Probabilidad y Estadística

Clase 3

Variables aleatorias condicionadas

Motivación

Cuando tenemos diferentes variables, que se encuentran vinculadas, saber qué ocurrió con una variable nos da información extra sobre las otras.

Las variables condicionadas aparecen en el corazón de ML,

- Dado el valor de una muestra, cual es la probabilidad de que pertenezca a cierta clase?
- Modelos de grafos, que se basan en probabilidades condicionales
- Estimación paramétrica (enfoque Bayesiano)

Variables discretas

Sean X, Y dos variables aleatorias, y sean $p_X(x) > 0$ y $p_{X,Y}(x, y)$ las función de probabilidad marginal de X y la función de probabilidad conjunta respectivamente. Se define la función de probabilidad condicionada de Y dado que $X = x$ como:

$$\begin{aligned} \boxed{p_{Y|X=x}(y)} &= \mathbb{P}\left[Y = y | X = x\right] \\ &= \frac{\mathbb{P}(X = x, Y = y)}{\mathbb{P}(X = x)} \\ &\boxed{= \frac{p_{X,Y}(x, y)}{p_X(x)}} \end{aligned}$$

Variables continuas

Sean X, Y dos variables aleatorias con función de densidad conjunta $f_{X,Y}(x, y)$ y densidad marginal de X $f_X(x)$. Se define la función de densidad condicional de Y dado $X = x$ como

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

Obs: Si $f_{Y|X=x}(y) = f_Y(y)$ X e Y son independientes.

Ejercicio 3

1. La probabilidad de acertar a un blanco es $\frac{1}{5}$. Se realizan 10 tiros independientes y se cuenta la cantidad de aciertos. Sean X la cantidad de aciertos en los 10 tiros, e Y la cantidad de aciertos en el primer tiro. Hallar la distribución de $X|Y=y$ y $Y|X=x$.

2. Sean X,Y dos v.a. conjuntamente uniformes en el triángulo de vértices (0,0), (2,2), (0,2). Hallar la función de densidad de $Y|X=x$.

$n = 10$ $\quad X =$ "# de aciertos en 10 tiros", $X \in \{0, 1, \dots, 10\}$

$X \sim Bin(10, 1/5)$

$p = 1/5$ $\quad Y =$ " " " " " el 1º tiro", $Y \in \{0, 1\}$

$Y \sim Bin(1, 1/5) = Ber(1/5)$

$X | Y = 0 \sim Bin(9, 1/5)$ $\qquad X | Y = 1 = 1 + W, \; W \sim Bin(9, 1/5)$

$\underline{0}$ _ _ _ _ _ _ _ _ _ $\qquad \underline{1}$ _ _ _ _ _ _ _ _ _

$X | Y = y = y + W, \; W \sim Bin(9, 1/5)$

$Y | X = 0 = 0$

$$Y | X = 1 \Rightarrow p_{Y|X=1}(1) = P(Y=1 | X=1) = \frac{P(Y=1, X=1)}{P(X=1)} = \frac{\cancel{1/5}\,\cancel{(4/5)^9}}{\binom{10}{1} \cancel{1/5}\,\cancel{(4/5)^9}} = \frac{1}{10}$$

$\underline{1}\ \underline{0}\ \underline{0}\ \underline{0}\ \underline{1}\ \underline{0}\ \underline{1}\ \underline{0}\ \underline{0}\ \underline{0}$

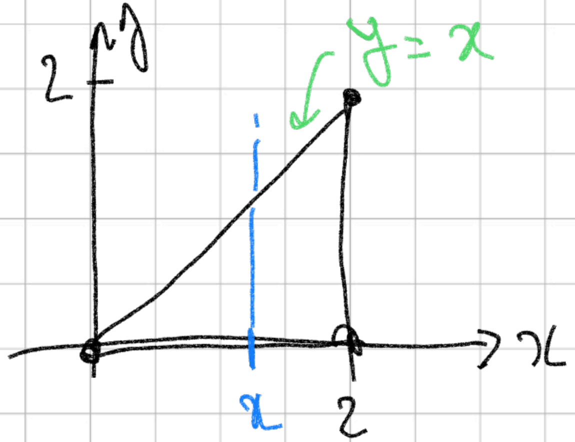
$$p_{Y|X=x}(1) = \frac{P(Y=1, X=x)}{P(X=x)} = \frac{\cancel{(1/5)^x}\,\cancel{(4/5)^{10-x}} \binom{9}{x-1}}{\binom{10}{x} \cancel{(1/5)^x}\,\cancel{(4/5)^{10-x}}}$$

$$\mathbb{P}_{Y|X=x}(1) = \frac{\binom{9}{x-1}}{\binom{10}{x}} = \frac{\dfrac{\cancel{9!}}{\cancel{(x-1)!}\,\cancel{(9-(x-1))!}}}{\dfrac{\cancel{10!}\ 10}{\cancel{x!}\ x\,\cancel{(10-x)!}}} = \frac{x}{10}$$

$$\mathbb{P}\big(\underbrace{X=5 \mid Y=1}_{X|Y=1 \,=\, 1+W}\big) = \mathbb{P}(1+W=5) = \mathbb{P}(W=4)$$

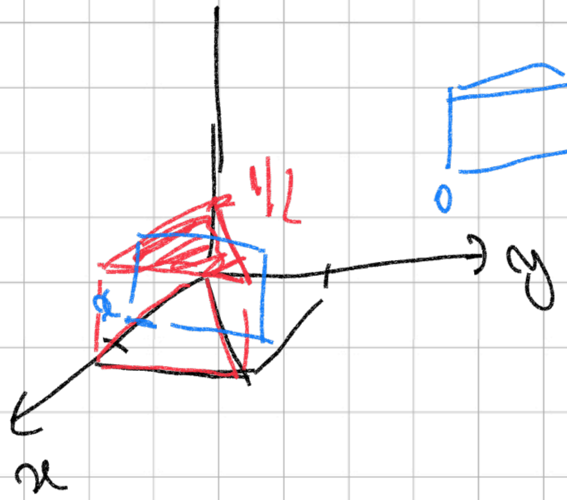
$$= \binom{9}{4}\left(1/5\right)^4\left(4/5\right)^5$$

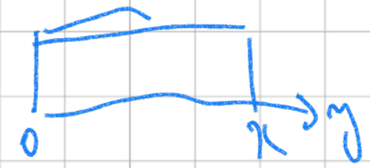
1.b



$$Y|X=x \sim \mathcal{U}(0,x)$$

$$\mathrm{Sop}_{Y|X=x} = (0,x)$$





$$f_{Y|X=x}(y) = \frac{1/2\ \mathbb{1}\{0<y<x<2\}}{x\ 1/2\ \mathbb{1}\{0<x<2\}} = 1/x\ \mathbb{1}\{0<y<x\}$$

$$f_X(x) = \int_0^x 1/2\ dy = x\, 1/2\ \mathbb{1}\{0<x<2\}$$

Factorización

Sean X, Y dos variables aleatorias con función. de densidad conjunta $f_{X,Y}(x,y)$, la misma puede descomponerse de la forma

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

$$= f_Y(y)\, f_{X|Y=y}(x)$$

Obs: Si $f_{X,Y}(x,y) = f_X(x) f_Y(y)$ X e Y son independientes.

Ejercicio 4

$f_{X,Y}(x,y) = f_{X|Y=y}(x)\, f_Y(y)$

Sean X, Y dos v.a. con función de densidad conjunta

$$f_{X,Y}(x,y) = \frac{e^{-x/2y}}{4y}\mathbf{1}\{0 < x, 1 < y < 3\}$$

e^{-x} $\mathbb{1}\{x>0\}$

Hallar la función de densidad de XIY=y

$X|Y=y \sim \mathcal{E}(1/2y)$

$\lambda e^{-\lambda x}\, \mathbb{1}\{x>0\}$

$$\underbrace{\frac{e^{-\frac{x}{2y}}}{2y}\mathbb{1}\{x>0\}}\;\underbrace{\frac{1}{2}\mathbb{1}\{1<y<3\}}_{Y\sim\mathcal{U}(1,3)}$$

$X|Y=y \sim \mathcal{E}(1/2y)$

Mezcla de v.a.

Sea M una v.a. discreta a valores $1, \ldots, n$, con función de probabilidad $p_M(m)$, y sea X una v.a. tal que se conocen las distribuciones $X|M = m,\ m = 1, 2, \ldots, n$. Luego, la distribución de X resulta

$$F_X(x) = \sum_{m=1}^{n} F_{X|M=m}(x) p_M(m)$$

Obs:

Si X es v.a.d: $p_X(x) = \sum_{m=1}^{n} p_{X|M=m}(x) p_M(m)$

Si X es v.a.c $f_X(x) = \sum_{m=1}^{n} f_{X|M=m}(x) p_M(m)$

Ejercicio 5

$$f_X(x) = \sum_m f_{X|M=m}\, p_M(m)$$

Para ir al trabajo Juan puede tomar el subte, o el tren, eligiendo viajar en tren el 60% de las veces. Si viaja en subte, el tiempo de viaje (en horas) es una v.a. con distribución U(0.75,1), mientras que si viaja en tren distribuye de manera uniforme en el intervalo (0.8, 1.25).

1. Calcular la función de densidad del tiempo viaje.

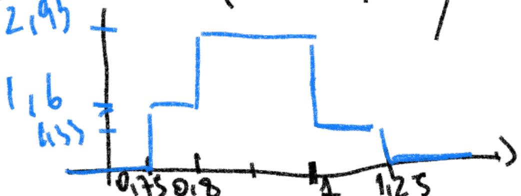
$0{,}6 \quad T \rightarrow X|T \sim \mathcal{U}(0{,}8\ ;\ 1{,}25)$

$0{,}4 \quad S \rightarrow X|S \sim \mathcal{U}(0{,}75\ ;\ 1)$

20/9

$$f_X(x) = 0{,}6\ \frac{1}{0{,}45}\ \mathbb{1}\{0{,}8;1{,}25\} + 0{,}4\ \frac{1}{0{,}25}\ \mathbb{1}\{0{,}75;1\}$$

4



Normal multivariada: dist. condicionales

$$\text{Sea } \boldsymbol{X} = [X_1, X_2]^T \sim \mathcal{N}(\boldsymbol{\mu}, \Sigma)$$

$$X_1|X_2 = x_2 \sim \mathcal{N}\left(\mu_1 + \frac{cov(X_1,X_2)}{\sigma_2^2}(x_2 - \mu_2), (1 - \frac{cov(X_1,X_2)^2}{\sigma_1^2\sigma_2^2})\sigma_1^2\right)$$

$$X_2|X_1 = x_1 \sim \mathcal{N}\left(\mu_2 + \frac{cov(X_1,X_2)}{\sigma_1^2}(x_1 - \mu_1), (1 - \frac{cov(X_1,X_2)^2}{\sigma_1^2\sigma_2^2})\sigma_2^2\right)$$

Esperanza condicional

Función de regresión

Def: Sean X, Y dos v.a. discretas, se llama función de regresión a

$$\varphi(x) = \mathbb{E}[Y|X=x] = \sum_{y \in R_y} y p_{Y|X=x}(y), \quad \forall x \in R_X$$

Def: Sean X, Y dos v.a. continuas, se llama función de regresión a

$$\varphi(x) = \mathbb{E}[Y|X=x] = \int_{y \in R_y} y f_{Y|X=x}(y), \quad \forall x \in R_X$$

Función de regresión

Def: Sean X, Y dos v.a. discretas, se llama función de regresión a

$$\varphi(x) = \mathbb{E}[Y|X=x] = \sum_{y \in R_y} y p_{Y|X=x}(y), \quad \forall x \in R_X$$

Def: Sean X, Y dos v.a. continuas, se llama función de regresión a

$$\varphi(x) = \mathbb{E}[Y|X=x] = \int_{y \in R_y} y f_{Y|X=x}(y), \quad \forall x \in R_X$$

Observar que es función de x

Ejercicio 1

Siguiendo con los ejercicios de la clase pasada calcular la función de regresión de X|Y=y o Y|X=x según corresponda

1. La probabilidad de acertar a un blanco es ⅕. Se realizan 10 tiros independientes y se cuenta la cantidad de aciertos. Sean X la cantidad de aciertos en los 10 tiros, e Y la cantidad de aciertos en el primer tiro. Hallar la distribución de X|Y=y y Y|X=x.

2. Sean X,Y dos v.a. conjuntamente uniformes en el triángulo de vértices (0,0), (2,2), (0,2). Hallar la función de densidad de Y|X=x.

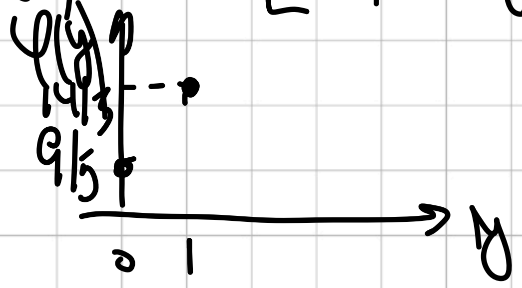
Sean X, Y dos v.a. con función de densidad conjunta

3) $$f_{X,Y}(x,y) = \frac{e^{-x/2y}}{4y}\mathbf{1}\{0 < x, 1 < y < 3\}$$

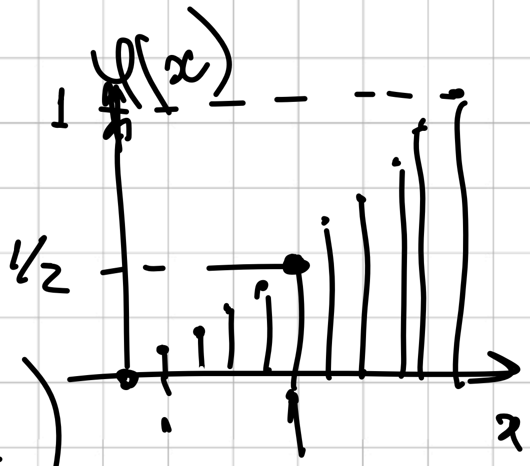
Hallar la función de densidad de X|Y=y

Caso 1 $\quad X \mid Y=y = y + W, \quad W \sim Bin(9, 1/5)$

$$\varphi(y) = \mathbb{E}[X \mid Y=y] = \mathbb{E}[y+W] = y + \mathbb{E}[W] = y + \frac{9}{5}$$



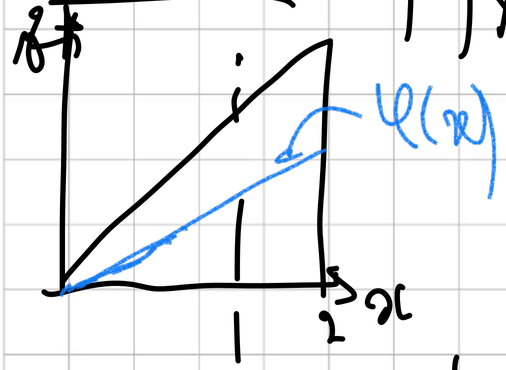
$$\varphi(x) = \mathbb{E}[Y \mid X=x] = x/10$$



$$Y \mid X=x \sim Ber\left(\frac{x}{10}\right)$$

caso 2

$Y \mid X=x \sim \mathcal{U}(0,x)$



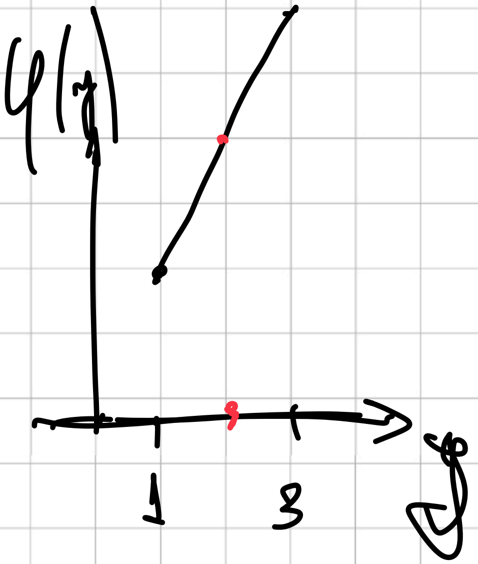
$$\varphi(x) = \mathbb{E}[Y \mid X=x] = x/2$$

caso 3

$X \mid Y=y \sim \mathcal{E}(1/2y)$

$$\varphi(y) = \mathbb{E}[X \mid Y=y] = 2y$$

$y \in (1,3)$



Esperanza condicional

Def: La variable aleatoria esperanza condicional de Y dada X se define como

$$\mathbb{E}[Y|X] = \varphi(X).$$

Además $\varphi(X)$ satisface que $\mathbb{E}[(Y - \varphi(X))\, t(X)] = 0$ para toda función medible $t : R_X \to \mathbb{R}$, tal que $\mathbb{E}[t(X)] < \infty$.

Esperanza condicional

Def: La variable aleatoria esperanza condicional de Y dada X se define como

$$\mathbb{E}[Y|X] = \varphi(X).$$

Además $\varphi(X)$ satisface que $\mathbb{E}[(Y - \varphi(X))\, t(X)] = 0$ para toda función medible $t : R_X \to \mathbb{R}$, tal que $\mathbb{E}[t(X)] < \infty$.

$\mathbb{E}[Y|X]$ es el mejor predictor de Y basado en X (i.e. es la proyección ortogonal de Y en el espacio de funciones de X)

Propiedades

1. $\mathbb{E}[Y] = \mathbb{E}[\mathbb{E}[Y|X]]$

2. $\mathbb{E}[r(X)s(Y)|X] = r(X)\mathbb{E}[s(Y)|X]$, para r, s tal que $r(X)s(X)$, $r(X)$ y $s(Y)$ tienen esperanza finita

3. $\mathbb{E}[aY_1 + bY_2|X] = a\mathbb{E}[Y_1|X] + b\mathbb{E}[Y_2|X]$

4. $\mathbb{E}[Y|X] = \mathbb{E}[Y]]$ si X y Y son independientes

Ejercicio 2

1. Siguiendo con los ejemplos del Ej. 1 calcular la esperanza condicional de X|Y o Y|X según corresponda
2. Para el último ejemplo, calcular E[X]

Caso 1 : $\varphi(x) = \mathbb{E}[Y \mid X = x] = \frac{x}{10} \to \mathbb{E}[Y \mid X] = \frac{X}{10}$

$\psi(y) = \mathbb{E}[X \mid Y = y] = y + \frac{9}{5} \to \mathbb{E}[X \mid Y] = Y + \frac{9}{5}$

Caso 2: $\varphi(x) = \mathbb{E}[Y \mid X = x] = x/2 \to \mathbb{E}[Y \mid X] = X/2$.

Caso 3: $\psi(y) = \mathbb{E}[X \mid Y = y) = 2y \to \mathbb{E}[X \mid Y] = 2Y$ v.a.

$\mathbb{E}[X] = \mathbb{E}[\mathbb{E}[X \mid Y]] = \mathbb{E}[2Y] = 2\,\mathbb{E}[Y]$

$Y \sim \mathcal{U}(1,3)$

$= 2 \cdot 2 = 4$

Varianza condicional

Def: Dada $\tau(x) = var(Y|X = x)$, se define la varianza condicional como

$$\mathbb{V}(Y|X) = \tau(X) = \mathbb{E}[Y^2|X] - \mathbb{E}[Y|X]^2$$

Propiedad: (Pitágoras)

$$var(Y) = \mathbb{E}[\mathbb{V}(Y|X)] + var(\mathbb{E}[Y|X])$$

Ejercicio 3

1. Siguiendo con los ejemplos del Ej. 1 calcular la varianza condicional de X|Y o Y|X según corresponda
2. Para el último ejemplo, calcular var[X]

Algunos comportamientos límite

LGN + TCL

Ley (débil) de los grandes números

Sea $\underline{X} = (X_1, \ldots, X_n)$ una muestra aleatoria, con Xi i.i.d con media μ y varianza σ^2. Para $\epsilon > 0$ se tiene que

$$\frac{\sum^{m} X_t}{m} \to \mu$$

$$P\{|\frac{\Sigma_t X_t}{N} - \mu| > \epsilon\} \to 0 \text{ cuando } N \to \infty$$

Es decir, a medida que crece la cantidad de muestras, el promedio tiende a la media real de la distribución.

Ejercicio 1

Simular la LGN usando Python para:

1. $X_1, \ldots, X_n \overset{i.i.d}{\sim} \mathcal{N}(5, 9)$

2. $X_1, \ldots, X_n \overset{i.i.d}{\sim} \mathcal{U}(2, 4)$

   $X_1, \ldots, X_n \overset{i.i.d}{\sim} \mathcal{B}in(10, 1/4)$

Teorema central del límite

Sea $\underline{X} = (X_1, \ldots, X_n)$ una muestra aleatoria, con Xi i.i.d con media μ y varianza σ^2. Luego,

$$\mathbb{P}\left(\frac{\sum_{i}^{n} 1\, X_i - n\mu}{\sqrt{n}\sigma} \leq z\right) \overset{n\to\infty}{\to} \Phi(z), \quad \Phi(z) = \mathbb{P}(Z \leq z), Z \sim \mathcal{N}(0,1)$$

$$\sqrt{n\sigma}\left(\sum X_i\right)$$

En general decimos que para n finitos diremos que

$$\frac{\sum_{i=1}^{n} X_i - n\mu}{\sqrt{n}\,\sigma} \longrightarrow N(0,1)$$

$$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$$

$$\mathbb{P}\left(\frac{\sum_{i}^{n} 1\, X_i - n\mu}{\sqrt{n}\sigma} \leq z\right) \approx \Phi(z),$$

$$\mathbb{P}\left(\sqrt{n}\frac{\bar{X}_i - \mu}{\sigma} \leq z\right) \approx \Phi(z),$$

$$\sqrt{n\sigma}(\bar{X})$$

$$\frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \longrightarrow N(0,1)$$

Ejercicio 2

Applet

Estadística

¿Qué es la estadística?

Ahora lo que tenemos son observaciones (realizaciones) de las variables aleatorias.

El objetivo es poder hacer algún tipo de inferencia a partir de los valores observados.

Muestra aleatoria

Supongamos que tenemos un experimento aleatorio relacionado con una v.a. X. La v.a. X representa un observable del experimento aleatorio.

Los valores de X son la población de estudio, y el objetivo es saber como se comporta esa población.

Una muestra aleatoria de tamaño n, es una sucesión de n v.a **independientes** $\underline{X}_n = (X_1, \ldots, X_n)$, tal que $X_i \sim X$

Estimador

Def: Un estimador para una cierta magnitud θ (desconocida) de la distribución de cierta población es una función $\delta(\underline{X})$ de la muestra aleatoria, que devuelve un valor aproximado de θ.

Error cuadrático medio

Def: El error cuadrático medio (ECM) como $\mathbb{E}[(\delta(\underline{X}_n) - \theta)^2]$

Def: Un estimador $\delta^*(\underline{X})$es óptimo si

$ECM(\delta^*(\underline{X})) \leq ECM(\delta(\underline{X}))$ para todo $\delta(\underline{X})$.

Ejercicio 3

$$\mathbb{E}[\bar{X}] = \mathbb{E}\left[\frac{1}{n}\sum X_i\right] = \frac{1}{n}\sum_{i=1}^{n} \overbrace{\mathbb{E}[X_i]}^{\mu} = \frac{1}{n} n\mu = \mu$$

Se desea estimar la media de una variable con distribución N(5,9) a partir del promedio de n realizaciones. Hallar el ECM para distintos valores de n.

$$\underline{X} = (X_1 \dots X_n)$$

$$\hat{\mu} = \frac{1}{n}\sum_{i=1}^{n} X_i = \bar{X}$$

$$\boxed{ECM = \mathbb{E}\left[(\bar{X}-\mu)^2\right]} = \mathbb{E}\left[\bar{X}^2 + \mu^2 - 2\bar{X}\mu\right] = \mathbb{E}[\bar{X}^2] + \mu^2 - 2\mu\,\mathbb{E}[\bar{X}]$$

$$= var(\bar{X}) = var\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) = \frac{1}{n^2} var\left(\sum_{i=1}^{n} X_i\right) = \frac{1}{n^2}\sum_{i=1}^{n} var(X_i)$$

X_i indep

(porque $\mathbb{E}[\bar{X}] = \mu$)

$$= \frac{1}{n^2} n \cdot 9 = \boxed{\frac{9}{n}}$$

Estimadores de mínimos cuadrados

Estimador de mínimos cuadrados

Es común querer estimar el valor de una v.a. X a partir de una medición Y. Ejemplo: Y es una versión ruidosa de X.

Buscamos un estimador $\hat{X}$ de X tal que tenga mínimo error cuadrático medio

$$ECM = \mathbb{E}[(X - \hat{X})^2]$$

Observar que se corresponde con la distnacia asociada al p.i. canónico para v.a.

Estimador de mínimos cuadrados

En otras palabras, queremos $\hat{X} = g^*(Y)$ tal que

$$\mathbb{E}[(X - \hat{X})^2] \leq \mathbb{E}[(X - g(Y))^2] \quad \forall\, g(Y) \text{ (medible).}$$

Estimador de mínimos cuadrados

En otras palabras, queremos $\hat{X} = g^*(Y)$ tal que

$$\mathbb{E}[(X - \hat{X})^2] \leq \mathbb{E}[(X - g(Y))^2] \quad \forall\, g(Y) \text{ (medible).}$$

¿Quién era $\hat{X}$?

$$\hat{X} = \mathbb{E}[X|Y]$$

Estimador de mínimos cuadrados

En otras palabras, queremos $\hat{X} = g^*(Y)$ tal que

$$\mathbb{E}[(X - \hat{X})^2] \leq \mathbb{E}[(X - g(Y))^2] \quad \forall\, g(Y) \text{ (medible).}$$

¿Quién era $\hat{X}$? $\qquad \hat{X} = \mathbb{E}[X|Y]$

Idea de demostración: [Ejercicio]

1. Probar que el mejor estimador constante es $\mathbb{E}[X]$
2. Probar que el mejor estimador condicional es $\mathbb{E}[X|Y = y]$.
3. Dejar que Y tome todos los valores posibles (i.e. reemplazo y por Y), recupero la esperanza condicional.

Mínimos cuadrados: caso lineal

A veces obtener $\mathbb{E}[X|Y]$ puede ser muy complicado, entonces nos restringimos a los estimadores lineales.

Buscamos a, b tq $\mathbb{E}[(X-(aY+b))^2]$ sea mínima.

Resulta que $a = \frac{cov(X,Y)}{var(Y)}$ y $b = \frac{cov(X,Y)}{var(Y)}\mathbb{E}[Y] + \mathbb{E}[X]$

$$\hat{X} = \frac{cov(X,Y)}{var(Y)}\left(Y - \mathbb{E}[Y]\right) + \mathbb{E}[X]$$

$$\mathbb{E}[\hat{X}] = \mathbb{E}[X]$$

Ejercicio 4

$$\operatorname{var}(W) = \mathbb{E}[W^2] - \mathbb{E}[W]^2$$

Sea $X \sim U(0,1)$ e $Y = X^2$. Hallar la mejor aproximación lineal de Y basada en X. Comparar con la mejor estimación de Y basada en X.

$$Y = X^2 \rightarrow \mathbb{E}[Y|X] = X^2 \rightarrow \text{lo mejor } \cancel{\text{aprox}}$$

$$\operatorname{var}(X) = 1/12, \quad \mathbb{E}[X] = 1/2, \quad \mathbb{E}[Y] = \mathbb{E}[\mathbb{E}[X^2]] = \mathbb{E}[X^2] = 1/12 + (1/2)^2 = 1/3$$

$$\operatorname{cov}(X,Y) = \operatorname{cov}(X, X^2) = \mathbb{E}[X \cdot X^2] - \mathbb{E}[X^3] = \int_0^1 x^3 \cdot 1\, dx = 1/4$$

$$\Rightarrow \hat{Y} = \frac{1/4}{1/12}\left(X - 1/2\right) + 1/3 = 3X - 7/6$$