Probabilidad

Foto: José Juan Castro

Los Del DGIIM, `losdeldgiim.github.io`

Doble Grado en Ingeniería Informática y Matemáticas
Universidad de Granada

Probabilidad

Los Del DGIIM, `losdeldgiim.github.io`

Arturo Olivares Martos

Granada, 2024-2025

Índice general

1. Relaciones de problemas

1.1. Esperanza Condicionada

Ejercicio 1.1.1. Sea X una variable aleatoria que se distribuye uniformemente en el intervalo $]0,1[$. Comprobar si las variables aleatorias X y $|{}^1\!/\!_2 - X|$ son incorreladas.

Ejercicio 1.1.2. Calcular las curvas de regresión y las razones de correlación para las siguientes distribuciones, comentando los resultados.

1. Considerar las distribución conjunta (X,Y) con función de masa de probabilidad dada por:

$X \setminus Y$	10	15	20	
1	0	$2/6$	0	$2/6$
2	$1/6$	0	0	$1/6$
3	0	0	$3/6$	$3/6$
	$1/6$	$2/6$	$3/6$	1

Tras haber calculado las distribuciones marginales, calculamos ahora las distribuciones condicionadas. La siguiente tabla muestra la distribución condicionada de Y dado X, $P[Y=y \mid X=x]$:

$X \setminus Y$	10	15	20
1	0	1	0
2	1	0	0
3	0	0	1

La distribución condicionada de X dado Y, $P[X=x \mid Y=y]$ viene dada por la misma tabla, ya que en este caso tenemos que:

$$P[X=x \mid Y=y] = P[Y=y \mid X=x] \qquad \forall x,y$$

Calculemos ahora las curvas de regresión. La curva de regresión de Y sobre X es:

$$\widehat{Y}(x) = E[Y \mid X=x] = \sum_y yP[Y=y \mid X=x] \qquad \forall x \in E_x$$

Por tanto, la curva de regresión de Y sobre X es:

$$\widehat{Y}(1) = 15$$
$$\widehat{Y}(2) = 10$$
$$\widehat{Y}(3) = 20$$

La curva de regresión de X sobre Y es:

$$\widehat{X}(y) = E[X \mid Y=y] = \sum_x xP[X=x \mid Y=y] \qquad \forall y \in E_y$$

Por tanto, la curva de regresión de X sobre Y es:

$$\widehat{X}(10) = 2$$
$$\widehat{X}(15) = 1$$
$$\widehat{X}(20) = 3$$

Calculemos ahora las razones de correlación. La razón de correlación de Y sobre X es:

$$\eta_{Y/X} = \frac{\mathrm{Var}(E[Y \mid X])}{\mathrm{Var}(Y)}$$

- Calculemos $E[Y]$:

$$\begin{aligned} E[Y] &= \sum_y yP[Y=y] \\ &= 10 \cdot {}^1\!/_6 + 15 \cdot {}^2\!/_6 + 20 \cdot {}^3\!/_6 \\ &= {}^{10}\!/_6 + 5 + 10 \\ &= \frac{50}{3} \end{aligned}$$

- Calculemos ahora $E[Y^2]$:

$$\begin{aligned} E[Y^2] &= \sum_y y^2 P[Y=y] \\ &= 10^2 \cdot {}^1\!/_6 + 15^2 \cdot {}^2\!/_6 + 20^2 \cdot {}^3\!/_6 \\ &= 100 \cdot {}^1\!/_6 + 225 \cdot {}^2\!/_6 + 400 \cdot {}^3\!/_6 \\ &= {}^{100}\!/_6 + 75 + 200 \\ &= \frac{875}{3} \end{aligned}$$

- Calculemos ahora $E[(E[Y \mid X])^2]$:

$$\begin{aligned} E[(E[Y \mid X])^2] &= \sum_x (E[Y \mid X=x])^2 P[X=x] \\ &= 10^2 \cdot {}^1\!/_6 + 15^2 \cdot {}^2\!/_6 + 20^2 \cdot {}^3\!/_6 \\ &= \frac{875}{3} = E[Y^2] \end{aligned}$$

- Calculemos ahora $E[E[Y \mid X]]$.

$$E[E[Y \mid X]] = E[Y]$$

Por tanto, usando lo anterior, tenemos:

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

2. Considerar las distribución conjunta (X, Y) con función de masa de probabilidad dada por:

$X \setminus Y$	10	15	20	25
1	0	$^3/_7$	0	$^1/_7$
2	0	0	$^1/_7$	0
3	$^2/_7$	0	0	0

Ejercicio 1.1.3. Sea X el número de balanzas e Y el número de dependientes en los puntos de venta de un mercado. Determinar las rectas de regresión y el grado de ajuste a la distribución, si la función masa de probabilidad de (X, Y) viene dada por:

$X \setminus Y$	1	2	3	4
1	$1/24$	$2/24$	0	0
2	$1/24$	$2/24$	$3/24$	$1/24$
3	0	$1/24$	$2/24$	$6/24$
4	0	0	$2/24$	$3/24$

Ejercicio 1.1.4. Sea (X, Y) un vector aleatorio con valores en $\{(x, y) \in \mathbb{R}^2/0 < x < y < 2\}$ y función de densidad constante. Calcular:

1. Curvas y rectas de regresión de X sobre Y y de Y sobre X.
2. Razones de correlación y coeficiente de correlación lineal.
3. Error cuadrático medio asociado a cada una de las funciones de regresión.

Ejercicio 1.1.5. Dada la función masa de probabilidad del vector aleatorio (X, Y)

$X \setminus Y$	0	1	2	3
0	0,2	0,2	0,05	0
1	0,1	0,1	0,1	0,05
2	0	0,05	0,05	0,1

1. Determinar la aproximación lineal mínimo cuadrática de Y para $X = 1$.
2. Determinar la aproximación mínimo cuadrática de Y para $X = 1$.

Ejercicio 1.1.6. Dadas las siguientes distribuciones, determinar qué variable, X ó X', aproxima mejor a la variable Y:

$X \setminus Y$	0	1	2
0	$1/5$	0	0
2	0	$1/5$	0
3	$1/5$	0	$1/5$
4	0	0	$1/5$

$X' \setminus Y$	0	1	2
0	$1/5$	0	$1/5$
2	0	$1/5$	0
3	$1/5$	0	0
4	0	0	$1/5$

Ejercicio 1.1.7. Probar que las variables $X = U + V$ e $Y = U - V$ son incorreladas, pero no independientes, si U y V son variables aleatorias con función de densidad conjunta:

$$f_{U,V}(u, v) = \exp(-u - v), \quad u, v > 0.$$

Ejercicio 1.1.8. Sea X una variable aleatoria con distribución uniforme en el intervalo $[0, 1]$, y sea Y una variable aleatoria continua tal que

$$f_{Y|X=x}(y) = \begin{cases} 1/x^2 & y \in [0, x^2] \\ 0 & \text{en caso contrario} \end{cases}$$

1. Calcular la función de densidad de probabilidad conjunta de X e Y. Calcular la función de densidad de probabilidad marginal de Y.

2. Calcular $E[X \mid Y = y]$ y $E[Y \mid X = x]$.

3. Para la misma densidad de probabilidad condicionada del apartado 1, considerando ahora que X es una variable aleatoria continua con función de densidad de probabilidad:

$$f_X(x) = \begin{cases} 3x^2 & x \in [0,1] \\ 0 & \text{en caso contrario} \end{cases}$$

Calcular de nuevo la función de densidad de probabilidad conjunta de X e Y, y la función de densidad de probabilidad marginal de Y, así como $E[X \mid Y = y]$ y $E[Y \mid X = x]$.

Ejercicio 1.1.9. Sean X e Y variables aleatorias con función de densidad conjunta:

$$f_{(X,Y)}(x,y) = \begin{cases} x + y & (x,y) \in [0,1] \times [0,1] \\ 0 & \text{en caso contrario} \end{cases}$$

1. Calcular la predicción mínimo cuadrática de Y a partir de X y el error cuadrático medio asociado.

2. Si se observa $X = 1/2$, qué predicción de Y tiene menor error cuadrático medio? Calcular dicho error.

3. Supóngase que una persona debe pagar una cantidad C por la oportunidad de observar el valor de X antes de predecir el valor de Y, o puede simplemente predecir el valor de Y sin observar X. Si la persona considera que su pérdida total es la suma de C y el error cuadrático medio de su predicción, qué valor máximo de C estaría dispuesta a pagar?

Ejercicio 1.1.10. Sea (X,Y) un vector aleatorio con función de densidad:

$$f_{(X,Y)}(x,y) = \exp(-y), \quad 0 < x < y$$

Obtener y representar las rectas y curvas de regresión. Calcular el coeficiente de correlación lineal, las razones de correlación y el error cuadrático medio cometido al predecir cada variable según cada una de las funciones de regresión. Interpretar los resultados.

Ejercicio 1.1.11. Sea (X,Y) un vector aleatorio con función de densidad uniforme sobre el cuadrado unidad. Obtener y representar las rectas y curvas de regresión. Calcular el coeficiente de correlación lineal, las razones de correlación y el error cuadrático medio cometido al predecir cada variable según cada una de las funciones de regresión. Interpretar los resultados.

Ejercicio 1.1.12. Supongamos que (X,Y) tiene función de densidad de probabilidad conjunta dada por:

$$f_{(X,Y)}(x,y) = \begin{cases} 1, & |y| < x, x \in]0,1[\\ 0, & \text{en otro caso} \end{cases}$$

Obtener y representar las rectas y curvas de regresión. Calcular el coeficiente de correlación lineal, las razones de correlación y el error cuadrático medio cometido al predecir cada variable según cada una de las funciones de regresión. Interpretar los resultados.

Ejercicio 1.1.13. Sea (X, Y) un vector aleatorio distribuido uniformemente en el paralelogramo de vértices $(0, 0)$; $(2, 0)$; $(3, 1)$ y $(1, 1)$. Calcular el error cuadrático medio asociado a la predicción de X a partir de la variable Y y a la predicción de Y a partir de la variable aleatoria X. Determinar la predicción más fiable a la vista de los resultados obtenidos.

Ejercicio 1.1.14. Sea (X, Y) un vector aleatorio con rectas de regresión

$$x + 4y = 1 \qquad x + 5y = 2$$

1. Cuál es la recta de regresión de Y sobre X?
2. Calcular el coeficiente de correlación lineal y la proporción de varianza de cada variable que queda explicada por la regresión lineal.
3. Calcular las medias de ambas variables