Probabilidad y Estadística

Clase 3

Variables aleatorias condicionadas

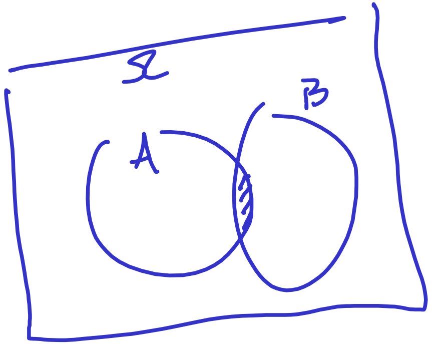
Motivación

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,

- Modelos de regresión lineal/logística
- Modelos de árboles
- Modelos de grafos, que se basan en probabilidades condicionales
- Estimación paramétrica (enfoque Bayesiano)
- NLP: por ejemplo análisis de sentimientos basado en frecuencia de aparición de palabras en un texto.

Recordemos



Probabilidad condicional:

$$\mathbb{P}(A \mid B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)}$$ ← Normaliza "$B = \Omega$"

Variables discretas

Sean X, Y dos v.a.d.

$p_{X,Y}(x,y)$: func. de probabilidad conjunta

$p_X(x)$: func. de probabilidad marginal de X

Función de probabilidad condicional de Y dado $X = x$

$$\boxed{p_{Y|X=x}(y)} = \mathbb{P}(\underline{Y = y} \mid \underline{X = x})$$

$$= \frac{\mathbb{P}(Y=y, X=x)}{\mathbb{P}(X=x)}$$

$$= \boxed{\frac{p_{X,Y}(x,y)}{p_X(x)}}$$

Qué pasa si X,Y son independientes?

$P_X = \sum_y P_{XY}(y)$

Variables Continuas

Sean X, Y dos v.a.c.

$f_{X,Y}(X,Y)$: func. de densidad conjunta

$f_X(X)$: func. de densidad marginal de X

Función de densidad condicional de Y dado $X = x$

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

Qué pasa si X,Y son independientes?

$f_{xy}(x,y) = f_x(x)\, f_y(y)$

$\Rightarrow f_{y|x} = f_y$.

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 $X|Y=y$.

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 $Y|X=x$.

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.

$Y \sim \text{Bernoulli}(1/5)$

$X = Y + \underbrace{\sum_{i=1}^{9} Y_i}_{W \sim Bin(9, 1/5)}$ con $Y, Y_1, \ldots, Y_9$ $\underline{\text{i.i.d.}}$ $\left(X \sim Bin(10, 1/5) \right)$

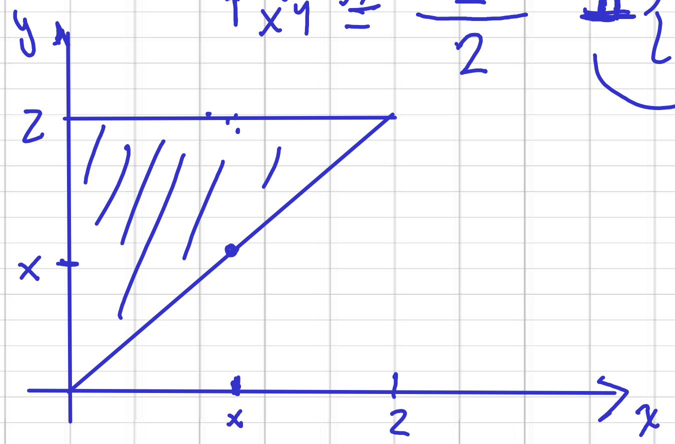
sabemos que $Y \perp\!\!\!\perp W$

$\Downarrow$

$$\boxed{X|Y=y = y + W}$$

$$\begin{cases} X \mid \underline{Y=0} = W \sim Bin(9, 1/5) \\ X \mid Y=1 = 1 + W \quad \text{con } W \sim Bin(9; 1/5) \end{cases}$$

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.



$$f_{XY}(x,y) = \frac{1}{2}\, \underbrace{1\!\!1_{\{0<x<y<2\}}(x,y)}$$

$$= \begin{cases} 1 & \text{si } (x,y) \in \underbrace{\{(x,y) : 0<x<y<2\}} \\ 0 & \text{si no} \end{cases}$$

$= (x,y) \in$ [triángulo] $= sop(f_{XY})$

$$f_X(x) = \int_{-\infty}^{\infty} f_{XY}(x,y)\, dy = \int_x^2 \frac{1}{2}\, dy = \frac{2-x}{2}\, \underline{1\!\!1_{\{0<x<2\}}}$$

$$f_{Y|X}(y) = \frac{f_{XY}(x,y)}{f_X(x)} = \frac{1\!\!1_{\{0<x<y<2\}}}{2} \cdot \frac{2}{(2-x)\,1\!\!1_{\{\}}} = \frac{1}{2-x}\, 1\!\!1_{\{x<y<2\}}$$

Factorización

$$\overbrace{P(A|B)}^{f_{X|Y}} = \frac{\overbrace{P(A\cap B)}^{f_{XY}}}{\underbrace{P(B)}_{f_Y}}$$

X

Sean X, Y dos v.a.c.

$f_{X,Y}(x,y)$: func. de densidad conjunta

$f_X(x)$: func. de densidad marginal de X

$f_Y(y)$: func. de densidad marginal de Y

$$\begin{aligned} f_{X,Y}(x,y) &= f_{Y|X=x}(y) f_X(x) \\ &= f_{X|Y=Y}(x) f_Y(y) \end{aligned}$$

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

Ejercicio 4

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

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

Hallar la función de densidad de X|Y=y

$$f_Y(y) = \int_0^\infty f_{XY}(x,y)\, dx = \dots$$

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

Mezcla de v.a.

$F_{X|m=1}$ $F_{X|m=2}$

$p_M(m) \geq 0$

$\sum_{m=1}^{n} p_M(m) = 1$

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

$$F_X(x) = \underbrace{\sum_{m=1}^{n} \overbrace{F_{X|M=m}(x)}^{P(X \leq x \mid M=m)} p_M(m)}_{\text{combinación convexa de fun. de distrib.}}$$

Obs:

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

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

Ejercicio 5

comb. convexa $\sum \alpha_i Y_i$

con $\sum \alpha_i = 1$ y $\alpha_i \geq 0$

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).

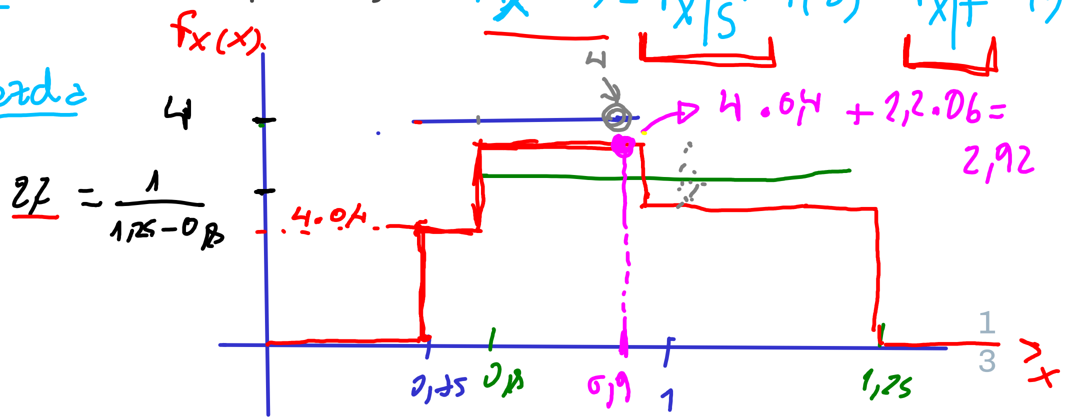
1. Calcular la función de densidad del tiempo viaje.

$$f_X(x) = f_{X|S}(x) \cdot \overset{0,4}{P(S)} + f_{X|T} \cdot \overset{0,6}{P(T)}$$

S (0,4): $X|S \sim U(0{,}75, 1)$

T (0,6): $X|S \sim U(0{,}8, 1{,}25)$

mezcla



Bayes para mezclas

Sea M una v.a. discreta a valores $1, \ldots, n$, con función de probabilidad $p_M(m)$, y sea X una v.a. continua tal que se conocen las distribuciones $f_{X|M=m}(x),\ m = 1, 2, \ldots, n$, la función de probabilidad de M dado que $X = x$ será:

$$p_{M|X=x}(m) = \frac{f_{X|M=m}(x) p_M(m)}{\sum_{m=1}^{n} f_{X|M=m}(x) p_M(m)}$$

$= f(x)$ (mezcla)

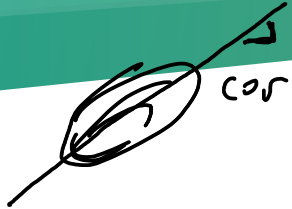
Ejercicio 5

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.
2. Calcular la probabilidad de que haya viajado en sube si un cierto día tardó exactamente 0.9 hs.

$$P(S \mid X=0{,}9) = \frac{f_{X|M=S}(0{,}9) \cdot P_M(S)}{f_X(0{,}9)} = \frac{4 \cdot 0{,}4}{2{,}92} = 0{,}55$$

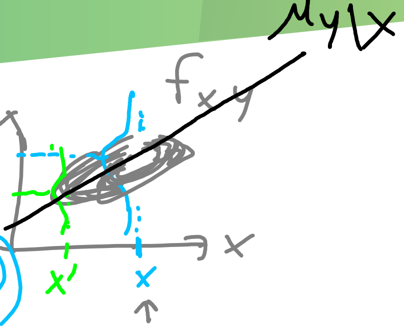
Normal multivariada: dist. condicionales



$\mathrm{cor} > 0$

$f_{xy}(x,y) = \frac{1}{(2\pi)^{?}|\Sigma|^{1/2}} e^{-(x-\mu)^T\Sigma^{-1}(x-\mu)}$

$\begin{pmatrix}\mu_1\\ \mu_2\end{pmatrix}$ $\begin{pmatrix}\sigma_1^2 & cov(x_1,x_2)\\ = & \sigma_2^2\end{pmatrix}$



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

$f_{x_1|x_2} = \frac{f_{x_1,x_2}}{f_{x_2}}$ $\quad f_{x_2} \sim N(\mu_2, \sigma_2^2)$

$= \frac{cor(x_1,x_2)\cdot\sigma_1\sigma_2}{\sigma_2^2} = cor(x_1,x_2)\frac{\sigma_1}{\sigma_2}$

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

$\frac{E[X_1|X_2]-\mu_1}{\sigma_1} = cor(x_1,x_2)\frac{x_2-\mu}{\sigma_2} \Big| X_2 = x_2$

es un estimador lineal de mín cuadrado $X_1|X_2$

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

(Ej) encontrar las densidades condicionales

Esperanza condicional

Motivación

En dónde podemos aplicar la esperanza condicional?

- Regresión: se usa para predecir el valor de la variable objetivo, dado un conjunto de variables condicionales (obervaciones)
- Reinforcement learning: usamos la esperanza condicional para estimar la recompensa esperada al tomar una acción en un estado dado. Esta estimación se usa para guiar al agente en la toma de decisiones
- Detección de anomalías: se usa para predecir el valor esperado de una variable basado en sus valores previos (y otras variables). Si el valor actual se desvía mucho, puede indicar una anomalía
- Sistemas de recomendación: se usa para predecir la preferencia del usuario por cierto contenido o servicio, por ejemplo en técnicas de collaborative filtering models

Función de regresión

$E[Y] = cte$

$\mathbb{E}[Y|x] = \int_{\mathbb{R}} y \underbrace{f_{Y|x}(y)} dy = \varphi(x)$

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

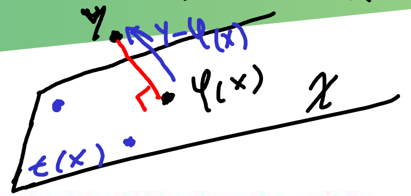
$$\varphi(x) = \mathbb{E}[Y|X = x] = \int_{y \in R_y} y f_{Y|X=x}(y), \quad \forall x \in R_X$$

Observar que es función de x

Esperanza condicional

Espacio de Hilbert (e.p.i)

$\langle X, Y \rangle = E[XY]$, $\mathcal{X} = \{X : Var(X) < \infty\}$



$P_{\mathcal{X}} Y = E\, Y|X$

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

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

$\langle Y - \varphi(X), t(X) \rangle = 0$ $\equiv$ Proyección ortogonal.

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. $\mathbb{E}[Y] = \underset{x}{\mathbb{E}}[\underbrace{\mathbb{E}[Y|X]}_{\varphi(x)}]$ $\quad \int s(y)\, f_{y|x}\, dy$

2. $\mathbb{E}[\overbrace{r(X)}^{cte}s(Y)|X] = r(X)\mathbb{E}[s(Y)|X]$, 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 $\to f_{y|x} = f_y$

$\int y\, f_{y|x}\, dy = \int y\, f_y(y)\, dy$

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.

$X|y = y + W$

$W \sim Bin(9, 1/5)$

$\boxed{P_{X|y}(x) = P_W(x-y)}$

Si $X|y = x \Leftrightarrow W = x-y$

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.

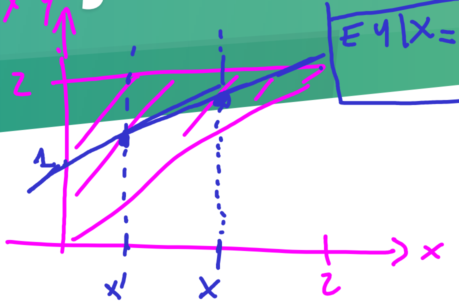
3. Sean X, Y dos v.a. con función de densidad conjunta

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

Hallar la función de densidad de X|Y=y

Ejercicio 2

f_{XY}



$\boxed{E\,Y|X = \frac{x}{2} + 1}$

$f_{Y|X}$

$E\,Y|X = \int_{-\infty}^{\infty} y\, f_{Y|X}\,dy = \int_{-\infty}^{\infty} \frac{y}{2-x}\,\mathbb{1}\{x<y<2\}\,dy$

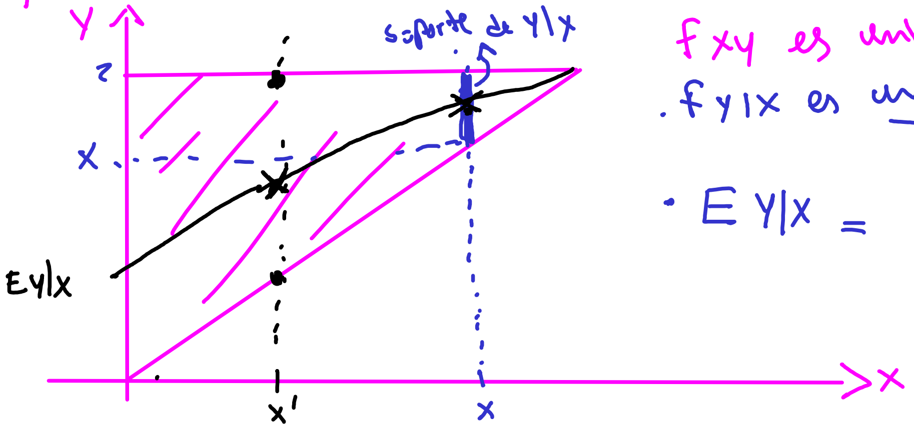
$= \int_{x}^{2} \frac{y\,dy}{2-x} = \frac{1}{2-x}\left.\frac{y^2}{2}\right|_{x}^{2} = \frac{1}{2}\,\frac{4-x^2}{2-x}$

$= \frac{1}{2}\,\frac{\cancel{(2-x)}(2+x)}{\cancel{2-x}}$

$= 1 + \frac{x}{2}$ ✓

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]

f_{XY}



f_{XY} es uniforme con soporte en , acá $f_{XY} = \frac{1}{2}$

· $f_{Y|X}$ es uniforme con soporte , acá $f_{Y|X} = \frac{1}{2-x}$

· $E\,Y|X =$

$$P_{Y|X}(y) = \frac{P_{X|Y}(x) \cdot P_Y(y)}{P_X(x)} \qquad P_Y(y) \to \text{Bernoulli}(1/5), \quad P_X(x) \to \text{Bin}(10, 1/5)$$

$$\boxed{p = 1/5}$$

$$P_{X|Y}(x) = \underline{P_W(x-y)}$$

$$= \frac{\binom{9}{x-y} \cancel{p^{x-y}} \cdot \cancel{(1-p)^{1+9-x+y-1}} \cdot \cancel{p^{y}} \cancel{(1-p)^{1-y}}}{\binom{10}{x} \cancel{p^{x}} \cancel{(1-p)^{10-x}}} = \frac{\cancel{9!} \quad x! \, (10-x)!}{(x-y)! \, (9-x+y)! \, \cancel{10!}\, 10}$$

$$P_{Y|X} = \begin{cases} \frac{10-x}{10} & y=0 \\ \frac{x}{10} & y=1 \end{cases} \qquad x = 0, \dots, 10$$

$$E[Y|X] = \sum_y y \, P_{Y|X}(y) = 1 \cdot P_{Y|X}(1) = \frac{X}{10}$$

Algunos comportamientos límite

LGN + TCL

Ley (débil) de los grandes números

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

$$\mathbb{P}(|\bar{X}_n - \mu| > \varepsilon) \xrightarrow[n\to\infty]{} 0$$

Es decir, a medida que crece la cantidad de muestras, el promedio $\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$ tiende a la media real de la distribución.

Teorema central del límite

Sea $\underline{X} = (X_1, \ldots, X_n)$ una muestra aleatoria, con X_i i.i.d. con media μ y varianza $\sigma^2 < \infty$. Sea $\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$, luego

$$\boxed{Z_n} = \underbrace{\frac{\bar{X}_n - \mu}{\sqrt{\mathbb{V}(\bar{X}_n)}}}_{\text{estandarizar}} = \frac{\sqrt{n}(\bar{X} - \mu)}{\sigma} \rightsquigarrow Z$$

con $Z \sim \underline{\mathcal{N}(0,1)}$, o equivalentemente

$$\lim_{n\to\infty} \underbrace{\mathbb{P}(Z_n \leq z)}_{F_{Z_n}(z)} = \Phi(z) = \int_{-\infty}^{z} \frac{1}{\sqrt{2\pi}} e^{-x^2/2} dx$$

Es decir, $\bar{X}_n$ tiene una distribución aproximadamente Normal, con media μ y varianza σ^2/n. Para n finito, podemos aproximar

$$\mathbb{P}\left(\underbrace{\frac{\sum_{i1}^{n} X_i - n\mu}{\sqrt{n}\sigma}}_{Z_n} \leq z\right) \approx \underline{\Phi(z)}, \qquad \mathbb{P}\left(\sqrt{n}\frac{\bar{X}_i - \mu}{\sigma} \leq z\right) \approx \underline{\Phi(z)},$$

Ejercicio 2

[Applet]()

Supongamos que la cantidad de errores en un programa tiene una distribución Poisson(4). Tenemos corriendo 25 procesos del programa y queremos saber la probabilidad de que el total de los errores sea menor a 200.

$W_i = \underline{\underline{\text{Bin}}} \quad = \quad W = \sum X_i \quad , X_i \sim \text{Bernoulli}$

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

$f_X(x|\underbrace{\mu,\Sigma}_{\text{parámetros}})$

$f_X(x|\theta)$

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

Props: el sesgo o bias de un estimador $\overbrace{\delta}^{\text{es una v.a.}}(\underline{X})$ de θ se define como la diferencia $\mathbb{E}_X\delta(\underline{X}) - \theta$, si es cero se dice insesgado.
La varianza de un estimador se define como

$$\mathbb{V}[\delta(\underline{X})] = \mathbb{E}[(\delta(\underline{X}) - \mathbb{E}\delta(\underline{X}))^2]$$

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

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

$\iff \delta^* = \arg\min_{\delta} ECM(\delta(x))$

Prop: el ECM puede escribirse como

$$\text{ECM} = \text{bias}^2[\delta(\underline{X})] + \mathbb{V}[\delta(\underline{X})]$$

Ej.

Ejercicio 3

$$\delta(\underline{x}) = \overline{X} = \frac{1}{n} \sum X_i$$

Se desea estimar la media de una variable con distribución Normal con varianza 9 a partir del promedio de n realizaciones. Hallar el ECM para distintos valores de n.

$$ECM = E\left[\left(\delta(\underline{x}) - \mu\right)^2\right]$$

$$\sigma^2 = E\left[(\bar{x}_i - \mu)^2\right]$$

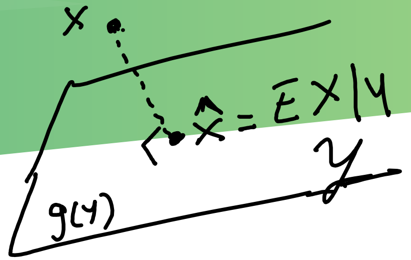
$$= E\left[\left(\underbrace{\frac{1}{n}\sum x_i}_{Z_n} - \mu\right)^2\right] = \frac{\sigma^2}{n} = \text{var } Z_n$$

$$Z_n \sim N\left(\mu, \frac{\sigma^2}{n}\right)$$

$$\frac{1}{n}\sum_{i=1}^{n} \mu \qquad n \cdot \left(\left(\frac{1}{n}\right)^2 \cdot \sigma^2\right)$$

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

$$\mathbf{ECM} = \mathbb{E}[(X - \hat{X})^2 \mid Y] \quad = \|X - \hat{X}\|^2 = \langle X - \hat{X}, X - \hat{X} \rangle$$

Observar que se corresponde con la distnacia asociada al p.i. canónico para v.a. $\langle X, Y \rangle = E[XY]$ (esp. de Hilbert)

Estimador de mínimos cuadrados

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

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

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

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.

1) considerando $\boxed{g(y) = c \in \mathbb{R}}$

$$\min E\left[(X - g(y))^2 \mid Y = y\right] = \min E\left[(X - c)^2 \mid Y = y\right]$$

$$0 = \frac{d}{dc} E\left[(x - c)^2 \mid Y = y\right] = E\left[\frac{\partial}{dc}(x-c)^2 \mid Y = y\right]$$

$$0 = E\left[2(x-c) \mid Y = y\right]$$

$$0 = EX|_{Y=y} - c$$

$$\boxed{C = EX}$$

2)

$$c(y) = C = EX \mid Y = y$$

Mínimos cuadrados: caso lineal

$\underline{g(Y) \text{ es lineal}}$

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

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

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

$$g^*(Y) = \hat{X} = \frac{cov(X,Y)}{\boxed{var\, Y}} (Y - E[Y]) + E[X]$$

Ejercicio 4

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~~ X basada en ~~X~~ Y.

Ver simu
y pensar porqué
una no es la inversa de la otra

Regresión lineal

Contamos con n observaciones conjuntas $(y_1, x_1), (y_2, x_2), \cdots, (y_n, x_n)$
el modelo de regresión lineal es

$$\boxed{Y = X\beta + \varepsilon} \quad \text{con} \quad \mathbb{E}(\varepsilon_i \mid X_i) = 0$$

donde

$$Y = \begin{pmatrix} Y_1 \\ \vdots \\ Y_n \end{pmatrix}, \qquad X = \begin{pmatrix} 1 & X_1 \\ \vdots & \vdots \\ 1 & X_n \end{pmatrix}, \qquad \varepsilon = \begin{pmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_n \end{pmatrix}$$

El estimador de cuadrados mínimos de $\hat{\beta}$ está dado por

$$\min \|\varepsilon\|^2 = \min \|Y - X\beta\|^2$$

$$\hat{\beta} = \underbrace{(X^\top X)^{-1} X^T}_{\text{Proyección ortogonal en el espacio col}(X)} Y$$