ApLabAD

Apuntes de Laboratorio de Análisis de Datos

RICUIB

2025-02-01

Tabla de contenidos

Prefacio

Este libro en la web es una versión de las notas de clase de asignaturas introductorias al análisis de datos.

Ha sido elaborado con Quarto RStudio, PBC. (2022). Quarto (Version 1.0). Hemos utilizado el formato formato book.

¿Quarto book o bookdown? https://yihui.org/en/2022/04/quarto-r-markdown/

Parte I

Parte 1: Probabilidad y variables aleatorias

En esta sección, veremos la teoría básica de la probabilidad y de las variables aleatorias. Resolveremos problemas prácticos para comprender mejor los conceptos y utilizaremos R para realizar cálculos y gráficos.

También exploraremos los modelos de probabilidad discretos y continuos más conocidos.

En ocasiones, trabajaremos con problemas de cálculo más complejos.

Además, introduciremos problemas sencillos de modelización con probabilidades. Estos consistirán en un enunciado, real o inventado, en el que se pedirá modelizar el problema mediante una variable aleatoria y responder a una serie de preguntas.

1 Preliminares: 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).

1.1 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 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 compleja y presenta varias paradojas como la paradoja de Russell.

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

Estas colecciones o conjuntos se pueden definir por:

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

1.1.1 Conjuntos básicos

Los conjuntos suelen tener un conjunto madre como por ejemplo

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

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

1.1.2 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 B 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 $\emptyset \subseteq A$.

Ejemplo

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

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

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

Cardinal de las partes de un conjunto

Propiedad

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 $\omega \in \Omega$

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

1.1.3 Operaciones entre conjuntos

Intersección de conjuntos

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

El conjunto **intersección** de A y B es el formado por todos los elementos que 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 de conjuntos

El conjunto **unión** de A y B es el formado por todos los elementos que perteneces a A **O** 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 de conjuntos

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

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

1.1.4 Más propiedades

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.
- $\left(A^c\right)^c = A$ doble complementario.
- $\left(A \cup B\right)^c = A^c \cap B^c$, $\left(A \cap B\right)^c = A^c \cup B^c$ leyes de De Morgan.

1.1.5 Con R, ejemplos.

Con R los conjuntos de pueden definir como vectores

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

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

```r
A
```

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

```r
B
```

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

```r
C
```

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

$A \cap B$

```r
A
```

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

```r
B
```

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

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

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

$A \cup B$

```r
A
```

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

```r
B
```

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

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

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

$B - C$

```r
B
```

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

```r
C
```

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

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

```
[1] 1 5
```

$A^c = \Omega - A$

```r
Omega
```

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

```r
A
```

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

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

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

1.1.6 Con python

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

```
{1, 2, 3, 4, 5, 6, 7, 8, 9, 10}
```

```python
A
```

```
{1, 2, 3, 4, 5}
```

```python
B
```

```
{1, 4, 5}
```

```python
C
```

```
{8, 4, 6, 7}
```

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

```
{1, 4, 5}
```

```python
A | B   # unión (|: or/o)
```

```
{1, 2, 3, 4, 5}
```

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

```
{1, 2, 3, 5}
```

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

```
{1, 2, 3, 5, 9, 10}
```

1.2 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?

1.2.1 Número Binonial

Número combinatorio o número binomial

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

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

Recordemos que

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

Ejercicio: el paquete `gtools`

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

1.2.2 Combinaciones con repetición

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

Combinaciones con repetición

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

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

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

Ejemplo: caramelos

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

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

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

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

$$CR_{n=4}^{k=12} = \binom{4+12-1}{12} = 455.$$

1.2.3 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, estas son las variaciones de orden $k = 2$ de un conjunto de $n = 3$ elementos.

Variaciones

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

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

Ejemplo

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

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

Ejercicio

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

1.2.4 Variaciones con repetición.

¿Y si en el caso anterior permitimos que se repita algún dígito?

Variaciones con repetición

Las variaciones de orden k de un conjunto de n elementos permitiendo que se repitan los elementos. Las denotamos y valen:

$$VR_n^k = n^k$$

Ejemplo

Efectivamente en nuestro caso

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

$$VR_{n=3}^{k=2} = n^k = 3^2 = 9.$$

1.2.5 Permutaciones

Permutaciones

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

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

Ejemplo: dígitos

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

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

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

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

Ejercicio

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

Investiga también 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 $x = \{1, 2, 3, \ldots, 10\}$.

1.2.6 Números multinomiales. Permutaciones con repetición.

Ejercicio

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

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

$$PR_n^{n_1,n_2,\ldots,n_k} = \binom{n}{n_1 \quad n_2 \quad \ldots \quad n_k} = \frac{n!}{n_1! \cdot n_2! \cdot \ldots \cdot n_k!},$$

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

Ejercicio

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

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

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

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

1.3 Para acabar

1.3.1 Principios básicos para contar cardinales de conjuntos

El principio de la suma

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

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

Principio de unión exclusión

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

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

El principio del producto

Sean $A_1, A_2, \dots A_n$

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

1.3.2 Otros aspectos a tener en cuenta

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

- Propiedades de los números combinatorios.
- Binomio de Newton.
- Multinomio de Newton.

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

2 Teoría elemental de la probabilidad

2.1 Definiciones básicas

Definición: experimento aleatorio

Un experimento que repetido en las mismas condiciones puede dar resultados diferentes, pero que a largo plazo son predecibles recibe el nombre de **experimento aleatorio**.

Daremos nombres a distintos tipos de sucesos:

Definición: espacio muestral y tipos de sucesos

- Llamaremos **suceso elemental** a cada uno de los posibles resultados del experimento aleatorio.
- Llamaremos **espacio muestral** (Ω,E) al conjunto formado por todos los sucesos elementales del experimento aleatorio.
- Llamaremos **suceso** a cualquier subconjunto del espacio muestral.
- **Suceso seguro o cierto** $A \subseteq \Omega$
- **Suceso imposible o vació**: $\emptyset$
- **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)

A continuación describimos el clásico experimento del lanzamiento de un dado.

Ejemplo

Consideremos el experimento aleatorio que consiste en lanzar un dado. El espacio muestral de este experimento es $\Omega = \{1, 2, 3, 4, 5, 6\}$ o las figuras de las caras del dado.

Si lo representamos gráficamente, tendríamos:

Por comodidad y conveniencia se opta por representar el espacio muestral por

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

Recordemos la notación $\mathcal{P}(\Omega)$ que usamos para referirnos al conjunto de todos los subconjuntos de Ω. Este conjunto se llama **conjunto de partes** de Ω.

Ejercicio

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

Veamos algún ejemplo menos clásico. Podemos considerar el experimento aleatorio que consiste en calcular los n gramas de una palabra escogida al azar.

Ejemplo n-gramas

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: $\{Bal, ale, lea\}$.

Existen bases de datos que estudian la frecuencias de n-gramas de caracteres en textos en diferentes idiomas; generalmente de palabras. Por ejemplo, en español, los bi-gramas de sílabas más frecuentes son "EN" (3.01%) y "DE" (2.77%) y los tri-gramas de sílabas son "QUE" (1.66%) y "ENT" (1.38%). Podéis consultar más estadísticas, por ejemplo, en Stefan Trost Media frecuencias de sílabas en español.

2.1.1 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

Dos sucesos cualesquiera A y B son *incompatibles* (o *disjuntos*) cuando $A \cap B = \emptyset$.

Otro ejemplo se observa el sexo y la lateralidad de los estudiantes de una clase.

Ejemplo

Supongamos que el sexo se divide entre Mujeres y Hombres y la lateralirad en diestros y zurdos. 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 sucesos anteriores serían:

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

2.1.2 Propiedades

Propiedades

Conmutativas:

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

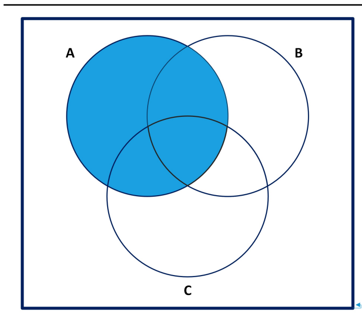
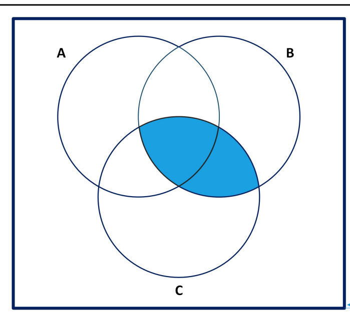
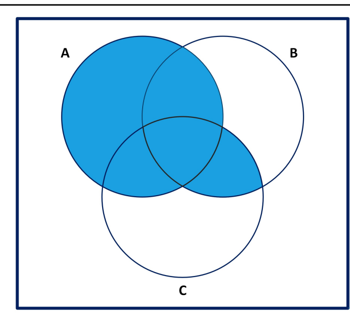
Asociativas:

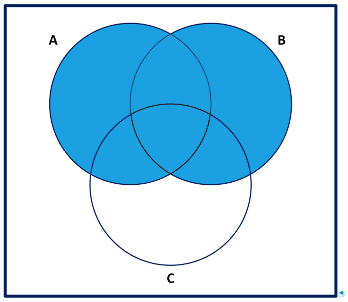
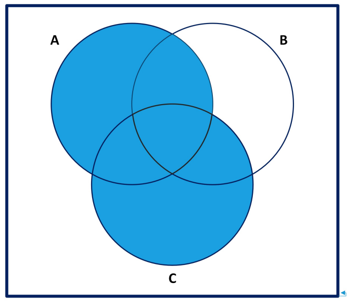
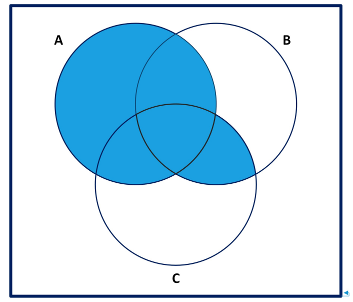
$$\begin{aligned} A \cup (B \cup C) = (A \cup B) \cup C, \\ A \cap (B \cap C) = (A \cap B) \cap C. \end{aligned}$$

Distributivas

$$\begin{aligned} A \cap (B \cup C) = (A \cap B) \cup (A \cap C), \\ A \cup (B \cap C) = (A \cup B) \cap (A \cup C). \end{aligned}$$

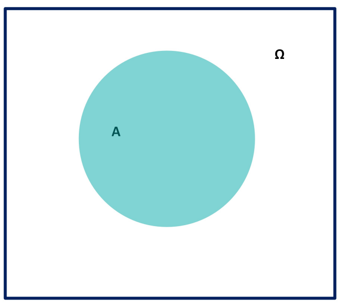
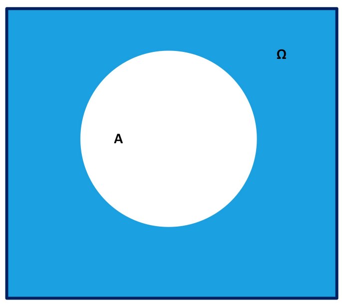
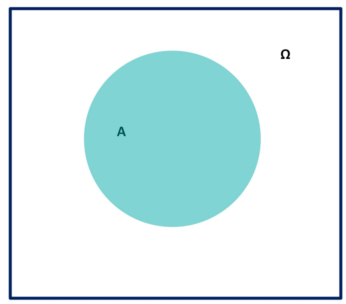
Veamos algunos diagramas que nos ayuda a demostrar las propiedades anteriores.

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

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

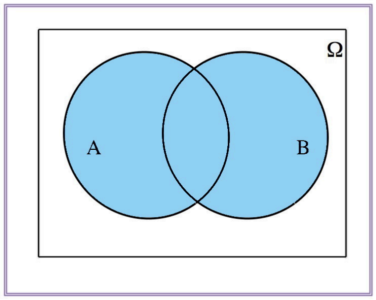
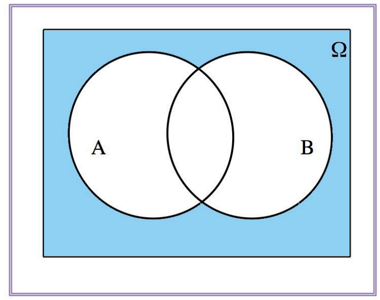
Complementario del complementario

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

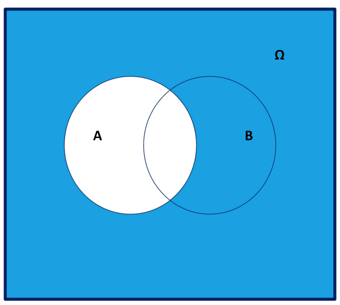
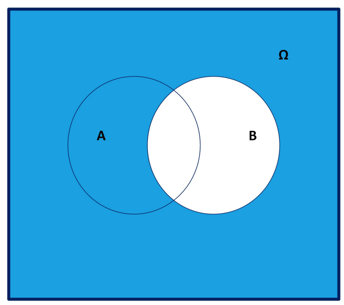
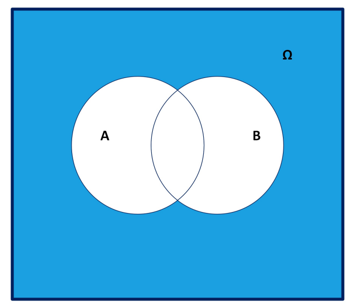
A	A^c	$(A^c)^c$
		

Leyes de De Morgan

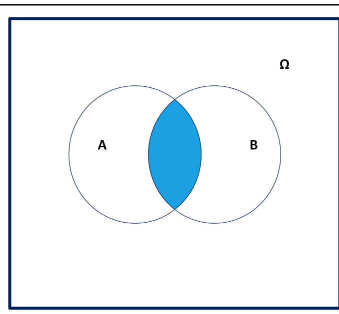
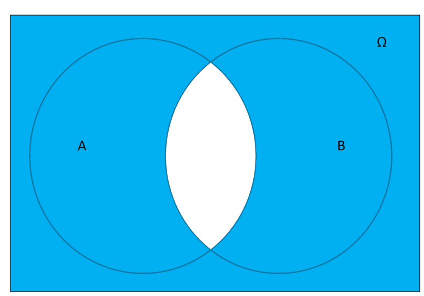
$$(A \cup B)^c = A^c \cap B^c$$

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

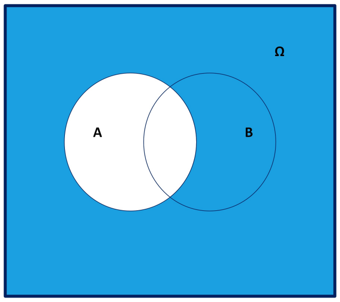
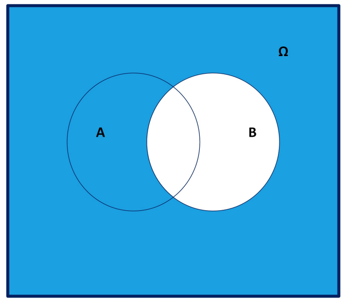
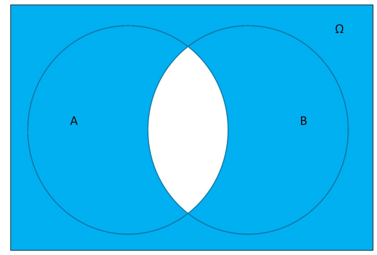
$$(A \cup B)^c = A^c \cap B^c$$

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

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

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

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

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

2.2 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

Definición formal de probabilidad

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

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

1. $0 \leq P(A) \leq 1$, para todo suceso A.
2. $P(\Omega) = 1$.
3. Si $\{A_1, A_2, \ldots, A_n\}$ son sucesos disjuntos dos a dos, entonces

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

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

Veamos un ejemplo real de cómo se calcula la probabilidad de un suceso.

Ejemplo

En la página de la Fundación Banco de Sangre y Tejidos de las Islas Baleares (17-08-2023) 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.

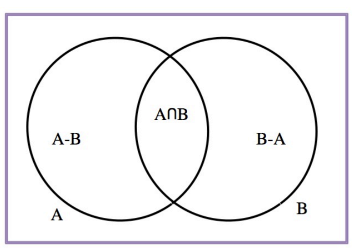
Suceso: $\{\text{O}\}^c = \{\text{A,B,AB}\}$.

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

Necesitaremos tener propiedades y fórmulas prácticas para poder calcular probabilidades de sucesos más complejos. Veamos algunas de ellas.

Propiedades básicas de la probabilidad

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



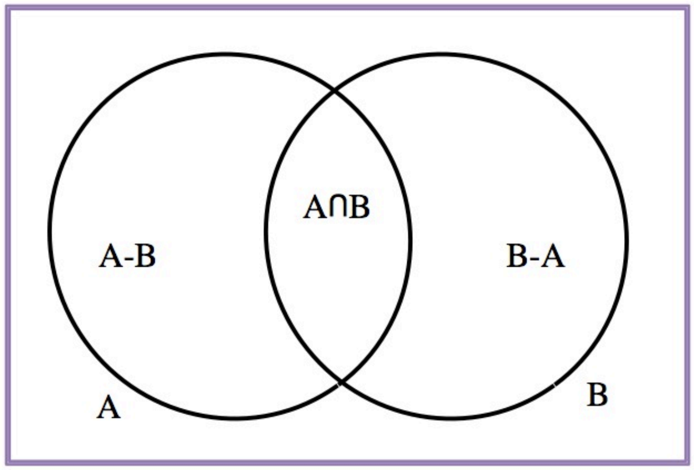
- Si $B \subseteq A$, entonces $0 \leq P(B) \leq P(A)$.

- $P(A^c) = 1 - P(A)$.

Una identidad muy utilizada es la de la probabilidad de la unión de dos sucesos cualesquiera.

La Probabilidad de la unión de dos sucesos

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



La demostración analítica de esta propiedad es la siguiente:

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

Probabilidad de la unión de n conjuntos

Sean $A_1, A_2, \dots A_n$ sucesos. Entonces:

$$P(\cup_{i=1}^{n} A_i) = \sum_{i=1}^{n} P(A_i) - \sum_{1 \leq i < j \leq n} P(A_i \cap A_j) + \cdots + (-1)^{n-1} P(A_1 \cap A_2 \cap \cdots \cap A_n).$$

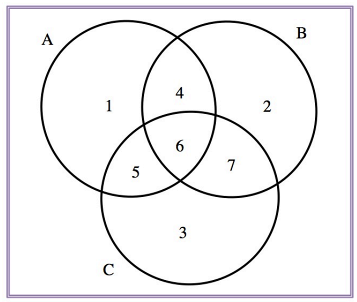
La demostración es sencilla mediante inducción: partimos del caso base de dos sucesos, suponemos que es cierta para n sucesos y luego la extendemos al caso de $n + 1$ sucesos.

Como comprobación, consideremos un ejemplo genérico con tres sucesos.

$A = \{1, 4, 5, 6\}$, $B = \{2, 4, 6, 7\}$ y $C = \{3, 5, 6, 7\}$. En este caso la fórmula nos da

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

Gráficamente tenemos esta situación:



Ahora podemos comprobar la fórmula para este caso.

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

Efectivamente tenemos que:

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

Una de las formas más intuitiva de asignación de probabilidades es hacer el cociente entre los casos favorables a que acontezca el evento y los casos posibles del experimento; la llamada fórmula de Laplace.

Propiedad

- En general dado un suceso $A = \{a_1, a_2, \dots, a_k\}$, entonces

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

- **Fórmula de Laplace**: Si todos los sucesos elementales tienen la misma probabilidad,

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

En el procesamiento del lenguaje se suelen estudiar las frecuencias de palabras o letras de un determinado idioma. Veamos un ejemplo sobre las frecuencias de las vocales en castellano.

< Ejemplo: Frecuencia de vocales/i>

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

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

Otro ejemplo en este caso es sobre un test de drogas en el que se analiza la presencia de cocaína y cannabis en la sangre de los conductores, inspirado en un caso real.

Ejemplo: Consumo de drogas

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

Pregunta: ¿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,por tanto:

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

Pregunta: ¿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:

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

Pregunta: ¿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; $P(A \cap B) = 0.0005$.
- $A - B$: dar positivo en cocaína pero no en cannabis, por lo tanto tenemos que :

$$P(A - B) = P(A) - P(A \cap B) = 0.001 - 0.0005 = 0.0005.$$

2.3 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: Probabilidad condicionada

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?

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

¡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.

2.3.1 Probabilidad condicionada. Propiedades

La probabilidad condicionada es una probabilidad, en el sentido de la siguiente propiedad.

Propiedad

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:

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

También se cumplirán el resto de propiedades mientras se condicionen cada una de las probabilidades al mismo suceso A.

Ejercicio

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

Veamos un ejemplo donde se aplica la probabilidad condicionada, en este caso, para calcular la probabilidad de que un adulto sea hipertenso, dado que cree que lo es.

Ejemplo

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

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

Sean los sucesos

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

Ahora 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:

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

Otra pregunta 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

Otro ejemplo de probabilidad condicionada en este caso un ejemplo simple de dígito de control de error.

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:

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

La probabilidad condicional también es útil en la resolución de problemas de clasificación, como el siguiente ejemplo.

Ejemplo: SPAM

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

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

Asignemos sucesos y probabilidades

- 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; $P((A \cup S)^c) = 0.15$,

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

- ¿Cuál es la probabilidad que un correo lleve adjunto si es SPAM?
- $P(A) = 0.5, P(S) = 0.65, P(A^c \cap S^c) = P((A \cup S)^c) = 0.15,$
- $P(A \cup S) = 1 - P((A \cup S)^c) = 0.85,$
- $P(A \cap S) = P(A) + P(S) - P(A \cup S) = 0.3,$

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

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

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

2.4 Teorema de la probabilidad total

Teorema de la probabilidad total

Dados dos sucesos A y B se tiene que

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

Vamos a generalizar el resultado anterior a una colección de sucesos $A_1, A_2, \dots, A_n$ que forman una partición del espacio muestral Ω.

Partición del espacio espacio muestral

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

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

Ahora podemos volver a enunciar el teorema anterior pero en esta ocasión para particiones arbitrarias.

Teorema de la probabilidad total generalizado

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

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

Revisitemos el ejemplo de los mensajes con dígitos de control de error.

Ejemplo

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

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

Sean los sucesos del enunciado:

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

entonces obtenemos las probabilidades a partir del enunciado:

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

y por tanto,

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

2.5 Clasificación o diagnostico caso binario

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).

SPAM continuación

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.

Falsos Positivos y Falsos Negativos

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.

Veamos el siguiente ejemplo:

Falsos Positivos y Negativos

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.

Necesitamos algunas denominaciones adicionales:

Falsos Positivos y Negativos

- **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**: $P(T^c|M)$.

Falsos Positivos y Negativos

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

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

3 Teorema de Bayes

Teorema de Bayes para dos sucesos

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

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

Ejemplo

Demostrar el teorema de Bayes utilizando que

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

Generalicemos este resultado para una partición arbitraria del espacio muestral.

Teorema de Bayes para una partición

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

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

Podéis demostrar el teorema de Bayes utilizando que

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

Test de 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?

Los sucesos del ejemplo son:

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

de donde podemos calcular:

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

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

Los sucesos del ejemplo son:

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

de donde podemos calcular:

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

Ejemplo: Tipos de clientes

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 y

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

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

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

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

Ejemplo: Fidelización de clientes

Para fidelizar a sus clientes una empresa implementa 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?

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

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

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

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

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

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

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 $P(T^c|B) = 1 - P(T|B) = 0.025$ y $P(T^c|B^c) = 1 - P(T|B^c) = 0.88$.

4 Independencia de sucesos

Sucesos Independientes

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

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

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

Propiedad

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. $P(B|A) = P(B)$.
4. A^c y B son independientes.
5. A y B^c son independientes.
6. A^c y B^c son independientes.

Veamos un sencillo ejemplo de compras de billetes de avión y alojamiento en hotel.

Ejemplo 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

$P(A \cap B) = P(A) + P(B) - P(A \cup B) = 0.55 + 0.2 - 0.6 = 0.15$ y $P(A) \cdot P(B) = 0.55 \cdot 0.2 = 0.11$.

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

4.1 Sucesos independientes vs disjuntos

sucesos disjuntos e independencia

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?

5 Variables aleatorias

5.1 Introducción

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.

5.2 Definición de variable aleatoria

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

Variable Aleatoria (definición práctica)

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

Notación

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

Ejemplo

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

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

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

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

- 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 $\{X \leq 3\}$.

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

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

5.3 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.

5.4 Ejemplo

Ejemplo

Son variables *aleatorias discretas*:

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

Son variables *aleatorias continuas*:

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

5.5 Variables aleatorias 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.

5.6 Distribuciones de probabilidad discretas

< Función de Probabilidad

La **función de probabilidad** (*probability mass function* o incluso abusando de notación *probability density function*) de una variable aleatoria discreta X a la que denotaremos por $P_X(x)$ está definida por

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

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

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

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: Dado de parchís

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

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

y la 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

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

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

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:

$$X(\omega) = \left\{ \begin{array}{ll} 1 & \text{si } \omega = c \\ 0 & \text{si } \omega = + \end{array} \right.$$

Su función de probabilidad es:

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

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:

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

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

5.7 Propiedades de la función de probabilidad.

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

Ejemplo: lanzmiento moneda

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

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

(expresados en orden de aparición).

Este espacio tiene todos los sucesos elementales equiprobables.

Consideremos la variable aleatoria asociada a este experimento:

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

Su función de probabilidad es:

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

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

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

Efectivamente los valores de la función de distribución suman 1:

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

5.8 Función de distribución de variables aleatorias

Función de distribución de Probabilidad (acumuladada)

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, si X se encuentra entre dos valores a y b $\{a < X \leq b\} = \{X \leq b\} - \{X \leq a\}$. Ahora podemos hacer

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

Lo que finaliza la demistración de la propiedad.

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 $F_X(x_0^-) = \lim\limits_{x \to x_0^-} F(x)$, 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 $\lim\limits_{x \to \infty} F_X(x) = 1$; $\lim\limits_{x \to -\infty} F_X(x) = 0$.
- Toda función F verificando las propiedades anteriores es función de distribución de alguna v.a. X.

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 $F_X(x_0^-) = \lim\limits_{x \to x_0^-} F_X(x)$,, entonces se cumplen las siguientes igualdades:

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

Más propiedades de la función de distribución

- Si F_X es continua en x se tiene que $P(X = x) = 0$. Así que si la v.a. es continua $P(X \leq a) = P(X < a) + P(X = a) = P(X < a)$ y propiedades similares.

- Sea X una variable aleatoria discreta que con dominio D_X y que tiene por función de probabilidad $P_X(x)$ entonces su función de distribución $F_X(x_0)$ es

$$F_X(x_0) = \sum_{x \leq x_0} P_X(x),$$

donde $\sum\limits_{x \leq x_0}$ 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

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

Lo que demuestra estas dos propiedades.

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 tambiñen podemos prodeder así:

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

Propiedades

Sea X una variable con función de distribución F_X entonces:

- $0 \leq F_X(x) \leq 1$ para todo x,
- Si $x < x'$, entonces
$$F_X(x) \leq F_X(x').$$
Es una función creciente,es decir, 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 $\lim\limits_{x \to x_0^+} F_X(x) = F_X(x_0)$.

5.9 Momentos de variables aleatorias discretas

Al igual que en la estadística descriptiva se utilizan distintas medidas para resumir los valores centrales y para medir la dispersión de una muestra, podemos definir las correspondiente medidas para variables aleatorias.

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

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

5.9.1 Esperanza de un variable aleatoria discreta

Definición: 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

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

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

Ejemplo: Intepretación de de la media

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

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

Resulta que

$$\begin{aligned} P(X = 0) = 0.42,\ P(X = 1) = 0.4,\ P(X = 2) = 0.18,\ \text{por lo tanto,} \\ E(X) = 0 \cdot 0.42 + 1 \cdot 0.4 + 2 \cdot 0.18 = 0.76. \end{aligned}$$

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?

Propiedad: 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

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

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

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

5.9.2 Series geométricas

Series geométricas

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

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

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

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

Propiedades

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

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

.

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

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

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

$$\left(\sum_{k=0}^{+\infty} r^k\right)' = \sum_{k=1}^{+\infty} k r^{k-1}; \qquad \left(\frac{1}{1-r}\right)' = \frac{1}{(1-r)^2}$$
$$\left(\sum_{k=0}^{+\infty} r^k\right)'' = \sum_{k=2}^{+\infty} k(k-1) r^{k-2}; \qquad \left(\frac{1}{1-r}\right)'' = \frac{2}{(1-r)^3}$$

.

Ejemplo: paleta de colores (continuación)

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

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

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

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

Calculad el valor esperado de la variable

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

5.9.3 Momentos de una variable aleatoria. Varianza

Definición: 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.

5.9.4 Medidas de la variabilidad

Definición: 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

Demostración de b)

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

Se deja como ejercicio la primera afirmación.

Ejemplo: número de errores (continuación)

Calculemos en el ejemplo del contero de errores la varianza de estos.

Recordemos que:

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

y que

$$E(X)=0.76.$$

Entonces:

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

Ahora necesitamos calcular

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

y por lo tanto

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

y

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

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

5.9.5 Propiedades de la varianza

Propiedades de la varianza

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

Ejercicio

Se deja como ejercicio la demostración de estas propiedades.

5.10 Transformaiones lineales de v.a.

Transformación lineal

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

Propiedad: 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 $Y = a + b \cdot X$:

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

Demostración:

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

Lo que demuestra esta propiedad, las emás se dejan comom ejercicio.

5.11 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: Distribución uniforme en $[0, 1]$

Ejemplo: distancia dardo 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 que cualquier punto de la diana sea *equiprobable*).

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

Su función de distribución es

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

consideremos

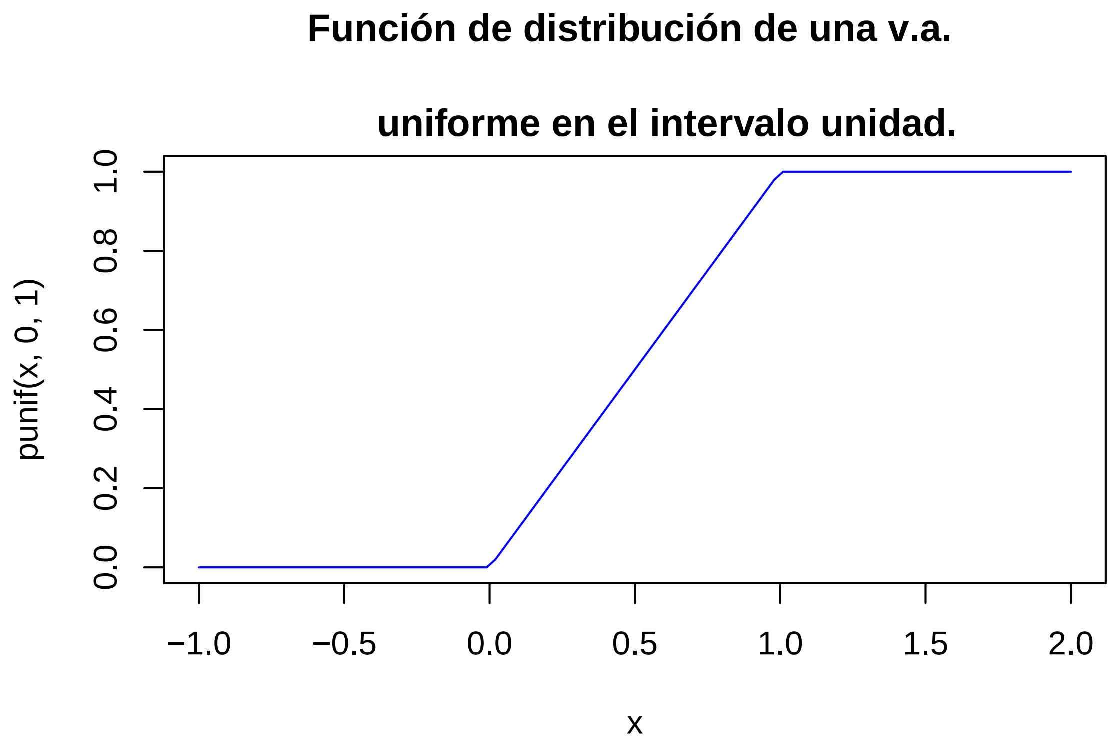
- C.F. *longitud favorable* que es $x - 0$,
- C.P. *longitud posible* que es $1 - 0$,

luego

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

el siguente código grafica la función de distribución uniforme

```
curve(punif(x,0,1),xlim=c(-1,2),col="blue",
      main="Función de distribución de una v.a. \n
      uniforme en el intervalo unidad.")
```



Propiedades

En las variables continuas los sucesos del tipo $\{X \leq x\}$ y $\{X < x\}$ tendrán la misma probabilidad. Otras identidades similares son :

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

Demostración:

Algunas identidades son evidentes $P(X \leq b) = P(X < b) + P(X = b) = P(X < b)$.

Para otras, como la siguiente, podemos hacer

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

entonces

\begin{eqalign} *P(X< b) = & P({X≤a}∪{a<X<b}) = P(X≤a)+P(a<X<b)= \ & P(X< a)+P(a<X<b).$ \end{align}*

La demostración de las otras propiedades las 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:

Dada una variable aleatoria continua se tiene que:

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

Se deja la demostración como ejercicio.

Ejemplo

Ejemplo: diana (continuación)

En el ejemplo de la diana:

$$P(0.25 < X < 0.3) = F_X(0.3) - F_X(0.25) = 0.3 - 0.25 = 0.05.$$

Definición: Función de densidad

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

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

Propiedad: Relación entre la función de distribución y la densidad

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.

Además el los valores de F_X son continuos y derivables en los puntos donde f_X es continua y la derivada de la función de distribución es una densidad $F_x'(x) = f_X(x)$.

Definición: Dominio de una variable aleatoria continua

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

Ejemplo: diana (continuación)

En nuestra ejemplo de la diana, la función f es una densidad

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

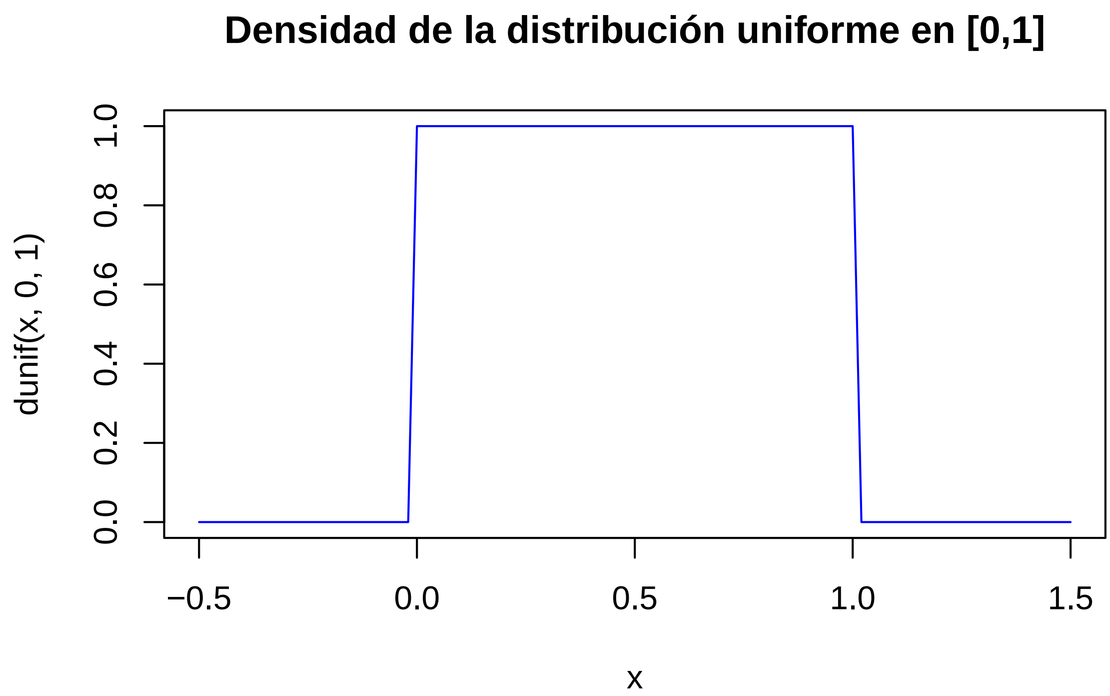
que es la densidad de X, en efecto:

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

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

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

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



Propiedades

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

$$P(a < X < b) = P(a < X \leq b) = P(a \leq X < b) =$$
$$P(a \leq X \leq b) = \int_a^b f_X(x)dx.$$

- Si A es un subconjunto adecuado 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

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

Luego su función de distribución es:

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

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

Ejercicio: Tiempo de un proceso

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

5.12 Esperanza y varianza para variables aleatorias continuas

Algunas de estas propiedades ya han sido estudiadas en el caso de variables aleatorias discretas. Por ello, en esta sección nos centraremos en presentar sus definiciones, métodos de cálculo y algunos ejemplos en el contexto continuo.

A partir de ahora, salvo indicación en contrario, consideraremos que X es una variable aleatoria continua con función de densidad $f_X(x)$

Definición: Esperanza y Varianza 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**

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

- Su desviación típica es:

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

Propiedades

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

Ejemplo: Diana (continuación)

$$\mu_X = \int_0^1 xdx = \left[\frac{x}{2}\right]_0^1 = \frac{1}{2},$$

$$E(X^2) = \int_0^1 x^2dx = \left[\frac{x^3}{3}\right]_0^1 = \frac{1}{3},$$

$$Var(X) = E(X^2) - E(X)^2 = \frac{1}{3} - \left(\frac{1}{2}\right)^2 = \frac{1}{12}.$$

Podemos comprobar que con la definición directa el resultado es el mismo

$$\begin{aligned} Var(X) =& E\left((X - E(X))^2\right) = \int_0^1 \left(x - \frac{1}{2}\right)^2 dx = \\ & \left[\frac{1}{3}\left(x - \frac{1}{2}\right)^3\right]_0^1 = \frac{1}{3} \cdot \left(\left(1 - \frac{1}{2}\right)^3 - \left(0 - \frac{1}{2}\right)^3\right) = \frac{1}{3}\left(\frac{1}{8} - \left(-\frac{1}{8}\right)\right) = \frac{1}{3} \cdot \frac{2}{8} = \frac{1}{12}. \end{aligned}$$

5.13 Esperanza de transformaciones lineales de v.a. continuas

Propiedades

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

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

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

Ejemplo

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

Entonces la esperanza del beneficio es

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

y

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

5.14 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.

Propiedad: Transformaciones 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 el anterior, pues la transformación de, por ejemplo, una v.a. continua puede ser continua, discreta, mixta,…

Propiedad: 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:

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

5.15 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,

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

5.16 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.

5.16.1 Desigualdad de Markov

Propiedad: 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

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

5.16.2 Desigualdad de Markov

Propiedad: Desigualdad de Markov

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

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

Se deja como **ejercicio** la prueba del corolario anterior a partir de la desigualdad de Markov.

5.17 Desigualdad de Chebychev

Propiedad: Desigualdad de Chebychev

La **desigualdad de Chebychev** también se escribe 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.

5.17.1 Uso de la desigualdad de Chebychev

Utilidad básica de la desigualdad de Chebychev

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

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

lo que implica que

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

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

Ejemplo: tiempo de respuesta

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

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

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

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

que es otra manera de expresar la desigualdad de Chebychev.

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

5.18 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 $a = k\cdot\sigma$,

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

5.18.1 La varianza como medida de dispersión

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

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 $2\cdot\sigma$ ¡*Sea cual sea la distribución de la v.a.*!

6 Teoría de la probabilidad

En esta sección definiremos de forma rigurosa los conceptos básicos de la teoría de la probabilidad.

Lo haremos de forma esquemática dejando las demostraciones que sean fáciles o medias al lector, como ejercicio.

Se supone que el lector conoce la teoría básca de conjuntos, operaciones entre conjuntos, aplicaciones entre conjuntos y pricipios básicos de numerabilidad.

Definición: Álgebra y σ-Álgebra

Sea Ω un conjunto y $\mathcal{A} \subseteq \mathcal{P}(\Omega)$ una familia de subconjuntos de Ω diremos que $\mathcal{A}$ es una **álgebra** de Ω si:

1. $\Omega \in \mathcal{A}$
2. $A \in \mathcal{A} \Rightarrow A^c \in \mathcal{A}$
3. $A_1, A_2, \dots, A_n \in \mathcal{A} \Rightarrow \bigcup_{i=1}^{\infty} A_i \in \mathcal{A}$ para todo $n \in \mathbb{N}$

Sí además se cumple que $\mathcal{A}$ es cerrada respecto a la unión numerable de conjuntos entonces diremos que $\mathcal{A}$ es una σ**-álgebra** de Ω.

El par formado por $(\Omega, \mathcal{A})$ se llama espacio medible y los elemento de $\mathcal{A}$ se denominan conjuntos medibles (en esta σ-álgebra) o sucesos.

Proposición

Sea $\mathcal{A}$ una σ-álgebra de Ω entonces:

1. Es cerrada respecto a la diferencia e conjuntos e intersecciones finitas o numerables. y además $\emptyset \in \mathcal{A}$.
2. Si $\mathcal{A}$ es un álgebra será tambiñen una σ-álgebra si

a. Dada un familia ${A_i}_{i=1}^{\infty}$ de elemento de $\mathcal{A}$ tales que $A_i \subseteq A_{i+1}$ para todo i entonces $\bigcup_{i=1}^{\infty} A_i \in \mathcal{A}$
b. Dada un familia ${A_i}_{i=1}^{\infty}$ de elemento de $\mathcal{A}$ tales que $A_i \cap A_j$ para todo $i, j\ i \neq j$ entonces $\bigcup_{i=1}^{\infty} A_i \in \mathcal{A}$

Proposición

Sean $\mathcal{A}$ y $\mathcal{B}$ dos σ-álgebras de Ω entonces $\mathcal{C} = A \cap B : A \in \mathcal{A}, B \in \mathcal{B}$ es una σ-álgebra de Ω.

Si $\mathcal{F} \in \mathcal{P}(\Omega)$ particular la intersección de todas as σ-álgebras de Ω que continene a $\mathcal{F}$ recibe el nombre de σ-álgebra generada por $\mathcal{F}$.

Algunas σ-álgebras notables

La σ-álgebra más grande de Ω es $\mathcal{P}(\Omega)$ y la más pequeña o trivial es $\{\emptyset, \Omega\}$.

Dado $A \subseteq \Omega$, la familia $\{\emptyset, A, A^c, \Omega\}$ es la σ-álgebra de Ω generada por A.

La σ-álgebra de Borel de $\mathbb{R}$ es la σ-álgebra generada por los semi-intervalos abiertos por derecha $\{x \in R : x < a\} =]-\infty, a]$ con $a \in \mathbb{R}$.

La anterior σ-álgebra se puede generalizar a $\mathbb{R}^n$ como la σ-álgebra generada por $\{(x_1, \dots, x_n) \in \mathbb{R}^n : x_i < a\}$ con $a \in \mathbb{R}$ y para $i \in \{1, 2, \dots, n\}$.

Proposición

La intersección arbitraria de σ-álgebras de partes de un conjunto Ω es una σ-álgebra en Ω. Así, si C es una familia de subconjuntos de Ω, existe la más pequeña σ-álgebra de Ω que contiene a C; la llamaremos σ-álgebra engendrada por Cy la denotaremos $\sigma(C)$."

Vamos a esbozar la algunas ideas para construir σ-álgebras

σ-álgebra producto Ejercicio

Sean $(\Omega_1, \mathcal{A}_1)$ y $(\Omega_2, \mathcal{A}_2)$ dos espacios medibles.

Podemos construir una **σ-álgebra producto** para un espacio medible producto $\Omega = \Omega_1 \times \Omega_2$.

Tomaremos como **conjunto producto medible** los conjuntos de la forma
$A_1 \times A_2$ con $A_i \in \mathcal{A}_i$ para $i = 1, 2$. La σ- álgebra generada por estos productos de conjuntos medibles se llama **σ-álgebra producto** y la denotaresmos por $\mathcal{A}_1 \times \mathcal{A}_2$.

Estudiad si la σ-álgebra de Borel en $\mathbb{R}^n$ es igual al producto de n copias de la σ-álgebra de Borel en $\mathbb{R}$.

σ-álgebra inducida

Dado un espacio medible $(\Omega, \mathcal{A})$ y un subconjunto $B \subset \Omega$ no vacío, podemos construir una σ-álgebra σ-álgebra en B a la que llamaremos inducida

$$\mathcal{A}_B = \{A \cap B : A \in \mathcal{A}\}$$

Demostrar que $(B, \mathcal{A}_B)$ es un espacio medible.

Definición: pi y d sistemas

Se define π-**sistema** con una familia subconjuntos de Ω cerrada bajo intersecciones finitas.

Definimos d-**sistema** o clase de Dynkin con una familia de subconjuntos de Ω $\mathcal{D} \subseteq \mathcal{P}(\Omega)$ que cumplan las siguientes propiedades:

1. $\Omega \in \mathcal{D}$.
2. Si $\{A_n\}_{n \geq 1}$ es una sucesión creciente de conjuntos de $\mathcal{D}$, entonces $\cup_{n=1}^{\infty} A_n \in \mathcal{D}$.
3. Si $A, B \in \mathcal{D}$ y $A \subset B$, entonces $B - A \in \mathcal{D}$.

Ejercicio

Demostrar que la intersección arbitraria de d-sistemas sigue siendo un d-sistema, lo que facilita la demostración de que una familia de subconjuntos forma una σ-álgebra.

Problemas propuestos

7 Medidas y Probabilidades

La imagen muestra la **Definición 3.1** sobre **medidas en espacios medibles** en teoría de la medida. A continuación, te hago un resumen y explicación detallada del contenido:

7.0.1 Definición de Medida

Sea $(\Omega, \mathcal{A})$ un **espacio medible**, donde:

- Ω es el **espacio de referencia**.
- $\mathcal{A}$ es una σ-**álgebra** de subconjuntos de Ω.

Una **medida** es una función:

$$\mu : \mathcal{A} \to [0, +\infty]$$

que cumple dos propiedades fundamentales:

1. **Aditividad numerable (σ-aditividad):**
   Si (A_n)es una sucesión de conjuntos **disjuntos** en $\mathcal{A}$, entonces:

$$\mu\left(\bigcup_n A_n\right) = \sum_n \mu(A_n).$$

2. **Medida del conjunto vacío:**

$$\mu(\emptyset) = 0.$$

A la σ-álgebra y su espacio de medida $(\Omega, \mathcal{A}, \mu)$ la llamaremos **espacio de medida**.

7.0.2 Tipos de medidas

Según el comportamiento de $⊠$, podemos clasificar las medidas en:

1. **Medida finita:**
   Si $\mu(\Omega) < +\infty$, es decir, la medida total del espacio es finita.
2. **Medida σ-finita:**
   Si Ω puede cubrirse con una colección **numerable** de conjuntos medibles de **medida finita**.
3. **Medida de probabilidad:**
   Si $⊠(⊠) = 1 $, se llama **probabilidad**, y el espacio medible $(\Omega, \mathcal{A}, \mu)$ se llama **espacio de probabilidad**.

7.0.3 Ejemplos

1. **Medida de conteo:**
   Si $\mu(A)$ es simplemente el número de elementos de$A$ (cuando$A$ es finito o numerable), entonces es una medida finita si Ω es finito, pero puede ser infinita si Ωes infinito.
2. Medida de Lebesgue en $\mathbb{R}$:
   - Es la **medida estándar de la longitud** en $\mathbb{R}$.
   - Es una **medida σ-finita** porque $\mathbb{R}$ puede cubrirse con intervalos de medida finita como $[-n, n]$.
3. Medida de probabilidad uniforme en $[0, 1]$:
   - Es una **medida de probabilidad** porque $\mu([0, 1]) = 1$.
   - La probabilidad de un subconjunto A es simplemente la longitud de A en $[0, 1]$.
4. **Medida de Dirac:**
   - Es una **medida de probabilidad**.
   - La probabilidad de un conjunto A es 1 si $0 \in A$ y 0 en caso contrario.

La imagen muestra la **Proposición 3.1**, que establece propiedades fundamentales de las medidas en espacios medibles. A continuación, te resumo y explico cada una de las propiedades junto con su demostración.

7.1 Proposición 3.1: Propiedades fundamentales de la medida

Sea μ una medida definida en un espacio medible $(\Omega, \mathcal{A})$, se cumplen las siguientes propiedades:

7.1.1 (a) Monotonía

Si $A \subseteq B$, entonces:

$$\mu(A) \leq \mu(B).$$

Si además $\mu(A) < +\infty$, entonces:

$$\mu(B - A) = \mu(B) - \mu(A).$$

Demostración:
Como B puede descomponerse como $B = A \cup (B - A)$ con $A \cap (B - A) = \emptyset$, la propiedad de **aditividad de la medida** nos dice que:

$$\mu(B) = \mu(A) + \mu(B - A) \geq \mu(A).$$

Si $\mu(A) < +\infty$, despejando obtenemos $\mu(B - A) = \mu(B) - \mu(A)$.

7.1.2 (b) Continuidad hacia arriba

Si (A_n) es una sucesión creciente de conjuntos, es decir,

$$A_1 \subseteq A_2 \subseteq A_3 \subseteq \dots,$$

entonces:

$$\mu\left(\bigcup_n A_n\right) = \lim_{n \to \infty} \mu(A_n).$$

Demostración:
Sea $A = \bigcup_n A_n$. Si $\mu(A_n) = +\infty$ para algún n, entonces $\mu(A_k) = \infty$ para todo $k \geq n$ y el resultado es trivial.
Si todos los A_n tienen **medida finita**, reescribimos la unión como:

$$A = A_1 \cup (A_2 - A_1) \cup (A_3 - A_2) \cup \dots$$

Dado que los términos son **disjuntos**, la aditividad numerable implica:

$$\mu(A) = \mu(A_1) + (\mu(A_2) - \mu(A_1)) + (\mu(A_3) - \mu(A_2)) + \dots$$

y esto converge a $\lim_n \mu(A_n)$.

7.1.3 (c) Continuidad hacia abajo

Si (A_n) es una sucesión decreciente de conjuntos, es decir:

$$A_1 \supseteq A_2 \supseteq A_3 \supseteq \dots,$$

y al menos uno de los A_n tiene medida finita, entonces:

$$\mu\left(\bigcap_n A_n\right) = \lim_{n\to\infty} \mu(A_n).$$

Demostración:
Podemos suponer sin pérdida de generalidad que A_1 tiene medida finita. Definimos $A = \bigcap_n A_n$ y consideramos la sucesión creciente $B_n = A_1 - A_n$, la cual satisface:

$$\bigcup_n B_n = A_1 - A.$$

Aplicando la **continuidad hacia arriba** al conjunto B_n, se obtiene:

$$\mu(A_1 - A) = \lim_n \mu(A_1 - A_n).$$

Restando de $\mu(A_1)$, se obtiene la ecuación deseada:

$$\mu(A) = \lim_n \mu(A_n).$$

7.1.4 (d) Subaditividad numerable (σ-subaditividad)

Si (A_n) es una sucesión arbitraria de conjuntos, entonces:

$$\mu\left(\bigcup_n A_n\right) \leq \sum_n \mu(A_n).$$

Demostración:
Sea $A := \bigcup_{n=1}^{\infty} A_n$. Descomponemos A en términos de las uniones de diferencias:

$$A = \bigcup_{n=1}^{\infty}\left(A_n - \bigcup_{i=1}^{n-1} A_i\right).$$

Por **aditividad numerable** y el hecho de que la medida es no negativa:

$$\mu(A) = \sum_n \mu(A_n - \bigcup_{i=1}^{n-1} A_i) \leq \sum_n \mu(A_n).$$

7.2 Conclusión

Estas propiedades son fundamentales en la teoría de la medida y se aplican en muchas áreas como **análisis real, probabilidad y teoría de integración**. En particular:

- La **monotonía** muestra que medidas más grandes contienen subconjuntos de medida más pequeña.
- La **continuidad hacia arriba y hacia abajo** garantiza que la medida respeta límites de secuencias de conjuntos.
- La **σ-subaditividad** es una propiedad crucial para la construcción de medidas y prueba de teoremas como el **teorema de Carathéodory**.

8 Distribuciones notables 1

8.1 Introducción

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,...

8.2 Distribución Bernoulli

Distribución Bernoulli

Consideremos un experimento con dos resultados posibles éxito (E) y fracaso (F). El espacio de sucesos será $\Omega = \{E, F\}$.

Supongamos que la probabilidad de éxito es $P(E) = p$, y naturalmente $P(F) = 1 - p = q$ con $0 < p < 1$.

Consideremos la aplicación

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

definida por

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

Su función de probabilidad es

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

Su función de distribución es

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

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

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

Este tipo de experimentos (éxito/fracaso)se reciben el nombre de experimentos Bernoulli.

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

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 $E(X^2)$

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

La **varianza** de una va.a Bernouilli es

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

Su desviación típica es

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

Pongamos en una tabla el resumen de v.a con distribución Bernoulli

Ejemplo : Cálculos con R

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

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

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

```
[1] 0.75
```

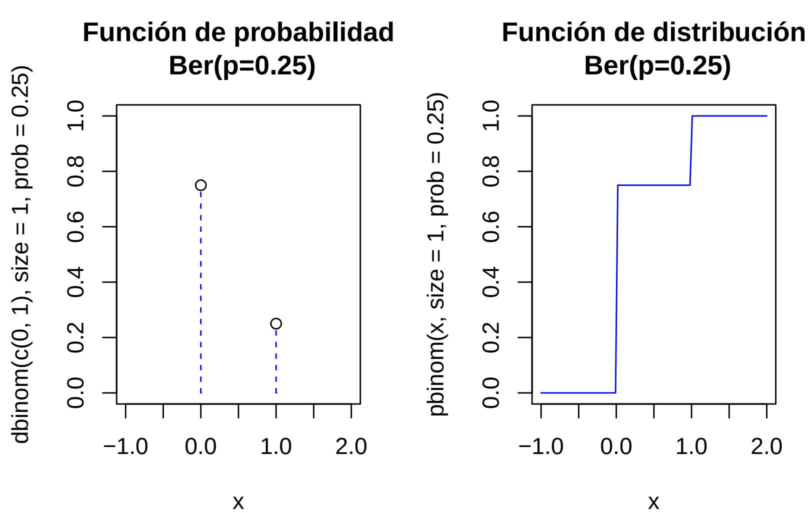
```
dbinom(1,size=1,prob=0.25)
```

```
[1] 0.25
```

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

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

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



8.3 Distribución binomial

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 $X \sim B(n,p)$.

Obviamente se tiene que una v.a. Bernoulli es una binomial con $n=1$

$$B(1,p) = Ber(p).$$

La **función de probabilidad** de una binomial es

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

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

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

Recordemos los números binomiales con un ejemplo

Números binomiales

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

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

```
[1] 6
```

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

```r
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!!

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

```
[1] 646646
```

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

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, hoja de cálculo dispone de funciones para el cálculo de estas probabilidades, así que el **uso de las tablas** queda **totalmente anticuado**.

La **esperanza** de una v.a. X con distribución $B(n, p)$ 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

X binomial: $B(n,p)$
$D_X=\{0,1,\ldots n\}$
$P_X(x)=P(X=x)=\begin{cases}\binom{n}{x}\cdot p^x\cdot(1-p)^{n-x} & \text{si } x=0,1,\ldots,n\\ 0 & \text{en otro caso.}\end{cases}$
$F_X(x)=P(X\leq X)=\begin{cases}0 & \text{si } x<0\\ \sum_{i=0}^{k}\binom{n}{i}\cdot p^i\cdot(1-p)^{n-i} & \text{si } k\leq x<k+1; \quad k=0,1,\ldots,n\\ 1 & \text{si } x\geq n\end{cases}$.
$E(X)=n\cdot p; Var(X)=n\cdot p\cdot(1-p)$

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

Por lo tanto su **varianza** se calcula así

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

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 $X=\displaystyle\sum_{i=1}^{n}X_i$ sigue una distribución $B(n,p)$.

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

8.3.1.1 Cálculos distribución binomial con R

Veamos los cálculos básicos con funciones de R para una v.a X con distribución binomial $B(n=10,p=0.25)$.

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

```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
```

8.3.1.2 Generación de muestras aleatorias con R

Generaremos una muestra aleatoria de 100 valores de una población con distribución $B(20, 0.5)$

```r
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
```

Que corresponde a los resultados de repetir 100 veces el experimento de lanzar una moneda 20 veces y contar el número de caras.

8.3.1.3 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)
```

```
0.056313514709472684
```

```
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)
```

```
0.056313514709472656
```

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

```
0.14599800109863284
```

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([ 7,  8,  8,  6,  5,  7,  5,  9,  7,  6,  5,  6,  7,  5,  4,  7,  4,
        6,  4,  4,  4,  4,  3,  7,  5,  7,  4,  5,  4,  7,  5,  1,  9,  6,
        6,  6,  5,  8,  1,  7,  8,  2,  7,  3,  5,  4,  6,  4,  8,  5,  4,
        5,  5,  1,  5,  7,  5,  3,  3,  5,  3,  8,  9,  4,  4,  3,  7,  5,
        3,  4,  6,  3,  4,  4,  6,  9,  5,  4,  5,  3,  3,  4,  7,  4,  4,
         5,  7,  4,  5,  5,  0, 10,  8,  5,  3,  5, 12,  7,  3,  3],
        dtype=int64)
```

Observación

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

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

```
array([ 6,  3,  3,  2,  5,  6,  7,  6,  7,  2,  7,  6,  3,  9,  3,  3,  3,
        4,  4,  5,  8,  7,  1,  3,  5,  6,  5,  7,  2,  8,  7,  7,  3,  5,
        3,  6,  3,  4,  1,  6,  5,  5,  6,  6,  4,  5,  8,  3,  4,  5,  3,
        9,  4,  5,  8,  5,  6,  3,  5,  4,  3,  3,  4,  6,  5,  5,  3,  4,
        3,  2,  4,  4,  5,  5,  5,  7,  2,  7, 10,  7,  9,  3,  7,  5,  6,
         5,  8,  6,  4,  3,  8,  5,  6,  6,  4,  6,  8,  3,  7,  6],
        dtype=int64)
```

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)
```

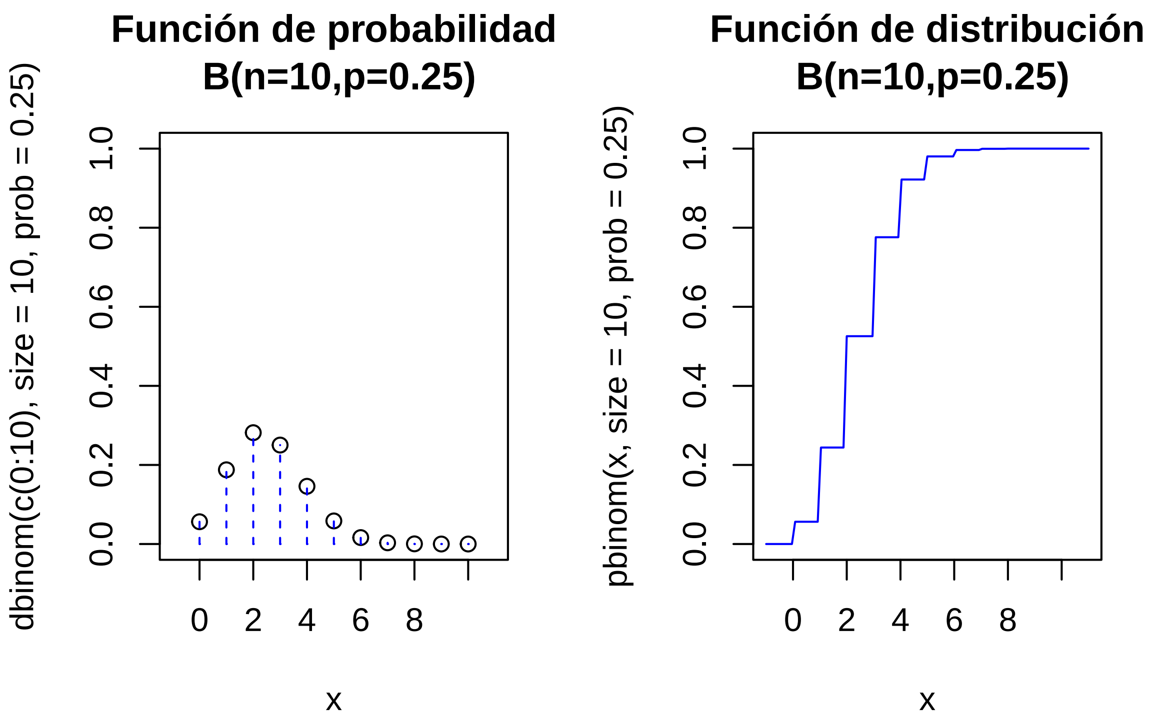
```
0.1877117156982421
```

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

```
array([3, 6, 6, 7, 6, 9, 5, 4, 5, 8], dtype=int64)
```

8.3.1.4 Gráficas de la distribución binomial con R

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



Ahora con `ggplot2`

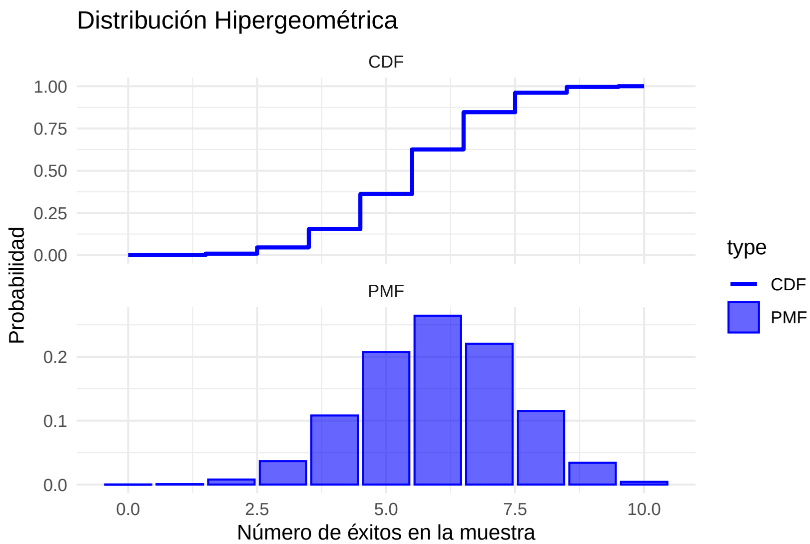
```
library(ggplot2)
library(dplyr)
```

```
Adjuntando el paquete: 'dplyr'
```

```
The following objects are masked from 'package:stats':

    filter, lag

The following objects are masked from 'package:base':

    intersect, setdiff, setequal, union
```

```
library(tidyr)
library(stats)

# Parámetros de la distribución hipergeométrica
M <- 100  # Tamaño total de la población
n <- floor(0.6 * M)  # Número de elementos de éxito en la población
N <- 10  # Tamaño de la muestra extraída

# Valores de la variable aleatoria
x <- 0:N

# Función de masa de probabilidad (PMF)
pdf <- dhyper(x, n, M - n, N)

# Función de distribución acumulativa (CDF)
cdf <- phyper(x, n, M - n, N)

# Crear un dataframe para ggplot
data <- data.frame(x = rep(x, 2),
                   prob = c(pdf, cdf),
                   type = rep(c("PMF", "CDF"), each = length(x)))

# Crear los gráficos con ggplot2
ggplot(data, aes(x = x, y = prob, fill = type, color = type)) +
  geom_col(data = subset(data, type == "PMF"), alpha = 0.6) +
  geom_step(data = subset(data, type == "CDF"), direction = "mid", size = 1)
  facet_wrap(~type, ncol = 1, scales = "free_y") +
  labs(title = "Distribución Hipergeométrica",
       x = "Número de éxitos en la muestra",
       y = "Probabilidad") +
  scale_fill_manual(values = c("blue", "blue")) +
  scale_color_manual(values = c("blue", "blue")) +
  theme_minimal()
```

```
Warning: Using `size` aesthetic for lines was deprecated in ggplot2 3.4.0.
i Please use `linewidth` instead.
```



8.3.1.5 Gráficos de la distribución binomial con python

Necesitaremos usar estas librerías:

```python
import numpy as np
import matplotlib.pyplot as plt
import scipy.stats as stats
```

```python
# Parámetros de la distribución binomial
n = 10
p = 0.7

# Valores posibles de X
x = np.arange(0, n+1)

# Función de probabilidad (PMF)
pmf_values = stats.binom.pmf(x, n, p)

# Función de distribución acumulada (CDF)
cdf_values = stats.binom.cdf(x, n, p)

# Graficar la función de probabilidad (PMF)
```

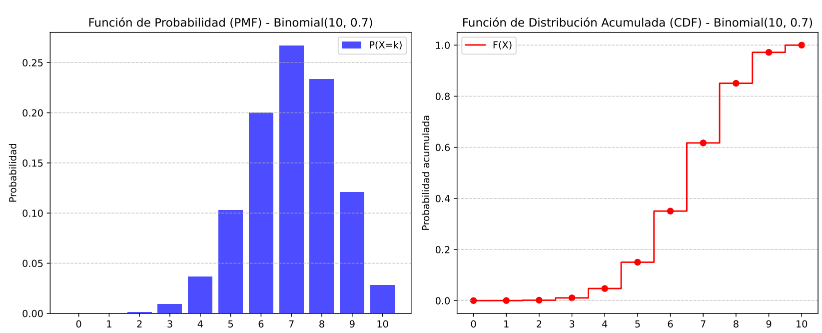
```python
plt.figure(figsize=(12, 5))
plt.subplot(1, 2, 1)
plt.bar(x, pmf_values, color='blue', alpha=0.7, label='P(X=k)')
plt.xlabel('Valores de X')
plt.ylabel('Probabilidad')
plt.title('Función de Probabilidad (PMF) - Binomial(10, 0.7)')
plt.xticks(x)
```

```
([<matplotlib.axis.XTick object at 0x000001A8174F8520>, <matplotlib.axis.XTic
```

```python
plt.legend()
plt.grid(axis='y', linestyle='--', alpha=0.7)

# Graficar la función de distribución acumulada (CDF)
plt.subplot(1, 2, 2)
plt.step(x, cdf_values, where='mid', color='red', label='F(X)')
plt.scatter(x, cdf_values, color='red')
plt.xlabel('Valores de X')
plt.ylabel('Probabilidad acumulada')
plt.title('Función de Distribución Acumulada (CDF) - Binomial(10, 0.7)')
plt.xticks(x)
```

```
([<matplotlib.axis.XTick object at 0x000001A81AC6ED00>, <matplotlib.axis.XTic
```

```python
plt.legend()
plt.grid(axis='y', linestyle='--', alpha=0.7)

# Mostrar gráficos
plt.tight_layout()
plt.show()
```



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

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

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

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

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

$$P(Roja) = P(xito) = p = \frac{40}{100} = 0.4.$$

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

```
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

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

```
[1] 0.3822806
```

Así que

$$
P(X \geq 4) = 1 - P(X < 4) = 1 - P(X \leq 3) \approx 1 - 0.3823 = 0.6177.
$$

Otra manera usando R sería:

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

```
[1] 0.6177194
```

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

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

```
[1] 0.6177194
```

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

$$
\begin{aligned}
P(X < 3) =& P(X \leq 2) = P(X=0)+P(X=1)+P(X=2) \\
=& \binom{10}{0} \cdot 0.4^{0} \cdot (1-0.4)^{10-0} + \binom{10}{1} \cdot 0.4^{1} \cdot (1-0.4)^{10-1} \\
& + \binom{10}{2} \cdot 0.4^{2} \cdot (1-0.4)^{10-2} \\
=& 0.1672898.
\end{aligned}
$$

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
```

8.4 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 es $D_X = \{0, 1, 2, \ldots\}$.

Ejemplo

Por ejemplo calculemos 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:

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

Lo que demuestra la siguiente propiedad

Función de distribución de una $Ge(p)$

Si X es una v.a. que sigue una distribución geométrica de parámetro p, entonces su función de distribución es:

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

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 $0<1-p<1$.

Hagamos los cálculos necesarios para los momentos básicos de la distribución geométrica. Recordemos del tema de variables aleatorias que

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

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

Ahora podemos calular el valor esperado y la varianza de una v.a. $Ge(p)$.vRecordemos que $P(X = x) = (1-p)^x \cdot p$ si $x = 0, 1, 2, \ldots$ y aplicado la fórmula anterior con $r = 1 - p$

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

Necesitamos calcular $E(X^2)$, aplicado la fórmula anterior con $r = 1 - p$

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

Ahora la varianza es

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

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

Por último su desviación típica es

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

8.4.1 Resumen distribución geométrica $Ge(p)$ empezando en 0

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= número de intentos para obtener el primer éxito. Entonces $Y = X + 1$ donde $X \sim Ge(p)$. 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 $Var(Y) = Var(X+1) = Var(X) = \frac{1-p}{p^2}$.

8.4.2 Resumen distribución geométrica $Ge(p)$ empezando en 1.

8.4.3 Propiedad de la falta de memoria

Propiedad de la falta de memoria

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

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

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

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

Es decir 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

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

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

De forma similar obtenemos

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

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

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

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

Lo que demuestra la propiedad de la falta de memoria

Observación: Interpretación
de la propiedad

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*).

Un ejemplo muy sencillo nos aclarará el alcance de esta propiedad

Ejercicio: la llave que abre la puerta

Tenemos un llavero con 10 llaves, solo 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: El clásico del fútbol

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.

Nos hacemos las siguientes preguntas:

- Si Aina solo tiene dinero para ir a ver 3 partidos, ¿cuál es la probabilidad de no ver ganar al Barça en al menos tres partidos consecutivos?
- ¿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, que cuenta el número de veces consecutivas que en un partido de fútbol de la liga el Barça no gana al Madrid, con una ley geométrica empezando en cero con probabilidad de éxito $p = P(\text{Barça}) = \frac{72}{178}$,

$$X = Ge\left(p = \frac{72}{178} = 0.4045\right)$$

Así que lo que nos pregunta Aina es la siguiente probabilidad

$$P(X \geq 3) = 1 - P(X \leq 2) = 1 - \left(1 - \frac{72}{178}\right)^{2+1} = 0.7888.$$

Así que Aina tiene una probabilidad del 78.88% de no ver ganar al Barça en al menos 3 partidos antes de ver uno en el sí que gane.

Para responder a la segunda pregunta, usando que la distribución de X es:

$$X = Ge\left(p = \frac{72}{178} = 0.4045\right)$$

entonces

$$E(X) = \frac{1-p}{p} = \frac{1-0.4045}{0.4045} = 1.4722$$

y

$$Var(X) = \frac{1-p}{p^2} = \frac{1-0.4045}{0.4045^2} = 3.6397$$

La desviación típica es

$$\sqrt{3.6397} = 1.9078.$$

8.4.3.1 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
```

$P(X \leq 0) = 1 - (1 - 0.25)^{0+1} = 1 - 0.75 = 0.25$

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

```
[1] 0.25
```

$P(X \leq 4) = 1 - (1-0.25)^{4+1} = 1 - 0.75 = 1 - 0.75^5 = 0.7626953.$

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

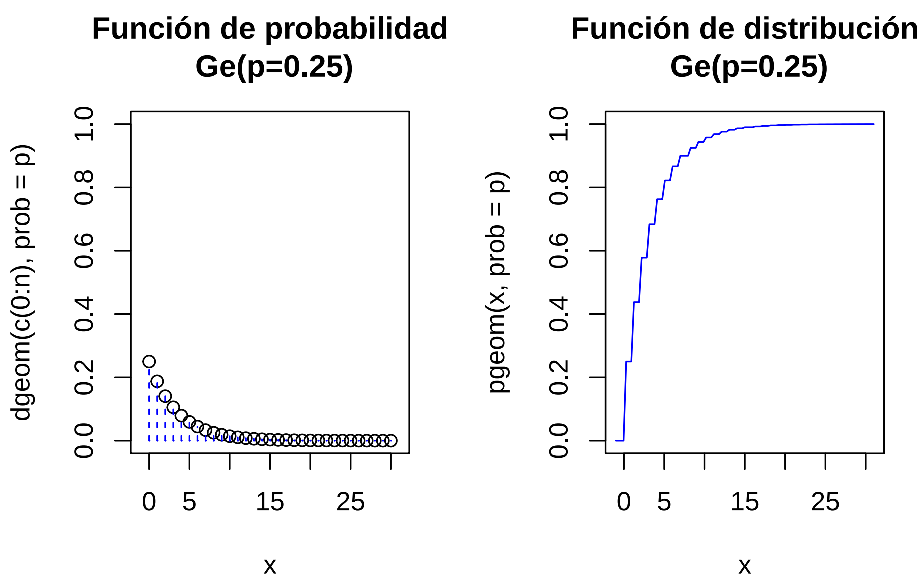
```
[1] 0.7626953
```

Una muestra aleatoria de tamaño 25 de una $Ge(0.25)$

```r
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
```

8.4.3.2 Gráficos con R el código



8.4.3.3 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 $Ge(0.25)$

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

```
array([ 0, 11,  2,  7,  0,  0, 10,  1,  6,  3,  0,  5,  1,  0,  5, 16,  2,
        4,  4,  3], dtype=int64)
```

Ejercicio

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

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

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

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

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

Con python también podemos calcular directamente algunos parámetros asociado a una función de distribución predefinida

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

```
(array(4.), array(12.))
```

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

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

Ejercicio

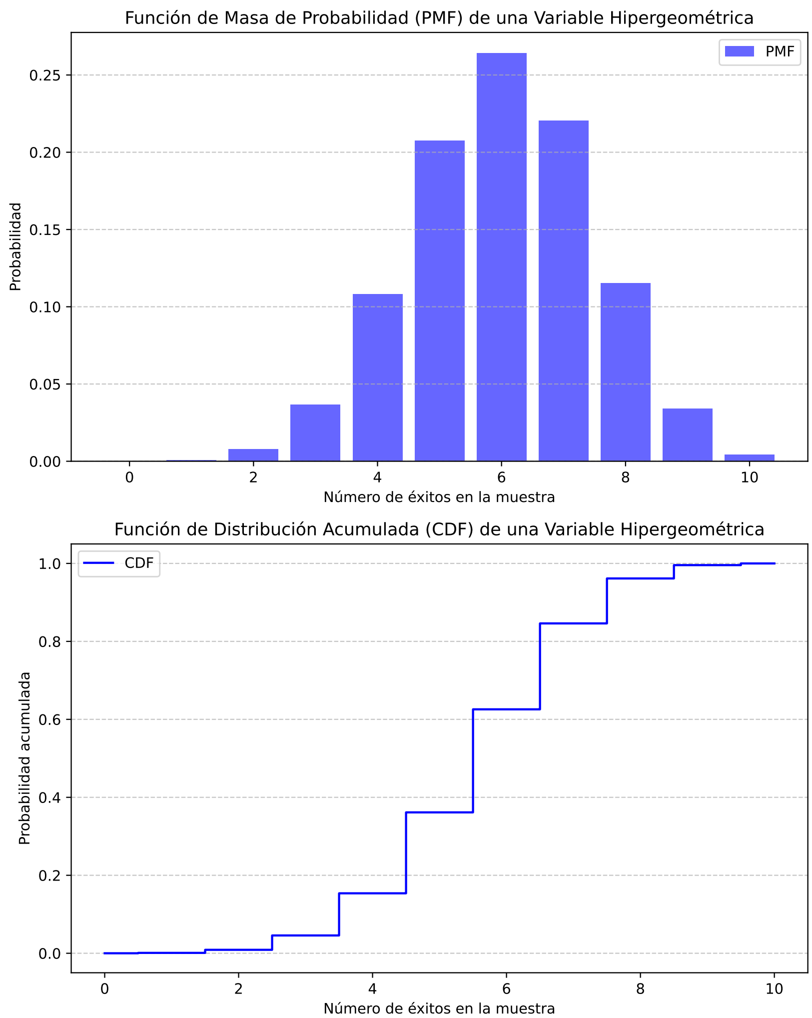
Comprobad que las medias y las varianzas calculadas en el código anterior, corresponden a una $Ge(p = 0.3)$ empezando en 1 y a una $Ge(p = 0.3)$ empezando en 0.

¿Son las varianzas siempre iguales?

8.4.3.4 Gráficos distribución geométrica con python

```
import numpy as np
import matplotlib.pyplot as plt
from scipy.stats import hypergeom

# Parámetros de la distribución hipergeométrica
M = 100  # Tamaño total de la población
n = int(0.6 * M)  # Número de elementos de éxito en la población
N = 10  # Tamaño de la muestra extraída

# Valores de la variable aleatoria
x = np.arange(0, N+1)

# Función de masa de probabilidad (PMF)
pdf = hypergeom.pmf(x, M, n, N)

# Función de distribución acumulativa (CDF)
cdf = hypergeom.cdf(x, M, n, N)

# Crear la figura y los subgráficos
fig, axes = plt.subplots(2, 1, figsize=(8, 10))

# Gráfico de la PMF
axes[0].bar(x, pdf, color='blue', alpha=0.6, label='PMF')
axes[0].set_title('Función de Masa de Probabilidad (PMF) de una Variable Hip
axes[0].set_xlabel('Número de éxitos en la muestra')
axes[0].set_ylabel('Probabilidad')
axes[0].legend()
axes[0].grid(axis='y', linestyle='--', alpha=0.7)

# Gráfico de la CDF
axes[1].step(x, cdf, color='blue', where='mid', label='CDF')
axes[1].set_title('Función de Distribución Acumulada (CDF) de una Variable H
axes[1].set_xlabel('Número de éxitos en la muestra')
axes[1].set_ylabel('Probabilidad acumulada')
axes[1].legend()
axes[1].grid(axis='y', linestyle='--', alpha=0.7)

# Ajustar el layout para evitar solapamientos
plt.tight_layout()

# Mostrar los gráficos
```

```
plt.show()
```



8.5 Distribución binomial negativa

Podemos volver al probalma de de la puerta pero en esta ocasión tiene 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 $n = 2$:

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

Distribución binomial negativa

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

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

Propiedade

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

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

Demostración

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

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

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

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:

$$(t+1)^{-n} = \sum_{k=0}^{+\infty} \left(\begin{array}{c} -n \\ \\ k \end{array} \right) t^k$$

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

$$\begin{aligned}
\binom{-6}{4} &= \tfrac{-6\cdot(-6-1)\cdot\cdot(-6-2)\cdot(-6-3)}{4!} \\
&= \tfrac{-6\cdot(-7)\cdot(-8)\cdot(-9)}{24} \\
&= \tfrac{3024}{24} = 126.
\end{aligned}$$

Si realizamos el cálculo con R obtenemos el mismo resultado:

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

```
[1] 126
```

8.5.1 Esperanza y varianza 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**

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

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 por tanto la desviación típica es

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

8.5.2 Resumen distribución Binomial Negativa $BN(n,p)$

8.5.3 Ejemplo de la puerta con dos cerraduras

: Ejemplo: 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 de X?

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

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 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

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

Si aplicamos la expresión anterior para $n = 2$ y $p = 0.1$, obtenemos:

$$P_X(k) = P(X = k) = \begin{cases} \binom{k+2-1}{2-1} \cdot 0.9^k \cdot 0.1^2 & \text{si } k = 0, 1, 2, \ldots \\ 0 & \text{en otro caso.} \end{cases}$$

Simplificando

$$P_X(X=k)=P(X=k)=\begin{cases} 0.01\cdot(k+1)\cdot 0.9^k, & \text{si } k=0,1,2,\ldots \\ 0 & \text{en otro caso.}\end{cases}$$

La función de distribución en general es

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

Simplificando para $n=2, p=0.1$.

$$F_X(x)=P(X\leq x)=\begin{cases} 0, & \text{si } x<0, \\ \sum_{i=0}^{k} 0.01\cdot(i+1)\cdot 0.9^{i+1}, & \text{si } \begin{cases} k\leq x<k+1, \\ k=0,1,2,\ldots\end{cases}\end{cases}$$

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

$$P(X=5)=0.01\cdot(5+1)\cdot 0.9^5=0.06\cdot 0.9^5=0.0354294.$$

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

Nos piden que

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

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 es

$$Var(X) = n \cdot \frac{1-p}{p^2} = 2 \cdot \frac{1-0.1}{0.1^2} = 180,$$

y su desviación típica $\sqrt{180} = 13.41641$.

8.5.3.1 Cálculos de la distribución $BN(n, p)$ 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
```

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

```
[1] 0.885735
```

```
pnbinom(4,size=2,p=0.1,lower.tail=FALSE)
```

```
[1] 0.885735
```

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

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

Recordemos que de nuevo se cumple que

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

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

```
0.0354294
```

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

```
0.0354294
```

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

```
0.11426500000000002
```

```
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.

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

```
array([ 3,  8, 14, 11,  7, 57, 48, 14,  2, 71, 16,  0, 38,  0,  4, 10, 48,
       12,  9, 20, 29, 11, 10, 37, 11, 13,  7, 18,  6, 44,  9, 18,  2, 25,
       11,  6,  7, 11, 27,  6, 24,  3,  2,  6, 10,  5,  7, 40, 35,  3, 19,
       61, 13,  1,  6,  6,  9,  8, 13, 32, 47, 12, 19, 22, 30, 41,  9, 30,
       14, 16, 92, 36, 26,  8, 20, 21, 27,  5, 24, 18, 19, 18, 20, 14,  4,
        8, 31,  1, 22, 29,  3, 28,  7, 14,  6, 20,  3, 36,  3,  8],
      dtype=int64)
```

8.5.3.2 Cálculos de la distribución $BN(n, p)$ con python

La **esperanza** y la **varianza**de una $BN(n = 2, 0.1)$ valen:

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

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

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

```
Var(X)=179.99999999999997
```

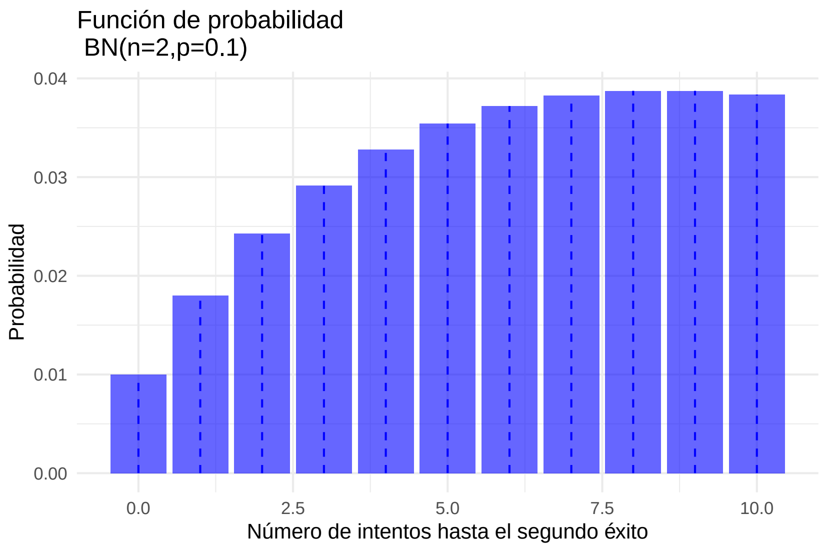
8.5.3.3 Gráficas de la binomial negativa con R

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))
```

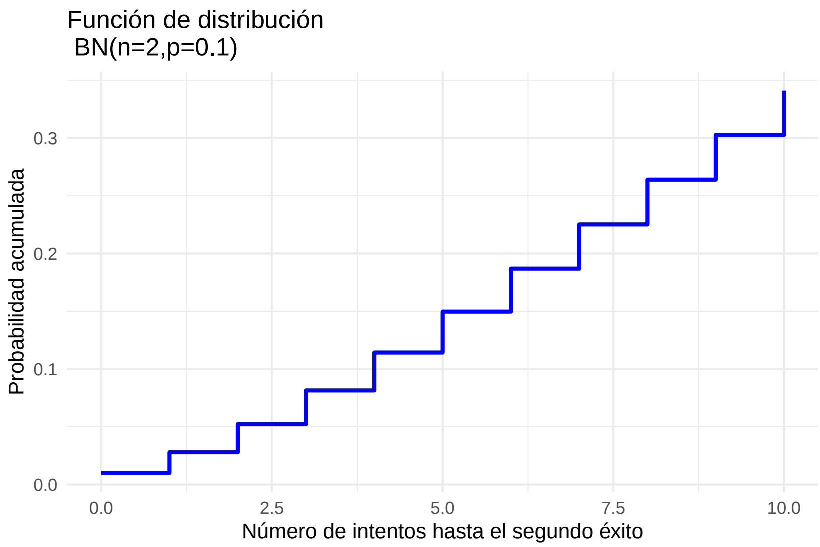
Y con ggplot2:

```
library(ggplot2)
library(dplyr)

# Parámetros de la distribución binomial negativa
size <- 2
prob <- 0.1
x <- 0:10

# Crear dataframes para PMF y CDF
df_pmf <- data.frame(x = x, y = dnbinom(x, size, prob))
df_cdf <- data.frame(x = x, y = pnbinom(x, size, prob))

# Gráfico de PMF
ggplot(df_pmf, aes(x = x, y = y)) +
  geom_bar(stat = "identity", fill = "blue", alpha = 0.6) +
  geom_segment(aes(x = x, xend = x, y = 0, yend = y), linetype = "dashed", c
  labs(title = "Función de probabilidad\n BN(n=2,p=0.1)",
       x = "Número de intentos hasta el segundo éxito", y = "Probabilidad")
  theme_minimal()
```




```
# Gráfico de CDF
ggplot(df_cdf, aes(x = x, y = y)) +
  geom_step(color = "blue", size = 1) +
```

```
  labs(title = "Función de distribución\n BN(n=2,p=0.1)",
       x = "Número de intentos hasta el segundo éxito", y = "Probabilidad ac
  theme_minimal()
```

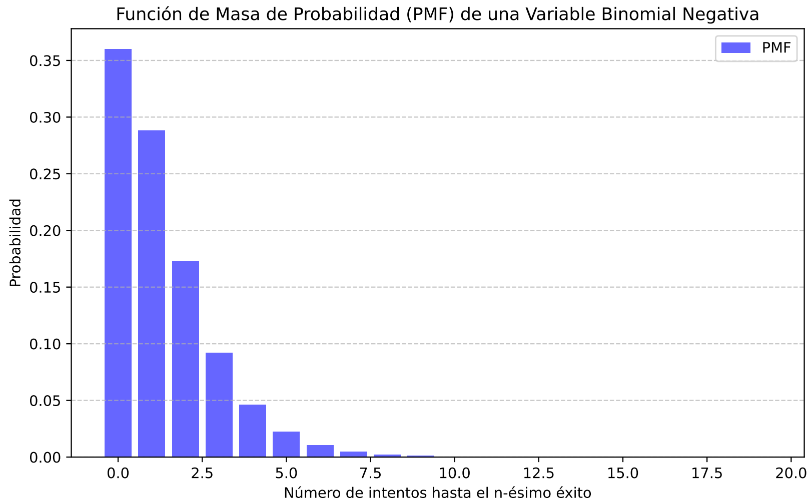


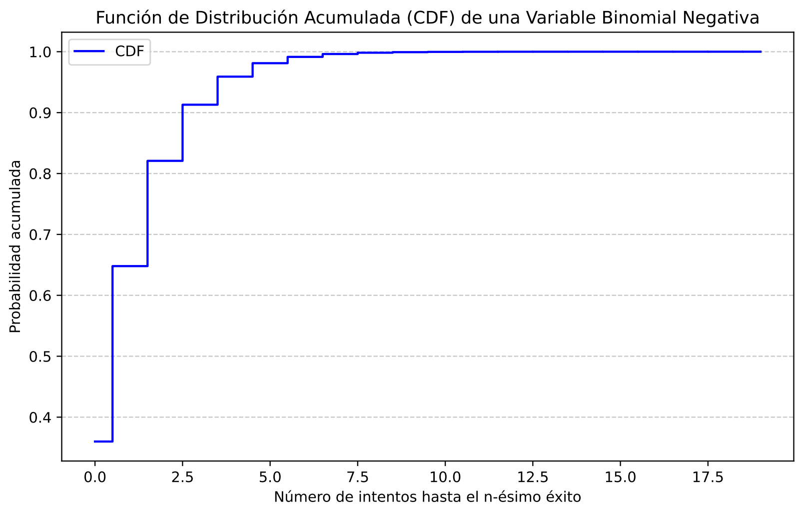
8.5.3.4 Gráficos de la binomial negativa con python

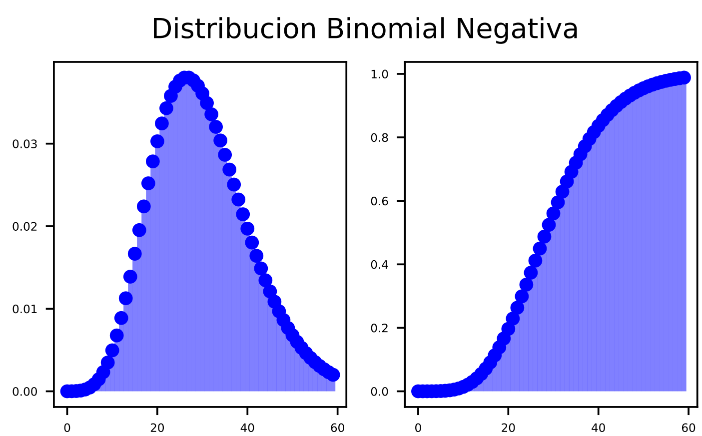
Este es el código con python que dibuja las función de probabilidad y la de distribución de una $BN(n = 2, p = 0.1)$

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

# Parámetros de la distribución binomial negativa
n = 2   # Número de fracasos antes de alcanzar el éxito
p = 0.6   # Probabilidad de éxito en cada intento

# Valores de la variable aleatoria
x = np.arange(0, 20)

# Función de masa de probabilidad (PMF)
pdf = nbinom.pmf(x, n, p)

# Función de distribución acumulativa (CDF)
```

```
cdf = nbinom.cdf(x, n, p)

# Crear la figura y los subgráficos
fig, axes = plt.subplots(2, 1, figsize=(8, 10))

# Gráfico de la PMF
axes[0].bar(x, pdf, color='blue', alpha=0.6, label='PMF')
axes[0].set_title('Función de Masa de Probabilidad (PMF) de una Variable Bin
axes[0].set_xlabel('Número de intentos hasta el n-ésimo éxito')
axes[0].set_ylabel('Probabilidad')
axes[0].legend()
axes[0].grid(axis='y', linestyle='--', alpha=0.7)

# Gráfico de la CDF
axes[1].step(x, cdf, color='blue', where='mid', label='CDF')
axes[1].set_title('Función de Distribución Acumulada (CDF) de una Variable B
axes[1].set_xlabel('Número de intentos hasta el n-ésimo éxito')
axes[1].set_ylabel('Probabilidad acumulada')
axes[1].legend()
axes[1].grid(axis='y', linestyle='--', alpha=0.7)

# Ajustar el layout para evitar solapamientos
plt.tight_layout()

# Mostrar los gráficos
plt.show()
```


Función de Masa de Probabilidad (PMF) de una Variable Binomial Negativa
PMF
Probabilidad
Número de intentos hasta el n-ésimo éxito


Función de Distribución Acumulada (CDF) de una Variable Binomial Negativa
CDF
Probabilidad acumulada
Número de intentos hasta el n-ésimo éxito


Distribucion Binomial Negativa

8.5.3.5 Caso práctico: Acceso aleatorio a un sistema con triple clave.

Sistema con tres claves de acceso

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 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

$$P(X=150) = \binom{152}{2} \cdot 0.999^{150} \cdot 0.001^3.$$

Lo calcularemos operando con R

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

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

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

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

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

Nos piden, lo resolveremos con python

$$P(X=150) = \binom{152}{2} \cdot 0.999^{150} \cdot 0.001^3$$

```
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)
```

```
9.876743459670532e-06
```

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

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

Calculemos $P(X \leq 150)$

$$\begin{aligned}
P(X \leq 150) =& P(X = 0) + P(X = 1) + P(X = 2) + \ldots + P(X = 150) \\
=& \sum_{k=0}^{150} \binom{k+3-1}{3-1} \cdot (0.999)^k \cdot 0.001^3 \ldots = \ldots = 5.2320035 \times 10^{-4}
\end{aligned}$$

Con R

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

```
[1] 0.0005232003
```

Con python

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

```
0.0005232003490824064
```

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 desviación típica?

Tenemos que $E(X) = n \cdot \frac{1-p}{p} = 3 \cdot \frac{1-0.001}{0.001} = 2997$ y $Var(X) = n \cdot \frac{1-p}{p^2} = 3 \cdot \frac{1-0.001^2}{0.001^2} = 2.997 \times 10^6$.

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
```

8.5.3.6 Caso práctico: ¿Tres claves de tres dígitos o una de 9 dígitos?

Supongamos que hemos decidido tener 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.

Como en el caso anteriro supongamos que cada vez que alguien intenta acceder lo hace con una 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 $Ge(p = \frac{1}{10^9} = 0.000000001)$.

Qué da más seguridad ¿tres claves de tres dígitos o una de 9 dígitos?

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 es

$$3 \cdot \frac{1-0.001}{0.001} = 2997.$$

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.

8.6 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:

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

8.6.1 La distribución de Poisson como "límite" de una binomial.

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

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

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

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

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

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

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

$$
\lim_{n\to\infty} \frac{n!}{(n-k)!\cdot n^k} = \lim_{n\to\infty} \frac{n\cdot(n-1)\cdots(n-k-1)}{n^k} = \lim_{n\to\infty} \frac{n^k+\cdots}{n^k} = 1.
$$

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

Y también teniendo en cuanta que k es constante.

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

Para acabar

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

Lo que confirma que límite de una serie de variables $B(n, p_n = \frac{\lambda}{n})$ sigue una ley $Po(\lambda)$.

X con distribución Poisson de media o promedio λ, $Po(\lambda)$
$D_X = \{0, 1, ...\}$
$P_X(x) = P(X = x) = \begin{cases} \frac{\lambda^x}{x!} e^{-\lambda} & \text{si } x = 0, 1, ... \\ 0 & \text{en otro caso.} \end{cases}$
$F_X(x) = P(X \leq X) = \begin{cases} 0 & \text{si } x < 0 \\ \sum_{i=0}^{k} P(X=i) = \sum_{i=0}^{k} \frac{\lambda^i}{i!} \cdot e^{-\lambda} & \text{si } \begin{cases} k \leq x < k+1 \\ k = 0, 1, 2, ... \end{cases} \end{cases}$
$E(X) = \lambda$; $Var(X) = \lambda$

8.6.2 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.

Claro que esto no tiene que poder ser así.

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

Así que podremos definir una serie de variables X_t de distribución $Po(\lambda \cdot t)$.

Definición 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=numero 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**.

8.6.3 Resumen distribución Poisson $X \sim Po(\lambda)$

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

8.6.4.1 Aproximación de la distribución binomial por la Poisson

Bajo el punto de vista anterior y si p es pequeño y n suficientemente grande la distribución $B(n, p)$ se aproxima a una $Po(\lambda = n \cdot p)$.

$X_t =$ número de eventos en el intervalo $(0,t]$ $Po(\lambda\cdot t)$ donde λ promedio por u.t.
$D_X = \{0, 1, ...\}$
$P_X(x) = P(X = x) = \begin{cases} \frac{(\lambda\cdot t)^x}{x!} e^{-\lambda\cdot t} & \text{si } x = 0, 1, ... \\ 0 & \text{en otro caso.} \end{cases}$
$F_X(x)=P(X\leq X)=\begin{cases} 0 & \text{si } x < 0 \\ \sum_{i=0}^{k} P(X=i)=\sum_{i=0}^{k} \frac{(\lambda\cdot t)^i}{i!}\cdot e^{-\lambda\cdot t} & \text{si } \begin{cases} k\leq x<k+1 \\ k=0,1,2,... \end{cases} \end{cases}$
$E(X) = \lambda \cdot t; Var(X) = \lambda \cdot t$

Existen distintos criterios (ninguno perfecto) de cuando la aproximación es buena.

Por ejemplo si

$$n \geq 20 \text{ o mejor } n \geq 30, n\cdot p < 10 \text{ y } p \leq 0.05,$$

la aproximación de una $B(n, p)$ por una $Po(n \cdot p)$ es buena. Sobre todo para los valores cercanos a $E(X) = \lambda$.

Condición deseable $n \geq 20$, $n \cdot p < 10$, $p \leq 0.05$.

8.6.5 Caso práctico: Trampa de insectos

Ejemplo: Trampa 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

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

En nuestro caso

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

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

Nos piden la probabilidad siguiente:

$$P(X = 21) = \frac{20^{21}}{21!} e^{-20} = 0.0846051.$$

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

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

```
[1] 0.08460506
```

```r
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

$$\sqrt{Var(X)} = +\sqrt{\lambda} = +\sqrt{20} = 4.47214.$$

8.6.6 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 $F_X(10) = \sum\limits_{x=0}^{10} P_X(x)$:

```
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 $P(X = 21)$:

```r
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 $P(X \geq 6) = 1 - P(X < 6) = 1 - P(X \leq 5)$:

```r
ppois(5,lambda=20)
```

```
[1] 7.190884e-05
```

```r
1-ppois(5,lambda=20) # es 1-P(X<=5)=P(X>=6)
```

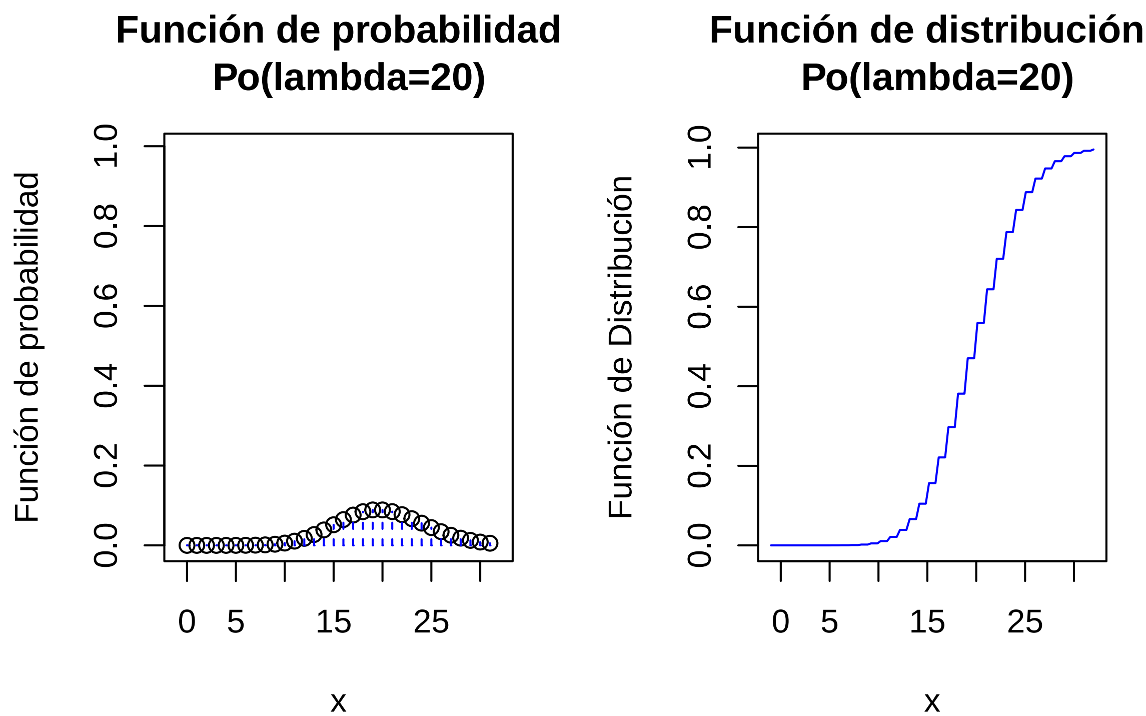
```
[1] 0.9999281
```

```r
ppois(5,lambda=20,lower.tail =FALSE ) # acumula hacia arriba
```

```
[1] 0.9999281
```

```r
# P(X>5)=P(X>=6)=P(X=6)+P(X=7)+...
```

8.6.6.1 Gráficos de la distribución Poisson con R



8.6.7 Cálculos con python

Sea X un una v.a. $Po(\lambda = 3)$. Entonces

$P_X(0) = P(X = 0), P_X(1) = P(X = 1)$ en este orden son

```
from scipy.stats import poisson
poisson.pmf(0,mu = 3)
```

```
0.049787068367863944
```

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

```
0.14936120510359185
```

Sea X un una v.a. $Po(\lambda = 3)$. Entonces

$F_X(0) = P(X \leq 0), F_X(1) = P(X \leq 1)$ en este orden son

```
poisson.cdf(0,mu = 3)
```

```
0.04978706836786395
```

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

```
0.1991482734714558
```

```
poisson.pmf(0,mu = 3)+poisson.pmf(1,mu= 3)
```

```
0.1991482734714558
```

```
## es igual a poisson.cdf(1,lambda=3)
```

Por ejemplo podemos comprobar que $F_X(10) = \sum_{0}^{10} P_X(x)$

```
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
```

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.

La respuesta a la pregunta 3 es calcular $P(X = 21)$

```python
poisson.pmf(21,mu=20)
```

```
0.08460506418293791
```

```python
# P(X=21)
```

Pregunta 4. Calculad la probabilidad de que en una hora caigan en la trampa al menos 6 insectos.

La pregunta 4 nos pide calcular $P(X \geq 6) = 1 - P(X \leq 5)$

```python
1-poisson.cdf(5,mu=20)
```

```
0.9999280911594716
```

```python
# es 1-P(X<=5)=P(X>=6)
```

Como ya hemos visto con `scipy.stats` podemos pedir los momentos de una variable aleatoria $Po(3)$

```python
poisson.stats(mu=3, moments='mv')
```

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

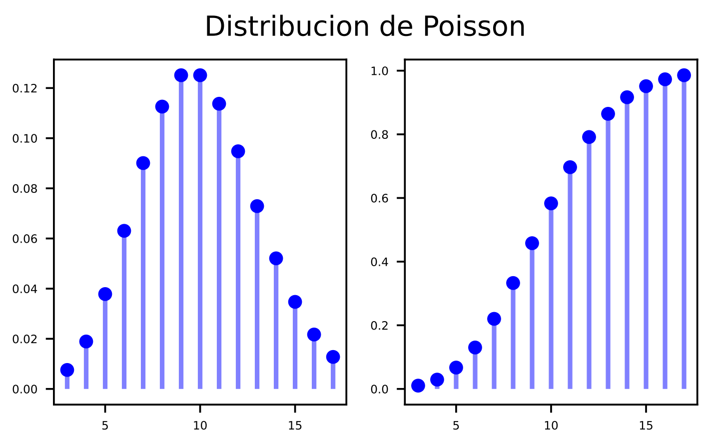
Y también generar secuencias de observaciones aleatorias de una población $Po(3)$

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

```
array([ 3,  8,  5,  0,  5,  2,  3,  5,  3,  0,  2,  4,  5,  6,  2,  0,  4,
        2,  2,  2,  2,  3,  1,  4,  0,  6,  4,  1,  0,  2,  5,  3,  6,  2,
        1,  1,  5,  4,  3, 11], dtype=int64)
```

8.6.7.1 Gráficos con python

```
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()
```



8.6.8 Caso práctico: Proceso Poisson visera de un casco

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 $Po(3 \cdot t)$.

Ahora estamos en condiciones de preguntar al proceso de Poisson.

¿Cuál es la probabilidad de que en 10 minutos impacten más de 25 insectos?

En este caso $t = 10$ X_{10}= número de insectos que impactan en 10 minutos, el intervalo $[0, 10)$ que sigue una $P(3 \cdot 10 = 30)$. Por lo tanto

$$P(X > 25) = 1 - P(X \leq 25)$$

lo resolvemos con R

```
1-ppois(25,lambda=30)
```

```
[1] 0.7916426
```

Otra pregunta interesante es que tengamos que esperar más de 2 minutos para observar el primer impacto

$$P(X_2 = 0) = \frac{(3 \cdot 2)^0}{0!} \cdot e^{-3 \cdot 2} = e^{-6} = 0.002479.$$

Con R

```
6^0/factorial(0)*exp(-6)
```

```
[1] 0.002478752
```

```
ppois(0,lambda=3*2)
```

```
[1] 0.002478752
```

8.7 Distribución hipergeométrica

Modelo de la 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 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".

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

- $H(m=2, n=5, k=3)$. 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:

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 así

$$H(N, k, p) \text{ donde } p = \frac{m}{N}.$$

8.7.1 Resumen distribución Hipergeométrica $H(m, n, k)$.

Urna con bolas blancas y rojas

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 $H(m=15, n=10, k=3)$.

La función de probabilidad es

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

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

$X = \begin{cases} \text{número de bolas blancas en } k \text{ extracciones} \\ \text{sin reposición de una urna conm} \\ \text{bolas blancas y n negras.} \end{cases} ; H(m,n,k)$	
$D_X = \{x \in \mathbb{N} \mid \max\{0, k-n\} \leq x \leq \min\{m,k\}\}$	
$P_X(x) = P(X = x) = \begin{cases} \frac{\binom{m}{x}\cdot\binom{n}{k-x}}{\binom{m+n}{k}}, & \text{si } \max\{0, k-n\} \leq x \leq \min\{m,k\}, \\ 0, & \text{en otro caso.} \end{cases}$	
$F_X(x) = P(X \leq x)$.	
$E(X) = \frac{k\cdot m}{m+n}; Var(X) = k \cdot \frac{m}{m+n} \cdot \left(1 - \frac{m}{m+n}\right) \cdot \frac{m+n-k}{m+n-1}$	

8.7.2 Ejemplo clásico urna $m = 15$ blancas, $n = 10$ rojas y $k = 3$ extracciones sin reposición.

La probabilidad de sacar 2 blancas será

$$P(X = 2) = \frac{\binom{15}{2} \cdot \binom{10}{3-2}}{\binom{25}{3}}$$

```
c(choose(15,2), choose(10,1), choose(25,3))
```

```
[1]  105   10 2300
```

$P(X = 2) = \frac{105\cdot 10}{2300} = 0.4565217$.

La probabilidad de que saquemos más de 1 bola blanca es

$$\begin{aligned} P(X > 1) &= 1 - P(X \leq 1) = 1 - (P(X = 0) + P(X = 1)) \\ &= 1 - \left(\frac{\binom{15}{0}\cdot\binom{10}{3}}{\binom{25}{3}} + \frac{\binom{15}{1}\cdot\binom{10}{2}}{\binom{25}{3}} \right) \\ &= 1 - \left(\frac{1\cdot 120}{2300} + \frac{15\cdot 45}{2300} \right) = 1 - \frac{120+15\cdot 45}{2300} = 0.6543478. \end{aligned}$$

El número esperado de bolas blancas extraídas para una v.a. X $H(m = 15, n = 10, k = 3)$ 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 $+\sqrt{Var(X)} = +\sqrt{0.66} = 0.812404$.

8.7.3 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
```

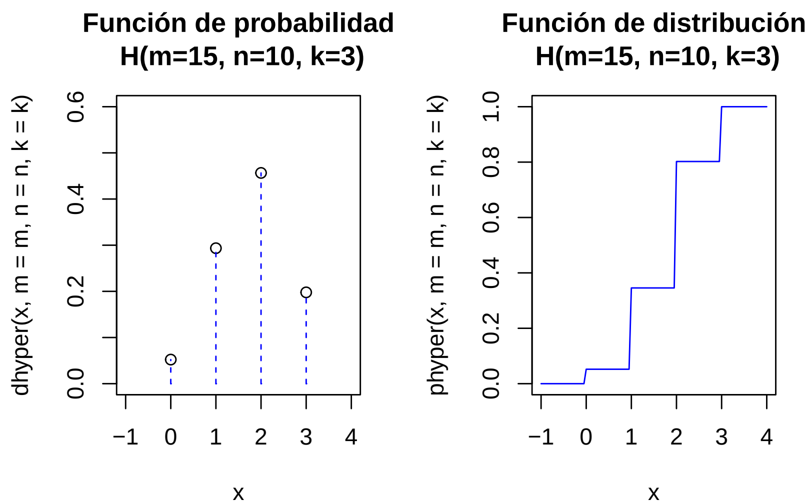
```
1-phyper(q=1,m=15,n=10,k=3)
```

```
[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
```



8.7.4 Cálculos con python

Sea X una $H(m, n, k)$, las funciones de `scipy.stats` cambian los parámetros

- 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 $N = k$.

```python
from scipy.stats import hypergeom
```

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

```
0.2934782608695652
```

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

```
0.3456521739130434
```

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

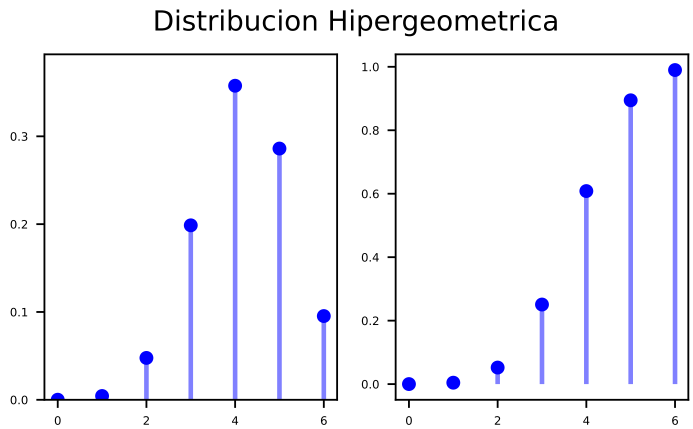
```
0.6543478260869566
```

Una muestra aleatoria de este experimento sería…

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

```
array([2, 2, 2, 2, 1, 1, 1, 2, 2, 1, 0, 2, 1, 1, 2, 1, 3, 1, 2, 0, 2, 2,
       1, 2, 2, 2, 2, 2, 2, 3, 1, 1, 2, 2, 1, 2, 1, 1, 1, 2, 1, 1, 1, 2,
       2, 2, 2, 2, 1, 3, 1, 2, 1, 0, 2, 2, 2, 1, 3, 1, 2, 3, 2, 2, 2, 1,
       2, 3, 3, 3, 1, 2, 3, 2, 1, 3, 1, 2, 2, 2, 2, 1, 2, 1, 2, 2, 0, 2,
         1, 3, 2, 0, 2, 1, 2, 0, 2, 2, 3, 3], dtype=int64)
```

```
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()
```


Distribucion Hipergeometrica
0.0
0.1
0.2
0.3
0
2
4
6
0.0
0.2
0.4
0.6
0.8
1.0
0
2
4
6

9 Distribuciones notables 2. Notables Continuas.

9.1 Introducción

En esta segunda parate del tema de distribucione notable veremos las distribuciones continuas másusuales: uniforme, exponencial y normal.

El lector debe tener en cuenta que exiten muchas otras distribuciones que deberá estudiar y algunas ya las hemos visto como la χ^2

9.2 Distribución uniforme

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.

$$\int_{-\infty}^{+\infty} f_x(x)\cdot dx = \int_a^b \frac{1}{b-a}\cdot dx = \left.\frac{x}{b-a}\right]_{x=a}^{x=b} = \frac{b}{b-a}-\frac{a}{b-a}=\frac{b-a}{b-a}=1.$$

Ahora calculemos su función de distribución

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

- 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 $U(a, b)$.

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 $E(X^2)$.

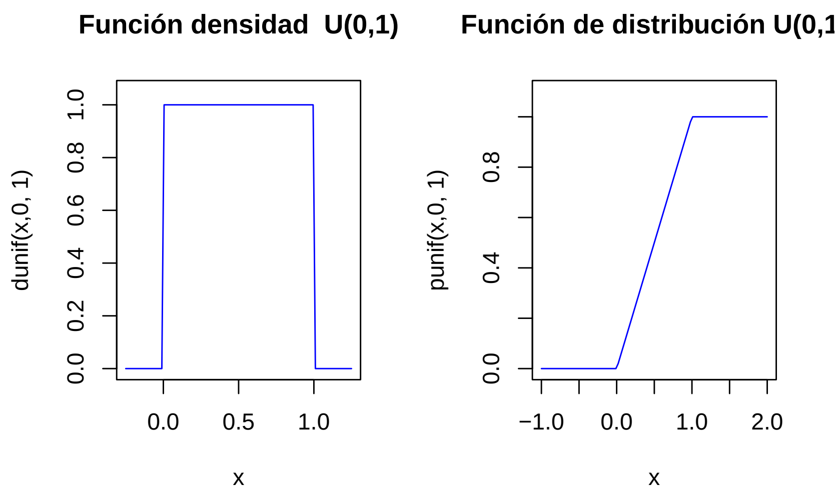
Calculemos ahora $Var(X)$.

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

9.2.1 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))
```



9.2.2 Transformación lineal de la v.a. uniforme

Si X sigue una distribución $U(a, b)$ entonces $Z = \frac{X-a}{b-a}$ sigue una distribución $U(0, 1)$.

Propiedad: Transformación lineal de la v.a. uniforme

Sea X una v.a $U(a, b)$

Si $scale \neq 0$ y loc son dos constantes reales entonces

- si $scale > 0, T = scale \cdot X + loc$ sigue una ley $U(scale \cdot a + loc, scale \cdot b + loc)$
- si $scale < 0, T = scale \cdot X + loc$ sigue una ley $U(scale \cdot b + loc, scale \cdot a + loc)$

Demostración

Supongamos que X sigue una ley $U(a, b)$, que $scale > 0$ y que $T = scale \cdot X + loc$. Dejamos el caso $scale < 0$ como ejercicio.

La función de distribución de X es:

$$F_X(x) = P(X \leq x) = \begin{cases} 0 & \text{si } x \leq a \\ \frac{x-a}{b-a} & \text{si } a \leq x \leq b \\ 1 & \text{si } b \leq x \end{cases}$$

Distribución uniforme $U(a,b)$
Dominio $D_X=(a,b)$
$f_X(x)\begin{cases}\frac{1}{b-a}, & \text{si } a<x<b,\\ 0, & \text{en cualquier otro caso.}\end{cases}$
$F_X(x)=P(X\leq X)=\begin{cases}0, & \text{si } x\leq a\\ \frac{x-a}{b-a}, & \text{si } a\leq x\leq b,\\ 1, & \text{si } b\leq x.\end{cases}$
$E(X)=\frac{a+b}{2}; Var(X)=\frac{(b-a)^2}{12}$

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

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

por lo que T sigue una ley $U(scale\cdot a+loc, scale\cdot b+loc)$.

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?

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

9.2.4 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] -0.2204948 -0.3947785  0.1296429 -0.2226486  0.5411477
```

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] 0.9721063 0.1612653 0.6459652 0.2367076 0.5451255
```

9.2.5 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:

```python
from scipy.stats import uniform
uniform.pdf(0.5,loc=-1,scale=2)
```

```
0.5
```

```python
uniform.ppf(0.5,loc=-1,scale=2)
```

```
0.0
```

Para generar una muestra de valores aleatorios, hay que usar la función `uniform.rvs`:

```python
uniform.rvs(size=30,loc=-1,scale=2)
```

```
array([-0.7623501 ,  0.22430153, -0.32348461,  0.6272097 , -0.402084  ,
        0.75731113, -0.2666446 ,  0.97386339,  0.63758113, -0.7786142 ,
        0.55934524,  0.08008232,  0.67605919, -0.45275724,  0.17331332,
        0.93083844, -0.79498604,  0.90215401,  0.44268231,  0.17866161,
        0.74578496, -0.69757779,  0.11416953,  0.66385692,  0.15291337,
       -0.06157477,  0.68600145, -0.12787417,  0.26643886,  0.94353823])
```

Los valores de los parámetros por defecto son `loc=0, scale=1`:

```python
uniform.pdf(0.5)
```

```
1.0
```

```python
uniform.ppf(0.5)
```

```
0.5
```

```python
uniform.rvs(size=5)
```

```
array([0.79015385, 0.05686863, 0.64363505, 0.11133606, 0.87299565])
```

9.3 Cuantiles de variables aleatorias

Definición Cuantiles

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

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

En 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 $p = 0.3$ 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 $F_X(2) = P(X \leq 2) = 0.5 \geq 0.3$.

Cálculo de cuantiles

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

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

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

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

Ejemplo: cuantiles en un 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)} els
  {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, construimos 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)))))}}
)
}
```

Efectivamente los cuantiles del dado X son

```r
qdado(1.5)
```

```
[1] NA
```

```r
qdado(-1)
```

```
[1] NA
```

```r
qdado(c(0.1,0.5,0.6,1,1.01,2))
```

```
[1]  1  3  3  6 NA NA
```

Ejemplo: Cuantiles Binomial

Por ejemplo si X es una $B(n=10, p=0.3)$

```r
set.seed(2222)
(q=runif(10,0,1))
```

```
 [1] 0.36765818 0.18187591 0.82617679 0.58497444 0.95886983 0.10179894
 [7] 0.75688767 0.24369144 0.67806543 0.06275295
```

```r
qbinom(q,10,0.3)
```

```
 [1] 2 2 4 3 6 1 4 2 4 1
```

```r
set.seed(2222)
rbinom(10,10,0.3)
```

```
 [1] 2 2 4 3 6 1 4 2 4 1
```

Por ejemplo si X es una $BN(n=3, p=0.1)$

```r
set.seed(2222)
(q=runif(10,0,1))
```

```
 [1] 0.36765818 0.18187591 0.82617679 0.58497444 0.95886983 0.10179894
 [7] 0.75688767 0.24369144 0.67806543 0.06275295
```

```r
qnbinom(q,3,0.1)
```

```
 [1] 19 12 41 27 61  9 36 15 32  7
```

```r
set.seed(2222)
rnbinom(10,3,0.1)
```

```
 [1] 18  9  6 46 66 49 24 44 19 26
```

9.4 Distribución exponencial

La distribución exponencial está asociadad al tiempo que transcurre entre dos eventos Poisson consecutivos.

Supongamos que tenemos un proceso Poisson con parámetro λ en una unidad de tiempo.

Dado un tiempo t, definimos N_t como el número de eventos en el intervalo de tiempo $(0, t]$. La distribución de N_t es una $Po(\lambda \cdot t)$. Consideremos la v.a. T como el tiempo transcurrido entre dos eventos Poisson consecutivos.

Sea $t > 0$, entonces

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

Tomando complementarios, la función de distribución de T será:

$$F_T(t) = P(T \leq t) = 1 - P(T > t) = \begin{cases} 0, & \text{si } t \leq 0, \\ 1 - e^{-\lambda t}, & \text{si } t > 0, \end{cases}$$

Para hallar la función de densidad de T, basta derivar la expresión anterior:

$$f_T(t) = \begin{cases} \lambda \cdot e^{-\lambda t}, & \text{si } t > 0, \\ 0, & \text{si } t \leq 0. \end{cases}$$

Llamaremos a la variable T exponencial de parámetro λ y la denotaremos por $Exp(\lambda)$.

9.4.1 Propiedad de la falta de memoria

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,

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

Ejemplo: El clásico problema del peluquero.

Una pequeña peluquería es regentada por un único peluquero. El peluquero está esperando al próximo cliente mientras lee el periódico.

Supongamos que $N_T =$ 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.

En este ejemplo la propiedad de la pérdida de memoria significa que 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 $P(T > 0.25 + 1/6 | T > 0.25) = P(T > 1/6)$.

El tiempo esperado (en horas) hasta el siguiente cliente es

$$E(X) = \frac{1}{\lambda} = \frac{1}{4} = 0.25.$$

y la varianza es

$$Var(X) = \frac{1}{\lambda^2} = \frac{1}{4^2} = 0.0625.$$

Por último ¿Cuál es la probabilidad de que nuestro peluquero esté sin clientes (leyendo el periódico) más de 30 minutos (0.5 horas)?

$$P(X > 0.5) = 1 - P(X \leq 0.5) = 1 - (1 - e^{-4 \cdot 0.5}) = e^{-2} = 0.1353353.$$

Si queremos hacer los cálculos con R,

```
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
```

9.4.2 Cálculos con R y python

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)# no es una probabilidad es una densidad y puede ser >1
```

```
[1] 2.991013
```

```
pexp(0.5,rate=3) # P(X<0.5)
```

```
[1] 0.7768698
```

```
rexp(8,3)# ocho tiempos de una exponencial
```

```
[1] 0.5069426 0.4497573 0.2876943 0.5514840 1.0552252 0.3168070 0.2488148
[8] 0.2377065
```

Y en python con:

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

```
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)
```

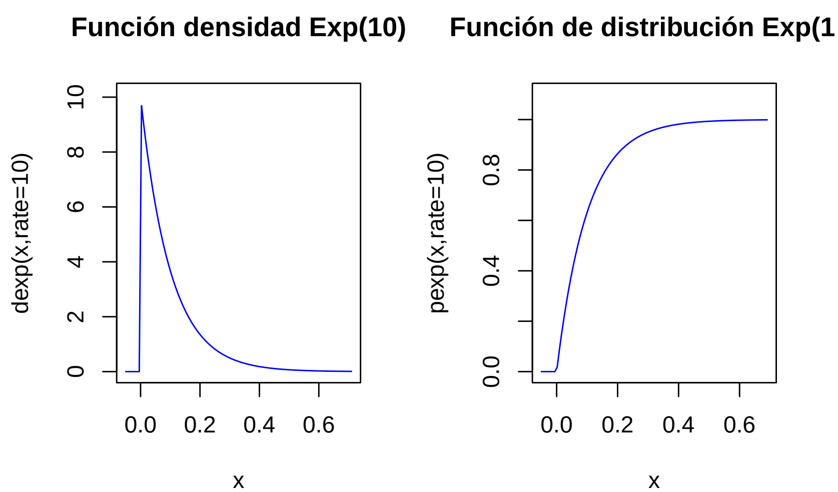
```
array([0.31900047, 0.57918153, 0.07598563, 0.05315   , 0.01322694,
       1.45092053, 0.63475084, 0.03854183, 0.3378667 , 0.70529797])
```

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

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

```
lambda=10
par(mfrow=c(1,2))
curve(dexp(x,rate=lambda)
      xlim=c(-0.05,round(qexp(0.99,rate=lambda,2),2)+0.25),
```

```
      ylim=c(0,dexp(0,lambda)+0.1),col="blue",
      main=paste0("Función densidad Exp(",lambda,")"),
      ylab=paste0("dexp(x,rate=",lambda,")"))
curve(pexp(x,rate=lambda),xlim=c(-0.05,qexp(0.999,10)),
      ylim=c(0,1.1),col="blue",
      main=paste0("Función de distribución Exp(",lambda,")"),
      ylab=paste0("pexp(x,rate=",lambda,")"))
par(mfrow=c(1,1))
```



Ejercicio

Consultad en el manual de python scipy.stats.

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

Ejercicio: las bombillas que no envejecen

Supongamos que compramos una bombilla led que promete un **valor esperado** de duración de 10000 (1.14 años) horas de funcionamiento continuo. Además, nos aseguran que la distribución de X, el número de horas de funcionamiento continuo de una bombilla led, sigue una ley exponencial.

- Si X es $Exp(\lambda)$ ¿cuál es el valor del parámetro λ?.
- ¿Cuál es la probabilidad de que una bombilla led ilumine más de 2 años?
- Supongamos que ya tengo una bombilla led funcionando 1 año ¿Cuál es la probabilidad de que dure dos años más?
- ¿Cuál es la varianza de la duración en horas de este tipo de bombillas?

9.5 Distribución normal o Gaussiana

Una de las variables aleatorias continua más populares es la llamada distribución normal o Gaussiana
.

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

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

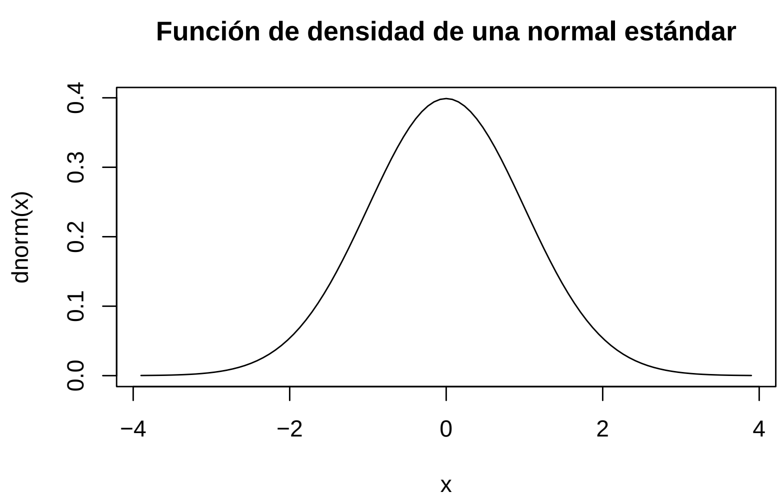
para todo $x \in \mathbb{R}$.

La gráfica de esta función de densidad es conocida como **campana de Gauss.**

La v.a. normal con $\mu = 0$ y $\sigma = 1$ recibe el nombre de normal estándar y se suele denotar por la letra Z normal $N(0, 1)$. El siguiente código la dibuja.

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

9.6 Gráfica distribución normal o Gaussiana



X sigue una distribución $N(\mu,\sigma)$
$D_X = \mathbb{R} = (-\infty,+\infty)$
$f_X(x) = \frac{1}{\sqrt{2\pi\cdot\sigma^2}}\cdot e^{\frac{-(x-\mu)^2}{2\cdot\sigma^2}}$ para todo $x\in\mathbb{R}$.
$F_X(x) = P(X\leq X) = \int_{-\infty}^{x} f_X(t)dt.$
$E(X)=\mu; Var(X)=\sigma^2.$

9.6.1 Propiedades de la función de densidad de la distribución normal

Propiedades de la función de densidad de la distribución normal

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

- La función f_X verifica todas las propiedades de las funciones de densidad: $f_X(x)>0$, para todo $x\in\mathbb{R}$ y $\displaystyle\int_{-\infty}^{\infty} f_X(x)\,dx = 1$.
- La función $f_X(x)$ es simétrica respecto de la recta $x=\mu$: $f_X(\mu-x)=f_X(\mu+x)$, para todo $x\in\mathbb{R}$.
- f_X tiene un único máximo absoluto en $x=\mu$ que vale $f_X(\mu)=\frac{1}{\sqrt{2\pi\sigma^2}}$.
- Si F_X es la función de distribución de X, entonces $F_X(\mu+x)=1-F_X(\mu-x)$, para todo $x\in\mathbb{R}$.
- En particular si Z es una $N(0,1)$ entonces $F_Z(-x)=1-F_Z(x)$, para todo $x\in\mathbb{R}$.
- $Z=\frac{X-\mu}{\sigma}$ es una v.a. $N(0,1)$ y $X=\sigma\cdot Z+\mu$ es una $N(\mu,\sigma)$ donde Z es la normal estándar.

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

La función $F(x)$ no tiene ninguna expresión algebraica "decente". Es por esta razón, y por comodidad, que esta función está tabulada o hay que calcularla usando un software estadístico.

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

9.6.3 Cálculos con R

Las funciones que calculan la función de densidad y de distribución de una variable $N(\mu,\sigma)$ en un valor x son `dnorm(x,mean=mu,sd=sigma)` y `pnorm(x,mean=mu,sd=sigma)`, respectivamente. Por ejemplo, para una variable $X\sim N(\mu=1,\sigma=2)$ la función de densidad $f_X(2)$ se puede calcular de la forma siguiente:

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

```
[1] 0.1760327
```

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

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

```
[1] 0.6914625
```

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]  0.9806747  5.5415845  2.8174087 -2.4085639  0.8920435
```

9.6.4 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([ 1.27851168,  2.21598804, -2.02721711,  1.82936516,  4.1621825 ])
```

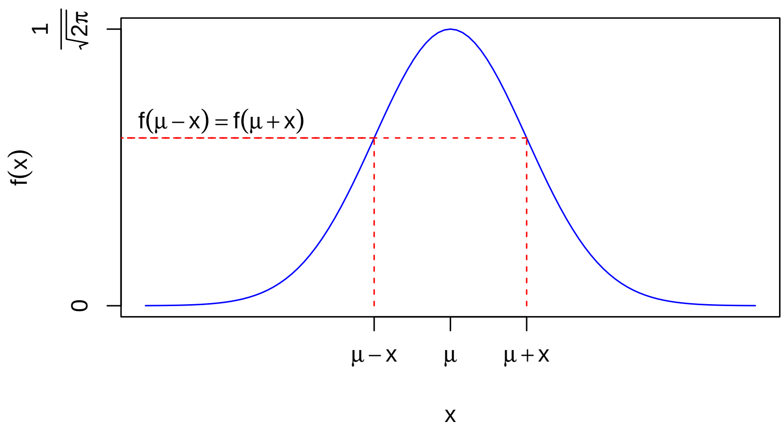
Consultad SciPy.org para dibujar las funciones de densidad y de distribución con python.

9.6.5 Resumen de la distribución normal.

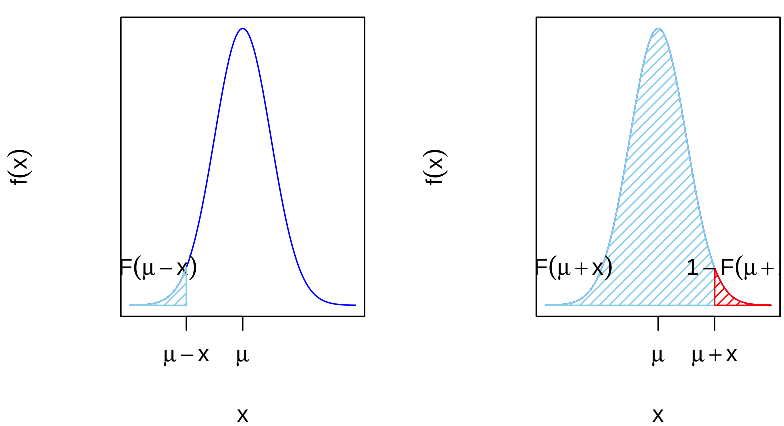
Propiedades de la función de densidad de la distribución normal

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).
- $f(\mu + x) = f(\mu - x)$.
- $F(\mu - x) = 1 - F(\mu + x)$.



- $\lim\limits_{x \to +\infty} f(x) = \lim\limits_{x \to -\infty} f(x) = 0$ 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 $x = \mu - \sigma$.



9.6.6 Transformaciones lineales de variables aleatorias normales

Propiedad: transformación lineal la distribución normal

Sea X una variable $N(\mu, \sigma)$ entonces la variable $Y = aX + b$ con $a \neq 0, b \in \mathbb{R}$ tiene distribución $N(a\mu + b, |a|\sigma)$

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:

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

9.6.7 Propiedades de la distribución normal estándar

Proiedades normal estándar

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

En este caso, $\mu = 0$ y $\sigma = 1$. Podemos escribir algunas de las propiedades vistas para una distribución normal cualquiera de la forma siguiente:

- La propiedad $f_X(\mu - x) = f_X(\mu + x)$ se traduce a $f_Z(-x) = f_Z(x)$
- La propiedad $F_X(\mu - x) = 1 - F_X(\mu + x)$ se traduce a $F_Z(-x) = 1 - F(x)$.
- Dado $\delta > 0$,

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

Ejercicio

Cálculos con la distribución normal estándar

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

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

- $P(Z \geq -2)$.

Resolución:

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

9.6.8 Relación entre una distribución normal y la normal estándar.

Para hallar la probabilidad de que X esté en un intervalo (a, b) cualquiera, podemos usar la función de distribución de Z de la siguiente manera:

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

Para el caso particular en que el intervalo esté centrado en la media μ, o sea existe un valor $\delta > 0$ tal que $(a, b) = (\mu - \delta, \mu + \delta)$, obtenemos:

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

9.6.9 Ejemplo cálculo probabilidades normal

Ejercicio

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

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

Solución

La primera probabilidad se calcula de la forma siguiente:

$$\begin{aligned} P(1 < X < 2) & = P\left(\tfrac{1-2}{2} < \tfrac{X-2}{2} < \tfrac{2-2}{2}\right) = P\left(\tfrac{-1}{2} < Z < 0\right) \\ & = F_Z(0) - F_Z(-0.5) = \tfrac{1}{2} - 1 + F_Z(0.5) = -\tfrac{1}{2} + F_Z(0.5). \end{aligned}$$

La segunda probabilidad se calcular de la forma siguiente:

$$P(X > 3) = P\left(\frac{X-2}{2} > \frac{3-2}{2}\right) = P(Z > 0.5) = 1 - F_Z(0.5).$$

Ejercicio

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

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

Solución con R

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

```
[1] 0.1914625
```

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

```
[1] 0.3085375
```

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

```
[1] 0.3085375
```

Solución con python

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

```
0.19146246127401312
```

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

```
0.3085375387259869
```

9.7 La distribución normal aproxima otras distribuciones

En los temas que siguen veremos como, bajo determinadas condiciones,

- la distribución normal puede aproximar la distribución binomial,
- la distribución normal puede aproximar la distribución Poisson
- la distribución normal es la distribución límite de la media aritmética de una muestra de variables aleatorias.