

# Contents

<b>Presentación</b>	<b>1</b>
Objetivo . . . . .	1
<b>1 Probabilidad y Experimentos aleatorios</b>	<b>2</b>
1.1 Problema 1 . . . . .	2
1.2 Problema 2 . . . . .	2
1.3 Problema 3 . . . . .	5
1.4 Problema 4 . . . . .	5
1.5 Problema 5 . . . . .	7
<b>2 Variables aleatorias y Distribuciones de probabilidad</b>	<b>8</b>
2.1 Ejercicio 2.1 . . . . .	8
2.2 Ejercicio 2.2 . . . . .	10
2.3 Ejercicio 2.3 . . . . .	13
2.4 Ejercicio 2.4 . . . . .	18
2.5 Ejercicio 2.5 . . . . .	19
2.6 Ejercicio 2.6 . . . . .	21
2.7 Ejercicio 2.7 . . . . .	24
2.8 Ejercicio 2.8 . . . . .	26
2.9 Ejercicio 2.9 . . . . .	26
<b>3 Distribuciones de probabilidad multidimensionales</b>	<b>26</b>
3.1 Ejercicio 1 . . . . .	26
3.2 Ejercicio 2 . . . . .	26
3.3 Ejercicio 3 . . . . .	26
3.4 Ejercicio 4 . . . . .	27
3.5 Ejercicio 5 . . . . .	27
3.6 Ejercicio 6 . . . . .	27
3.7 Ejercicio 7 . . . . .	28
3.8 Ejercicio 8 . . . . .	28

## Presentación

### Objetivo

El objetivo de estos ejercicios es proporcionar unos materiales de soporte para la asignatura de “Inferencia Estadística” del Máster interuniversitario de Bioestadística y Bioinformática impartido conjuntamente por la Universitat Oberta de Catalunya (UOC) y la Universidad de Barcelona (UB).

Esta asignatura adolece de las características habituales de las asignaturas de posgrado, y especialmente de un posgrado de estadística (y bioinformática), que muestran algunas de las cosas que no debe de ser esta asignatura:

Tal como se indica en la introducción a las notas de soporte del curso, este debería:

- Servir para repasar y consolidar los conceptos básicos que la mayoría de estudiantes traerán consigo.
- Además, y sobretodo, debe proporcionar una visión general, lo más completa posible dentro de las limitaciones de tiempo, del campo de la inferencia estadística

Y, naturalmente, una de las formas de consolidar conocimientos, como en cualquier disciplina cuantitativa, es a través de la resolución de ejercicios que permiten reflexionar, comprender y ver como se aplican los conceptos teóricos introducidos.

Para ello, estos materiales contienen una serie de ejercicios similares a los que se proponen en las actividades y pruebas de evaluación continua de la asignatura.

La mayoría de los ejercicios están resueltos, pero *es importante intentar resolverlos de forma autónoma antes de consultar la solución*.

En general los ejercicios no presuponen ningún conocimiento especial de matemáticas, más allá de las habilidades básicas que se adquieren durante los estudios de una carrera de ciencias o de ingeniería.

## 1 Probabilidad y Experimentos aleatorios

### 1.1 Problema 1

Sean  $A$  y  $B$  dos sucesos. Suponiendo que  $P(A) = 0.3$ ,  $P(B) = 0.6$ , y  $P(A \cap B) = 0.1$ , calcula las siguientes probabilidades:

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

#### 1.1.1 Solución

- $P(A \cup B) = P(A) + P(B) - P(A \cap B) = 0.3 + 0.6 - 0.1 = 0.8$
- $P(A^c) = 1 - P(A) = 1 - 0.3 = 0.7$
- $P(A^c \cap B) = P(B) - P(A \cap B) = 0.6 - 0.1 = 0.5$
- $P(A \cap B^c) = P(A) - P(A \cap B) = 0.3 - 0.1 = 0.2$
- $P(A^c \cap B^c) = 1 - P(A \cup B) = 1 - 0.8 = 0.2$

### 1.2 Problema 2

Una población está afectada por tres enfermedades diferentes  $A$ ,  $B$  i  $C$ . La probabilidad de que una persona sufra  $A$  es 0.30, la probabilidad de que sufra  $B$  es 0.20 y la probabilidad de que sufra  $C$  es 0.15. La probabilidad de que una persona sufra  $A$  y  $B$  es 0.12, la que sufra  $A$  y  $C$  es 0.09 y la que sufra  $B$  y  $C$  es 0.06. La probabilidad de que una persona sufra las tres enfermedades es 0.03. Se piden las probabilidades de que una persona escogida al azar:

1. padezca al menos una enfermedad
2. sólo sufra  $A$
3. sufra  $B$  o  $C$ , pero no sufra  $A$
4. sufra  $A$  o no sufra ni  $B$  ni  $C$ .

### 1.2.1 Solución

---

a) **¿Cuál es la probabilidad de que una persona padezca al menos una enfermedad?**

Queremos calcular la probabilidad de que una persona sufra al menos una de las tres enfermedades, es decir,  $P(A \cup B \cup C)$ .

Para calcular  $P(A \cup B \cup C)$ , usamos la regla de inclusión-exclusión:

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

Sustituyendo los valores dados en el enunciado:

$$P(A \cup B \cup C) = 0.30 + 0.20 + 0.15 - 0.12 - 0.09 - 0.06 + 0.03 = 0.41$$

Por lo tanto, la probabilidad de que una persona padezca al menos una enfermedad es **0.41**.

---

b) **¿Cuál es la probabilidad de que una persona solo sufra  $A$ ?**

Para resolver esto, necesitamos calcular la probabilidad de que la persona sufra  $A$ , pero no  $B$  ni  $C$ , es decir,  $P(A \cap B^c \cap C^c)$ .

Podemos calcular  $P(A \cap B^c \cap C^c)$  restando de  $P(A)$  la probabilidad de que la persona sufra  $A$  junto con alguna de las otras dos enfermedades:

$$P(A \cap B^c \cap C^c) = P(A) - P(A \cap B) - P(A \cap C) + P(A \cap B \cap C)$$

Sustituyendo los valores:

$$P(A \cap B^c \cap C^c) = 0.30 - 0.12 - 0.09 + 0.03 = 0.12$$

Por lo tanto, la probabilidad de que una persona solo sufra  $A$  es **0.12**.

---

c) **¿Cuál es la probabilidad de que una persona sufra  $B$  o  $C$ , pero no sufra  $A$ ?**

Aquí buscamos la probabilidad  $P(A^c \cap (B \cup C))$ , es decir, la probabilidad de que la persona no tenga  $A$ , pero tenga  $B$  o  $C$ .

Primero, calculamos  $P(B \cup C)$  utilizando la regla de inclusión-exclusión:

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

Sustituyendo los valores:

$$P(B \cup C) = 0.20 + 0.15 - 0.06 = 0.29$$

Ahora, para calcular  $P(A^c \cap (B \cup C))$ , restamos de  $P(B \cup C)$  la probabilidad de que la persona tenga  $A$  y alguna de las enfermedades  $B$  o  $C$ , es decir,  $P(A \cap (B \cup C))$ :

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

Sustituyendo los valores:

$$P(A \cap (B \cup C)) = 0.12 + 0.09 - 0.03 = 0.18$$

Finalmente, restamos:

$$P(A^c \cap (B \cup C)) = P(B \cup C) - P(A \cap (B \cup C)) = 0.29 - 0.18 = 0.11$$

Por lo tanto, la probabilidad de que una persona sufra  $B$  o  $C$ , pero no  $A$ , es **0.11**.

---

d) **¿Cuál es la probabilidad de que una persona sufra  $A$  o no sufra ni  $B$  ni  $C$ ?**

Aquí buscamos la probabilidad  $P(A \cup (B^c \cap C^c))$ , es decir, que la persona sufra  $A$  o que no sufra ni  $B$  ni  $C$ .

Primero, calculamos  $P(B^c \cap C^c)$ , que es la probabilidad de que la persona no sufra ni  $B$  ni  $C$ . Esto es simplemente  $1 - P(B \cup C)$ , que ya calculamos previamente:

$$P(B^c \cap C^c) = 1 - P(B \cup C) = 1 - 0.29 = 0.71$$

Ahora, aplicamos la regla de la unión para calcular  $P(A \cup (B^c \cap C^c))$ :

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

Ya calculamos  $P(B^c \cap C^c)$ , y sabemos que  $P(A \cap B^c \cap C^c)$  es la probabilidad de que una persona solo sufra  $A$ , que también calculamos previamente:

$$P(A \cap B^c \cap C^c) = 0.12$$

Sustituyendo los valores:

$$P(A \cup (B^c \cap C^c)) = 0.30 + 0.71 - 0.12 = 0.89$$

Por lo tanto, la probabilidad de que una persona sufra  $A$  o no sufra ni  $B$  ni  $C$  es **0.89**.

**Resumiendo:**

1. La probabilidad de que una persona padezca al menos una enfermedad es **0.41**.
2. La probabilidad de que una persona solo sufra  $A$  es **0.12**.
3. La probabilidad de que una persona sufra  $B$  o  $C$ , pero no  $A$ , es **0.11**.
4. La probabilidad de que una persona sufra  $A$  o no sufra ni  $B$  ni  $C$  es **0.89**.

### 1.3 Problema 3

Por los síntomas observados en un enfermo, y según la experiencia acumulada en un gran número de situaciones similares, se deduce que ha podido coger la enfermedad  $A$  con probabilidad  $1/3$ , o la enfermedad  $B$  con probabilidad  $2/3$ . Con el fin de precisar el diagnóstico, se hace un análisis clínico al enfermo con dos resultados posibles, positivo o negativo. Se sabe, también por experiencia, que en los pacientes que tienen la enfermedad  $A$  el análisis es positivo con probabilidad  $0.99$ , y en los que padecen la enfermedad  $B$  lo es con probabilidad  $0.06$ .

- a) ¿Cuál es la probabilidad de que el análisis dé un resultado negativo?
- b) Si el resultado ha sido positivo, ¿cuál es la probabilidad de que el paciente sufra la enfermedad  $A$ ? ¿Y la probabilidad de que padezca la enfermedad  $B$ ?

#### 1.3.1 Solución

a.

$$\begin{aligned} P(Neg) &= P(Neg|A) \cdot P(A) + P(Neg|B) \cdot P(B) = \\ &= 0.01 \cdot 1/3 + 0.94 \cdot 2/3 = 0.63 \end{aligned}$$

b.

$$\begin{aligned} P(A|Pos) &= \frac{P(Pos|A)P(A)}{P(Pos)} = 0.8919, \quad \text{para } A, \\ P(B|Pos) &= 1 - P(A|Pos) = 0.1081, \quad \text{para } B. \end{aligned}$$

Las probabilidades las hemos calculado con R a partir de la información del enunciado:

```
pA<-1/3
pB<-2/3
ppA<-0.99
ppB<-0.06
pn<-(1-ppA)*pA+(1-ppB)*pB
pn
```

```
## [1] 0.63
```

### 1.4 Problema 4

El embolismo pulmonar es una condición relativamente común que necesita hospitalización y que a menudo ocurre en pacientes hospitalizados. La presión arterial menor de 90 mm HG es uno de los criterios importantes para diagnosticar esta condición. Supongamos que la sensibilidad del test es del 95% y la especificidad del test es del 75% y la prevalencia es del 20%.

- a) Calcula el valor predictivo positivo del test.
- b) Calcula el valor predictivo negativo del test.
- c) Responde a las preguntas anteriores si la prevalencia fuera del 80%.

### 1.4.1 Solución

a) Calcula el valor predictivo positivo del test

$$VP+ = P(\text{Embolismo} / \text{Test} +) = \frac{\text{Sens} \times \text{Prev}}{\text{Sens} \times \text{Prev} + (1 - \text{Esp})(1 - \text{Prev})}$$

```
sens<-0.95
esp<-0.75
prev<-0.20
vpp<-(sens*prev)/(sens*prev+(1-esp)*(1-prev))
vpp
```

```
## [1] 0.4871795
```

b) Calcula el valor predictivo negativo del test

$$VP- = \frac{\text{Esp}(1 - \text{Prev})}{\text{Esp}(1 - \text{Prev}) + (1 - \text{Sens}) \text{Prev}}$$

```
vpn<-(esp*(1-prev))/(esp*(1-prev)+(1-sens)*prev)
vpn
```

```
## [1] 0.9836066
```

Como se observa al tratarse de una prueba muy sensible y poco específica hay pocos falsos negativos y cuando el test da negativo hay una probabilidad muy alta (0.98) de que el individuo sea sano. No así cuando da positivo. Sólo el 48% serán verdaderos enfermos.

c) Responde a las preguntas anteriores si la prevalencia fuera del 80%

```
prev<-0.80
vpp<-(sens*prev)/(sens*prev+(1-esp)*(1-prev))
vpp
```

```
## [1] 0.9382716
```

```
vpn<-(esp*(1-prev))/(esp*(1-prev)+(1-sens)*prev)
vpn
```

```
## [1] 0.7894737
```

Si la prevalencia es más alta, el VP- sigue siendo alto, aunque no tanto pero hemos aumentado el VP+ hasta el 93% y no habrá tantos falsos positivos. Lo que está claro es el VPN y el VPP dependen de la prevalencia de la enfermedad.

## 1.5 Problema 5

Un índice que evalúa el síndrome de la muerte súbita (SMS) tiene una sensibilidad del 68% y una especificidad del 82%. ¿Cuáles son los valores predictivos positivo y negativo del índice si se aplica a una población donde se producen un 0,21% de muertes súbitas sobre el total de nacimientos?

### 1.5.1 Solución

La prevalencia del síndrome de la muerte súbita en la población es del 0.21%, es decir 0.0021.

Nos piden que calculemos respectivamente los valores predictivos positivo y negativo del test. Es decir, que tan bien funciona el test para detectar la enfermedad ( $SMS$ ) cuando da un resultado positivo ( $T+$ ) y para indicar su ausencia ( $SMS^c$ ), mediante un resultado negativo ( $T-$ ).

$$VP+ = P[SMS|T+], \quad VP- = P[SMS^c|T-],$$

Puede hacerse el cálculo directamente a partir de las probabilidades condicionadas.

$$\begin{aligned} VP+ &= P[SMS|T+] = \frac{P[T+|SMS] \times P[SMS]}{P[T+]} = \\ &= \frac{P[T+|SMS] \times P[SMS]}{P[T+|SMS] \times P[SMS] + P[T+|SMS^c] \times P[SMS^c]} = \\ &= \frac{\text{Sensibilidad} \times \text{Prevalencia}}{\text{Sensibilidad} \times \text{Prevalencia} + 1 - \text{Especificidad} \times 1 - \text{Prevalencia}} \end{aligned}$$

De forma análoga:

$$\begin{aligned} VP- &= P[SMS^c|T-] = \frac{P[T-|SMS^c] \times P[SMS^c]}{P[T-]} = \\ &= \frac{P[T-|SMS^c] \times P[SMS^c]}{P[T-|SMS^c] \times P[SMS^c] + P[T-|SMS] \times P[SMS]} = \\ &= \frac{\text{Especificidad} \times 1 - \text{Prevalencia}}{\text{Especificidad} \times 1 - \text{Prevalencia} + 1 - \text{Sensibilidad} \times \text{Prevalencia}} \end{aligned}$$

Estos cálculos se reañlizan de forma inmediata usando R:

```
sensi <- 0.68
espec <- 0.82
prev <- 0.0021
vp.pos <- (sensi * prev) / (sensi * prev + (1-espec) * (1-prev))
cat ("El valor predictivo positivo es: ", vp.pos)
```

```
## El valor predictivo positivo es: 0.007887324
```

```
vp.neg <- (espec * (1-prev)) / (espec * (1-prev) + (1-sensi) * (prev))
cat ("El valor predictivo negativo es: ", vp.neg)
```

```
## El valor predictivo negativo es: 0.9991794
```

Como en el caso anterior, podemos ver que. al ser la prevalencia muy baja, el valor predicpositivo del test también lo es puesto que un test + tan solo indica en un 0,79% de veces la presencia del síndrome, correctamente.

## 2 Variables aleatorias y Distribuciones de probabilidad

### 2.1 Ejercicio 2.1

Se sabe que la presencia de algunas mutaciones en una región genómica puede influir en la sobreexpresión (“Up”) o la inhibición (“Down”) de dos genes distintos. Se conocen 6 variantes de dicha mutación y, dado que los efectos de la sobreexpresión de los dos genes son muy similares se ha optado por contar únicamente cuántos genes se sobre-expresan en presencia de cada una de ellas (un individuo puede presentar una única variante). Un estudio realizado sobre 300 pacientes ha permitido estimar las siguientes probabilidades de aparición de cada mutación así como el número de genes sobre-expresados asociados a las mismas. Los resultados se encuentran disponibles en la tabla siguiente:

Mutación	Probabilidad	$N^\circ$ de genes
$e_1$	0.15	0
$e_2$	0.13	1
$e_3$	0.07	1
$e_4$	0.30	2
$e_5$	0.20	2
$e_6$	0.15	0

Consideremos la variable aleatoria:  $X$  = “Número de genes sobre expresados”

1. Obtener su distribución de probabilidad y representarla gráficamente
2. Calcular la esperanza y la varianza de dicha variable

### SOLUCIÓN

La variable aleatoria que nos interesa es  $X$  = “Número de genes sobre-expresados”.

#### 2.1.1 Distribución de probabilidad

Para obtener la distribución de probabilidad de  $X$ , necesitamos sumar las probabilidades de las mutaciones que tienen el mismo número de genes sobre-expresados.

Los posibles valores de  $X$  son 0, 1 y 2. A continuación calculamos la probabilidad de cada uno:

- Para  $X = 0$ , las mutaciones son  $e_1$  y  $e_6$ :

$$P(X = 0) = P(e_1) + P(e_6) = 0.15 + 0.15 = 0.30$$

- Para  $X = 1$ , las mutaciones son  $e_2$  y  $e_3$ :

$$P(X = 1) = P(e_2) + P(e_3) = 0.13 + 0.07 = 0.20$$

- Para  $X = 2$ , las mutaciones son  $e_4$  y  $e_5$ :



$$P(X = 2) = P(e_4) + P(e_5) = 0.30 + 0.20 = 0.50$$

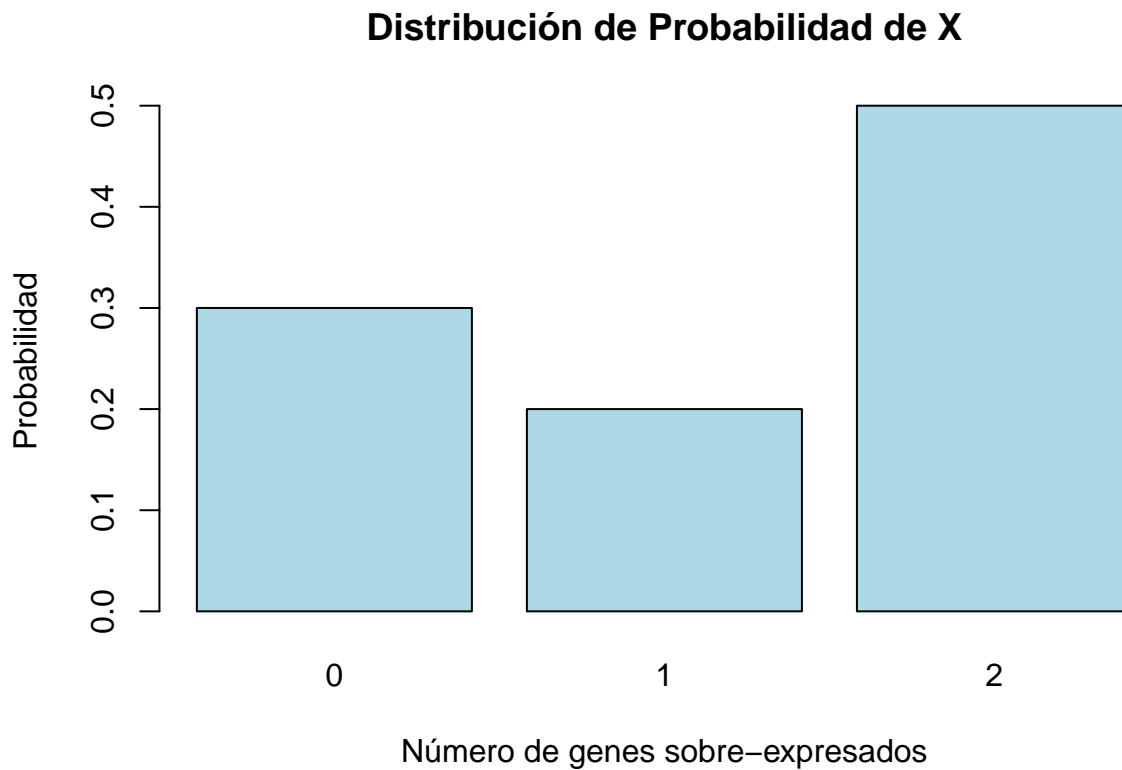
La distribución de probabilidad de  $X$  es la siguiente:

$$P(X = x) = \begin{cases} 0.30 & \text{si } x = 0, \\ 0.20 & \text{si } x = 1, \\ 0.50 & \text{si } x = 2. \end{cases}$$

Podemos representarla gráficamente usando R:

```
# Valores de X y sus probabilidades
X_values <- c(0, 1, 2)
probabilities <- c(0.30, 0.20, 0.50)

# Crear el gráfico
barplot(probabilities, names.arg = X_values, col = "lightblue",
        main = "Distribución de Probabilidad de X",
        xlab = "Número de genes sobre-expresados", ylab = "Probabilidad")
```



### 2.1.2 Esperanza y varianza

La **esperanza** (o valor esperado) de una variable aleatoria discreta  $X$  se calcula como:

$$E(X) = \sum_x x \cdot P(X = x)$$

Sustituyendo los valores:

$$E(X) = 0 \cdot 0.30 + 1 \cdot 0.20 + 2 \cdot 0.50 = 0 + 0.20 + 1.00 = 1.20$$

La **varianza** de  $X$  se calcula como:

$$\text{Var}(X) = E(X^2) - [E(X)]^2$$

Primero calculamos  $E(X^2)$ :

$$E(X^2) = \sum_x x^2 \cdot P(X = x)$$

$$E(X^2) = 0^2 \cdot 0.30 + 1^2 \cdot 0.20 + 2^2 \cdot 0.50 = 0 + 0.20 + 2.00 = 2.20$$

Entonces, la varianza es:

$$\text{Var}(X) = 2.20 - (1.20)^2 = 2.20 - 1.44 = 0.76$$

Verificamos los cálculos con R:

```
# Calcular esperanza y varianza
esperanza <- sum(X_valores * probabilities)
esperanza_cuadrado <- sum(X_valores^2 * probabilities)

varianza <- esperanza_cuadrado - esperanza^2

esperanza
```

```
## [1] 1.2
```

```
varianza
```

```
## [1] 0.76
```

## 2.2 Ejercicio 2.2

Para describir el número de mutaciones presentes en un volumen estándar de un tumor unos investigadores han propuesto el modelo siguiente

$$p(x) = \frac{K}{2+x}, x = 0, 1, 2, 3, 4, 5$$

1. Determinar qué valor debe de tener  $K$  para que  $p(x)$  sea una función de masa de probabilidad
2. Calcular su esperanza y su varianza
3. Calcular las probabilidades de los sucesos:

- 1 Un tumor presenta exactamente tres mutaciones
- 2 Un tumor presenta al menos una mutación
- 3 Un tumor presenta como máximo dos mutaciones.

## SOLUCIÓN

Se considera el modelo para la distribución de probabilidades de mutaciones en un tumor dado por:

$$p(x) = \frac{K}{2+x}, x = 0, 1, 2, 3, 4, 5$$

### 2.2.1 Valor de $K$

Para que  $p(x)$  sea una función de masa de probabilidad, la suma de todas las probabilidades debe ser igual a 1. Es decir:

$$\sum_{x=0}^5 p(x) = 1$$

Sustituyendo la fórmula de  $p(x)$ :

$$\sum_{x=0}^5 \frac{K}{2+x} = 1$$

Simplificamos la suma:

$$K \sum_{x=0}^5 \frac{1}{2+x} = 1$$

La suma es:

$$\sum_{x=0}^5 \frac{1}{2+x} = \frac{1}{2} + \frac{1}{3} + \frac{1}{4} + \frac{1}{5} + \frac{1}{6} + \frac{1}{7}$$

Podemos calcular esta suma numéricamente en R:

```
# Valores de la suma
suma <- sum(1 / (2 + 0:5))

# Calcular el valor de K
K <- 1 / suma
K
```

```
## [1] 0.6278027
```

### 2.2.2 Esperanza y la varianza

La **esperanza** de  $X$  se calcula como:

$$E(X) = \sum_{x=0}^5 x \cdot p(x) = \sum_{x=0}^5 x \cdot \frac{K}{2+x}$$

La **varianza** se calcula usando:

$$\text{Var}(X) = E(X^2) - [E(X)]^2$$

Para esto, primero calculamos  $E(X^2)$ :

$$E(X^2) = \sum_{x=0}^5 x^2 \cdot p(x) = \sum_{x=0}^5 x^2 \cdot \frac{K}{2+x}$$

Podemos calcular la esperanza y la varianza en R de la siguiente forma:

```
# Calcular la esperanza
esperanza <- sum((0:5) * K / (2 + 0:5))

# Calcular la esperanza al cuadrado
esperanza_cuadrado <- sum((0:5)^2 * K / (2 + 0:5))

# Calcular la varianza
varianza <- esperanza_cuadrado - esperanza^2

esperanza
```

```
## [1] 1.766816
```

```
varianza
```

```
## [1] 2.761769
```

### 2.2.3 Probabilidades

**Probabilidad de que un tumor presente exactamente tres mutaciones**

La probabilidad de que  $X = 3$  es:

$$P(X = 3) = p(3) = \frac{K}{2+3}$$

Podemos calcularlo en R:

```
# Probabilidad de X = 3
P_X_3 <- K / (2 + 3)
P_X_3
```

```
## [1] 0.1255605
```

### Probabilidad de que un tumor presente al menos una mutación

La probabilidad de que  $X \geq 1$  es:

$$P(X \geq 1) = 1 - P(X = 0)$$

Podemos calcularlo en R:

```
# Probabilidad de X >= 1
P_X_1 <- 1 - K / (2 + 0)
P_X_1
```

```
## [1] 0.6860987
```

### Probabilidad de que un tumor presente como máximo dos mutaciones

La probabilidad de que  $X \leq 2$  es:

$$P(X \leq 2) = P(X = 0) + P(X = 1) + P(X = 2)$$

Podemos calcularlo en R:

```
# Probabilidad de X <= 2
P_X_2 <- sum(K / (2 + 0:2))
P_X_2
```

```
## [1] 0.6801196
```

## 2.3 Ejercicio 2.3

Un modelo simplificado del tiempo de supervivencia, en años, tras un diagnóstico de una variante de leucemia es el siguiente:

$$f_x(x) = -0.5 \cdot x + 1, \quad \text{donde } 0 \leq x \leq 2$$

1. Comprobar que  $f_X$  es una densidad. Representarla gráficamente.
2. Calcular  $F_X$  y representarla gráficamente.
3. Calcular  $P(X \geq 1)$ ,  $P(X > 1)$ ,  $P(X = 1)$ ,  $f_x(1)$ .
4. Calcular la probabilidad de que un individuo diagnosticado con leucemia sobreviva :  
(i) menos de seis meses, (ii) entre seis meses y un año, (iii) más de dos años.
5. Calcular  $E(X)$  i  $\text{Var}(X)$ .
6. En vista que el modelo anterior no ha resultado satisfactorio una bioestadística ha propuesto un modelo alternativo consistente en modelizar la variable como:

$$g_X(x) = \exp(-kx), \quad \text{donde } x \geq 0$$

Calcular la constante  $k$  para que  $g_X$  sea una función de densidad de probabilidad. Repetir los cálculos de los apartados b), c), d) y e) con el nuevo modelo. Discutir adecuación de ambos modelos a una situación real.

### SOLUCIÓN

### 2.3.1 $f_X(x)$ es una densidad

Para comprobar que  $f_X(x)$  es una función de densidad, necesitamos verificar que cumple las dos condiciones básicas:

1.  $f_X(x) \geq 0$  para todo  $x$  en su dominio.
2. La integral de  $f_X(x)$  sobre todo su dominio debe ser 1, es decir:

$$\int_0^2 f_X(x) dx = 1$$

La función de densidad dada es  $f_X(x) = -0.5 \cdot x + 1$  con  $0 \leq x \leq 2$ .

Primero, comprobamos que  $f_X(x) \geq 0$  para  $x \in [0, 2]$ . Evaluamos los valores extremos:

- $f_X(0) = -0.5 \cdot 0 + 1 = 1$
- $f_X(2) = -0.5 \cdot 2 + 1 = 0$

La función es no negativa en el intervalo dado.

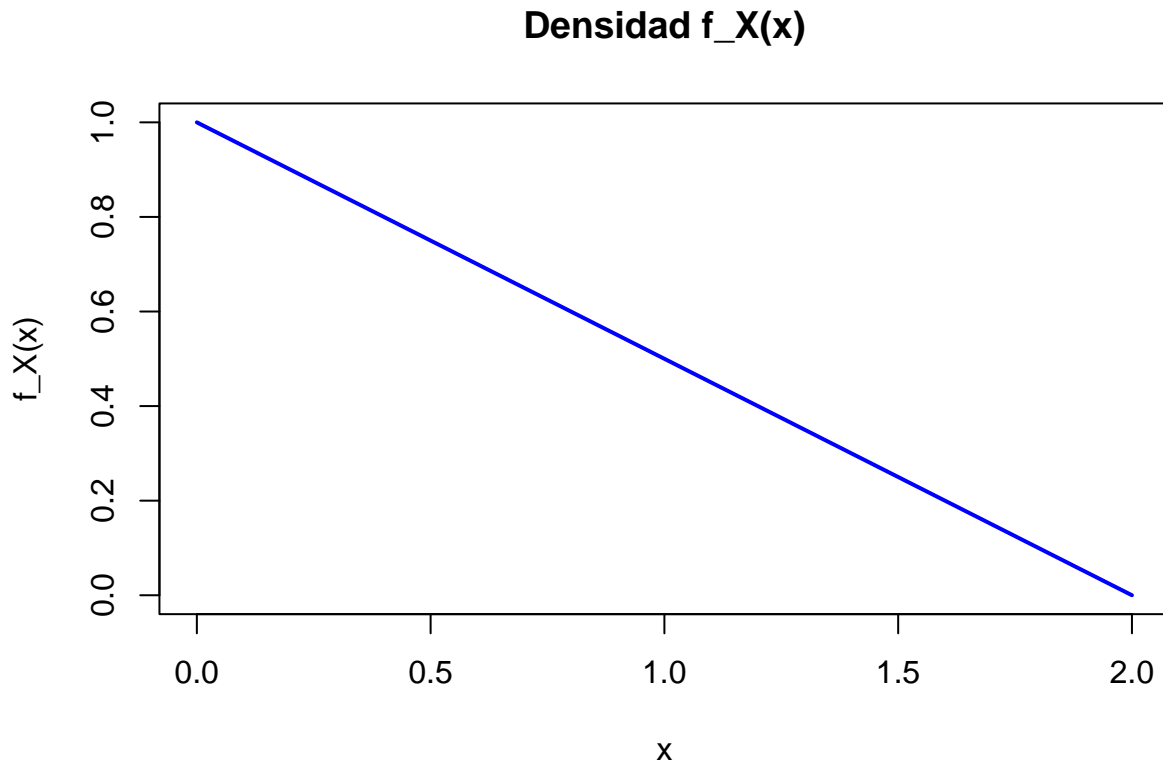
Ahora, calculamos la integral:

$$\int_0^2 (-0.5 \cdot x + 1) dx = [-0.25 \cdot x^2 + x]_0^2 = (-0.25 \cdot 4 + 2) - (0) = 1$$

Por lo tanto,  $f_X(x)$  cumple con ambas condiciones y es una función de densidad.

### 2.3.2 Gráfica de $f_X(x)$

```
# R code to plot the density function
f_x <- function(x) -0.5 * x + 1
curve(f_x, from = 0, to = 2, col = "blue", lwd = 2, ylab = "f_X(x)", xlab = "x",
      main = "Densidad f_X(x)")
```



### 2.3.3 Función de distribución

Calcular  $F_X(x)$  y representarla gráficamente

La función de distribución acumulada (CDF)  $F_X(x)$  se obtiene integrando la función de densidad:

$$F_X(x) = \int_0^x (-0.5 \cdot t + 1) dt$$

Para  $x \in [0, 2]$ , tenemos:

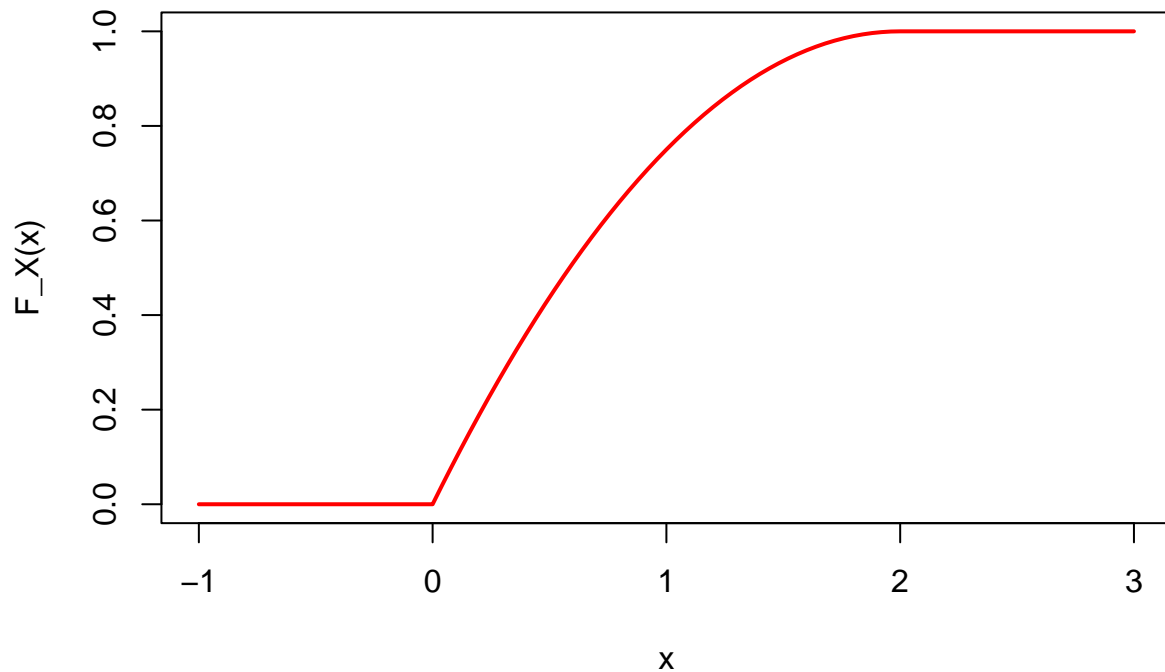
$$F_X(x) = [-0.25 \cdot t^2 + t]_0^x = -0.25 \cdot x^2 + x$$

Para  $x < 0$ ,  $F_X(x) = 0$ , y para  $x > 2$ ,  $F_X(x) = 1$ .

Gráfica de  $F_X(x)$

```
# R code to plot the CDF function
F_x <- function(x) ifelse(x < 0, 0, ifelse(x > 2, 1, -0.25 * x^2 + x))
curve(F_x, from = -1, to = 3, col = "red", lwd = 2, ylab = "F_X(x)", xlab = "x",
      main = "Distribución acumulada F_X(x)")
```

### Distribución acumulada $F_X(x)$



#### 2.3.4 Probabilidades y $f_X(1)$

1.  $P(X \geq 1) = 1 - F_X(1)$ :

$$F_X(1) = -0.25 \cdot 1^2 + 1 = 0.75$$

Por lo tanto,  $P(X \geq 1) = 1 - 0.75 = 0.25$ .

2.  $P(X > 1)$ : Como  $X$  es una variable continua,  $P(X > 1) = P(X \geq 1) = 0.25$ .
3.  $P(X = 1)$ : Para una variable continua, la probabilidad puntual es 0, es decir,  $P(X = 1) = 0$ .
4.  $f_X(1)$ :

$$f_X(1) = -0.5 \cdot 1 + 1 = 0.5$$

#### 2.3.5 Probabilidad de supervivencia

1. Menos de seis meses ( $x = 0.5$ ):

$$P(X < 0.5) = F_X(0.5) = -0.25 \cdot 0.5^2 + 0.5 = 0.4375$$



2. Entre seis meses y un año ( $x \in [0.5, 1]$ ):

$$P(0.5 \leq X \leq 1) = F_X(1) - F_X(0.5) = 0.75 - 0.375 = 0.375$$

3. Más de dos años ( $x > 2$ ): Como el dominio de  $X$  es  $[0, 2]$ ,  $P(X > 2) = 0$ .

### 2.3.6 $E(X)$ y $\text{Var}(X)$

1. La esperanza de  $X$  es:

$$E(X) = \int_0^2 x \cdot f_X(x) dx = \int_0^2 x \cdot (-0.5 \cdot x + 1) dx$$

Desarrollamos:

$$E(X) = \int_0^2 (-0.5 \cdot x^2 + x) dx = \left[ -\frac{0.5}{3} \cdot x^3 + 0.5 \cdot x^2 \right]_0^2$$

Calculamos:

$$E(X) = -\frac{0.5}{3} \cdot 8 + 0.5 \cdot 4 = -\frac{4}{3} + 2 = \frac{2}{3}$$

2. La varianza de  $X$  es:

$$\text{Var}(X) = E(X^2) - E(X)^2$$

Primero calculamos  $E(X^2)$ :

$$E(X^2) = \int_0^2 x^2 \cdot f_X(x) dx = \int_0^2 x^2 \cdot (-0.5 \cdot x + 1) dx$$

Desarrollamos y calculamos:

$$E(X^2) = \int_0^2 (-0.5 \cdot x^3 + x^2) dx = \left[ -\frac{0.5}{4} \cdot x^4 + \frac{1}{3} \cdot x^3 \right]_0^2$$

$$E(X^2) = -\frac{0.5}{4} \cdot 16 + \frac{1}{3} \cdot 8 = -2 + \frac{8}{3} = \frac{2}{3}$$

Finalmente:

$$\text{Var}(X) = E(X^2) - E(X)^2 = \frac{2}{3} - \left(\frac{2}{3}\right)^2 = \frac{2}{3} - \frac{4}{9} = \frac{2}{9}$$

### 2.3.7 Modelo alternativo $g_X(x)$

Dado el modelo alternativo  $g_X(x) = \exp(-k \cdot x)$  para  $x \geq 0$ , la constante  $k$  se determina imponiendo que la integral de la función de densidad debe ser 1:

$$\int_0^{\infty} \exp(-k \cdot x) dx = 1$$

Resolviendo:

$$\frac{1}{k} = 1 \implies k = 1$$

Por lo tanto, el nuevo modelo de densidad es  $g_X(x) = \exp(-x)$ .

## 2.4 Ejercicio 2.4

Para estudiar la regulación hormonal de una línea metabólica se inyectan ratas albinas con un fármaco que inhibe la síntesis de proteínas del organismo. En general, 4 de cada 20 ratas mueren a causa del fármaco antes de que el experimento haya concluido. Si se trata a 10 animales con el fármaco, ¿cuál es la probabilidad de que al menos 8 lleguen vivas al final del experimento?

### SOLUCION

En este problema en el que tenemos grupos de 10 animales independientes, cada uno de los cuales puede sobrevivir o no resulta apropiada **la distribución binomial**.

- La probabilidad de que una rata sobreviva al fármaco es  $p = \frac{16}{20} = 0.8$ , dado que 4 de cada 20 ratas mueren.
- El experimento se realiza con 10 ratas, por lo que tenemos  $n = 10$ .
- Queremos calcular la probabilidad de que al menos 8 ratas sobrevivan. Matemáticamente, esto corresponde a:

$$P(X \geq 8)$$

donde  $X$  es el número de ratas que sobreviven y sigue una **distribución binomial**:

$$X \sim \text{Binomial}(n = 10, p = 0.8)$$

#### 2.4.1 Cálculo de la probabilidad

La probabilidad de que exactamente  $k$  ratas sobrevivan está dada por la fórmula de la binomial:

$$P(X = k) = \binom{n}{k} p^k (1 - p)^{n-k}$$

Para responder la pregunta debemos calcular:

$$P(X \geq 8) = P(X = 8) + P(X = 9) + P(X = 10)$$

Esto puede calcularse:

- directamente usando la función de probabilidad acumulada implementada en R
- indirectamente calculando las probabilidades individuales y sumándolas.

En todo caso debemos recordar que al tratarse de una variable discreta si queremos usar  $F_X(x)$  para calcular  $P(X \geq k)$  deberemos tener en cuenta que:

$$P(X \geq k) = 1 - P(X \leq k - 1)$$

En primer lugar calculamos esta suma utilizando la función de masa de probabilidad:

```
# Parámetros del problema
n <- 10
p <- 0.8

# Probabilidades P(X = 8), P(X = 9) y P(X = 10)
prob_8 <- dbinom(8, size = n, prob = p)
prob_9 <- dbinom(9, size = n, prob = p)
prob_10 <- dbinom(10, size = n, prob = p)

# Probabilidad total P(X >= 8)
prob_total <- prob_8 + prob_9 + prob_10
prob_total
```

```
## [1] 0.6777995
```

Si usamos la función de distribución, `pbinom`

```
1-pbinom (7, size = n, prob = p)
```

```
## [1] 0.6777995
```

Naturalmente ambos resultados coinciden. Obsérvese que al ser  $p = 0.8$  valores altos resultan bastante probables, con lo que la

## 2.5 Ejercicio 2.5

En una cierta población se ha observado un número medio anual de 12 muertes por cáncer de pulmón. Si el número de muertes causadas por la enfermedad sigue una distribución de Poisson, ¿cuál es la probabilidad de que durante el año en curso: 1. haya exactamente 10 muertes por cáncer de pulmón? 2. 15 o más personas mueran a causa de la enfermedad? 3. 10 o menos personas mueran a causa de la enfermedad?

El número de muertes por cáncer de pulmón sigue una distribución de Poisson, que se usa para modelar la ocurrencia de eventos discretos dentro de un intervalo de tiempo, donde el valor esperado es proporcional al tamaño del intervalo. En este caso, el valor esperado es el número medio de muertes por año, que es 12. La función de masa de probabilidad (PMF) de una variable aleatoria  $X$  con distribución de Poisson y parámetro  $\lambda$  es:

$$P(X = k) = \frac{\lambda^k e^{-\lambda}}{k!}$$

donde  $k$  es el número de eventos,  $\lambda$  es el valor esperado (12 en nuestro caso) y  $k!$  es el factorial de  $k$ . Usaremos este modelo para resolver los apartados.

### 2.5.1 Probabilidad de que haya exactamente 10 muertes

La probabilidad de observar exactamente  $k = 10$  muertes se puede calcular usando la PMF de la distribución de Poisson con  $\lambda = 12$ :

$$P(X = 10) = \frac{12^{10}e^{-12}}{10!}$$

Podemos calcular este valor con R.

```
lambda <- 12
k <- 10
prob_10_muertes <- dpois(k, lambda)
prob_10_muertes
```

```
## [1] 0.1048373
```

### 2.5.2 Probabilidad de que 15 o más personas mueran

Para obtener la probabilidad de que 15 o más personas mueran, necesitamos calcular la probabilidad acumulada de  $X \geq 15$ . Esto se puede obtener restando de 1 la probabilidad acumulada de  $X < 15$ , es decir:

$$P(X \geq 15) = 1 - P(X < 15) = 1 - P(X \leq 14)$$

Usamos la función de probabilidad acumulada (CDF) de la Poisson en R.

```
k_15 <- 14
prob_15_o_mas <- 1 - ppois(k_15, lambda)
prob_15_o_mas
```

```
## [1] 0.2279755
```

### 2.5.3 Probabilidad de que 10 o menos personas mueran

La probabilidad de que 10 o menos personas mueran es simplemente la probabilidad acumulada de  $X \leq 10$ , que se puede calcular directamente con la CDF de la distribución de Poisson.

$$P(X \leq 10)$$

Calculamos esto en R:

```
prob_10_o_menos <- ppois(k, lambda)
prob_10_o_menos
```

```
## [1] 0.3472294
```

### 2.5.4 Conclusión

1. La probabilidad de que haya exactamente 10 muertes es:

```
prob_10_muertes
```

```
## [1] 0.1048373
```

2. La probabilidad de que 15 o más personas mueran es:

```
prob_15_o_mas
```

```
## [1] 0.2279755
```

3. La probabilidad de que 10 o menos personas mueran es:

```
prob_10_o_menos
```

```
## [1] 0.3472294
```

## 2.6 Ejercicio 2.6

Los daños a los cromosomas del óvulo o del espermatozoide, pueden causar mutaciones que conducen a abortos, defectos de nacimiento, u otras deficiencias genéticas. Un estudio sobre los efectos teratogénicos de la radiación ha determinado que la probabilidad de que tal mutación se produzca por radiación es del 10%. El resto son atribuibles a otras causas. Una vez detectadas 150 mutaciones,

1. ¿cuántas se esperaría que se debiesen a radiaciones?
2. ¿Cuál es la probabilidad de que solamente 10 se debiesen a radiaciones?

### Solución

Para analizar el número de mutaciones que se deben a radiaciones, podemos considerar dos modelos diferentes: uno basado en la distribución binomial y otro en la distribución de Poisson.

#### 2.6.1 Justificación del uso de distribución binomial

La distribución binomial es adecuada cuando tenemos un número fijo de ensayos independientes y cada ensayo tiene dos posibles resultados: éxito (la mutación es debida a radiación) o fracaso (la mutación no es debida a radiación). En cada ensayo, la probabilidad de éxito es constante.

Esto se ajusta perfectamente a las condiciones del problema: - Hay 150 ensayos independientes (cada mutación observada puede estar o no causada por radiación). - Cada ensayo tiene dos posibles resultados: mutación por radiación o mutación por otra causa. - La probabilidad de éxito es constante y pequeña ( $p = 0.1$ ). Por tanto, el número de mutaciones debidas a radiación se puede modelizar bien mediante una distribución binomial  $X \sim \text{Binomial}(n = 150, p = 0.1)$ .

#### 2.6.2 Justificación del uso de distribución de Poisson

La distribución de Poisson es adecuada para modelar el número de eventos raros que ocurren en un intervalo de tiempo, espacio, o cualquier otra unidad, cuando estos eventos ocurren de forma independiente y su probabilidad de ocurrencia es baja.

En este caso las “mutaciones debidas a radiación” pueden considerarse eventos raros dentro de un gran conjunto de mutaciones (150 mutaciones observadas, pero solo un 10% de ellas son debidas a radiación).

Puede considerarse además, que las mutaciones individuales pueden ocurrir de forma independiente entre sí, ya que la probabilidad de que una mutación se deba a radiación no afecta a la probabilidad de que otra mutación sea causada por radiación.

Estas condiciones son características de los *procesos de Poisson* y por tanto la distribución de Poisson es una elección natural para describir procesos en los que los eventos ocurren de manera aleatoria en un intervalo dado (por ejemplo, en un periodo de tiempo o un espacio), siempre que:

- Los eventos ocurran con una tasa promedio constante (en este caso, la tasa de mutaciones debidas a radiaciones es proporcional a la tasa global de mutaciones, multiplicada por la probabilidad  $p = 0.1$ ).
- No haya límite teórico en el número de eventos que puedan ocurrir en un intervalo (aunque observamos un total de 150 mutaciones, teóricamente podríamos seguir detectando más mutaciones).

En el modelo de Poisson, el parámetro  $\lambda$  representa la tasa promedio de ocurrencia de los eventos (en este caso, mutaciones debidas a radiación). Si conocemos la tasa promedio de aparición de mutaciones por radiación ( $\lambda = n \cdot p$  en el contexto binomial, pero también se puede calcular directamente si conocemos la tasa de aparición de eventos raros), entonces podemos usar directamente la distribución de Poisson para modelar el número de eventos.

En este caso,  $\lambda = 150 \cdot 0.1 = 15$ , que representa el número esperado de mutaciones debidas a radiación en el total observado de mutaciones.

### 2.6.3 Aproximación del modelo binomial por el de Poisson

La distribución de Poisson puede considerarse una aproximación de la binomial cuando el número de ensayos ( $n$ ) es grande y la probabilidad de éxito ( $p$ ) es pequeña. En este caso, el número esperado de éxitos,  $n \cdot p$ , se mantiene moderado (en este caso,  $n \cdot p = 15$ ).

Este resultado que se conoce como *límite de Poisson* establece que si:

- $n$  es grande (muchos ensayos),
- $p$  es pequeño (baja probabilidad de éxito),
- el producto  $n \cdot p = \lambda$  es moderado,

entonces la binomial  $X \sim \text{Binomial}(n, p)$  se puede aproximar por una distribución de Poisson con parámetro  $\lambda = n \cdot p$ .

En este caso:

- $n = 150$  es suficientemente grande.
- $p = 0.1$  es pequeño.
- $n \cdot p = 15$ , lo cual es un valor razonable para usar la aproximación de Poisson.

Por tanto, el número de mutaciones debidas a radiaciones puede aproximarse por una distribución de Poisson  $X \sim \text{Poisson}(\lambda = 15)$ .

### 2.6.4 Número esperado de mutaciones

En ambos modelos, la esperanza del número de mutaciones debidas a radiaciones es  $E[X] = n \cdot p$ . Esto representa el número promedio de mutaciones debidas a radiaciones. Lo calculamos:

$$E[X] = 150 \cdot 0.1 = 15$$

Por lo tanto, se espera que alrededor de 15 mutaciones se deban a radiaciones.

## 2.6.5 Probabilidad de que exactamente 10 mutaciones se deban a radiaciones

**2.6.5.1 Usando la distribución Binomial** La probabilidad de que exactamente 10 mutaciones se deban a radiaciones se puede calcular usando la PMF de la binomial:

$$P(X = 10) = \binom{150}{10} (0.1)^{10} (0.9)^{140}$$

Usando R tenemos:

```
n <- 150
p <- 0.1
k <- 10
prob_binom_10 <- dbinom(k, n, p)
prob_binom_10
```

```
## [1] 0.04591681
```

**2.6.5.2 Usando la aproximación de Poisson** La distribución de Poisson con  $\lambda = n \cdot p = 15$  también se puede usar para aproximar esta probabilidad. La probabilidad de obtener exactamente 10 mutaciones se calcula como:

$$P(X = 10) = \frac{15^{10} e^{-15}}{10!}$$

Con R:

```
lambda <- 15
prob_pois_10 <- dpois(k, lambda)
prob_pois_10
```

```
## [1] 0.04861075
```

## 2.6.6 Conclusión

- Se espera que 15 de las 150 mutaciones se deban a radiaciones.
- La probabilidad de que exactamente 10 mutaciones se deban a radiaciones es:
  - Usando la distribución binomial:

```
prob_binom_10
```

```
## [1] 0.04591681
```

- Usando la aproximación de Poisson:

```
prob_pois_10
```

```
## [1] 0.04861075
```

Ambos métodos dan resultados similares, pero el modelo de Poisson es útil para simplificar los cálculos cuando el número total de mutaciones es grande y la probabilidad de cada evento es pequeña.

## 2.7 Ejercicio 2.7

Entre los diabéticos, el nivel de glucosa en sangre  $X$ , en ayunas, puede suponerse de distribución aproximadamente normal, con media 106mg/100ml y desviación típica 8mg/100ml, es decir :  $X \sim N(\mu = 106, \sigma^2 = 64)$ .

Hallar; 1. El porcentaje de diabéticos con niveles de glucosa inferiores a 120 (  $P[X \leq 120]$  ) 2. ¿Qué porcentaje de diabéticos tienen niveles comprendidos entre 90 y 120? 3. Hallar el nivel de glucosa “p25”, caracterizado por la propiedad de que el 25% de todos los diabéticos tiene un nivel de glucosa en ayunas inferior o igual a  $x$ .

### SOLUCIÓN

Según el enunciado el nivel de glucosa  $X$  se distribuye según una distribución normal con media  $\mu = 106$  y varianza  $\sigma^2 = 64$ , es decir,  $X \sim N(106, 64)$ , o equivalentemente  $X \sim N(106, 8^2)$ .

#### 2.7.1 Porcentaje de diabéticos con niveles de glucosa inferiores a 120 ( $P[X \leq 120]$ )

Para calcular esta probabilidad, necesitamos estandarizar la variable  $X$  a una normal estándar  $Z \sim N(0, 1)$ . La fórmula de estandarización es:

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

Sustituyendo los valores de  $\mu = 106$  y  $\sigma = 8$ :

$$Z = \frac{120 - 106}{8} = 1.75$$

Ahora calculamos  $P(Z \leq 1.75)$ , es decir, la probabilidad de que la variable estándar normal sea menor o igual que 1.75. Esta probabilidad la obtenemos a partir de la tabla de la normal estándar o usando R.

```
# Calculamos la probabilidad con la función pnorm
p1 <- pnorm(1.75)
p1
```

```
## [1] 0.9599408
```

#### 2.7.2 Porcentaje de diabéticos con niveles de glucosa comprendidos entre 90 y 120

En este caso queremos calcular  $P(90 \leq X \leq 120)$ . Para hacerlo, calculamos las probabilidades individuales de  $P(X \leq 120)$  y  $P(X \leq 90)$ , y restamos la segunda de la primera:

$$P(90 \leq X \leq 120) = P(X \leq 120) - P(X \leq 90)$$

Primero estandarizamos ambas variables:

$$Z_{120} = \frac{120 - 106}{8} = 1.75$$
$$Z_{90} = \frac{90 - 106}{8} = -2.00$$

Ahora calculamos  $P(Z \leq 1.75)$  y  $P(Z \leq -2.00)$  usando R.



```
# Calculamos ambas probabilidades
p2_120 <- pnorm(1.75)
p2_90 <- pnorm(-2.00)
p2 <- p2_120 - p2_90
p2
```

```
## [1] 0.9371907
```

### 2.7.3 Hallar el nivel de glucosa “p25”

Para encontrar el percentil 25 de la distribución, necesitamos resolver la ecuación:

$$P(X \leq p_{25}) = 0.25$$

Sabemos que  $X \sim N(106, 64)$ , así que estandarizamos el valor  $p_{25}$ :

$$Z_{p_{25}} = \frac{p_{25} - 106}{8}$$

Luego, encontramos el valor de  $Z_{p_{25}}$  que corresponde al percentil 25 de la distribución normal estándar, es decir,  $P(Z \leq Z_{p_{25}}) = 0.25$ . Esto lo obtenemos con la función inversa de la distribución normal estándar.

```
# Calculamos el valor z correspondiente al percentil 25
z_p25 <- qnorm(0.25)
# Calculamos el p25 en la escala original
p25 <- 106 + z_p25 * 8
p25
```

```
## [1] 100.6041
```

### 2.7.4 Resumen de resultados:

1. La probabilidad de que el nivel de glucosa sea menor o igual a 120 es aproximadamente:

$$P[X \leq 120] = 0.9599$$

2. El porcentaje de diabéticos con niveles de glucosa comprendidos entre 90 y 120 es aproximadamente:

$$P[90 \leq X \leq 120] = 0.9104$$

3. El nivel de glucosa correspondiente al percentil 25, es decir, el valor  $p_{25}$ , es aproximadamente:

$$p_{25} \approx 100.61 \text{ mg/100ml}$$

## 2.8 Ejercicio 28

Se supone que la glucemia basal en individuos sanos,  $X_s$  sigue una distribución  $X \sim N(\mu = 80, \sigma = 10)$ , mientras que en los diabéticos  $X_d$ , sigue una distribución  $X \sim N(\mu = 160, \sigma = 31.4)$ . Si se conviene en clasificar como sanos al 2% de los diabéticos: a) ¿Por debajo de qué valor se considera sano a un individuo? ¿Cuántos sanos serán clasificados como diabéticos? b) Se sabe que en la población en general el 10% de los individuos son diabéticos ¿cuál es la probabilidad de que un individuo elegido al azar y diagnosticado como diabético, realmente lo sea?

## 2.9 Ejercicio 2.9

Supóngase que se van a utilizar 20 ratas en un estudio de agentes coagulantes de la sangre. Como primera experiencia, se dio un anticoagulante a 10 de ellos, pero por inadvertencia se pusieron todas sin marcas en el mismo recinto. Se necesitaron 12 ratas para la segunda fase del estudio y se le tomó al azar sin reemplazamiento. ¿Cuál es la probabilidad de que de las 12 elegidas 6 tengan la droga y 6 no la tengan?

# 3 Distribuciones de probabilidad multidimensionales

## 3.1 Ejercicio 1

Se tienen dos estudios clínicos importantes, cuyos análisis genéticos deben ser asignados aleatoriamente a uno o más de tres laboratorios, A, B y C. Denote con  $Y_1$  el número de estudios asignados al laboratorio A y con  $Y_2$  el número de estudios asignados al laboratorio B. Cada laboratorio puede recibir 0, 1 o 2 estudios.

- Encuentre la función de probabilidad conjunta para  $Y_1$  y  $Y_2$ .
- Encuentre  $F(1, 0)$ , es decir, la probabilidad de que el laboratorio A reciba como máximo un estudio y el laboratorio B no reciba ninguno.

## 3.2 Ejercicio 2

Tres monedas balanceadas se lanzan en forma independiente al aire. Una de las variables de interés es  $Y_1$ , el número de caras. Denote con  $Y_2$  la cantidad de dinero ganado en una apuesta colateral en la siguiente forma. Si la primera cara aparece en el primer tiro, usted gana 1€. Si la primera cara aparece en el tiro segundo o en el tercero gana 2€ o 3€, respectivamente. Si no aparece una cara, usted pierde 1€ (esto es, gana - 1€). a) Encuentre la función de probabilidad conjunta para  $Y_1$  y  $Y_2$ . b) ¿Cuál es la probabilidad de que haya menos de tres caras y usted gane 1€ o menos? [Esto es, encuentre  $F(2, 1)$ ].

## 3.3 Ejercicio 3

En el Ejercicio 1 determinamos que la distribución conjunta de  $Y_1$ , el número de análisis asignados al laboratorio A, y  $Y_2$ , el número de análisis asignados al laboratorio B, está dada por las entradas en la siguiente tabla.

	$y_1$		
$y_2$	0	1	2
0	1/9	2/9	1/9
1	2/9	2/9	0
2	1/9	0	0

- a. Encuentre la distribución de probabilidad marginal de  $Y_1$ .
- b. De acuerdo con los resultados vistos anteriormente  $Y_1$  tiene una distribución binomial con  $n = 2$  y  $p = 1/3$ . ¿Hay algún conflicto entre este resultado y la respuesta dada en el punto a?

### 3.4 Ejercicio 4

Un ingeniero ambiental mide la cantidad (en peso) de partículas contaminantes en muestras de aire de cierto volumen recolectado en dos chimeneas en una planta de energía alimentada con carbón. Una de las chimeneas está equipada con un aparato limpiador. Denote con  $Y_1$  la cantidad de contaminante por muestra recolectada arriba de la chimenea que no tiene aparato limpiador y denote con  $Y_2$  la cantidad de contaminante por muestra recolectada arriba de la chimenea que está equipada con el aparato limpiador.

Suponga que el comportamiento de frecuencia relativa de  $Y_1$  y  $Y_2$  puede ser modelado por

$$f(y_1, y_2) = \begin{cases} k, & 0 \leq y_1 \leq 2, \quad 0 \leq y_2 \leq 1, \quad 2y_2 \leq y_1 \\ 0, & \text{en cualquier otro punto.} \end{cases}$$

Esto es,  $Y_1$  y  $Y_2$  están uniformemente distribuidas sobre la región dentro del triángulo limitado por  $y_1 = 2$ ,  $y_2 = 0$  y  $2y_2 = y_1$ .

- a. Encuentre el valor de  $k$  que haga de ésta una función de densidad de probabilidad.
- b. Encuentre  $P(Y_1 \geq 3Y_2)$ . Esto es, encuentre la probabilidad de que el aparato limpiador reduzca la cantidad de contaminante en un tercio o más.

### 3.5 Ejercicio 5

En el Ejercicio 4 hemos establecido que

$$f(y_1, y_2) = \begin{cases} k, & 0 \leq y_1 \leq 2, \quad 0 \leq y_2 \leq 1, \quad 2y_2 \leq y_1 \\ 0, & \text{en cualquier otro punto.} \end{cases}$$

es una función de densidad de probabilidad conjunta válida para  $Y_1$ , la cantidad de contaminante por muestra recolectada arriba de la chimenea que no tenía el aparato limpiador, y para  $Y_2$ , la cantidad recolectada arriba de la chimenea con el aparato limpiador.

- a. Si consideramos la chimenea con el limpiador instalado, encuentre la probabilidad de que la cantidad de contaminante en una muestra determinada sea mayor que 0.5.
- b. Dado que se observa que la cantidad de contaminante en una muestra tomada arriba de la chimenea con el limpiador es 0.5, encuentre la probabilidad de que la cantidad de contaminante exceda de 1.5 arriba de la otra chimenea (la que no tiene limpiador).

### 3.6 Ejercicio 6

En el ejercicio 1 determinamos que la distribución conjunta de  $Y_1$ , el número de análisis asignados al laboratorio A, y  $Y_2$ , el número de análisis asignados al laboratorio B, está dada por las entradas en la siguiente tabla.

	$y_1$		
$y_2$	0	1	2
0	1/9	2/9	1/9
1	2/9	2/9	0
2	1/9	0	0

- Encuentre  $\text{Cov}(Y_1, Y_2)$ .
- Le sorprende que  $\text{Cov}(Y_1, Y_2)$  sea negativa? ¿Por qué?

### 3.7 Ejercicio 7

Las variables aleatorias  $Y_1$  y  $Y_2$  son tales que  $E(Y_1) = 4$ ,  $E(Y_2) = -1$ ,  $V(Y_1) = 2$  y  $V(Y_2) = 8$ .

- ¿Cuál es  $\text{Cov}(Y_1, Y_1)$ ?
- Suponiendo que las medias y las varianzas sean correctas, ¿es posible que  $\text{Cov}(Y_1, Y_2) = 7$ ? [Sugerencia: si  $\text{Cov}(Y_1, Y_2) = 7$ , ¿cuál es el valor de  $\rho$ , el coeficiente de correlación?]
- Suponiendo que las medias y las varianzas sean correctas, ¿cuál es el máximo valor posible para  $\text{Cov}(Y_1, Y_2)$ ? Si  $\text{Cov}(Y_1, Y_2)$  alcanza este valor máximo, ¿qué implica eso acerca de la relación entre  $Y_1$  y  $Y_2$ ?

### 3.8 Ejercicio 8

Un experimento de aprendizaje requiere que una rata corra por un laberinto (una red de pasillos) hasta que localice una de tres posibles salidas. La salida 1 presenta una recompensa de alimento, no así las salidas 2 y 3. (Si la rata finalmente selecciona la salida 1 casi siempre, puede tener lugar el aprendizaje.) Denote con  $Y_i$  el número de veces que la salida  $i$  es seleccionada en corridas sucesivas. Para lo siguiente, suponga que la rata escoge una salida aleatoriamente en cada corrida.

- Encuentre la probabilidad de que  $n = 6$  corridas resulte en  $Y_1 = 3$ ,  $Y_2 = 1$  y  $Y_3 = 2$ .
- Para  $n$  general, encuentre  $E(Y_1)$  y  $V(Y_1)$ .
- Encuentre  $\text{Cov}(Y_2, Y_3)$  para  $n$  general.
- Para comprobar la preferencia de la rata entre las salidas 2 y 3, podemos buscar en  $Y_2 - Y_3$ . Encuentre  $E(Y_2 - Y_3)$  y  $V(Y_2 - Y_3)$  para  $n$  general.