Modelos Estadísticos para la Regresión y la Clasificación
Continuación Práctico 2 - Probabilidad

Micaela Long

Instituto de Matemática y Estadística Prof. Rafael Laguardia (IMERL)
Facultad de Ingeniería, Universidad de la República, Uruguay

23 de agosto de 2024

Práctico 16/8:

- Ejercicio 1: Cambio de Variable.
- Cómo hacer simulaciones en R y Python (distribución uniforme).
- Esperanza condicional, fórmula de la esperanza total, ejemplos.
- Ejercicio 6: Esperanza condicional.

Ejercicio 2: Densidades condicionales.

Sugerencias:

- *Hallar K que asegure que la función de densidad conjunta $f(x, y)$ es una densidad.*
- *Usar las definiciones vistas en el teórico.*
- *Hacer cuentas!*

Recordar que el **coeficiente de correlación de Pearson** se define como

$$\rho_{XY} = \frac{\mathrm{Cov}(X, Y)}{\sqrt{\mathrm{Var}(X)\mathrm{Var}(Y)}} = \frac{\mathbb{E}[(X - \mathbb{E}[X])(Y - \mathbb{E}[Y])]}{\sqrt{\mathrm{Var}(X)\mathrm{Var}(Y)}}$$

- Magnitud de la covarianza está influenciada por la relación entre las variables y también por sus varianzas.
- La covarianza es difícil de interpretar directamente.
- $\rho_{XY} \in [-1, 1]$ y no tiene unidad de medida.
- ρ_{XY} indica la fuerza y dirección de la relación lineal entre dos variables:
  - $\rho_{XY} = 1$ las variables aumentan o disminuyen juntas de manera lineal.
  - $\rho_{XY} = 0$ no hay correlación lineal.
  - $\rho_{XY} = -1$ una variable aumenta mientras la otra disminuye de manera lineal.

Ejercicio 3: Descomposición de la varianza I

Se extraen sucesivamente, sin reposición, dos bolas de una urna que contiene tres blancas y dos negras. En relación a este experimento, se definen las siguientes variables:

$$X = \begin{cases} 0 & \text{si la primera bola es blanca} \\ 1 & \text{si la primera bola es negra} \end{cases}$$

$$Y = \begin{cases} 0 & \text{si la segunda bola es blanca} \\ 1 & \text{si la segunda bola es negra} \end{cases}$$

Descomponga la varianza de la variable X teniendo en cuenta el condicionamiento a Y.

Ejercicio 3

Descomposición de la varianza de la variable X condicionada a Y:

$$\mathrm{Var}(X) = \mathbb{E}[\mathrm{Var}(X \mid Y)] + \mathrm{Var}(\mathbb{E}[X \mid Y])$$

Ejemplo:

Si X fuera rendimiento académico de estudiantes de diferentes clases (definidas por Y):

- $\mathbb{E}[\mathrm{Var}(X \mid Y)]$ es la **variabilidad dentro de cada clase**.

  Calculamos la varianza de X dentro de cada subgrupo, y luego tomamos el promedio de esas varianzas.

  Es la cantidad promedio de variabilidad en el rendimiento académico, que no se explica por la variabilidad entre las distintas clases.

- $\mathrm{Var}(\mathbb{E}[X \mid Y])$ es la **variabilidad entre las clases**.

  Calculamos el valor promedio de X para cada valor de Y, y luego medimos cuánta variación hay entre estos promedios.

Ejercicio 3

3 bolas blancas, 2 negras. Sacamos 2.

$$X = \begin{cases} 0 & \text{si la primera bola es blanca} \\ 1 & \text{si la primera bola es negra} \end{cases}$$

$$Y = \begin{cases} 0 & \text{si la segunda bola es blanca} \\ 1 & \text{si la segunda bola es negra} \end{cases}$$

$$\mathrm{Var}(X) = \underbrace{\mathbb{E}[\mathrm{Var}(X \mid Y)]}_{(2)} + \underbrace{\mathrm{Var}(\mathbb{E}[X \mid Y])}_{(1)}$$

(1) $\mathbb{E}[X \mid Y]$:

$$\mathbb{E}[X \mid Y = y] = \sum_x x \cdot \mathbb{P}(X = x \mid Y = y)$$

$$\begin{aligned} \mathbb{E}[X \mid Y = 0] &= 0 \cdot \mathbb{P}(X = 0 \mid Y = 0) + 1 \cdot \mathbb{P}(X = 1 \mid Y = 0) \\ &= \frac{\mathbb{P}(X = 1, Y = 0)}{\mathbb{P}(Y = 0)} = \frac{3/10}{6/10} = \frac{1}{2} \end{aligned}$$

donde

$$\mathbb{P}(X = 1, Y = 0) = \mathbb{P}(X = 1)\mathbb{P}(Y = 0 \mid X = 1) = \frac{2}{5} \cdot \frac{3}{4} = \frac{6}{20} = \frac{3}{10}$$

$$\begin{aligned} \mathbb{P}(Y = 0) &= \mathbb{P}(Y = 0, X = 0) + \mathbb{P}(Y = 0, X = 1) \\ &= \mathbb{P}(X = 0)\mathbb{P}(Y = 0 \mid X = 0) + \mathbb{P}(X = 1)\mathbb{P}(Y = 0 \mid X = 1) \\ &= \frac{3}{5} \cdot \frac{2}{4} + \frac{3}{10} = \frac{6}{10} \end{aligned}$$

Ejercicio 3

De la misma forma se calcula $\mathbb{E}(X \mid Y = 1) = \frac{1}{4}$. Luego

$$\mathbb{E}(X \mid Y) = \begin{cases} \frac{1}{2} & \text{si } Y = 0 \\ \frac{1}{4} & \text{si } Y = 1 \end{cases}$$

Ahora calculamos $\mathrm{Var}(\mathbb{E}(X \mid Y))$:

$$\begin{aligned}
\mathrm{Var}(\mathbb{E}(X \mid Y)) &= \overbrace{\mathbb{E}((\mathbb{E}(X \mid Y) - \mathbb{E}(X))^2)}^{\mathbb{E}(g(Y))} \\
&= g(0) \cdot \mathbb{P}(Y = 0) + g(1) \cdot \mathbb{P}(Y = 1) \\
&= (\mathbb{E}(X \mid Y = 0) - \mathbb{E}(X))^2 \cdot \mathbb{P}(Y = 0) + (\mathbb{E}(X \mid Y = 1) - \mathbb{E}(X))^2 \cdot \mathbb{P}(Y = 1) \\
&= \left(\frac{1}{2} - \mathbb{E}(X)\right)^2 \cdot \frac{6}{10} + \left(\frac{1}{4} - \mathbb{E}(X)\right)^2 \cdot \mathbb{P}(Y = 1)
\end{aligned}$$

donde

$$\mathbb{E}(X) = 0 \cdot \mathbb{P}(X = 0) + 1 \cdot \mathbb{P}(X = 1) = \frac{2}{5}$$

$$\mathbb{P}(Y = 1) = \frac{4}{10} \quad \text{chequear!}$$

y por tanto

$$\mathrm{Var}(\mathbb{E}(X \mid Y)) = \left(\frac{1}{2} - \frac{2}{5}\right)^2 \cdot \frac{6}{10} + \left(\frac{1}{4} - \frac{2}{5}\right)^2 \cdot \frac{4}{10} = \frac{3}{200}$$

Ejercicio 3

Otra forma de calcular $\mathrm{Var}(\mathbb{E}(X \mid Y))$:

$$\mathrm{Var}(\mathbb{E}(X \mid Y)) = \mathbb{E}\left(\mathbb{E}(X \mid Y)^2\right) - [\mathbb{E}(\mathbb{E}(X \mid Y))]^2$$

$$\mathbb{E}(X \mid Y) = \begin{cases} \frac{1}{2} & \text{si } Y = 0 \\ \frac{1}{4} & \text{si } Y = 1 \end{cases} \quad \Rightarrow \quad \mathbb{E}(X \mid Y)^2 = \begin{cases} \frac{1}{4} & \text{si } Y = 0 \\ \frac{1}{16} & \text{si } Y = 1 \end{cases}$$

$$\mathbb{E}\left(\mathbb{E}(X \mid Y)^2\right) = \mathbb{P}(Y = 0)\frac{1}{4} + \mathbb{P}(Y = 1)\frac{1}{16} = \frac{6}{10} \cdot \frac{1}{4} + \frac{4}{10} \cdot \frac{1}{16} = \frac{7}{40}$$

Por otro lado, usando la fórmula de la esperanza total:

$$[\mathbb{E}(\mathbb{E}(X \mid Y))]^2 = [\mathbb{E}(X)]^2 = \left(\frac{2}{5}\right)^2 = \frac{4}{25}$$

Entonces

$$\mathrm{Var}(\mathbb{E}(X \mid Y)) = \mathbb{E}\left(\mathbb{E}(X \mid Y)^2\right) - [\mathbb{E}(\mathbb{E}(X \mid Y))]^2 = \frac{7}{40} - \frac{4}{25} = \frac{15}{1000} = \frac{3}{200}$$

Ejercicio 3

(2) $\mathbb{E}(\mathrm{Var}(X \mid Y))$

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

Hay que calcular

$$\mathrm{Var}(X \mid Y=0) = \mathbb{E}(X^2 \mid Y=0) - [\mathbb{E}(X \mid Y=0)]^2$$

$$\mathrm{Var}(X \mid Y=1) = \mathbb{E}(X^2 \mid Y=1) - [\mathbb{E}(X \mid Y=1)]^2$$

Notar que

$$X = \begin{cases} 0 & \text{con probabilidad } \frac{3}{5} \\ 1 & \text{con probabilidad } \frac{2}{5} \end{cases} = X^2$$

Entonces

$$\begin{aligned} \mathrm{Var}(X \mid Y=0) = \mathbb{E}(X^2 \mid Y=0) - [\mathbb{E}(X \mid Y=0)]^2 &= \mathbb{E}(X \mid Y=0) - [\mathbb{E}(X \mid Y=0)]^2 \\ &= \frac{1}{2} - \frac{1}{4} \\ &= \frac{1}{4} \end{aligned}$$

$$\begin{aligned} \mathrm{Var}(X \mid Y=1) = \mathbb{E}(X^2 \mid Y=1) - [\mathbb{E}(X \mid Y=1)]^2 &= \mathbb{E}(X^2 \mid Y=1) - [\mathbb{E}(X \mid Y=1)]^2 \\ &= \frac{1}{4} - \frac{1}{16} \\ &= \frac{3}{16} \end{aligned}$$

Ejercicio 3

Luego

$$\begin{aligned}\mathbb{E}(\mathrm{Var}(X \mid Y)) &= \mathbb{P}(Y=0)\mathrm{Var}(X \mid Y=0) + \mathbb{P}(Y=1)\mathrm{Var}(X \mid Y=1) \\ &= \frac{6}{10}\cdot\frac{1}{4} + \frac{4}{10}\cdot\frac{3}{16} \\ &= \frac{6}{40} + \frac{3}{40} \\ &= \frac{9}{40}\end{aligned}$$

Por tanto

$$\begin{aligned}\mathrm{Var}(X) &= \mathbb{E}[\mathrm{Var}(X \mid Y)] + \mathrm{Var}(\mathbb{E}[X \mid Y]) \\ &= \frac{9}{40} + \frac{3}{200} \\ &= \frac{6}{25}\end{aligned}$$

Calcular $\mathrm{Var}(X)$ directamente para ver que da lo mismo!

Ejercicio 4: Descomposición de la varianza II

Sea (X, Y) un vector aleatorio con función de densidad $f(x, y) = 2$ para $0 < x < y < 1$.

1. Calcular $E(X^3 \mid Y)$.
2. Descomponga la varianza de la variable Y.

El dominio de integración es el interior de un triángulo con vértices (0,0), (0,1) y (1,1).
Para la parte 2

$$\mathrm{Var}(Y) = \mathbb{E}(\mathrm{Var}(Y|X)) + \mathrm{Var}(\mathbb{E}(Y|X))$$

Pensar como se distribuye $Y|X = x$.