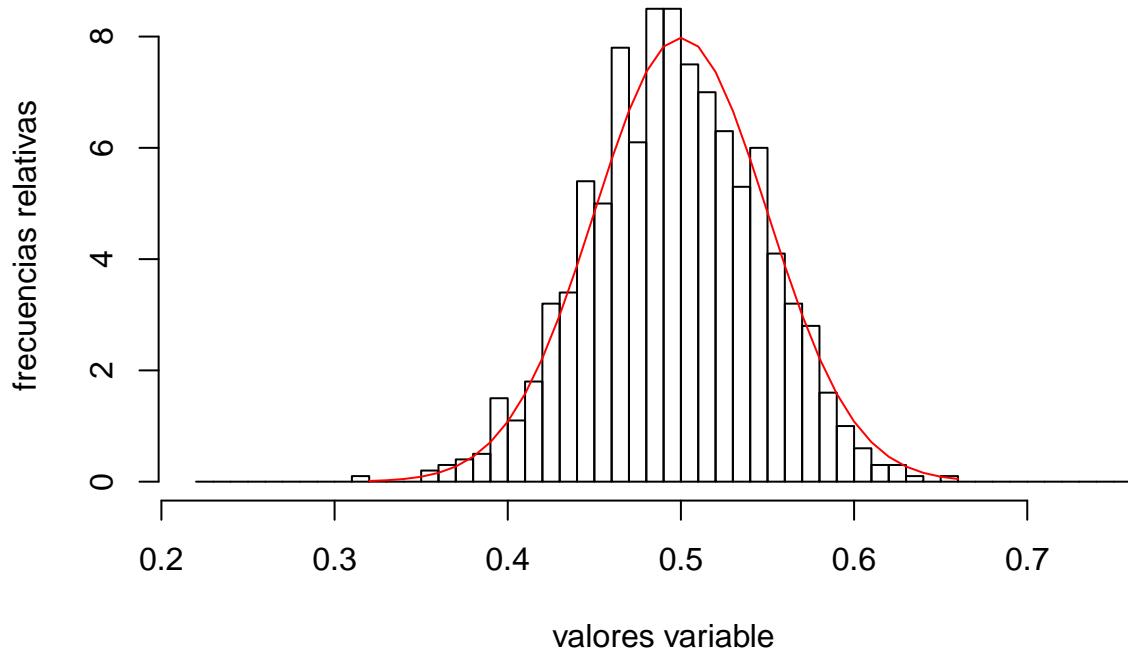
$$\bar{X}_n = \frac{1}{n} B\left(n = 100, p = \frac{1}{2}\right) \approx N\left(\mu = p = \frac{1}{2} = 0,5, \sigma = \sqrt{\frac{\frac{1}{2} \cdot \frac{1}{2}}{100}} = 0,05\right).$$

Para ver dicha aproximación, en primer lugar vamos a generar una muestra de $N = 1000$ valores de una binomial de parámetros $n = 100$ y $p = \frac{1}{2}$ y dividiendo por $n = 100$, tenemos una muestra de \bar{X}_n :

```
n=100
p=1/2
sigma=p*(1-p)
set.seed(2019)
muestra.binomial = rbinom(1000,n,p)
muestra.xnbarra = muestra.binomial/n
```

Para ver si la aproximación funciona, dibujaremos en una misma gráfica el histograma de frecuencias relativas de la muestra anterior y la curva de la función de densidad de la distribución normal de parámetros $\mu = \frac{1}{2}$ y $\sigma = 0,05$:

```
hist(muestra.xnbarra,freq=FALSE,
      breaks=seq(from=min(muestra.xnbarra)-0.1,
                 to=max(muestra.xnbarra)+0.1,by=0.01),
      main="Histograma de la distribución de las medias muestrales",
      xlab="valores variable",ylab="frecuencias relativas")
mu=p
sigma.xnbarra=sqrt(p*(1-p)/n)
x=seq(from=min(muestra.xnbarra),to=max(muestra.xnbarra),by=0.01)
lines(x,dnorm(x,mu,sigma.xnbarra),col='red')
```

Histograma de la distribución de las medias muestrales

Observamos que la aproximación es bastante buena.