

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:

$$\hat{Y}(1) = 15 \cdot \frac{3}{4} + 25 \cdot \frac{1}{4} = 17,5$$

$$\hat{Y}(2) = 20$$

$$\hat{Y}(3) = 10$$

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:

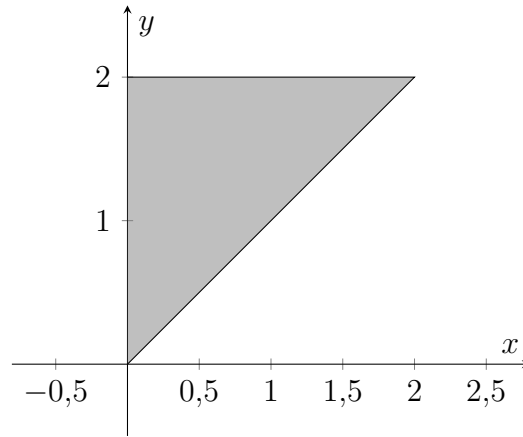
$$\rho_{Y/X}^2 = \frac{\text{Cov}[X, Y]^2}{\text{Var}[X] \text{Var}[Y]} = \frac{1849}{4408} \approx 0,419$$

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:

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

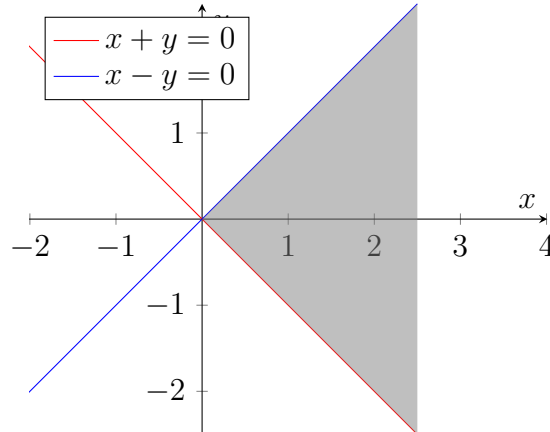
Además, veamos que el jacobiano no se anula:

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

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:

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

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 .

Para calcular $f_{X,Y}(x, y)$, usamos el siguiente resultado:

$$f_{Y|X=x}(y) = \frac{f_{X,Y}(x, y)}{f_X(x)} \implies f_{X,Y}(x, y) = f_{Y|X=x}(y) \cdot f_X(x)$$

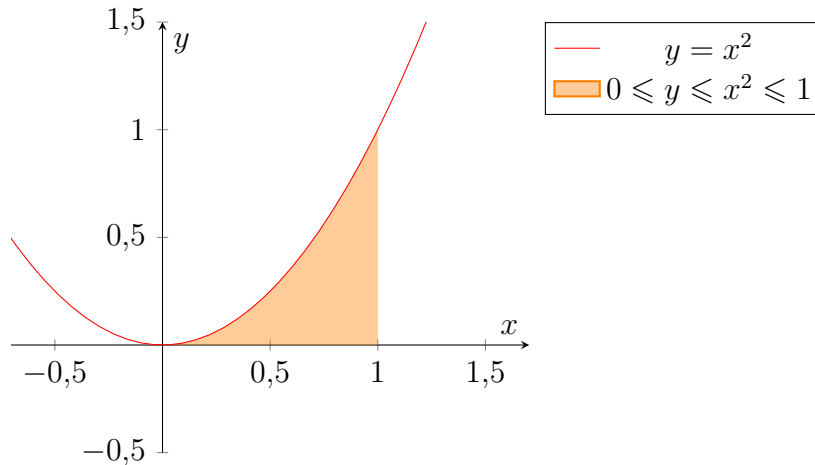
Por tanto, calculamos la marginal de X . Como esta es uniforme en $[0, 1]$, tenemos que:

$$1 = \int_0^1 f_X(x) dx = \int_0^1 k dx = k \implies f_X(x) = 1 \quad \forall x \in [0, 1]$$

Por tanto, tenemos que:

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

Veamos gráficamente el conjunto donde $f_{X,Y}(x, y)$ no se anula:



Para calcular la marginal de Y , para cada $y \in [0, 1]$, tenemos que:

$$f_Y(y) = \int_{\sqrt{y}}^1 f_{X,Y}(x, y) dx = \int_{\sqrt{y}}^1 1/x^2 dx = \left[-\frac{1}{x} \right]_{\sqrt{y}}^1 = -1 + \frac{1}{\sqrt{y}}$$

Por tanto, tenemos que:

$$f_Y(y) = \begin{cases} -1 + 1/\sqrt{y} & 0 \leq y \leq 1 \\ 0 & \text{en caso contrario} \end{cases}$$

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

Calculamos en primer lugar la distribución condicionada de X dado un valor $Y = y \in [0, 1]$.

$$f_{X|Y=y}(x) = \frac{f_{X,Y}(x, y)}{f_Y(y)} = \frac{1/x^2}{-1 + 1/\sqrt{y}} \quad \forall x \in [0, 1] \mid \sqrt{y} \leq x$$

Calculamos ahora por tanto las esperanzas condicionadas:

$$\begin{aligned}
 E[X \mid Y = y] &= \int_{\sqrt{y}}^1 x \cdot f_{X|Y=y}(x) \, dx = \int_{\sqrt{y}}^1 x \cdot \frac{1/x^2}{-1 + 1/\sqrt{y}} \, dx = \frac{1}{-1 + 1/\sqrt{y}} \int_{\sqrt{y}}^1 1/x \, dx = \\
 &= \frac{1}{-1 + 1/\sqrt{y}} [\ln(x)]_{\sqrt{y}}^1 = \frac{1}{-1 + 1/\sqrt{y}} [\ln(1) - \ln(\sqrt{y})] = \\
 &= -\frac{\ln(\sqrt{y})}{-1 + 1/\sqrt{y}} \quad \forall y \in [0, 1] \\
 E[Y \mid X = x] &= \int_0^{x^2} y \cdot f_{Y|X=x}(y) \, dy = \int_0^{x^2} y \cdot \frac{1}{x^2} \, dy = \frac{1}{x^2} \int_0^{x^2} y \, dy = \\
 &= \frac{1}{x^2} \left[\frac{y^2}{2} \right]_0^{x^2} = \frac{1}{x^2} \cdot \frac{x^4}{2} = \frac{x^2}{2} \quad \forall x \in [0, 1]
 \end{aligned}$$

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

Repetimos todo el proceso anterior, pero con la nueva función de densidad de X .

$$f_{X,Y}(x, y) = f_{Y|X=x}(y) \cdot f_X(x) = \begin{cases} 3 & 0 \leq y \leq x^2 \leq 1 \\ 0 & \text{en caso contrario} \end{cases}$$

Para calcular la marginal de Y , para cada $y \in [0, 1]$, tenemos que:

$$f_Y(y) = \int_{\sqrt{y}}^1 f_{X,Y}(x, y) \, dx = 3 \int_{\sqrt{y}}^1 1 \, dx = 3 [x]_{\sqrt{y}}^1 = 3 - 3\sqrt{y} \quad \forall y \in [0, 1]$$

Calculamos ahora la distribución condicionada de X dado un valor $Y = y \in [0, 1]$.

$$f_{X|Y=y}(x) = \frac{f_{X,Y}(x, y)}{f_Y(y)} = \frac{3}{3 - 3\sqrt{y}} = \frac{1}{1 - \sqrt{y}} \quad \forall x \in [0, 1] \mid \sqrt{y} \leq x$$

Calculamos ahora las esperanzas condicionadas:

$$\begin{aligned}
 E[X | Y = y] &= \int_{\sqrt{y}}^1 x \cdot f_{X|Y=y}(x) dx = \int_{\sqrt{y}}^1 x \cdot \frac{1}{1 - \sqrt{y}} dx = \frac{1}{1 - \sqrt{y}} \int_{\sqrt{y}}^1 x dx = \\
 &= \frac{1}{1 - \sqrt{y}} \left[\frac{x^2}{2} \right]_{\sqrt{y}}^1 = \frac{1}{1 - \sqrt{y}} \left[\frac{1}{2} - \frac{y}{2} \right] = \frac{1}{2(1 - \sqrt{y})} (1 - y) = \\
 &= \frac{1 + \sqrt{y}}{2} \quad \forall y \in [0, 1] \\
 E[Y | X = x] &= \int_0^{x^2} y \cdot f_{Y|X=x}(y) dy = \int_0^{x^2} y \cdot \frac{1}{x^2} dy = \frac{1}{x^2} \int_0^{x^2} y dy = \\
 &= \frac{1}{x^2} \left[\frac{y^2}{2} \right]_0^{x^2} = \frac{1}{x^2} \cdot \frac{x^4}{2} = \frac{x^2}{2} \quad \forall x \in [0, 1]
 \end{aligned}$$

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.

Para calcular la predicción mínimo cuadrática de Y a partir de X , hemos de calcular $E[Y | X]$. Para ello, hemos de calcular la distribución marginal de X y la distribución condicionada de Y dado $X = x \in [0, 1]$. condicionada de Y dado $X = x \in [0, 1]$.

$$\begin{aligned}
 f_X(x) &= \int_0^1 f_{(X,Y)}(x, y) dy = \int_0^1 x + y dy = \left[xy + \frac{y^2}{2} \right]_0^1 = x + \frac{1}{2} \quad \forall x \in [0, 1] \\
 f_{Y|X=x}(y) &= \frac{f_{(X,Y)}(x, y)}{f_X(x)} = \frac{x + y}{x + 1/2} \quad \forall y \in [0, 1]
 \end{aligned}$$

Calculamos ahora la esperanza condicionada de Y dado $X = x$:

$$\begin{aligned}
 E[Y | X = x] &= \int_0^1 y \cdot f_{Y|X=x}(y) dy = \int_0^1 y \cdot \frac{x + y}{x + 1/2} dy = \frac{1}{x + 1/2} \int_0^1 y \cdot (x + y) dy = \\
 &= \frac{1}{x + 1/2} \int_0^1 xy + y^2 dy = \frac{1}{x + 1/2} \left[\frac{xy^2}{2} + \frac{y^3}{3} \right]_0^1 = \frac{1}{x + 1/2} \left(\frac{x}{2} + \frac{1}{3} \right) = \\
 &= \frac{2}{2x + 1} \left(\frac{3x + 2}{6} \right) = \frac{3x + 2}{6x + 3} \quad \forall x \in [0, 1]
 \end{aligned}$$

Calculamos su error cuadrático medio:

$$\text{ECM}(E[Y | X]) = E[\text{Var}[Y | X]]$$

Calculamos por tanto en primer lugar $\text{Var}[Y | X] = E[Y^2 | X] - E[Y | X]^2$:

$$\begin{aligned} E[Y^2 | X = x] &= \int_0^1 y^2 \cdot f_{Y|X=x}(y) dy = \int_0^1 y^2 \cdot \frac{x+y}{x+1/2} dy = \frac{1}{x+1/2} \int_0^1 y^2 \cdot (x+y) dy = \\ &= \frac{1}{x+1/2} \int_0^1 xy^2 + y^3 dy = \frac{1}{x+1/2} \left[\frac{xy^3}{3} + \frac{y^4}{4} \right]_0^1 = \frac{1}{x+1/2} \left(\frac{x}{3} + \frac{1}{4} \right) = \\ &= \frac{4x+3}{12x+6} \quad \forall x \in [0, 1] \end{aligned}$$

$$\begin{aligned} \text{Var}[Y | X = x] &= E[Y^2 | X = x] - E[Y | X = x]^2 = \frac{4x+3}{12x+6} - \left(\frac{3x+2}{6x+3} \right)^2 = \\ &= \frac{4x+3}{6(2x+1)} - \frac{(3x+2)^2}{9(2x+1)^2} = \frac{3(4x+3)(2x+1) - 2(3x+2)^2}{18(2x+1)^2} = \\ &= \frac{3(8x^2 + 10x + 3) - 2(9x^2 + 4 + 12x)}{18(2x+1)^2} = \\ &= \frac{24x^2 + 30x + 9 - 18x^2 - 8 - 24x}{18(2x+1)^2} = \\ &= \frac{6x^2 + 6x + 1}{18(2x+1)^2} \quad \forall x \in [0, 1] \end{aligned}$$

Por tanto, tenemos que:

$$\begin{aligned} \text{ECM}(E[Y | X]) &= E[\text{Var}[Y | X]] = \int_0^1 \frac{6x^2 + 6x + 1}{18(2x+1)^2} \cdot f_X(x) dx = \\ &= \int_0^1 \frac{6x^2 + 6x + 1}{18(2x+1)^2} \cdot (x+1/2) dx = \int_0^1 \frac{6x^3 + 6x^2 + x}{36(2x+1)} dx \end{aligned}$$

Para resolver dicha integral, realizamos la división de polinomios:

$$\begin{array}{r} (6x^3 + 6x^2 + x) : (2x + 1) = 3x^2 + \frac{3}{2}x - \frac{1}{4} + \frac{\frac{1}{4}}{2x+1} \\ \underline{-6x^3 - 3x^2} \\ 3x^2 + x \\ \underline{-3x^2 - \frac{3}{2}x} \\ -\frac{1}{2}x \\ \underline{-\frac{1}{2}x + \frac{1}{4}} \\ \frac{1}{4} \end{array}$$

Por tanto, tenemos que:

$$\begin{aligned} \text{ECM}(E[Y | X]) &= \frac{1}{36} \int_0^1 3x^2 + \frac{3}{2}x - \frac{1}{4} + \frac{1/4}{2x+1} dx = \\ &= \frac{1}{36} \left[x^3 + \frac{3}{4}x^2 - \frac{x}{4} + \frac{1}{8} \ln(2x+1) \right]_0^1 = \frac{1}{36} \left(1 + \frac{3}{4} - \frac{1}{4} + \frac{1}{8} \ln(3) \right) = \\ &= \frac{1}{36} \left(\frac{3}{2} + \frac{1}{8} \ln(3) \right) \approx 0,045 \end{aligned}$$

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

En este caso, tenemos que la predicción de Y a partir de $X = 1/2$ es:

$$E[Y | X = 1/2] = \frac{3 \cdot 1/2 + 2}{6 \cdot 1/2 + 3} = \frac{7}{12}$$

Calculamos su error cuadrático medio. Para esto, hay dos opciones:

Usando el apartado anterior En este caso, tenemos que:

$$\begin{aligned} \text{ECM}(E[Y | X = 1/2]) &= E[\text{Var}[Y | X = 1/2]] = \\ &= E\left[\frac{6 \cdot 1/4 + 6 \cdot 1/2 + 1}{18(2 \cdot 1/2 + 1)^2}\right] = E\left[\frac{11}{144}\right] = \frac{11}{144} \end{aligned}$$

Usando la definición de ECM En este caso, tenemos que:

$$\text{ECM}(E[Y | X = 1/2]) = E[(Y - E[Y | X = 1/2])^2] = E\left[\left(Y - \frac{7}{12}\right)^2\right] = E[Y^2] - \frac{7}{6}E[Y] + \frac{49}{144}$$

Para calcular las esperanzas, calculamos en primer lugar la marginal de Y . Para $y \in [0, 1]$:

$$f_Y(y) = \int_0^1 f_{(X,Y)}(x, y) dx = \int_0^1 x + y dx = \left[\frac{x^2}{2} + xy\right]_0^1 = \frac{1}{2} + y \quad \forall y \in [0, 1]$$

Calculamos por tanto las esperanzas necesarias:

$$\begin{aligned} E[Y] &= \int_0^1 y \cdot f_Y(y) dy = \int_0^1 y \cdot \left(\frac{1}{2} + y\right) dy = \int_0^1 \frac{y}{2} + y^2 dy = \\ &= \left[\frac{y^2}{4} + \frac{y^3}{3}\right]_0^1 = \frac{1}{4} + \frac{1}{3} = \frac{7}{12} \\ E[Y^2] &= \int_0^1 y^2 \cdot f_Y(y) dy = \int_0^1 y^2 \cdot \left(\frac{1}{2} + y\right) dy = \int_0^1 \frac{y^2}{2} + y^3 dy = \\ &= \left[\frac{y^3}{6} + \frac{y^4}{4}\right]_0^1 = \frac{1}{6} + \frac{1}{4} = \frac{5}{12} \end{aligned}$$

Por tanto, tenemos que:

$$\text{ECM}(E[Y | X = 1/2]) = E[Y^2] - \frac{7}{6}E[Y] + \frac{49}{144} = \frac{5}{12} - \frac{7}{6} \cdot \frac{7}{12} + \frac{49}{144} = \frac{11}{144}$$

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?

Estudiemos cada una de las opciones por separado.

Observar X antes de predecir Y : En este caso, la pérdida total es:

$$C + \text{ECM}(E[Y | X]) = C + \frac{1}{36} \left(\frac{3}{2} + \frac{1}{8} \ln(3) \right)$$

Predecir Y sin observar X : En este caso, la pérdida total es (ya que no pagamos C):

$$\text{ECM}(E[Y]) = E[(Y - E[Y])^2] = \text{Var}[Y] = E[Y^2] - E[Y]^2 = \frac{5}{12} - \left(\frac{7}{12} \right)^2 = \frac{11}{144}$$

Por tanto, la persona estaría dispuesta a pagar un máximo de:

$$C = \frac{11}{144} - \frac{1}{36} \left(\frac{3}{2} + \frac{1}{8} \ln(3) \right) \approx 0,03090$$

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

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

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.

Obtenemos en primer lugar las marginales:

$$\begin{aligned} f_X(x) &= \int_x^\infty e^{-y} dy = [-e^{-y}]_x^\infty = e^{-x} \quad \forall x > 0 \\ f_Y(y) &= \int_0^y e^{-y} dx = ye^{-y} \quad \forall y > 0 \end{aligned}$$

Obtenemos ahora las condicionadas. Dados $x^*, y^* \in \mathbb{R}^+$, tenemos que:

$$\begin{aligned} f_{Y|X=x^*}(y) &= \frac{f_{(X,Y)}(x^*, y)}{f_X(x^*)} = \frac{e^{-y}}{e^{-x^*}} = e^{x^*-y} \quad \forall y > x^* \\ f_{X|Y=y^*}(x) &= \frac{f_{(X,Y)}(x, y^*)}{f_Y(y^*)} = \frac{e^{-y^*}}{y^* e^{-y^*}} = \frac{1}{y^*} \quad \forall x \in]0, y^*[\end{aligned}$$

Calculamos ahora las curvas de regresión. Para la de Y sobre X , tenemos que:

$$\hat{Y}(x) = E[Y | X = x] = \int_x^\infty y \cdot e^{x-y} dy = e^x \int_x^\infty y \cdot e^{-y} dy$$

Para calcular la integral, realizamos integración por partes:

$$\begin{aligned} \left[\begin{array}{l} u(y) = y \\ v'(y) = e^{-y} \end{array} \quad \begin{array}{l} u'(y) = 1 \\ v(y) = -e^{-y} \end{array} \right] &\implies \int_x^\infty y \cdot e^{-y} dy = [-ye^{-y}]_x^\infty - \int_x^\infty -e^{-y} dy = \\ &= [-ye^{-y} - e^{-y}]_x^\infty = e^{-x}(x + 1) \end{aligned}$$

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

$$\hat{Y}(x) = e^x e^{-x}(x+1) = x+1 \quad \forall x > 0$$

Como vemos, la curva de regresión de Y sobre X es $y = x+1$, que es una recta y por tanto coincide con la recta de regresión.

Calculamos ahora la curva de regresión de X sobre Y . Para ello, tenemos que:

$$\hat{X}(y) = E[X | Y = y] = \int_0^y x \cdot \frac{1}{y} dx = \frac{1}{y} \int_0^y x dx = \frac{1}{y} \left[\frac{x^2}{2} \right]_0^y = \frac{y}{2} \quad \forall y > 0$$

Como vemos, la curva de regresión de X sobre Y es $x = y/2$, que es una recta y por tanto coincide con la recta de regresión.

Calculamos ahora el coeficiente de determinación lineal. Para ello, calculamos las varianzas y covarianzas:

$$\begin{aligned} E[X] &= \int_0^\infty x \cdot e^{-x} dx = [-xe^{-x} - e^{-x}]_0^\infty = 1 \\ E[Y] &= \int_0^\infty y^2 \cdot e^{-y} dy = [-y^2 e^{-y}]_0^\infty + 2 \int_0^\infty y \cdot e^{-y} dy = 0 + 2 \cdot 1 = 2 \\ E[X^2] &= \int_0^\infty x^2 \cdot e^{-x} dx = 2 \\ E[Y^2] &= \int_0^\infty y^3 \cdot e^{-y} dy = [-y^3 e^{-y}]_0^\infty + 3 \int_0^\infty y^2 \cdot e^{-y} dy = 0 + 3 \cdot 2 = 6 \\ E[XY] &= \int_0^{+\infty} \int_0^y xy f_{(X,Y)}(x,y) dx dy = \int_0^{+\infty} \int_0^y xy e^{-y} dx dy = \\ &= \int_0^{+\infty} ye^{-y} \cdot \frac{y^2}{2} dy = \frac{1}{2} \int_0^{+\infty} y^3 e^{-y} dy = \frac{1}{2} \cdot 6 = 3 \\ \text{Var}[X] &= E[X^2] - E[X]^2 = 2 - 1 = 1 \\ \text{Var}[Y] &= E[Y^2] - E[Y]^2 = 6 - 4 = 2 \\ \text{Cov}[X, Y] &= E[XY] - E[X]E[Y] = 3 - 2 \cdot 1 = 1 \end{aligned}$$

Por tanto, podemos calcular el coeficiente de determinación lineal. Además, como las rectas de regresión coinciden con las curvas de regresión, el coeficiente de determinación lineal coincide con las razones de correlación:

$$\eta_{X/Y}^2 = \eta_{Y/X}^2 = \rho_{X/Y}^2 = \frac{\text{Cov}[X, Y]^2}{\text{Var}[X] \text{Var}[Y]} = \frac{1^2}{1 \cdot 2} = \frac{1}{2}$$

Como podemos ver, tenemos que la mitad de la varianza de X y Y se explica por la regresión, lo que nos indica que los ajustes no son muy buenos.

Calculamos ahora el coeficiente de correlación lineal. Como la covarianza es positiva, el coeficiente de correlación lineal también lo será:

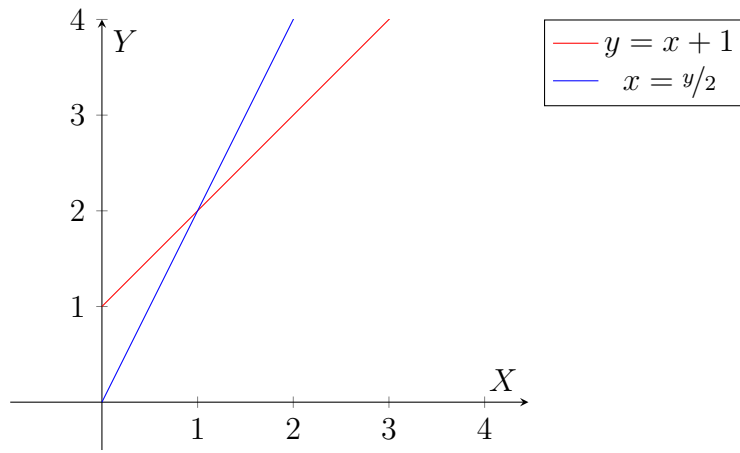
$$\rho_{X/Y} = +\sqrt{\rho_{X/Y}^2} = +\sqrt{\frac{1}{2}} = +\frac{1}{\sqrt{2}} = \frac{\sqrt{2}}{2}$$

Como podemos ver, el coeficiente de correlación lineal es positivo, lo que nos indica que las variables están correlacionadas positivamente.

Calculamos ahora el error cuadrático medio cometido al predecir cada variable según cada una de las funciones de regresión.

$$\begin{aligned} \text{ECM}(\hat{Y}(X)) &= E[(Y - \hat{Y}(X))^2] = E[(Y - X - 1)^2] = E[Y^2 + X^2 + 1 - 2Y + 2X - 2YX] = \\ &= E[Y^2] + E[X^2] + 1 - 2E[Y] + 2E[X] - 2E[XY] = 6 + 2 + 1 - 4 + 2 - 6 = 1 \\ \text{ECM}(\hat{X}(Y)) &= E[(X - \hat{X}(Y))^2] = E[(X - Y/2)^2] = E[X^2 + Y^2/4 - YX] = \\ &= E[X^2] + \frac{1}{4}E[Y^2] - E[XY] = 2 + \frac{1}{4} \cdot 6 - 3 = \frac{1}{2} \end{aligned}$$

Por último, representamos las rectas de regresión:



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.

Dado que la función de densidad es uniforme en el cuadrado unidad, tenemos que (compruébese):

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

Definimos la siguiente función auxiliar:

$$\begin{aligned} g : \mathbb{R} &\longrightarrow \mathbb{R} \\ t &\longmapsto \begin{cases} 1, & t \in [0, 1] \\ 0, & \text{en caso contrario} \end{cases} \end{aligned}$$

Tenemos que $f_{(X,Y)}(x, y) = g(x)g(y)$ para todo $(x, y) \in \mathbb{R}^2$, por lo que X e Y

son independientes. Por tanto, tenemos que las curvas de regresión son:

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

Al ser independientes, en particular son incorreladas, por lo que:

$$\rho_{X/Y}^2 = \eta_{X/Y}^2 = \eta_{Y/X}^2 = 0$$

Calculamos ahora el error cuadrático medio cometido al predecir cada variable según cada una de las funciones de regresión.

$$\begin{aligned}\text{ECM}(\hat{Y}(X)) &= E[(Y - \hat{Y}(X))^2] = E[(Y - E[Y])^2] = \text{Var}[Y] \\ \text{ECM}(\hat{X}(Y)) &= E[(X - \hat{X}(Y))^2] = E[(X - E[X])^2] = \text{Var}[X]\end{aligned}$$

Calculamos por tanto ambas varianzas:

$$\begin{aligned}E[X^2] &= \int_0^1 x^2 \, dx = \left[\frac{x^3}{3} \right]_0^1 = \frac{1}{3} \\ E[Y^2] &= \int_0^1 y^2 \, dy = \left[\frac{y^3}{3} \right]_0^1 = \frac{1}{3} \\ \text{Var}[X] &= E[X^2] - E[X]^2 = \frac{1}{3} - \left(\frac{1}{2} \right)^2 = \frac{1}{12} \\ \text{Var}[Y] &= E[Y^2] - E[Y]^2 = \frac{1}{3} - \left(\frac{1}{2} \right)^2 = \frac{1}{12}\end{aligned}$$

Por tanto, tenemos que:

$$\begin{aligned}\text{ECM}(\hat{Y}(X)) &= \text{Var}[Y] = \frac{1}{12} \\ \text{ECM}(\hat{X}(Y)) &= \text{Var}[X] = \frac{1}{12}\end{aligned}$$

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.

Calculamos las marginales:

$$f_X(x) = \int_{-x}^x 1 \, dy = 2x \quad \forall x \in]0, 1[$$

$$f_Y(y) = \int_{|y|}^1 1 \, dx = 1 - |y| \quad \forall y \in]-1, 1[$$

Calculamos ahora las condicionadas. Dados $x^* \in]0, 1[$ e $y^* \in]-1, 1[$, tenemos que:

$$f_{Y|X=x^*}(y) = \frac{f_{(X,Y)}(x^*, y)}{f_X(x^*)} = \frac{1}{2x^*} \quad \forall y \in]-x^*, x^*[$$

$$f_{X|Y=y^*}(x) = \frac{f_{(X,Y)}(x, y^*)}{f_Y(y^*)} = \frac{1}{1 - |y^*|} \quad \forall x \in]|y^*|, 1[$$

Calculamos ahora las curvas de regresión. Tenemos que:

$$\hat{Y}(x) = E[Y | X = x] = \int_{-x}^x y \cdot \frac{1}{2x} \, dy = \frac{1}{2x} \int_{-x}^x y \, dy = \frac{1}{2x} \left(\frac{x^2}{2} - \frac{x^2}{2} \right) = 0 \quad \forall x \in]0, 1[$$

$$\hat{X}(y) = E[X | Y = y] = \int_{|y|}^1 x \cdot \frac{1}{1 - |y|} \, dx = \frac{1}{1 - |y|} \int_{|y|}^1 x \, dx = \frac{1}{1 - |y|} \left(\frac{1}{2} - \frac{y^2}{2} \right) =$$

$$= \frac{1 - y^2}{2(1 - |y|)} = \frac{(1 - y^2)(1 + |y|)}{2(1 - |y|)(1 + |y|)} = \frac{1 + |y|}{2} \quad \forall y \in]-1, 1[$$

Como la curva de regresión de Y sobre X es constante, la recta de regresión de Y sobre X es la misma que la curva de regresión. Calculamos ahora la recta de regresión X sobre Y . Para ello, previamente calculamos:

$$E[X] = \int_0^1 x \cdot 2x \, dx = 2 \int_0^1 x^2 \, dx = 2 \left[\frac{x^3}{3} \right]_0^1 = \frac{2}{3}$$

$$E[Y] = \int_{-1}^1 y \cdot (1 - |y|) \, dy = \int_{-1}^0 y(1 + y) \, dy + \int_0^1 y(1 - y) \, dy =$$

$$= \left[\frac{y^2}{2} + \frac{y^3}{3} \right]_{-1}^0 + \left[\frac{y^2}{2} - \frac{y^3}{3} \right]_0^1 = -\frac{1}{2} + \frac{1}{3} + \frac{1}{2} - \frac{1}{3} = 0$$

$$E[X^2] = \int_0^1 x^2 \cdot 2x \, dx = 2 \int_0^1 x^3 \, dx = 2 \left[\frac{x^4}{4} \right]_0^1 = \frac{1}{2}$$

$$E[Y^2] = \int_{-1}^1 y^2 \cdot (1 - |y|) \, dy = \int_{-1}^0 y^2(1 + y) \, dy + \int_0^1 y^2(1 - y) \, dy =$$

$$= \left[\frac{y^3}{3} + \frac{y^4}{4} \right]_{-1}^0 + \left[\frac{y^3}{3} - \frac{y^4}{4} \right]_0^1 = \frac{1}{3} - \frac{1}{4} + \frac{1}{3} - \frac{1}{4} = \frac{1}{6}$$

$$E[XY] = \int_0^1 \int_{-x}^x xy \, dy \, dx = \int_0^1 x \left[\frac{y^2}{2} \right]_{-x}^x \, dx = \int_0^1 x \left(\frac{x^2}{2} - \frac{x^2}{2} \right) \, dx = 0$$

$$\text{Var}[X] = E[X^2] - E[X]^2 = \frac{1}{2} - \left(\frac{2}{3} \right)^2 = \frac{1}{18}$$

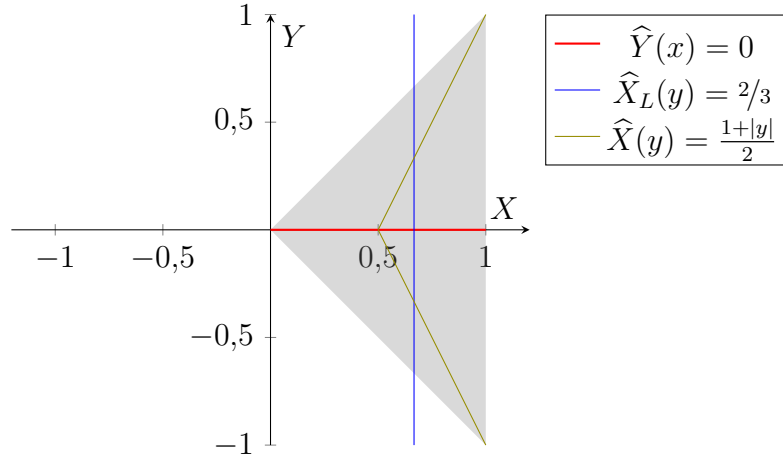
$$\text{Var}[Y] = E[Y^2] - E[Y]^2 = \frac{1}{6} - 0 = \frac{1}{6}$$

$$\text{Cov}[X, Y] = E[XY] - E[X]E[Y] = 0$$

Por tanto, tenemos que la recta de regresión de X sobre Y es:

$$\hat{X}_L(y) = E[X] + \frac{\text{Cov}[X, Y]}{\text{Var}[Y]} (y - E[Y]) = \frac{2}{3} \quad \forall y \in]-1, 1[$$

La representación gráfica de las rectas y curvas de regresión es:



El coeficiente de determinación lineal es:

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

Como vemos, el ajuste lineal es nulo, ya que no explica nada de la variabilidad de X ni de Y .

El coeficiente de correlación lineal es:

$$\rho_{X/Y} = 0$$

Como vemos, las variables son linealmente incorreladas, como era de esperar tras el cálculo del coeficiente de determinación lineal.

Calculamos las razones de correlación. Como la curva de regresión de Y sobre X coincide con la recta de regresión, tenemos que:

$$\eta_{Y/X}^2 = \rho_{Y/X}^2 = 0$$

Respecto de la curva de regresión de X sobre Y , tenemos que:

$$\eta_{X/Y}^2 = \frac{\text{Var}[\hat{X}(Y)]}{\text{Var}[X]} = \frac{\text{Var}\left[\frac{1+|Y|}{2}\right]}{\text{Var}[X]}$$

Calculemos la varianza del numerador:

$$\begin{aligned}
 E[|Y|] &= \int_{-1}^1 |y| \cdot (1 - |y|) dy = \int_{-1}^0 -y(1 + y) dy + \int_0^1 y(1 - y) dy = \\
 &= -\left[\frac{y^2}{2} + \frac{y^3}{3}\right]_{-1}^0 + \left[\frac{y^2}{2} - \frac{y^3}{3}\right]_0^1 = \frac{1}{2} - \frac{1}{3} + \frac{1}{2} - \frac{1}{3} = \frac{1}{3} \\
 E\left[\frac{1 + |Y|}{2}\right] &= \frac{1}{2} + \frac{1}{2}E[|Y|] = \frac{1}{2} + \frac{1}{2} \cdot \frac{1}{3} = \frac{1}{2} + \frac{1}{6} = \frac{2}{3} \\
 E\left[\left(\frac{1 + |Y|}{2}\right)^2\right] &= \frac{1}{4} [1 + 2E[|Y|] + E[Y^2]] = \frac{1}{4} \left[1 + 2 \cdot \frac{1}{3} + \frac{1}{6}\right] = \frac{11}{24} \\
 \text{Var}\left[\frac{1 + |Y|}{2}\right] &= E\left[\left(\frac{1 + |Y|}{2}\right)^2\right] - E\left[\frac{1 + |Y|}{2}\right]^2 = \frac{11}{24} - \left(\frac{2}{3}\right)^2 = \frac{1}{72}
 \end{aligned}$$

Por tanto, tenemos que:

$$\eta_{X/Y}^2 = \frac{\text{Var}\left[\frac{1+|Y|}{2}\right]}{\text{Var}[X]} = \frac{1/72}{1/18} = \frac{1}{4}$$

En este caso, tan solo se explica un cuarto de la variabilidad de X por la regresión, por lo que el ajuste sigue siendo malo.

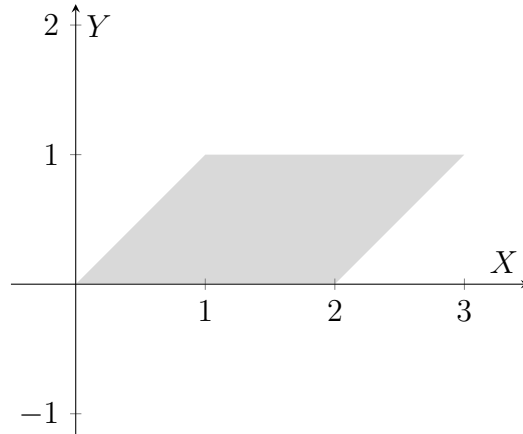
Calculamos ahora el error cuadrático medio de cada una de las funciones de regresión:

$$\begin{aligned}
 \text{ECM}(\hat{Y}(X)) &= E[(Y - \hat{Y}(X))^2] = E[Y^2] = \frac{1}{6} \\
 \text{ECM}(\hat{X}(Y)) &= E[(X - \hat{X}(Y))^2] = E\left[\left(X - \frac{1 + |Y|}{2}\right)^2\right] = \\
 &= E[X^2] + E\left[\left(\frac{1 + |Y|}{2}\right)^2\right] - E[X(1 + |Y|)] = \\
 &= E[X^2] + E\left[\left(\frac{1 + |Y|}{2}\right)^2\right] - E[X] - E[|Y| \cdot X] = \\
 &= \frac{1}{2} + \frac{11}{24} - \frac{2}{3} - \frac{1}{4} = \frac{1}{24} \\
 E[X \cdot |Y|] &= \int_0^1 \int_{-x}^x x|y| \cdot 1 dy dx = \int_0^1 2x \int_0^x y dy dx = \\
 &= \int_0^1 2x \cdot \frac{x^2}{2} dx = \int_0^1 x^3 dx = \left[\frac{x^4}{4}\right]_0^1 = \frac{1}{4} \\
 \text{ECM}(\hat{X}_L(Y)) &= E[(X - E[X])^2] = \text{Var}[X] = \frac{1}{18}
 \end{aligned}$$

Aquí podemos ver que el error cuadrático medio cometido con la regresión lineal es mayor que el cometido con la curva de regresión, lo que indica que la regresión lineal es peor que la curva de regresión (como es de esperar).

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.

Veamos gráficamente el paralelogramo:



Notemos por P al paralelogramo. Tenemos que la función de densidad conjunta es:

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

Para calcular k , tenemos que:

$$1 = \int_P f_{(X,Y)} = \int_P k = k\lambda(P) = k \cdot 2 \cdot 1 = 2k \implies k = \frac{1}{2}$$

Calculamos en primer lugar las marginales. Para $x \in [0, 3]$, tenemos que:

- Si $x \in [0, 1]$, tenemos que:

$$f_X(x) = \int_0^x \frac{1}{2} dy = \frac{x}{2}$$

- Si $x \in [1, 2]$, tenemos que:

$$f_X(x) = \int_0^1 \frac{1}{2} dy = \frac{1}{2}$$

- Si $x \in [2, 3]$, tenemos que:

$$f_X(x) = \int_{x-2}^1 \frac{1}{2} dy = \frac{1}{2}(1 - x + 2) = \frac{3-x}{2}$$

Por tanto, tenemos que:

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

Para $y \in [0, 1]$, tenemos que:

$$f_Y(y) = \int_y^{y+2} \frac{1}{2} dx = \frac{1}{2}(y+2-y) = 1$$

Calculamos ahora las condicionadas. Dado $y^* \in [0, 1]$, tenemos que:

$$f_{X|Y=y^*}(x) = \frac{f_{(X,Y)}(x, y^*)}{f_Y(y^*)} = \frac{1/2}{1} = \frac{1}{2} \quad \forall x \in [0, 3]$$

Respecto a la condicionada de Y dado $X = x^*$, distinguimos en función de x^* :

$$f_{Y|X=x^*}(y) = \frac{f_{(X,Y)}(x^*, y)}{f_X(x^*)} = \begin{cases} 1/x^*, & x^* \in [0, 1], y \in [0, x^*] \\ 1/2, & x^* \in [1, 2], y \in [0, 1] \\ \frac{1}{3-x^*}, & x^* \in [2, 3], y \in [x^*-2, 1] \end{cases}$$

Tenemos que calcular los siguientes errores cuadráticos medios:

$$\text{ECM}(\hat{X}(Y)) = E[\text{Var}[X | Y]]$$

$$\text{ECM}(\hat{Y}(X)) = E[\text{Var}[Y | X]]$$

Para calcular $E[\text{Var}[X | Y]]$, tenemos que:

$$E[X | Y = y] = \int_y^{y+2} x \cdot \frac{1}{2} dx = \frac{1}{2} \left[\frac{x^2}{2} \right]_y^{y+2} = \frac{1}{2} \left(\frac{(y+2)^2}{2} - \frac{y^2}{2} \right) = \frac{4y-4}{4} = y-1$$

$$\begin{aligned} E[X^2 | Y = y] &= \int_y^{y+2} x^2 \cdot \frac{1}{2} dx = \frac{1}{2} \left[\frac{x^3}{3} \right]_y^{y+2} = \frac{1}{2} \left(\frac{(y+2)^3}{3} - \frac{y^3}{3} \right) = \frac{6y^2 + 12y + 8}{6} = \\ &= y^2 + 2y + 4/3 \end{aligned}$$

$$\begin{aligned} \text{Var}[X | Y = y] &= E[X^2 | Y = y] - E[X | Y = y]^2 = y^2 + 2y + \frac{4}{3} - (y-1)^2 = \\ &= y^2 + 2y + 4/3 - y^2 + 2y - 1 = 4y + 1/3 \end{aligned}$$

$$E[Y] = \int_0^1 y dy = \left[\frac{y^2}{2} \right]_0^1 = \frac{1}{2}$$

$$E[\text{Var}[X | Y]] = E \left[4Y + \frac{1}{3} \right] = 4E[Y] + \frac{1}{3} = 2 + \frac{1}{3} = \frac{7}{3}$$

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 ?

Suponemos que la recta de regresión de Y sobre X es $x + 4y = 1$. Por tanto, tenemos que:

$$y = \hat{Y}(x) = \frac{1}{4} - \frac{x}{4} = E[Y] + \frac{\text{Cov}[X, Y]}{\text{Var}[X]} (x - E[X]) \implies \frac{\text{Cov}[X, Y]}{\text{Var}[X]} = -\frac{1}{4}$$

Por otro lado, tenemos que la recta de regresión de X sobre Y es $x + 5y = 2$. Por tanto, tenemos que:

$$x = \hat{X}(y) = 2 - 5y = E[X] + \frac{\text{Cov}[X, Y]}{\text{Var}[Y]} (y - E[Y]) \implies \frac{\text{Cov}[X, Y]}{\text{Var}[Y]} = -5$$

Por tanto, el coeficiente de determinación lineal es:

$$\rho_{X/Y}^2 = \frac{\text{Cov}[X, Y]^2}{\text{Var}[X] \text{Var}[Y]} = -\frac{1}{4} \cdot (-5) = \frac{5}{4} > 1$$

Esto es un absurdo, luego nuestra suposición era errónea. La recta de regresión de Y sobre X es $x + 5y = 2$ y la de X sobre Y es $x + 4y = 1$.

2. Calcular el coeficiente de correlación lineal y la proporción de varianza de cada variable que queda explicada por la regresión lineal.

Como la recta de regresión de Y sobre X es $x + 5y = 2$, tenemos que:

$$y = \hat{Y}(x) = \frac{2-x}{5} = E[Y] + \frac{\text{Cov}[X, Y]}{\text{Var}[X]} (x - E[X]) \implies \frac{\text{Cov}[X, Y]}{\text{Var}[X]} = -\frac{1}{5}$$

Por otro lado, como la recta de regresión de X sobre Y es $x + 4y = 1$, tenemos que:

$$x = \hat{X}(y) = 1 - 4y = E[X] + \frac{\text{Cov}[X, Y]}{\text{Var}[Y]} (y - E[Y]) \implies \frac{\text{Cov}[X, Y]}{\text{Var}[Y]} = -4$$

Por tanto, el coeficiente de determinación lineal es:

$$\rho_{X/Y}^2 = \frac{\text{Cov}[X, Y]^2}{\text{Var}[X] \text{Var}[Y]} = -\frac{1}{5} \cdot (-4) = \frac{4}{5} = 0,8$$

Por tanto, la proporción de varianza de cada variable que queda explicada por la regresión lineal es un 80 %. Además, como la covarianza es negativa, el coeficiente de correlación lineal es:

$$\rho_{X/Y} = -\sqrt{\rho_{X/Y}^2} = -\sqrt{\frac{4}{5}} = -\frac{2\sqrt{5}}{5}$$

3. Calcular las medias de ambas variables.

De la forma de las rectas de regresión, planteamos el siguiente sistema:

$$\begin{cases} E[Y] + 1/5 E[X] = 2/5 \\ E[X] + 4E[Y] = 1 \end{cases}$$

Resolviendo el sistema, llegamos a que:

$$E[X] = -3 \quad E[Y] = 1$$