Tema 3: Variable Aleatoria

(Nota acerca de la notación: log *se refiere al logaritmo neperiano*. Aunque no hay problema si preferís usar ln, log es la notación matemática internacional y es la más recomendada)

1. [Spiegel 2.2, 2.4]
   (a) Encuentra la función de probabilidad de niños y niñas en familas con 3 hijos, suponiendo igual probabilidad para niños y niñas.
   (b) Encuentra la función de distribución para la variable aleatoria del apartado a).
   (c) Dibuja la función de distribución.
   (d) Imagínate que el único dato del que partes es la función de distribución. ¿Cómo podrías hallar la función de probabilidad?

Solution:

X	0	1	2	3
$P(x)$	1/8	3/8	3/8	1/8

$$F(x) = \begin{cases} 0 & x < 0 \\ 1/8 & 0 \leq x < 1 \\ 4/8 & 1 \leq x < 2 \\ 7/8 & 2 \leq x < 3 \\ 1 & x \geq 3 \end{cases}$$

2. 3 baterías se eligen al azar entre un grupo de 3 nuevas, 4 usadas pero que funcionan, y 5 baterías defectuosas. Sea X "el número de baterías nuevas"; sea Y "el número de baterías usadas pero funcionando". ¿Cuál es la distribución conjunta de X e Y ? ¿Y las marginales? Calcúlalas mediante razonamiento directo y marginalizando p(x, y) en la tabla.

Solution:

$$p(x,y) = \begin{cases} \frac{\binom{3}{x}\binom{4}{y}\binom{5}{3-x-y}}{\binom{12}{3}} & \text{para } 0 \leq x+y \leq 3. \\ 0 & e.o.c. \end{cases}$$

$$p(x) = \frac{\binom{3}{x}\binom{9}{3-x}}{\binom{12}{3}} \qquad p(y) = \frac{\binom{4}{y}\binom{8}{3-y}}{\binom{12}{3}}.$$

3. Se realizan 12 lanzamientos de una moneda equilibrada. Determinar la distribución (a) del número de caras obtenidas. (b) de la diferencia entre caras y cruces (pista: nº de cruces = 12 - nº de caras). En cada caso, representar la función de probabilidad para caracterizar la distribución.

Solution: (a) $P(X = k) = \binom{12}{k}(0.5)^k(0.5)^{12-k} = \binom{12}{k}(0.5)^{12}$.
(b) $P(X = k) = \binom{12}{6+k/2}0.5^{12} \qquad k = -12, -10, \cdots, 10, 12.$

4. Se lanzan dos monedas con probabilidades de cara p_1 y p_2. a) Estudiar el número de lanzamientos N que se realizan hasta que aparece cara en ambas monedas simultáneamente. b) ¿Cuál es número esperando de lanzamientos si $p_1 = 0.5 = p_2$.

Solution: a) $P(N = n) = (1 - p_1p_2)^{n-1}p_1p_2 \qquad n = 1, 2, 3, \cdots$ b) 4.

5. La función de probabilidad conjunta de dos variables aleatorias X e Y está dada por

$$f(x, y) = \begin{cases} cxy & x = 1, 2, 3 \text{ e } y = 1, 2, 3 \\ 0 & \text{en otro caso} \end{cases}$$

Calcula (a) la constante c, (b) $P(X = 2, Y = 3)$, (c) $P(X \leq 2, Y \leq 2)$, (d) $P(X \geq 2)$, (e) $P(Y \leq 2)$, (f) $P(X = 1)$, (g) $P(Y = 3)$

Solution: (a) 1/36, (b) 1/6, (c) 1/4, (d) 5/6, (e) 1/2, (f) 1/6, (g) 1/2

6. Para la distribución del problema 5, encuentra la función de densidad condicional de (a) X dado Y, (b) Y dado X.

Solution: (a) $f(x \mid y) = f_x(x)$ para $x = 1, 2, 3$. (b) $f(y \mid x) = f_y(y)$ para $y = 1, 2, 3$.

7. Encuentra las funciones de probabilidad marginal de (a) X e (b) Y, para las variables aleatorias del problema anterior. (c) Determina si X e Y son independientes.

Solution: (a)

$$f_x(x) = \begin{cases} x/6 & x = 1, 2, 3 \\ 0 & \text{en otro caso} \end{cases}$$

(b)

$$f_y(y) = \begin{cases} y/6 & y = 1, 2, 3 \\ 0 & \text{en otro caso} \end{cases}$$

(c) sí lo son.

8. Una máquina fabrica tornillos de los cuales el 10% son defectuosos. Encuentra la probabilidad de que en una muestra tomada al azar de 400 tornillos (a) como máximo 30, (b) entre 30 y 50, (c) 65 o más, de los tornillos sean defectuosos.

Solution: (a) 0.0524, (b) 0.9207, (d) $7.036756 \cdot 10^{-5}$

9. Cinco bolas se colocan al azar en tres urnas. Hallar la distribución del número de urnas que contienen una bola.

Solution: El problema se soluciona al darse cuenta que, de los 3^5 casos posibles:

- Hay 3 que corresponden a que todas las bolas estén en la misma urna (¿Por qué?).

- Hay 30 casos con 4 bolas en una urna (¿Por qué?).
- Hay 60 casos con 3 bolas en una urna y dos en otra (¿Por qué?).
- Hay 60 casos con 4 bolas en una urna y una en cada una de las dos restantes (¿Por qué?).
- Hay 90 casos con dos bolas en dos urnas y una en la tercera.

Por tanto:

$$p(x) = \begin{cases} (3+60)/3^5 & x = 0 \\ (30+90)/3^5 & x = 1 \\ (60)/3^5 & x = 2 \end{cases}$$

10. De una urna que contiene a bolas blancas y b negras se hacen extracciones sin reemplazamiento. Sea X_i el número de extracción en que aparece la i-ésima bola blanca. Calcula (a) La distribución conjunta de X_1 y X_2. (b) La distribución de X_1 condicionada por X_2.

Solution:

$$P(X_1 = j, X_2 = k) = \binom{a+b-k}{a-2} \Big/ \binom{a+b}{a}$$

$$P(X_1 = j \mid X_2 = k) = \frac{1}{k-1}$$

11. Una variable aleatoria (VA) X tiene función de densidad

$$f(x) = \begin{cases} ce^{-3x} & x \geq 0 \\ 0 & x < 0. \end{cases}$$

Calcula (a) la constante c, (b) $P(1 < X < 2)$, (c) $P(X \geq 3)$, (d) $P(X < 1)$, (e) La función de distribución de X, (f) Representa gráficamente las funciones de distribución y densidad de manera que describa la relación entre ellas.

Solution: (a) 3, (b) $e^{-3} - e^{-6}$, (c) e^{-9}, (d) $1 - e^{-3}$, (e) $F(x) = \begin{cases} 0 & x < 0 \\ 1 - e^{-3x} & x \geq 0 \end{cases}$

12. Una variable aleatoria (VA) X tiene función de densidad

$$f(x) = \begin{cases} cx^2 & 1 \leq x \leq 2 \\ cx & 2 \leq x \leq 3 \\ 0 & x < 0. \end{cases}$$

Calcula (a) la constante c, (b) $P(X > 2)$, (c) $P(1/2 < X < 3/2)$, (d) la función de distribución de X, (e) Representa gráficamente las funciones de distribución y densidad de manera que describa la relación entre ellas.

Solution: (a) 6/29, (b), 15/29, (c) 19/116 (d)

$$F(x) = \begin{cases} 0 & x < 1 \\ (2x^3 - 2)/29 & 1 \le x \le 2 \\ \frac{2}{29} + \frac{3}{29}x^2 & 2 \le x \le 3 \\ 1 & x \ge 3 \end{cases}$$

13. Un puente es bombardeado. Por ser muy estrecho, podemos aproximarlo por una línea unidimensional y, por tanto, describirlo mediante una sola coordenada. El puente se encuentra entre los puntos a y b. ¿Cuáles son las fuciones de densidad y de densidad acumulada del puente? ¿Cuál es la probabilidad de alcanzar un segmento del puente de tamaño l?

Solution: Se trata de un problema de probabilidad geométrica:
$f(x) = \frac{1}{b-a}$,
$$F(x) = \begin{cases} 0 & x \le a \\ \frac{x-a}{b-a} & a < x < b \\ 1 & x \ge b \end{cases}$$
$P(\text{alcanzar}) = \frac{l}{b-a}$.

14. Una persona está jugando a los dardos con una diana de radio A. El centro de la diana tiene radio B. ¿Cuál es la distribución del radio? ¿Cuál es la probabilidad de dar en el centro de la diana? ¿Y de acertar en el segundo anillo (comprendido entre los radios B y C)?

Solution:

- Este es un problema de probabilidad geométrica (¿Por qué?). Para resolverlo fácilmente, consideremos
$$F(r) = P(R \le r) = \frac{\text{Área favorable}}{\text{Área total}} = \frac{r^2}{A^2},$$
lo que caracteriza la distribución de R.
- La probabilidad de acertar en el centro de la diana será $\frac{B^2}{A^2}$.
- $P(B \le R \le C) = F(C) - F(B) = \frac{C^2 - B^2}{A^2}$.

15. La demanda de combustible de una gasolinera, en m^3, tiene distribución

$$f(x) = \begin{cases} kx^2 & 0 \le x \le 27 \\ k27^4/x^2 & x \ge 27. \end{cases}$$

(a) Si la gasolinera tiene un depósito de 100 m^3, que se rellena al principio de cada semana, hallar la probabilidad de que se pueda atender toda la demanda semanal. (b) Cuál debería ser la capacidad del depósito para que haya probabilidad 0.95 de que no se agote el combustible en una semana.

Solution: (a) 0.798, (b) $c = 405\ m^3$.

16. La función de probabilidad conjunta de dos variables aleatorias X e Y está dada por

$$f(x,y) = \begin{cases} c(x^2+y^2) & 0 \le x \le 1 \text{ e } 0 \le y \le 1 \\ 0 & \text{en otro caso} \end{cases}$$

Calcula (a) la constante c, (b) $P(X < 1/2, Y > 1/2)$, (c) $P(1/4 < X < 3/4)$, (d) $P(Y < 1/2)$.

Solution: (a) 3/2, (b) 1/4, (c) 29/64, (d) 5/16.

17. Encuentra las funciones de distribución marginal de (a) X e (b) Y, para las variables aleatorias del problema anterior. (c) Determina si X e Y son independientes.

Solution: (a)

$$F_x(x) = \begin{cases} 0 & x < 0 \\ \frac{1}{2}(x^3+x) & 0 \le x < 1 \\ 1 & x \ge 1 \end{cases}$$

(b)

$$F_y(y) = \begin{cases} 0 & y < 0 \\ \frac{1}{2}(y^3+y) & 0 \le y < 1 \\ 1 & y \ge 1 \end{cases}$$

(c) sí lo son

18. Para la distribución del problema 16, encuentra la función de densidad condicional de (a) X dado Y, (b) Y dado X.

Solution: (a)

$$f(x \mid y) = \begin{cases} \frac{x^2+y^2}{y^2+1/3} & 0 \le x \le 1; 0 \le y \le 1 \\ 0 & x < 0 \text{ o } x > 1; 0 \le y \le 1 \end{cases}$$

(b)

$$f(y \mid x) = \begin{cases} \frac{x^2+y^2}{x^2+1/3} & 0 \le x \le 1; 0 \le y \le 1 \\ 0 & 0 \le x \le 1; y < 0 \text{ o } y > 1. \end{cases}$$

19. [Spiegel 2.14] La función de densidad conjunta de X e Y es

$$f(x,y) = \begin{cases} \frac{xy}{96} & 0 < x < 4, \quad 1 < y < 5 \\ 0 & \text{en otro caso} \end{cases}$$

Calcula $P(X + Y < 3)$

Solution: La clave para resolver este ejercicio es dibujar la región donde se verifica la desigualdad $X + Y < 3$, para luego hacer la integral adecuada. Obtendremos: $P(X + Y < 3) = \frac{1}{48}$.

20. Una persona tiene dos prótesis de pierna. Sea X la variable aleatoria: "duración de la prótesis izquierda (en años)" e Y la variable aleatoria: "duración de la prótesis derecha (en años)". La distribución conjunta de ambas variables es

$$f(x,y) = \begin{cases} ce^{-2x}e^{-3y} & x \geq 0, y \geq 0 \\ 0 & \text{en otro caso} \end{cases}$$

   (a) Calcula c.
   (b) Calcula la función de distribución marginal de X.
   (c) Calcula $Pr(1 < X < 3)$.
   (d) Calcula la probabilidad de que la prótesis izquierda dure al menos 3 años más que la prótesis derecha.

21. Sean X e Y variables aleatorias con función de densidad:

$$f(x,y) = \begin{cases} \frac{3}{5}x(x+y) & 0 \leq x \leq 1; 0 \leq y \leq 2 \\ 0 & \text{en otro caso} \end{cases}$$

   Calcula (a) $\mathbb{E}[X]$, (b) $\mathbb{E}[Y]$, (c) $\mathbb{E}[X+Y]$, (d) $\mathbb{E}[XY]$.

Solution: (a) 7/10, (b) 6/5, (c) 19/10, (d) 5/6.

22. En el problema 21, (a) $\mathbb{E}[X+Y] = \mathbb{E}[X] + \mathbb{E}[Y]$, (b) $\mathbb{E}[XY] = \mathbb{E}[X]\mathbb{E}[Y]$

Solution: (a) Sí, es una propiedad general de la esperanza. (b) No, ya que X e Y no son independientes.

23. Encuentra la media y la varianza de una distribución uniforme entre a y b

Solution: $\mu = (a+b)/2$, $\sigma^2 = (b-a)^2/12$.

24. Calcula la media y la varianza de X si

$$f(x) = \begin{cases} e^{-x} & x \geq 0 \\ 0 & \text{en otro caso} \end{cases}$$

Solution: $\mu = 1$, $\sigma^2 = 1$.

25. Un globo se infla hasta alcanzar un radio aleatorio R que toma valores de forma uniforme en $(0,2)$, sea V el volumen del globo (recuerda que $V = \frac{4}{3}\pi R^3$). Calcula el volumen esperado analíticamente y mediante simulaciones.

26. Sean X e Y variables aleatorias con función de densidad conjunta

$$f(x, y) = \begin{cases} x + y & 0 \leq x \leq 1, 0 \leq y \leq 1 \\ 0 & \text{en otro caso} \end{cases}$$

Calcula (a) la varianza de X, (b) la varianza de Y, (c) la covarianza entre X e Y, (d) la correlación entre X e Y.

Solution: (a) 11/144, (b) 11/144, (c) -1/144, (f) -1/11.

27. Tres jugadores A, B, C lanzan, por turnos, dos dados y gana el primero que consiga obtener 9 como suma de ambas puntuaciones. En cada tirada, el que lanza debe poner un euro en la mesa y el ganador se lo lleva todo. ¿Cuál es el beneficio esperado de cada jugador?

Solution: La probabilidad de que el ganador se produzca con el lanzamiento n es $(8/9)^{n-1} \cdot (1/9)$ (¿Por qué?). Para el caso de A:

- Si A gana en la tirada $n = 3k + 1$ gana $2k$ euros (con $k = 0, 1, ...$).
- Si A pierde en la tirada $n = 3k + 2$ paga $k + 1$ euros (con $k = 0, 1, ...$).
- Si A pierde en la tirada $n = 3k$ paga k euros (¡ojo!, con $k = 1, 2, ...$).

Esto nos permite plantear un sumatorio con R para calcular $\mathbb{E}[X_A]$.

```r
ganancias <- rep(NA, 300)
ks <- 0:99
ganancias[3 * ks + 1] <- 2 * ks
ganancias[3 * ks + 2] <- -(ks + 1)
ganancias[3 * (ks + 1)] <- -(ks + 1)

probs <- (8 / 9) ^ (1:length(ganancias) - 1) * (1 / 9)

esperanza_a <- sum(ganancias * probs)
print(esperanza_a)
```

Repitiendo para B y C obtenemos: $\mathbb{E}[X_A] \approx -0.344$, $\mathbb{E}[X_B] \approx 0.026$, $\mathbb{E}[X_C] \approx 0.318$. Nótese que $\mathbb{E}[X_A] + \mathbb{E}[X_B] + \mathbb{E}[X_C] = 0$. A esto se le llama **juego de suma cero** (lo que gana un jugador lo pierden los otros) y tiene importantes aplicaciones en **teoría de juegos**.

28. El número de pacientes que llega a las urgencias de un hospital tiene distribución de Poisson de parámetro $\lambda = 5$. Cada paciente tiene probabilidad $p = 0.1$ de requerir hospitalización. Calcula (a) La distribución del número de enfermos que ingresan en el hospital. (b) Si un día han ingresado 10 pacientes, hallar la probabilidad de que hayan llegado n personas al servicio de urgencia. Calcula el número esperado de llegadas en estas circunstancias.

Solution: (a) Para resolver este apartado debemos emplear probabilidades condicionadas para calcular la distribución conjunta de el número de pacientes que llegan a urgencias y el número de pacientes ingresados. Marginalizando con R llegamos a la solución del apartado.

```r
# x: ingresados, y: llegadas
f <- function(x, y) {
  ifelse(
    x > y,  # no puede haber más ingresados que llegadas
    0,
    dbinom(x, size=y, prob = 0.1) * dpois(y, lambda = 5)
  )
}

# Y puede llegar hasta infinito, pero dpois(lambda=5) es prácticamente cero por
# encima de 50. Usamos el siguiente grid como aproximación
x <- 0:20
y <- 0:50
tabla_ps <- outer(x, y, f)
p_x_values <- rowSums(tabla_ps)
```

La función de probabilidad se puede aproximar por la siguiente tabla:

X	0	1	2	3	4	≥ 5
$P(x)$	$\approx$0.607	$\approx$0.303	$\approx$0.076	$\approx$ 0.0126	$\approx$0.00158	$\approx$0

(b) Tenemos que calcular una esperanza usando probabilidades condicionadas (¿Por qué?). Una vez calculada la probabilidad condicionada necesaria podemos calcular la esperanza con un sumatorio. Obtenemos 14.5.

29. Una pieza de un ordenador y su repuesto tienen duraciones aleatorias X e Y, respectivamente. Su densidad es
$$f(x,y) = \lambda e^{-\lambda x} \mu e^{-\mu y} \qquad x > 0, y > 0.$$
Utilizadas sucesivamente, las piezas duran $T = X + Y$. De este periodo de tiempo, la pieza 1 es responsable en la proporción $P = \frac{X}{X+Y}$.
    - ¿Son independientes X e Y?
    - Escribe una función en R para calcular $f(t|p)$ mediante simulaciones.
    - Usando la función anterior, ¿cuál es la duración más probable, si sabemos que la pieza principal has sido responsable del 50% del tiempo total?

30. Cada vez que se tira de una manivela, una máquina produce cierta pieza redonda con un radio medio de $10\,cm$ y varianza $4\,cm^2$. La distribución de las longitudes es acampanada y simétrica.
    (a) Calcula el número de piezas con un radio entre 7.5 y $11\,cm$ que esperarías encontrarte entre un conjunto de 500 piezas.
    (b) Un trabajador necesita 10 piezas con un radio mayor de $10.5\,cm$ para construir una máquina. ¿Cuál es la probabilidad que necesite tirar más de 20 veces de la manivela?
    (c) Si obtienes dos piezas de la máquina, calcula la probabilidad de que una sea mayor que la otra por más de 2 cm.

    Resuelve el ejercicio analíticamente y mediante simulaciones.

31. Gosset cultiva T tomates en su huerto, teniendo una media de 20 tomates en él. Gosset verifica si hay defectos en sus tomates. La probabilidad de que un tomate tenga defectos es $p = 0.1$. Suponiendo que el hecho de que un tomate tenga defectos no influye en el resto de tomates, encuentra la cantidad esperada de tomates defectuosos mediante simulaciones.

Solution: $T \sim \mathcal{P}(20)$, ya que noy hay un límite superior para T. Sea D: nº de tomates defectuosos. D se distribuye como $D|T \sim \mathcal{B}(T,p)$. Simulando primero un vector de Ts y usando estos valores para muestrear $D \mid T$, podemos obtener muchas muestras de D. La media es $\mathbb{E}[D] = \lambda p = 2$.

32. Un agente de movilidad está monitorizando una intersección con el fin de detectar infracciones al volante (que ocurren con probabilidad 0.01). El agente quiere saber cuántos autos pasarán antes de que ocurra la primera de las infracciones. En particular,
    (a) le gustaría encontrar la probabilidad de que la primera infracción ocurra después de que haya pasado el automóvil número 30.
    (b) Después de que hayan pasado 20 coches sin una infracción, un segundo agente sustituye al primero. También le gustaría saber la probabilidad de que la primera violación en movimiento ocurra después del automóvil número 30 (contando desde su llegada). Dado que ya han pasado 20 coches sin ninguna infracción, esta probabilidad es igual a una probabilidad condicional.

Solution: La variable X: "nº de coches que han pasado antes de la primera infracción" tiene distribución $X \sim$ Geom(p=0.01). De donde a) $P(X \geq 30) = 1 - P(X < 30) = 1 - pgeom(29, 0.01) \approx 0.74$. b) Aquí se pide $P(X \geq 50 \mid X \geq 20) = \frac{1-pgeom(49,0.01)}{1-pgeom(19,0.01)} \approx 0.74$. Este ejercicio ilustra que la distribución geométrica *no tiene memoria* de eventos pasados (ver Wikipedia: *Memorylessness*).

33. El *Old Faithful* es un geyser en el parque de Yellowstone, en Estados Unidos. El *Old Faithful* tiene dos tipos de erupciones: erupciones largas, con probabilidad 0.6; y erupciones cortas, con probabilidad 0.4. Si la erupción es larga, la duración de la misma se puede aproximar por una distribución Normal de media 4 minutos y varianza 0.15. Si la erupción es corta, la duración de la misma se puede aproximar por una distribución Normal de media 2 minutos y varianza 0.2. Se pide:
    (a) La distribución de la variable aleatoria "duración de una erupción". (Pista: haz uso de la regla de la probabilidad total y de la probabilidad condicionada).
    (b) La duración media de una erupción.
    (c) Calcula la probabilidad de que una erupción dure más de 3 minutos.

Solution: (a)

$$f(x) = P(\text{Erupción corta}) f(x \mid \text{Erupción corta}) + P(\text{Erupción larga}) f(x \mid \text{Erupción larga})$$

$$f(x) = 0.4 \cdot \mathcal{N}(x \mid 2, 0.2) + 0.6 \cdot \mathcal{N}(x \mid 4, 0.15),$$

donde $\mathcal{N}(x \mid \mu, \sigma^2)$ representa la función de densidad (que es una función de x) de una normal con media μ y varianza σ^2.
(b)

$$\mathbb{E}[X] = \int_{-\infty}^{\infty} x f(x) dx = 0.4 \int_{-\infty}^{\infty} x \mathcal{N}(x \mid 2, 0.2) dx + 0.6 \int_{-\infty}^{\infty} x \mathcal{N}(x \mid 4, 0.15) dx = 0.4 \cdot 2 + 0.6 \cdot 4.$$

Nótese que no hace falta hacer las integrales, ya que son la definición de la esperanza de las normales.
(c) $P(X > 3) = \int_3^\infty f(x)dx = 0.4 \int_3^\infty \mathcal{N}(x \mid 2, 0.2)dx + 0.6 \int_3^\infty \mathcal{N}(x \mid 4, 0.15)dx$, de donde la probabilidad pedida es ≈ 0.6021.

34. Cierto test de Coeficiente Intelectual (CI) se divide en 2 partes. En cierta escuela, las puntuaciones de la parte de matemáticas, X, tienen media 25 y desviación típica 5. Las puntuaciones de la parte de lectura, Y, tienen distribución 55 y desviación típica 10 y, sorprendentemente, parecen ser independientes de los resultados en matemáticas. EL CI final se calcula como $CI = 2X + Y$. ¿Cuál es la probabilidad de que un alumno de la escela sea superdotado, $CI > 130$?

> **Solution:** Las puntuaciones de CI son uno de los ejemplos clásicos de normalidad, por lo que asumiremos normalidad para X e Y. Por otra parte, el enunciado afirma que X y Y son independientes, de donde se sigue que $CI \sim \mathcal{N}(2\mu_x + \mu_y, 4\sigma_x^2 + \sigma_y^2) = \mathcal{N}(105, 200)$. Por tanto,
>
> $$P(CI > 130) \approx 0.039.$$

35. Una bolsa tiene N bombones, siendo $N-1$ de chocolate negro y uno de ellos de riquísimo licor. Antonio está obsesionado con los bombones de licor. Razona detalladamente cuál es la distribución de la variable aleatoria X: número de extracciones hasta que se obtiene el bombón de licor...
    (a) ... si cada vez que Antonio saca un bombón de chocolate negro, lo devuelve a la bolsa.
    (b) ... si cada vez que Antonio saca un bombón de chocolate negro, lo deja fuera de la bolsa para descartarlo.

> **Solution:** (a) $P(X = k) = \left(1 - \frac{1}{N}\right)^{k-1} \frac{1}{N}$. (b)$P(X = k) = 1/N$

36. **La dama degustadora de té**. Este es un famoso problema enunciado por el gran Sir Ronald A. Fisher: Una dama afirma que al probar una taza de té con leche es capaz de saber si la leche se echó antes o después que la infusión de té. Consideremos el problema de diseñar un experimento mediante el cual esta afirmación pueda ser probada. Nuestro experimento consiste en mezclar ocho tazas de té, cuatro de una manera y cuatro en la otra, y presentarlas a la dama para que las pruebe en un orden aleatorio. La dama debe dividir las 8 tazas en dos conjuntos de 4: uno para té+leche y otro para leche+té.
    (a) Sea X: número de tazas de leche+té identificadas correctamente. Identifica la distribución de X asumiendo que la dama hace la selección al azar (la dama no es capaz de distinguir los dos tipos de infusiones, solo estaba presumiendo).
    (b) En base al punto anterior. Halla $P(X = k)$.

> **Solution:** (a) X es hipergeométrica: piensa que tenemos una urna con 4 leche+té y 4 té+leche; tras probarlos, y dado que la dama no es capaz de distinguir entre los dos grupos, simplemente saca de la urna 4 infusiones al azar.
> (b) Se sigue que $P(X = k) = \binom{4}{k}\binom{4}{4-k}/\binom{8}{4}$ para $k = 0, 1, 2, 3, 4$.

37. Una solicitud para un puesto de trabajo se acepta con probabilidad 0.15. Las evaluaciones se realizan de una en una, en orden de llegada, hasta completar el número de puestos que se ofrecen (10 puestos de trabajo). Encuentra la probabilidad de que no se necesiten evaluar más de 100 solicitantes para asignar todos los puestos.

> **Solution:** Se trata de una binomial negativa de parámetros 10 y 0.15. La probabilidad pedida es 0.9449.

38. Gertrude Cox va todos los días al banco. En el banco hay dos colas, una cola estándar y una cola prioritaria. Independientemente del número de personas en el banco, la cola estándar tiene un tiempo medio de espera de 10 minutos, mientras que la prioritaria tiene un tiempo medio de espera de 5 minutos. Desgraciadamente, el banco está intentando ahorrar costes y solo abre ocasionalmente la cola prioritaria. Cada día se decide si se abre la cola prioritaria con independencia de lo que haya sucedido en días anteriores. La probabilidad de abrir la cola prioritaria es 0.7. Gertrude es cliente VIP, así que tiene derecho a usar la cola prioritaria. Si Gertrude va al banco y encuentra la cola prioritaria abierta se pone en ella, en otro caso usa la cola estándar.
    (a) Sabiendo que los tiempos de espera se pueden modelar con una distribución exponencial, obtén la distribución del tiempo de espera (diario) de Gertrude.
    (b) Calcula la probabilidad de que Gertrude tenga que esperar más de 4 minutos en una visita al banco. Calcula el valor sin usar R, haciendo uso de las funciones de densidad/distribución de los chuletarios.
    (c) Calcula la probabilidad de que, en 7 visitas al banco, Gertrude sea atendida en menos de 4 minutos en tres ocasiones.
        i. Calcula el valor sin usar R, haciendo uso de las funciones de densidad/distribución de los chuletarios.
        ii. Indica el (único) comando de R que tendrías que usar para calcular esta misma probabilidad.

Solution:

(a) $f(x) = \frac{0.7}{5}\exp(-x/5) + \frac{0.3}{10}\exp(-x/10)$

(b) 0.5156.

(c) i. 0.28097; ii. *dbinom(3, 7, 0.4843).*

39. De cara a unas elecciones, se estima que los partidos A, B, C y D recibirán el 20, 25, 30 y 25% de los votos, respectivamente. En una muestra de 10 votantes, encuentra la probabilidad de que haya dos votantes del partido A, 2 votantes de B, tres votantes de C y tres votantes de D.

Solution: Si X_A, X_B, X_C, X_D denota el número de votantes en la muestra, entonces

$$(X_A, X_B, X_C, X_D) \sim \text{Multi}(10, 0.20, 0.25, 0.30, 0.25).$$

Por tanto $P(X_A = 2, X_B = 2, X_C = 3, X_D = 3) \approx 0.0266$.

40. Cierto sistema se basa en dos módulos independientes, X e Y. Un fallo de cualquier módulo provoca un fallo de todo el sistema. La vida útil de cada módulo tiene una distribución exponencial con parámetros $\lambda_x = 1$, $\lambda_y = 2$. Es decir

$$f(x) = \begin{cases} e^{-x} & x \geq 0 \\ 0 & x < 0. \end{cases} \qquad f(y) = \begin{cases} 2e^{-2y} & y \geq 0 \\ 0 & y < 0. \end{cases}$$

donde x e y se miden en años.
    (a) ¿Cuál es la probabilidad de que el sistema funcione al menos 2 años sin fallos? Haz los cálculos a mano.
    (b) Si el sistema falla durante los primeros 2 años, ¿cuál es la probabilidad de que falle debido a un fallo del componente Y pero no del componente X?

(c) En la fábrica del componente X se fabrican 1000 componentes al día. Se dice que cada uno de estos componentes X es defectuoso si falla en menos de 1 semana (puedes asumir que un año tiene 52 semanas). ¿Cuál es el número esperado de componentes defectuosos fabricados en un día? Resuelve el problema identificando todas las variables aleatorias de interés y sin usar simulaciones.

41. Jacob Bernoulli propuso el siguiente juego de dados. El jugador paga un euro y lanza un dado. Luego lanza un conjunto de n dados, donde n es el número que muestra el primer dado. El número total de puntos que muestran los n dados se utiliza para determinar el resultado del juego. Si el número es menor que doce pierde el euro apostado, mientras que si el número es igual o mayor que doce recibe dos euros. Calcula, usando simulaciones en R, la cantidad de dinero que esperarías ganar al jugar a este juego.

42. Los coches pasan por cierta intersección a un ritmo de 16 automóviles por minuto. ¿Cuál es la probabilidad de que al menos 1000 coches crucen la intersección en la próxima hora?

Solution: Sea X: nº de coches en una hora. Elegimos una distribución de Poisson para modelar X, ya que no hay un límite superior claro en los valores que puede tomar y además, nos dan un valor medio que permite calcular λ. Para λ, el número medio de coches en una hora será 960 de donde $X \sim \mathcal{P}(960)$. La probabilidad por la que se pregunta es

$$P(X \geq 1000) \approx 0.1018.$$

43. El tiempo medio que un estudiante tarda en completar un examen es 45 minutos. En una clase de 10 estudiantes, ¿cuál es la probabilidad de que al menos un estudiante acabe en menos de 20 minutos?

Solution: 0.9883 (asumiendo independencia)

44. Se tira una moneda (sin trucar) hasta que aparecen cuatro caras. Calcula, analíticamente y mediante simulaciones,

    (a) La probabilidad de que se necesiten 10 lanzamientos.

    (b) La probabilidad de que se necesiten al menos 10 lanzamientos.

    - Halla la función de probabilidad de Y: “número de cruces obtenidas” para el caso general en el que la probabilidad de cara sea p.

Solution: (a) 0.082, (b) 0.254, (c) $P(Y = y) = \binom{y+4-1}{3} p^4 (1-p)^y$.

45. ¿Cuántos lanzamientos, en promedio, debemos lanzar un dado hasta obtener el primer 6?

Solution: 6

46. Un método muy empleado para estimar el número de individuos de poblaciones esquivas es el de *captura-recaptura*. Para estimar la población total N, se capturan r individuos que se etiquetan y se devuelven a la población total. Más tarde, los investigadores recapturan, en d días distintos, muestras de tamaño n y se cuentan el número de individuos $K_1, K_2, ..., K_d$ que están etiquetados. Promediando los resultados de los d días, se obtiene la media $\bar{K}$.

Se aplica este método para contar el número de peces en un lago. En la “captura” se cuentan y etiquetan 600 peces. En las recapturas se pescan 800 peces, resultando en un media final de $\bar{K} = 100$ peces etiquetados. ¿Cuál es la población total de peces?¿En qué distribución te has basado para tus cálculos?

Solution: La distribución relevante es la geométrica. Dado que $\bar{K}$ tiende, por el teorema de los grandes números a la media real μ de la distribución geométrica

$$\bar{K} \approx \mu \Leftrightarrow 100 \approx 800\frac{600}{N}.$$

De aquí

$$N \approx 4800.$$

47. Las personas con tipo sanguíneo 0-negativo son donantes universales. En cierto país, el 7.2% de la población es de tipo 0-negativo. Un médico quiere encontrar a 10 personas 0-negativas, para lo que hace un cribado entre sus pacientes. ¿Cuál es la probabilidad de encontrar a estas 10 personas entre las primeras 100 personas cribadas?

Solution: 0.183

48. Cierta franquicia de comida rápida regala una figura de tu película favorita con cada comida. Hay 10 personajes en total. ¿Cuántas comidas tienes que comprar hasta conseguir los 10 personajes? Resuelve el problema mediante simulaciones.

Solution: 29.3 comidas

49. Cuando Fisher va al gimnasio realiza una de las siguientes actividades: correr, andar en bicicleta estática o remar. Él elige cada una de estas actividades con probabilidades 0.5, 0.3 y 0.2. Si elige correr o andar en bicicleta, la cantidad de tiempo (en minutos) que pasa haciendo ejercicio tiene distribución exponencial con parámetros 0.05 y 0.025, respectivamente. Si elige remar, solo rema durante 10 minutos, se detiene para tomar un vaso de agua y comienza de nuevo, eligiendo una de las tres actividades como si acabara de entrar por la puerta del gimnasio. Programa un script en R que permita calcular la esperanza del tiempo que pasa Fisher haciendo deporte en una visita al gimnasio.

50. (a) Aunque el peso y la altura de las personas suelen estar correladas, asumamos por el momento que ambas variables son independientes y con distribuciones normales. Concretamente, la altura de cierta población tiene media 1.65 m y varianza 0.0025 m^2, mientras que el peso tiene media 64 Kg y varianza 9 Kg^2. Por otra parte, el Índice de Masa Corporal (IMC) se define como

$$IMC = \frac{\text{peso}}{\text{estatura}^2}.$$

Escribe una función con prototipo *calc_imc_prob(imc, n_sims)* que calcule y devuelva, mediante *n_sims* simulaciones, la probabilidad de que el IMC exceda el valor indicado por el argumento *imc*.

(b) Para obtener un modelo más realista de las distribuciones del peso y altura necesitamos tener en cuenta las correlaciones entre ambas variables. Seguimos asumiendo que la altura tiene media 1.65 m y varianza 0.0025 m^2. Sin embargo, asumimos ahora que una persona que mida H tendrá un peso P con distribución normal con media $40H$ y varianza 1 Kg^2. Queremos verificar que la distribución de P de la población es normal con media 66 Kg y varianza 0.1 Kg^2. Para ello, dibuja la distribución de muestras obtenidas mediante simulaciones y compárala con la distribución teórica (superponiéndola al gráfico anterior). En este ejercicio, no es necesario implementar el código dentro de una función.

Fuentes

- [Spiegel]: Spiegel et al. Probabilidad y estadística, Schaum, segunda edición, 2003.