Probabilidad y variables aleatorias para ML con R y Python

Ricardo Alberich, Juan Gabriel Gomila y Arnau Mir

2019-12-02

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

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

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

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

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

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

Pre requisitos: Teoría de conjuntos y combinatoria

Para aprender cálculo de probabilidades son necesarios conocimientos de:

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

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

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

0.1. Teoría de conjuntos

Definición de conjunto

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

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

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

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

Estas colecciones o conjuntos se pueden definir por:

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

0.1.1. Conjuntos básicos

Los conjuntos suelen tener un conjunto madre como por ejemplo

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

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

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

0.1.2. Características y propiedades básicas de los conjuntos

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

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

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

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

Cardinal de las partes de un conjunto

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

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

Efectivamente

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

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

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

dado un $\omega \in \Omega$

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

0.1.3. Operaciones con conjuntos

Intersección.

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

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

Más formalmente

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

Unión.

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

Más formalmente

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

Diferencia.

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

Más formalmente

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

Complementario

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

Más formalmente

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

0.1.4. 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 $A \cap B = \emptyset$.
- $\Omega^c = \emptyset$.
- $\emptyset^c = \Omega$.
- $A \cup B = B \cup A$, $A \cap B = B \cap A$ conmutativas.
- $(A \cup B) \cup C = A \cup (B \cup C)$, $(A \cap B) \cap C = A \cap (B \cap C)$ asociativas.
- $A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$, $A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$ distributivas.
- $(A^c)^c = A$ doble complementario.
- $(A \cup B)^c = A^c \cap B^c$, $(A \cap B)^c = A^c \cup B^c$ leyes de De Morgan.

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
```

Con python

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

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

```
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. Número binomial.

Número combinatorio o número binomial

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

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

Recordemos que

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

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

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

Puedo formar 21 equipos distintos.

Ejercicio

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

0.2.2. Variaciones.

Variaciones

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

La podemos escribir

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

Luego hay seis casos.

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

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

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

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

Ejercicio

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

0.2.3. Variaciones con repetición.

Variaciones con repetición

¿Y repitiendo algún dígito?

$$VR_k^n = n^k$$

Efectivamente en nuestro caso

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

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

0.2.4. Permutaciones

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

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

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 funcion `permn` para calcular todas las permutaciones anteriores.

Ejercicio

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

Ejercicio

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

0.3. Para acabar

Otros asuntos

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

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

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

Nota

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

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

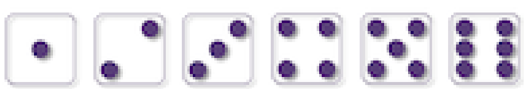
Ejemplo

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

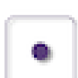
Los sucesos elementales del ejemplo anterior serían:

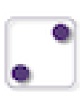


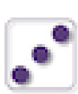
Espacio muestral: el conjunto Ω formado por todos los sucesos elementales del experimento aleatorio

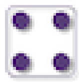
Ejemplo

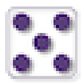
El espacio muestral del ejemplo anterior del dado es

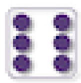
$\Omega = \Big\{$  ,

 ,

 ,

 ,

 ,

 $\}$

pero por comodidad, a partir de ahora pondremos

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

Suceso : Cualquier subconjunto del espacio muestral.

Alguno sucesos notables que merece la pena nombrar son:

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

Ejercicio

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

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:

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

Algunos sucesos serían:

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

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

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

Algunas operaciones entre los conjuntos:

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

1.1.2. Propiedades

Conmutativas:

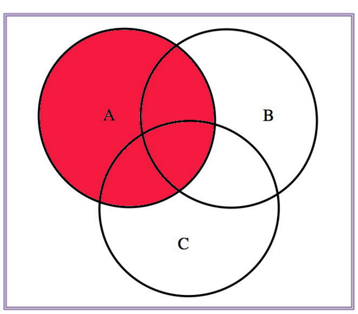
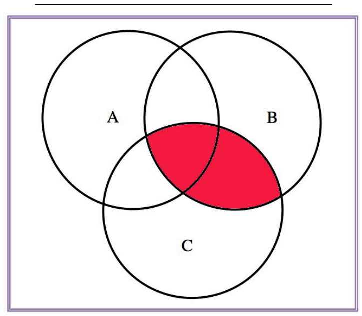
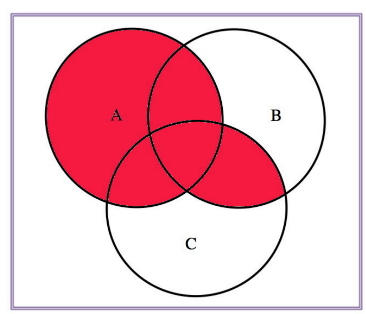
$$A \cup B = B \cup A, \quad A \cap B = B \cap A$$

Asociativas:

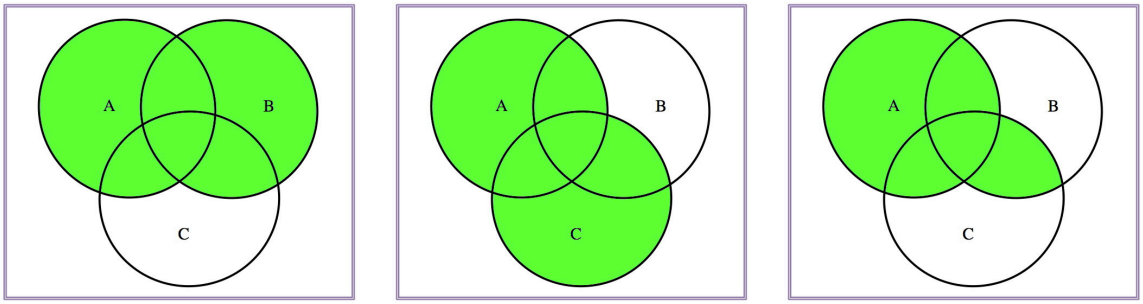
$$A \cup (B \cup C) = (A \cup B) \cup C, \quad A \cap (B \cap C) = (A \cap B) \cap C$$

Distributivas:

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

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

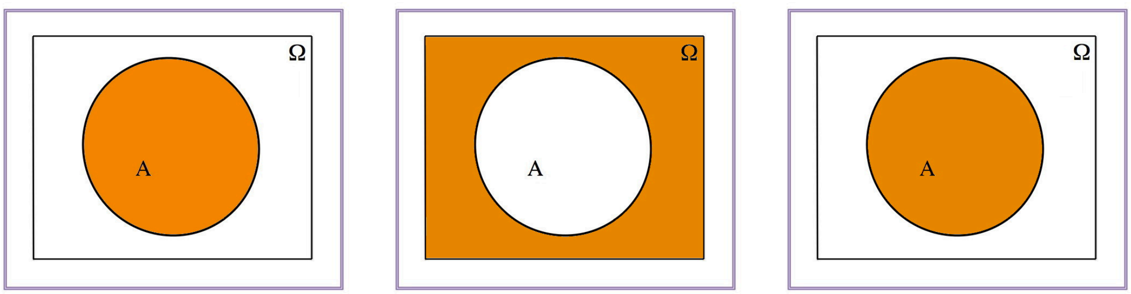
$A \cup B$	$A \cup C$	$(A \cup B) \cap (A \cup C)$



Complementario del complementario

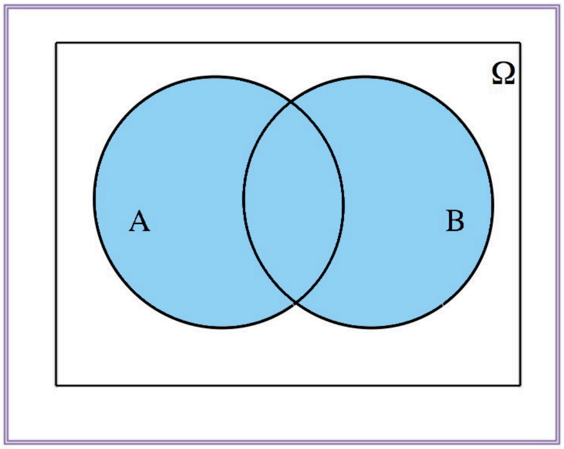
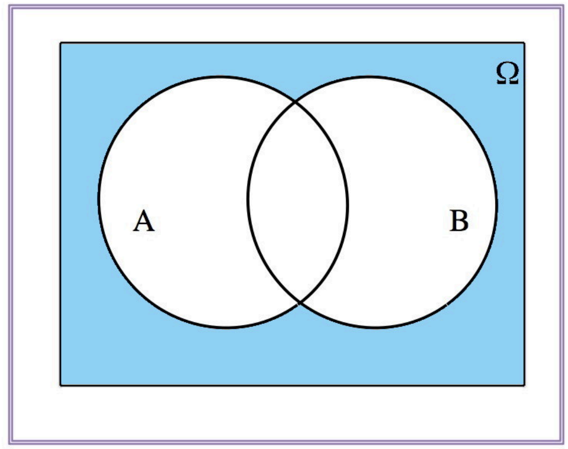
$$(A^c)^c = A$$

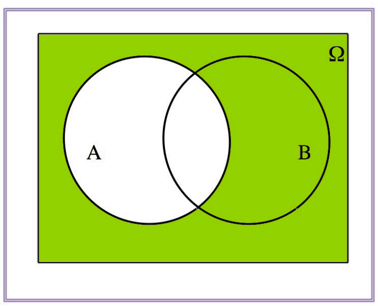
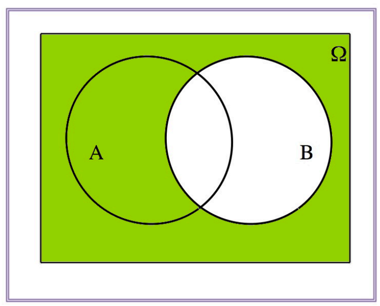
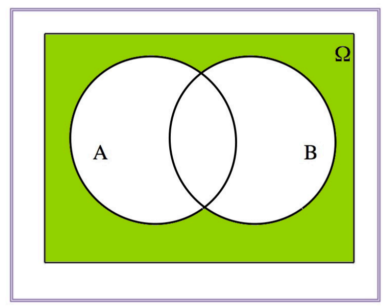
A	A^c	$(A^c)^c$



Leyes de De Morgan

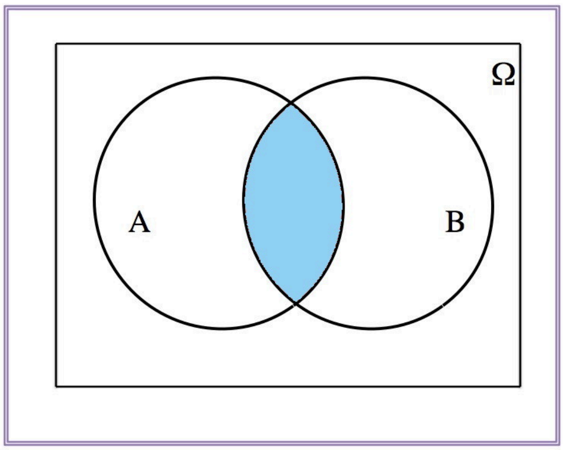
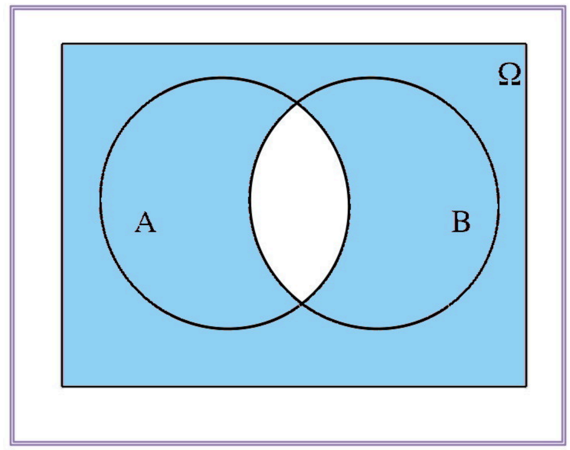
$$(A \cup B)^c = A^c \cap B^c$$

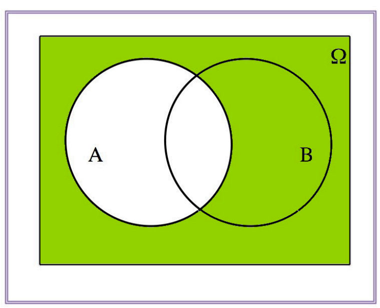
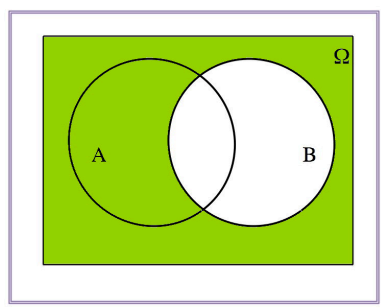
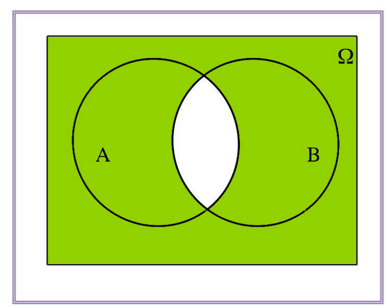
$A \cup B$	$(A \cup B)^c$
	

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*) numérico entre 0 y 1 que mide la verosimilitud de que este evento se produzca.

Esta verosimilitud puede estar justificada por:

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

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

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

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

3. Si $\{A_1, A_2, \ldots, A_n\}$ son sucesos disjuntos dos a dos, entonces

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

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

Ejemplo: grupos sangíneos

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

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

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

Experimento aleatorio: tipo de sangre de un paciente humano

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

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

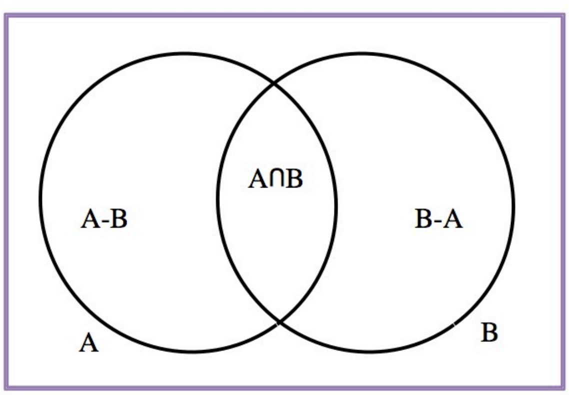
Suceso: $\{\text{O}\}^c = \{\text{A,B,AB}\}$

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

1.1.4. Propiedades

Propiedades básicas de la probabilidad

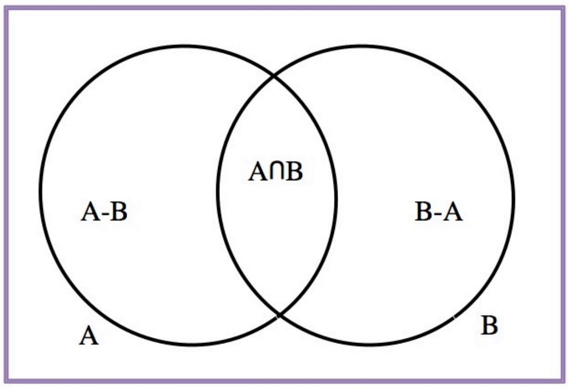
- $P(\emptyset) = 0$.
- $P(A - B) = P(A) - P(A \cap B)$ porque $P(A) = P(A - B) + P(A \cap B)$.



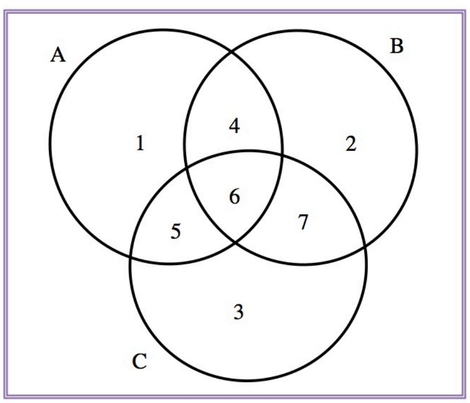
- Si $B \subseteq A$, entonces $0 \leq P(B) \leq P(A)$.
- $P(A^c) = 1 - P(A)$.

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

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



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



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

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

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

- Si todos los sucesos elementales tienen la misma probabilidad,

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

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 $\{E, 0\}$.

Y su probabilidad es

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

Ejemplo: Consumo de drogas

Segun un árticulo de El País, en un control especial de la policía el 0,1 % de todos los conductores analizados en un control de tráfico dan positivo en un el test en cocaína, y el 1 % da positivo en cannabis. Un 1,05 % da positivo en alguno de los dos test.

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

Los sucesos elementales del enunciado del problema son:

- A: dar positivo en cocaína; $P(A) = 0{,}001$.
- B: dar positivo en cannabis; $P(B) = 0{,}01$.

En este caso nos interesa estudiar los sucesos:

- $A \cup B$: dar positivo en alguno de los dos test; $P(A \cup B) = 0{,}0105$.
- $(A \cup B)^c$: no dar positivo en ninguno de los test,

de donde, por tanto:

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

Ejemplo

En un control especial de la policía el 0,1 % de todos los conductores analizados en un control de tráfico dan positivo en un el test en cocaína, y el 1 % da positivo en cannabis. Un 1,05 % da positivo en alguno de los dos test.

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

Los sucesos elementales son:

- A: dar positivo en cocaína; $P(A) = 0{,}001$.
- B: dar positivo en cannabis; $P(B) = 0{,}01$.

En este caso nos interesa estudiar los sucesos:

- $A \cup B$: dar positivo en algún de los dos test; $P(A \cup B) = 0{,}0105$.

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

de donde, por tanto:

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

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

Los sucesos elementales son:

- A: dar positivo en cocaína; $P(A) = 0{,}001$.
- B: dar positivo en cannabis; $P(B) = 0{,}01$.

En este caso nos interesa estudiar los sucesos:

- $A \cap B$: dar positivo en los dos test; $P(A \cap B) = 0{,}0005$.
- $B - A$: dar positivo en cocaína pero no en cannabis,

de donde, por tanto:

$$P(B - A) = P(B) - P(A \cap B) = 0{,}01 - 0{,}0005 = 0{,}0095.$$

1.2. Probabilidad condicionada

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 sea mujer y lleve gafas.*
- $P(A|B)$: probabilidad de que si pasa B, entonces pase A, *Probabilidad de que, si es mujer, lleve gafas.*

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

1.2.2. Propiedades

La probabilidad condicionada es una probabilidad

Proposición

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

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

satisface las propiedades de las probabilidades, como por ejemplo:

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

Ejercicio

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

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

Si tenemos los sucesos:

- A: ser hipertenso,
- B: creer ser hipertenso,

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

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

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 de que un correo no lleve 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 $A_1, A_2, \ldots, A_n$ son una **partición** del espacio muestral Ω de un determinado experimento aleatorio, si cumplen las condiciones siguientes:

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

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

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

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

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

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

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

- Si el resultado de una exploración es P y el valor dado es también P, entonces se conoce como un Verdadero Positivo (VP).
- Sin embargo si el valor real es N entonces se conoce como un Falso Positivo (FP).

- De igual modo, tenemos un Verdadero Negativo (VN) cuando tanto la exploración como el valor dado son N.
- Un Falso Negativo (FN) cuando el resultado de la predicción es N pero el valor real es P.

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

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

Ejercicio

Demostrar el teorema de Bayes utilizando que

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

Teorema de Bayes generalizado

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

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

Ejercicio

Demostrar el teorema de Bayes utilizando que

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

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

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

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

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

Buscamos estudiar el suceso E: el cliente compra, se tiene que:

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

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

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

Definimos los sucesos y datos del ejercicio:

- T: Dar positivo al test,
- B: darse de baja; $P(B) = 0{,}02$,
- $P(T|B) = 0{,}975, P(T|B^c) = 0{,}12$.

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

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

$$\begin{aligned} P(T) &= P(B) \cdot P(T|B) + P(B^c) \cdot P(T|B^c) \\ &= 0{,}02 \cdot 0{,}975 + 0{,}98 \cdot 0{,}12 = 0{,}1371. \end{aligned}$$

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

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

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

$$\begin{aligned} P(B|T^c) &= \frac{P(B \cap T^c)}{P(T^c)} = \frac{P(B) - P(B \cap T)}{1 - P(T)} \\ &= \frac{0{,}02 - 0{,}0195}{1 - 0{,}1371} \approx 0{,}00058. \end{aligned}$$

O también se obtiene así

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

donde

$$\begin{aligned} P(T^c|B) &= 1 - P(T|B) = 0{,}025, \\ P(T^c|B^c) &= 1 - P(T|B^c) = 0{,}88. \end{aligned}$$

1.4. Independencia de sucesos

1.4.1. Sucesos independientes

Sucesos Independientes

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

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

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

Proposición:

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

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

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 si 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; esta funciones son las variables aleatorias.

2.1.1. Definición de variable aleatoria

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

Variable Aleatoria (definición práctica)

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

Notación:

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

Ejemplo: Dado seis caras

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

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

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

$$X(1) = 1, X(2) = 2, X(3) = 3,$$
$$X(4) = 4, X(5) = 5, X(6) = 6.$$

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

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:

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

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

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

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

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

Supongamos que el dado está bien balanceado. Entonces

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

Concretamente:

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

Su dominio es

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

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

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

Ejemplo: Lanzamiento moneda

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

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

(expresados en orden de aparición).

Este espacio tiene todos los sucesos elementales equiprobables.

Consideremos la variable aleatoria asociada a este experimento:

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

Su función de probabilidad es:

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

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 $F_X(x_0^-) = \lim_{x \to x_0^-} F(x)$, entonces se cumple que $P(X < x_0) = F_X(x_0^-)$ y que $P(X = x_0) = F_X(x_0) - F_X(x_0^-)$.
- Se cumple que $\lim_{x \to \infty} F_X(x) = 1$; $\lim_{x \to -\infty} F_X(x) = 0$.
- Toda función F verificando las propiedades anteriores es función de distribución de alguna v.a. X.
- $P(X > x) = 1 - F_X(x)$.
- Dados $a, b \in \mathbb{R}$ con $a < b$,
$$P(a < X \leq b) = F_X(b) - F_X(a).$$

Advertencia desigualdades estrictas

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

Veamos que propiedades tenemos cuando se cambian estas desigualdades.

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

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

entonces se cumplen las siguientes igualdades:

- $P(X = x) = F_X(x) - F_X(x^-)$.
- $P(a < X < b) = F_X(b^-) - F_X(a)$.
- $P(a \leq X < b) = F_X(b^-) - F_X(a^-)$.
- $P(X < a) = F_X(a^-)$.
- $P(a \leq X \leq b) = F_X(b) - F_X(a^-)$.
- $P(X \geq a) = 1 - F_X(a^-)$.
- 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 $\sum\limits_{x \leq x_0}$ 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.
- $\lim\limits_{x\to-\infty} F_X(x) = 0$ y $\lim\limits_{x\to+\infty} F_X(x) = 1$.
- Es continua por la derecha: $\lim\limits_{x\to x_0^+} F_X(x) = F_X(x_0)$.

2.2.2. 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 correspondiente medidas para variables aleatorias.

A estas medidas se les suele añadir el adjetivo **poblacionales** mientras que a las que provienen de la muestra se las adjetiva como **muestrales**.

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

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 $i = 1, \ldots, 6$. Sea X la v.a. que nos representa el valor de una tirada del dado.

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

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

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

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

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

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

entonces

$$E(X) = 0 \cdot 0{,}42 + 1 \cdot 0{,}4 + 2 \cdot 0{,}18 = 0{,}76.$$

Elegida una página del texto al azar esperamos encontrar 0,76 errores por página.

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

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

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

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

Series geométricas

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

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

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

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

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

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

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

Momentos de orden m

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

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

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

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:

$$P(X = 0) = 0{,}42,\ P(X = 1) = 0{,}4,\ P(X = 2) = 0{,}18,$$

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 más en una cantidad finita o numerable de puntos).

En lo que sigue supondremos que la función de distribución de variables aleatorias continuas cumplen estas propiedades.

Notemos que si X es una v.a. con función de distribución continua se tiene que $P(X=x_0)=F_X(x_0)-F(x_0^-)=0$. Por lo que no tiene sentido definir *función de probabilidad*.

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*).

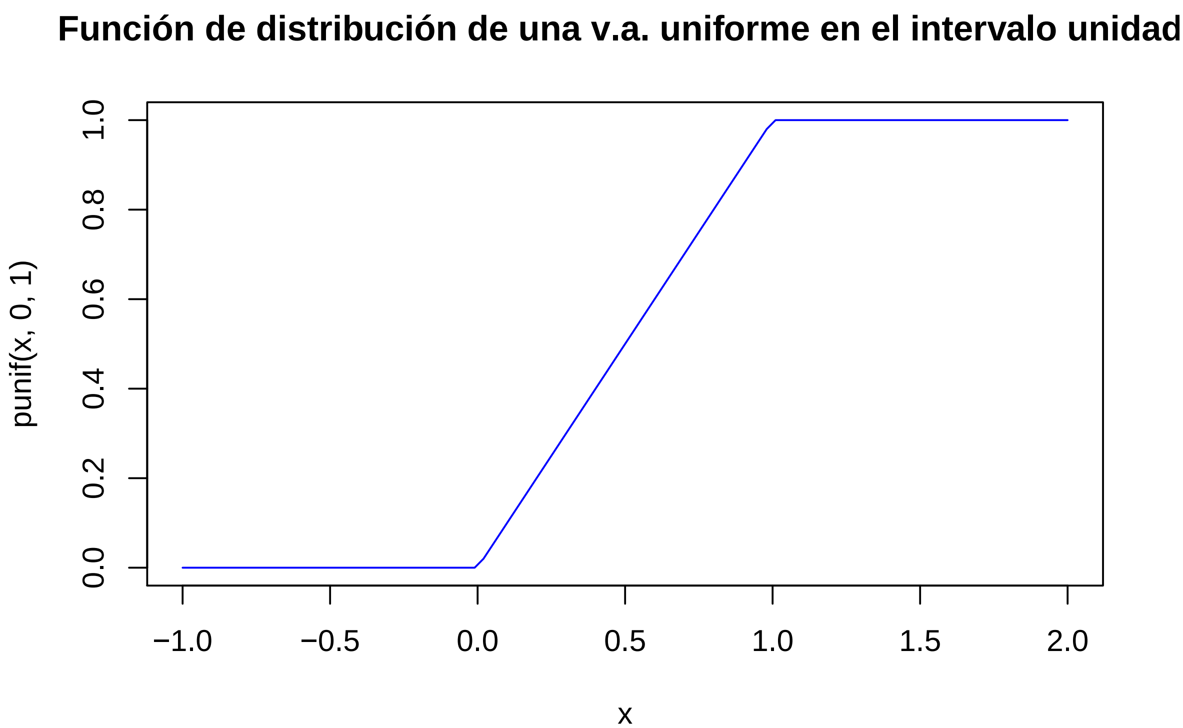
Consideremos la v.a. continua $X=$ distancia del dardo al centro de la diana.

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

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



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

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

Función de distribución de una variable aleatoria

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

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

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

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

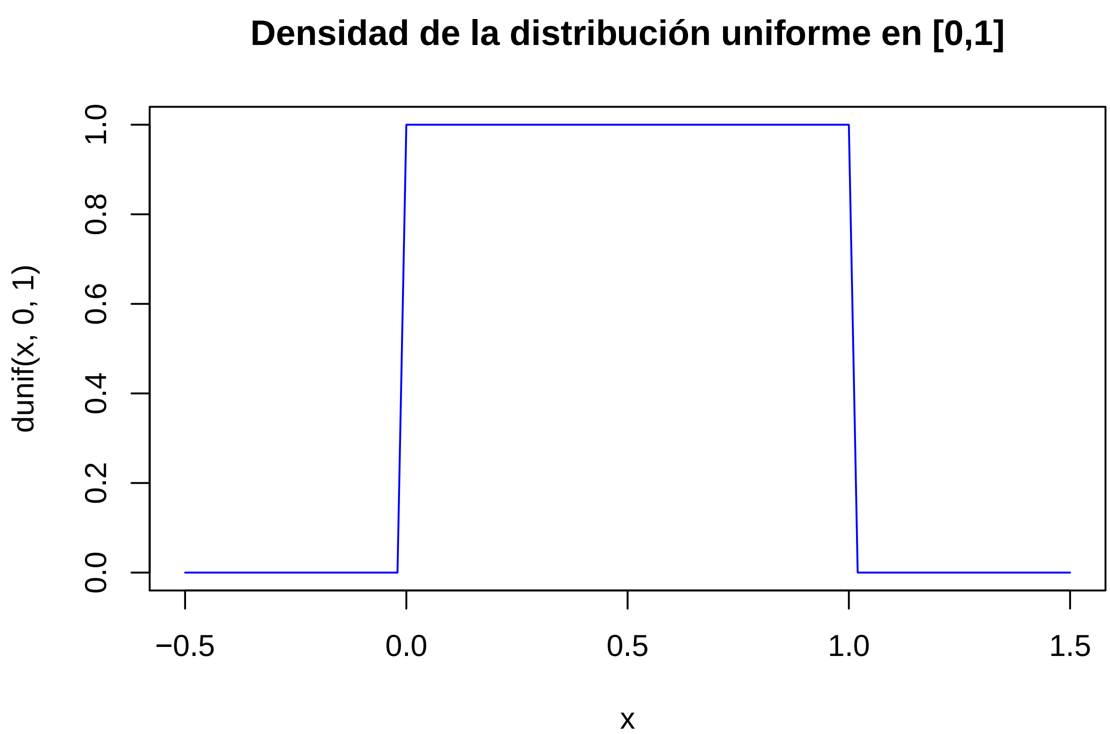
$$f_X(x) = \begin{cases} 0, & \text{si } x \leq 0, \\ 1, & \text{si } 0 < x < 1, \\ 0, & \text{si } 1 \leq x, \end{cases}$$

que es la densidad de X. En efecto:

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

Por lo tanto, $F_X(x) = \displaystyle\int_{-\infty}^{x} f_X(t)dt$ 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]")
```



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.)
- $\int\limits_{-\infty}^{+\infty} f_X(x)dx = \int\limits_{0}^{2} \frac{1}{2}dx = \left[\frac{x}{2}\right]_0^2 = \frac{2}{2} - \frac{0}{2} = 1.$

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

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

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.

Tranformaciones de v.a. discretas

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

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

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

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

Densidad de una transformación de una v.a. continua

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

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

una aplicación, no necesariamente monótona tal que :

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

entonces:

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

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 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.

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

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

Apliquemos la consecuencia de la desigualdad de Markov a la v.a. no negativa $Y^2 = (X - \mu)^2$. Entonces

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

Por otra parte

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

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

Utilidad básica de la desigualdad de Chebychev

Supongamos que X es una v.a. con $Var(X) = 0$. Entonces, aplicando la desigualdad anterior,

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

lo que implica que

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

por lo que probabilidad de que X sea constantemente $E(X)$ es 1, 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 desgualdad de Chebychev

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

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

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

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

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.*!

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 numero 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

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

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

Su función de probabilidad es:

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

Su función de distribución es:

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

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

Lo denotaremos por $X \equiv Ber(p)$ o también $X \equiv B(1, p)$.

A este tipo de experimentos (éxito/fracaso) se les denomina experimentos Bernoulli.

Fue su descubridor un científico suizo Jacob Bernoulli, uno más de la de la conocida familia de científicos suizos Bernoulli.

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)
```

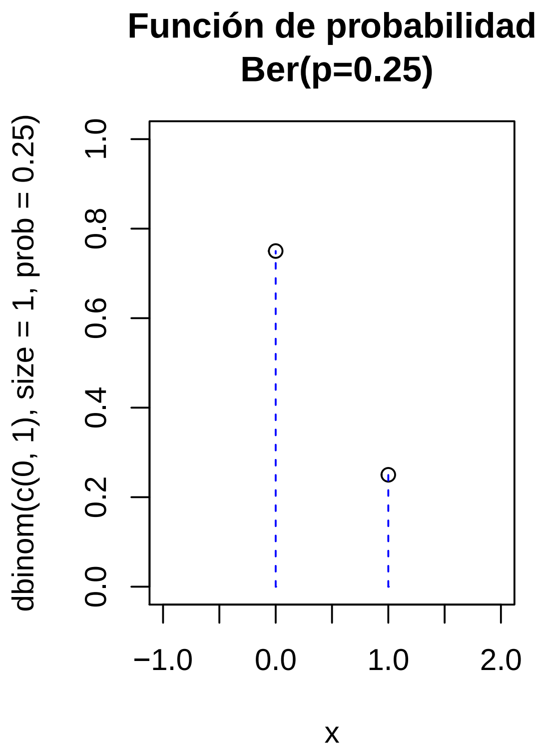
```
## [1] 0.25
```

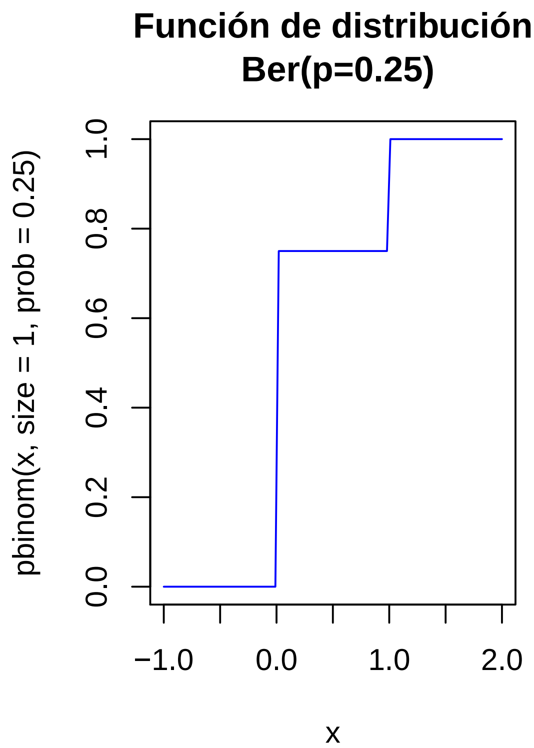
```
rbinom(n=20,size = 1,prob=0.25)
```

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

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))
```





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=pbinom(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.:

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

A la variable aleatoria anterior se le 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:

$$P_X(x) = \left\{ \begin{array}{ll} \binom{n}{x} \cdot p^x \cdot (1-p)^{n-x}, & \text{si } x = 0,1,\ldots,n, \\ 0, & \text{en otro caso.} \end{array} \right.$$

Función de distribución de una binomial

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

$$\begin{array}{lll} F_X(x) = P(X \leq x) & = & \displaystyle\sum_{i=0}^{x} P_X(i) \\ & = & \left\{ \begin{array}{ll} 0, & \text{si } x \leq 0, \\ \displaystyle\sum_{i=0}^{k} \binom{n}{i} \cdot p^i \cdot (1-p)^{n-i} & \text{si } \left\{ \begin{array}{l} k \leq x < k+1, \\ k = 0,1,\ldots,n, \end{array} \right. \\ 1, & \text{si } n \leq x. \end{array} \right. \end{array}$$

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 comprobar 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:

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

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 las suma de n v.a. $Ber(p)$ independientes.

Ejercicio

Justificar de forma intuitiva que si X_i con $i = 1, 2, \ldots, n$ 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	$B(n,p)$
$D_X =$	$\{0, 1, \ldots n\}$
$P_X(x) = P(X = x) =$	$\begin{cases} \binom{n}{x} \cdot p^x \cdot (1-p)^{n-x} & \text{si } x = 0, 1, \ldots, n \\ 0 & \text{en otro caso.} \end{cases}$
$F_X(x) = P(X \leq X) =$	no tiene fórmula (utilizad funciones de `R` o python)
$E(X) =$	$n \cdot p$
$Var(X) =$	$n \cdot p \cdot (1-p)$

Cálculos 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:

```r
pbinom(0,size=10,prob=0.25)
```

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

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

```r
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:

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

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

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

```r
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
##  [24]  8  8  5 11 14 11 10 11  5 12  8  6  7  9 10  5 12 11  9 12 11 12 10
##  [47] 13 13  8  8  9  7  6  9 10  9 16 13  6  6  8  8 11  9 12 15  9  7 12
##  [70] 11  9  8  9  8 11 15  7 10  9 12  6 13 14  8 10  8 10 11 11  9 10 11
##  [93] 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 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.stat`:

```
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)
```

```
## 0.9218730926513672
```

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([ 6,  2,  4,  6,  5,  3,  3,  5,  6,  4,  2,  5,  3,  6,  5,  3,  7,
##         1,  8,  7,  7,  7,  4,  5,  6,  2,  7,  5, 10,  4,  5,  4,  3,  5,
##         0,  4,  6,  4,  7,  5,  3,  3,  3,  6,  2,  5,  5,  3,  5,  5,  4,
##         7,  9,  4,  6,  5,  3,  7,  7,  3,  4,  5,  3,  5,  6,  5,  5,  5,
##         5,  5,  4,  6,  5,  7,  9,  5,  9,  5,  3,  4,  6,  7,  7,  7,  7,
##         4,  2,  4,  6,  2,  8,  5,  7,  1,  9,  3,  4,  5,  3,  4])
```

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([ 3,  6,  6,  4,  4,  1,  3,  4,  5,  3,  7,  4,  5,  4,  4,  7,  6,
##         2,  3,  4,  4,  7,  5,  7,  4,  7,  7,  4,  6,  6,  4,  2,  5,  2,
##         4,  4,  4,  6,  0,  4,  6,  3,  4,  8,  3,  4,  7,  4, 12,  4,  4,
##         6,  4,  3,  9,  5,  7,  3,  5,  7,  4,  4,  4,  6,  6,  5,  7,  7,
##         4,  4,  8,  4,  7,  4,  5,  6,  7,  6,  5,  5,  2,  5,  5,  2,  8,
##         8,  5,  9,  6,  7,  6,  7,  1,  5,  3,  8,  7,  5,  5,  5])
```

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)
```

```
## 0.9802722930908203
```

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

```
## 0.18771171569824247
```

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

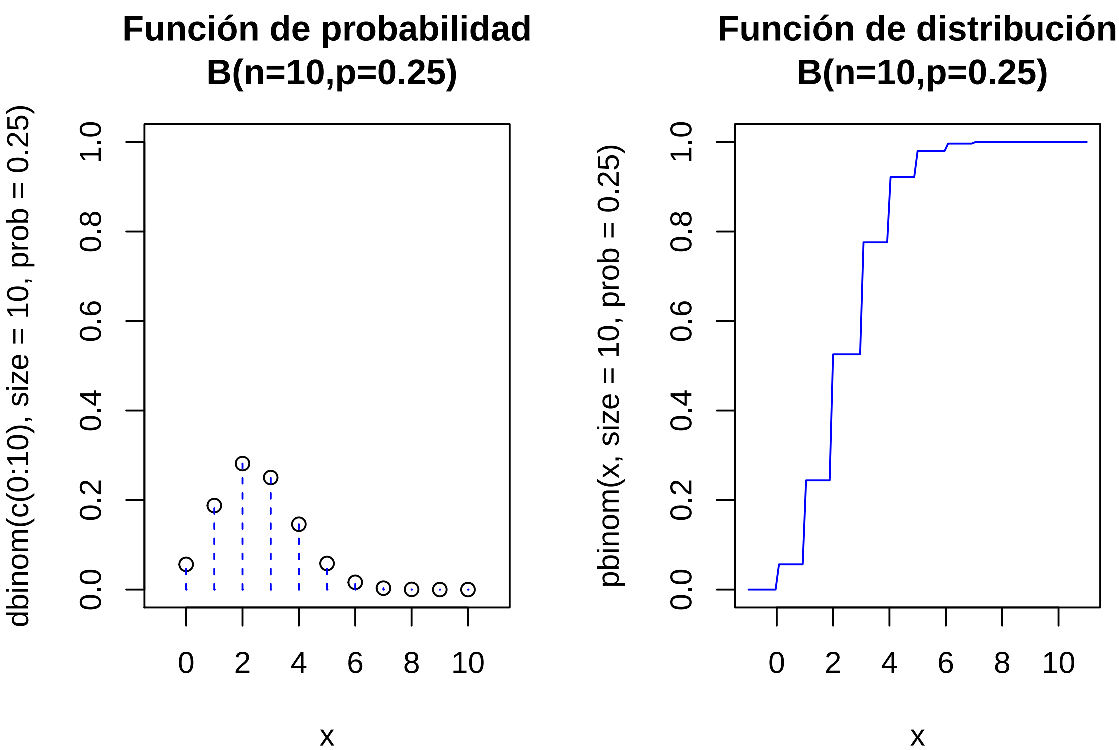
```
## array([4, 3, 3, 6, 6, 5, 8, 4, 5, 4])
```

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))
```



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

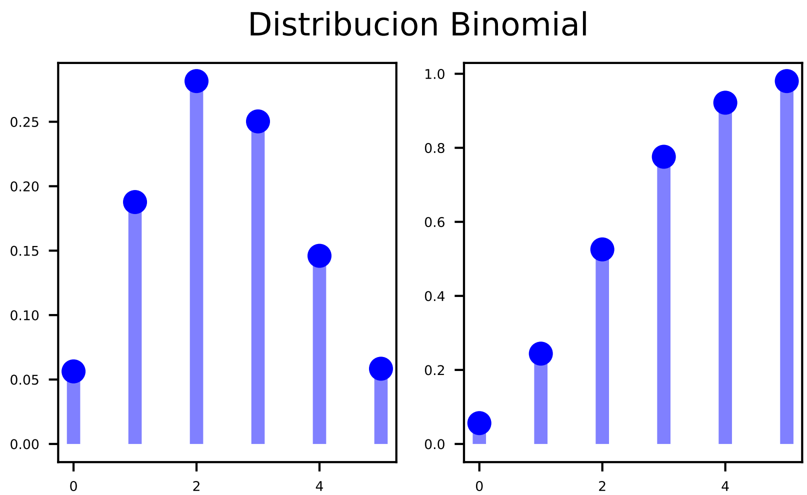
Ejercicio

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 Bernouilli con probabilidad de éxito (sacar bola roja)

$$P(Roja) = P(xito) = p = \frac{40}{100} = 0{,}4.$$

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`:

```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 $P(X\geq 4)=1-P(X<4)=1-P(X\leq 3):$

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

Con R:

```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:

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

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

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

```r
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) = {m}".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) = {v}".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

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

Su función de probabilidad es:

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

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

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 $k = 0,1,2,\ldots$, tenemos:

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

En general, tendremos que:

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

De forma más compacta, tendremos que

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

Notemos que el límite de la función de distribución es:

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

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 $x = 0,1,2,\ldots$ y aplicado 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$**

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

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 $D_Y = \{1, 2, \ldots\}$

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
$D_Y =$	$\{1, 2, \ldots n, \ldots\}$
$P_Y(y) = P(Y = y) =$	$\begin{cases} (1-p)^{y-1} \cdot p, & \text{si } y = 1, 2, 3, \ldots \\ 0, & \text{en otro caso.} \end{cases}$
$F_Y(y) = P(Y \leq y) =$	$\begin{cases} 0, & \text{si } y < 1 \\ 1 - (1-p)^k, & \text{si } \begin{cases} k \leq y < k+1, \\ \text{para } k = 1, 2, 3, \ldots \end{cases} \end{cases}$
$E(X) = \frac{1}{p}$	$Var(X) = \frac{1-p}{p^2}$

Propiedad de la falta de memoria

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

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

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

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

Demostración

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

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

y el lado de izquierdo es:

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

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:

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

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

Combinado las igualdades, tenemos que:

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

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 $P(X = 0) = p$:

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

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

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

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 historicos 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
```

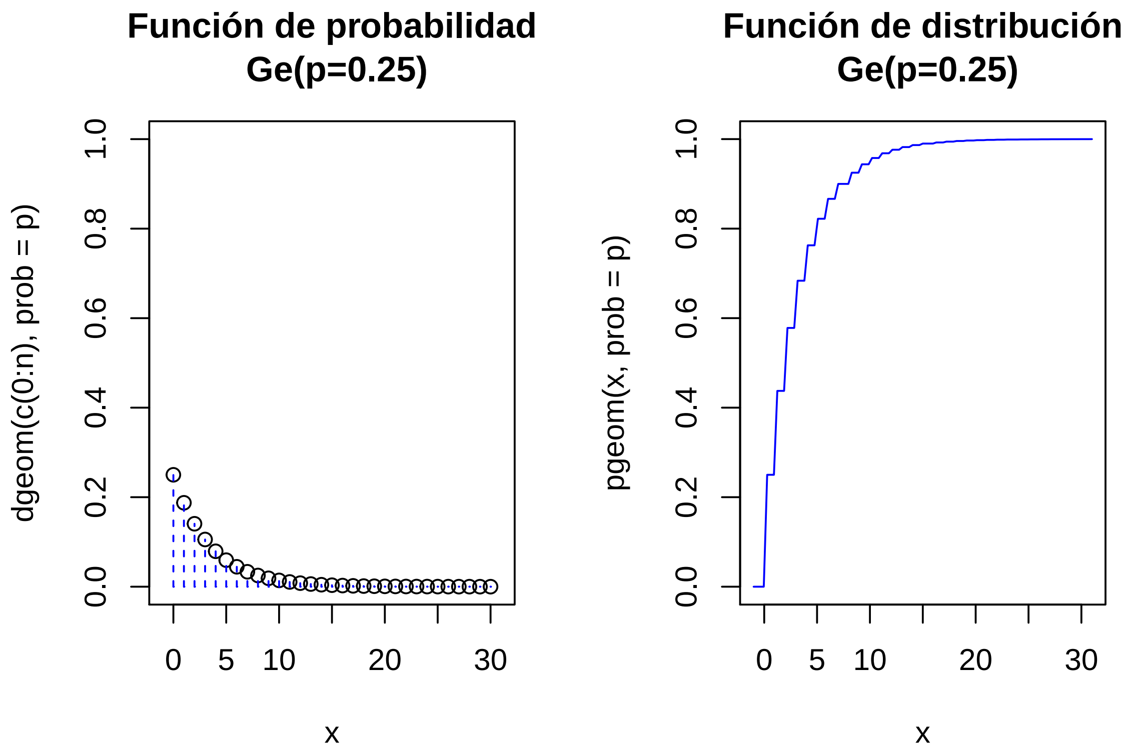
Una muestra aleatoria de tamaño 25 de una $Ge(0{,}25)$:

```
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
## [24]  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 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 un geométrica que cuenta el número de intentos para obtener el primer éxito 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)
```

```
## 0.24999999999999997
```

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

```
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([7, 4, 1, 4, 0, 2, 1, 8, 3, 0, 5, 5, 5, 3, 2, 4, 0, 2, 0, 1])
```

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.), array(12.))
```

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

```
## (array(3.), array(12.))
```

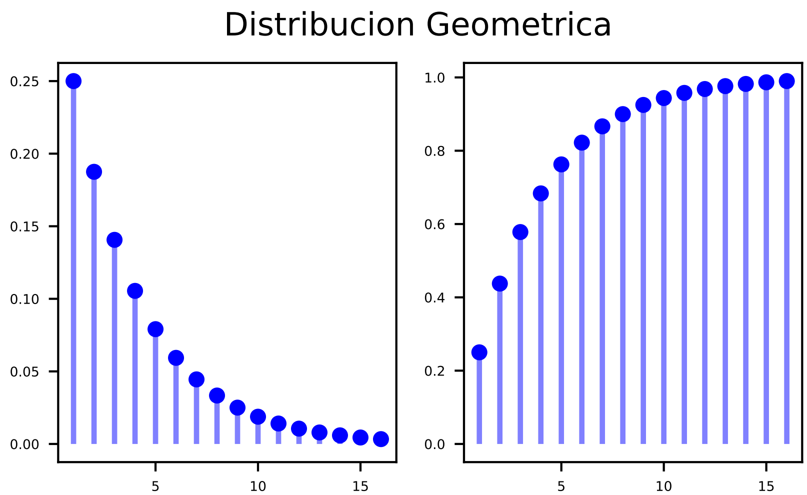
Ejercicio

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

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

Calculemos algunas probabilidades para $n = 2$:

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

En general su función de probabilidad es

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

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 $k = 0, 1, 2, \ldots$

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

$$\overbrace{\overbrace{\underbrace{EFFF \ldots EEF}_{k \quad \text{Fracasos}}}^{n-1 \quad \text{Éxitos.}}}^{k+n-1 \text{ posiciones}} E$$

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:

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

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
$D_X =$	$\{0,1,2,3\ldots\}$
$P_X(k) = P(X=k) =$	$\begin{cases} \binom{k+n-1}{n-1} \cdot (1-p)^k \cdot p^n, & \text{si } k = 0,1,\ldots \\ 0, & \text{en otro caso.} \end{cases}$
$F_X(x) = P(X \leq x) =$	$\begin{cases} 0, & \text{si } x < 0 \\ \displaystyle\sum_{i=0}^{k} P(X=i) & \text{si } \begin{cases} k \leq x < k+1, \\ k = 0,1,2,\ldots \end{cases} \end{cases}$ 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 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 del 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 estados 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 del X?

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

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

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

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

Simplificando,

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

La función de distribución en general es

$$F_X(x) = P(X \leq x) = \left\{ \begin{array}{ll} 0, & \text{si } x < 0, \\ \displaystyle\sum_{i=0}^{k} \binom{i+n-1}{n-1} \cdot (1-p)^{i+n-1} \cdot p^n, & \text{si } \left\{ \begin{array}{l} k \leq x < k+1, \\ k = 0, 1, 2, \ldots \end{array} \right. \end{array} \right.$$

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

$$F_X(x) = P(X \leq x) = \left\{ \begin{array}{ll} 0, & \text{si } x < 0, \\ \displaystyle\sum_{i=0}^{k} 0{,}01 \cdot (i+1) \cdot 0{,}9^{i+1}, & \text{si } \left\{ \begin{array}{l} k \leq x < k+1, \\ k = 0, 1, 2, \ldots \end{array} \right. \end{array} \right.$$

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 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
```

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

```
## [1] 0.885735
```

```r
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 `scpi.stats`

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

Recordemos que de nuevo se cumple que

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

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

```
## 0.03542940000000002
```

```python
nbinom.pmf(k=5,n=2,p=0.1,loc=0)
```

```
## 0.03542940000000002
```

```python
nbinom.cdf(k=4,n=2,p=0.1)
```

```
## 0.11426500000000003
```

```python
1-nbinom.cdf(k=4,n=2,p=0.1)
```

```
## 0.8857349999999999
```

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.

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

```
## array([ 3, 11, 13, 12, 26,  4, 39,  3, 55, 31,  9, 32, 23, 16, 35,  5, 17,
##         6, 30,  1, 15, 18,  8, 13,  2,  2, 16, 16, 26,  3,  1, 15,  2, 14,
##        19, 14,  7, 11, 24, 11, 13, 23, 17, 32, 10, 24, 23,  4, 21, 25, 24,
##        14, 10,  6, 26, 22,  1, 24, 23,  3,  2, 22, 36, 31, 11, 32, 11, 18,
##        21,  0, 16,  8, 18, 18, 11,  5, 20, 29, 16,  7, 21,  9, 53, 62, 55,
##        29, 35,  2,  5, 21, 19,  2, 39,  6,  5, 25, 20,  4, 22, 15])
```

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

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

```
## E(X)=18.0
```

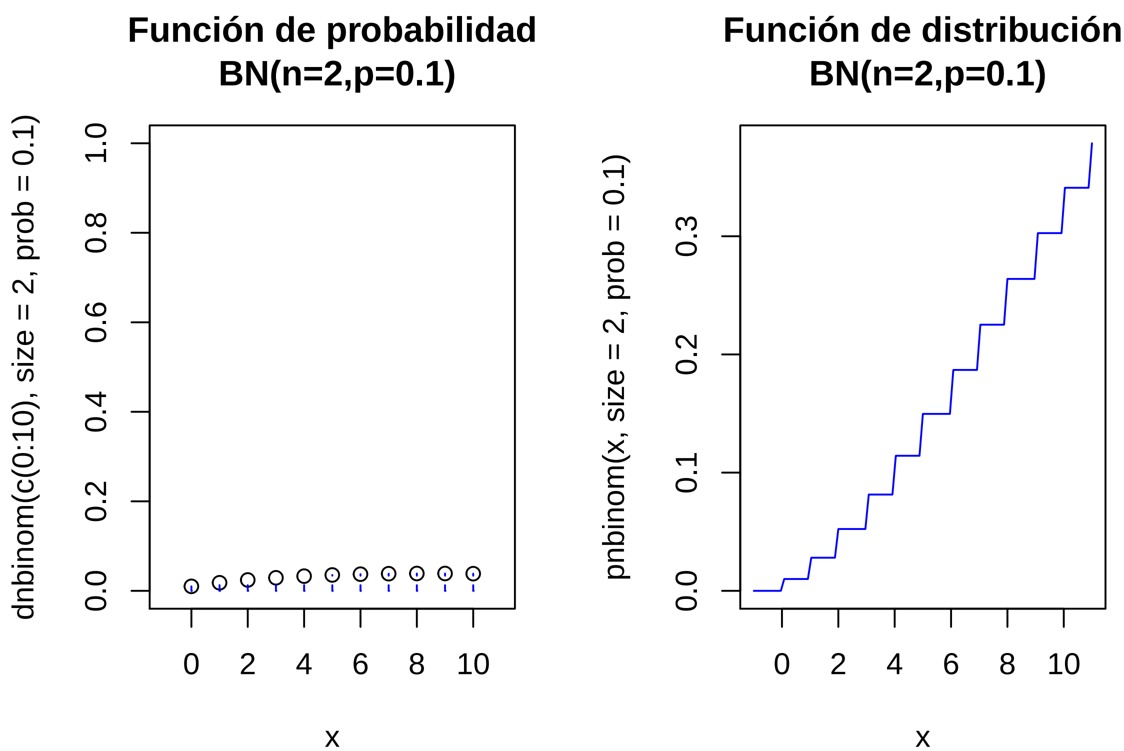
```python
print("Var(X)={v}".format(v=params[1]))
```

```
## 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

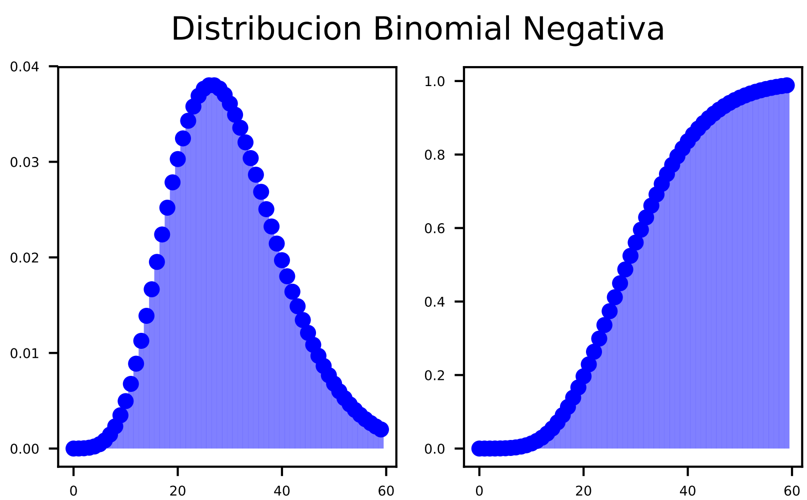
Buscad en los manuales de python cómo se dibuja la función de probabilidad y de distribución de una binomial. negativa

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 tiene un programa de seguridad que genera accesos con claves de 3 dígitos $000, 001, \ldots 999$. 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 estados condiciones tenemos que la probabilidad de "éxito" de cada intento es $p = \frac{1}{1000} = 0{,}001$. Y como cada vez *olvidamos* en 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:

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

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

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

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:

```
dnbinom(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
```

```
## 9.876743459670526e-06
```

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

```
## 9.876743459670217e-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 $P(X \leq 150)$

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

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)
```

```
## 0.0005232003490824058
```

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

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

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

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

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

Ejercicio

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 entra 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 $X(\Omega) = \mathbf{N}$ tiene distribución de Poisson con parámetro $\lambda > 0$, y lo denotaremos por $Po(\lambda)$ si su función de probabilidad es:

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

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

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

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

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

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 $p_n = \frac{\lambda}{n}\cdot$

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

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

Si hacemos tender n hacia ∞, obtenemos:

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

Calculemos el límite de algunos de los factores de la expresión

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

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:

$$\lim_{n\to\infty} P(X_n = k) = \lim_{n\to\infty} \left(\begin{array}{c} n \\ k \end{array} \right) \cdot p_n^k \cdot (1-p_n)^{n-k} = \frac{\lambda^k}{k!} \cdot 1 \cdot e^{-\lambda} \cdot 1 = \frac{\lambda^k}{k!} \cdot e^{-\lambda}.$$

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
$D_X =$	$\{0, 1, \ldots\}$
$P_X(x) = P(X = x) =$	$\begin{cases} \frac{\lambda^x}{x!} e^{-\lambda}, & \text{si } x = 0, 1, \ldots \\ 0, & \text{en otro caso.} \end{cases}$
$F_X(x) = P(X \leq X) =$	$\begin{cases} 0, & \text{si } x < 0, \\ \displaystyle\sum_{i=0}^{k} P(X = i) = \sum_{i=0}^{k} \frac{\lambda^i}{i!} \cdot e^{-\lambda}, & \text{si } \begin{cases} k \leq x < k+1, \\ k = 0, 1, 2, \ldots \end{cases} \end{cases}$ 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.
$D_X =$	$\{0, 1, \ldots\}$
$P_X(x) = P(X = x) =$	$\begin{cases} \frac{(\lambda \cdot t)^x}{x!} e^{-\lambda \cdot t} & \text{si } x = 0, 1, \ldots \\ 0 & \text{en otro caso.} \end{cases}$
$F_X(x) = P(X \leq X) =$	$\begin{cases} 0, & \text{si } x < 0, \\ \displaystyle\sum_{i=0}^{k} P(X = i) = \sum_{i=0}^{k} \frac{(\lambda \cdot t)^i}{i!} \cdot e^{-\lambda \cdot t}, & \text{si } \begin{cases} k \leq x < k+1, \\ k = 0, 1, 2, \ldots \end{cases} \end{cases}$ 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 és 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",ylab="Función de probabilidad",
       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),bty = "n")
  curve(pbinom(x,size=n,p=pr),
        xlim=c(-1,n+1),col="blue",ylab="Función de Distribución",
         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_l="Condición\n TRUE"} else
    {aux_l="Condición\n FALSE"}
  legend("topleft",legend=c(aux_l,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 explicita 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 $p_n = \frac{\lambda}{n}\cdot$ 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 explicita la función de probabilidad y de distribución de X.

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

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

En nuestro caso, $\lambda = 20$:

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

La función de distribución es:

$$F_X(x) = P(X \leq X) = \left\{ \begin{array}{ll} 0, & \text{si } x < 0, \\ \displaystyle\sum_{i=0}^{k} P(X = i) = \sum_{i=0}^{k} \frac{\lambda^i}{i!} \cdot e^{-\lambda}, & \text{si } \left\{ \begin{array}{l} k \leq x < k+1, \\ k = 0, 1, 2, \ldots \end{array} \right. \end{array} \right.$$

En nuestro caso:

$$F_X(x) = P(X \leq X) = \left\{ \begin{array}{ll} 0, & \text{si } x < 0, \\ \displaystyle\sum_{i=0}^{k} P(X = i) = \sum_{i=0}^{k} \frac{20^i}{i!} \cdot e^{-20}, & \text{si } \left\{ \begin{array}{l} k \leq x < k+1, \\ k = 0, 1, 2, \ldots \end{array} \right. \end{array} \right.$$

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:

$$\begin{aligned} P(X \geq 6) &= 1 - P(X < 6) = 1 - P(X \leq 5) = 1 - F_X(5) = 1 - \sum_{x=0}^{5} \frac{20^x}{x!} \cdot e^{-20} \\ &= 1 - \left(\tfrac{20^0}{0!} \cdot e^{-20} + \tfrac{20^1}{1!} \cdot e^{-20} + \tfrac{20^2}{2!} \cdot e^{-20} + \tfrac{20^3}{3!} \cdot e^{-20} + \tfrac{20^4}{4!} \cdot e^{-20} + \tfrac{20^5}{5!} \cdot e^{-20} \right) \\ &= 1 - e^{-20} \cdot \left(1 + 20 + \tfrac{400}{4} + \tfrac{8000}{6} + \tfrac{160000}{24} + \tfrac{3200000}{120} \right) \\ &= 1 - e^{-20} \cdot \left(\tfrac{1 \cdot 120 + 20 \cdot 120 + 400 \cdot 30 + 8000 \cdot 20 + 160000 \cdot 24 + 3200000 \cdot 1}{120} \right) \\ &= 1 - e^{-20} \cdot \left(\tfrac{4186520}{120} \right) = 1 - 7{,}1908841 \times 10^{-5} = 0{,}9999281. \end{aligned}$$

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 2 3 3 5 3 3 2 4 2 3 2 1 1 3 4 6 2 5 3
##  [36] 4 1 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
##  [71] 4 5 0 4 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)
```

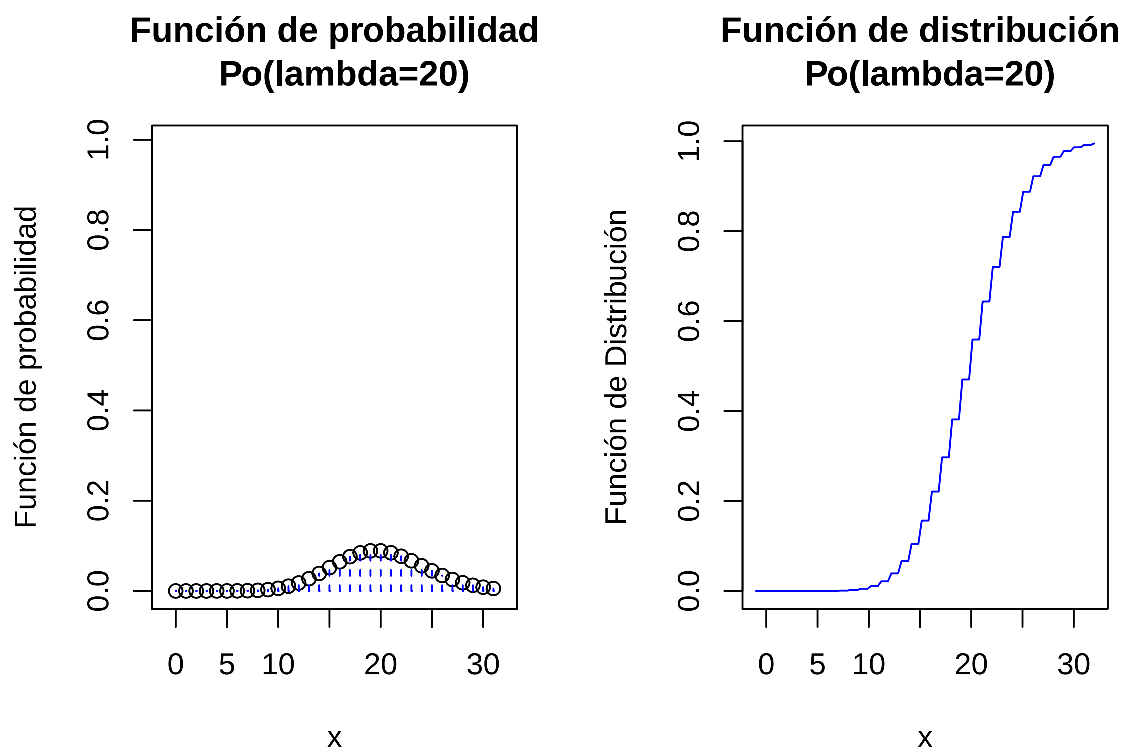
```
## [1] 0.9999281
```

```
ppois(5,lambda=20,lower.tail =FALSE ) ## acumula hacia arriba P(X>5)=P(X>=6)=P(X=6)+P(X=7)+...
```

```
## [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)*2,2)]=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",ylab="Función de probabilidad",
     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",ylab="Función de Distribución",
      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",ylab="Función de probabilidad",
       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",ylab="Función de Distribución",
```

```
        main=paste0(c("Función de distribución \n Po(lambda=",lambda,")"),collapse = ""))
  par(mfrow=c(1,1))
})
```

Cálculos con python

Realicemos los mismos cálculos realizados con `R` pero ahora usando pyhton. 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)
```

```
## 0.04978706836786395
```

```
poisson.cdf(1,mu = 3)
```

```
## 0.1991482734714558
```

```
poisson.pmf(0,mu = 3)+poisson.pmf(1,mu= 3) ### es igual a poisson.cdf(1,lambda=3)
```

```
## 0.1991482734714558
```

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))
```

```
## 0.9988975118698846
```

```
poisson.cdf(10,mu=3)
```

```
## 0.9997076630493527
```

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)
```

```
## 0.08460506418293791
```

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)
```

```
## 0.9999280911594716
```

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.), array(3.))
```

Y también generar secuencias de observaciones aleatorias de una población $Po(3)$:

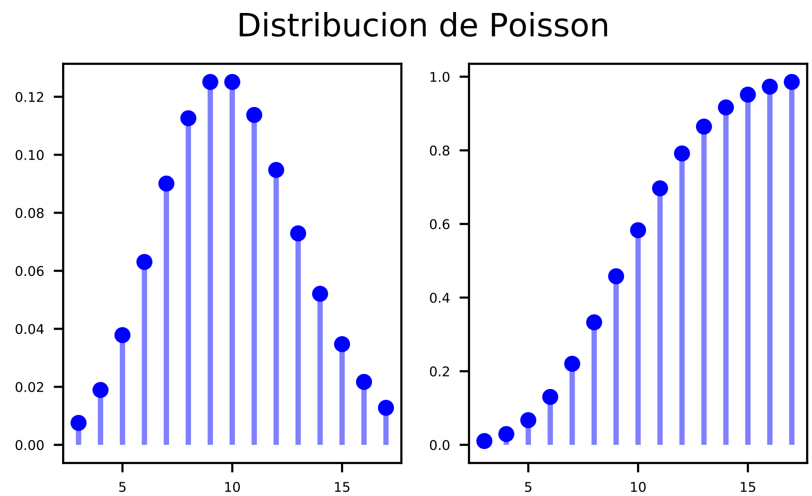
```
poisson.rvs(mu=3,size=40)
```

```
## array([10,  3,  4,  3,  3,  5,  4,  0,  3,  3,  3,  4,  2,  4,  5,  4,  3,
##         8,  2,  2,  1,  7,  2,  3,  4,  3,  6,  3,  3,  2,  2,  4,  4,  5,
##         5,  4,  1,  1,  3,  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",ylab="Función de probabilidad",
```

```
     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",ylab="Función de Distribución",
       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 un 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 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 la v.a. que cuenta el número de bolas blancas extraídas.

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 número 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

$$D_X = \{x \in \mathbf{N} \mid \max\{0, k - n\} \leq x \leq \min\{m, k\}\}$$

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 si 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 si 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:

$$P_X(x) = \begin{cases} \frac{\binom{m}{x}\cdot\binom{n}{k-x}}{\binom{m+n}{k}}, & \text{si } \max\{0, k-n\} \leq x \leq \min\{m, k\}, \text{ para } x \in \mathbf{N}, \\ 0, & \text{en otro caso.} \end{cases}$$

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 $H(m, n, k)$

$X =$número de bolas blancas en k extracciones sin reposición de una urna con m bolas blancas y n negras.	$H(m, n, k)$
$D_X=$	$\{x \in \mathbf{N} \mid \max\{0, k-n\} \leq x \leq \min\{m, k\}\}$
$P_X(x) = P(X = x) =$	$\begin{cases} \frac{\binom{m}{x}\cdot\binom{n}{k-x}}{\binom{m+n}{k}}, & \text{si } \max\{0, k-n\} \leq x \leq \min\{m, k\}, \\ 0, & \text{en otro caso.} \end{cases}$
$F_X(x) = P(X \leq x)$	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 esta 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:

$$P_X(x) = P(X = x) = \begin{cases} \frac{\binom{m}{x}\cdot\binom{n}{k-x}}{\binom{m+n}{k}}, & \text{si } \max\{0, k-n\} \leq x \leq \min\{m, k\}, \text{ para } x \in \mathbf{N}, \\ 0, & \text{en otro caso.} \end{cases}$$

Sustituyendo los parámetros m, n y k por $m = 15$, $n = 10$ y $k = 3$, obtenemos:

$$P_X(x) = P(X = x) = \begin{cases} \frac{\binom{15}{x}\cdot\binom{10}{3-x}}{\binom{25}{3}} = \frac{\binom{15}{x}\cdot\binom{10}{3-x}}{2300}, & \text{si } 0 \leq x \leq 3, \text{ para } x \in \mathbf{N}, \\ 0, & \text{en otro caso.} \end{cases}$$

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, serà, 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:

$$\begin{aligned} Var(X) &= k\cdot\frac{m}{m+n}\cdot\left(1 - \frac{m}{m+n}\right)\cdot\frac{m+n-k}{m+n-1} \\ &= 3\cdot\frac{15}{15+10}\cdot\left(1 - \frac{15}{15+10}\right)\cdot\frac{15+10-3}{15+10-1} \\ &= 3\cdot\frac{15}{25}\cdot\left(1 - \frac{15}{25}\right)\cdot\frac{22}{24} = 3\cdot\frac{15}{25}\cdot\frac{25-15}{25}\cdot\frac{22}{24} \\ &= 3\cdot\frac{15}{25}\cdot\frac{10}{25}\cdot\frac{22}{24} = 0{,}66. \end{aligned}$$

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=15,10,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)
```

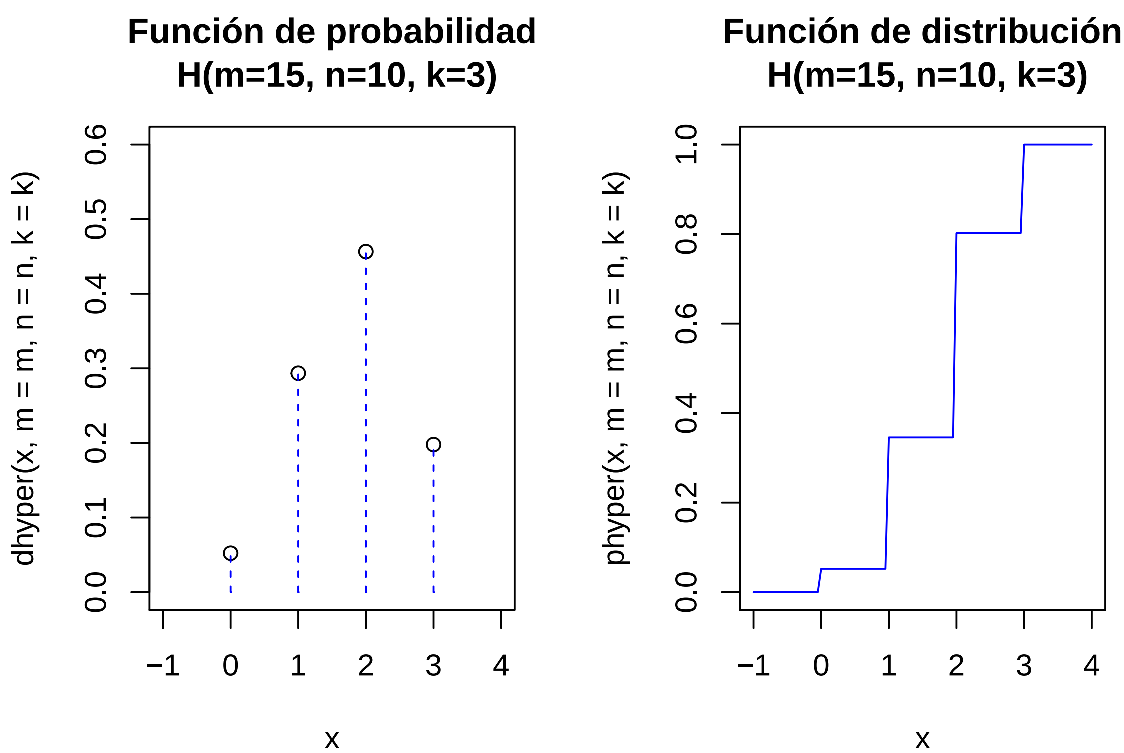
```
##   [1] 2 3 1 3 1 2 2 3 2 2 1 2 1 2 2 3 3 1 1 1 1 0 2 3 2 1 3 2 2 2 2 3 2 3 3
##  [36] 2 0 1 2 1 3 2 2 3 2 3 2 2 3 2 3 1 2 2 2 2 3 2 2 1 3 2 2 3 1 2 2 2 2 2
##  [71] 3 0 2 0 3 2 2 2 1 2 2 3 1 1 1 2 2 2 2 1 1 3 2 2 3 2 2 1 1 1 3 3 2 2 2
## [106] 1 3 2 2 2 1 1 2 3 2 2 1 2 2 2 2 2 2 3 1 2 3 3 1 1 2 2 1 1 3 2 1 1 2 2
## [141] 3 1 1 1 2 1 1 3 1 2 2 3 3 2 3 1 2 1 2 2 2 1 2 3 1 3 3 3 2 2 1 3 3 1 1
## [176] 2 2 2 2 2 3 2 1 2 1 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(phyper(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(phyper(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",ylab="Función de probabilidad",
       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", ylab="Función de Distribución",
         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)
```

```
## 0.29347826086956585
```

```
hypergeom.cdf(1,M=15+10,n=15,N=3)
```

```
## 0.3456521739130442
```

```
1-hypergeom.cdf(1,M=15+10,n=15,N=3)
```

```
## 0.6543478260869557
```

Una muestra aleatoria de este experimento sería:

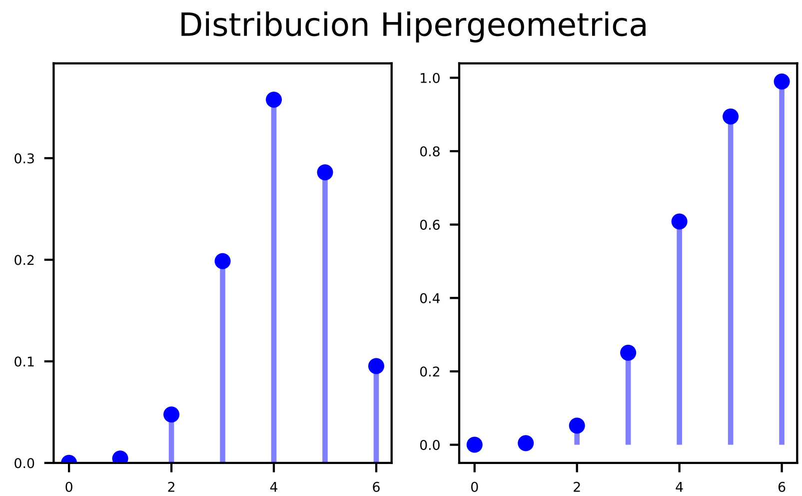
```
hypergeom.rvs(M=15+10,n=15,N=3,size=100)
```

```
## array([3, 2, 2, 1, 2, 3, 1, 3, 0, 1, 2, 2, 3, 3, 3, 1, 3, 3, 2, 1, 1, 2,
##        2, 1, 2, 1, 1, 1, 2, 2, 3, 2, 2, 3, 2, 2, 2, 0, 1, 2, 0, 1, 0, 2,
##        2, 2, 2, 3, 1, 2, 1, 2, 2, 1, 0, 3, 1, 2, 2, 3, 1, 3, 2, 2, 3, 2,
##        1, 2, 1, 2, 3, 2, 2, 1, 0, 1, 1, 2, 2, 2, 2, 1, 2, 2, 3, 3, 2, 1,
##        1, 1, 0, 2, 2, 1, 3, 1, 3, 2, 2, 3])
```

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. Distribuciones continuas

3.2.1. 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 $E(X^2)$.

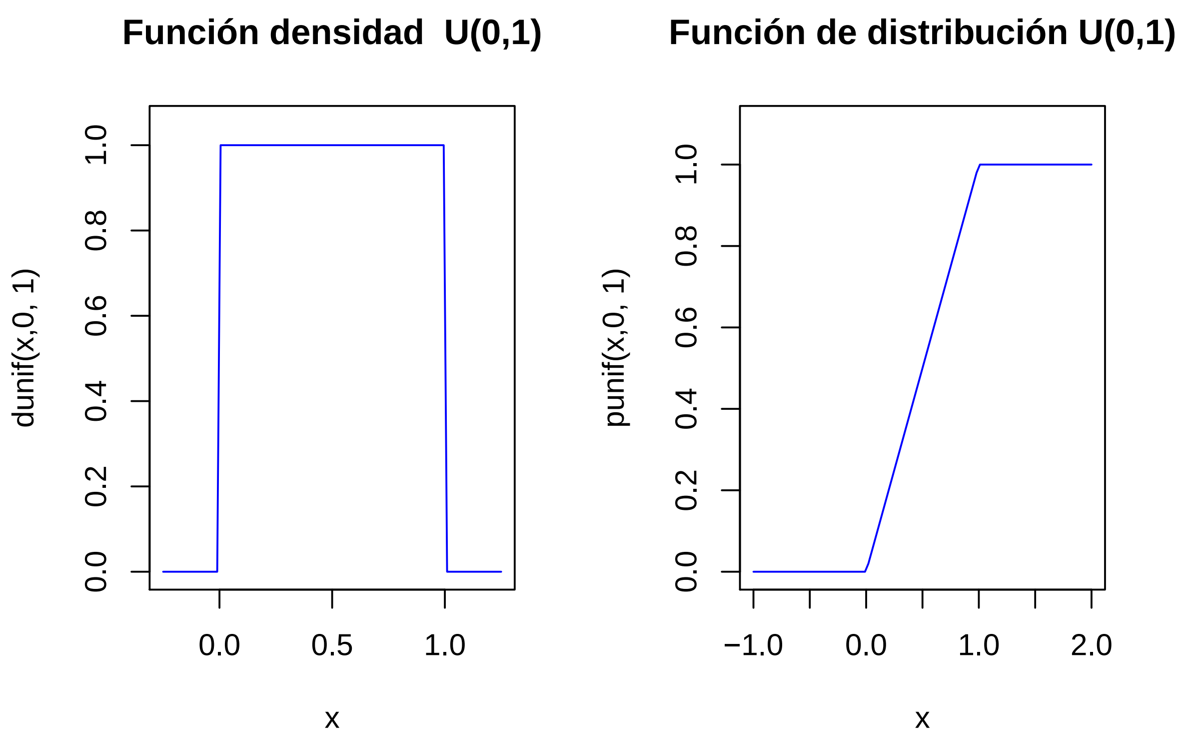
Calculemos $Var(X)$.

$$\begin{aligned} Var(X) &= E(X^2) - (E(X))^2 = \frac{b^2+ab+a^2}{3} - \left(\frac{b+a}{2}\right)^2 \\ &= \frac{b^2+ab+a^2}{3} - \frac{b^2+2ab+a^2}{4} \\ &= \frac{4\cdot(b^2+ab+a^2) - 3\cdot(b^2+2ab+a^2)}{4\cdot 3} \\ &= \frac{b^2-2ab+a^2}{12} = \frac{(b-a)^2}{12}. \end{aligned}$$

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))
```



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
  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),2), 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:

$$F_X(x) = P(X \leq x) = \left\{\begin{array}{ll} 0, & \text{si } x \leq a, \\ \frac{x-a}{b-a}, & \text{si } a \leq x \leq b, \\ 1, & \text{si } b \leq x. \end{array}\right.$$

Si T vale $T = scale \cdot X + loc$, su función de distribución será:

$$\begin{aligned} F_T(t) &= P(T \leq t) = P(scale \cdot X + loc \leq t) = P\left(X \leq \frac{t-loc}{scale}\right) = F_X\left(\frac{t-loc}{scale}\right) \\ &= \left\{\begin{array}{ll} 0, & \text{si } \frac{t-loc}{scale} \leq a \\ \frac{\frac{t-loc}{scale}-a}{b-a}, & \text{si } a \leq \frac{t-loc}{scale} \leq b, \\ 1, & \text{si } b \leq \frac{t-loc}{scale}, \end{array}\right. \\ &= \left\{\begin{array}{ll} 0, & \text{si } t \leq scale \cdot a + loc, \\ \frac{t-(scale\cdot a+loc)}{scale\cdot(b-a)}, & \text{si } scale \cdot a + loc \leq t \leq scale \cdot b + loc, \\ 1, & \text{si } scale \cdot b + loc \leq t, \end{array}\right. \\ &= \left\{\begin{array}{ll} 0, & \text{si } t \leq scale \cdot a + loc, \\ \frac{t-(scale\cdot a+loc)}{scale\cdot b+loc-(scale\cdot a+loc)}, & \text{si } scale \cdot a + loc \leq t \leq scale \cdot b + loc, \\ 1, & \text{si } scale \cdot b + loc \leq t, \end{array}\right. \end{aligned}$$

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 probabilidad sigue T?

3.2.2. Resumen v.a con distribución uniforme, $U(a,b)$

Distribución uniforme	$U(a,b)$
Dominio	$D_X = (a,b)$
$f_X(x)$	$\left\{\begin{array}{ll} \frac{1}{b-a}, & \text{si } a < x < b, \\ 0, & \text{en cualquier otro caso.} \end{array}\right.$
$F_X(x) = P(X \leq X) =$	$\left\{\begin{array}{ll} 0, & \text{si } x \leq a \\ \frac{x-a}{b-a}, & \text{si } a \leq x \leq b, \\ 1, & \text{si } b \leq x. \end{array}\right.$
$E(X) =$	$\frac{a+b}{2}$

Distribución uniforme	$U(a,b)$
$Var(X) =$	$\frac{(b-a)^2}{12}$

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 un 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([-0.55776847, -0.22118256,  0.9128915 ,  0.0203245 ,  0.77311668,
##         0.56240535, -0.11745023,  0.78547743, -0.82681453, -0.2237993 ,
##        -0.87446538, -0.73568611,  0.51654938, -0.37151817, -0.36370293,
##         0.98612411, -0.33899483,  0.13763265,  0.81284179, -0.49191018,
##        -0.28115256,  0.83016209, -0.06956841,  0.22300919, -0.24117142,
##         0.2250286 ,  0.79239454, -0.12031712, -0.47082215, -0.96032849])
```

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([0.44969675, 0.33455238, 0.09714847, 0.41024777, 0.09912737])
```

Cuantiles de variables aleatorias

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

$$P(X \leq x_q) \geq q.$$

En `R`, cada distribución X tiene la función `qX(p,...)` que devuelve precisamente el cuantil x_p tal que $P(X \leq x_p) \geq p$.

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 $B(5, 0{,}5)$.

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

En caso de no existir la inversa, dado q, definimos el conjunto $F_X^{-1}(q)$ como:

$$F_X^{-1}(q) = \{x \in \mathbb{R}, \mid F_X(x) = q\}.$$

Entonces el cuantil q sería el mínimo del conjunto anterior considerando sólo valores del dominio de la variable: $x_q = \min_{x \in D_X}(F_X^{-1}(q))$.

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:

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

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(1.5,n=6)
```

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

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

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

```
##  [1] 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 $i = 0, \ldots, 11$ 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, construímos la función `qdado` que nos calcula el cuantil p, para $0 \leq p \leq 1$, de la variable X como el mínimo de la antiimagen de p mediante la función de distribución $F_X^{-1}(p)$

```
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
```

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))
})
```

3.2.3. 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

$$\begin{aligned} P(T > t) &= P(\text{Cero eventos en el intervalo}(0, t]) \\ &= P(N_t = 0) = \tfrac{(\lambda t)^0}{0!} e^{-\lambda t} = e^{-\lambda t}. \end{aligned}$$

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

$$P(X > s + t \big| X > s) = P(X > t) \text{ para todo } s, t \in \mathbb{R}$$

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,

$$\begin{aligned} P(X > s + t \big| X > s) &= \tfrac{P(\{X > s+t\} \cap \{X > s\})}{P(X > s)} = \tfrac{P(X > s+t)}{P(X > s)} = \tfrac{e^{-\lambda \cdot (s+t)}}{e^{-\lambda \cdot s}} = \tfrac{e^{-\lambda \cdot s} \cdot e^{-\lambda \cdot t}}{e^{-\lambda \cdot s}} \\ &= e^{-\lambda \cdot t} = P(X > t). \end{aligned}$$

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

La función de densidad, de distribución y la generación aleatoria de valores de una exponencial, se pueden obtener en R con:

```
dexp(0.001,rate=3)### alerta no es una probailidad 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
```

```
##  [1] 0.32072389 0.57998323 0.26019628 0.48896233 0.31041354 0.11871455
##  [7] 0.10798351 0.18699902 0.10023100 0.03619832
```

3.2.4. Cálculos con python

Y en python con:

```
from scipy.stats import expon
expon.pdf(0.0001,scale= 1./3)
```

```
## 2.9991001349865014
```

```
expon.cdf(0.5,scale= 1./3)
```

```
## 0.7768698398515702
```

```
expon.rvs(scale=1./3,size=10)
```

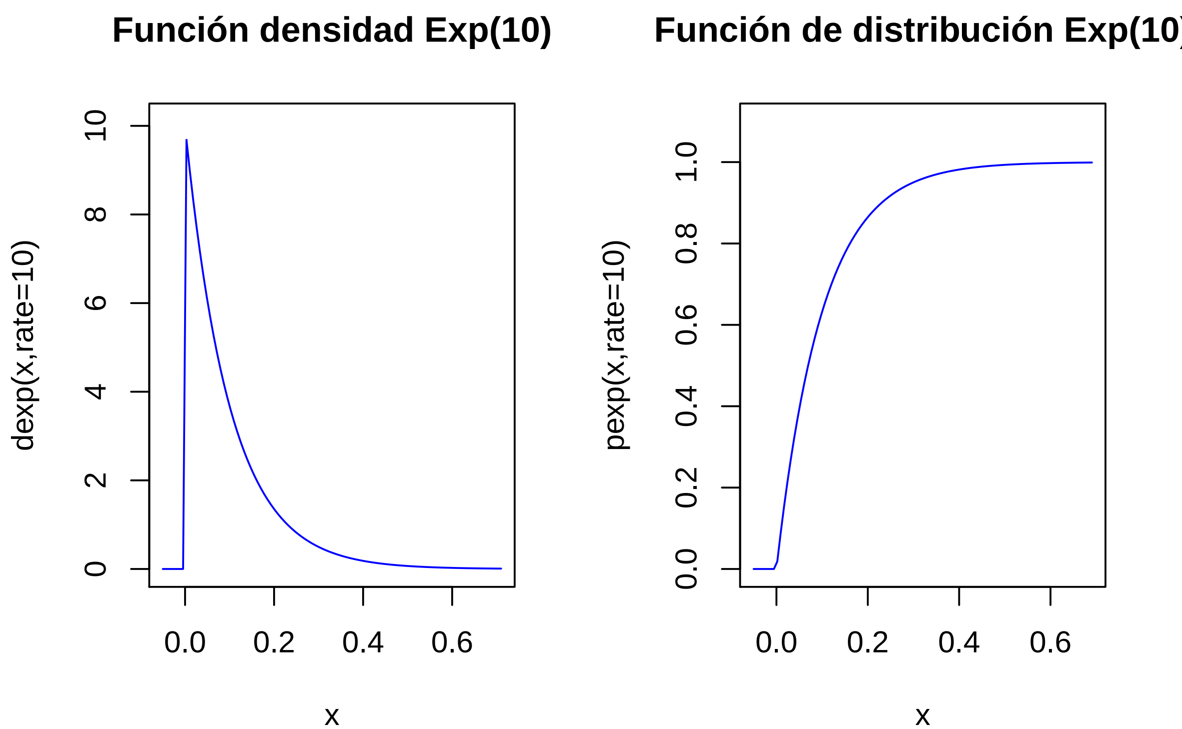
```
## array([0.80408068, 0.59997115, 0.923264  , 0.35575296, 0.62634933,
##        0.25718931, 0.11540355, 0.14608897, 0.54389597, 0.37842878])
```

3.2.5. Resumen v.a con distribución exponencial, $Exp(\lambda)$

X	$Exp(\lambda)$
$D_X =$	$(0,+\infty)$
$f_X(x) =$	$\begin{cases} \lambda e^{-\lambda x} & \text{si } x > 0 \\ 0 & \text{si } x \leq 0 \end{cases}$
$F_X(x) = P(X \leq X) =$	$\begin{cases} 0 & \text{si } x \leq 0 \\ 1 - e^{-\lambda x} & \text{si } x > 0 \end{cases}$
$E(X) = \frac{1}{\lambda}$	$Var(X) = \frac{1}{\lambda^2}$

```
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))
```

3.2.7. Gráficas densidad y distribución $Exp(\lambda = 10)$



Ejercicio

Consultad en el manual de python scipy.stats.

Dibujad la función de densidad y de distribución de una $Exp(10)$.

3.2.8. Gráficas interactivas de una $Exp(\lambda)$.

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))
})
```

3.2.9. Ejercicio: las bombillas que no envejecen.

Ejercicio

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 si X es 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. Distribución normal o Gaussiana

Una de las más variables aleatorias continua más populares son las llamadas distribuciones normales o de Gaussiana .

Distribución normal o de Gauss Diremos que una v.a. X sigue una ley normal de parámetros μ y σ^2 y lo denotaremos por $N(\mu, \sigma)$ si tiene por función de densidad

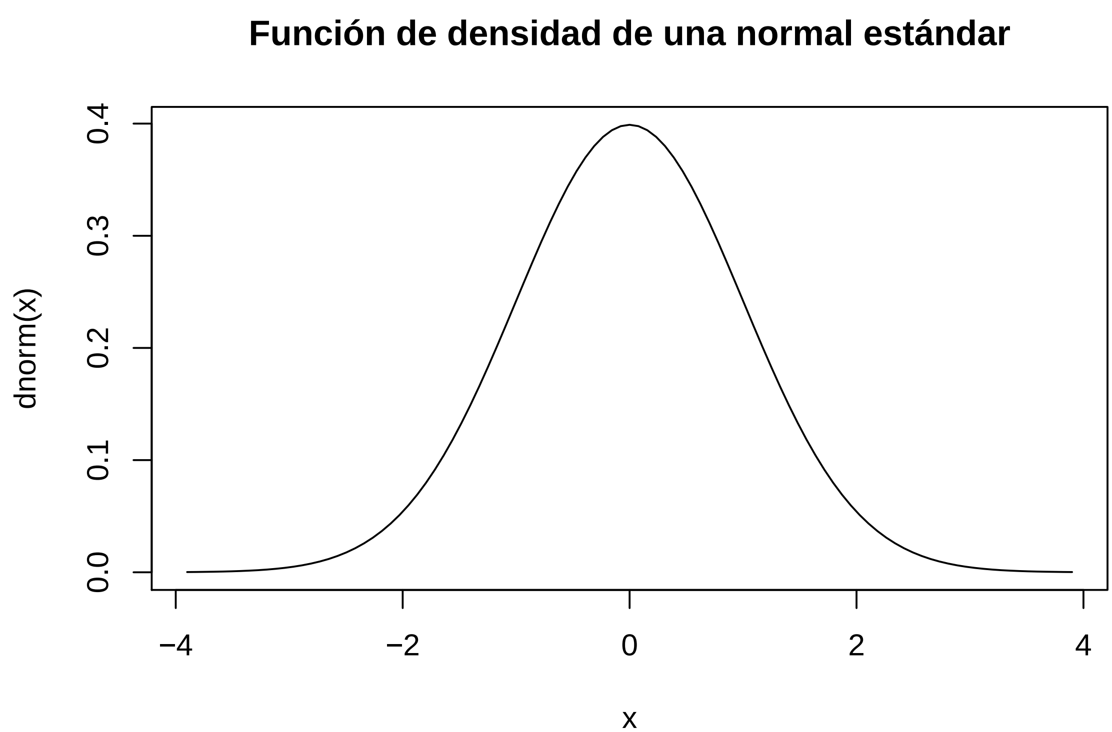
$$f_X(x) = \frac{1}{\sqrt{2 \cdot \pi \cdot \sigma^2}} e^{-\frac{1}{2} \cdot \left(\frac{x-\mu}{\sigma}\right)^2}$$

para todo $x \in \mathbb{R}$.

La gráfica de esta función es la conocida 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 normal $N(0, 1)$.

```
curve(dnorm(x),main="Función de densidad de una normal estándar",xlim=c(-3.9,3.9))
```



3.3.1. Propiedades de la densidad normal

Propiedades de la densidad normal

Sea X una v.a. $N(\mu, \sigma)$ y sea f_X su función de densidad. Entonces:

1. Evidentemente f_X verifica todas las propiedades de las funciones de densidad.
2. $f_X(\mu - x) = f_X(\mu + x)$ es simétrica respecto de la recta $x = \mu$
3. f_X alcanza el máximo en $x = \mu$
4. Si F_X la función de distribución de X entonces $F_X(\mu + x) = 1 - F_X(\mu - x)$.
5. En particular si Z es una $N(0, 1)$ entonces $F_Z(-x) = 1 - F_Z(x)$
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.

3.3.2. Función de distribución N(0,1)

Su función de distribución es, como sabemos :

$$F(x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2 \cdot \pi \cdot \sigma^2}} e^{-\frac{1}{2}\left(\frac{t-\mu}{\sigma}\right)^2} dt.$$

Que no tiene ninguna expresión algebraica "decente". Es por esta razón, y por comodidad, que esta función está tabulada o hay que calcular con un programa.

Cuando una variable tiene distribución normal con parámetros μ y σ la denotamos con X sigue un ley de distribución $N(\mu, \sigma)$.

3.3.3. Resumen v.a con distribución normal, $N(\mu, \sigma)$

X	$N(\mu, \sigma)$
$D_X =$	$\mathbb{R} = (-\infty, +\infty)$
$f_X(x)$	$= \frac{1}{\sqrt{2\pi \cdot \sigma^2}} \cdot e^{\frac{-(x-\mu)^2}{2 \cdot \sigma^2}}$ para todo $x \in \mathbb{R}$.
$F_X(x) = P(X \leq X) =$	`pnorm(x,mean=mu,sd=sigma)`. Utilizad funciones de R o python
$E(X) = \mu$.	$Var(X) = \sigma^2$.

3.3.4. Cálculos con R

De forma la forma habitual los parámetros son `mean` y `sd` la media μ y la desviación estándar σ. Por ejemplo para una $X \sim N(\mu = 1, \sigma = 2)$ la función de densidad $f_X(2)$ se puede calcular como

```r
dnorm(2,mean=1,sd=2)
```

```
## [1] 0.1760327
```

y la función de distribución $F_X(2) = P(X \leq 2)$ como

```r
pnorm(2,mean=1,sd=2)
```

```
## [1] 0.6914625
```

El cuantil $x_{0,95}$ es el valor que cumple $P(X \leq x_{0,95}) = 0{,}95$ como

```
qnorm(0.95,mean=1,sd=2)
```

```
## [1] 4.289707
```

Y la generación aleatoria de valores según X como

```
rnorm(n=5,mean=1,sd=2)
```

```
## [1] -1.7813382  0.7320555  2.7209162  2.9098518  2.5098235
```

3.3.5. Cálculos con python

De forma la forma habitual importaremos `norm` de `scipy.stas` los parámetros son `loc` y `scale` la media μ y la desviación estándar σ.

```
from scipy.stats import norm
```

Por ejemplo para una $X\sim N(\mu=1,\sigma=2)$, la función de densidad $f_X(2)$:

```
norm.pdf(2,loc=1,scale=2)
```

```
## 0.17603266338214976
```

y la función de distribución $F_X(2)=P(X\leq 2)$:

```
norm.cdf(2,loc=1,scale=2)
```

```
## 0.6914624612740131
```

El cuantil $x_{0,95}$ es el valor que cumple $P(X\leq x_{0,95})=0{,}95$ como

```
norm.ppf(0.95,loc=1,scale=2)
```

```
## 4.289707253902945
```

Y la generación aleatoria de valores según X como

```
norm.rvs(loc=1,scale=2,size=5)
```

```
## array([ 2.1138545 ,  1.8882694 ,  2.12838486, -0.88005777,  0.27439974])
```

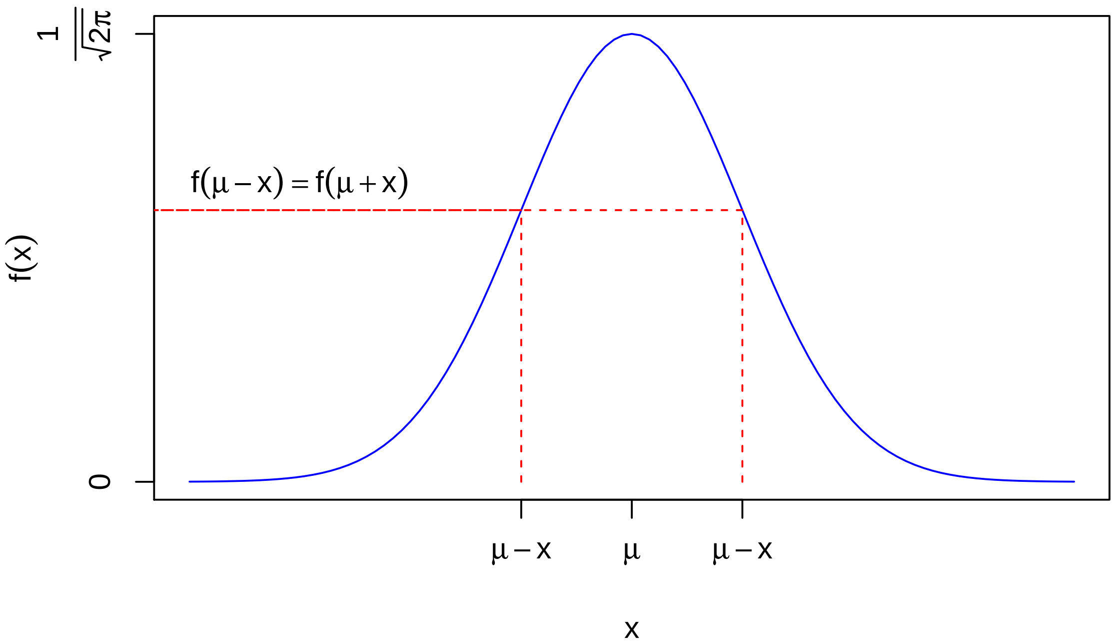
Ejercicio

Consultad SciPy.org para dibujar las funciones de densidad y de distribución con python.

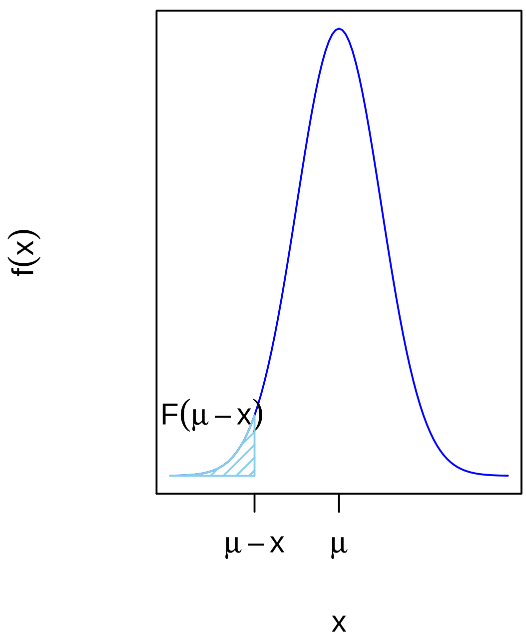
3.3.6. Propiedades de la distribución normal.

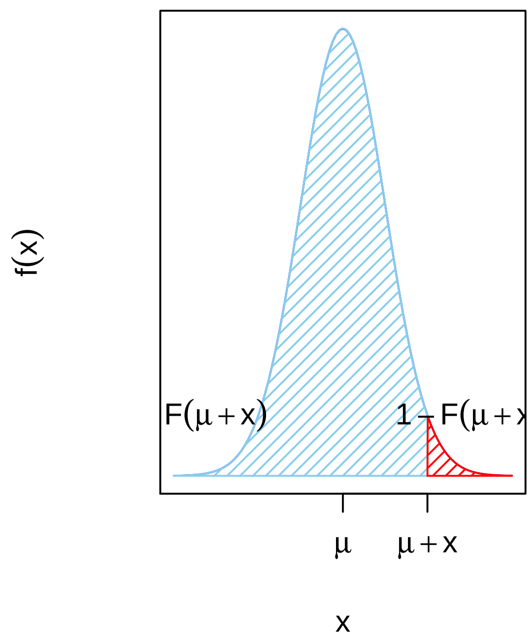
La función de densidad de la distribución normal tiene las siguientes propiedades:

- La función f_X es continua.
- $\int_{-\infty}^{+\infty}\frac{1}{\sqrt{2\cdot\pi\cdot\sigma^2}}\cdot e^{-\frac{1}{2}\cdot\left(\frac{x-\mu}{\sigma}\right)^2}dx=1$. (propiedad de todas las densidades).
- $f(\mu+x)=f(\mu-x)$.
- $F(\mu-x)=1-F(\mu+x)$.



- $\lim_{x\to+\infty} f(x) = \lim_{x\to-\infty} f(x) = 0$ es decir tiene asíntota horizontal a derecha e izquierda.
- f es estrictamente creciente si $x < \mu$ y decreciente si $x > \mu$.
- Alcanza el máximo en $x = \mu$ y en este punto vale $f(\mu) = \frac{1}{\sqrt{2\pi}\sigma}$
- Tiene dos puntos de inflexión en $x = \mu + \sigma$ y en $x = \mu - \sigma$.





3.3.7. Gráficas interactivas parámetros 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("toplef",legend=c(expression(N(mu[1],sigma[1])),expression(N(mu[2],sigma[2]))),col=c("red","blue
curve(dnorm(x,m2,s2),add=TRUE,col="blue",lty=2)
})
```

3.3.8. 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 $a \neq 0, b \in \mathcal{R}$ tiene distribución $N(a\mu + b, |a|\sigma)$

En particular si X sigue una $N(\mu,\sigma)$, tomando $a = \frac{1}{\sigma}$ y $b = \frac{-\mu}{\sigma}$ obtenemos la tipificación o estandarización de la v.a.

$$Z = \frac{X-\mu}{\sigma}$$

se distribuye $N(0,1)$, es decir $E(X) = 0$ y $Var(X) = 1$.

Esta propiedad es muy útil, ya que utilizándola sólo necesitaremos tabular la $N(0,1)$.

Si Z sigue una distribución $N(0,1)$ diremos que Z sigue una distribución normal estándar.

Por lo tanto podemos calcular cualquier distribución normal desde la distribución normal estándar:

$$F_X(x) = F_Z\left(\frac{x-\mu}{\sigma}\right).$$

3.3.9. Propiedades de la distribución normal estándar

Sea Z una $N(0,1)$.

Como en este caso $\mu = 0$ y $\sigma = 1$ tenemos que algunas de las propiedades anteriores se simplifican incluso más:

- De $f_X(\mu - x) = f_X(\mu + x)$ obtenemos $f_Z(-x) = f_Z(x)$

- De $F_X(\mu - x) = 1 - F_X(\mu + x)$ obtenemos $F_Z(-x) = 1 - F(x)$.
- Dado $\delta > 0$,

$$P(-\delta \leq Z \leq \delta) = F_Z(\delta) - F_Z(-\delta) = F_Z(\delta) - (1 - F_Z(\delta)) = 2 \cdot F_Z(\delta) - 1.$$

3.3.10. Cálculos con la distribución normal

Ejercicio Cálculos con la distribución normal estándar

Sea Z una distribución $N(0,1)$, calcular las siguientes probabilidades en función de F_Z.

- $P(-4 \leq Z \leq 4)$.
- $P(-2 \leq Z \leq 2)$.
- $P(Z \leq -2)$.
- $P(Z \leq 2)$.
- $P(Z \geq 2)$.
- $P(Z > 2)$.
- $P(Z = 2)$.
- $P(Z \geq -2)$.

Resolvamos el ejercicio

- $P(-4 \leq Z \leq 4) = F_Z(4) - F_Z(-4) = 2 \cdot F_Z(4) - 1$
- $P(-2 \leq Z \leq 2) = F_Z(2) - F_Z(-2) = 2 \cdot F_Z(2) - 1$
- $P(Z \leq -2) = F_Z(-2) = 1 - F_Z(2)$
- $P(Z \leq 2) = F_Z(2)$
- $P(Z \geq 2) = 1 - P(Z < 2) = 1 - F_Z(2)$
- $P(Z > 2) = 1 - P(Z \leq 2) = 1 - F_Z(2)$
- $P(Z = 2) = 0$ ya que es una distribución continua.
- $P(Z \geq -2) = 1 - P(Z < -2) = 1 - F_Z(-2) = 1 - (1 - F_Z(2)) = F_Z(2)$.

3.3.11. Relación entre una normal y la normal estándar.

Propiedad

Si X es una normal $N(\mu, \sigma)$ y Z es su variable tipificada, es decir, $Z = \frac{X-\mu}{\sigma}$ es una $N(0,1)$ entonces:

$$P(X \leq x) = P\left(\frac{X-\mu}{\sigma} \leq \frac{x-\mu}{\sigma}\right) = F_Z\left(\frac{x-\mu}{\sigma}\right).$$

Propiedad

- Cuando tengamos un intervalo

$$\begin{aligned} P(a < X < b) &= P\left(\frac{a-\mu}{\sigma} < \frac{X-\mu}{\sigma} < \frac{b-\mu}{\sigma}\right) = \\ &= P\left(\frac{a-\mu}{\sigma} < Z < \frac{b-\mu}{\sigma}\right) \\ &= F_Z\left(\frac{b-\mu}{\sigma}\right) - F_Z\left(\frac{a-\mu}{\sigma}\right) \end{aligned}$$

- Si $\delta > 0$ $P\left(\mu-\delta \leq X \leq \mu+\delta\right) = 2\cdot F_Z\left(\frac{\delta}{\sigma}\right)-1$

3.3.12. Ejemplo cálculo probabilidades normal

Ejercicio

Sea X una normal com media 2 y varianza 4. Calcular

- $P(1 < X < 2)$.
- $P(X > 3)$.

3.3.13. Ejemplo cálculo probabilidades normal

Solución

El cálculo de la primera pregunta es

$$\begin{aligned} P(1<X<2) &= P\left(\frac{1-2}{2}<\frac{X-2}{2}<\frac{2-2}{2}\right) = P\left(\frac{-1}{2}<Z<0\right) \\ &= F_Z(0)-F_Z(-0{,}5) = \frac{1}{2}-1+F_Z(0{,}5) = -\frac{1}{2}+F_Z(0{,}5). \end{aligned}$$

La segunda cuestión se resuelve así

$$P(X>3) = P\left(\frac{X-2}{2}>\frac{3-2}{2}\right) = P(Z>0{,}5) = 1-F_Z(0{,}5).$$

3.3.14. Ejemplo normal con R y python

Ejercicio

Sea X una normal com media 2 y varianza 4. Calcular con R y con python las probabilidades

- $P(1 < X < 2)$.
- $P(X > 3)$.

Solución con R

```r
pnorm(2,mean=2,sd=4)-pnorm(1,mean=2,sd=4)#P(1< X< 2)
```

```
## [1] 0.09870633
```

```r
pnorm(3,mean=2,sd=4,lower.tail =FALSE)#P(X>3)
```

```
## [1] 0.4012937
```

```r
1-pnorm(3,mean=2,sd=4,lower.tail=TRUE)#P(X>3) = 1-P(X<=3)
```

```
## [1] 0.4012937
```

Solución con Python

```
norm.cdf(2,loc=2,scale=4)-norm.cdf(1,loc=2,scale=4)#P(1< X< 2)
```

```
## 0.0987063256829237
```

```
1-norm.cdf(3,loc=2,scale=4)#P(X>3) = 1-P(X<=3)
```

```
## 0.4012936743170763
```

```
norm.stats(loc=2,scale=4,moments="mv")#E(X), Var(X)
```

```
## (array(2.), array(16.))
```

3.3.15. 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.

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 $m_n = E\left(X^n\right)$.

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.

4.1.1.1. 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:

$$P_X(0) = q = 1 - p,\ p_X(1) = p.$$

Su momento de orden n será:

$$m_n = E\left(X^n\right) = p \cdot 1^n + (1-p) \cdot 0^n = p.$$

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: $f_X(x) = \lambda \mathrm{e}^{-\lambda x}$, para $x \geq 0$ y 0, en caso contrario.

Su momento de orden n será:

$$m_n = E\left(X^n\right) = \int_0^\infty \lambda \mathrm{e}^{-\lambda x} x^n \, dx = \frac{n!}{\lambda^n}.$$

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: $\lim_{n\to\infty} m_n = \lim_{n\to\infty} \frac{n!}{\lambda^n} = \infty$.

Ejemplo: momento de orden n de una variable normal de parámetros $m = 0$ y $\sigma = 1$

Recordemos que su función de densidad es: $f_X(x) = \frac{1}{\sqrt{2\pi}}\mathrm{e}^{-\frac{x^2}{2}}$, 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: $m_2 = E\left(X^2\right) = \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}}\mathrm{e}^{-\frac{x^2}{2}} \cdot x^2\, dx = 1$. La integral anterior se resuelve usando técnicas de integrales de dos variables. Dicho valor también se puede obtener usando que su varianza vale 1: $m_2 = \mathrm{Var}(X)+E(X)^2 = \sigma^2+0^2 = 1$.

Los momentos de orden impar n serán cero ya que integramos una función impar: $m_n = E\left(X^n\right) = \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}}\mathrm{e}^{-\frac{x^2}{2}} \cdot x^n\, dx = 0$. O sea, si consideramos $g(x) = \frac{1}{\sqrt{2\pi}}\mathrm{e}^{-\frac{x^2}{2}} \cdot x^n$, se verifica $g(-x) = -g(x)$, para todo $x \in \mathbb{R}$.

Si intentamos calcular el momento de orden 4, obtenemos: $m_4 = E\left(X^4\right) = \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}}\mathrm{e}^{-\frac{x^2}{2}} \cdot x^4\, dx = 3$, 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 $\mu_n = E\left((X-\mu)^n\right)$, 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:

$$\mu_1 = E\left((X-\mu)\right) = E(X) - E(\mu) = E(X) - E(X) = 0.$$

Observación. El momento central de orden 2 de una variable aleatoria es la varianza:

$$\mu_2 = E\left((X-\mu)^2\right) := \mathrm{Var}(X).$$

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**:

$$\mu_n = E\left((X-\mu)^n\right) = E\left(\sum_{k=0}^{n}(-1)^{n-k}\binom{n}{k}X^k\mu^{n-k}\right).$$

Aplicando la propiedad de la esperanza que la esperanza de la suma es la suma de esperanzas, obtenemos la expresión dada por la proposición:

$$\mu_n = \sum_{k=0}^{n} (-1)^{n-k} \binom{n}{k} \mu^{n-k} E\left(X^k\right) = \sum_{k=0}^{n} (-1)^{n-k} \binom{n}{k} \mu^{n-k} m_k.$$

4.1.2.1. 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:

$$P_X(0) = q = 1 - p,\ P_X(1) = p.$$

Usando que $E(X) = p$, su momento central de orden n será:

$$\mu_n = E\left((X-p)^n\right) = p \cdot (1-p)^n + (1-p) \cdot (0-p)^n = p(1-p)^n + (-1)^n (1-p) p^n.$$

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: $f_X(x) = \lambda \mathrm{e}^{-\lambda x}$, para $x \geq 0$.

Usando que $E(X) = \frac{1}{\lambda}$, su momento central de orden n será:

$$\mu_n = E\left(\left(X - \frac{1}{\lambda}\right)^n\right) = \int_0^{\infty} \lambda \mathrm{e}^{-\lambda x} \left(x - \frac{1}{\lambda}\right)^n dx = \frac{a_n}{\lambda^n},$$

donde $a_n = n! \sum\limits_{k=0}^{n} \frac{(-1)^k}{k!}$.

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: $\lim\limits_{n\to\infty} \mu_n = \lim\limits_{n\to\infty} \frac{a_n}{\lambda^n} = \infty$:

$$\lim_{n\to\infty} \mu_n = \lim_{n\to\infty} \frac{n! \sum\limits_{k=0}^{n} \frac{(-1)^k}{k!}}{\lambda^n} = \lim_{n\to\infty} \sum_{k=0}^{n} \frac{(-1)^k}{k!} \cdot \lim_{n\to\infty} \frac{n!}{\lambda^n} = \mathrm{e}^{-1} \cdot \infty = \infty.$$

Ejemplo: momento central de orden n de una variable normal de parámetros μ y σ

Recordemos que su función de densidad es: $f_X(x) = \frac{1}{\sqrt{2\pi}\sigma} \mathrm{e}^{-\frac{(x-\mu)^2}{2\sigma^2}}$, 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 $x = \mu$: $\mu_n = E\left((X-\mu)^n\right) = \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}\sigma}\mathrm{e}^{-\frac{(x-\mu)^2}{2\sigma^2}} \cdot (x-\mu)^n\, dx = 0$. O sea, si consideramos $g(x) = \frac{1}{\sqrt{2\pi}\sigma}\mathrm{e}^{-\frac{(x-\mu)^2}{2\sigma^2}} \cdot (x-\mu)^n$, se verifica $g(\mu - x) = -g(\mu + x)$, para todo $x \in \mathbb{R}$.

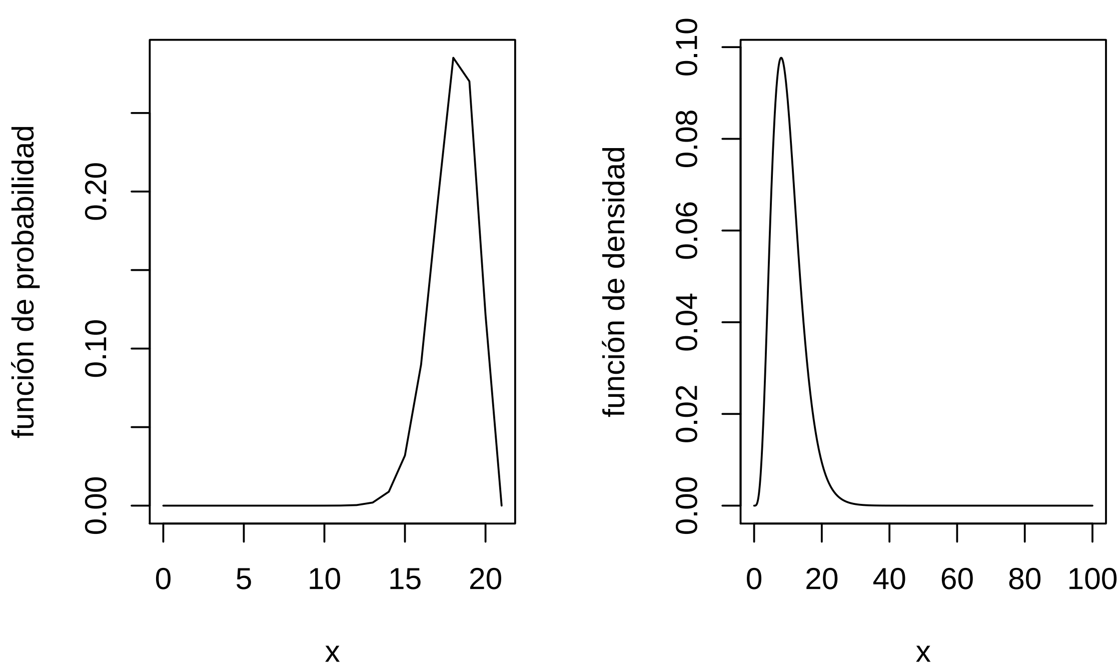
Si intentamos calcular el momento central de orden 4, obtenemos: $\mu_4 = E\left((X-\mu)^4\right) = \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}\sigma}\mathrm{e}^{-\frac{(x-\mu)^2}{2\sigma^2}} \cdot (x-\mu)^4\, dx = 3\sigma^4$. La integral anterior puede resolverse con el cambio de variable $t = \frac{x-\mu}{\sigma}$ y usando que: $\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}}\mathrm{e}^{-\frac{x^2}{2}} \cdot x^4\, dx = 3$.

4.2. Asimetría de una variable aleatoria

4.2.1. Definición

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 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:



4.2.2. ¿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 $\sigma^2 = \mathrm{Var}(X)$.

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 $\gamma_1 < 0$.

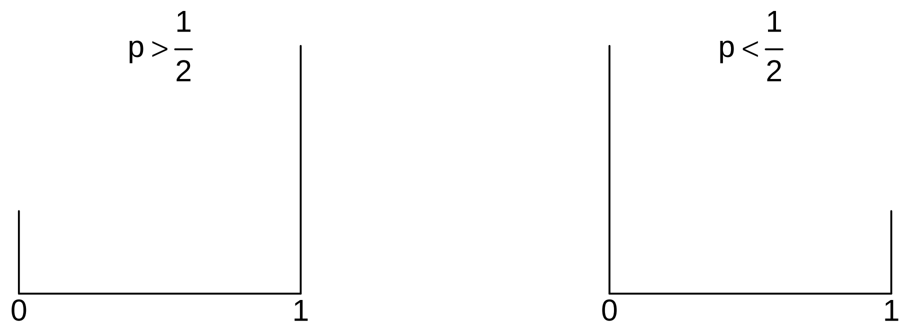
4.2.2.1. Ejemplo de cálculo de asimetría

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:

$$\gamma_1 = \frac{p - 3p(p-p^2) - p^3}{\sqrt{(p-p^2)^3}} = \frac{p(1-p)(1-2p)}{\sqrt{(p-p^2)^3}}.$$

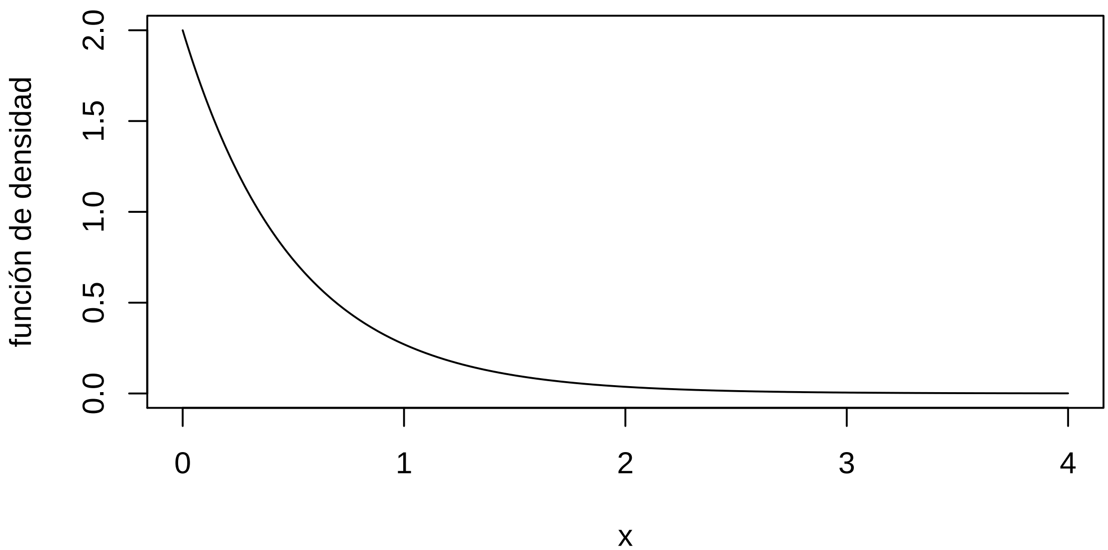
Por tanto, la variable de Bernoulli de parámetro p tendrá simetria negativa si $p > \frac{1}{2}$ y positiva, si $p < \frac{1}{2}$:



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: $\gamma_1 = \frac{\frac{2}{\lambda^3}}{\frac{1}{\lambda^3}} = 2$.

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

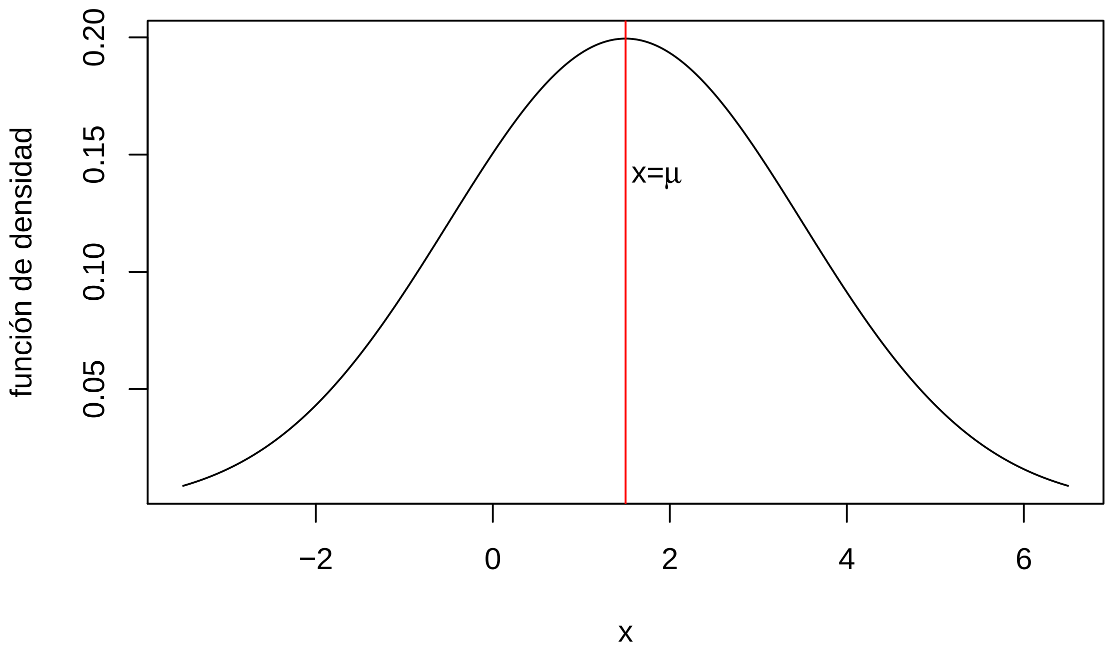
Sea X una variable aleatoria normal de parámetros μ y σ.

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 $f_X(x) = \frac{1}{\sqrt{2\pi}\sigma} \mathrm{e}^{-\frac{(x-\mu)^2}{2\sigma^2}}$, 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

4.3.1. Definición

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 estándard 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 $\sigma^2 = \mathrm{Var}(X)$.

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 $\gamma_2 < 3$.

4.3.1.1. Ejemplo de cálculo de curtosis

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 $\mu_4 = \frac{a_4}{\lambda^3} = \frac{9}{\lambda^4}$, su coeficiente de asimetría de Pearson será: $\gamma_2 = \frac{\frac{9}{\lambda^4}}{\frac{1}{\lambda^4}} = 9$.

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: $m_X(t) = E\left(\mathrm{e}^{tX}\right)$.

4.4.1.1. Ejemplo de cálculo de función generatriz de momentos

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:

$$P_X(0) = q = 1 - p,\ p_X(1) = p.$$

Su función generatriz de momentos será:

$$m_X(t) = E\left(\mathrm{e}^{tX}\right) = p\mathrm{e}^{t\cdot 1} + (1-p)\mathrm{e}^{t\cdot 0} = p\mathrm{e}^{t} + (1-p) = 1 + p\left(\mathrm{e}^{t} - 1\right).$$

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: $f_X(x) = \lambda\mathrm{e}^{-\lambda x}$, para $x \geq 0$ y 0, en caso contrario.

Su función generatriz de momentos será:

$$m_X(t) = E\left(\mathrm{e}^{tX}\right) = \int_0^\infty \mathrm{e}^{tx}\lambda\mathrm{e}^{-\lambda x}\,dx = \lambda\int_0^\infty \mathrm{e}^{(t-\lambda)x}\,dx = \lambda\left[\frac{\mathrm{e}^{(t-\lambda)x}}{t-\lambda}\right]_{x=0}^{x=\infty} = \frac{\lambda}{\lambda - t},\ \text{si } t < \lambda.$$

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: $f_X(x) = \frac{1}{\sqrt{2\pi}\sigma}\mathrm{e}^{-\frac{(x-\mu)^2}{2\sigma^2}}$, para $x \in \mathbb{R}$.

Su función generatriz de momentos será:

$$\begin{aligned} m_X(t) &= E\left(\mathrm{e}^{tX}\right) = \int_{-\infty}^{\infty} \mathrm{e}^{tx}\frac{1}{\sqrt{2\pi}\sigma}\mathrm{e}^{-\frac{(x-\mu)^2}{2\sigma^2}}\,dx = \frac{1}{\sqrt{2\pi}\sigma}\int_{-\infty}^{\infty} \mathrm{e}^{tx-\frac{(x-\mu)^2}{2\sigma^2}}\,dx \\ &= \frac{1}{\sqrt{2\pi}\sigma}\int_{-\infty}^{\infty} \mathrm{e}^{-\frac{1}{2\sigma^2}\left((x-(\sigma^2 t+\mu))^2 - 2\sigma^2 t\mu - \sigma^4 t^2\right)}\,dx = \frac{1}{\sqrt{2\pi}\sigma}\mathrm{e}^{\frac{1}{2}(2t\mu+\sigma^2 t^2)}\int_{-\infty}^{\infty} \mathrm{e}^{-\frac{1}{2\sigma^2}(x-(\sigma^2 t+\mu))^2}\,dx \\ &= \mathrm{e}^{\frac{1}{2}(2t\mu+\sigma^2 t^2)}\left(\frac{1}{\sqrt{2\pi}\sigma}\int_{-\infty}^{\infty} \mathrm{e}^{-\frac{1}{2\sigma^2}(x-(\sigma^2 t+\mu))^2}\,dx\right) = \mathrm{e}^{t\mu+\frac{\sigma^2 t^2}{2}}. \end{aligned}$$

La integral del último paréntesis se resuelve hacienta el cambio de variable $u = x - \sigma^2 t$ y usando que la integral de la función de densidad de X sobre todo $\mathbb{R}$ vale 1.

4.4.1.2. 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:

$$m_n = E\left(X^n\right) = \frac{d}{dt^n} m_X(t)|_{t=0} = m_X^{(n)}(0).$$

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: $m_X(t) = E\left(\mathrm{e}^{tX}\right)$.

La idea de la demostración es probar por inducción que $m_X^{(n)}(t) = E\left(\mathrm{e}^{tX} \cdot X^n\right)$.

Veámoslo para $n = 1$: $m'_X(t) = E\left(\mathrm{e}^{tX} \cdot X\right)$.

Seguidamente, apliquemos inducción sobre n. Supongamos que $m_X^{(n)}(t) = E\left(\mathrm{e}^{tX} \cdot X^n\right)$ y veamos que $m_X^{(n+1)}(t) = E\left(\mathrm{e}^{tX} \cdot X^{n+1}\right)$: $m_X^{(n+1)}(t) = \frac{d}{dt}(m_X^{(n)}(t)) = \frac{d}{dt} E\left(\mathrm{e}^{tX} \cdot X^n\right) = E\left(\mathrm{e}^{tX} \cdot X^{n+1}\right)$, tal como queríamos demostrar.

Ahora si aplicamos la expresión demostrada $m_X^{(n)}(t) = E\left(\mathrm{e}^{tX} \cdot X^n\right)$ a $t = 0$, obtenemos: $m_X^{(n)}(0) = E\left(X^n\right) = m_n$, tal como dice la proposición.

4.4.1.3. Ejemplo

Ejemplo: aplicación de la proposición en el caso en que X es una variable de Bernoulli de parámetro p

En este caso, recordemos que: $m_X(t) = 1 + p\left(\mathrm{e}^t - 1\right)$.

Se puede comprobar que $m_X^{(n)}(t) = p\mathrm{e}^t$. 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: $m_X(t) = \mathrm{e}^{t\mu + \frac{\sigma^2 t^2}{2}}$.

Aplicando la fórmula de los momentos para $n = 1$ obtenemos: $m'(t) = \mathrm{e}^{t\mu + \frac{\sigma^2 t^2}{2}} \left(\mu + t\sigma^2\right)$, que en $t = 0$ vale: $m'(0) = \mu = E(X)$, tal como ya sabemos.

Si la aplicamos para $n = 2$, obtenemos: $m''(t) = \mathrm{e}^{t\mu + \frac{\sigma^2 t^2}{2}} \left((\mu + t\sigma^2)^2 + \sigma^2\right) = \mathrm{e}^{t\mu + \frac{\sigma^2 t^2}{2}} \left(t^2\sigma^4 + \mu^2 + \sigma^2 + 2t\mu\sigma^2\right)$, que en $t = 0$ vale: $m''(0) = \mu^2 + \sigma^2 = E\left(X^2\right)$, tal como ya sabemos.

Para $n = 3$m obtenemos: $m'''(t) = e^{\mu t + \frac{\sigma^2 t^2}{2}} \left(\mu + \sigma^2 t\right) \left(\left(\mu + \sigma^2 t\right)^2 + 3\sigma^2\right)$, que en $t = 0$ vale: $m'''(0) = 3\sigma^2\mu = E\left(X^3\right)$, valor que correspondería al momento de tercer orden de X.

Por último, para $n = 4$, obtenemos: $m^{(iv)}(t) = e^{\mu t + \frac{\sigma^2 t^2}{2}} \left(6\sigma^2 \left(\mu + \sigma^2 t\right)^2 + \left(\mu + \sigma^2 t\right)^4 + 3\sigma^4\right)$, que en $t = 0$ vale: $m^{(iv)}(0) = 6\sigma^2\mu^2 + \mu^4 + 3\sigma^4 = E\left(X^4\right)$, 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: $\phi_X(w) = E\left(\mathrm{e}^{\mathrm{i}wX}\right)$, donde i es el número complejo $\mathrm{i} = \sqrt{-1}$.

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: $\phi(w) = \int_{-\infty}^{\infty} f_X(x)\mathrm{e}^{\mathrm{i}wx}\, dx$.

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, $\phi(w)$: $f_X(x) = \frac{1}{2\pi}\int_{-\infty}^{\infty} \phi_X(w)\mathrm{e}^{-\mathrm{i}wx}\, dw$.

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: $\phi(w) = \sum_k P_X(x_k)\mathrm{e}^{\mathrm{i}wx_k}$.

En los casos en que los x_k sean enteros, $x_k = k$, que son la mayoría, la ecuación anterior es la **tranformada de Fourier de la secuencia** $P_X(k)$. Dicha función es una *función periódica* en w de periodo 2π ya que $\mathrm{e}^{\mathrm{i}(w+2\pi)k} = \mathrm{e}^{\mathrm{i}wk}$.

Por tanto, usando la fórmula de **inversión**, podemos escribir la **función de probabilidad** $P_X(k)$ como función de la función característica de X, $\phi(w)$: $P_X(k) = \frac{1}{2\pi}\int_0^{2\pi} \phi_X(w)\mathrm{e}^{-\mathrm{i}wk}\, dw$.

4.4.2.1. Ejemplos de cálculo de función característica

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:

$$P_X(0) = q = 1-p,\ p_X(1) = p.$$

Su función característica será:

$$\phi_X(w) = E\left(\mathrm{e}^{\mathrm{i}wX}\right) = p\mathrm{e}^{\mathrm{i}w\cdot 1} + (1-p)\mathrm{e}^{\mathrm{i}w\cdot 0} = p\mathrm{e}^{\mathrm{i}w} + (1-p) = 1 + p\left(\mathrm{e}^{\mathrm{i}w} - 1\right).$$

Comprobemos la fórmula de la inversión:

$$\begin{aligned}
P_X(1) &= \tfrac{1}{2\pi}\int_0^{2\pi}\left(1+p\left(\mathrm{e}^{\mathrm{i}w}-1\right)\right)e^{-\mathrm{i}w\cdot 1}\,dw = \tfrac{1}{2\pi}\left(\int_0^{2\pi}(1-p)e^{-\mathrm{i}w}\,dw + \int_0^{2\pi}p\,dw\right)\\
&= \tfrac{1}{2\pi}\left((1-p)\left[\tfrac{\mathrm{e}^{-\mathrm{i}w}}{-\mathrm{i}}\right]_0^{2\pi} + 2\pi p\right) = \tfrac{1}{2\pi}\left((1-p)\cdot 0 + 2\pi p\right) = p,\\
P_X(0) &= \tfrac{1}{2\pi}\int_0^{2\pi}\left(1+p\left(\mathrm{e}^{\mathrm{i}w}-1\right)\right)e^{-\mathrm{i}w\cdot 0}\,dw = \tfrac{1}{2\pi}\left(\int_0^{2\pi}(1-p)\,dw + \int_0^{2\pi}p\mathrm{e}^{\mathrm{i}w}\,dw\right)\\
&= \tfrac{1}{2\pi}\left((1-p)\cdot 2\pi + p\left[\tfrac{\mathrm{e}^{\mathrm{i}w}}{\mathrm{i}}\right]_0^{2\pi}\right) = \tfrac{1}{2\pi}\left((1-p)\cdot 2\pi + p\cdot 0\right) = 1-p.
\end{aligned}$$

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: $f_X(x) = \lambda\mathrm{e}^{-\lambda x}$, para $x \geq 0$ y 0, en caso contrario.

Su función característica será:

$$\phi_X(w) = E\left(\mathrm{e}^{\mathrm{i}wX}\right) = \int_0^\infty \mathrm{e}^{\mathrm{i}wx}\lambda\mathrm{e}^{-\lambda x}\,dx = \lambda\int_0^\infty \mathrm{e}^{(\mathrm{i}w-\lambda)x}\,dx = \lambda\left[\frac{\mathrm{e}^{(\mathrm{i}w-\lambda)x}}{\mathrm{i}w-\lambda}\right]_{x=0}^{x=\infty} = \frac{\lambda}{\lambda - \mathrm{i}w}.$$

La expresión anterior es válida para todo $w \in \mathbb{R}$ ya que su valor sería: $\phi_X(w) = \frac{\lambda}{\lambda-\mathrm{i}w}\cdot\frac{\lambda+\mathrm{i}w}{\lambda+\mathrm{i}w} = \frac{\lambda^2+\mathrm{i}\lambda w}{\lambda^2+w^2} = \frac{\lambda^2}{\lambda^2+w^2} + \mathrm{i}\frac{\lambda w}{\lambda^2+w^2}$. En la última expresión hemos separado la parte real de la imaginaria.

Calculemos la función de densidad a partir de la función característica:

$$f_X(x) = \frac{1}{2\pi}\int_{-\infty}^{\infty}\frac{\lambda}{\lambda-\mathrm{i}w}\mathrm{e}^{-\mathrm{i}wx}\,dw = a\mathrm{e}^{-ax},$$

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: $f_X(x) = \frac{1}{\sqrt{2\pi}\sigma}\mathrm{e}^{-\frac{(x-\mu)^2}{2\sigma^2}}$, para $x \in \mathbb{R}$.

Su función característica será:

$$\begin{aligned}
\phi_X(w) &= E\left(\mathrm{e}^{\mathrm{i}wX}\right) = \int_{-\infty}^{\infty}\mathrm{e}^{\mathrm{i}wx}\tfrac{1}{\sqrt{2\pi}\sigma}\mathrm{e}^{-\frac{(x-\mu)^2}{2\sigma^2}}\,dx = \tfrac{1}{\sqrt{2\pi}\sigma}\int_{-\infty}^{\infty}\mathrm{e}^{\mathrm{i}wx-\frac{(x-\mu)^2}{2\sigma^2}}\,dx\\
&= \tfrac{1}{\sqrt{2\pi}\sigma}\int_{-\infty}^{\infty}\mathrm{e}^{-\frac{1}{2\sigma^2}\left((x-(\sigma^2\mathrm{i}w+\mu))^2-2\sigma^2\mathrm{i}w\mu+\sigma^4w^2\right)}\,dx = \tfrac{1}{\sqrt{2\pi}\sigma}\mathrm{e}^{\frac{1}{2}(2\mathrm{i}w\mu-\sigma^2w^2)}\int_{-\infty}^{\infty}\mathrm{e}^{-\frac{1}{2\sigma^2}(x-(\sigma^2\mathrm{i}w+\mu))^2}\,dx\\
&= \mathrm{e}^{\frac{1}{2}(2\mathrm{i}w\mu-\sigma^2w^2)}\left(\tfrac{1}{\sqrt{2\pi}\sigma}\int_{-\infty}^{\infty}\mathrm{e}^{-\frac{1}{2\sigma^2}(x-(\sigma^2\mathrm{i}w+\mu))^2}\,dx\right) = \mathrm{e}^{\mathrm{i}w\mu-\frac{\sigma^2w^2}{2}}.
\end{aligned}$$

La integral del último paréntesis se resuelve haciento el cambio de variable $u = x - \sigma^2 \mathrm{i}w$ 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:

$$\begin{aligned} f_X(x) &= \tfrac{1}{2\pi}\int_{-\infty}^{\infty} \mathrm{e}^{\mathrm{i}w\mu - \frac{\sigma^2 w^2}{2}} \mathrm{e}^{-\mathrm{i}wx}\, dw = \tfrac{1}{2\pi}\int_{-\infty}^{\infty} \mathrm{e}^{\left(\frac{\mathrm{i}w\sigma}{\sqrt{2}} + \frac{\mu-x}{\sigma\sqrt{2}}\right)^2 - \frac{(\mu-x)^2}{2\sigma^2}}\, dw = \tfrac{1}{2\pi}\mathrm{e}^{-\frac{(\mu-x)^2}{2\sigma^2}}\int_{-\infty}^{\infty} \mathrm{e}^{\left(\frac{\mathrm{i}w\sigma}{\sqrt{2}} + \frac{\mu-x}{\sigma\sqrt{2}}\right)^2}\, dw \\ &= \tfrac{1}{2\pi}\mathrm{e}^{-\frac{(x-\mu)^2}{2\sigma^2}}\int_{-\infty}^{\infty} \mathrm{e}^{-\left(\frac{w\sigma}{\sqrt{2}} + \frac{\mu-x}{\mathrm{i}\sigma\sqrt{2}}\right)^2}\, dw \overset{\text{cambio de variable } u=\frac{w\sigma}{\sqrt{2}}+\frac{\mu-x}{\mathrm{i}\sigma\sqrt{2}}}{=} \tfrac{1}{2\pi}\mathrm{e}^{-\frac{(x-\mu)^2}{2\sigma^2}}\int_{-\infty}^{\infty} \tfrac{\sqrt{2}}{\sigma}\mathrm{e}^{-u^2}\, du \\ &\overset{\int_{-\infty}^{\infty} \mathrm{e}^{-u^2}\, du = \sqrt{\pi}}{=} \frac{1}{\sqrt{2\pi}\sigma}\mathrm{e}^{-\frac{(x-\mu)^2}{2\sigma^2}}\sqrt{\pi} = \frac{1}{\sqrt{2\pi}\sigma}\mathrm{e}^{-\frac{(x-\mu)^2}{2\sigma^2}}, \end{aligned}$$

función que coincide con la densidad de la distribución $N(\mu, \sigma)$.

4.4.2.2. 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:

$$m_n = E\left(X^n\right) = \frac{1}{\mathrm{i}^n}\frac{d}{dw^n}\phi_X(w)|_{w=0} = \frac{1}{\mathrm{i}^n}\phi_X^{(n)}(0).$$

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, $m_n = E\left(X^n\right) = p$.
- Si X es una variable exponencial de parámetro λ, demostrar usando la función característica que para todo n, $m_n = E\left(X^n\right) = \frac{n!}{\lambda^n}$.
- Si X es una variable normal de parámetros μ y σ, demostrar usando la función característica que $E(X) = \mu$, $E\left(X^2\right) = \mu^2 + \sigma^2$, $E\left(X^3\right) = 3\sigma^2\mu$ y $E\left(X^4\right) = 6\sigma^2\mu^2 + \mu^4 + 3\sigma^4$.

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:

$$\begin{aligned} E(T) &= \int_{t=0}^{t=\infty} 1 - F_T(t)\, dt = \int_{t=0}^{t=\infty} \int_{u=t}^{u=\infty} f_T(u)\, du\, dt \\ &= \int_{u=0}^{u=\infty} f_T(u) \int_{t=0}^{t=u} dt\, du = \int_{u=0}^{u=\infty} f_T(u) \cdot u\, du = E(T), \end{aligned}$$

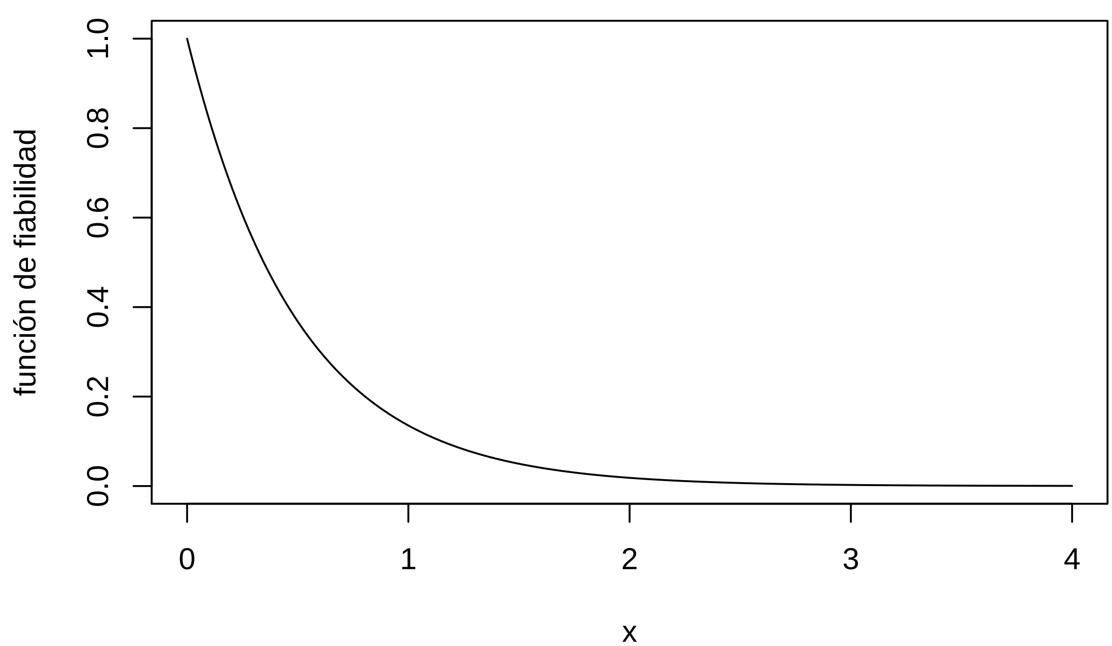
donde $f_T(u)$ seria la función de densidad de la variable T en el valor u.

4.5.1.1. Ejemplo

Ejemplo

Sea T una variable aleatoria exponencial de parámetro λ.

La fiabilidad de T sería: $R(t) = P(T > t) = 1 - F_T(t) = \mathrm{e}^{-\lambda t}$:



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 $y \in [0,1]$:

$$\begin{aligned} F_Y(y) &= P(Y \leq y) = P(F_X(X) \leq y) \overset{\text{usando que } F_X \text{ es estrictamente creciente}}{=} P(X \leq F_X^{-1}(y)) \\ &= F_X(F_X^{-1}(y)) = y. \end{aligned}$$

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.

4.6.2. Ejemplos

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: $F_X(x) = 1 - \mathrm{e}^{-\lambda x}$.

Hallemos a continuación F_X^{-1}:

$$y = 1 - \mathrm{e}^{-\lambda x}, \Leftrightarrow 1 - y = \mathrm{e}^{-\lambda x}, \Leftrightarrow \ln(1-y) = -\lambda x, \Leftrightarrow x = -\frac{1}{\lambda}\ln(1-y).$$

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
```

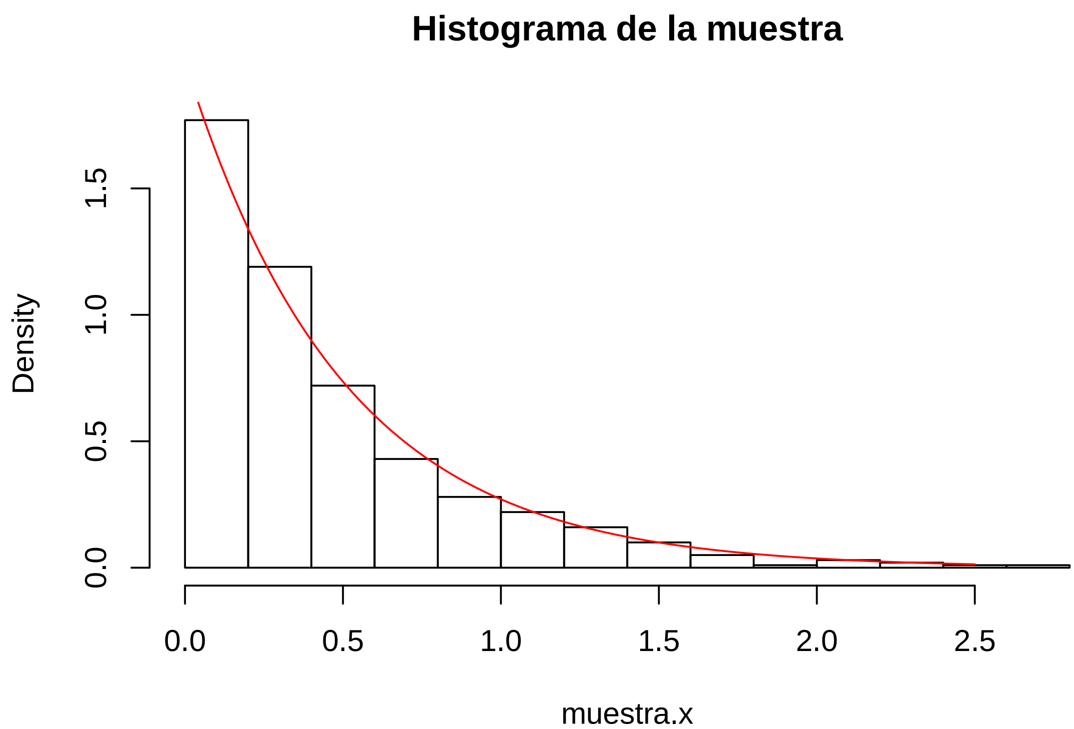
```
##  [1] 0.35877893 0.25280668 0.11355513 0.71885498 1.19013394 0.49444154
##  [7] 0.23443983 0.07049329 1.48215515 0.09206527 0.14783308 2.35604428
## [13] 0.41728297 0.43470135 0.02910931 0.36337987 0.08532398 0.27981457
## [19] 0.40904272 0.02482599 0.25061234 1.67562739 0.07212948 0.27227039
## [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")
```



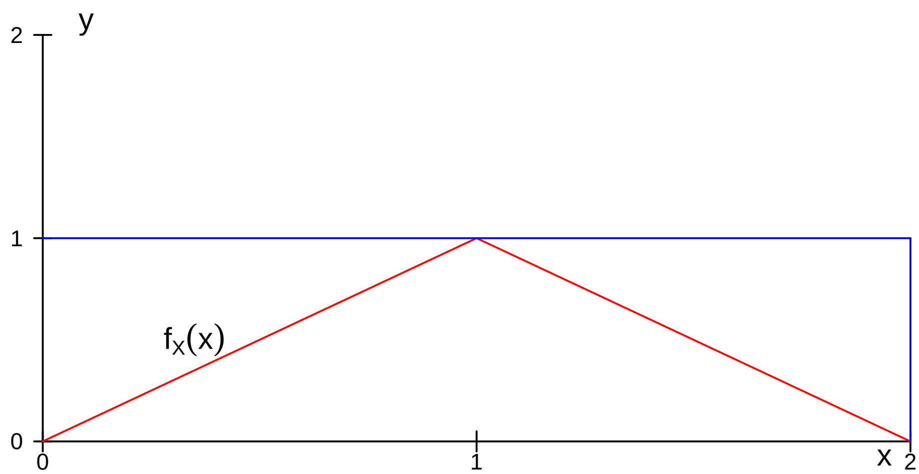
4.6.3. Método de rechazo

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.

4.6.3.1. Ejemplos

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**:

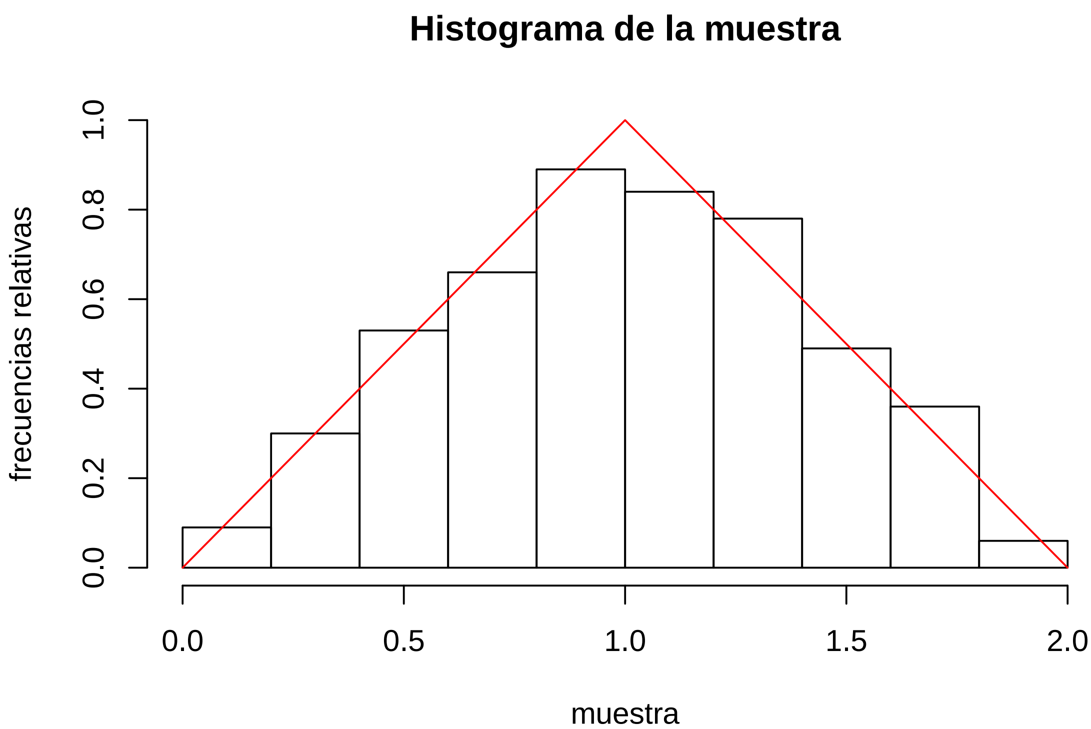
```
a=0; b=2; c=0; d=1; n=25; i=1;
f = function(x){ifelse(x>=0 & x<=1,x,ifelse(x>=1&x<=2,2-x,0))}
muestra=c()
```

```
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")
```



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: $D_X = \{1, 2, \ldots, N\}$.

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: $I(A_k) = I(\{X = k\}) = \ln\left(\frac{1}{P(X=k)}\right) = -\ln\left(P(X = k)\right)$.

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: $H_X = E\left(-\ln(f_X)\right) = \int_{-\infty}^{\infty} -\ln(f_X(x)) f_X(x)\, dx$, en el caso continuo y, $H_X = E\left(-\ln(P_X)\right) = \sum_{x_k \in D_X} -\ln(P_X(x_k)) P_X(x_k)$, en el caso discreto.

4.7.1.1. Ejemplos

Ejemplo: entropía de una variable de Bernoulli de parámetro p

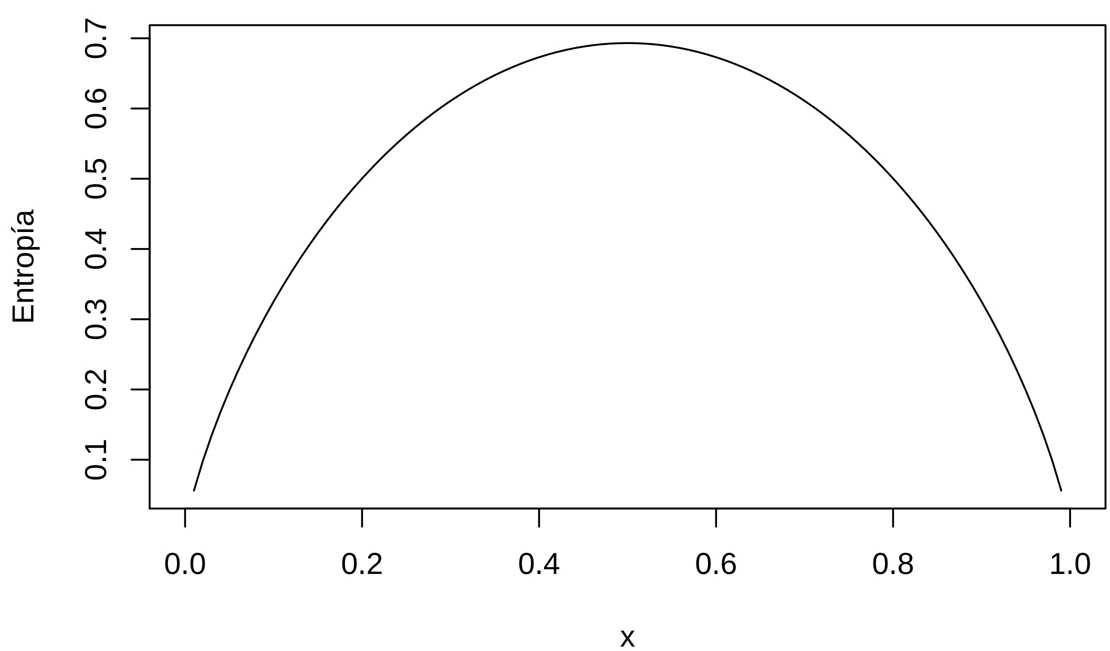
Sea X una variable de Bernoulli de parámetro p.

Recordemos que su función de probabilidad P_X es: $P_X(0) = 1 - p = q$, $P_X(1) = p$.

La entropía de X será:

$$H_X = E\left(-\ln(P_X)\right) = -(1-p) \cdot \ln(1-p) - p \cdot \ln p.$$

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

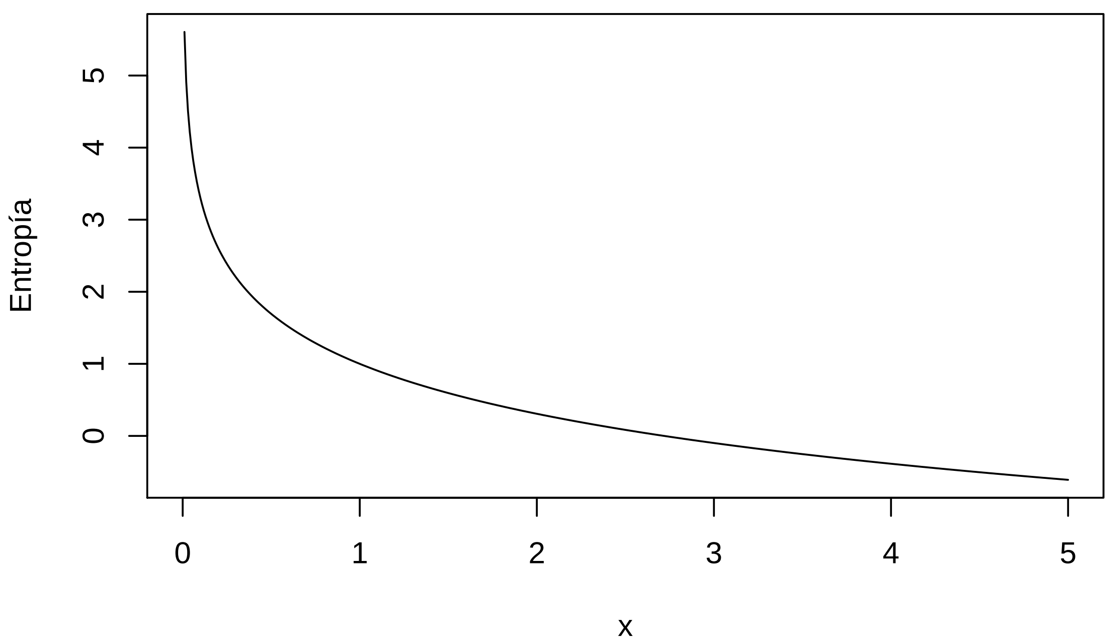
Sea X una variable aleatoria exponencial de parámetro λ.

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:

$$\begin{aligned} H_X &= E\left(-\ln(f_X)\right) = -\int_0^\infty \ln\left(\lambda e^{-\lambda x}\right)\lambda e^{-\lambda x}\,dx = -\lambda\int_0^\infty (\ln(\lambda) - \lambda x)e^{-\lambda x}\,dx \\ &= -\ln(\lambda)\int_0^\infty \lambda e^{-\lambda x}\,dx + \lambda\int_0^\infty \lambda x e^{-\lambda x}\,dx = -\ln(\lambda)\int_0^\infty f_X(x)\,dx + \lambda E(X) \\ &= -\ln(\lambda) + \lambda\frac{1}{\lambda} = 1 - \ln(\lambda). \end{aligned}$$

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 sería 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 sería 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.