PEC-3 Otoño 2025

UOC

Os puede ser útil consultar el siguiente material:

1. Reto 2. Probabilidad y variables aleatorias: Módulo Probabilidad y variables aleatorias.
2. Reto 2. Probabilidad y variables aleatorias: Variables aleatorias. Actividades Resueltas.
3. Reto 2. Probabilidad y variables aleatorias: Distribuciones de probabilidad e inferencia estadística con R.

NOMBRE:

PEC3

Muy importante:

En todas las preguntas, tenéis que indicar qué tipo de variable aleatoria usáis y plantear la probabilidad que se pide, como por ejemplo $P(X = 4)$, $P(X \leq 3)$, etc, antes de poner las instrucciones y hacer los cálculos con R. Procurad usar las funciones: dnorm, pnorm, dbinom, …

Toda respuesta que no lleve todas estas explicaciones, con los comentarios pertinentes, será considerada incorrecta aunque se dé un resultado numérico correcto.

Pregunta 1 (60%)

En cada uno de los siguientes apartados se tiene que indicar cuál es la variable aleatoria considerada, cuál es su distribución, así como todos los cálculos.

1) Una empresa de desarrollo de software de código abierto monitorea la calidad de los nuevos parches que se integran al repositorio principal. En un día determinado, el equipo de integración ha programado revisar un total de $n = 20$ parches de código. Suponiendo que cada parche es independiente de los demás y que la probabilidad de que un parche contenga un error crítico (critical bug) es $p = 0.10$.

a) ¿Cuál es la probabilidad de que se encuentren exactamente 3 parches con errores críticos? (10%)
b) ¿Cuál es la probabilidad de que se encuentren más de 3 parches con errores críticos? (10%)

c) ¿Cuántos parches se puede esperar que contengan errores críticos? (10%)

2) Ahora el equipo revisa los parches secuencialmente. La probabilidad de que un parche contenga un error crítico es $p = 0.10$.

a) ¿Cuál es la probabilidad de que el quinto parche revisado sea el primer parche en contener un error crítico? (10%)

b) ¿Cuál es la probabilidad de que el primer parche con un error crítico que aparezca sea antes del sexto parche? (10%)

Observación: Si Y es una variable aleatoria geométrica de parámetro p para calcular $P(Y = k)$ usaremos en R la función $dgeom(x = k - 1, prop = p)$. Fijaos que en la función tenemos que poner $k - 1$.

3) Se ha determinado que el número de errores críticos que se detectan en una semana sigue una distribución de Poisson con una media de 3 errores.

¿Cuál es la probabilidad de que, en una semana seleccionada al azar, se detecte más de 4 errores críticos en los parches integrados? (10%)

Solución:

1) Sea Y="número de parches con errores críticos de los 20 que se revisan". La probabilidad de que un parche contenga un error crítico (critical bug) es $p = 0.10$. Entonces la variable aleatoria Y sigue una distribución binomial de parámetros n=20 y p=0.10.

$$Y \sim \mathrm{B}(n = 20, p = 0.10)$$

Su función de masa de probabilidad viene dada por:

$$P(Y = k) = \binom{20}{k} \cdot 0.10^k \cdot (1 - 0.10)^{20-k}, \qquad k = 0,1,2,\ldots,20$$

a) Nos piden $P(Y = 3)$. La podemos calcular a partir de la función de masa de probabilidad de la distribución binomial:

$$P(Y = 3) = \frac{20!}{3! \cdot (20 - 3)!} 0.10^3 \cdot (1 - 0.10)^{20-3} = 0.1901199$$

```
factorial(20)/(factorial(3)*factorial(17))*(0.10^3)*
(1-0.10)^{17}

## [1] 0.1901199
```

Esta probabilidad, con las funciones de R para la distribución binomial, se puede calcular así:

- $P(Y = 3) = 0.1901199$

```
dbinom(3, size = 20, prob = 0.10)

## [1] 0.1901199
```

b) Ahora nos piden:

- $P(Y > 3) = 0.1329533$
- Con la instrucción pbinomial de R:

```
pbinom(3,size=20, prob=0.10,lower.tail=FALSE)

## [1] 0.1329533
```

c) El valor esperado será: $E(X) = n \cdot p = 20 \cdot 0.10 = 2$

```
20*0.10

## [1] 2
```

2) Sea Y la variable aleatoria que indica el número de parches que tenemos que examinar hasta encontrar el primer parche con un error crítico. Ahora tenemos que considerar la distribución geométrica con parámetro $p = 0.10$ y cuya función de masa de probabilidad viene dada por:

$$P(Y = k) = (1 - 0.10)^{k-1} \cdot 0.10, \qquad k = 1,2,\ldots$$

a) Nos piden la probabilidad:

$$P(Y = 5) = (1 - 0.10)^{5-1} \cdot 0.10 = 0.06561$$

```
  dgeom(4, prob=0.10)

## [1] 0.06561
```

b) Ahora nos piden:

$$P(Y < 6) = P(Y \leq 5) = \sum_{k=1}^{5} (1 - 0.10)^{k-1} \cdot 0.10$$

que podemos calcular con R mediante las siguientes instrucciones:

- $P(Y < 6) = P(Y \leq 5) = P(Y = 1) + P(Y = 2) + P(Y = 3) + P(Y = 4) + P(Y = 5) =$

```
  sum(dgeom(0:4, prob=0.10))

## [1] 0.40951
```

O también directamente:

- $P(Y < 6) = P(Y \leq 5) = 0.40951$

```
  pgeom(4, prob=0.10)

## [1] 0.40951
```

3) Sea X = “número de errores críticos que se detectan en una semana”

$$X \sim \text{Poisson}(\lambda = 3)$$

La fórmula de probabilidad de Poisson es: $P(X = k) = \frac{\lambda^k e^{-\lambda}}{k!}$

En nuestro caso:

$$P(X = k) = \frac{e^{-3} \cdot 3^k}{k!}, \qquad k = 0,1,2,\ldots$$

Nos piden la probabilidad: $P(X > 4)$ que podemos calcular con R de las siguientes formas:

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

Donde:

- $P(X \leq 4) = P(X = 0) + P(X = 1) + P(X = 2) + P(X = 3) + P(X = 4) = 0.8152632$

```
sum(dpois(0:4, lambda=3))

## [1] 0.8152632
```

Con lo que:

$$P(X > 4) = 1 - 0.8152632 = 0.1847368$$

O directamente con:

```
ppois(4, lambda=3, lower.tail=FALSE)

## [1] 0.1847368
```

Pregunta 2 (40%)

Un equipo de desarrollo ha implementado un nuevo algoritmo de ordenación. Tras realizar múltiples pruebas en un conjunto de datos grande y homogéneo, han determinado que el tiempo de ejecución del algoritmo, X (medido en milisegundos), sigue una distribución normal con una media (μ) de 450 ms y una desviación típica (σ) de 25 ms.

a) ¿Cuál es la probabilidad de que el algoritmo tarde menos de 400 ms en ejecutarse? (10%)
b) ¿Cuál es la probabilidad de que el algoritmo tarde más de 500 ms en ejecutarse? (10%)
c) ¿Entre qué dos tiempos (simétricos respecto a la media) se encuentra el 90% de los tiempos de ejecución?(20%)

Solución:

a) La probabilidad pedida será: $P(X < 400) = 0.02275013$

```
pnorm(400, mean= 450, sd = 25,lower.tail = TRUE)

## [1] 0.02275013
```

b) Ahora la probabilidad pedida se calcula así:

$$P(X > 500) = 0.02275013$$

```r
pnorm(500, mean= 450, sd = 25,lower.tail = FALSE)

## [1] 0.02275013
```

c) Buscamos los valores x_1 y x_2 tales que $P(x_1 < X < x_2) = 0.90$. Esto significa que las colas deben sumar $1 - 0.90 = 0.10$. Como es simétrico, cada cola es de $0.10/2 = 0.05$.

Utilizando la instrucción de R tenemos que:

```r
qnorm(c(0.05, 0.95), mean = 450, sd = 25)

## [1] 408.8787 491.1213
```

Con lo que el 90% de los tiempos de ejecución del algoritmo se encuentran entre 408.875 ms y 491.125 ms.