Esperanza Condicional (continuación)

Pablo L. De Nápoli

Departamento de Matemática
Facultad de Ciencias Exactas y Naturales
Universidad de Buenos Aires

Probabilidades y Estadística para Matemática
Segundo cuatrimestre de 2021

Parte I

La esperanza condicional como proyección ortogonal

Introducción

En la clase anterior, introdujimos el concepto de esperanza condicional de una variable aleatoria con respecto a otra.

Recordamos que la idea al definir $E[X/Y]$ es estimar X por medio de una función de Y. Hoy formalizaremos esta intuición usando las ideas que introdujimos en la clase 12.

El enfoque que vamos a desarrollar sólo funciona para variables aleatorias con segundo momento finito (mientras que como vimos en la clase anterior $E[X/Y]$ se puede definir en general con sólo asumir que $E[|X|] < \infty$).

Contexto abstracto en el que vamos a trabajar

Consideramos un espacio de probabilidad $(\Omega, \mathcal{E}, P)$. Consideramos el espacio vectorial de las variables aleatorias con segundo momento finito

$$L^2(\Omega) = \{\text{variables aleatorias } X : \Omega \to \overline{\mathbb{R}} : E(X^2) < \infty\}$$

Recordamos que si $X \in L^2(\Omega)$,

$$E(|X|) \leq E(X^2)^{1/2} \text{ por la desigualdad de Jensen}$$

y

$$\text{Var}(X) = E(X^2) - E(X)^2$$

Por lo que las variables aleatorias en L^2 tienen esperanza y varianza finitas. $L^2(\Omega)$ es un espacio normado con la norma

$$\|X\| = E(X^2)^{1/2}$$

que proviene del producto interno

$$\langle X, Y \rangle = E(X \cdot Y)$$

Como ya mencionamos varias veces, $L^2(\Omega)$ es un espacio de Hilbert.

Contexto abstracto en el que vamos a trabajar (2)

Para que $L^2(\Omega)$ sea realmente un espacio vectorial normado, hay considerar iguales a las variables aleatorias X e Y tales que

$$P\{X = Y\} = 1$$

Con esta convención,

$$\|X\| = 0 \Rightarrow X = 0$$

Planteo del problema

En la clase 12, planteamos el problema de aproximar una variable $Y \in L^2(\Omega)$ por un elemento $\widehat{Y}$ de un subespacio S, de modo de minimizar el error cuarático medio

$$\mathrm{ECM}(Y, \widehat{Y}) = E(|Y - \widehat{Y}|^2) = \|Y - \widehat{Y}\|^2$$

Dadas otra variable aleatoria X, vamos a considerar ahora el subespacio:

$$S = \{Y \in L^2(\Omega) : Y = h(X) \text{ donde } h : \mathbb{R} \to \mathbb{R}\}$$

Por razones técnicas tenemos que pedir que h sea una función boreliana como mencionamos en la clase pasada: es decir que $h^{-1}(I)$ sea un conjunto boreliano para cada intervalo abierto I en $\mathbb{R}$. Por ejemplo, cualquier función continua va a cumplir esto.
Entonces, si $X, Y \in L^2(\Omega)$ podemos definir la esperanza condicional $E[Y/X]$ como la solución $\widehat{Y}$ de este problema de optimización.

Un lema de álgebra lineal

Recordamos que la solución de este problema de encontrar la mejor aproximación, viene dado por la proyección ortogonal:

Lema

Sea V un espacio con producto interno y $S \subset V$ un subespacio. Consideramos $x_0 \in V$. Entonces $s_0 \in S$ es el elemento de S que minimiza la distancia a x_0

$$d(x, s) = \|x - s\| \quad x \in S$$

si y sólo si s_0 es la proyección ortogonal de x_0 sobre s es decir:

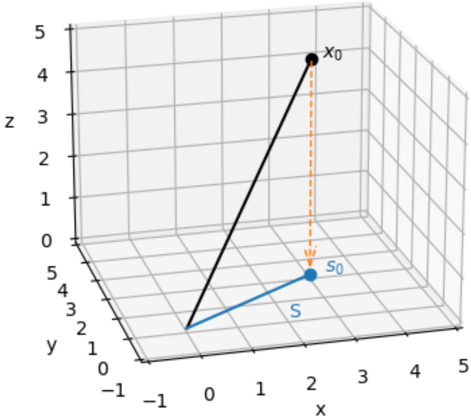
$$x_0 - s_0 \in S^{\perp}$$

o sea

$$\langle x_0 - s_0, s \rangle = 0 \text{ para todo } s \in S \tag{1}$$

Nuestro $V = L^2(\Omega)$ es un espacio de dimensión infinita, pero este lema funciona exactamente igual que en dimensión finita (y con la misma prueba).

Ilustración gráfica de la proyección ortogonal



La esperanza condicional

Aplicando este lema al ejemplo del subespacio

$$S = \{Y \in L^2(\Omega) : Y = h(X) \text{ donde } h : \mathbb{R} \to \mathbb{R}\}$$

vemos que la esperanza condicional $\widehat{Y} = E[Y/X]$ se define por las siguientes dos propiedades:

- $\widehat{Y} \in S$.
- $E[(Y - \widehat{Y}) \cdot Z] = 0$ para toda $Z \in S$, o sea:

$$E[Y \cdot Z] = E[\widehat{Y} \cdot Z] \text{ para todo } Z \in S$$

Son esencialmente las mismas condiciones de la definición axiomática de la esperanza condicional que vimos en la clase pasada. La única diferencia, es que allí trabajamos con variables con esperanza finita, entonces tuvimos que pedir que $Z = f(Y)$ con $f : \mathbb{R} \to \mathbb{R}$ acotada (para poder garantizar que las esperanzas que aparecen aquí sean finitas).

El caso en que la variable Y es discreta (1)

Como vimos en la clase pasada, el caso más sencillo de la esperanza condicional $E[Y/X]$ es cuando la variable X es discreta. Para simplificar vamos a suponer que

$$Im(X) = \{x_1, x_2, \dots, x_n\}$$

es finita y que $p_j = P\{X = x_j\} > 0$. Notamos que los eventos

$$A_j = \{X = x_j\} = \{\omega \in \Omega : X(\omega) = x_j\}$$

forman una partición de Ω. Vamos a suponer que $P(A_j) > 0$. En este caso S es de dimensión finita, y una base de S está formada por sus funciones indicadoras

$$B = \{I_{A_1}, I_{A_2}, \dots, I_{A_n}\}$$

La condición de que $\widehat{Y} = E[Y/X] \in S$ dice que

$$E[Y/X] = \sum_{k=1}^{n} c_j \cdot I_{A_j}$$

para ciertos escalares c_j que queremos determinar.

El caso en que la variable Y es discreta (2)

Ahora miramos la condición

$$E[Y \cdot Z] = E[\widehat{Y} \cdot Z] \text{ para todo } Z \in S$$

Como B es una base de S, alcanza mirar esta condición para $Z = I_{A_j}$. Por otra parte, como los A_j son disjuntos, resulta que B es una base ortogonal (pero no ortonormal) de S, pues

$$\langle I_{A_j}, I_{A_k} \rangle = E(I_{A_j} \cdot I_{A_k}) = \begin{cases} P(A_j) & \text{si} \quad j = k \\ 0 & \text{si} \quad j \neq k \end{cases}$$

Nos queda:

$$E[Y \cdot I_{A_j}] = c_j \cdot P(A_j)$$

entonces:

$$c_j = \frac{1}{P(A_j)} E[Y \cdot I_{A_j}] = E[Y/A_j]$$

que coincide con la defición que vimos en la clase pasada.

El caso en que la variable Y es discreta (3)

En resumen, cuando X es dicreta con imagen finita:

$$E[Y/X] = \sum_{i=1}^{n} E[Y/A_j] \cdot I_{A_j}$$

Esta fórmula puede generalizarse al caso en que X tiene imagen numerable (en este caso S no es de dimensión finita, y en lugar de una suma finita tenemos una serie, pero esencialmente funciona igual).

$$E[Y/X] = \sum_{i=1}^{\infty} E[Y/A_j] \cdot I_{A_j}$$

El caso en que la variable Y es discreta (4)

En este caso, debemos comprobar que esta fórmula define en efecto una función en $L^2(\Omega)$. Como las I_{A_j} son ortogonales

$$E[E[Y/X]^2] = \|E[Y/X]\|^2 = \sum_{j=1}^{\infty} \|E[Y/A_j] \cdot I_{A_j}\|^2$$

$$= \sum_{j=1}^{\infty} |E[Y/A_j]|^2 \|I_{A_j}\|^2$$

$$= \sum_{j=1}^{\infty} |E[Y/A_j]|^2 P(A_j)$$

Ahora por la desigualdad de Cauchy-Schwarz:

$$|E[Y/A_j]| \leq \frac{1}{P(A_j)} E(|YI_{A_j}|) \leq \frac{1}{P(A_j)} \|YI_{A_j}\| \cdot \|I_{A_j}\| = \|Y\| = E(I_{A_j} Y^2)^{1/2} \frac{1}{P(A_j)^{1/2}}$$

Entonces

$$E[E[Y/X]^2] \leq \sum_{i=1}^{\infty} |E[I_{A_j} Y]|^2 = E(|Y|^2) < +\infty$$

Parte II

Esperanzas condicionales en el caso continuo

El caso en que X e Y son variables continuas con una densidad conjunta

El otro caso que vimos en la clase anterior es cuando X e Y son variables continuas con una densidad conjunta $f_{XY}(x,y)$.

Recordamos que en este caso definimos la densidad condicional

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

donde

$$f_X(x) = \int_{-\infty}^{\infty} f_{XY}(x,y)\, dy$$

es la densidad marginal de X, suponiendo que $f_X(y) > 0$ para todo y. En la clase anterior, definimos $E[Y/X] = h(X)$ donde

$$h(x) = \int_{-\infty}^{\infty} y \cdot f_{Y/X=x}(y)\, dy$$

El caso en que X e Y son variables continuas con una densidad conjunta (2)

Si Y tiene esperanza finita, podemos calcular

$$\begin{aligned}
E[|E[Y/X]] = E[|h(X)|] &= \int_{-\infty}^{\infty} |h(x)|\ f_X(x)\ dx \\
&= \int_{-\infty}^{\infty} \left| \int_{-\infty}^{\infty} y \cdot f_{Y/X=x}(x)\ dy \right|\ f_X(x)\ dx \\
&\leq \int_{-\infty}^{\infty} \left[\int_{-\infty}^{\infty} |y| \cdot f_{Y/X=x}(x)\ dy \right]\ f_X(x)\ dx \\
&= \int_{-\infty}^{\infty} |y| \left[\int_{-\infty}^{\infty} f_{Y/X=x}(x)\ f_X(x) dx \right]\ dy \\
&= \int_{-\infty}^{\infty} |y| \left[\int_{-\infty}^{\infty} f_{XY}(x,y) dx \right]\ dy \\
&= \int_{-\infty}^{\infty} |y| \cdot f_Y(y)\ dy = E(|Y|) < \infty
\end{aligned}$$

Se deduce en particular que $h(x)$ es finita para casi todo x, por lo que $E[X/Y]$ está bien definida.

El caso en que X e Y son variables continuas con una densidad conjunta (2)

Vamos a comprobar ahora que si $Z = g(X)$ con $g : \mathbb{R} \to \mathbb{R}$ acotada, entonces $E[Y \cdot Z] = E[\widehat{Y} \cdot Z]$
Para calcular $E[Y \cdot Z]$ la pensamos como la esperanza de una función del vector aleatorio (X, Y).

$$E[Y \cdot Z] = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} y \cdot g(x) \cdot f_{XY}(x, y) \; dx \; dy$$

mientras que como $\widehat{Y} \cdot Z$ es una función de X sola,

$$E[\widehat{Y} \cdot Z] = \int_{-\infty}^{\infty} h(x) \cdot g(x) \cdot f_X(x) \; dx$$

pero por la definición de h,

$$h(x) \; f_X(x) = \left[\int_{-\infty}^{\infty} y \cdot f_{Y/X=x}(y) \; dy \right] \cdot f_X(x) = \int_{-\infty}^{\infty} y \cdot f_{XY}(x, y) \; dy$$

El caso en que X e Y son variables continuas con una densidad conjunta (3)

Reemplazando vemos que

$$E[\widehat{Y} \cdot Z] = \int_{-\infty}^{\infty} g(x) \left[\int_{-\infty}^{\infty} y \cdot f_{XY}(x,y)\, dy \right]\, dx$$
$$= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} y\, g(x) f_{XY}(x,y)\;\, dx\, dy = E[Y \cdot Z]$$

Vemos que $h(Y)$ verfica las dos condiciones de la definición axiomática de esperanza condicional que vimos en la clase anterior.

Si supiéramos que $Y \in L^2$ va a resultar que $\widehat{Y}$ está en L^2 y que la misma cuenta se puede hacer suponiendo que Z está en L^2 (aunque g no fuera acotada).

El caso en que X e Y son variables continuas con una densidad conjunta (4)

Para probar que en efecto $\widehat{Y} = E[Y/X] \in L^2$ cuando $Y \in L^2$, la idea es aproximar h por funciones acotadas:

$$h_n(x) = \begin{cases} h(x) & \text{si} \quad |h(x)| \leq n \\ n & \text{si} \quad h(x) > n \\ -n & \text{si} \quad h(x) < -n \end{cases} \qquad n \in \mathbb{N}$$

Entonces tomando $Z_n = h_n(X)$, tenemos que:

$$E[Y \cdot Z_n] = E[\widehat{Y} \cdot Z_n]$$

por lo que ya probamos. Luego por la desigualdad de Cauchy-Schwarz:

$$|E[\widehat{Y} \cdot Z_n]| \leq E(Y^2)^{1/2} \cdot E(Z_n^2)^{1/2}$$

pero como $|h_n(x)| \leq |h(x)|$, $|Z_n| \leq |\widehat{Y}|$, luego como $Y \in L^2$:

$$|E[\widehat{Y} \cdot Z_n]| \leq E(Y^2)^{1/2} \cdot E(\widehat{Y}^2)^{1/2}$$

El caso en que X e Y son variables continuas con una densidad conjunta (5)

Ahora bien, explícitamente

$$E[\widehat{Y} \cdot Z_n] = \int_{-\infty}^{\infty} h_n(x)\; h(x)\; f_X(x)\; dx$$

cuando $n \to +\infty$ esta integral va a converger a

$$E[\widehat{Y}^2] = \int_{-\infty}^{\infty} h(x)^2\; f_X(x)\; dx$$

porque $h_n(x)$ converge en forma monótona creciente hacia h. [Por el teorema de convergencia monótona, otro resultado clave de análisis real). Resulta que:

$$E[\widehat{Y}^2] \leq E[Y^2]$$

o sea:

$$E(E(Y/X)^2) \leq E(Y^2)$$

que es la misma desigualdad que obtuvimos antes en el caso discreto.

Parte III

Propiedades de la esperanza condicional

Definición axiomática de la esperanza condicional

Como vimos en la clase anterior, la esperanza condicional $\widehat{Y} = E[Y/X]$ se puede definir de forma axiomática cuando $Y \in L^1$. por las siguientes propiedades:

- $\widehat{Y} = h(X)$ donde $h : \mathbb{R} \to \mathbb{R}$ es una función boreliana.
- $\widehat{Y} \in L^1$ o sea $E(|\widehat{Y}|) < \infty$.
- Si $Z = f(X)$ donde $f : \mathbb{R} \to \mathbb{R}$ es una función boreliana acotada, entonces:

$$E[\widehat{Y}Z] = E[YZ]$$

Unicidad

Proposición

Si $\widehat{Y}_1$ y $\widehat{Y}_2$ verifican la definición axiomática de la esperanza condicional, entonces $\widehat{Y}_1 = \widehat{Y}_2$ con probabilidad 1.

Demostración Sea $W = \widehat{Y}_1 - \widehat{Y}_2$. $\widehat{Y}_1 = h_1(X)$, $\widehat{Y}_2 = h_2(X)$. Entonces

$$W = h(X) \text{ con } h = h_1 - h_2$$

$$E[WZ] = E[\widehat{Y}_1 Z] - E[\widehat{Y}_2 Z] = E[YZ] - E[YZ] = 0$$

para toda $Z = f(X)$ con f boreliana acotada. Elegimos $f(x) = I_{\{h(x) > \delta\}}$. Tenemos

$$\delta \cdot P(A_\delta) \leq E[W \cdot I_{A_\delta}] = 0 \text{ donde } A_\delta = \{W > \delta\}$$

Luego $P(A_\delta) = 0$ para todo *delta* > 0, se deduce que $W \leq 0$ con probabilidad 1. Cambiando W por $-W$, vemos que también $W \geq 0$ con probabilidad 1, luego $W = 0$ o sea $\widehat{Y}_1 = \widehat{Y}_2$ con probabilidad 1.

Linealidad

Las propiedades de la esperanza condicional se pueden deducir de la definición axiomática.

Linealidad

Sean $Y_1, Y_2 \in L^1(\Omega)$. Si $c_1, c_2 \in \mathbb{R}$,

$$E[c_1 \cdot Y_1 + c_2 \cdot Y_2/X] = c_1 \cdot E[Y_1/X] + c_2 E[Y_2/X]$$

Demostración Sean $\widehat{Y_1} = E[Y_1/X]$, $\widehat{Y_2} = E[Y_2/X]$, $Y = c_1 \cdot Y_1 + c_2 \cdot Y_2$ Hay que verificar que $\widehat{Y} = c_1 \widehat{Y_1} + c_2 \widehat{Y_2}$ cumple con la definción axiomática de esperanza condicional.

- $\widehat{Y}$ es función de X porque $\widehat{Y_1}$ e $\widehat{Y_2}$ lo son. ✓
- $\widehat{Y}$ tiene esperanza finita, pues $E[\widehat{Y}] = c_1 E[\widehat{Y_1}] + c_2 E[\widehat{Y_2}]$✓
- Si $Z = f(X)$ con f acotada, entonces

$$E[\widehat{Y} Z] = c_1 E[\widehat{Y_1} Z] + c_2 [\widehat{Y_2} Z] = c_1 E[Y_1 Z] + c_2 [Y_2 Z] = E[YZ]✓$$

Por la unicidad de la esperanza condicional, vale la propiedad.

Una generalización

Proposición

Sea $Y \in L^1$, $g : \mathbb{R} \to \mathbb{R}$ boreliana acotada, X otra variable aleatoria:

$$E[Yg(X)/X] = g(X)E[Y/X]$$

Otras propiedades de la esperanza condicional:

- Si $Y \in L^1$, $E[E[Y/X]] = E[Y]$
- Monotonía: si $Y_1, Y_2 \in L^1$, $Y_1 \leq Y_2$ con probabilidad 1,

$$E[Y_1/X] \leq E[Y_2/X]$$

- Desigualdad de Cauchy-Schwartz: Si $X, Y \in L^2$,

$$E(Y_1 Y2/X) \leq E[Y_1/X]^{1/2} \cdot E[Y_2/X]^{1/2}$$

- Desigualdad de Jensen: si $Y \in L^1$, $\varphi : \mathbb{R} \to \mathbb{R}$ es convexa y $\varphi(Y) \in L^1$,

$$\varphi(E[Y/X]) \leq E[\varphi(Y)/X)$$

En particular:

$$|E[Y/X])|^p \leq E[|Y|^p/X) \text{ si } p \geq 1$$

Condicionamiento reiterado

$$E[E[X/Y]/Z] = E[X/Z]$$