Ejercicios Inferencia Unidad2. SOLUCIONES

Alex Sanchez y Santi Perez-Hoyos

2024-11-08

Table of contents

1 Ejercicio 2.1 **2**
1.1 1. Distribución de probabilidad . . . 3
1.2 2. Esperanza y varianza . . . 4

2 Ejercicio 2.2 **5**
2.1 Determinar el valor de K . . . 5
2.2 Calcular la esperanza y la varianza . . . 6
2.3 3. Calcular las probabilidades . . . 7
2.3.1 1. Probabilidad de que un tumor presente exactamente tres mutaciones . . . 7
2.3.2 2. Probabilidad de que un tumor presente al menos una mutación . . . 7
2.3.3 3. Probabilidad de que un tumor presente como máximo dos mutaciones . . . 8

3 Ejercicio 2.3 **8**
3.1 Comprobar que $f_X(x)$ es una densidad . . . 9
3.1.1 Gráfica de $f_X(x)$ . . . 9
3.2 Calcular $F_X(x)$ y representarla gráficamente . . . 10
3.2.1 Gráfica de $F_X(x)$ . . . 10
3.3 Calcular probabilidades y $f_X(1)$ . . . 11
3.4 Probabilidad de supervivencia . . . 11
3.5 Calcular $E(X)$ y $\text{Var}(X)$ . . . 12
3.6 Modelo alternativo $g_X(x)$ . . . 13

4 Ejercicio 2.4 **13**
4.1 Cálculo de la probabilidad . . . 14

5 Ejercicio 2.5 **15**
5.1 Probabilidad de que haya exactamente 10 muertes . . . 15
5.2 Probabilidad de que 15 o más personas mueran . . . 15
5.3 Probabilidad de que 10 o menos personas mueran . . . 16
5.4 Conclusión . . . 16

6 Ejercicio 2.6 **17**
6.1 Justificación del uso de distribución binomial . . . 17
6.2 Justificación del uso de distribución de Poisson . . . 17
6.3 Aproximación del modelo binomial por el de Poisson . . . 18
6.4 Número esperado de mutaciones . . . 18

6.5 Probabilidad de que exactamente 10 mutaciones se deban a radiaciones 19
6.5.1 Usando la distribución Binomial 19
6.5.2 Usando la aproximación de Poisson 19
6.6 Conclusión 19

7 Ejercicio 2.7 **20**
7.1 Porcentaje de diabéticos con niveles de glucosa inferiores a 120 20
7.2 Porcentaje de diabéticos con niveles de glucosa entre 90 y 120 21
7.3 Nivel de glucosa "p25" 21

8 Ejercicio 2.8 **22**
8.1 ¿Por debajo de qué valor se considera sano a un individuo? ¿Cuántos sanos serán clasificados como diabéticos? 22
8.2 ¿Cuál es la probabilidad de que un individuo elegido al azar y diagnosticado como diabético, realmente lo sea? 23

9 Ejercicio 2.9 **24**

1 Ejercicio 2.1

Se sabe que la presencia de algunas mutaciones en una región genómica puede influir en la sobre-expresión ("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° 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".

1.1 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:

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

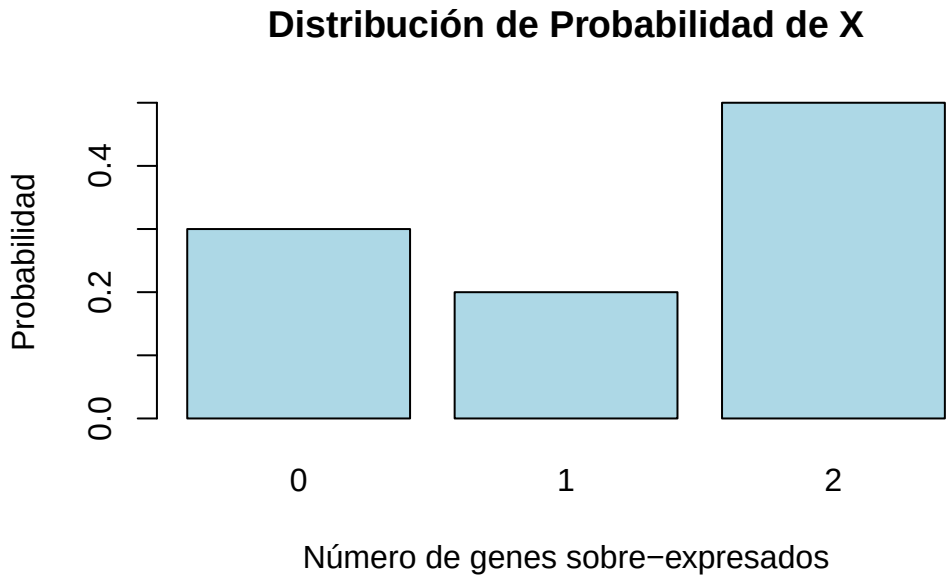
La distribución de probabilidad de X es la siguiente:

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

Podemos representarla gráficamente usando R:

```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")
```



1.2 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:

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

varianza <- esperanza_cuadrado - esperanza^2

esperanza
```

```
[1] 1.2
```

```r
varianza
```

```
[1] 0.76
```

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:

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

2.1 Determinar el 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 Calcular la 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:

```r
# 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
```

```r
varianza
```

```
[1] 2.761769
```

2.3 3. Calcular las probabilidades

2.3.1 1. 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:

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

```
[1] 0.1255605
```

2.3.2 2. 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:

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

```
[1] 0.6860987
```

2.3.3 3. 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
```

3 Ejercicio 2.3

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

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

a) Comprobar que f_X es una densidad. Representarla gráficamente.
b) Calcular F_Xy representarla gráficamente.
c) Calcular $P(X \geq 1), P(X > 1), P(X = 1), f_x(1)$.
d) 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.

e) Calcular $E(X)$ i $\text{Var}(X)$.
f) En vista que el modelo anterior no ha resultado satisfactorio una bioestadística ha propuesto un modelo alternativo consistente en modelizar la variable como:

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

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

3.1 Comprobar que $f_X(x)$ 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.

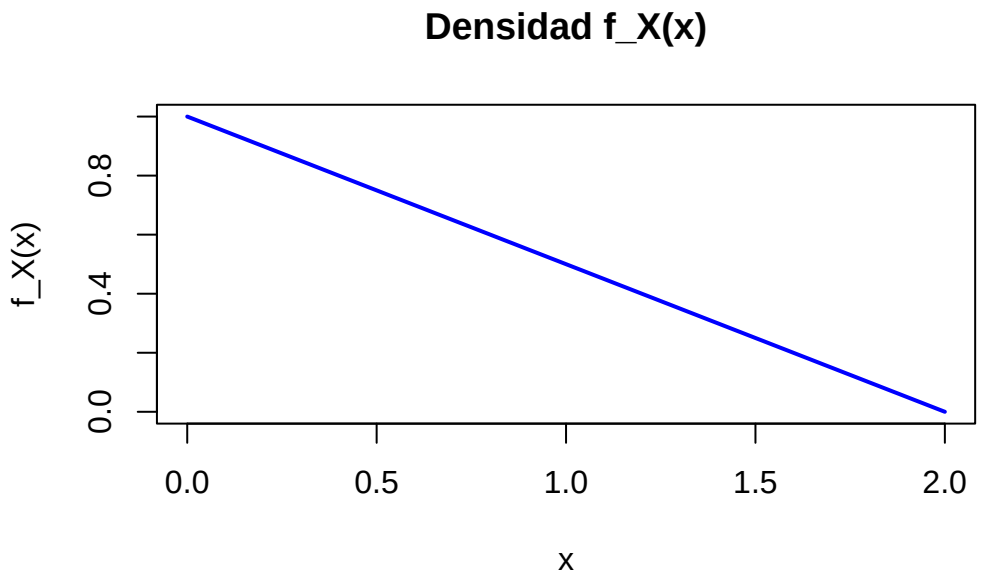
Ahora, calculamos la integral:

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

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

3.1.1 Gráfica de $f_X(x)$

```r
# 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)")
```



3.2 Calcular $F_X(x)$ y representarla gráficamente

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

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

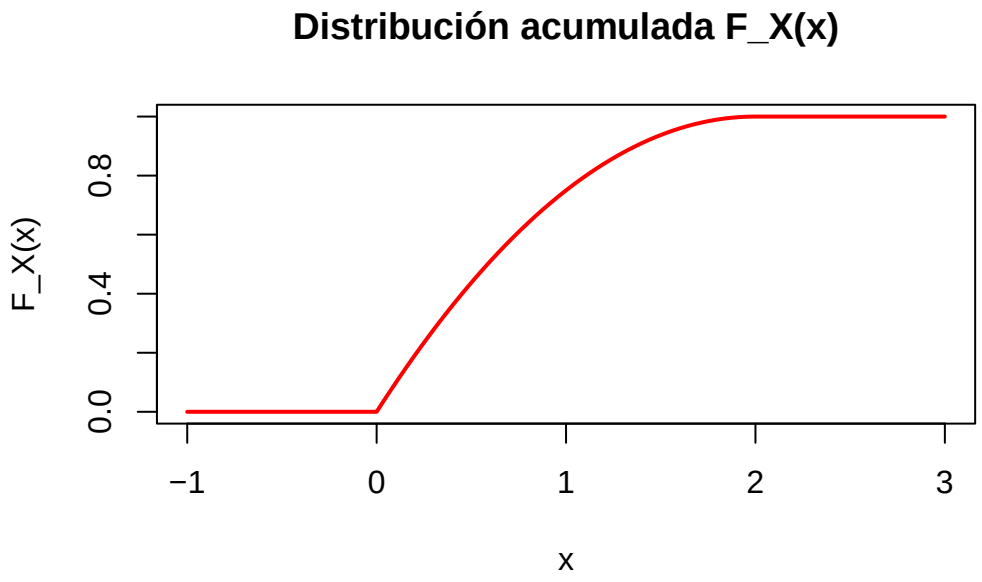
Para $x \in [0, 2]$, tenemos:

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

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

3.2.1 Gráfica de $F_X(x)$

```r
# 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)")
```



3.3 Calcular 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. $f_X(1)$:

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

3.4 Probabilidad de supervivencia

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

$$P(X < 0.5) = F_X(0.5) = -0.25 \cdot 0.5^2 + 0.5 = 0.375$$

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 $[0, 2]$, $P(X > 2) = 0$.

3.5 Calcular $E(X)$ y $\mathrm{Var}(X)$

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:

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

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:

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

3.6 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 $g_X(x) = \exp(-x)$.

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**:

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

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 funcion 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

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 λ es:

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

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

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
```

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.

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

```
[1] 0.2279755
```

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:

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

```
[1] 0.3472294
```

5.4 Conclusión

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

```r
prob_10_muertes
```

```
[1] 0.1048373
```

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

```r
prob_15_o_mas
```

```
[1] 0.2279755
```

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

```r
prob_10_o_menos
```

```
[1] 0.3472294
```

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 considera dos modelos diferentes: uno basado en la distribución binomial y otro en la distribución de Poisson.

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 $X \sim \text{Binomial}(n = 150, p = 0.1)$.

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 λ 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.

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 $X \sim \text{Poisson}(\lambda = 15)$.

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.

6.5 Probabilidad de que exactamente 10 mutaciones se deban a radiaciones

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
```

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
```

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.

7 Ejercicio 2.7

Entre los diabéticos, el nivel de glucosa en sangre X, en ayunas, puede suponerse de distribución aproximadamente normal, con media[1] 106mg/100ml y desviación típica 8mg/100ml, es decir : $X \sim N\left(\mu = 106, \sigma^2 = 64\right)$.

1. Hallar 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.

[1] Esto es un ejercicio de cálculo de probabilidades, por lo que el objetivo es realizar el cálculo que se solicita. El nivel medio de glucosa en diabéticos de tipo 2 suele ser algo más elevado que el indicado aquí.

Solución

7.1 Porcentaje de diabéticos con niveles de glucosa inferiores a 120

Dado que la variable X sigue una distribución normal con media $\mu = 106$ y varianza $\sigma^2 = 64$ (desviación estándar $\sigma = 8$), calcularemos $P(X \leq 120)$. Para resolverlo, estandarizamos X:

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

y sustituimos $X = 120$:

$$P(X \leq 120) = P\left(Z \leq \frac{120 - 106}{8}\right) = P(Z \leq 1.75)$$

Usamos la función de distribución acumulada de la normal estándar para calcular esta probabilidad:

```
pnorm(1.75)
```

```
[1] 0.9599408
```

El resultado nos da aproximadamente 0.9599, lo que indica que el **95.99%** de los diabéticos tienen niveles de glucosa inferiores a 120 mg/100 ml.

7.2 Porcentaje de diabéticos con niveles de glucosa entre 90 y 120

Para calcular el porcentaje de diabéticos con niveles de glucosa entre 90 y 120, evaluamos $P(90 \leq X \leq 120)$. Esto se puede expresar como la diferencia de dos probabilidades:

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

Ya calculamos $P(X \leq 120)$ en la parte a. Ahora, calculamos $P(X \leq 90)$, estandarizando:

$$P(X \leq 90) = P\left(Z \leq \frac{90 - 106}{8}\right) = P(Z \leq -2)$$

Usamos la función de distribución acumulada de la normal estándar para calcular esta probabilidad:

```
pnorm(-2)
```

```
[1] 0.02275013
```

Con ambos resultados, calculamos la probabilidad total:

```
# Probabilidad entre 90 y 120
pnorm(1.75) - pnorm(-2)
```

```
[1] 0.9371907
```

Esto nos da aproximadamente 0.9575, por lo que el **95.75%** de los diabéticos tienen niveles de glucosa entre 90 y 120 mg/100 ml.

7.3 Nivel de glucosa "p25"

Para hallar el nivel de glucosa "p25" que cumple con que el 25% de los diabéticos tienen un nivel de glucosa en ayunas inferior o igual a x, buscamos el valor x tal que $P(X \leq x) = 0.25$. Esto corresponde al **cuantil** del 25% de la distribución normal de X.

Usamos la relación:

$$x = \mu + z_{0.25} \cdot \sigma$$

donde $z_{0.25}$ es el percentil del 25% de la distribución normal estándar, que se puede obtener con la función de cuantil para una normal estándar:

```
# Calculo de p25
p25 <- 106 + qnorm(0.25) * 8
p25
```

```
[1] 100.6041
```

El resultado es aproximadamente 100.61 mg/100 ml. Por lo tanto, el nivel de glucosa **p25** es **100.61 mg/100 ml**, lo que significa que el 25% de los diabéticos tienen un nivel de glucosa en ayunas menor o igual a este valor.

8 Ejercicio 2.8

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:

1. ¿Por debajo de qué valor se considera sano a un individuo? ¿Cuántos sanos serán clasificados como diabéticos?
2. 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?

Solución

8.1 ¿Por debajo de qué valor se considera sano a un individuo? ¿Cuántos sanos serán clasificados como diabéticos?

Para determinar el valor por debajo del cual un individuo se considera sano, necesitamos encontrar el cuantil correspondiente al 2% superior de la distribución de la glucemia basal en diabéticos, $X_d \sim N(\mu = 160, \sigma = 31.4)$. Es decir, debemos encontrar el valor de X_d tal que:

$$P(X_d \leq x) = 0.02$$

Esto corresponde al percentil 2% de la distribución normal de X_d. Utilizamos la función cuantil de la distribución normal para calcularlo:

```
# Nivel de glucosa límite para clasificar como sano
percentil_sano <- qnorm(0.02, mean = 160, sd = 31.4)
percentil_sano
```

```
[1] 95.51228
```

El valor obtenido nos da el límite de glucosa por debajo del cual un individuo será considerado sano. Ahora, para calcular cuántos individuos sanos serán clasificados como diabéticos, necesitamos calcular el porcentaje de individuos sanos cuya glucemia será superior a este valor límite. Dado que $X_s \sim N(\mu = 80, \sigma = 10)$, calculamos la probabilidad:

$$P(X_s > \text{percentil_sano})$$

Esto se puede hacer usando la función de distribución acumulada de la normal para la distribución de los sanos:

```r
# Probabilidad de que un sano sea clasificado como diabético
prob_sano_como_diabetico <- 1 - pnorm(percentil_sano, mean = 80, sd = 10)
prob_sano_como_diabetico
```

```
[1] 0.06042348
```

Esto nos dará la probabilidad de que un individuo sano sea clasificado erróneamente como diabético. Para encontrar el número de sanos clasificados como diabéticos en una población, multiplicamos esta probabilidad por el número total de sanos en la población.

8.2 ¿Cuál es la probabilidad de que un individuo elegido al azar y diagnosticado como diabético, realmente lo sea?

Este problema es un caso de probabilidad condicional. Sabemos que:

- El 10% de la población son diabéticos, por lo que la probabilidad de que un individuo seleccionado al azar sea diabético es $P(D) = 0.10$.
- La probabilidad de que un diabético sea diagnosticado como diabético es $P(\text{diagnóstico}|D)$, lo cual corresponde a la probabilidad de que un individuo con glucosa $X_d \sim N(160, 31.4)$ sea diagnosticado como diabético, es decir, que su nivel de glucosa sea superior al valor límite de clasificación.

Para calcular $P(\text{diagnóstico}|D)$, primero calculamos la probabilidad de que un diabético sea diagnosticado como tal. Es decir, que su nivel de glucosa sea superior al valor que hemos calculado anteriormente para los sanos:

```r
# Probabilidad de que un diabético sea diagnosticado como diabético
prob_diabetico_diagnosticado <- 1 - pnorm(percentil_sano, mean = 160, sd = 31.4)
prob_diabetico_diagnosticado
```

```
[1] 0.98
```

La probabilidad de que un individuo seleccionado al azar y diagnosticado como diabético realmente lo sea se puede obtener usando la fórmula de probabilidad condicional de Bayes:

$$P(D|\text{diagnóstico}) = \frac{P(\text{diagnóstico}|D)P(D)}{P(\text{diagnóstico})}$$

Donde $P(\text{diagnóstico})$ es la probabilidad total de que un individuo sea diagnosticado como diabético. Esta probabilidad es la probabilidad de que un diabético sea diagnosticado como tal más la probabilidad de que un sano sea diagnosticado erróneamente como diabético:

$$P(\text{diagnóstico}) = P(\text{diagnóstico}|D)P(D) + P(\text{diagnóstico}|S)P(S)$$

Donde $P(\text{diagnóstico}|S)$ es la probabilidad de que un sano sea diagnosticado erróneamente como diabético, y $P(S) = 0.90$ es la probabilidad de que un individuo seleccionado al azar sea sano.

Primero, calculamos $P(\text{diagnóstico})$:

```r
# Probabilidad de diagnóstico en la población general
prob_sano_diagnosticado <-
  pnorm(percentil_sano, mean = 80, sd = 10)
prob_diagnostico_total <-
  prob_diabetico_diagnosticado * 0.10 +
  prob_sano_diagnosticado * 0.90
prob_diagnostico_total
```

```
[1] 0.9436189
```

Luego, calculamos la probabilidad condicional:

```r
# Probabilidad de que un individuo diagnosticado como diabético realmente lo sea
prob_diagnostico_given_diabetico <-
  (prob_diabetico_diagnosticado * 0.10) /
  prob_diagnostico_total
prob_diagnostico_given_diabetico
```

```
[1] 0.1038555
```

Este valor es la probabilidad de que un individuo diagnosticado como diabético realmente lo sea, teniendo en cuenta tanto los errores de diagnóstico como las probabilidades de ser diabético o sano en la población.

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 estas fueron extraídas al azar y sin reemplazamiento del recinto en el que se encontraban las 20 ratas. ¿Cuál es la probabilidad de que de las 12 elegidas 6 tengan la droga y 6 no la tengan?

Solución

Este ejercicio puede resolverse usando una distribución **hipergeométrica** debido a que estamos extrayendo ratas sin reemplazo de un grupo finito y buscamos la probabilidad de que haya una combinación específica de ratas con y sin droga en una muestra.

La fórmula de la distribución hipergeométrica se expresa de la siguiente forma:

$$P(X = k) = \frac{\binom{M}{k}\binom{N-M}{n-k}}{\binom{N}{n}}$$

Donde:

- N es el tamaño total de la población (20 ratas).
- M es el número de ratas que tienen la droga (10 ratas con droga).
- n es el tamaño de la muestra seleccionada (12 ratas).

- k es el número de ratas con la droga que queremos en nuestra muestra (6 ratas con droga).

La probabilidad de que 6 ratas de las 12 seleccionadas tengan la droga y 6 no la tengan se calcula usando la fórmula anterior, con los siguientes valores:

- $N = 20$
- $M = 10$
- $n = 12$
- $k = 6$

Podemos calcular esta probabilidad utilizando la función `dhyper` en R, que calcula la probabilidad de un valor específico en una distribución hipergeométrica.

```
# Calcular la probabilidad usando la distribución hipergeométrica
probabilidad <- dhyper(6, 10, 10, 12)
probabilidad
```

```
[1] 0.3500834
```

Este cálculo nos dará la probabilidad de que, al seleccionar al azar 12 ratas sin reemplazo, exactamente 6 tengan la droga y 6 no la tengan.