Probabilidad y Estadística

Clase 3

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:

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

Función generadora de momentos

Definimos a la función generadora de momentos de X como

$$\frac{\partial m(t)}{\partial t} = \frac{d E(e^{tx})}{\partial t} = E\left(\frac{\partial e^{tx}}{\partial t}\right) = M_X = \mathbb{E}[e^{tX}] = m(t)$$

Teorema: Sean M_X y M_Y dos funciones generadoras de momentos de las v.a. X e Y, respectivamente. Si

$$M_X(t) = M_Y(t)$$

Para todo t, entonces X e Y tienen la misma distribución.

- $E(xe^{tx}) = \psi(t)$ $\psi(0) = E(x)$

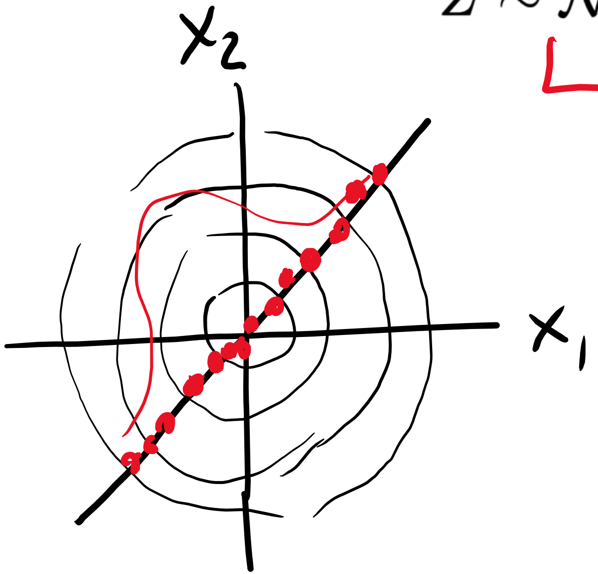
V.A Gaussianas: Proyección

Sea $X \sim \mathcal{N}(\mu, \Sigma)$ de dimensión n. Y sea $w \in \mathbb{R}^n$.
Definamos $Z = w^T X$ la proyección de X en w

$X \sim N(\mu, \Sigma)$

$X \in \mathbb{R}^2$

$$Z \sim \mathcal{N}(w^T \mu, w^T \Sigma w)$$



$w^T X = Z$

$w = \begin{bmatrix} w_1 \\ w_2 \end{bmatrix}$

$\mu = \begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix}$

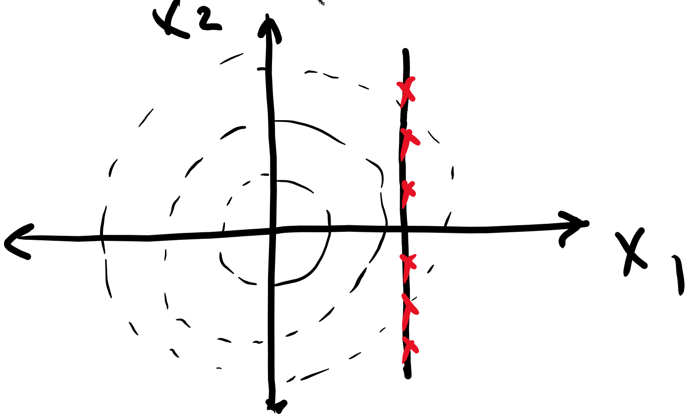
$\Sigma = \begin{bmatrix} \sigma_1^2 & \mathrm{Cov}\, x_1 x_2 \\ \mathrm{Cov}\, x_2 x_1 & \sigma_2^2 \end{bmatrix}$

Normal multivariada: dist. condicionales

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

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

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



$X_2 | X_1 = x_1$

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

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

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

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

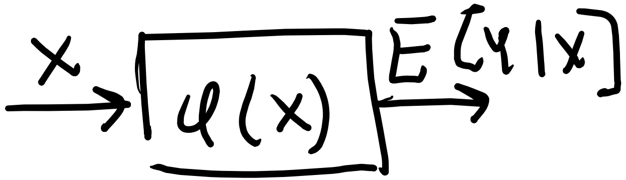
$$f_{X,Y}(x,y) = \frac{e^{-x/2y}}{4y}\mathbf{1}\{0 < x, 1 < y < 3\}$$

Hallar la función de densidad de X|Y=y

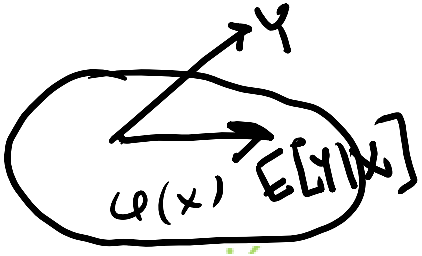
Esperanza condicional

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 $t : R_X \to \mathbb{R}$, tal que $\mathbb{E}[t(X)] < \infty$.



¿A qué nos recuerda esto?

$\mathbb{E}[Y|X]$ 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, calcular la esperanza condicional de X|Y o Y|X según corresponda.
2. Calcular E(X) para el ultimo ejemplo, usando propiedades de la esperanza.

$$E[X] = E[\underbrace{E[X|Y]}] = E[2Y]$$

$$= 2E[Y] = 2.2 = 4$$

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

sucesión de v.a $(X_1 X_2 X_3 \dots X_n) \overset{iid}{\sim} F$ $E(X) = \mu$

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

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

$$\bar{X}_n = \frac{\sum^n X_i}{n}$$

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

$$\bar{X}_n \longrightarrow \mu$$

Teorema central del límite

$$(X_1, X_2, \ldots\ldots X_n) \overset{iid}{\sim} F$$

$$\sum^{n} X_i \overset{(a)}{\approx} N(n\mu, n\sigma^2)$$ si n es suficientem. grande.

$$\frac{\sum^{n} X_i - n\mu}{\sqrt{n\sigma^2}} \longrightarrow N(0,1)$$

Teorema central del límite

$$(X_1, X_2 \ldots\ldots X_n) \overset{iid}{\sim} F$$

$$\bar{X} = \frac{\sum X}{n} \overset{(a)}{\sim} N\left(\mu ; \frac{\sigma^2}{n}\right)$$

para n suf. grande

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

Simulación

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.

$$(x_1, x_2, x_3 \ldots\ldots x_n) \overset{iid}{\sim} F$$

muestra

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 $X_i \sim X$

$\rightarrow (X_1 X_2 \ldots X_n) \overset{iid}{\sim} F$ $\boxed{\text{MUESTRA}}$

$\boxed{\text{DATOS}}$ $\rightarrow$ REALIZACIONES

$(x_1 x_2 x_3 \ldots x_n)$

Estimador

#MUESTRA #DATOS

#ESTIMADOR #ESTIMACIÓN

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.

$\theta \rightarrow$ MAGNITUD DESCONOCIDA

$\hat{\theta} = \delta(X_1, X_2 \ldots X_n) \rightarrow$ ESTIMADOR ES UNA VARIABLE ALEATORIA

$= \delta(\underline{X})$

$(X_1, X_2 \ldots\ldots X_n)$ X_i = consumo eléctrico i-ésimo usuario

$\theta = \mu$ $\hat{\theta} = \delta(X_1 X_2 \ldots X_n) = \bar{X} = \frac{\sum X_i}{n} \rightarrow$ variable aleatoria

Error cuadrático medio

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

$\hat{\theta}$ θ

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

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.

$X \sim N(5,9)$ $\qquad (X_1, X_2, \dots, X_n) \overset{iid}{\sim} N(5,9)$

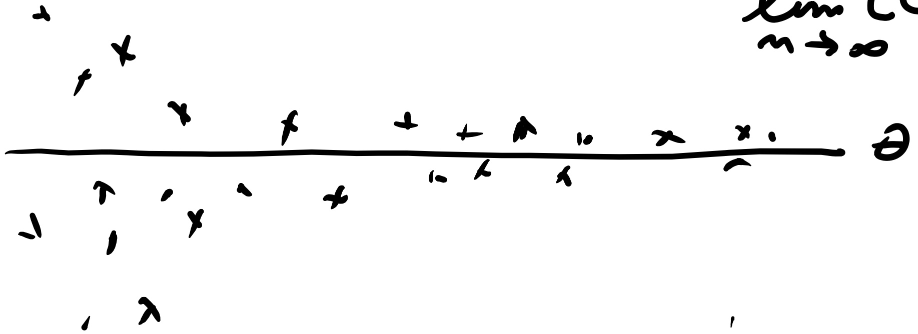
$\bar{X} = \frac{\sum X}{n} = \hat{\mu} \qquad E(\bar{X}) = 5$

$ECM(\bar{X}) = E[(\bar{X} - 5)^2] = V(\bar{X}) = V\left(\frac{\sum x_i}{n}\right)$

$= \frac{V(\sum x_i)}{n^2} = \frac{\sum V(x_i)}{n^2} = \frac{\cancel{n}\sigma^2}{n^{\cancel{2}}} = \frac{\sigma^2}{n} \qquad \underbrace{\lim_{n \to \infty} V(\bar{X}) = 0}$

$$\lim_{n\to\infty} ECM(\hat{\theta}) = 0$$

ESTIMADOR INSESGADO CONSISTENTE



$$\lim_{n\to\infty} ECM(\hat{\theta}) \neq 0$$

ESTIMADOR SESGADO CONSISTENTE

Estimadores de mínimos cuadrados

Estimador de mínimos cuadrados

$\hat{Y} \rightarrow$ PRONÓSTICO de Y

$\hat{Y} = g(X)$ función de X

$ECM(\hat{Y}, Y) = E[(\hat{Y} - Y)^2]$ SE PUEDE PROBAR QUE

EL MEJOR PREDICTOR EN TÉRMINOS DEL ECM

ES $\underbrace{E[Y|X]}$ LA ESPERANZA CONDICIONAL

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

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.

Estimador de mínimos cuadrados

Si me restrinjo a $g(x) = aX + b = \hat{Y}$

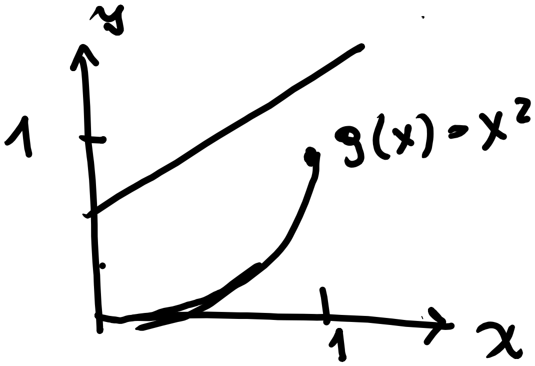
y minimizo $E[(Y - (ax+b))^2]$

obtengo $a = \dfrac{cov(X,Y)}{var(X)}$

$$b = -\frac{cov(X,Y)}{V(X)} \cdot E(X) + E(Y)$$

Ejercicio 4

$X \sim U[0,1]$ $\underbrace{Y = X^2}$



$\hat{Y} = X^2$

BUSCAMOS ESPERANZA CONDICIONAL

$E[\underbrace{Y|X=x}] = E(x^2) = x^2$

QUÉ PASA SI YO ME RESTRINJO a UNA RECTA

$\hat{Y} = (aX+b) =$

Ejercicio 4

C. AUX = $Y = X^2$

$$E(X) = 1/2$$

$$V(X) = \frac{(1-0)^2}{12} = 1/12$$

$$E(Y) = E(X^2) =$$

$$= V(X) + \underbrace{E(X)^2}$$

$$= 1/12 + 1/4 = 1/3$$

$$a = \frac{\operatorname{cov}(X,Y)}{V(X)}$$

$$b = \frac{\operatorname{cov}(X,Y)}{V(X)} \cdot E(X) + E(Y)$$

$$\operatorname{Cov}(X,Y) = E(X \cdot Y) - E(X) \cdot E(Y) = E(X^3) - 1/2 \cdot 1/3$$

$$= \int_0^1 x^3 \, dx - 1/6 = 1/12$$

Regresión lineal

$(x_1\ y_2 \ldots x_n)$ $(y_1 \ldots y_n)$

Tengo las observaciones $(x_1, \ldots, x_n)$ e $(y_1, \ldots, y_n)$, observaciones de dos v.a. X e Y, y queremos hallar la mejor relación lineal $\underbrace{Y = aX + b}$.

$Y = X^T \beta \rightarrow \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix}$

Definiendo $A = \begin{bmatrix} x_1 & 1 \\ \vdots & \vdots \\ x_n & 1 \end{bmatrix}$, tenemos que

$[a, b]^T = \underbrace{(A^T A)^{-1} A^T}_{\hat{\beta}} y$ Nuevamente, la demostración la vimos en AM.