Ejercicios 3
Análisis de Datos

Denisse Garnica Sánchez
denisse.garnica@cimat.mx
Maestría en Computación
5 de Septiembre 2025

Problema 1.
Considera una secuencia de lanzamientos independientes de una moneda equilibrada.

1. Calcula la probabilidad de que en el lanzamiento número 20 se obtenga *por cuarta vez* águila.
2. En promedio, ¿cuántas veces se debe lanzar la moneda para obtener *por segunda vez* águila?

Solución.

a) Asumiendo una moneda justa, sea N el número de lanzamientos necesarios para observar la cuarta águila. El evento "$N = 20$" exige dos condiciones:

   i) en los primeros 19 lanzamientos hubo exactamente 3 águilas
   ii) el lanzamiento 20 es águila.

   Usando esto y observando que el evento puede ser modelado por medio de la distribución binomial, notemos que
$$\mathbb{P}(N = 20) = \binom{19}{3} \left(\tfrac{1}{2}\right)^{19} \left(\tfrac{1}{2}\right) = \binom{19}{3} 2^{-20}.$$
   Al saber que $\binom{19}{3} = 969$, llegamos a que $\mathbb{P}(N = 20) = 969 \cdot 2^{-20} \approx 9.24 \times 10^{-4}$. La cual es la probabilidad buscada.

b)
   - (Opción a:) El número de lanzamientos para observar la segunda águila se modela con la **distribución binomial negativa**. Esta distribución describe la probabilidad de que el r-ésimo éxito ocurra en el lanzamiento número n. En general, si cada intento tiene probabilidad de éxito p y de fracaso $q = 1 - p$, entonces
$$\mathbb{P}(N = n) = \binom{n-1}{r-1} p^r q^{n-r}, \qquad n = r, r+1, \dots$$
     donde N es el número de intentos hasta el r-ésimo éxito.
     En nuestro caso, $r = 2$ (queremos la segunda águila) y $p = \frac{1}{2}$. La función de probabilidad toma la forma
$$\mathbb{P}(N = n) = \binom{n-1}{1} \left(\tfrac{1}{2}\right)^2 \left(\tfrac{1}{2}\right)^{n-2} = (n-1)\, 2^{-n}, \qquad n = 2, 3, 4, \dots$$
     La **esperanza** de una variable binomial negativa está dada por la fórmula general
$$\mathbb{E}[N] = \frac{r}{p}.$$
     Sustituyendo $r = 2$ y $p = \frac{1}{2}$ obtenemos
$$\mathbb{E}[N] = \frac{2}{1/2} = 4.$$

Es decir, en promedio se requieren 4 lanzamientos para obtener la segunda águila.
Como verificación, podemos calcular directamente la esperanza mediante la suma infinita

$$\mathbb{E}[N] = \sum_{n=2}^{\infty} n\,\mathbb{P}(N=n) = \sum_{n=2}^{\infty} n(n-1)2^{-n}.$$

Esta serie converge a 4, lo cual coincide con la fórmula. Siendo este el resultado buscado.

- (Opción b:) Sea G_1 el número de lanzamientos hasta obtener la *primera* águila y G_2 el número de lanzamientos, a partir de ese momento, hasta obtener la *segunda* águila. Por independencia entre lanzamientos, G_1 y G_2 tienen la misma distribución y son independientes.
  La probabilidad de éxito por lanzamiento es p, luego usando la distribución geométrica

  $$\mathbb{P}(G_i = k) = (1-p)^{k-1}p, \quad k = 1, 2, \ldots$$

  y calculando su esperanza, obtenemos

  $$\mathbb{E}[G_i] = \sum_{k\geq 1} k(1-p)^{k-1}p = \frac{1}{p}.$$

  El número total de lanzamientos hasta la segunda águila es

  $$N = G_1 + G_2$$

  luego, por linealidad de la esperanza,

  $$\mathbb{E}[N] = \mathbb{E}[G_1] + \mathbb{E}[G_2] = \frac{1}{p} + \frac{1}{p} = \frac{2}{p}.$$

  Para moneda equilibrada $p = \frac{1}{2}$ se obtiene $\mathbb{E}[N] = 4$.

■

Problema 2.
El número de accidentes por semana en una glorieta es una v.a. X con la siguiente distribución

$$\mathbb{P}(X = x) = \begin{cases} 0.20, & x = 0, \\ 0.30, & x = 1, \\ 0.30, & x = 2, \\ 0.15, & x = 3, \\ 0.05, & x = 4, \\ 0, & \text{en otro caso.} \end{cases}$$

Suponiendo independencia entre semanas, calcula la probabilidad de que en un año (52 semanas) haya exactamente 3 semanas con menos de 2 accidentes.

Solución. Definimos el evento semanal $A = \{X < 2\} = \{X = 0 \text{ o } X = 1\}$. Su probabilidad es

$$q = \mathbb{P}(A) = \mathbb{P}(X=0) + \mathbb{P}(X=1) = 0.20 + 0.30 = 0.50.$$

Sea Y el número de semanas (entre 52) en que ocurre A. Por independencia semanal, $Y \sim \mathrm{Bin}(n = 52, p = 0.5)$. Buscamos $\mathbb{P}(Y = 3)$:

$$\mathbb{P}(Y=3) = \binom{52}{3}(0.5)^3(0.5)^{49} = \binom{52}{3}2^{-52}.$$

Con $\binom{52}{3} = 22100$ se obtiene

$$\mathbb{P}(Y=3) = 22100 \cdot 2^{-52} \approx 4.89 \times 10^{-12}.$$

■

Problema 3.
Tienes 100 *pares* diferentes de calcetines (en total 200 calcetines). Un incendio destruye 30 calcetines al azar de esta colección ("al azar" significa que se elige uniformemente un subconjunto de tamaño 30 sin reemplazo de entre los 200 calcetines).
Para cada par $i \in \{1, \ldots, 100\}$, define la variable indicadora

$$X_i = \mathbf{1}_{\{\text{el par } i \text{ queda intacto}\}} = \begin{cases} 1, & \text{si ninguno de los dos calcetines del par } i \text{ es destruido,} \\ 0, & \text{en otro caso.} \end{cases}$$

Calcula $\mathbb{E}[X_i]$. ¿Son independientes las variables X_i (da un argumento informal, o muestra que no lo son)? ¿Cuántos pares de calcetines se quedan intactos en promedio?

Solución. **1) Cálculo de $\mathbb{E}[\mathbf{X_i}]$.** El evento $\{X_i = 1\}$ equivale a que, entre los 30 calcetines destruidos, no aparezca ninguno de los dos calcetines del par i. Como la elección es uniforme entre todos los subconjuntos de tamaño 30 de una población de 200 calcetines, la probabilidad de que *ninguno* de los dos caiga en el subconjunto destruido es

$$\mathbb{P}(X_i = 1) = \frac{\binom{200-2}{30}}{\binom{200}{30}} = \frac{\binom{198}{30}}{\binom{200}{30}}.$$

Dado que X_i es indicadora,

$$\mathbb{E}[X_i] = \mathbb{P}(X_i = 1) = \frac{\binom{198}{30}}{\binom{200}{30}} = \prod_{t=0}^{29} \frac{198 - t}{200 - t}.$$

(Aproximación numérica) $\mathbb{E}[X_i] \approx 0.7218592965$.

2) (No) independencia entre los $\mathbf{X_i}$. Intuitivamente, las X_i no son independientes porque el muestreo es *sin reemplazo*: si ya sabemos que un par quedó intacto (ninguno de sus 2 calcetines se destruyó), esa información modifica ligeramente las probabilidades de destrucción para los demás pares.
Una manera formal y breve de ver que no hay independencia es comparar

$$\mathbb{P}(X_i = 1, X_j = 1) = \frac{\binom{196}{30}}{\binom{200}{30}} \quad \text{con} \quad \mathbb{P}(X_i = 1)\,\mathbb{P}(X_j = 1) = \left(\frac{\binom{198}{30}}{\binom{200}{30}} \right)^2.$$

Si fueran independientes, estas cantidades serían iguales. En realidad,

$$\mathbb{P}(X_j = 1 \mid X_i = 1) = \frac{\binom{196}{30}}{\binom{198}{30}} < \frac{\binom{198}{30}}{\binom{200}{30}} = \mathbb{P}(X_j = 1),$$

porque para cada término del producto $\frac{196-t}{198-t} < \frac{198-t}{200-t}$ cuando $t = 0, 1, \ldots, 29$. Por lo tanto, $\mathbb{P}(X_i = 1, X_j = 1) \neq \mathbb{P}(X_i = 1)\mathbb{P}(X_j = 1)$ y las X_i *no* son independientes (de hecho están *negativamente* correlacionadas).

3) Número esperado de pares intactos. Sea $S = \sum_{i=1}^{100} X_i$ el número total de pares que quedan intactos. Por linealidad de la esperanza (la cual no requiere independencia),

$$\mathbb{E}[S] = \sum_{i=1}^{100} \mathbb{E}[X_i] = 100 \cdot \frac{\binom{198}{30}}{\binom{200}{30}} = 100 \cdot \prod_{t=0}^{29} \frac{198 - t}{200 - t}.$$

(Aproximación numérica) $\mathbb{E}[S] \approx 72.18592965$.

Resumen.

$$\boxed{\mathbb{E}[X_i] = \frac{\binom{198}{30}}{\binom{200}{30}} \approx 0.72186\,, \qquad X_i \text{ no son independientes,} \qquad \mathbb{E}[S] = 100\,\frac{\binom{198}{30}}{\binom{200}{30}} \approx 72.186\,.}$$

■