

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. Esperanza Condicionada	5
1.1. Esperanza condicionada	5
1.2. Momentos condicionados	11
1.2.1. Momentos condicionados no centrados	11
1.2.2. Momentos condicionados centrados	11
1.3. Varianza condicionada	12
1.4. Regresión Mínimo Cuadrática	13
1.4.1. Búsqueda de la función de regresión óptima, φ_{opt}	14
1.5. Razones de Correlación	17
2. Relaciones de problemas	21

1. Esperanza Condicionada

Cuando se considera un conjunto de variables aleatorias definidas en relación a un determinado experimento, es usual que existan relaciones entre ellas. El problema de regresión consiste en encontrar una función matemática que permita aproximar el valor de determinadas variables, conocidos los valores del resto.

Observación. Este concepto de regresión es el análogo al visto en la asignatura de EDIP. En este caso, en vez de establecer un carácter Y en función de un carácter X en función de sus frecuencias, buscamos establecer una relación entre dos variables aleatorias.

En el presente curso, por simplicidad, nos limitaremos a analizar el problema de regresión bidimensional; es decir, dadas dos variables aleatorias X e Y , queremos encontrar una función matemática que nos permita aproximar el valor de Y conocido el valor de X . Para encontrar esta función matemática usaremos el criterio de optimalidad denominado mínimos cuadrados, visto ya en EDIP.

1.1. Esperanza condicionada

Para encontrar esta función matemática, necesitamos introducir el concepto de esperanza condicionada.

Definición 1.1 (Esperanza condicionada). Dadas dos variables aleatorias X e Y sobre el mismo espacio de probabilidad, se define la esperanza condicionada de X a Y , notada por $E[X | Y]$, como la variable aleatoria siguiente (en el caso de que exista):

$$E[X | Y](y) := E[X | Y = y]$$

Notemos que esta esperanza es una esperanza normal, solo que considerando la distribución condicionada de X a $Y = y$.

Veamos en función de si es discreta o continua:

- Si X e Y son discretas, entonces:

$$E[X | Y = y] = \sum_{x \in E_x} x \cdot P[X = x | Y = y] \quad \text{con } P[Y = y] > 0$$

- Si X e Y son continuas, entonces:

$$E[X | Y = y] = \int_{-\infty}^{\infty} x \cdot f_{X|Y=y}(x) dx \quad \text{con } f_Y(y) > 0$$

Notemos que la definición de $E[X | Y = y]$ requiere que la distribución condicionada de X a $Y = y$ esté bien definida. Por este motivo, es preciso exigir que $P[Y = y] \neq 0$ en el caso discreto, y $f_Y(y) \neq 0$ en el caso continuo. Así, la variable aleatoria $E[X | Y]$ es una función de la variable Y , definida sobre el conjunto de sus valores, E_Y .

Veamos ahora qué hemos de imponer para que exista la esperanza condicionada de X a Y .

Proposición 1.1. *Sean dos variables aleatorias X e Y sobre el mismo espacio de probabilidad. Entonces:*

$$\exists E[X] \implies \exists E[X | Y]$$

Demostración. Distinguimos en función de si X e Y son discretas o continuas:

- Si X e Y son discretas, entonces:

$$\begin{aligned} E[X | Y = y] &= \sum_{x \in E_x} |x| \cdot P[X = x | Y = y] = \sum_{x \in E_x} |x| \cdot \frac{P[X = x, Y = y]}{P[Y = y]} = \\ &= \frac{\sum_{x \in E_x} |x| \cdot P[X = x, Y = y]}{P[Y = y]} \leq \frac{\sum_{x \in E_x} |x| \cdot P[X = x]}{P[Y = y]} = \frac{E[|X|]}{P[Y = y]} < \infty \end{aligned}$$

donde en la desigualdad hemos hecho uso de que

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

y, al final, hemos usado que $E[|X|] < \infty$ por tener $E[X]$ finita. Por tanto, $E[X | Y]$ existe.

- Si X e Y son continuas, entonces:

$$\begin{aligned} E[X | Y = y] &= \int_{-\infty}^{\infty} |x| \cdot f_{X|Y=y}(x) dx = \int_{-\infty}^{\infty} |x| \cdot \frac{f_{X,Y}(x, y)}{f_Y(y)} dx = \\ &= \frac{\int_{-\infty}^{\infty} |x| \cdot f_{X,Y}(x, y) dx}{f_Y(y)} \end{aligned}$$

Como $E[|X|] < \infty$, entonces:

$$\begin{aligned} E[|X|] &= \int_{-\infty}^{\infty} |x| \int_{-\infty}^{\infty} f_{X,Y}(x, y) dx dy = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} |x| \cdot f_{X,Y}(x, y) dx dy < \infty \implies \\ &\implies \int_{-\infty}^{\infty} |x| \cdot f_{X,Y}(x, y) dx < \infty \end{aligned}$$

Por tanto, $E[X | Y]$ existe. □

Ejemplo. Sea (X, Y) un vector aleatorio bidimensional con función de densidad conjunta:

$$f_{X,Y}(x, y) = \begin{cases} 2 & \text{si } 0 < x < y < 1 \\ 0 & \text{en otro caso} \end{cases}$$

Calcular $E[X | Y]$ y $E[Y | X]$.

Tenemos que:

$$E[X | Y = y] = \int_{-\infty}^{\infty} x \cdot f_{X|Y=y}(x) dx$$

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

Para calcular ambas funciones de densidad condicionada, necesitamos calcular las marginales de X e Y :

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

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

Así, las funciones de densidad condicionada son:

$$f_{X|Y=y}(x) = \frac{f_{X,Y}(x, y)}{f_Y(y)} = \frac{2}{2y} = \frac{1}{y} \quad \text{si } 0 < x < y < 1$$

$$f_{Y|X=x}(y) = \frac{f_{X,Y}(x, y)}{f_X(x)} = \frac{2}{2(1-x)} = \frac{1}{1-x} \quad \text{si } 0 < x < y < 1$$

Por tanto, las esperanzas condicionadas son:

$$E[X | Y = y] = \int_{-\infty}^{\infty} x \cdot f_{X|Y=y}(x) dx = \int_0^y x \cdot \frac{1}{y} dx = \frac{1}{y} \int_0^y x dx = \frac{1}{y} \left[\frac{x^2}{2} \right]_0^y = \frac{y}{2}$$

$$E[Y | X = x] = \int_{-\infty}^{\infty} y \cdot f_{Y|X=x}(y) dy = \int_x^1 y \cdot \frac{1}{1-x} dy = \frac{1}{1-x} \int_x^1 y dy = \frac{1}{1-x} \left[\frac{y^2}{2} \right]_x^1 =$$

$$= \frac{1}{1-x} \left(\frac{1}{2} - \frac{x^2}{2} \right) = \frac{1}{2} \cdot \frac{1-x^2}{1-x} = \frac{1+x}{2}$$

Al igual que considerábamos $E[g(X)]$, consideramos ahora la esperanza condicionada de una función de X a Y .

Definición 1.2 (Esperanza condicionada de una función). Dadas dos variables aleatorias X e Y sobre el mismo espacio de probabilidad, y una función $g : \mathbb{R} \rightarrow \mathbb{R}$ medible, se define la esperanza condicionada de $g(X)$ a Y , notada por $E[g(X) | Y]$, como la variable aleatoria siguiente (en el caso de que exista):

$$E[g(X) | Y](y) := E[g(X) | Y = y]$$

Veamos en función de si es discreta o continua:

- Si X e Y son discretas, entonces:

$$E[g(X) | Y = y] = \sum_{x \in E_x} g(x) \cdot P[X = x | Y = y] \quad \text{con } P[Y = y] > 0$$

- Si X e Y son continuas, entonces:

$$E[g(X) | Y = y] = \int_{-\infty}^{\infty} g(x) \cdot f_{X|Y=y}(x) dx \quad \text{con } f_Y(y) > 0$$

Al igual que ocurría con la esperanza condicionada de X a Y , para que exista basta con imponer que $E[g(X)] < \infty$. Es decir, la demostración de la siguiente proposición es análoga a la del proposición anterior.

Proposición 1.2. Sean dos variables aleatorias X e Y sobre el mismo espacio de probabilidad, y una función $g : \mathbb{R} \rightarrow \mathbb{R}$ medible. Entonces:

$$\exists E[g(X)] \implies \exists E[g(X) | Y]$$

Ejemplo. En el experimento aleatorio del lanzamiento de tres monedas se consideran las variables:

- X : Número de caras.
- Y : Diferencia, en valor absoluto, entre el número de caras y el número de cruces.

Calcular $E[X^2 | Y]$.

El espacio muestral es el siguiente:

$$\Omega = \{CCC, CC+, C+C, +CC, C++ , +C+, ++C, +++\}$$

La función masa de probabilidad conjunta viene dada por:

$X \setminus Y$	1	3	$P[X = x_i]$
0	0	$1/2^3$	$1/2^3$
1	$3/2^3$	0	$3/2^3$
2	$3/2^3$	0	$3/2^3$
3	0	$1/2^3$	$1/2^3$
$P[Y = y_j]$	$3/2^2$	$1/2^2$	1

donde además hemos calculado las marginales de X e Y . Entonces, tenemos que:

$$\begin{aligned}
 E[X^2 | Y = 1] &= \sum_{x \in E_x} x^2 \cdot P[X = x | Y = 1] = \\
 &= 1^2 \cdot P[X = 1 | Y = 1] + 2^2 \cdot P[X = 2 | Y = 1] = \\
 &= 1 \cdot \frac{P[X = 1, Y = 1]}{P[Y = 1]} + 4 \cdot \frac{P[X = 2, Y = 1]}{P[Y = 1]} = \\
 &= 1 \cdot \frac{3/2^3}{3/2^2} + 4 \cdot \frac{3/2^3}{3/2^2} = \frac{5}{2} \\
 E[X^2 | Y = 3] &= \sum_{x \in E_x} x^2 \cdot P[X = x | Y = 3] = \\
 &= 0^2 \cdot P[X = 0 | Y = 3] + 3^2 \cdot P[X = 3 | Y = 3] = \\
 &= 0 \cdot \frac{P[X = 0, Y = 3]}{P[Y = 3]} + 9 \cdot \frac{P[X = 3, Y = 3]}{P[Y = 3]} = \\
 &= 9 \cdot \frac{1/2^3}{1/2^2} = \frac{9}{2}
 \end{aligned}$$

Por tanto, la esperanza condicionada de X^2 a Y es:

$$E[X^2 | Y] = \begin{cases} 5/2 & \text{si } Y = 1 \\ 9/2 & \text{si } Y = 3 \end{cases}$$

Propiedades de la esperanza condicionada

Introducimos las siguientes propiedades, que se deducen de forma directa en la mayoría de los casos haciendo uso de que la esperanza condicionada de X a Y es la esperanza de X con respecto a la distribución condicionada de X a Y .

Proposición 1.3. *Sea Y una variable aleatoria discreta sobre el espacio de probabilidad (Ω, \mathcal{A}, P) y $c \in \mathbb{R}$. Entonces:*

$$E[c | Y] = c$$

Demostración. Como Y es discreta, entonces:

$$E[c | Y = y] = \sum_{x \in E_x} x \cdot P[X = x | Y = y] = c \cdot P[X = c | Y = y] = c$$

□

Notemos que no consideramos Y continua, ya que c es una variable aleatoria constante (discreta) y solo hemos definido la esperanza condicionada para variables aleatorias del mismo tipo.

Proposición 1.4 (Linealidad). *Sean X e Y dos variables aleatorias sobre el espacio de probabilidad (Ω, \mathcal{A}, P) . Entonces:*

$$E[(aX + b) | Y] = aE[X | Y] + b \quad \forall a, b \in \mathbb{R}$$

Proposición 1.5. *Sean X e Y dos variables aleatorias sobre el espacio de probabilidad (Ω, \mathcal{A}, P) . Entonces:*

$$E[aX_1 + bX_2 | Y] = aE[X_1 | Y] + bE[X_2 | Y] \quad \forall a, b \in \mathbb{R}$$

Proposición 1.6. *Sean X e Y dos variables aleatorias sobre el espacio de probabilidad (Ω, \mathcal{A}, P) . Si $X \geq 0$, entonces:*

$$E[X | Y] \geq 0 \wedge E[X | Y] = 0 \iff P[X = 0] = 1$$

Proposición 1.7 (Conservación del Orden). *Sean X_1, X_2 e Y tres variables aleatorias sobre el espacio de probabilidad (Ω, \mathcal{A}, P) . Si $X_1 \leq X_2$, entonces:*

$$E[X_1 | Y] \leq E[X_2 | Y]$$

Proposición 1.8. *Sean X e Y dos variables aleatorias sobre el espacio de probabilidad (Ω, \mathcal{A}, P) , y consideramos una función $g : \mathbb{R} \rightarrow \mathbb{R}$ medible. Entonces:*

$$\text{Si } X, Y \text{ independientes} \implies E[g(X) | Y] = E[g(X)]$$

En particular, tomando $g = Id$, tenemos:

$$\text{Si } X, Y \text{ independientes} \implies E[X | Y] = E[X]$$

Demostración. Distinguimos en función de si X e Y son discretas o continuas:

- Si X e Y son discretas, como X e Y son independientes, entonces:

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

Por tanto, tenemos:

$$E[g(X) \mid Y = y] = \sum_{x \in E_X} g(x) \cdot P[X = x \mid Y = y] = \sum_{x \in E_X} g(x) \cdot P[X = x] = E[g(X)]$$

- Si X e Y son continuas, como X e Y son independientes, entonces:

$$f_{X|Y=y}(x) = f_X(x) \quad \forall x, y$$

Por tanto, tenemos:

$$E[g(X) \mid Y = y] = \int_{-\infty}^{\infty} g(x) \cdot f_{X|Y=y}(x) dx = \int_{-\infty}^{\infty} g(x) \cdot f_X(x) dx = E[g(X)]$$

□

Proposición 1.9. Sean X e Y dos variables aleatorias sobre el espacio de probabilidad (Ω, \mathcal{A}, P) , y consideramos una función $g : \mathbb{R} \rightarrow \mathbb{R}$ medible. Entonces:

$$E[E[g(X) \mid Y]] = E[g(X)]$$

En particular, tomando $g = Id$, tenemos:

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

Demostración. Distinguimos en función de si X e Y son discretas o continuas:

- Si X e Y son discretas, entonces, como $E[g(X) \mid Y]$ es una variable aleatoria en función de Y , usando la esperanza de una función de una variable aleatoria, tenemos:

$$\begin{aligned} E[E[g(X) \mid Y]] &= \sum_{y \in E_Y} E[g(X) \mid Y = y] \cdot P[Y = y] = \\ &= \sum_{y \in E_Y} \sum_{x \in E_X} g(x) \cdot P[X = x \mid Y = y] \cdot P[Y = y] = \\ &= \sum_{x \in E_X} g(x) \cdot \sum_{y \in E_Y} P[X = x, Y = y] = \\ &= \sum_{x \in E_X} g(x) \cdot P[X = x] = E[g(X)] \end{aligned}$$

- Si X e Y son continuas, entonces, como $E[g(X) \mid Y]$ es una variable aleatoria en función de Y , usando la esperanza de una función de una variable aleatoria,

tenemos:

$$\begin{aligned}
 E[E[g(X) | Y]] &= \int_{-\infty}^{\infty} E[g(X) | Y = y] \cdot f_Y(y) dy = \\
 &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(x) \cdot f_{X|Y=y}(x) dx \cdot f_Y(y) dy = \\
 &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(x) \cdot f_{X,Y}(x, y) dx dy = \\
 &= \int_{-\infty}^{\infty} g(x) \cdot \int_{-\infty}^{\infty} f_{X,Y}(x, y) dy dx = \\
 &= \int_{-\infty}^{\infty} g(x) \cdot f_X(x) dx = E[g(X)]
 \end{aligned}$$

□

1.2. Momentos condicionados

Como una variable aleatoria que es, se pueden definir los momentos condicionados de una variable aleatoria a otra. Comenzamos con los momentos condicionados no centrados.

1.2.1. Momentos condicionados no centrados

Definición 1.3 (Momento condicionado). Dadas dos variables aleatorias X e Y sobre el mismo espacio de probabilidad y $k \in \mathbb{N}$, se define el momento condicionado de orden k de X a Y como la variable aleatoria siguiente (en el caso de que exista):

$$E[X^k | Y]$$

Usando la esperanza condicionada de una función, tenemos que $\exists E[X^k]$ implica que exista $E[X^k | Y]$, y se calcula como:

- Si X e Y son discretas, entonces:

$$E[X^k | Y = y] = \sum_{x \in E_x} x^k \cdot P[X = x | Y = y] \quad \text{con } P[Y = y] > 0$$

- Si X e Y son continuas, entonces:

$$E[X^k | Y = y] = \int_{-\infty}^{\infty} x^k \cdot f_{X|Y=y}(x) dx \quad \text{con } f_Y(y) > 0$$

1.2.2. Momentos condicionados centrados

Respecto de los momentos condicionados centrados, tenemos la siguiente definición.

Definición 1.4 (Momento condicionado centrado). Dadas dos variables aleatorias X e Y sobre el mismo espacio de probabilidad y $k \in \mathbb{N}$, se define el momento condicionado de orden k centrado de X a Y como la variable aleatoria siguiente (en el caso de que exista):

$$E[(X - E[X | Y])^k | Y]$$

En caso de existencia, los valores de estas variables se calculan teniendo en cuenta que al considerar el condicionamiento a un valor arbitrario, $Y = y$, la variable $E[X | Y]$ toma el valor $E[X | Y = y]$ y, por tanto:

$$E[(X - E[X | Y])^k | Y = y] = E[(X - E[X | Y = y])^k | Y = y] \quad \forall y \in E_Y$$

Entonces, ya que dado $Y = y$, $E[X | Y = y]$ es una constante, la variable $(X - E[X | Y = y])^k$ sólo depende de X , y aplicando de nuevo la expresión de la esperanza condicionada de una función de una variable aleatoria, se tiene:

- Si X e Y son discretas, entonces:

$$E[(X - E[X | Y])^k | Y = y] = \sum_{x \in E_X} (x - E[X | Y = y])^k \cdot P[X = x | Y = y] \quad \text{con } P[Y = y] > 0$$

- Si X e Y son continuas, entonces:

$$E[(X - E[X | Y])^k | Y = y] = \int_{-\infty}^{\infty} (x - E[X | Y = y])^k \cdot f_{X|Y=y}(x) dx \quad \text{con } f_Y(y) > 0$$

Las propiedades de los momentos condicionados son similares a las de los momentos sin condicionar. En particular, la existencia de los momentos condicionados no centrados equivale a la de los momentos condicionados centrados.

1.3. Varianza condicionada

La varianza condicionada es un caso particular de los momentos condicionados centrados, y es de especial relevancia en el estudio de la regresión, por lo que estudiaremos su definición y propiedades.

Definición 1.5. Dadas dos variables aleatorias X e Y sobre el mismo espacio de probabilidad, se define la varianza condicionada de X a Y como el momento condicionado centrado de orden 2 de X a Y :

$$\text{Var}[X | Y] := E[(X - E[X | Y])^2 | Y]$$

Veamos ahora algunas de sus propiedades.

Proposición 1.10. Sean X e Y dos variables aleatorias sobre el espacio de probabilidad (Ω, \mathcal{A}, P) . Si $\exists E[X^2]$, entonces:

- $\exists \text{Var}[X | Y]$ y $\text{Var}[X | Y] \geq 0$.
- $\text{Var}[X | Y] = E[X^2 | Y] - (E[X | Y])^2$.

Demostración. Demostramos cada uno de los apartados:

- Dado que $\exists E[X^2]$, entonces $\exists E[X^2 | Y]$. Como la existencia de momentos no centrados implica la de momentos centrados, entonces $\exists E[(X - E[X | Y])^2 | Y] = \text{Var}[X | Y]$.

Por otro lado, como $E[(X - E[X | Y])^2 | Y]$, es la esperanza condicionada de una variable aleatoria no negativa, entonces $\text{Var}[X | Y] \geq 0$.

- Partiendo de la definición de varianza condicionada, fijado $Y = y$, tenemos:

$$\begin{aligned}
 \text{Var}[X | Y = y] &= E[(X - E[X | Y = y])^2 | Y = y] = \\
 &= E[X^2 - 2XE[X | Y = y] + (E[X | Y = y])^2 | Y = y] \stackrel{(*)}{=} \\
 &\stackrel{(*)}{=} E[X^2 | Y = y] - 2E[E[X | Y = y] \cdot X | Y = y] + E[(E[X | Y = y])^2 | Y = y] \stackrel{(**)}{=} \\
 &\stackrel{(**)}{=} E[X^2 | Y = y] - 2E[X | Y = y] \cdot E[X | Y = y] + (E[X | Y = y])^2 = \\
 &= E[X^2 | Y = y] - 2(E[X | Y = y])^2 + (E[X | Y = y])^2 = \\
 &= E[X^2 | Y = y] - (E[X | Y = y])^2
 \end{aligned}$$

donde en (*) hemos usado la linealidad de la esperanza condicionada, y en (**) hemos usado que, como ya hemos condicionado a $Y = y$, $E[X | Y = y]$ es una constante.

□

Proposición 1.11 (Descomposición de la varianza). Sean X e Y dos variables aleatorias sobre el espacio de probabilidad (Ω, \mathcal{A}, P) . Si $\exists E[X^2]$, entonces:

$$\text{Var}[X] = \text{Var}[E[X | Y]] + E[\text{Var}[X | Y]]$$

1.4. Regresión Mínimo Cuadrática

Explicamos de nuevo el problema de regresión, que ya se introdujo para Estadística Descriptiva. Dadas dos variables aleatorias, X e Y , definidas sobre el mismo espacio de probabilidad, el problema de regresión de Y sobre X consiste en determinar una función de X que proporcione una representación, lo más precisa posible, de la variable Y . Esto es, se trata de aproximar la variable Y por una función de X , de manera que la aproximación sea óptima en algún sentido preestablecido.

$$Y \approx \varphi(X)$$

donde:

- Y es la variable dependiente, explicada o endógena.
- X es la variable independiente, explicativa o exógena.
- φ es la función de regresión.

El criterio de optimalidad más usual para abordar el problema de regresión es el basado en el principio de mínimos cuadrados, y consiste en encontrar la función que minimiza la media de las desviaciones cuadráticas de las aproximaciones respecto de los verdaderos valores de la variable aproximada. Esto es, se trata de encontrar una función, φ , que minimice $E[(Y - \varphi(X))^2]$.

Definición 1.6 (Error cuadrático medio). Dadas dos variables aleatorias X e Y sobre el mismo espacio de probabilidad, el error cuadrático medio asociado a la función de regresión φ de forma que $Y \approx \varphi(X)$ es:

$$\text{E.C.M.}(\varphi) = E[(Y - \varphi(X))^2]$$

Por tanto, el problema de regresión abordado bajo esta perspectiva consiste en encontrar una función, φ_{opt} , que minimice el E.C.M.

1.4.1. Búsqueda de la función de regresión óptima, φ_{opt}

En la presente subsección, suponemos que estamos buscando aproximar Y por una función de X , $\varphi(X)$, y que hemos decidido que la función de regresión óptima es aquella que minimiza el E.C.M. asociado a la aproximación.

Observación. Se deja para el lector generalizar el caso para aproximar X por una función de Y .

Así, el problema de regresión se reduce a encontrar la función de regresión óptima, φ_{opt} , que minimiza el E.C.M. El siguiente teorema no se demostrará por ser un resultado que excede los conocimientos de la asignatura, pero nos proporciona la solución al problema de regresión bajo el criterio de mínimos cuadrados.

Teorema 1.12. Sea Y una variable aleatoria sobre el espacio de probabilidad (Ω, \mathcal{A}, P) , y X una variable aleatoria sobre el mismo espacio de probabilidad. Entonces, $E[Y | X]$ minimiza el E.C.M. Es decir:

$$\varphi_{\text{opt}}(X) = E[Y | X]$$

Además, esta función de regresión óptima es única.

El E.C.M. asociado a la función de regresión óptima, que es el mínimo error cuadrático medio cometido al aproximar Y a partir de X , es:

$$\begin{aligned} \text{E.C.M.}(\varphi_{\text{opt}}) &= \text{E.C.M.}(E[Y | X]) = E[(Y - E[Y | X])^2] = \\ &= E[E[(Y - E[Y | X])^2 | X]] = E[\text{Var}[Y | X]] \end{aligned}$$

Considerando la descomposición de la varianza, tenemos:

$$\text{Var}[Y] = \text{Var}[E[Y | X]] + E[\text{Var}[Y | X]] = \text{Var}[\varphi_{\text{opt}}(X)] + \text{E.C.M.}(\varphi_{\text{opt}})$$

Definición 1.7 (Curva de regresión mínimo cuadrática). Dadas dos variables aleatorias X e Y sobre el mismo espacio de probabilidad, la curva de regresión mínimo cuadrática de Y sobre X es la curva obtenida empleando $\varphi_{\text{opt}}(X)$. Es decir:

- Curva de Regresión Mínimo Cuadrática de Y sobre X :

$$\hat{Y}(x) = E[Y | X = x] \quad \forall x \in E_X$$

- Curva de Regresión Mínimo Cuadrática de X sobre Y :

$$\hat{X}(y) = E[X | Y = y] \quad \forall y \in E_Y$$

Veamos algunos casos particulares, para los que antes debemos introducir las siguientes definiciones.

Definición 1.8 (Dependencia funcional). Dadas dos variables aleatorias X e Y sobre el mismo espacio de probabilidad, se dice que Y depende funcionalmente de X si existe una función $f : E_X \rightarrow E_Y$ tal que:

$$Y = f(X) \quad \forall x \in E_X$$

Dicha función f se denomina curva de dependencia.

Proposición 1.13. Sean X, Y dos variables aleatorias sobre el mismo espacio de probabilidad.

$$\text{Var}[Y | X] = 0 \iff Y \text{ depende funcionalmente de } X$$

Demostración. Demostramos mediante la doble implicación:

\implies) Partimos de:

$$\text{Var}[Y | X = x] = 0 \quad \forall x \in E_X$$

Por tanto, fijado $x \in E_X$, $\exists!$ $c_x \in E_Y$ tal que:

$$[Y | X = x] = c_x$$

Por tanto, definimos la función $f : E_X \rightarrow E_Y$ como:

$$f(x) = c_x \quad \forall x \in E_X$$

Por tanto, Y depende funcionalmente de X .

\impliedby) Partimos de que Y depende funcionalmente de X , es decir, $\exists f : E_X \rightarrow E_Y$ tal que:

$$Y = f(X)$$

Entonces, tenemos que:

$$\begin{aligned} \text{Var}[Y | X = x] &= \text{Var}[f(X) | X = x] = \\ &= \text{Var}[f(x) | X = x] = 0 \quad \forall x \in E_X \end{aligned}$$

donde hemos usado que la varianza de una constante es 0.

□

Definición 1.9 (Dependencia recíproca). Dadas dos variables aleatorias X e Y sobre el mismo espacio de probabilidad, se dice que hay dependencia recíproca entre X e Y si Y depende funcionalmente de X y X depende funcionalmente de Y con la misma función. Es decir, existe una función $f : E_X \rightarrow E_Y$ tal que:

$$\begin{aligned} Y &= f(X) & \forall x \in E_X \\ X &= f^{-1}(Y) & \forall y \in E_Y \end{aligned}$$

Dicha función f se denomina curva de dependencia.

Corolario 1.13.1. Sean X, Y dos variables aleatorias sobre el mismo espacio de probabilidad.

$$\text{Var}[Y | X] = \text{Var}[X | Y] = 0 \iff Y \text{ y } X \text{ dependen recíprocamente}$$

Demostración. Demostramos mediante doble implicación:

\implies) Como corolario de la proposición anterior, sabemos que $\exists f : E_X \rightarrow E_Y$, $g : E_Y \rightarrow E_X$ tales que:

$$\begin{aligned} Y &= f(X) & \forall x \in E_X \\ X &= g(Y) & \forall y \in E_Y \end{aligned}$$

Sea $c_x \in E_y$, y consideramos su preimagen

$$f^{-1}(c_x) = \{x \in E_X \mid f(x) = c_x\} = \{x \in E_X \mid Y = c_x\}$$

No obstante, usando la demostración de la proposición anterior tenemos que dicho c_x es único. Por tanto, $f^{-1} = g$ y, por tanto, Y y X dependen recíprocamente.

\impliedby) Se tiene de forma directa como corolario de la proposición anterior.

□

Algunos casos particulares son:

- Si Y depende funcionalmente de X con curva de dependencia $Y = f(X)$, entonces la curva de regresión mínimo cuadrática de Y sobre X es la propia curva de dependencia:

$$y = f(x) = E[Y | X = x] \quad \forall x \in E_X$$

Demostración. Dado que $Y = f(X)$, entonces:

$$E[Y | X = x] = E[f(X) | X = x] = E[f(x) | X = x] = f(x) \quad \forall x \in E_X$$

□

- Si hay dependencia recíproca entre X e Y , es decir, $Y = f(X)$ y $X = f^{-1}(Y)$, entonces ambas curvas de regresión mínimo cuadrática coinciden con las curvas de dependencia:

$$\begin{aligned} y = f(x) &= E[Y | X = x] & \forall x \in E_X \\ x = f^{-1}(y) &= E[X | Y = y] & \forall y \in E_Y \end{aligned}$$

- Si X e Y son independientes, entonces la curva de regresión mínimo cuadrática de Y sobre X es la esperanza de Y , y la de X sobre Y es la esperanza de X :

$$\begin{aligned} \hat{Y}(x) &= E[Y] & \forall x \in E_X \\ \hat{X}(y) &= E[X] & \forall y \in E_Y \end{aligned}$$

Como vemos, estas curvas de regresión mínimo cuadrática son constantes y paralelas a los ejes, lo que muestra que no tiene sentido plantear un problema de regresión en este caso.

Demostración. Dado que X e Y son independientes, sabemos de forma directa que:

$$\begin{aligned} E[Y | X = x] &= E[Y] & \forall x \in E_X \\ E[X | Y = y] &= E[X] & \forall y \in E_Y \end{aligned}$$

□

Por tanto, a modo de resumen, en la Tabla 1.1 se recogen las predicciones de Y según lo que estemos observando.

	Sin observar X	Observando X	Para $X = x$
\hat{Y}	$E[Y]$	$E[Y X]$	$E[Y X = x]$
E.C.M.	$\text{Var}[Y]$	$\text{Var}[Y X]$	$\text{Var}[Y X = x]$

Tabla 1.1: Resumen de las predicciones de Y sobre en función de lo que se observe.

1.5. Razones de Correlación

Buscamos ahora estudiar el grado de bondad de la aproximación mínimo cuadrática de cada variable a partir de la otra. Partimos de la siguiente igualdad:

$$Y = E[Y | X] + (Y - E[Y | X])$$

Sabemos que $\hat{Y} = E[Y | X]$ es la mejor aproximación que hemos obtenido. Vemos por tanto que el error es la variable aleatoria $(Y - E[Y | X])$ unidades, que denominaremos *residuo*.

Definición 1.10 (Residuo). Dadas dos variables aleatorias X e Y sobre el mismo espacio de probabilidad, el residuo asociado a la aproximación de Y a partir de X es la siguiente variable aleatoria:

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

La aproximación será mejor cuanto menor sea $|R|$. Esta es una variable aleatoria, cuya comparación para estudiar la bondad no es sencilla. Usaremos por tanto valores numéricos que resuman la bondad de la aproximación.

Como primer intento, buscamos comparar mediante $E[R]$. Tenemos que:

$$E[R] = E[Y - E[Y | X]] = E[Y] - E[E[Y | X]] = E[Y] - E[Y] = 0$$

Por tanto, siempre será nulo, lo que no nos aporta información sobre la bondad de la aproximación. Buscamos por tanto comparar mediante la varianza de R , conocida como *varianza residual*.

Definición 1.11 (Varianza residual). Dadas dos variables aleatorias X e Y sobre el mismo espacio de probabilidad, la varianza residual asociada a la aproximación de Y a partir de X es varianza del residuo, es decir:

$$\text{Var}[R] = \text{Var}[Y - E[Y | X]]$$

Calculemos el valor de la varianza residual.

$$\begin{aligned} \text{Var}[R] &= \text{Var}[Y - E[Y | X]] = E[(Y - E[Y | X])^2] - (E[Y - E[Y | X]])^2 = \\ &= E[(Y - E[Y | X])^2] - \cancel{E[R]^2} = E[(Y - E[Y | X])^2] = \\ &= E[E[(Y - E[Y | X])^2 | X]] = E[\text{Var}[Y | X]] = \text{E.C.M.}(\varphi_{\text{opt}}) \end{aligned}$$

Por tanto, usando la descomposición de la varianza, tenemos que:

$$\text{Var}[Y] = \text{Var}[E[Y | X]] + E[\text{Var}[Y | X]] = \text{Var}[\hat{Y}] + \text{Var}[R] = \text{Var}[\hat{Y}] + \text{E.C.M.}(\varphi_{\text{opt}}) \quad (1.1)$$

Vemos que el $\text{E.C.M.}(\varphi_{\text{opt}})$ será menor conforme mayor sea $\text{Var}[\hat{Y}]$, por lo que la aproximación será mejor conforme mayor sea $\text{Var}[\hat{Y}] = \text{Var}[E[Y | X]]$. No obstante, usar este valor para comparar la bondad de la aproximación introduce los siguientes problemas:

- $\text{Var}[E[Y | X]]$ no es adimensional, por lo que no es un valor que podamos comparar.
- No es invariante frente a cambios de escala, lo cual puede llevar a conclusiones engañosas. Por ejemplo, sean:
 - Y una variable aleatoria que mide la altura de una persona en metros.
 - Y' una variable aleatoria que mide la altura de una persona en centímetros.

Entonces, es razonable pensar que cualquier variable X debe aproximar igual de bien a Y que a $Y' = 100Y$. No obstante, tenemos que:

$$\text{Var}[E[Y' | X]] = \text{Var}[E[100Y | X]] = \text{Var}[100E[Y | X]] = 100^2 \text{Var}[E[Y | X]]$$

De esta forma, midiendo la bondad de la aproximación por su varianza sin tener en cuenta la unidad de medida de las variables aproximadas, podríamos concluir que la variable X aproxima mucho mejor a Y' que a Y .

Estos inconvenientes se salvan normalizando la varianza de la función de regresión y usando el siguiente coeficiente, que es el que buscábamos presentar en esta sección.

Definición 1.12 (Razón de correlación). Dadas dos variables aleatorias X e Y sobre el mismo espacio de probabilidad, se define la razón de correlación de Y sobre X :

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

De forma análoga se define la razón de correlación de X sobre Y :

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

Usando la Ecuación 1.1, tenemos:

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

Veamos que $\eta_{Y/X}^2$ solventa los dos problemas que presentábamos al usar $\text{Var}[E[Y | X]]$ para medir la bondad de la aproximación.

- $\eta_{Y/X}^2$ es adimensional, ya que las unidades de medida de $\text{Var}[E[Y | X]]$ y $\text{Var}[Y]$ son las mismas.
- Veamos que es invariante frente a cambios de escala. Sean Y e Y' dos variables aleatorias, y $Y' = aY$ con $a \in \mathbb{R}$. Entonces:

$$\eta_{Y'/X}^2 = \eta_{aY/X}^2 = \frac{\text{Var}[E[aY | X]]}{\text{Var}[aY]} = \frac{\text{Var}[aE[Y | X]]}{a^2 \text{Var}[Y]} = \frac{a^2 \text{Var}[E[Y | X]]}{a^2 \text{Var}[Y]} = \eta_{Y/X}^2$$

Por tanto, la razón de correlación es un valor adimensional e invariante frente a cambios de escala que nos permite comparar la bondad de la aproximación de una variable a partir de la otra. En particular, $\eta_{Y/X}^2$ mide la proporción de la varianza de Y que queda explicada por la función de regresión $\hat{Y} = E[Y | X]$. En este sentido, se puede interpretar como una medida de la bondad del ajuste de la distribución a la curva de regresión correspondiente.

Proposición 1.14. Sean X e Y dos variables aleatorias sobre el espacio de probabilidad (Ω, \mathcal{A}, P) . Entonces:

1. $0 \leq \eta_{Y/X}^2, \eta_{X/Y}^2 \leq 1$.

2. $\eta_{Y/X}^2 = 0 \iff$ La curva de regresión de Y sobre X es $Y = E[Y]$.
 $\eta_{X/Y}^2 = 0 \iff$ La curva de regresión de X sobre Y es $X = E[X]$.
3. $\eta_{Y/X}^2 = 1 \iff Y$ depende funcionalmente de X .
 $\eta_{X/Y}^2 = 1 \iff X$ depende funcionalmente de Y .
4. $\eta_{Y/X}^2 = \eta_{X/Y}^2 = 1 \iff X$ e Y tienen dependencia recíproca.

Demostración. Demostraremos tan solo los resultados para $\eta_{Y/X}^2$, ya que los resultados para $\eta_{X/Y}^2$ son análogos.

- Tenemos en primer lugar $\text{Var}[E[Y | X]] \geq 0$ y $\text{Var}[Y] \geq 0$ por ser varianzas. Por tanto lado, usamos que:

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

Tenemos que $\text{Var}[Y | X], \text{Var}[Y] \geq 0$ por ser varianzas. Además, como la esperanza de una variable aleatoria no negativa es no negativa, tenemos que $E[\text{Var}[Y | X]] \geq 0$. Por tanto, $\eta_{Y/X}^2 \leq 1$ y tenemos lo buscado.

- Tenemos que:

$$\eta_{Y/X}^2 = 0 \iff \text{Var}[E[Y | X]] = 0 \iff \exists c \in \mathbb{R} \text{ tal que } E[Y | X] = c \quad \forall x \in E_X$$

donde la última implicación se debe a las propiedades de la varianza. Por tanto, tenemos que:

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

Por tanto, tenemos que:

$$\eta_{Y/X}^2 = 0 \iff E[Y | X] = E[Y]$$

- Tenemos que:

$$\begin{aligned} \eta_{Y/X}^2 = 1 &\iff E[\text{Var}[Y | X]] = 0 \iff P[\text{Var}[Y | X] = 0] = 1 \iff \\ &\iff \text{Var}[Y | X] = 0 \end{aligned}$$

Por tanto, tenemos que Y depende funcionalmente de X .

- Por lo razonado anteriormente, tenemos que:

$$\eta_{Y/X}^2 = \eta_{X/Y}^2 = 1 \iff \text{Var}[Y | X] = 0 = \text{Var}[X | Y]$$

Por tanto, X e Y tienen dependencia recíproca.

□

2. Relaciones de problemas