Probabilidad II
Tema 7: Esperanza condicionada

José R. Berrendero

Departamento de Matemáticas
Universidad Autónoma de Madrid

Estructura de este tema

- Introducción y motivación
- Definición de esperanza condicionada
- Propiedades básicas
- La esperanza condicionada como proyección
- El concepto de martingala
- Teorema de Doob

Introducción

Objetivo

Dar una definición general, válida bajo condiciones mínimas, de la esperanza condicionada. Para obtener este grado de generalidad renunciaremos a definirla "punto a punto". La definiremos como una v.a. cuyos valores pueden alterarse en conjuntos de probabilidad cero y aún siguen cumpliendo las condiciones requeridas.

Utilidad del concepto de esperanza condicionada

(a) **Herramienta de cálculo:** $\mathrm{P}(A) = \int \mathrm{P}(A|X = x)dF_X(x)$.

(b) **Herramienta de predicción:** $\mathrm{E}(Y|X)$ es la mejor predicción (en el sentido del error cuadrático medio) de Y en términos de X.

(c) **Herramienta conceptual:** para definir y manejar algunas ideas de dependencia sumamente importantes (procesos markovianos, martingalas, etc.)

Motivación

- En el tema 1 definimos $\mathrm{P}(A|B)$ para el caso en que $\mathrm{P}(B) > 0$. Pero en muchos problemas es necesario trabajar con probabilidades condicionadas a sucesos con probabilidad cero.

- Sea X una v.a. uniforme en $[0,1]$. Supongamos que la v.a. Y tiene distribución normal de media X y varianza 1 y sea $A \in \mathcal{B}$.

- La probabilidad condicionada $\mathrm{P}(Y \in A|X = x)$ interviene en la fórmula de la probabilidad total:

$$\mathrm{P}(Y \in A) = \int_0^1 \mathrm{P}(Y \in A|X = x)dx,$$

- El análogo para valores esperados es:

$$\mathrm{E}(Y) = \int_0^1 \mathrm{E}(Y|X = x)dx = \int_{\mathbb{R}} \mathrm{E}(Y|X = x)dF_X(x).$$

Motivación

Supongamos que $X = I_B$, $B \in \mathcal{F}$ con $P(B) > 0$.

La σ-álgebra generada por X es $\sigma(X) = \{B, B^c, \Omega, \emptyset\}$.

Podemos contemplar $\mathrm{E}(Y|X)$ como una v.a. que toma dos posibles valores:

- $\mathrm{E}(Y|X)(\omega) = \mathrm{E}(Y|B)$, si $\omega \in B$.
- $\mathrm{E}(Y|X)(\omega) = \mathrm{E}(Y|B^c)$, si $\omega \notin B$.

$\mathrm{E}(Y|X)$ es medible respecto a $\sigma(X)$ (intuitivamente, es función de la información que proporciona la observación de X)

Motivación

Definimos $\mathrm{P}_B(A) = \mathrm{P}(A \cap B)/\mathrm{P}(B)$.

Intuitivamente, se debe cumplir

$$\mathrm{E}(Y|B) = \int_\Omega Y d\mathrm{P}_B \Leftrightarrow \mathrm{P}(B)\mathrm{E}(Y|B) = \int_B Y d\mathrm{P}.$$

o, equivalentemente,

$$\int_B \mathrm{E}(Y|X) d\mathrm{P} = \int_B \mathrm{E}(Y|B) dP = \int_B Y d\mathrm{P}.$$

Y análogamente para B^c,

$$\int_{B^c} \mathrm{E}(Y|X) d\mathrm{P} = \int_{B^c} \mathrm{E}(Y|B^c) d\mathrm{P} = \int_{B^c} Y d\mathrm{P}.$$

¿Qué ocurre si integramos $\mathrm{E}(Y|X)$ sobre $\emptyset$? ¿Y sobre Ω?

Definición de esperanza condicionada

Definición

Sea Y una v.a. definida sobre $(\Omega, \mathcal{F}, \mathrm{P})$ con $\mathrm{E}(|Y|) < \infty$. Sea $\mathcal{D}$ una sub-σ-álgebra de $\mathcal{F}$. La esperanza condicionada de Y dada $\mathcal{D}$ se define como una función $\mathrm{E}(Y|\mathcal{D}) : \Omega \to \mathbb{R}$, **medible respecto a** $\mathcal{D}$ y tal que, para todo $D \in \mathcal{D}$,

$$\int_D \mathrm{E}(Y|\mathcal{D}) d\mathrm{P} = \int_D Y d\mathrm{P}.$$

Observación

La función $\mathrm{E}(Y|\mathcal{D})$ existe y es única c.s. respecto a P ya que la medida signada finita $\lambda(D) = \int_D Y d\mathrm{P}$ es absolutamente continua respecto a P. Se aplica el teorema de Radon-Nikodym:

$$\lambda(D) = \int_D \frac{d\lambda}{d\mathrm{P}} d\mathrm{P}.$$

Ejercicio: Determina $\mathrm{E}(Y|\mathcal{D})$ si $\mathcal{D} = \{\Omega, \emptyset\}$.

Notación

Sean X e Y dos v.a. sobre $(\Omega, \mathcal{F}, \mathrm{P})$, con $\mathrm{E}(|Y|) < \infty$.

- $\mathrm{E}(Y|\sigma(X)) = \mathrm{E}(Y|X)$.
- $\mathrm{E}(Y|\sigma(X_t\; t \in T)) = \mathrm{E}(Y|X_t, t \in T)$, donde (recordamos)
$$\sigma(X_t,\, t \in T) = \sigma\left(\bigcup_{t \in T} \sigma(X_t)\right).$$
- Si $B \in \mathcal{F}$, $\mathrm{E}(I_B|\mathcal{D}) = \mathrm{P}(B|\mathcal{D})$.

Notación

- Si $\omega \in \Omega$ es tal que $X(\omega) = x$, entonces $\mathrm{E}(Y|X)(\omega) = \mathrm{E}(Y|X = x)$.
- Si $B \in \mathcal{B}(\mathbb{R})$,

$$\int_{\{X\in B\}} \mathrm{E}(Y|X)d\mathrm{P} = \int_{\{X\in B\}} Yd\mathrm{P}$$

  es equivalente (teorema de cambio de espacio de integración) a

$$\int_{B} \mathrm{E}(Y|X = x)dF_X(x) = \int_{\{X\in B\}} Yd\mathrm{P}.$$

- ¿Qué resulta si tomamos $B = \mathbb{R}$? ¿Y si además sustituimos Y por I_A?

Ejemplo

Sea (X, Y) un vector aleatorio con función de densidad conjunta $f(x, y)$.

Sea $f_1(x)$ la función de densidad marginal de X, $f_1(x) = \int_{-\infty}^{\infty} f(x, y)dy$.

Para x tal que $f_1(x) > 0$ definimos $h(y|x) = f(x, y)/f_1(x)$ (y 0 en caso contrario). Entonces,

$$E(Y|X = x) = \int_{-\infty}^{\infty} yh(y|x)dy.$$

$$\begin{aligned}\int_{\{X\in B\}} Yd\mathrm{P} &= \int\int_{\{(x,y):\, x\in B\}} yf(x, y)dxdy = \int_B f_1(x)\left[\int_{-\infty}^{\infty} yh(y|x)dy\right]dx \\ &= \int_B \left[\int_{-\infty}^{\infty} yh(y|x)dy\right] dF_X(x).\end{aligned}$$

Propiedades básicas de la esperanza condicionada

(a) Ley de la esperanza iterada: $\mathrm{E}(Y) = \mathrm{E}[\mathrm{E}(Y|\mathcal{D})]$.

(b) Si Y es medible respecto a $\mathcal{D}$, entonces $\mathrm{E}(Y|\mathcal{D}) = Y$ c.s.

(c) Linealidad: $\mathrm{E}(a_1 Y_1 + a_2 Y_2|\mathcal{D}) = a_1 \mathrm{E}(Y_1|\mathcal{D}) + a_2 \mathrm{E}(Y_2|\mathcal{D})$ c.s.

(d) Si $Y \geq 0$, entonces $\mathrm{E}(Y|\mathcal{D}) \geq 0$ c.s.

(e) TCM condicional: $0 \leq Y_n \uparrow Y$, entonces $\mathrm{E}(Y_n|\mathcal{D}) \uparrow \mathrm{E}(Y|\mathcal{D})$ c.s.

(f) TCD condicional: $|Y_n| \leq U$ para todo n, $\mathrm{E}U < \infty$ y $Y_n \to Y$ c.s., entonces $\mathrm{E}(Y_n|\mathcal{D}) \to \mathrm{E}(Y|\mathcal{D})$ c.s.

(g) Desigualdad de Jensen condicional: $g : \mathbb{R} \to \mathbb{R}$ convexa y $\mathrm{E}|g(Y)| < \infty$, entonces $\mathrm{E}(g(Y)|\mathcal{D}) \geq g(\mathrm{E}(Y|\mathcal{D}))$ c.s.

Propiedades básicas de la esperanza condicionada

(h) Si $\mathcal{C}$ es una sub-σ-álgebra de $\mathcal{D}$, entonces

$$\mathsf{E}(\mathsf{E}(Y|\mathcal{D})|\mathcal{C}) = \mathsf{E}(Y|\mathcal{C}) \text{ c.s.}$$

(i) Si X y XY son integrables, y X es medible respecto a $\mathcal{D}$, entonces $\mathsf{E}(XY|\mathcal{D}) = X\mathsf{E}(Y|\mathcal{D})$ c.s.

(j) Si $\mathcal{C}$ es independiente de $\sigma(\sigma(Y) \cup \mathcal{D})$, entonces $\mathsf{E}(Y|\sigma(\mathcal{D} \cup \mathcal{C})) = \mathsf{E}(Y|\mathcal{D})$ c.s.

(k) Si Y es independiente de $\mathcal{D}$, $\mathsf{E}(Y|\mathcal{D}) = \mathsf{E}(Y)$ c.s.

La esperanza condicionada como proyección

Proposición
Si $\mathrm{E}(Y^2) < \infty$, entonces $\mathrm{E}(Y|\mathcal{D})$ minimiza $\mathrm{E}[(Y-X)^2]$ en $L_2(\mathcal{D})$, el espacio de v.a. medibles respecto a $\mathcal{D}$ y de cuadrado integrable.

Demostración
Como $L_2(\mathcal{D})$ es un subespacio cerrado de $L_2(\Omega, \mathcal{F}, \mathrm{P})$, existe $X^* \in L_2(\mathcal{D})$ tal que

(a) $\mathrm{E}[(Y-X^*)^2] = \inf\{\mathrm{E}[(Y-X)^2] : \ X \in L_2(\mathcal{D})\}$.

(b) $\mathrm{E}[(Y-X^*)X] = 0$, para toda $X \in L_2(\mathcal{D})$.

Sea $D \in \mathcal{D}$. Si tomamos $X = I_D$ en (b),

$$\int_D X^* d\mathrm{P} = \int_D Y d\mathrm{P}.$$

Por lo tanto $X^* = \mathrm{E}(Y|\mathcal{D})$ c.s.

El concepto de martingala

Filtración

Sea $(\Omega, \mathcal{F}, \mathrm{P})$ un espacio de probabilidad. Una filtración $(\mathcal{F}_n : n \geq 0)$ es una sucesión creciente de sub-σ-álgebras de $\mathcal{F}$.

Sucesión adaptada a una filtración

Se dice que la sucesión de v.a. $(X_n : n \geq 0)$ definidas sobre $(\Omega, \mathcal{F}, \mathrm{P})$ está adaptada a la filtración $(\mathcal{F}_n)$ si para cada n, X_n es medible respecto a $\mathcal{F}_n$.

Martingala

La sucesión de v.a. $(X_n : n \geq 0)$ es una martingala [respecto a la filtración $(\mathcal{F}_n)$] si

(a) (X_n) está adaptada a $(\mathcal{F}_n)$.

(b) Para todo n, $\mathrm{E}|X_n| < \infty$.

(c) Para todo $n \geq 1$, $\mathrm{E}(X_n | \mathcal{F}_{n-1}) = X_{n-1}$ c.s.

Ejemplos

Sumas de v.a. independientes de media cero.
Sean $X_1, X_2, \ldots$ v.a. independientes con $\mathrm{E}(X_n) = 0$ para todo n. Definimos

- $\mathcal{F}_n = \sigma(X_1, \ldots, X_n)$, $\mathcal{F}_0 = \{\emptyset, \Omega\}$.
- $S_0 = 0$, $S_n = X_1 + \cdots + X_n$.

Entonces S_n es una martingala respecto a $\mathcal{F}_n$.

Productos de v.a. independientes no negativas de media uno.
Sean $X_1, X_2, \ldots$ v.a. independientes no negativas con $\mathrm{E}(X_n) = 1$ para todo n. Definimos

- $\mathcal{F}_n = \sigma(X_1, \ldots, X_n)$, $\mathcal{F}_0 = \{\emptyset, \Omega\}$.
- $M_0 = 1$, $M_n = X_1 \cdots X_n$.

Entonces M_n es una martingala respecto a $\mathcal{F}_n$.

Convergencia de submartingalas

Submartingalas y supermartingalas
La sucesión de v.a. $(X_n : n \geq 0)$ es una **submartingala** [respecto a la filtración $(\mathcal{F}_n)$] si

(a) (X_n) está adaptada a $(\mathcal{F}_n)$.

(b) Para todo n, $\mathrm{E}|X_n| < \infty$.

(c) Para todo $n \geq 1$, $\mathrm{E}(X_n|\mathcal{F}_{n-1}) \geq X_{n-1}$ c.s.

$(X_n : n \geq 0)$ es **supermartingala** si $\mathrm{E}(X_n|\mathcal{F}_{n-1}) \leq X_{n-1}$ c.s.

Teorema de Doob
Sea (X_n) una submartingala respecto a $(\mathcal{F}_n)$. Supongamos que $\sup_n \mathbb{E}|X_n| < \infty$. Entonces existe una v.a. X tal que $X_n \xrightarrow{c.s} X$. Además $\mathrm{E}|X| \leq \sup_n \mathrm{E}|X_n|$.