

Probabilidad

FACULTAD
DE
CIENCIAS
UNIVERSIDAD DE GRANADA



Los Del DGIIM, losdeldgiim.github.io

Doble Grado en Ingeniería Informática y Matemáticas
Universidad de Granada



Esta obra está bajo una Licencia Creative Commons Atribución-NoComercial-SinDerivadas 4.0 Internacional (CC BY-NC-ND 4.0).

Eres libre de compartir y redistribuir el contenido de esta obra en cualquier medio o formato, siempre y cuando des el crédito adecuado a los autores originales y no persigas fines comerciales.

Probabilidad

Los Del DGIIM, losdeldgiim.github.io

Arturo Olivares Martos

Granada, 2024-2025

Índice general

1. Relaciones de problemas	5
1.1. Esperanza Condicionada	6

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 la 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:

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

Calculemos ahora las curvas de regresión y las razones de correlación.

- Curva de regresión de Y sobre X :

$$\hat{Y}(x) = E[Y \mid X = x] = \sum_y y P[Y = y \mid X = x] \quad \forall x \in E_x$$

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

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

Para calcular la razón de correlación de Y sobre X , hay dos opciones:

Opción 1. Método rutinario.

Usamos la fórmula:

$$\eta_{Y/X}^2 = \frac{\text{Var}(E[Y \mid X])}{\text{Var}(Y)}$$

- Calculemos $E[Y]$:

$$E[Y] = \sum_y y P[Y = y] = 10 \cdot 1/6 + 15 \cdot 2/6 + 20 \cdot 3/6 = \frac{50}{3}$$

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

$$E[Y^2] = \sum_y y^2 P[Y = y] = 10^2 \cdot 1/6 + 15^2 \cdot 2/6 + 20^2 \cdot 3/6 = \frac{875}{3}$$

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

$$\begin{aligned} E[(E[Y | X])^2] &= \sum_x (E[Y | X = x])^2 P[X = x] = \\ &= 10^2 \cdot 1/6 + 15^2 \cdot 2/6 + 20^2 \cdot 3/6 = \frac{875}{3} = E[Y^2] \end{aligned}$$

- Calculemos ahora $E[E[Y | X]]$.

$$E[E[Y | X]] = E[Y]$$

Por tanto, usando lo anterior, tenemos:

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

Opción 2. Razonando por dependencia funcional.

En este caso, vemos que Y es función de X . Por tanto:

$$\eta_{Y/X}^2 = 1$$

- Curva de regresión de X sobre Y :

$$\hat{X}(y) = E[X | Y = y] = \sum_x x P[X = x | Y = y] \quad \forall y \in E_y$$

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

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

De nuevo, razonando ahora por dependencia funcional, tenemos que:

$$\eta_{X/Y}^2 = 1$$

Como vemos, en este caso, hay dependencia recíproca entre X e Y . Por tanto, el ajuste es el ideal, ya que $Y = f(X)$ y $X = g(Y)$. Cada una explica la totalidad de la variabilidad de la otra.

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

$X \setminus Y$	10	15	20	25	
1	0	$3/7$	0	$1/7$	$4/7$
2	0	0	$1/7$	0	$1/7$
3	$2/7$	0	0	0	$2/7$
	$2/7$	$3/7$	$1/7$	$1/7$	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	25
1	0	$\frac{3}{4}$	0	$\frac{1}{4}$
2	0	0	1	0
3	1	0	0	0

La distribución condicionada de X dado Y , $P[X = x \mid Y = y]$ viene dada por la siguiente tabla:

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

Calculemos ahora las curvas de regresión y las razones de correlación.

- Curva de regresión de Y sobre X :

$$\hat{Y}(x) = E[Y \mid X = x] = \sum_y y P[Y = y \mid X = x] \quad \forall x \in E_x$$

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

$$\begin{aligned}\hat{Y}(1) &= 15 \cdot \frac{3}{4} + 25 \cdot \frac{1}{4} = 17,5 \\ \hat{Y}(2) &= 20 \\ \hat{Y}(3) &= 10\end{aligned}$$

Para calcular la razón de correlación de Y sobre X , tenemos que:

$$\eta_{Y/X}^2 = \frac{\text{Var}(E[Y \mid X])}{\text{Var}(Y)}$$

- Calculemos $E[Y]$:

$$E[Y] = \sum_y y P[Y = y] = 10 \cdot \frac{2}{7} + 15 \cdot \frac{3}{7} + 20 \cdot \frac{1}{7} + 25 \cdot \frac{1}{7} = \frac{110}{7}$$

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

$$\begin{aligned}E[Y^2] &= \sum_y y^2 P[Y = y] \\ &= 10^2 \cdot \frac{2}{7} + 15^2 \cdot \frac{3}{7} + 20^2 \cdot \frac{1}{7} + 25^2 \cdot \frac{1}{7} = \frac{1900}{7}\end{aligned}$$

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

$$\begin{aligned}E[(E[Y \mid X])^2] &= \sum_x (E[Y \mid X = x])^2 P[X = x] \\ &= 17,5^2 \cdot \frac{4}{7} + 20^2 \cdot \frac{1}{7} + 10^2 \cdot \frac{2}{7} = \frac{1825}{7}\end{aligned}$$

- Calculemos ahora $E[E[Y | X]]$.

$$E[E[Y | X]] = E[Y]$$

Por tanto, usando lo anterior, tenemos:

$$\begin{aligned}\eta_{Y/X}^2 &= \frac{\text{Var}(E[Y | X])}{\text{Var}(Y)} = \frac{E[(E[Y | X])^2] - E[E[Y | X]]^2}{E[Y^2] - E[Y]^2} \\ &= \frac{E[(E[Y | X])^2] - E[Y]^2}{E[Y^2] - E[Y]^2} \\ &= \frac{\frac{1825}{7} - \left(\frac{110}{7}\right)^2}{\frac{1900}{7} - \left(\frac{110}{7}\right)^2} = \frac{9}{16} \approx 0,5625\end{aligned}$$

Por tanto, tenemos que X explica el 56,25 % de la variabilidad de Y . Tenemos entonces que no es un ajuste ideal.

- Curva de regresión de X sobre Y :

En este caso, tenemos que $X = f(Y)$, por lo que:

$$\hat{X}(10) = 3, \quad \hat{X}(15) = 1, \quad \hat{X}(20) = 2, \quad \hat{X}(25) = 1$$

Por tanto, como X es función de Y , tenemos que:

$$\eta_{X/Y}^2 = 1$$

Tenemos que Y explica la totalidad de la variabilidad de X , por lo que el ajuste es el ideal.

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	$3/24$
2	$1/24$	$2/24$	$3/24$	$1/24$	$7/24$
3	0	$1/24$	$2/24$	$6/24$	$9/24$
4	0	0	$2/24$	$3/24$	$5/24$
	$2/24$	$5/24$	$7/24$	$10/24$	1

Tras calcular las distribuciones marginales, calculamos ahora las esperanzas ne-

cesarias para que el resto de cálculos posteriormente sean directos.

$$\begin{aligned}
 E[X] &= 1 \cdot 3/24 + 2 \cdot 7/24 + 3 \cdot 9/24 + 4 \cdot 5/24 = 8/3 \\
 E[Y] &= 1 \cdot 2/24 + 2 \cdot 5/24 + 3 \cdot 7/24 + 4 \cdot 10/24 = 73/24 \\
 E[X^2] &= 1^2 \cdot 3/24 + 2^2 \cdot 7/24 + 3^2 \cdot 9/24 + 4^2 \cdot 5/24 = 8 \\
 E[Y^2] &= 1^2 \cdot 2/24 + 2^2 \cdot 5/24 + 3^2 \cdot 7/24 + 4^2 \cdot 10/24 = 245/24 \\
 E[XY] &= \sum_{\substack{x \in E_X \\ y \in E_Y}} xy P[X = x, Y = y] = 209/24 \\
 \text{Var}[X] &= E[X^2] - E[X]^2 = 8/9 \\
 \text{Var}[Y] &= E[Y^2] - E[Y]^2 = 551/576 \\
 \text{Cov}[X, Y] &= E[XY] - E[X]E[Y] = 43/72
 \end{aligned}$$

Calculamos ahora las rectas de regresión.

- Recta de regresión de Y sobre X :

$$\begin{aligned}
 \hat{Y}(x) &= E[Y] + \frac{\text{Cov}[X, Y]}{\text{Var}[X]}(X - E[X]) = \frac{73}{24} + \frac{43/72}{8/9}(X - 8/3) = \\
 &= \frac{73}{24} + \frac{43}{64}(X - 8/3)
 \end{aligned}$$

- Recta de regresión de X sobre Y :

$$\begin{aligned}
 \hat{X}(y) &= E[X] + \frac{\text{Cov}[X, Y]}{\text{Var}[Y]}(Y - E[Y]) = 8/3 + \frac{43/72}{551/576}(Y - 73/24) = \\
 &= 8/3 + \frac{344}{551}(Y - 73/24)
 \end{aligned}$$

Para determinar el grado de ajuste a la distribución, calculamos ahora el coeficiente de determinación lineal:

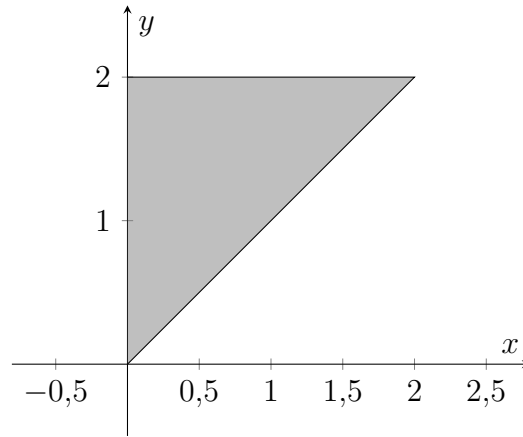
$$\rho_{Y/X}^2 = \frac{\text{Cov}[X, Y]^2}{\text{Var}[X] \text{Var}[Y]} = \frac{1849}{4408} \approx 0,419$$

Por tanto, tenemos que el 41,9% de la variabilidad de Y queda explicada por la regresión lineal de Y sobre X . Vemos por tanto que el ajuste no es bueno.

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

1. Función de densidad de probabilidad conjunta.

Veamos en primer lugar el conjunto en el que se distribuye el vector aleatorio (X, Y) :



Como la función de densidad es constante, tenemos que:

$$f_{(X,Y)}(x,y) = \begin{cases} k, & (x,y) \in T \\ 0, & \text{en otro caso} \end{cases}$$

Para calcular k , tenemos que:

$$1 = \int_T f_{(X,Y)} = \int_T k = k\lambda(T) = k \cdot 2 \implies k = 1/2$$

2. Curvas y rectas de regresión de X sobre Y y de Y sobre X .

Comenzamos calculando las curvas de regresión, ya que si estas son funciones lineales, las rectas de regresión coincidirán con las curvas de regresión. Para ello, calculamos las funciones de densidad marginales y condicionadas.

- Función de densidad de X . Para $x \in]0, 2[$:

$$f_X(x) = \int_{-\infty}^{\infty} f_{(X,Y)}(x,y) dy = \int_x^2 1/2 dy = 1/2(2-x) = 1 - \frac{x}{2}$$

- Función de densidad de Y . Para $y \in]0, 2[$:

$$f_Y(y) = \int_{-\infty}^{\infty} f_{(X,Y)}(x,y) dx = \int_0^y 1/2 dx = \frac{y}{2}$$

- Función de densidad condicionada de X dado $Y = y^* \in]0, 2[$. Para $x \in]0, y^*[$:

$$f_{X|Y=y^*}(x) = \frac{f_{(X,Y)}(x,y^*)}{f_Y(y^*)} = \frac{1/2}{y^*/2} = \frac{1}{y^*}$$

- Función de densidad condicionada de Y dado $X = x^* \in]0, 2[$. Para $y \in]x^*, 2[$:

$$f_{Y|X=x^*}(y) = \frac{f_{(X,Y)}(x^*,y)}{f_X(x^*)} = \frac{1/2}{1-x^*/2} = \frac{1}{2-x^*}$$

Calculamos ahora las curvas de regresión.

- Curva de regresión de X sobre Y :

$$\begin{aligned}\widehat{X}(y) &= E[X \mid Y = y] = \int_0^y x f_{X|Y=y}(x) dx = \int_0^y x \cdot \frac{1}{y} dx = \frac{1}{y} \int_0^y x dx = \\ &= \frac{1}{y} \left[\frac{x^2}{2} \right]_0^y = \frac{y}{2} \quad \forall y \in]0, 2[\end{aligned}$$

- Curva de regresión de Y sobre X :

$$\begin{aligned}\widehat{Y}(x) &= E[Y \mid X = x] = \int_x^2 y f_{Y|X=x}(y) dy = \int_x^2 y \cdot \frac{1}{2-x} dy = \frac{1}{2-x} \int_x^2 y dy = \\ &= \frac{1}{2-x} \left[\frac{y^2}{2} \right]_x^2 = \frac{2^2 - x^2}{2(2-x)} = \frac{4 - x^2}{2(2-x)} = \frac{2+x}{2} = 1 + \frac{x}{2} \quad \forall x \in]0, 2[\end{aligned}$$

Tenemos además que las curvas de regresión son funciones lineales, por lo que las rectas de regresión coinciden con las curvas de regresión.

3. Razones de correlación y coeficiente de correlación lineal.

Como las curvas de regresión son funciones lineales, tenemos que las razones de correlación coinciden y son iguales al coeficiente de determinación lineal.

$$\eta_{Y/X}^2 = \eta_{X/Y}^2 = \rho_{Y/X}^2 = \frac{\text{Cov}[X, Y]^2}{\text{Var}[X] \text{Var}[Y]}$$

Como este es el producto de las pendientes de las rectas de regresión, tenemos que:

$$\rho_{Y/X}^2 = \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{4}$$

Como la correlación es positiva por ser ambas rectas crecientes, tenemos que:

$$\rho_{Y/X} = +\sqrt{\rho_{Y/X}^2} = 1/2$$

4. Error cuadrático medio asociado a cada una de las funciones de regresión.

Por lo visto en teoría, sabemos que:

$$\begin{aligned}\text{ECM}(\widehat{X}) &= \text{Var}[X] - \frac{\text{Cov}[X, Y]^2}{\text{Var}[Y]} \\ \text{ECM}(\widehat{Y}) &= \text{Var}[Y] - \frac{\text{Cov}[X, Y]^2}{\text{Var}[X]} \end{aligned}$$

Calculamos por tanto los valores necesarios:

$$E[X] = \int_0^2 x f_X(x) dx = \int_0^2 x(1 - x/2) dx = \left[\frac{x^2}{2} - \frac{x^3}{6} \right]_0^2 = \frac{4}{2} - \frac{8}{6} = \frac{2}{3}$$

$$E[Y] = \int_0^2 y f_Y(y) dy = \int_0^2 y \cdot y/2 dy = \left[\frac{y^3}{6} \right]_0^2 = \frac{8}{6} = \frac{4}{3}$$

$$E[X^2] = \int_0^2 x^2 f_X(x) dx = \int_0^2 x^2(1 - x/2) dx = \left[\frac{x^3}{3} - \frac{x^4}{8} \right]_0^2 = \frac{8}{3} - \frac{16}{8} = \frac{2}{3}$$

$$E[Y^2] = \int_0^2 y^2 f_Y(y) dy = \int_0^2 y^2 \cdot y/2 dy = \left[\frac{y^4}{8} \right]_0^2 = \frac{16}{8} = 2$$

$$\begin{aligned} E[XY] &= \int_0^2 \int_x^2 xy f_{(X,Y)}(x, y) dy dx = \int_0^2 \int_x^2 xy \cdot 1/2 dy dx = \frac{1}{2} \int_0^2 x \left[\frac{y^2}{2} \right]_x^2 dx = \\ &= \frac{1}{4} \int_0^2 x (4 - x^2) dx = \frac{1}{4} \left[2x^2 - \frac{x^4}{4} \right]_0^2 = \frac{1}{4} (8 - 4) = 1 \end{aligned}$$

$$\text{Var}[X] = E[X^2] - E[X]^2 = \frac{2}{3} - \left(\frac{2}{3} \right)^2 = \frac{2}{9}$$

$$\text{Var}[Y] = E[Y^2] - E[Y]^2 = 2 - \left(\frac{4}{3} \right)^2 = \frac{2}{9}$$

$$\text{Cov}[X, Y] = E[XY] - E[X]E[Y] = 1 - \frac{2}{3} \cdot \frac{4}{3} = \frac{1}{9}$$

Por tanto, tenemos que:

$$\text{ECM}(\hat{X}) = \text{Var}[X] - \frac{\text{Cov}[X, Y]^2}{\text{Var}[Y]} = \frac{2}{9} - \frac{1/9^2}{2/9} = \frac{2}{9} - \frac{1}{18} = \frac{3}{18} = \frac{1}{6}$$

$$\text{ECM}(\hat{Y}) = \text{Var}[Y] - \frac{\text{Cov}[X, Y]^2}{\text{Var}[X]} = \frac{2}{9} - \frac{1/9^2}{2/9} = \frac{2}{9} - \frac{1}{18} = \frac{3}{18} = \frac{1}{6}$$

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	0,45
1	0,1	0,1	0,1	0,05	0,35
2	0	0,05	0,05	0,1	0,2
	0,3	0,35	0,2	0,15	1

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

Nos piden la recta de regresión de Y sobre X , y evaluarla en el punto $X = 1$.

Calculamos las esperanzas necesarias para el cálculo de la recta de regresión.

$$E[X] = 0 \cdot 0,3 + 1 \cdot 0,35 + 2 \cdot 0,2 = 0,75$$

$$E[Y] = 0 \cdot 0,3 + 1 \cdot 0,35 + 2 \cdot 0,2 + 3 \cdot 0,15 = 1,2$$

$$E[X^2] = 0^2 \cdot 0,3 + 1^2 \cdot 0,35 + 2^2 \cdot 0,2 = 1,15$$

$$E[XY] = \sum_{\substack{x \in E_X \\ y \in E_Y}} xyP[X = x, Y = y] = 1,35$$

$$\text{Var}[X] = E[X^2] - E[X]^2 = 1,15 - 0,75^2 = 0,5875$$

$$\text{Cov}[X, Y] = E[XY] - E[X]E[Y] = 1,35 - 0,75 \cdot 1,2 = 0,45$$

Calculamos ahora la recta de regresión de Y sobre X .

$$\hat{Y}(x) = E[Y] + \frac{\text{Cov}[X, Y]}{\text{Var}[X]}(x - E[X]) = 1,2 + \frac{0,45}{0,5875}(x - 0,75) = 1,2 + \frac{36}{47}(x - 0,75)$$

Evaluando en $X = 1$, tenemos que:

$$\hat{Y}(1) = 1,2 + \frac{36}{47}(1 - 0,75) = 1,2 + \frac{36}{47} \cdot 0,25 = 1,2 + \frac{9}{47} \approx 1,391489$$

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

En este caso, piden $E[Y | X = 1]$. Para ello, hemos de calcular las distribución condicionada de Y dado $X = 1$.

$$P[Y = 0 | X = 1] = \frac{P[X = 1, Y = 0]}{P[X = 1]} = \frac{0,1}{0,35} = \frac{2}{7}$$

$$P[Y = 1 | X = 1] = \frac{P[X = 1, Y = 1]}{P[X = 1]} = \frac{0,1}{0,35} = \frac{2}{7}$$

$$P[Y = 2 | X = 1] = \frac{P[X = 1, Y = 2]}{P[X = 1]} = \frac{0,1}{0,35} = \frac{2}{7}$$

$$P[Y = 3 | X = 1] = \frac{P[X = 1, Y = 3]}{P[X = 1]} = \frac{0,05}{0,35} = \frac{1}{7}$$

Por tanto, tenemos que:

$$\begin{aligned} E[Y | X = 1] &= \sum_{y \in E_Y} yP[Y = y | X = 1] = \\ &= 0 + 1 \cdot P[Y = 1 | X = 1] + 2 \cdot P[Y = 2 | X = 1] + 3 \cdot P[Y = 3 | X = 1] = \\ &= \frac{2}{7} + 2 \cdot \frac{2}{7} + 3 \cdot \frac{1}{7} = \frac{9}{7} \end{aligned}$$

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	$\frac{1}{5}$	0	0	$\frac{1}{5}$
2	0	$\frac{1}{5}$	0	$\frac{1}{5}$
3	$\frac{1}{5}$	0	$\frac{1}{5}$	$\frac{2}{5}$
4	0	0	$\frac{1}{5}$	$\frac{1}{5}$
	$\frac{2}{5}$	$\frac{1}{5}$	$\frac{2}{5}$	1

$X' \setminus Y$	0	1	2	
0	$\frac{1}{5}$	0	$\frac{1}{5}$	$\frac{2}{5}$
2	0	$\frac{1}{5}$	0	$\frac{1}{5}$
3	$\frac{1}{5}$	0	0	$\frac{1}{5}$
4	0	0	$\frac{1}{5}$	$\frac{1}{5}$
	$\frac{2}{5}$	$\frac{1}{5}$	$\frac{2}{5}$	1

Hemos de obtener $\eta_{Y/X}^2$ y $\eta_{Y/X'}^2$ para compararlas.

$$\eta_{Y/X}^2 = \frac{\text{Var}[E[Y | X]]}{\text{Var}[Y]}, \quad \eta_{Y/X'}^2 = \frac{\text{Var}[E[Y | X']]}{\text{Var}[Y]}$$

Calculamos las esperanzas necesarias para el cálculo de las varianzas.

$$\begin{aligned} E[Y] &= 0 \cdot 2/5 + 1 \cdot 1/5 + 2 \cdot 2/5 = 1 \\ E[Y^2] &= 0^2 \cdot 2/5 + 1^2 \cdot 1/5 + 2^2 \cdot 2/5 = 9/5 \\ E[E[Y | X]] &= E[Y] = 1 \\ E[E[Y | X']] &= E[Y] = 1 \\ E[(E[Y | X])^2] &= \sum_{x \in E_X} (E[Y | X = x])^2 P[X = x] \\ E[(E[Y | X'])^2] &= \sum_{x \in E_{X'}} (E[Y | X' = x])^2 P[X' = x] \end{aligned}$$

Calculamos por tanto las esperanzas condicionadas.

$$\begin{aligned} E[Y | X = 0] &= 0 \cdot 1 + 1 \cdot 0 + 2 \cdot 0 = 0 \\ E[Y | X = 2] &= 0 \cdot 0 + 1 \cdot 1 + 2 \cdot 0 = 1 \\ E[Y | X = 3] &= 0 \cdot 1/2 + 1 \cdot 0 + 2 \cdot 1/2 = 1 \\ E[Y | X = 4] &= 0 \cdot 0 + 1 \cdot 0 + 2 \cdot 1 = 2 \\ E[Y | X' = 0] &= 0 \cdot 1/2 + 1 \cdot 0 + 2 \cdot 1/2 = 1 \\ E[Y | X' = 2] &= 0 \cdot 0 + 1 \cdot 1 + 2 \cdot 0 = 1 \\ E[Y | X' = 3] &= 0 \cdot 1/2 + 1 \cdot 0 + 2 \cdot 1/2 = 1 \\ E[Y | X' = 4] &= 0 \cdot 0 + 1 \cdot 0 + 2 \cdot 1 = 2 \end{aligned}$$

Por tanto, hemos comprobado que $E[Y | X] = E[Y | X']$, por lo que:

$$E[Y | X] = E[Y | X'] \implies \text{Var}[E[Y | X]] = \text{Var}[E[Y | X']] \implies \eta_{Y/X}^2 = \eta_{Y/X'}^2$$

Por tanto, ambas variables aproximan igual de bien a Y . Veamos cómo es la aproximación calculando aun así los coeficientes. Tenemos que:

$$\begin{aligned} E[(E[Y | X])^2] &= \sum_{x \in E_X} (E[Y | X = x])^2 P[X = x] = \\ &= 0^2 \cdot 1/5 + 1^2 \cdot 1/5 + 1^2 \cdot 2/5 + 2^2 \cdot 1/5 = 7/5 \end{aligned}$$

Por tanto, tenemos que:

$$\begin{aligned} \text{Var}[Y] &= E[Y^2] - E[Y]^2 = 9/5 - 1 = 4/5 \\ \text{Var}[E[Y | X]] &= E[(E[Y | X])^2] - E[E[Y | X]]^2 = 7/5 - 1 = 2/5 \end{aligned}$$

Por tanto, tenemos que:

$$\eta_{Y/X}^2 = \frac{\text{Var}[E[Y | X]]}{\text{Var}[Y]} = \frac{2/5}{4/5} = 1/2 = \eta_{Y/X'}^2$$

Por tanto, tenemos que ninguno de los ajustes es bueno, ya que solo explican la mitad de la variabilidad de Y .

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:

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

Veamos que no son independientes. Para ello, calculamos en primer lugar $f_{X,Y}(x, y)$. Para ello, definimos la transformación:

$$g : \begin{array}{ccc} \mathbb{R}^2 & \longrightarrow & \mathbb{R}^2 \\ (U, V) & \longmapsto & (X, Y) = (U + V, U - V) \end{array}$$

Para obtener g^{-1} , buscamos obtener U, V en función de X, Y :

$$\begin{cases} X = U + V \\ Y = U - V \end{cases} \implies \begin{cases} U = \frac{X+Y}{2} \\ V = \frac{X-Y}{2} \end{cases}$$

Por tanto, tenemos que $\exists g^{-1}$, con:

$$g^{-1} : \begin{array}{ccc} \mathbb{R}^2 & \longrightarrow & \mathbb{R}^2 \\ (X, Y) & \longmapsto & (U, V) = \left(\frac{X+Y}{2}, \frac{X-Y}{2}\right) \end{array}$$

Tenemos que todas las componentes de g^{-1} son derivables:

$$\frac{\partial U}{\partial X} = \frac{1}{2}, \quad \frac{\partial U}{\partial Y} = \frac{1}{2}, \quad \frac{\partial V}{\partial X} = \frac{1}{2}, \quad \frac{\partial V}{\partial Y} = -\frac{1}{2}$$

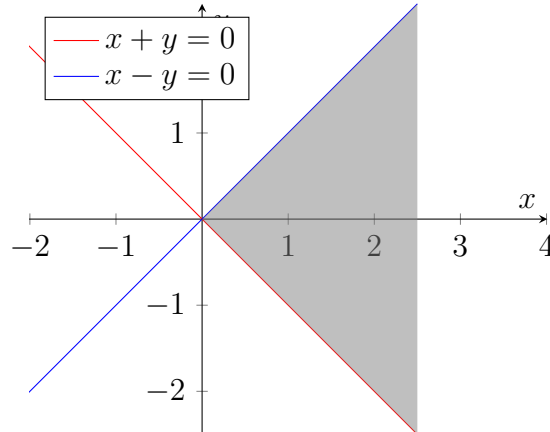
Además, veamos que el jacobiano no se anula:

$$\det Jg^{-1}(x, y) = \begin{vmatrix} 1/2 & 1/2 \\ 1/2 & -1/2 \end{vmatrix} = \frac{1}{2^2} \cdot \begin{vmatrix} 1 & 1 \\ 1 & -1 \end{vmatrix} = -\frac{1}{2} \neq 0 \quad \forall (x, y) \in \mathbb{R}^2.$$

Por tanto, podemos aplicar el teorema del cambio de variable para obtener $f_{X,Y}(x, y)$:

$$\begin{aligned} f_{X,Y}(x, y) &= f_{U,V}(g^{-1}(x, y)) \cdot |Jg^{-1}(x, y)| = \frac{1}{2} \cdot f_{U,V}\left(\frac{x+y}{2}, \frac{x-y}{2}\right) \\ &= \frac{1}{2} \cdot \exp\left(-\frac{x+y}{2} - \frac{x-y}{2}\right) = \frac{1}{2} \cdot \exp\left(\frac{-x-y-x+y}{2}\right) = \\ &= \frac{1}{2} \cdot \exp(-x) \quad \forall (x, y) \in \mathbb{R}^2 \text{ tal que } x+y, x-y > 0 \end{aligned}$$

Veamos el conjunto gráficamente:



Tenemos que el conjunto donde $f_{X,Y}(x, y)$ no se puede expresar como producto cartesiano, por lo que intuimos que X e Y no son independientes. Para comprobarlo, calculamos la función de densidad marginal de X y Y :

$$\begin{aligned} f_X(x) &= \int_{-\infty}^{\infty} f_{X,Y}(x, y) dy = \int_{-x}^x \frac{1}{2} \exp(-x) dy = \frac{1}{2} \exp(-x) [y]_{-x}^x = \\ &= \frac{1}{2} \exp(-x)(x + x) = x e^{-x} \quad \forall x > 0 \end{aligned}$$

Para calcular la marginal de Y , distinguimos en función de y :

■ Si $y > 0$:

$$\begin{aligned} f_Y(y) &= \int_{-\infty}^{\infty} f_{X,Y}(x, y) dx = \int_y^{\infty} \frac{1}{2} \exp(-x) dx = \frac{1}{2} [-\exp(-x)]_y^{\infty} = \\ &= \frac{1}{2} [0 - (-\exp(-y))] = \frac{e^{-y}}{2} \quad \forall y > 0 \end{aligned}$$

■ Si $y \leq 0$:

$$\begin{aligned} f_Y(y) &= \int_{-\infty}^{\infty} f_{X,Y}(x, y) dx = \int_{-y}^{\infty} \frac{1}{2} \exp(-x) dx = \frac{1}{2} [-\exp(-x)]_{-y}^{\infty} = \\ &= \frac{1}{2} [0 - (-\exp(y))] = \frac{e^y}{2} \quad \forall y \leq 0 \end{aligned}$$

Por tanto, tenemos que:

$$f_Y(y) = \begin{cases} \frac{e^{-y}}{2} & y > 0 \\ \frac{e^y}{2} & y \leq 0 \end{cases}$$

Para el $(1, 1)$, tenemos que:

$$\begin{aligned} f_X(1) &= e^{-1}, & f_{X,Y}(1, 1) &= \frac{1}{2} e^{-1} \\ f_Y(1) &= \frac{1}{2} e^{-1} \end{aligned}$$

Por tanto, no se cumple que $f_{X,Y}(x, y) = f_X(x)f_Y(y)$ para todo $(x, y) \in \mathbb{R}^2$, por lo que X e Y no son independientes.

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 $E[X | Y = y]$ y $E[Y | X = x]$.
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 $E[X | Y = y]$ y $E[Y | X = x]$.

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

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

- 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