

# Probabilidad y Estadística

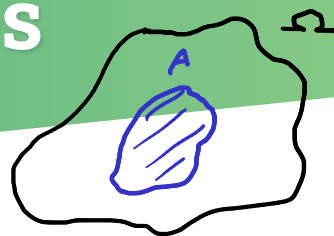
## Clase 2

Nociones de convergencia  $\epsilon, \delta, N$   $\forall N > 0, \exists \delta > 0 : \forall \epsilon > 0$   
+  $\{f_n\}_{n=1}^{\infty} ; f_n \xrightarrow{n \rightarrow \infty} f$   $\epsilon$ psilon  $\delta$ elta  $N$  fijo

# Algunos comportamientos límite

LGN + TCL

# Ley (débil) de los grandes números



$$\begin{aligned} X_1 &\sim F_X & \mu=1 \\ X_2 &\sim F_X \\ &\vdots \\ X_n &\sim F_X ; F_X \sim \mathcal{N}(\mu, \sigma^2) \end{aligned}$$

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

$$\mathbb{P}(A) = 0 \quad (\text{Evento Improbable})$$

$$\mathbb{P}(|\bar{X}_n - \mu| > \epsilon) \xrightarrow{n \rightarrow \infty} 0 \quad ; \quad \mu \text{ es un parámetro desconocido}$$

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

Normalización

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

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

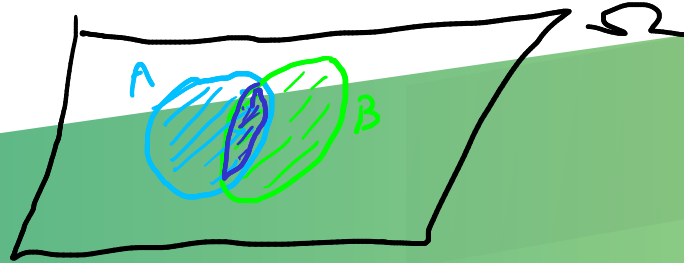
$$\underbrace{\lim_{n \rightarrow \infty} \mathbb{P}(Z_n \leq z)} = \underbrace{\Phi(z) = \int_{-\infty}^z \frac{1}{\sqrt{2\pi}} e^{-x^2/2} dx}$$

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

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

$P(\text{Evento dado que otro})$

$$P(A|B) = \frac{P(A \cap B)}{P(B)}$$



$$\frac{P(A \cap B)}{P(B)}$$

# 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:

$$\mathbb{P}(A \mid B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)}$$

# 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  $X$  ]

$X$  e  $Y$  son  
independientes

$$P(Y=y | X=x) =$$

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

$$\begin{aligned} p_{Y|X=x}(y) &= \mathbb{P}(Y = y \mid X = x) = \\ &= \frac{\mathbb{P}(Y=y, X=x)}{\mathbb{P}(X=x)} \\ &= \frac{p_{X,Y}(x, y)}{p_X(x)} \end{aligned}$$

Qué pasa si  $X, Y$  son independientes?



$$P(A|B) = \frac{P(A \cap B)}{P(B)} =$$

$$= \frac{P(B|A) \cdot P(A)}{P(B)} \sim$$

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

$$; P(B) > 0$$

$$\begin{aligned} p_{Y|X=x}(y) &= \mathbb{P}(\overset{A}{Y} = y \mid \overset{B}{X} = x) = \\ &= \frac{\mathbb{P}(Y=y, X=x)}{\mathbb{P}(X=x)} \\ &= \frac{p_{X,Y}(x,y)}{p_X(x)} \end{aligned}$$

# Variables Continuas

$X, Y$  v.a.c.

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?

# 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  $Y$

$$\begin{aligned} f_{X,Y}(x, y) &= \overbrace{f_{Y|X=x}(y)}^x \cdot \overbrace{f_X(x)}^y \\ &= \underbrace{f_{X|Y=y}(x)}_x \cdot \underbrace{f_Y(y)}_y \end{aligned} \quad \left\{ \begin{aligned} f_{X,Y}(x, y) &= f_X(x) f_Y(y) = \\ &= f_Y(y) \cdot f_X(x) \end{aligned} \right.$$

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

# Ejercicio 1

$X$  es una variable aleatoria discreta

$$Y \sim \text{Ber}(p) \quad \bullet \quad p = \frac{1}{5}$$

$$Z \sim \text{Bin}(n, p) \quad \bullet \quad n = 10$$

$$W \sim \text{Geo}(p)$$

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  $X|Y=y$ .

$$X: \Omega \rightarrow \mathbb{R}_{\geq 0}$$

$$\omega \mapsto X(\omega)$$

$$; \quad \mathbb{N} \subset \mathbb{R}_{\geq 0}$$

$$\begin{aligned} & \bullet \quad \Omega \in \mathcal{A} \\ & \bullet \quad A \in \mathcal{A} \Rightarrow A^c \in \mathcal{A} \\ & \bullet \quad \{A_i\}_{i=1}^{\infty} \in \mathcal{A} \Rightarrow \bigcup_{i=1}^{\infty} A_i \in \mathcal{A} \end{aligned}$$

$$X(\omega) = \begin{Bmatrix} 0 \\ 1 \\ 2 \\ \vdots \\ 10 \end{Bmatrix} \quad \left. \begin{array}{l} Y \sim \text{Ber}(\frac{1}{5}) \\ P(X|Y=y) \end{array} \right\}$$

Exp. Aleatorio

E E E ... E

E F ... F

F F ... F

$$X(\omega) = 10$$

$$X(\omega) = 1$$

$$X(\omega) = 0$$

$$\omega \in \Omega$$

# Ejercicio 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  $Y|X=x$ .

# Ejercicio 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  $X|Y=y$

# Mezcla de v.a.

Sea  $M$  una v.a. discreta a valores  $1, \dots, n$ , con función de probabilidad  $p_M(m)$ , y sea  $X$  una v.a. tal que se conocen las distribuciones  $X|M = m$ ,  $m = 1, 2, \dots, n$ . 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**  $f_X(x) = \sum_{m=1}^n f_{X|M=m}(x)p_M(m)$

# Bayes para mezclas

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

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



# 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 subte si un cierto día tardó exactamente 0.9 hs.

# Normal multivariada: dist. condicionales

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

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

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

# 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

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

**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 5

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

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  $Y|X=x$ .
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  $X|Y=y$

# 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  $t : R_X \rightarrow \mathbb{R}$ , 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 6

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

# 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**  $\underline{X}_n = (X_1, \dots, X_n)$ , tal que  $X_i \sim X$

# Estimador

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

**Props:** el **sesgo** o **bias** de un estimador  $\delta(\underline{X})$  de  $\theta$  se define como la diferencia  $\mathbb{E}\delta(\underline{X}) - \theta$ , 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

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

# Ejercicio 7

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

# 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

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

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

# Estimador de mínimos cuadrados

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

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

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

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  $\mathbb{E}[(X - (aY + b))^2 | Y]$  sea mínima.

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

# Ejercicio 8

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  $X$ .