Estimación LMS y LLMS

Clase 10: *least mean squares* y *linear least mean squares estimation*

6.431x — Probability: The Science of Uncertainty and Data

(versión de clase: las soluciones se comparten al terminar)

B.Sc. Cristian Lazo Quispe

BREIT (Brescia Institute of Technology) · Aporta — Grupo Breca

Programa SDS (Statistics and Data Science, MITx) · **BREIT** (Brescia Institute of Technology), iniciativa de Aporta —Grupo Breca— en alianza con MIT IDSS.
Basado en MIT 6.431x / MIT OCW RES.6-012 (CC BY-NC-SA 4.0).

clazoq@uni.pe · linkedin.com/in/cristian-lazo-quispe · github.com/CristianLazoQuispe

Dónde quedamos (Clase 9)

Inferencia bayesiana, en tres líneas

El desconocido es una v.a. Θ con **prior**; al observar $X = x$, Bayes da la **posterior** $f_{\Theta \mid X} \propto f_{\Theta}\, f_{X \mid \Theta}$. Para dar un número: el **MAP** (el pico de la posterior) o la **media posterior** $\mathbb{E}[\Theta \mid X = x]$. Moneda con prior uniforme: posterior Beta con media $\frac{k+1}{n+2}$. Modelo normal $X = \Theta + W$: posterior normal, media ponderada por precisiones.

Lo de hoy: dos preguntas que la Clase 9 dejó abiertas

- ¿En qué sentido es **óptima** la media posterior? Minimiza el error cuadrático medio: es el estimador **LMS**. ¿Cuánto error comete, antes y después de medir?
- ¿Y si la posterior es difícil de calcular, o no conocemos el modelo completo? Nos quedamos con la **mejor recta** $aX + b$: el estimador **LLMS**, que solo necesita medias, varianzas y covarianzas.

Cómo trabajaremos hoy

Primero la teoría de cada sección con ejemplos simples; al final, **cinco problemas difíciles** (dos de los seis problemas del examen final salen de esta clase): el enunciado (lo desarrollamos juntos en la pizarra) y después la solución escrita paso a paso.

¿Qué veremos hoy? (Unit 7 — Lectures 16 y 17)

1. El estimador LMS: la media condicional
2. La geometría del error
3. El estimador LLMS: la mejor recta
4. LMS frente a LLMS
5. Varias observaciones y la representación de los datos
6. Cinco problemas para resolver juntos
7. Cierre

Contenido

1. El estimador LMS: la media condicional
2. La geometría del error
3. El estimador LLMS: la mejor recta
4. LMS frente a LLMS
5. Varias observaciones y la representación de los datos
6. Cinco problemas para resolver juntos
7. Cierre

LMS: minimizar el error cuadrático medio

Sin observaciones

Estimar Θ con una constante $\hat{\theta}$. El MSE es

$$\mathbb{E}[(\Theta - \hat{\theta})^2] = \mathrm{Var}(\Theta) + (\mathbb{E}[\Theta] - \hat{\theta})^2,$$

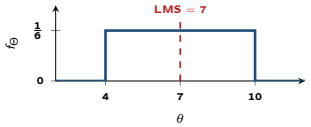
mínimo en $\hat{\theta} = \mathbb{E}[\Theta]$, con MSE $= \mathrm{Var}(\Theta)$. (Derivar: $-2\mathbb{E}\Theta + 2\hat{\theta} = 0$.)

Con una observación $X = x$

En el "universo" $X = x$ se repite el argumento:

$$\hat{\theta} = \mathbb{E}[\Theta \mid X = x] \quad \text{minimiza} \quad \mathbb{E}[(\Theta - \hat{\theta})^2 \mid X = x].$$

El estimador $\widehat{\Theta}_{\mathrm{LMS}} = \mathbb{E}[\Theta \mid X]$ minimiza $\mathbb{E}[(\Theta - g(X))^2]$ entre **todos** los estimadores $g(X)$.



El prior de la Lecture 16, $\Theta \sim \mathrm{Unif}[4, 10]$: **cualquier** punto de $[4, 10]$ es un MAP; el LMS es único, $\hat{\theta} = 7$, con MSE $= \frac{6^2}{12} = 3$. Estimar con 6 cuesta $3 + 1 = 4$.

Cuánto se equivoca el LMS

- **MSE condicional** (ya con el dato): $\mathrm{Var}(\Theta \mid X = x)$.
- **MSE global** (el diseño, antes de medir): $\mathbb{E}\big[\mathrm{Var}(\Theta \mid X)\big]$.

Error clásico

Usar el MAP cuando se pide minimizar el MSE. Coinciden solo si la posterior es unimodal y **simétrica** (como la normal del modelo lineal).

Ejemplo: el LMS a partir de una tabla

PMF conjunta de (Θ, X)

	$x=0$	$x=1$	$x=2$
$\theta=0$	0,2	0,2	0,1
$\theta=1$	0,1	0,2	0,2
p_X	0,3	0,4	0,3

El estimador, columna a columna

Θ es 0 o 1, así que $\mathbb{E}[\Theta \mid X=x] = \mathbb{P}(\Theta=1 \mid X=x)$:

$$\mathbb{E}[\Theta \mid X] = \tfrac{1}{3},\ \tfrac{1}{2},\ \tfrac{2}{3} \quad \text{si } X=0,1,2.$$

MSE condicional (varianza Bernoulli $p(1-p)$): $\frac{2}{9},\ \frac{1}{4},\ \frac{2}{9}$.

MSE global

$$0{,}3\cdot\tfrac{2}{9} + 0{,}4\cdot\tfrac{1}{4} + 0{,}3\cdot\tfrac{2}{9} = \tfrac{7}{30} \approx 0{,}233.$$

Sin medir, el MSE sería $\mathrm{Var}(\Theta) = \frac{1}{4}$: este X informa poco.

Dos lecturas

- El LMS da $\frac{1}{3}$, $\frac{1}{2}$ o $\frac{2}{3}$: valores que Θ **nunca** toma. No importa: minimiza el MSE.
- El MAP daría 0, empate o 1: otra regla, otro objetivo (probabilidad de error).

Contenido

1. El estimador LMS: la media condicional
2. La geometría del error
3. El estimador LLMS: la mejor recta
4. LMS frente a LLMS
5. Varias observaciones y la representación de los datos
6. Cinco problemas para resolver juntos
7. Cierre

El error del LMS es ortogonal a todo lo que sabemos

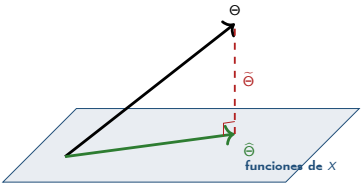
Propiedades ($\widehat{\Theta} = \mathbb{E}[\Theta \mid X]$, error $\widetilde{\Theta} = \widehat{\Theta} - \Theta$)

1. $\mathbb{E}[\widetilde{\Theta} \mid X = x] = 0$ y $\mathbb{E}[\widetilde{\Theta}] = 0$: sin sesgo.
2. $\mathbb{E}[\widetilde{\Theta}\, h(X)] = 0$ para **toda** función h; en particular $\mathrm{Cov}(\widetilde{\Theta}, \widehat{\Theta}) = 0$.
3. $\mathrm{Var}(\Theta) = \mathrm{Var}(\widehat{\Theta}) + \mathrm{Var}(\widetilde{\Theta})$.

La 2 en una línea: $\mathbb{E}[\widetilde{\Theta} h(X) \mid X] = h(X)\,\mathbb{E}[\widetilde{\Theta} \mid X] = 0$.

Es la varianza total de la Clase 8

$\mathrm{Var}(\widehat{\Theta}) = \mathrm{Var}(\mathbb{E}[\Theta \mid X])$ y $\mathrm{Var}(\widetilde{\Theta}) = \mathbb{E}[\mathrm{Var}(\Theta \mid X)]$. En la tabla: $\frac{1}{60} + \frac{7}{30} = \frac{15}{60} = \frac{1}{4} = \mathrm{Var}(\Theta)$. ✓



Con el "producto interno" $\mathbb{E}[UV]$, el LMS es la **proyección ortogonal** de Θ sobre todas las funciones de X; la propiedad 3 es Pitágoras.

Las dificultades del LMS

Necesita la posterior completa, $f_{\Theta \mid X} = f_\Theta f_{X \mid \Theta} / f_X$: (i) a veces no conocemos $f_{X \mid \Theta}$; (ii) con muchos desconocidos, $\mathbb{E}[\Theta_j \mid X]$ es una integral múltiple difícil. (Con varios Θ_j, se estima cada uno por separado.) $\Rightarrow$ **LLMS**.

MAP y LMS sobre una misma posterior asimétrica

Una moneda, un lanzamiento

$\Theta \sim \text{Unif}[0,1]$ es el sesgo; sale cara ($X = 1$). Posterior: $f_{\Theta \mid X}(\theta \mid 1) = 2\theta$.

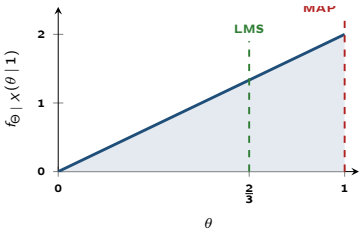
- MAP: el pico, $\hat{\theta} = 1$.
- LMS: la media, $\hat{\theta} = \frac{2}{3}$.

MSE global de cada regla

LMS: $\frac{1}{18} \approx 0{,}056$. MAP ($\widehat{\Theta} = X$): $\mathbb{E}[\Theta(1-\Theta)] = \frac{1}{6} \approx 0{,}167$, **el triple**.

Lectura

El MAP responde "¿cuál es el valor más probable?"; el LMS, "¿qué número me deja más cerca en promedio?". Con posteriores asimétricas, las respuestas se separan.



La posterior 2θ crece hasta el borde: el MAP queda en el extremo y la media, más adentro.

Contenido

1. El estimador LMS: la media condicional
2. La geometría del error
3. El estimador LLMS: la mejor recta
4. LMS frente a LLMS
5. Varias observaciones y la representación de los datos
6. Cinco problemas para resolver juntos
7. Cierre

LLMS: minimizar el MSE entre las rectas $aX + b$

La solución (Lecture 17)

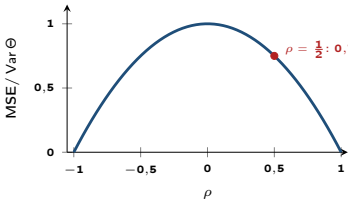
Con a fijo, el mejor b es $\mathbb{E}[\Theta] - a\mathbb{E}[X]$. Queda minimizar $\mathrm{Var}(\Theta - aX) = \mathrm{Var}\,\Theta + a^2\,\mathrm{Var}\,X - 2a\,\mathrm{Cov}(\Theta, X)$, y $a = \frac{\mathrm{Cov}(\Theta,X)}{\mathrm{Var}\,X}$:

$$\widehat{\Theta}_L = \mathbb{E}[\Theta] + \frac{\mathrm{Cov}(\Theta, X)}{\mathrm{Var}(X)}(X - \mathbb{E}[X]),$$

$$\mathbb{E}\big[(\widehat{\Theta}_L - \Theta)^2\big] = (1 - \rho^2)\,\mathrm{Var}(\Theta).$$

Lo que dice la fórmula

- **Solo importan medias, varianzas y covarianzas.**
- $\rho > 0$: X sobre su media $\Rightarrow$ estimación sobre $\mathbb{E}\Theta$.
- $\rho = 0$: $\widehat{\Theta}_L = \mathbb{E}\Theta$. $|\rho| = 1$: error cero.
- Ortogonalidad: $\mathbb{E}[\widetilde{\Theta}_L] = 0$ y $\mathrm{Cov}(\widetilde{\Theta}_L, X) = 0$ (no con X^2).



La parábola $1 - \rho^2$. ρ^2 es la fracción de $\mathrm{Var}\,\Theta$ que la recta explica (el R^2 de la regresión). Con $\rho = \frac{1}{2}$ solo se quita un cuarto del error.

Error clásico

Dividir entre $\mathrm{Var}(\Theta)$ en lugar de $\mathrm{Var}(X)$: la pendiente es Cov / Var(lo observado).

Ejemplo: la fórmula LLMS con números

Datos

$\mathbb{E}\Theta = 2$, $\mathbb{E}X = 5$, $\operatorname{Var}\Theta = 4$, $\operatorname{Var}X = 9$, $\operatorname{Cov}(\Theta, X) = 3$.

El estimador

$$a = \tfrac{3}{9} = \tfrac{1}{3}, \qquad \widehat{\Theta}_L = 2 + \tfrac{1}{3}(X - 5).$$

Si $X = 8$: $\hat{\theta} = 3$. Si $X = 5$ (su media): $\hat{\theta} = 2 = \mathbb{E}\Theta$.

Cuánto ayuda

$\rho = \frac{3}{2\cdot 3} = \frac{1}{2}$ y $\text{MSE} = \frac{3}{4}\cdot 4 = 3$, frente a 4 sin observar.

Sin término constante

Si se fuerza $b = 0$: $a = \frac{\mathbb{E}[\Theta X]}{\mathbb{E}[X^2]}$ con momentos **no centrados**: $\mathbb{E}X^2 = 34$, $\mathbb{E}[\Theta X] = 3 + 10 = 13$, $a = \frac{13}{34}$, $\text{MSE} = 8 - \frac{169}{34} \approx 3{,}03 > 3$. Una familia más chica de rectas nunca gana.

Cómo se obtienen los momentos en la práctica

Casi siempre condicionando en Θ (Clase 8):
$\mathbb{E}[X] = \mathbb{E}[\mathbb{E}[X \mid \Theta]]$,
$\operatorname{Var}X = \mathbb{E}[\operatorname{Var}(X \mid \Theta)] + \operatorname{Var}(\mathbb{E}[X \mid \Theta])$,
$\mathbb{E}[\Theta X] = \mathbb{E}[\Theta\, \mathbb{E}[X \mid \Theta]]$.

Contenido

1. El estimador LMS: la media condicional
2. La geometría del error
3. El estimador LLMS: la mejor recta
4. LMS frente a LLMS
5. Varias observaciones y la representación de los datos
6. Cinco problemas para resolver juntos
7. Cierre

Cuándo coinciden: $\mathbb{E}[\Theta \mid X]$ lineal

Regla

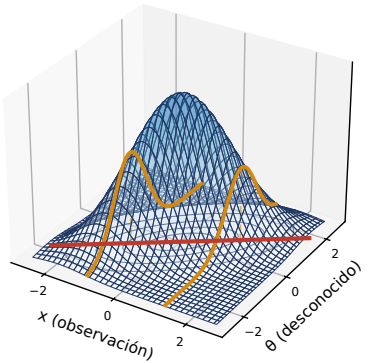
$\widehat{\Theta}_{\text{LMS}} = \widehat{\Theta}_{\text{LLMS}}$ si y solo si $\mathbb{E}[\Theta \mid X]$ es lineal. Siempre $\text{MSE}_{\text{LMS}} \le \text{MSE}_{\text{LLMS}}$.

La moneda (Lecture 17)

$\Theta \sim \text{Unif}[0,1]$, $X \sim \text{Bin}(n, \Theta)$. Condicionando en Θ: $\mathbb{E}X = \frac{n}{2}$, $\text{Var}\, X = \frac{n(n+2)}{12}$, $\text{Cov}(\Theta, X) = \frac{n}{3} - \frac{n}{4} = \frac{n}{12}$.

$$\widehat{\Theta}_L = \frac{1}{2} + \frac{X - n/2}{n+2} = \frac{X+1}{n+2} = \widehat{\Theta}_{\text{LMS}},$$

la media de la posterior Beta. $\text{MSE} = \frac{1}{6(n+2)}$.



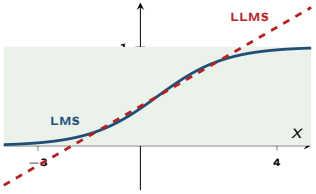
Normal conjunta con $\rho = 0{,}7$: el pico de cada rebanada $X = x$ (ámbar) cae sobre la recta roja $\mathbb{E}[\Theta \mid X = x] = 0{,}7x$. En el caso normal el LMS es una recta.

Cuándo no: la recta no conoce el rango de Θ

Un bit entre ruido

$\Theta \in \{0,1\}$ equiprobable, $X = \Theta + W$, $W \sim \mathcal{N}(0,1)$.

- LLMS: $\mathrm{Cov} = \frac{1}{4}$, $\mathrm{Var}\, X = \frac{5}{4}$: $\widehat{\Theta}_L = \frac{1}{2} + \frac{1}{5}(X - \frac{1}{2})$. En $X = 4$ da 1,2; en $X = -3$, $-0{,}2$.
- LMS: $\mathbb{P}(\Theta = 1 \,|\, X = x) = \frac{1}{1+e^{-(x-\mathbf{1/2})}}$, siempre en $(0,1)$; en $X = 4$, 0,971.



Franja verde: el rango de Θ. La sigmoide (LMS) nunca sale; la recta sí.

Otro: $\Theta = X^2$, $X \sim \mathrm{Unif}[0,1]$

LLMS: $X - \frac{1}{6}$, que da $-\frac{1}{6}$ en $X = 0$. Su error es incorrelado con X pero no con X^2 ($\mathrm{Cov} = \frac{1}{180}$). El LMS es X^2, sin error.

Lectura

Aun así el MSE del LLMS (0,2) es casi el del LMS: la recta se equivoca feo solo en datos raros.

Contenido

1. El estimador LMS: la media condicional
2. La geometría del error
3. El estimador LLMS: la mejor recta
4. LMS frente a LLMS
5. Varias observaciones y la representación de los datos
6. Cinco problemas para resolver juntos
7. Cierre

LLMS con varias observaciones: la precisión manda

Planteamiento

Minimizar $\mathbb{E}[(a_1X_1 + \cdots + a_nX_n + b - \Theta)^2]$: igualar derivadas a cero da un **sistema lineal** (el error con media 0 e incorrelado con cada X_j). Solo importan medias, varianzas y covarianzas. Con varios Θ_j, uno por uno.

El caso más simple (Lecture 17)

$X_i = \Theta + W_i$, con Θ, W_i incorreladas; $\mathbb{E}\Theta = x_0$, $\operatorname{Var}\Theta = \sigma_0^2$, $\operatorname{Var} W_i = \sigma_i^2$:

$$\widehat{\Theta}_L = \frac{x_0/\sigma_0^2 + \sum_i X_i/\sigma_i^2}{\sum_{i=0}^{n} 1/\sigma_i^2}, \quad \text{MSE} = \frac{1}{\sum_{i=0}^{n} \sigma_i^{-2}}.$$

Dos mediciones

$\mathbb{E}\Theta = 10$, $\operatorname{Var}\Theta = 4$; $X_1 = 12$ ($\operatorname{Var} W_1 = 1$) y $X_2 = 8$ ($\operatorname{Var} W_2 = 4$). Precisiones $\frac{1}{4}$, 1, $\frac{1}{4}$:

$$\hat{\theta} = \frac{2{,}5 + 12 + 2}{1{,}5} = 11, \qquad \text{MSE} = \tfrac{2}{3}.$$

Es la fórmula del modelo normal de la Clase 9

Pero ahora **sin suponer normalidad**: con las mismas medias y covarianzas, el LLMS no distingue una distribución de otra, y en el caso normal el LMS ya era esta recta.

La representación de los datos importa en el LLMS

LMS: da igual

$\mathbb{E}[\Theta \mid X] = \mathbb{E}[\Theta \mid X^3]$: X y X^3 contienen la misma información.

LLMS: no da igual

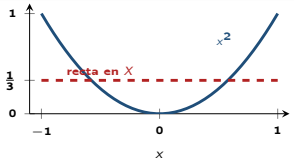
$aX + b$ y $aX^3 + b$ son familias distintas, con $\mathrm{Cov}(\Theta, X^3)$ y $\mathrm{Var}(X^3)$ distintas. Se pueden combinar:
$a_1X + a_2X^2 + a_3X^3 + b$, o $a_1X + a_2e^X + a_3 \ln X + b$.

Lectura

Elegir qué funciones de los datos entran en la recta es parte del **diseño** del estimador (como elegir las variables de una regresión).

Ejemplo: $X \sim \mathrm{Unif}[-1,1]$, $\Theta = X^2 + W$

- Con X: $\mathrm{Cov}(X^2, X) = \mathbb{E}[X^3] = 0$ por simetría $\Rightarrow$ recta horizontal $\frac{1}{3}$, MSE $\frac{4}{45} + \sigma^2$.
- Con $Z = X^2$: $\widehat{\Theta} = Z$, MSE σ^2.



Contenido

1. El estimador LMS: la media condicional
2. La geometría del error
3. El estimador LLMS: la mejor recta
4. LMS frente a LLMS
5. Varias observaciones y la representación de los datos
6. Cinco problemas para resolver juntos
7. Cierre

Cómo trabajaremos los cinco problemas

El formato

Cada problema ocupa un slide casi en blanco: lo **desarrollamos juntos en la pizarra**. La solución escrita, paso a paso, se comparte al **final de la clase** (versión completa de estos slides en OneDrive): hoy, atención a la pizarra. En la hoja de ejercicios son los problemas 19, 9, 10, 20 y 21, en versión completa.

Los cinco problemas y la sección de teoría que usan

1. El paralelogramo de la Lecture 16: **LMS por tramos, MSE condicional y global, y la recta** (secciones 1, 3 y 4).
2. ¿De qué tipo son las bombillas?: **MAP con umbral y LMS de una predicción** (secciones 1 y 2).
3. El parámetro de una geométrica: **posterior, LMS y MAP**, y por qué no hay LLMS (secciones 1 y 3).
4. LLMS con sumas aleatorias: **momentos por esperanzas iteradas** (sección 3).
5. El parámetro de una uniforme: **MAP, LMS y LLMS sobre la misma posterior** (secciones 1, 3 y 4).

Antes de cada uno

Pregúntate: ¿me piden el MAP, el LMS o el LLMS? Para el LMS hace falta la **posterior**; para el LLMS, solo **cinco números**: $\mathbb{E}\Theta$, $\mathbb{E}X$, $\mathrm{Var}\,X$, $\mathrm{Cov}(\Theta, X)$ y, para el MSE, $\mathrm{Var}\,\Theta$.

Problema 1: el paralelogramo de la Lecture 16

Adaptado del ejemplo de las Lectures 16 y 17 (MIT 6.431x). $\Theta \sim \text{Unif}[4, 10]$ y $X = \Theta + U$, con $U \sim \text{Unif}[-1, 1]$ independiente de Θ.

a Dibuja el soporte de (X, Θ). Halla el estimador LMS $\mathbb{E}[\Theta \mid X = x]$ por tramos, el MSE condicional $\text{Var}(\Theta \mid X = x)$ y el MSE global.

b Halla el estimador LLMS y su MSE. Compara con (a) y evalúa ambos en $X = 4$.

Problema 2: ¿de qué tipo son las bombillas?

Adaptado de "Determining the type of a lightbulb" (Problem Set 7b, MIT 6.431x). Una bombilla tipo A dura $\mathrm{Exp}(\lambda)$; una tipo B, $\mathrm{Exp}(\mu)$, con $\mu > \lambda > 0$. Una caja trae un solo tipo; a priori es B con probabilidad $\frac{1}{4}$. Dado el tipo, las duraciones son independientes. Se prueba una bombilla: dura $T_1 = t_1$.

a. Si $\mu \geq 3\lambda$, la regla MAP decide "A" si y solo si $t_1 \geq \alpha$. Halla α.

b. Con $\lambda = 3$ y $\mu = 4$ se observa $T_1 = 2$. Halla el estimador LMS de la duración T_2 de otra bombilla de la caja.

Problema 3: el parámetro de una geométrica

Adaptado de "Estimating the parameter of a geometric r.v." (Problem Set 7b, MIT 6.431x). k monedas tienen la misma probabilidad de cara, que es el valor de $Q \sim \text{Unif}[0,1]$; dado $Q = q$, todos los lanzamientos son independientes. T_i es el número de lanzamientos de la moneda i hasta su primera cara (incluida). Puede servir $\int_0^1 q^a(1-q)^b\,dq = \frac{a!\,b!}{(a+b+1)!}$.

a. Halla la PMF de T_1 y el estimador LMS de Q basado en $T_1 = t$.

b. Con las k monedas, halla el MAP de Q dados $T_1 = t_1, \ldots, T_k = t_k$. ¿Se puede usar LLMS con T_1?

Problema 4: LLMS con sumas aleatorias

Adaptado de "LLMS estimation with random sums" (Problem Set 7b, MIT 6.431x). N tiene media m y varianza v. $A_1, A_2, \ldots$ y $B_1, B_2, \ldots$ son i.i.d. con media 1 y varianza 1, independientes entre sí y de N. Sean $A = \sum_{i=1}^{N} A_i$ y $B = \sum_{i=1}^{N} B_i$.

a. Con esperanzas iteradas, calcula $\mathbb{E}[AB]$ y $\mathbb{E}[NA]$ en términos de m y v.

b. Halla el estimador LLMS $\widehat{N} = c_1 A + c_2$ de N basado en A.

Problema 5: el parámetro de una uniforme

Adaptado de "Estimating the parameter of a uniform r.v." (Problem Set 7b, MIT 6.431x). X es uniforme en $[\theta, 2\theta]$, donde θ es el valor de $\Theta \sim \text{Unif}[0, 1]$.

a. Para $0 \leq x \leq 1$, halla la posterior $f_{\Theta \mid X}(\theta \mid x)$, el MAP y el LMS.

b. Halla el LLMS $c_1 + c_2 X$ y compara las tres estimaciones en $x = 1$.

Contenido

1. El estimador LMS: la media condicional
2. La geometría del error
3. El estimador LLMS: la mejor recta
4. LMS frente a LLMS
5. Varias observaciones y la representación de los datos
6. Cinco problemas para resolver juntos
7. Cierre

Resumen de la Clase 10

Lo que hay que saber hacer

1. **LMS**: $\widehat{\Theta} = \mathbb{E}[\Theta \mid X]$ minimiza el MSE; sin datos, $\mathbb{E}\Theta$ con MSE $\operatorname{Var}\Theta$. MSE condicional $\operatorname{Var}(\Theta \mid X = x)$ y global $\mathbb{E}[\operatorname{Var}(\Theta \mid X)]$.
2. **El error LMS**: sin sesgo, incorrelado con toda función de X, y $\operatorname{Var}\Theta = \operatorname{Var}\widehat{\Theta} + \operatorname{Var}\widetilde{\Theta}$.
3. **LLMS**: $\mathbb{E}\Theta + \frac{\operatorname{Cov}(\Theta, X)}{\operatorname{Var} X}(X - \mathbb{E}X)$, MSE $(1 - \rho^2)\operatorname{Var}\Theta$; solo segundos momentos, que se calculan condicionando en Θ.
4. **LMS frente a LLMS**: coinciden si $\mathbb{E}[\Theta \mid X]$ es lineal (normal conjunta, moneda con prior uniforme); si no, el LMS gana y la recta puede salirse del rango de Θ.
5. **Varias observaciones**: ecuaciones normales; con ruidos incorrelados, promedio ponderado por precisiones. La representación $(X, X^2, \dots)$ cambia el LLMS, no el LMS.

Errores clásicos de esta unidad

Usar el MAP cuando se pide el MSE · confundir MSE condicional con global · dividir entre $\operatorname{Var}\Theta$ en la pendiente · tratar como independientes variables que solo lo son dado Θ (o N) · aplicar LLMS sin momentos finitos · creer que la recta respeta el rango de Θ.

Siguiente clase (Clase 11)

Unit 8: desigualdades de Markov y Chebyshev, ley de los grandes números y teorema central del límite.

¡Gracias!

Clase 10 — Estimación LMS y LLMS (6.431x, Unit 7)

clazoq@uni.pe · linkedin.com/in/cristian-lazo-quispe · github.com/CristianLazoQuispe

BREIT (Brescia Institute of Technology) · Aporta — Grupo Breca · en alianza con MIT IDSS

Basado en MIT 6.431x / MIT OCW RES.6-012 (Bertsekas & Tsitsiklis), CC BY-NC-SA 4.0.