Probabilidad y Estadística

Clase 2

Algunos comportamientos límite

LGN + TCL

Ley (débil) de los grandes números

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

$$E[\bar{X}_n] \overset{\text{lineales}}{=} \frac{1}{n}\sum_{i=1}^{n} E[X_i] = \mu$$

$$Var[\bar{X}_n] = Var\left(\frac{1}{n}\sum X_i\right) \overset{\text{indep}}{=} \frac{1}{n^2}\sum Var(X_i) = \frac{1}{n^2}\cdot n \cdot \sigma^2$$

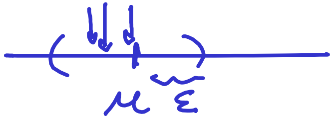
$$Var(\bar{X}_n) = \frac{\sigma^2}{n}$$

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

Convergencia en probabilidad

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

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

convergencia en distribución

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

"estandarización"

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

Normal estándar

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

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

Normal estándar

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

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

Variables aleatorias condicionadas

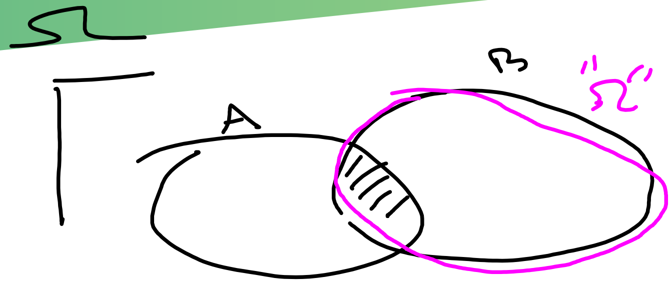
Motivación

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:

$$\mathbb{P}(A \mid \overset{\text{"}\Omega\text{"}}{\downarrow} B) = \frac{\mathbb{P}(A\cap B)}{\mathbb{P}(B)} \leftarrow \text{Normaliza}$$

$$\underbrace{P(\cdot \mid B)}_{} \text{ es una probabilidad}$$

$$P(B \mid B) = 1$$

Variables discretas

$\forall$ $\Omega \xrightarrow{X} \mathbb{R}$ $\{ P(X \in A)$

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 $X = x$

func. de prob de Y|X

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

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

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

← func. $= \boxed{p_Y(y)}$

Qué pasa si X,Y son independientes?

$p_{XY}(x,y) = p_X(x) \cdot p_Y(y)$

Variables Continuas

$$\boxed{F_X(x) = P(X \le x)}$$

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

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

caracteriza la densidad condicional

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?

$f_{XY}(x,y) = f_X(x) f_Y(y)$

Factorización

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 1

X 10 tiros $\begin{cases} Y \\ Y_2 \\ Y_3 \\ \vdots \\ Y_{10} \end{cases}$

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.

X: "# de aciertos en 10 tiros" $\Rightarrow X \sim Bin(10, 1/5)$

Y: " " " 1er tiro" $\Rightarrow Y \sim Ber(1/5)$

si $x=0 \Rightarrow P(Y=1)=0$

si $x=10 \Rightarrow P(Y=1)=1$. $\}$ No son indep.

$$X = \sum_{i=1}^{10} Y_i \quad \text{con} \quad Y_i \sim Ber(1/5) \text{ iid.}$$

$$X = Y + \sum_{i=2}^{10} Y_i = Y + W \quad \text{con} \quad W \sim Bin(9, 1/5)$$

$\Rightarrow X|Y=y = y + W$

con $W \sim Bin[9, 1/5]$

¿y cuál es la distrib. de $Y | X = x$?

$\text{con } Y, Y_i \sim Ber(1/5)$

si $\overline{x=0} \Rightarrow P(Y=0) = 1$ } No son indep.
si $x=10 \Rightarrow P(Y=1) = 1$.

$$\boxed{X = Y + \underbrace{\sum_{i=2}^{10} Y_i}_{W \sim Bin(9, 1/5)}}$$

$$\boxed{P(Y=1 | X=1)} = \frac{P(Y=1, X=1)}{P(X=1)} = \frac{\cancel{\overbrace{1/5}^{p} \cdot \overbrace{(4/5)}^{1-p}{}^{9}}}{\underbrace{\binom{10}{1}}_{10} \cancel{1/5\,(4/5)^{9}}} = \frac{1}{10}$$

$$\boxed{P(Y=1 / X=2)} = \frac{P(Y=1, X=2)}{P(X=2)} = \frac{1/5 \cdot P(W=1)}{\binom{10}{2}(1/5)^2(4/5)^8} = \frac{\cancel{(1/5)^2}\binom{9}{1}\cancel{(4/5)^8}}{\binom{10}{2}\cancel{(1/5)^2}\cancel{(4/5)^8}}$$

$$= \frac{\frac{\cancel{9!}}{1!\,\cancel{8!}}}{\frac{10! = 10 \cdot \cancel{9!}}{2!\,\cancel{8!}}} = \frac{2}{10}$$

$P(Y=1 | X=x) = \ldots$ completa $\ldots = \dfrac{x}{10}$.

$Y | X=x \sim Ber(x/10)$

Ejercicio 2

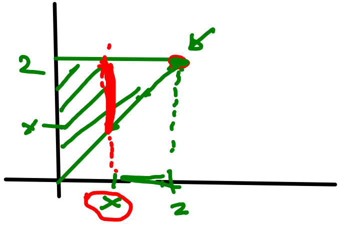
$P(X \in A) = \int_A f_X(x)\,dx$

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

$\boxed{P(Y=2 \mid X=2) = 1}$

$\cancel{\frac{P(Y=2, X=2)}{P(X=2)}}$ >0

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.



1) parametrizar la densidad conjunta.

$$f_{X,Y}(x,y) = \frac{1}{2}\mathbb{1}\{(x,y)\in \triangle\} = \frac{1}{2}\mathbb{1}\{0 \le x \le y \le 2\}$$

$$f_X(x) = \int_x^2 \frac{1}{2}\underbrace{\mathbb{1}\{0 \le x \le y \le 2\}}_{\mathbb{1}\{0 \le x \le 2\}}\,dy = \frac{1}{2}(2-x)\,\mathbb{1}\{0 \le x \le 2\}$$

$$f_{Y|X=x}(y) = \frac{\frac{1}{2}\mathbb{1}\{0 \le x \le y \le 2\}}{\frac{1}{2}(2-x)\,\mathbb{1}\{0 \le x \le 2\}} = \frac{1}{2-x}\mathbb{1}\{x \le y \le 2\}$$

$$\boxed{Y|X=x \sim \mathcal{U}(x,2)}$$

Ejercicio 3

$X \sim \mathcal{E}(\lambda) \Rightarrow f_X(x) = \lambda e^{-\lambda x} \mathbb{1}\{x > 0\}$

* Tarea: resolverlo integrando y encontrando la marginal $f_Y(y)$

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

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

Hallar la función de densidad de X|Y=y

$f_{X,Y}(x,y) = f_{X|Y=y}(x)\, f_Y(y)$ — solo depende de y

$$= \underbrace{\frac{e^{-x/2y}}{2y} \mathbb{1}\{x>0\}}_{X|Y=y \sim \mathcal{E}(1/2y)\ \checkmark} \cdot \underbrace{\frac{1}{2} \mathbb{1}\{1 < y < 3\}}_{\mathcal{U}(1,3)\ \checkmark}$$

← integra 1 y es no negativa

Mezcla de v.a.

$M \to X|M=1$

$X|M=2$

m $X|M=m$

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

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

$P(x,m) = P(x|m) P_m$

m son disjuntos

prob. total

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

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:

tengo

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

modelo un problema

$f_X \leftarrow (1)$

Ejercicio 4

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.

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

$P(M=1 \mid T=0{,}9)$

$$M = \begin{cases} 0 & \text{Juan tome el tren} \\ 1 & \text{Juan tome el subte} \end{cases}$$

$P(M=0)=0{,}6 \qquad P(M=1)=0{,}4$

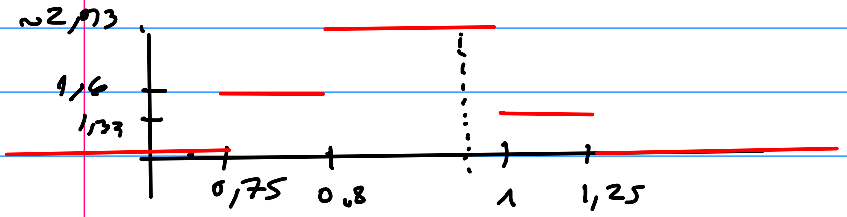
T: "tiempo de viaje"

$T/M=0 \sim \mathcal{U}(0{,}8;\ 1{,}25)$

$T/M=1 \sim N(0{,}75;\ 1)$

$$f_T(t) = f_{T/M=0}\cdot P(M=0) + f_{T/M=1}\cdot P(M=1)$$

$$= \underbrace{\frac{0{,}6}{0{,}45}}_{\sim 1{,}33} \mathbb{1}\{0{,}8 \le t \le 1{,}25\} + \underbrace{\frac{0{,}4}{0{,}25}}_{\sim 1{,}6} \mathbb{1}\{0{,}75 < t < 1\}$$



$T \mid M=1 \sim U(0{,}75;\,1)$

$f_{T\mid M=1}$

$$P(M=1 \mid T=0{,}9) = \frac{f_{T\mid M=1}(0{,}9)\cdot P(M=1)}{f_T(0{,}9)}$$

$$= \frac{\frac{0{,}4}{0{,}25}}{\frac{0{,}6}{0{,}45} + \frac{0{,}4}{0{,}25}} \approx \frac{1{,}6}{2{,}93} \sim 0{,}54$$

Normal multivariada: dist. condicionales

media Matriz de covarianza

marginales

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

condicionales

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

. $E[X_1 | X_2 = x_2]$ es lineal en x_2.

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

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 (observaciones)
- 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

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

$$\varphi(x) = \underline{\underline{\mathbb{E}}}[Y|X = x] = \underline{\sum_{y \in R_y}} y \underbrace{p_{Y|X=x}}_{\text{prob. condicional de } Y|X}(y), \quad \forall x \in R_X$$

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

$$\varphi(x) = \underline{\mathbb{E}}[Y|X = x] = \int_{y \in R_y} y \underbrace{f_{Y|X=x}}_{\text{densidad condicional de } Y|X}(y), \quad \forall x \in R_X$$

Observar que es función de x

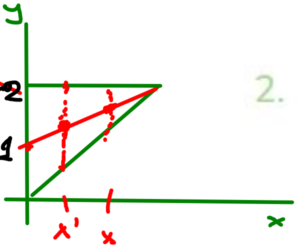
Ejercicio 5

Continuando con los ejercicios anteriores calcular la función de regresión de X|Y=y o Y|X=x según corresponda:

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

$E[Y|X=x] = \frac{x}{10}$

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.

$Y|X=x \sim \mathcal{U}(x,2)$

$E[Y|X=x] = \frac{2+x}{2} = 1 + \frac{x}{2}$

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

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

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

$E[X|Y=y] = 2y$

Hallar la función de densidad de X|Y=y

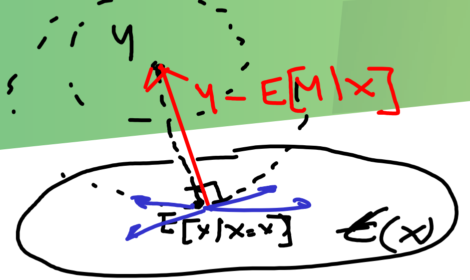
"Espacio de Hilbert"

Esperanza condicional

$\langle X, Y\rangle = E[XY]$

X, Y v.a con varianza finita

$\|X\|^2 = \langle X, X\rangle = E[X^2] = Var(X) + E[X]^2$



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

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

$\langle (Y - E[Y|X]), t(X)\rangle$ =

Además $\varphi(X)$ satisface que $\boxed{\mathbb{E}[(Y - \underbrace{\varphi(X)}_{E[Y|X]})\, t(X)] = 0}$ 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

$f_Y = f_{Y|X} \cdot f_X$

1. $\mathbb{E}[Y] = \mathbb{E}[\mathbb{E}[Y|X]]$ → esperanzas iteradas
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. $\mathbb{E}[aY_1 + bY_2|X] = a\mathbb{E}[Y_1|X] + b\mathbb{E}[Y_2|X]$ lineal
4. $\mathbb{E}[Y|X] = \mathbb{E}[Y]]$ si X y Y son independientes $f_{Y|X} = f_Y$