INTEGRACIÓN ESTOCÁSTICA

Sesión 06/16

Esperanza Condicional

Introducción

El concepto de Esperanza Condicional es básico en la Teoría de la Probabilidad Moderna. Su definición fue formulada por Kolmogorov en su libro de 1933 (*Foundations of the Theory of Probability*), en el cual estableció la formulación de la Teoría de la Probabilidad que prevalece hasta nuestros días. Como veremos más adelante, la definición de la esperanza condicional está basada en el teorema de Radon Nikodym, publicado en el año 1930. Ese teorema fue la conclusión de la investigación que inició Lebesgue acerca de la condición para que una función sea una integral indefinida, la cual consiste en que esa función tiene que ser absolutamente continua. Radon continuó esa investigación y estableció el ahora llamado teorema de Radon-Nikodym para el caso de medidas en $\mathbb{R}^n$. El resultado de Nikodym fue mucho más general ya que lo formuló habiéndose desarrollado ya una teoría general de la medida. Únicamente pasaron 3 años para que Kolmogorov hiciera ver la importancia del resultado de Nikodym en la teoría de la probabilidad.

Ahora el concepto de esperanza condicional es una de las principales bases para el estudio e incluso la definición misma de los procesos estocásticos, ya que precisamente se constituyó en la herramienta central para tratar con variables aleatorias dependientes.

Antes de la formulación moderna de la Teoría de la Probabilidad se trataba ya con distribuciones condicionales, aunque únicamente se hacía para el caso de variables aleatorias que admiten una función de densidad.

En la formulación moderna, el concepto central no es el de distribución condicional sino el de esperanza condicional. Se trata, por ejemplo, de definir $E[X \mid Y]$, donde X y Y son variables aleatorias; o, un poco más general, $E[X \mid Y_1, Y_2, \ldots, Y_n]$, donde $Y_1, Y_2, \ldots, Y_n$ son variables aleatorias.

Para entender mejor la idea de la esperanza condicional es conveniente tener en mente la definición de σ-álgebra generada por una familia de funciones.

Definición 1 (σ álgebra generada por una familia de funciones). *Sea $(\mathbb{E}, \mathfrak{E})$ un espacio medible y $\mathbb{F}$ un conjunto cualquiera. Dada una colección de funciones:*

$$\mathcal{H} = \{f_\gamma : \mathbb{F} \to (\mathbb{E}, \mathfrak{E}) : \gamma \in \Gamma\}$$

donde Γ es un conjunto de índices cualquiera, se define la σ-álgebra generada por $\mathcal{H}$ como la más pequeña σ-álgebra de subconjuntos de $\mathbb{F}$ tal que toda función $f \in \mathcal{H}$ es medible. Denotaremos a esta σ-álgebra por por $\sigma(\mathcal{H})$ o por $\sigma(\{f_\gamma : \gamma \in \Gamma\})$.

Obsérvese que si $f : \mathbb{F} \to (\mathbb{E}, \mathfrak{E})$ es cualquier función, la familia de conjuntos $\{f^{-1}(B) : B \in \mathfrak{E}\}$ es una σ-álgebra de subconjuntos de $\mathbb{F}$. Sin embargo, si $\{f_\gamma : \mathbb{F} \to (\mathbb{E}, \mathfrak{E}) : \gamma \in \Gamma\}$ es una colección de funciones, la familia de conjuntos $\left\{f_\gamma^{-1}(B) : \gamma \in \Gamma \text{ y } B \in \mathfrak{E}\right\}$ no es, en general, una σ-álgebra, pero la σ-álgebra generada por esa familia de conjuntos es la σ-álgebra generada por la familia de funciones $\{f_\gamma : \mathbb{F} \to (\mathbb{E}, \mathfrak{E}) : \gamma \in \Gamma\}$.

Por otra parte, si $f_1, f_2, \ldots, f_n$ son funciones de $\mathbb{F}$ en $(\mathbb{R}, \mathfrak{B}(\mathbb{R}))$ y definimos $f : \mathbb{F} \to \left(\overline{\mathbb{R}}, \mathfrak{B}\left(\overline{\mathbb{R}}^n\right)\right)$ mediante la relación:

$$f(x) = (f_1(x), f_2(x), \ldots, f_n(x))$$

sabemos que si $\Im$ es una σ-álgebra de subconjuntos de $\mathbb{F}$, entonces f es $\Im$-medible si y sólo si f_k es $\Im$-medible para cualquier $k \in \{1, 2, \ldots, n\}$. Por lo tanto:

$$\left\{f^{-1}(B) : B \in \mathfrak{B}\left(\overline{\mathbb{R}}^n\right)\right\} = \sigma(f) = \sigma(\{f_1, f_2, \ldots, f_n\})$$

Proposición 1. *Sean $Y_1, \ldots, Y_n$ n variables aleatorias con valores en $\overline{\mathbb{R}}$ y $Z : \Omega \mapsto \overline{\mathbb{R}}$ una función $\sigma(Y_1, \ldots, Y_n)$-medible. Entonces, existe una función boreliana $h : \overline{\mathbb{R}}^n \mapsto \overline{\mathbb{R}}$ tal que $Z = h(Y_1, \ldots, Y_n)$.*

Demostración

Si $Z = I_E$, donde $E \in \sigma(Y_1, \ldots, Y_n)$, entonces existe un boreliano $B \subset \overline{\mathbb{R}}^n$ tal que $Z = I_B(Y_1, \ldots, Y_n)$. Por lo tanto, se tiene el resultado para el caso de una función simple Z $\sigma(Y_1, \ldots, Y_n)$-medible.

Si Z es una variable aleatoria no negativa, sea $(Z_n)_{n \in \mathbb{N}}$ una sucesión no decreciente de funciones simples no negativas tales que $Z = \lim_{n \rightsquigarrow \infty} Z_n$ y, para cada $n \in \mathbb{N}$, sea $h_n : \overline{\mathbb{R}}^n \mapsto \overline{\mathbb{R}}$ una función boreliana no negativa tal que $Z_n = h_n(Y_1, \ldots, Y_n)$.

Sea $D = \left\{x \in \overline{\mathbb{R}}^n : \lim_{n \rightsquigarrow \infty} h_n(x) \text{ existe}\right\}$. Entonces D es un conjunto boreliano de $\overline{\mathbb{R}}^n$ y contiene a la imagen de Ω bajo la función $(Y_1, \ldots, Y_n)$. Definamos la función $h : \overline{\mathbb{R}}^n \mapsto \overline{\mathbb{R}}$ de la siguiente manera:

$$h(x) = \begin{cases} \text{lím}_{n \rightsquigarrow \infty} h_n(x) & \text{si } x \in D \\ 0 & \text{en otro caso} \end{cases}$$

h es entonces una función boreliana y $Z = h(Y_1, \ldots, Y_n)$.

■

Corolario 1. *Sean $Y_1, \ldots, Y_n$ n variables aleatorias con valores en $\mathbb{R}$ y $Z : \Omega \mapsto \mathbb{R}$ una función $\sigma(Y_1, \ldots, Y_n)$-medible. Entonces, existe una función boreliana $h : \mathbb{R}^n \mapsto \mathbb{R}$ tal que $Z = h(Y_1, \ldots, Y_n)$.*

Si $Y_1, Y_2, \ldots, Y_n$ son variables aleatorias, $E[X \mid Y_1, Y_2, \ldots, Y_n]$ expresa la esperanza de X dado que son conocidos los valores de $Y_1, Y_2, \ldots, Y_n$. Planteado de esta forma, con una buena definición, se esperaría que $E[X \mid Y_1, Y_2, \ldots, Y_n]$ sea una función de las variables aleatorias $Y_1, Y_2, \ldots, Y_n$; es decir: $E[X \mid Y_1, Y_2, \ldots, Y_n] = h(Y_1, Y_2, \ldots, Y_n)$, donde h es una función medible de $\mathbb{R}^n$ en $\mathbb{R}$. Obsérvese que, en particular, $E[X \mid Y_1, Y_2, \ldots, Y_n]$ sería una variable aleatoria medible con respecto a la σ-álgebra generada por las variables aleatorias $Y_1, Y_2, \ldots, Y_n$. Inicialmente, X es sólo medible con respecto a la σ-álgebra $\Im$ del espacio de probabilidad en el que estemos trabajando. Veremos en la exposición que sigue que esta nueva variable aleatoria es una especie de proyección de X sobre la σ-álgebra $\sigma(Y_1, Y_2, \ldots, Y_n)$, la cual, en general, es más pequeña que la σ-álgebra original $\Im$. Es una variable aleatoria que, en promedio, se comporta como la variable aleatoria original.

Para el estudio de los procesos estocásticos, en los cuales interviene una infinidad de variables aleatorias, no es suficiente contar con una definición de la esperanza de una variable aleatoria X dado que son conocidos los valores de un número finito de variables aleatorias que no son independientes de X; se requiere poder hacerlo también para el caso de una infinidad de variables aleatorias.

Como ejemplo, consideremos una familia de variables aleatorias $(X_u)_{u \in [0,\infty)}$, todas ellas definidas en el mismo espacio de probabilidad $(\Omega, \Im, P)$. Un problema común consiste en el siguiente:

Dados $s, t \in [0, \infty)$, con $s < t$, si se conocen los valores de X_u para cualquier $u \in [0, s]$, ¿qué se puede decir acerca de X_t? Una respuesta a esta pregunta la da el concepto de esperanza condicional ya que se puede definir $E\left[X_t \mid (X_u)_{u \in [0,s]}\right]$.

Como veremos más adelante, el resultado que permite definir una esperanza condicional de este tipo es el teorema de Radon-Nikodym, el cual trata con medidas (definidas sobre σ-álgebras); de manera que la definición de la esperanza condicional $E\left[X_t \mid (X_u)_{u \in [0,s]}\right]$ se da considerando la σ-álgebra generada por la familia de variables aleatorias $\{X_u : u \in [0, s]\}$. Denotando por $\Im_s$ a esa σ-algebra, lo que se define es la esperanza condicional $E[X_t \mid \Im_s]$.

Definición de la esperanza condicional

Definición 2. *Sean μ y ν dos medidas definidas sobre el mismo espacio medible $(\mathbb{E}, \mathcal{E})$. Se dice que ν es absolutamente continua con respecto a μ, lo cual será denotado por $\nu \ll \mu$, si $\nu(E) = 0$ para cualquier conjunto medible E tal que $\mu(E) = 0$.*

Teorema 1 (**Radon-Nikodym**). *Sean μ y ν dos medidas definidas sobre el mismo espacio medible $(\mathbb{E}, \mathcal{E})$. Supongamos que μ es finita y que ν es absolutamente continua con respecto a μ, entonces existe una función medible no negativa f tal que $\nu(E) = \int_E f d\mu$ para cualquier conjunto medible E.*

Teorema 2 (**Existencia de la esperanza condicional**). *Sea X una variable aleatoria de esperanza finita y $\mathcal{G}$ una sub σ-álgebra de $\mathcal{F}$, existe entonces una variable aleatoria Z, medible con respecto a $\mathcal{G}$, de esperanza finita y tal que $\int_B Z dP = \int_B X dP$ para cualquier conjunto $B \in \mathcal{G}$. Además, si Y es otra variable aleatoria con las mismas propiedades que Z, entonces $P[Y = Z] = 1$.*

Demostración

Sean X^+ y X^- la parte positiva y negativa, respectivamente, de X. Definamos entonces las medidas $Q_+ : \mathcal{G} \to \mathbb{R}$ y $Q_- : \mathcal{G} \to \mathbb{R}$ mediante las relaciones $Q_+(B) = \int_B X^+ dP$ y $Q_-(B) = \int_B X^- dP$, respectivamente. Q_+ y Q_- son absolutamente continuas con respecto a P (restringida a $\mathcal{G}$), de manera que, por el teorema de Radon-Nikodym, existen dos funciones no negativas $Z_+ : \Omega \to \overline{\mathbb{R}}$ y $Z_- : \Omega \to \overline{\mathbb{R}}$, $\mathcal{G}$-medibles, tales que, $Q_+(B) = \int_B Z_+ dP$ y $Q_-(B) = \int_B Z_- dP$ para cualquier $B \in \mathcal{G}$. En particular, tanto Z_+ como Z_- tienen esperanza finita. Así que $Z = Z_+ - Z_-$ satisface las condiciones del teorema.

Por otra parte, si Y es otra variable aleatoria con las mismas propiedades que Z, entonces $\int_B Y dP = \int_B Z dP$ para cualquier evento $B \in \mathcal{G}$, así que $P[Y = Z] = 1$.

$\blacksquare$

Obsérvese que la propiedad $\int_B Z dP = \int_B X dP$ para cualquier conjunto $B \in \mathcal{G}$, se puede escribir en la forma siguiente: $E[I_B Z] = E[I_B X]$ para cualquier conjunto $B \in \mathcal{G}$

Definición 3. *Sea X es una variable aleatoria de esperanza finita y $\mathcal{G}$ una sub σ-álgebra de $\mathcal{F}$. Se dice que Z es una versión de la* ***esperanza condicional de X con respecto a $\mathcal{G}$*** *y se escribe $E[X \mid \mathcal{G}] = Z$ c.s. si Z una variable aleatoria medible con respecto a $\mathcal{G}$, de esperanza finita y tal que $\int_B Z dP = \int_B X dP$ para cualquier evento $B \in \mathcal{G}$.*

Definición 4. *Sea X una variable aleatoria de esperanza finita y $Y_1, Y_2, \ldots, Y_n$ n variables aleatorias cualesquiera. Se define*

$$E[X \mid Y_1, Y_2, \ldots, Y_n] = E[X \mid \sigma(Y_1, Y_2, \ldots, Y_n)]$$

A un nivel elemental se define la esperanza condicional, $E[X \mid Y]$, de una variable aleatoria X, dado que se conoce el valor de otra variable aleatoria Y, únicamente en el caso en el que existe una función de densidad conjunta, definiendo primero una función de densidad condicional. Por ejemplo, en el caso aboslutamente continuo, se define la función $x \mapsto f_{X|Y}(x \mid y)$ de la siguiente manera:

$$f_{X|Y}(x \mid y) = \begin{cases} \frac{f_{X,Y}(x,y)}{f_Y(y)} & \text{si } f_Y(y) > 0 \\ f_X(x) & \text{en otro caso} \end{cases}$$

Para cualquier $y \in \mathbb{R}$, la función $x \mapsto f_{X|Y}(x \mid y)$ es una función de densidad.

En efecto, la función $x \mapsto f_{X|Y}(x \mid y)$ es no negativa y se tiene:

$$\int_{-\infty}^{\infty} f_{X|Y}(x \mid y)dx = \begin{cases} \frac{1}{f_Y(y)} \int_{-\infty}^{\infty} f_{X,Y}(x,y)dx & \text{si } f_Y(y) > 0 \\ \int_{-\infty}^{\infty} f_X(x)dx & \text{en otro caso} \end{cases}$$

$$= \begin{cases} \frac{1}{f_Y(y)} f_Y(y) & \text{si } f_Y(y) > 0 \\ \int_{-\infty}^{\infty} f_X(x)dx & \text{en otro caso} \end{cases} = 1$$

Entonces se define:

$$E[X \mid Y = y] = \int_{-\infty}^{\infty} x f_{X|Y}(x \mid y)dx$$

Ejemplo 1. *Supongamos que un cierto evento ocurre consecutivamente en los tiempos aleatorios X y Y de tal manera que X y $Y-X$ son variables aleatorias independientes, ambas con distribución exponencial de parámetro λ.*

Considerando la transformación $u=x$, $v=y-x$, se tiene:

$$f_{X,Y}(x,y)=f_{U,V}(x,y-x)=f_U(x)\,f_V(y-x)$$

$$=\begin{cases}\lambda e^{-\lambda x}\lambda e^{-\lambda(y-x)} & \text{si } 0<x<y\\ 0 & \text{en otro caso.}\end{cases}=\begin{cases}\lambda^2 e^{-\lambda y} & \text{si } 0<x<y\\ 0 & \text{en otro caso.}\end{cases}$$

Vamos a encontrar $E[Y\mid X]$.

Se tiene:

$$f_{Y|X}(y\mid x)=\frac{f_{X,Y}(x,y)}{f_X(x)}=\begin{cases}\frac{\lambda^2 e^{-\lambda y}}{\lambda e^{-\lambda x}} & \text{si } 0<x<y\\ 0 & \text{en otro caso}\end{cases}$$

$$=\begin{cases}\lambda e^{-\lambda(y-x)} & \text{si } 0<x<y\\ 0 & \text{en otro caso}\end{cases}$$

Por lo tanto:

$$E[Y\mid X=x]=\begin{cases}\int_x^\infty \lambda y e^{-\lambda(y-x)}dy & \text{si } x>0\\ 0 & \text{en otro caso}\end{cases}$$

Considerando el cambio de variable $z=y-x$ en la integral anterior, obtenemos:

$$E[Y\mid X=x]=\begin{cases}\int_0^\infty \lambda(x+z)\,e^{-\lambda z}dz & \text{si } x>0\\ 0 & \text{en otro caso}\end{cases}$$

$$=\begin{cases}x\int_0^\infty \lambda e^{-\lambda z}dz+\int_0^\infty \lambda z e^{-\lambda z}dz & \text{si } x>0\\ 0 & \text{en otro caso}\end{cases}=\begin{cases}x+\frac{1}{\lambda} & \text{si } x>0\\ 0 & \text{en otro caso}\end{cases}$$

Así que, $Z=E[Y\mid X]=X+\frac{1}{\lambda}$.

Veámos que Z satisface la condición que se pide en la definición general de la esperanza condicional.

La σ-álgebra generada por X está formada por todos los conjuntos de la forma $X^{-1}(B)$, donde B es un boreliano de $\mathbb{R}$.

Así que lo que tenemos que demostrar es que, para cualquier conjunto $B \in \mathcal{B}(\mathbb{R})$, se cumple la siguiente relación:

$$E\left[I_{X^{-1}(B)}Z\right] = E\left[I_{X^{-1}(B)}Y\right]$$

$E\left[I_{X^{-1}(B)}Z\right] = E\left[I_B(X)\left(X + \frac{1}{\lambda}\right)\right] = \int_0^{\infty} I_B(x)\left(x + \frac{1}{\lambda}\right)\lambda e^{-\lambda x}dx$

$= \int_0^{\infty} \lambda x I_B(x) e^{-\lambda x}dx + \int_0^{\infty} I_B(x) e^{-\lambda x}dx$

$E\left[I_{X^{-1}(B)}Y\right] = E\left[I_B(X)Y\right] = \int_0^{\infty}\int_x^{\infty} y I_B(x)\lambda^2 e^{-\lambda y}dydx$

$= \int_0^{\infty} \lambda I_B(x)\int_x^{\infty} y\lambda e^{-\lambda y}dydx$

Considerando el cambio de variable $z = y - x$, obtenemos:

$E\left[I_B(X)Y\right] = \int_0^{\infty} \lambda I_B(x)\int_0^{\infty}(z + x)\lambda e^{-\lambda z}e^{-\lambda x}dzdx$

$= \int_0^{\infty} \lambda I_B(x) e^{-\lambda x}\int_0^{\infty} z\lambda e^{-\lambda z}dzdx + \int_0^{\infty} \lambda x I_B(x) e^{-\lambda x}\int_0^{\infty} \lambda e^{-\lambda z}dzdx$

$= \int_0^{\infty} I_B(x) e^{-\lambda x}dx + \int_0^{\infty} \lambda x I_B(x) e^{-\lambda x}dx$

Ahora encontremos $E[X \mid Y]$.

Se tiene:

$$f_Y(y) = \int_{-\infty}^{\infty} f_{X,Y}(x,y)\,dx = \begin{cases} \int_0^y \lambda^2 e^{-\lambda y}dx & \text{si } y > 0 \\ 0 & \text{en otro caso.} \end{cases}$$

$$= \begin{cases} \lambda^2 y e^{-\lambda y} & \text{si } y > 0 \\ 0 & \text{en otro caso.} \end{cases}$$

Así que:

$$f_{X|Y}(x \mid y) = \frac{f_{X,Y}(x,y)}{f_Y(y)} = \begin{cases} \frac{\lambda^2 e^{-\lambda y}}{\lambda^2 y e^{-\lambda y}} & \text{si } 0 < x < y \\ 0 & \text{en otro caso} \end{cases} = \begin{cases} \frac{1}{y} & \text{si } 0 < x < y \\ 0 & \text{en otro caso} \end{cases}$$

Por lo tanto:

$$E[X \mid Y=y] = \begin{cases} \frac{1}{y}\int_0^y x dx & \text{si } y>0 \\ 0 & \text{en otro caso} \end{cases} = \begin{cases} \frac{1}{2}y & \text{si } y>0 \\ 0 & \text{en otro caso} \end{cases}$$

Así que, $W = E[X \mid Y] = \frac{1}{2}Y$.

Veámos que W satisface la condición que se pide en la definición general de la esperanza condicional.

Así que lo que tenemos que demostrar es que, para cualquier conjunto $B \in \mathcal{B}(\mathbb{R})$, se cumple la siguiente relación:

$$E\left[I_{Y^{-1}(B)}W\right] = E\left[I_{Y^{-1}(B)}X\right]$$

$E\left[I_{Y^{-1}(B)}W\right] = E\left[I_B(Y)W\right] = E\left[I_B(Y)\frac{1}{2}Y\right] = \frac{1}{2}\int_0^\infty yI_B(y)\lambda^2 ye^{-\lambda y}dy$

$= \frac{1}{2}\int_0^\infty \lambda^2 y^2 I_B(y)e^{-\lambda y}dy$

$E\left[I_{Y^{-1}(B)}X\right] = E\left[I_B(Y)X\right] = \int_0^\infty \int_0^y xI_B(y)\lambda^2 e^{-\lambda y}dxdy$

$= \int_0^\infty \lambda^2 e^{-\lambda y}I_B(y)\int_0^y xdxdy = \int_0^\infty \lambda^2 e^{-\lambda y}I_B(y)\frac{1}{2}y^2 dy$

$= \frac{1}{2}\int_0^\infty \lambda^2 y^2 I_B(y)e^{-\lambda y}dy$

Propiedades de la esperanza condicional

La siguiente proposición muestra que la esperanza condicional tiene propiedades similares a las de la esperanza no condicional. Se muestra también que tiene las propiedades que podrían esperarse con una buena definición, por corresponder a la idea intuitiva del concepto. Finalmente, se muestran otras propiedades específicas de la esperanza condicional, las cuales no resultan evidentes a partir de la idea intuitiva.

Proposición 2. *Sean X y Y dos variables aleatorias de esperanza finita, $\mathcal{G}$ y $\mathcal{H}$ dos subσ-álgebras de $\mathcal{F}$ y c una constante. Se tienen entonces las siguientes propiedades:*

i. $E[c \mid \mathcal{G}] = c$.

ii. $E[cX \mid \mathcal{G}] = cE[X \mid \mathcal{G}]$.

iii. $E[X + Y \mid \mathcal{G}] = E[X \mid \mathcal{G}] + E[Y \mid \mathcal{G}]$.

iv. Si X es $\mathcal{G}$-medible, entonces $E[X \mid \mathcal{G}] = X$.

v. $E[E[X \mid \mathcal{G}]] = E[X]$.

vi. Si $\mathcal{G} \subset \mathcal{H}$, entonces $E[E[X \mid \mathcal{H}] \mid \mathcal{G}] = E[X \mid \mathcal{G}]$.

vii. Si X es independiente de $\mathcal{G}$, entonces $E[X \mid \mathcal{G}] = E[X]$.

viii. Si $X \leq Y$, entonces $E[X \mid \mathcal{G}] \leq E[Y \mid \mathcal{G}]$.

ix. $|E[X \mid \mathcal{G}]| \leq E[|X| \mid \mathcal{G}]$.

Demostración

i y *ii* son inmediatas de la definición de esperanza condicional.

Para probar *iii*, definamos:

$Z = E[X + Y \mid \mathcal{G}]$

$Z_1 = E[X \mid \mathcal{G}]$

$Z_2 = E[Y \mid \mathcal{G}]$

Sea B cualquier conjunto en $\mathcal{G}$; entonces:

$\int_B (Z_1 + Z_2)\, dP = \int_B Z_1 dP + \int_B Z_2 dP = \int_B X dP + \int_B Y dP = \int_B (X + Y)\, dP$

Así que $Z_1 + Z_2$ es una versión de la esperanza condicional $E[X + Y \mid \mathcal{G}]$.

Por lo tanto:

$Z = Z_1 + Z_2$ con probabilidad 1.

iv. X es $\mathcal{G}$-medible y, para cualquier conjunto $B \in \mathcal{G}$, se tiene:

$$\int_B XdP = \int_B XdP$$

Por lo tanto:

$$E\left[X \mid \mathcal{G}\right] = X$$

v. Sabemos que $\int_B E\left[X \mid \mathcal{G}\right] dP = \int_B XdP$ para cualquier conjunto $B \in \mathcal{G}$, así que, en particular, se tiene:

$$E\left[E\left[X \mid \mathcal{G}\right]\right] = \int_\Omega E\left[X \mid \mathcal{G}\right] dP = \int_\Omega XdP = E\left[X\right]$$

Para probar *vi*, sean Z una versión de la esperanza condicional de X con respecto a $\mathcal{H}$, Y una versión de la esperanza condicional de Z con respecto a $\mathcal{G}$ y $B \in \mathcal{G}$. Entonces:

$$\int_B YdP = \int_B ZdP = \int_B XdP$$

Así que Y es una versión de la esperanza condicional de X con respecto a $\mathcal{G}$.

Para probar *vii*, sea $B \in \mathcal{G}$. Se tiene entonces:

$$\int_B E\left[X\right] dP = E\left[X\right] E\left[I_B\right] = E\left[XI_B\right] = \int_B XdP$$

Así que $E\left[X\right]$ es una versión de la esperanza condicional de X con respecto a $\mathcal{G}$.

Para demostrar *viii*, sea U una versión de la esperanza condicional de X con respecto a $\mathcal{G}$ y V una versión de la esperanza condicional de Y con respecto a $\mathcal{G}$, entonces, si $B \in \mathcal{G}$, se tiene:

$$\int_B UdP = \int_B XdP \leq \int_B YdP = \int_B VdP$$

Como U y V son $\mathcal{G}$-medibles, se tiene $U \leq V$ c.s.

Para demostar *ix*, se tiene $|X| \geq X$ y $|X| \geq -X$, así que $E\left[|X| \mid \mathcal{G}\right] \geq E\left[X \mid \mathcal{G}\right]$ y $E\left[|X| \mid \mathcal{G}\right] \geq -E\left[X \mid \mathcal{G}\right]$; por lo tanto, $E\left[|X| \mid \mathcal{G}\right] \geq \left|E\left[X \mid \mathcal{G}\right]\right|$

■

Las siguientes proposiciones generalizan la propiedad básica de la esperanza condicional.

Proposición 3. *Sea X es una variable aleatoria de esperanza finita, $\mathcal{G}$ una subσ-álgebra de $\mathcal{F}$ y Z una versión de la esperanza condicional de X con respecto a $\mathcal{G}$, entonces $E[WZ] = E[WX]$ para cualquier variable aleatoria acotada W $\mathcal{G}$-medible.*

Demostración

Si $\varphi = \sum_{k=1}^{n} x_k I_{B_k}$ es una variable aleatoria simple $\mathcal{G}$-medible, se tiene:

$E[\varphi Z] = E\left[\sum_{k=1}^{n} x_k I_{B_k} Z\right] = \sum_{k=1}^{n} x_k E\left[I_{B_k} Z\right]$

$= \sum_{k=1}^{n} x_k E\left[I_{B_k} X\right] = E\left[\sum_{k=1}^{n} x_k I_{B_k} X\right] = E[\varphi X]$

Si W es una variable aleatoria acotada $\mathcal{G}$-medible, existe una sucesión de variables aleatorias simples $\{\varphi_n\}_{n\in\mathbb{N}}$ tal que $\{\varphi_n(\omega)\}_{n\in\mathbb{N}}$ converge a $W(\omega)$ para cualquier $\omega \in \Omega$ y $|\varphi_n| \leq |W|$ para cualquier $n \in \mathbb{N}$. Así que, por el teorema de la convergencia dominada, se tiene:

$E[WZ] = E[\text{lím}_{n\rightsquigarrow\infty} \varphi_n Z] = \text{lím}_{n\rightsquigarrow\infty} E[\varphi_n Z] = \text{lím}_{n\to\infty} E[\varphi_n X]$

$= E[\text{lím}_{n\rightsquigarrow\infty} \varphi_n X] = E[WX]$

$\blacksquare$

Proposición 4. *Sea $\mathcal{G}$ una subσ-álgebra de $\mathcal{F}$, X una variable aleatoria de esperanza finita y Y una variable aleatoria $\mathcal{G}$-medible tal que YX tiene esperanza finita, entonces $YE[X \mid \mathcal{G}]$ tiene esperanza finita y $E[YX \mid \mathcal{G}] = YE[X \mid \mathcal{G}]$.*

Demostración

Supongamos primero que X y Y son no negativas y sea Z una versión de la esperanza condicional de X con respecto a $\mathcal{G}$. Se tiene entonces $E[WZ \mid \mathcal{G}] = E[WX]$ para cualquier variable aleatoria acotada W $\mathcal{G}$-medible. En particular, si $B \in \mathcal{G}$ y W es una variable aleatoria simple $\mathcal{G}$-medible, se tiene

$$\int_B WZdP = \int_B WXdP$$

Por otra parte, Y es el límite de una sucesión creciente, $\{W_n\}$, de variables aleatorias simples $\mathcal{G}$-medibles, así que, para cada $n \in N$ y $B \in \mathcal{G}$, se tiene

$$\int_B W_nZdP = \int_B W_nXdP$$

Tomando límites cuando n tiende a infinito se obtiene

$$\int_B YZdP = \int_B YXdP$$

lo cual implica que YZ tiene esperanza finita y es una versión de la esperanza condicional de YX con respecto a $\mathcal{G}$.

Para el resultado general, se puede escribir

$YX = (Y^+ - Y^-)(X^+ - X^-) = Y^+X^+ + Y^-X^- - Y^+X^- - Y^-X^+$

Además, $(YX)^+ = Y^+X^+ + Y^-X^-$ y $(YX)^- = Y^+X^- + Y^-X^+$, así que, como X y YX tienen esperanza finita, X^+, X^-, Y^+X^+, Y^-X^-, Y^+X^- y Y^-X^+ también tienen esperanza finita, por lo tanto se puede aplicar, en cada caso, la primera parte de la demostración.

$\blacksquare$

Proposición 5. *Sea $\mathcal{G}$ una sub σ-álgebra de $\mathcal{F}$ y X una variable aleatoria de varianza finita, entonces $Z = E[X \mid \mathcal{G}]$ también tiene varianza finita.*

Demostración

Definamos, para cada $n \in \mathbf{N}$, $W_n = I_{[-n,n]}(Z)$. Entonces W_n, $W_n Z$ y $W_n Z^2$ son variables aleatorias acotadas $\mathcal{G}$-medibles; así que $W_n (X - Z)^2 = W_n X^2 - 2W_n XZ + W_n Z^2$ tiene esperanza finita.

Además:

$E\left[W_n (X - Z)^2 \mid \mathcal{G}\right] = E[W_n X^2 - 2W_n XZ + W_n Z^2 \mid \mathcal{G}] = W_n E[X^2 \mid \mathcal{G}] - W_n Z^2$

Así que:

$$W_n Z^2 \leq W_n E[X^2 \mid \mathcal{G}]$$

Tomando límites cuando $n \rightsquigarrow \infty$, se obtiene $Z^2 \leq E[X^2 \mid \mathcal{G}]$.

Finalmente, tomando esperanzas en ambos miembros de la última desigualdad, se concluye que $E[Z^2] \leq E[X^2] < \infty$.

$\blacksquare$

Proposición 6 (**Teorema de la convergencia monótona**). *Sea $\mathcal{G}$ una sub σ-álgebra de $\mathcal{F}$ y $\{X_n\}_{n\in\mathbb{N}}$ una sucesión monótona creciente de variables aleatorias de esperanza finita tales que $X = \lim_{n\rightsquigarrow\infty} X_n$ tiene esperanza finita. Entonces:*

$$\lim_{n\to\infty} E\left[X_n \mid \mathcal{G}\right] = E\left[X \mid \mathcal{G}\right]$$

Demostración

Sea Z una versión de la esperanza condicional de X con respecto a $\mathcal{G}$ y, para cada $n \in \mathbb{N}$, Z_n una versión de la esperanza condicional de X_n con respecto a $\mathcal{G}$. Siendo la sucesión $\{Z_n\}_{n\in\mathbb{N}}$ monónotona creciente, el límite $Z^* = \lim_{n\rightsquigarrow\infty} Z_n$ existe. Además, como $Z_n \leq Z$ para cualquier $n \in N$, entonces $Z^* \leq Z$.

Ahora bien, las sucesiones $\{Z - Z_n\}_{n\in\mathbb{N}}$ y $\{X - X_n\}_{n\in\mathbb{N}}$ son no negativas, monótonas decrecientes y están acotadas por $Z - Z_1$ y $X - X_1$, respectivamente, de manera que, por el teorema de la convergencia dominada, se tiene:

$$\lim_{n\to\infty} E(Z - Z_n) = E(Z - Z^*)$$

$$\lim_{n\to\infty} E(X - X_n) = 0$$

Pero, como $Z - Z_n$ es una versión de la esperanza condicional de $X - X_n$ con respecto a $\mathcal{G}$, se tiene $E(Z - Z_n) = E(X - X_n)$. Así que $E(Z - Z^*) = 0$. Por lo tanto $Z^* = Z$ c.s.

$\blacksquare$

Corolario 2. *Sea $\mathcal{G}$ una sub σ-álgebra de $\mathcal{F}$ y $\{X_n\}_{n\in\mathbb{N}}$ una sucesión monótona decreciente de variables aleatorias de esperanza finita tales que $X = \lim_{n\to\infty} X_n$ tiene esperanza finita. Entonces:*

$$\lim_{n\to\infty} E\left[X_n \mid \mathcal{G}\right] = E\left[X \mid \mathcal{G}\right]$$

Proposición 7 (**Lema de Fatou**). *Sea $\mathcal{G}$ una sub σ-álgebra de $\mathcal{F}$ y $\{X_n\}_{n\in\mathbb{N}}$ una sucesión de variables aleatorias no negativas de esperanza finita tales que* $\liminf_{n\rightsquigarrow\infty} X_n$ *tiene esperanza finita. Entonces:*

$$E\left[\liminf_{n\rightsquigarrow\rightarrow\infty} X_n \mid \mathcal{G}\right] \leq \liminf_{n\rightsquigarrow\rightarrow\infty} E\left[X_n \mid \mathcal{G}\right]$$

Demostración

La sucesión $Y_n = \inf\{X_j : j \geq n\}$ es monótona creciente y $\liminf_{n\rightsquigarrow\infty} X_n = \lim_{n\to\infty} Y_n$, así que, por la proposición 6, $E\left[\liminf_{n\rightsquigarrow\infty} X_n \mid \mathcal{G}\right] = \lim_{n\to\infty} E\left[Y_n \mid \mathcal{G}\right]$.

Por otra parte, $Y_n \leq X_j$ para cualquier $j \geq n$, así que $E\left[Y_n \mid \mathcal{G}\right] \leq \inf\{E\left[X_j \mid \mathcal{G}\right] : j \geq n\}$. Por lo tanto,

$E\left[\liminf_{n\rightsquigarrow\infty} X_n \mid \mathcal{G}\right] = \lim_{n\rightsquigarrow\infty} E\left[Y_n \mid \mathcal{G}\right]$

$\leq \lim_{n\rightsquigarrow\infty} \inf\{E\left[X_j \mid \mathcal{G}\right] : j \geq n\} = \liminf_{n\rightsquigarrow\infty} E\left[X_n \mid \mathcal{G}\right]$.

∎

Proposición 8. *Sea* $\mathcal{G}$ *una sub* σ*-álgebra de* $\mathcal{F}$*,* Y *una variable aleatoria no negativa de esperanza finita y* $\{X_n\}_{n\in\mathbb{N}}$ *una sucesión de variables aleatorias tales que* $|X_n| \leq Y$ *y* $\lim_{n\rightsquigarrow\infty} X_n$ *existe c.s.. Entonces* $\lim_{n\rightsquigarrow\infty} E\left[|X_n - X| \mid \mathcal{G}\right] = 0$*, donde* $X = \lim_{n\to\infty} X_n$*.*

Demostración

Para cada $n \in \mathbb{N}$ sea $Z_n = 2Y - |X_n - X|$, entonces, por la proposición 7, se tiene:

$2E\left[Y \mid \mathcal{G}\right] = E\left[\lim_{n\to\infty} Z_n \mid \mathcal{G}\right]$

$\leq \liminf_{n\rightsquigarrow\infty} E\left[Z_n \mid \mathcal{G}\right] = 2E\left[Y \mid \mathcal{G}\right] - \limsup_{n\rightsquigarrow\infty} E\left[|X_n - X| \mid \mathcal{G}\right]$.

Así que,

$$\limsup_{n\rightsquigarrow\infty} E\left[|X_n - X| \mid \mathcal{G}\right] = 0$$

■

Corolario 3 (**Teorema de la convergencia dominada**)**.** *Sea* $\mathcal{G}$ *una sub* σ*-álgebra de* $\mathcal{F}$*,* Y *una variable aleatoria no negativa de esperanza finita y* $\{X_n\}_{n\in\mathbb{N}}$ *una sucesión de variables aleatorias tales que* $|X_n| \leq Y$ *y* $\lim_{n\to\infty} X_n$ *existe c.s.. Entonces* $E\left[\lim_{n\to\infty} X_n \mid \mathcal{G}\right] = \lim_{n\to\infty} E\left[X_n \mid \mathcal{G}\right]$*.*

Proposición 9. *Sea* $\{X_n\}_{n\in\mathbb{N}}$ *una sucesión de variables aleatorias la cual converge a* X *en* L^p*, con* $p \in [1,\infty)$*, y* $\mathcal{G}$ *una sub*σ*-álgebra de* $\mathcal{F}$*. Entonces la sucesión* $\{E[X_n \mid \mathcal{G}]\}_{n\in\mathbb{N}}$ *converge a* $E[X \mid \mathcal{G}]$ *en* L^p *.*

Demostración

Por la desigualdad de Jensen, se tiene:

$|E[X_n \mid \mathcal{G}] - E[X \mid \mathcal{G}]|^p = |E[X_n - X \mid \mathcal{G}]|^p \leq E[|X_n - X|^p \mid \mathcal{G}]$

Así que:

$E(|E[X_n \mid \mathcal{G}] - E[X \mid \mathcal{G}]|^p) \leq E(E[|X_n - X|^p \mid \mathcal{G}]) = E[|X_n - X|^p]$

Por lo tanto:

$$\lim_{n\rightsquigarrow\infty} E(|E[X_n \mid \mathcal{G}] - E[X \mid \mathcal{G}]|^p) = \lim_{n\rightsquigarrow\infty} E[|X_n - X|^p] = 0$$

■

Esperanzas condicionales como proyecciones

Las propiedades *ii*, *iii*, *vii* y *viii* de la proposición 2 muestran que la esperanza condicional se comporta como una proyección. En esta sección se mostrará que, efectivamente, limitándonos al caso de variables aleatorias en $L^2(\Omega, \mathcal{F}, P)$, la esperanza condicional puede verse como una proyección en el sentido riguroso del término.

Lema 1. *Sea $\mathcal{G}$ una subσ-álgebra de $\mathcal{F}$, X una variable aleatoria en $L^2(\Omega, \mathcal{F}, P)$ y $Z = E[X \mid \mathcal{G}]$. Entonces Z pertenece a $L^2(\Omega, \mathcal{G}, P)$ y $E\left[(X-Z)^2\right] = E[X^2 - Z^2]$.*

Demostración

Que Z pertenece a $L^2(\Omega, \mathcal{G}, P)$ es una consecuencia directa de la proposición 5. Por otra parte,

$E\left[(X-Z)^2 \mid \mathcal{G}\right] = E[X^2 - 2XZ + Z^2 \mid \mathcal{G}]$

$= E[X^2 \mid \mathcal{G}] - 2ZE[X \mid \mathcal{G}] + Z^2$

$= E[X^2 \mid \mathcal{G}] - 2Z^2 + Z^2 = E[X^2 - Z^2 \mid \mathcal{G}]$.

Así que, tomando esperanzas,

$$E\left[(X-Z)^2\right] = E[X^2 - Z^2]$$

∎

Proposición 10. *Sea $\mathcal{G}$ una subσ-álgebra de $\mathcal{F}$, X una variable aleatoria en $L^2(\Omega, \mathcal{F}, P)$ y $Z = E[X \mid \mathcal{G}]$. Entonces $E\left[(X-Z)^2\right] \leq E\left[(X-Y)^2\right]$ para cualquier variable aleatoria Y en $L^2(\Omega, \mathcal{G}, P)$.*

Demostración

Por el lema 1, se tiene

$$E\left[(X-Z)^2\right] = E[X^2] - E[Z^2]$$

Por otra parte,

$E\left[(X-Y)^2 \mid \mathcal{G}\right] = E[X^2 - 2XY + Y^2 \mid \mathcal{G}]$

$= E[X^2 \mid \mathcal{G}] - 2YE[X \mid \mathcal{G}] + Y^2 = E[X^2 \mid \mathcal{G}] - 2YZ + Y^2$.

Así que,

$$E\left[(X-Y)^2\right] = E[X^2] - 2E[YZ] + E[Y^2]$$

De manera que,

$$E\left[(X-Y)^2\right]-E\left[(X-Z)^2\right]=E\left[Y^2\right]-2E\left[YZ\right]+E\left[Z^2\right]$$

$$=E\left[Y^2-2YZ+Z^2\right]=E\left[(Y-Z)^2\right]\geq 0.$$

■

Obsérvese que la proposición 10 muestra que $E\left[X\mid\mathcal{G}\right]$ es un buen estimador de X en el sentido de que, entre todas las variables aleatorias Y en $L^2(\Omega,\mathcal{G},P)$, $E\left[X\mid\mathcal{G}\right]$ minimiza el valor de $E\left[(X-Y)^2\right]$. Por tal motivo se puede decir que $E\left[X\mid\mathcal{G}\right]$ es el **mejor estimador $\mathcal{G}$-medible de X en el sentido de la media cuadrática**.

Proposición 11. *Sea $\mathcal{G}$ una subσ-álgebra de $\mathcal{F}$, X una variable aleatoria en $L^2(\Omega,\mathcal{F},P)$ y $Z=E\left[X\mid\mathcal{G}\right]$. Entonces $E\left[Y\left(X-Z\right)\right]=0$ para cualquier variable aleatoria Y en $L^2(\Omega,\mathcal{G},P)$.*

Demostración

Para cualquier $t\in\mathbf{R}$, la variable aleatoria $Z+tY$ pertenece a $L^2(\Omega,\mathcal{F},P)$, así que, por la proposición 10, $E\left[(X-Z)^2\right]\leq E\left[(X-Z-tY)^2\right]$, de lo cual se sigue $2tE\left[Y\left(X-Z\right)\right]\leq t^2E\left[Y^2\right]$.

Por otro lado, si $E\left[Y\left(X-Z\right)\right]\neq 0$, se puede elegir t, suficientemente pequeño en valor absoluto, de tal manera que $2tE\left[Y\left(X-Z\right)\right]>t^2E\left[Y^2\right]$, lo cual es una contradicción.

■

Los resultados anteriores nos dan la base para hacer ver que la esperanza condicional es una proyección en un sentido riguroso. En efecto, si $\mathcal{G}$ es una subσ-álgebra de $\mathcal{F}$, los espacios vectoriales $\mathcal{H}=L^2(\Omega,\mathcal{F},P)$ y $\mathcal{S}=L^2(\Omega,\mathcal{G},P)$ son espacios de Hilbert con producto interior definido por $\langle X,Y\rangle=E\left[XY\right]$, de manera que si $\mathcal{S}^\perp$ es el complemento ortogonal de $\mathcal{S}$ en $\mathcal{H}$, se tiene $\mathcal{H}=\mathcal{S}\oplus\mathcal{S}^\perp$, donde $\oplus$ representa la suma directa de subespacios. Esto significa que cada elemento $X\in\mathcal{H}$ admite una única representación de la forma $X=Z+Y$, donde $Z\in\mathcal{S}$ y $Y\in\mathcal{S}^\perp$. Como es sabido, Y *y* Z son llamados las proyecciones ortogonales de X sobre los subespacios $\mathcal{S}^\perp$ y $\mathcal{S}$, respectivamente. Ahora bien, si $X\in\mathcal{H}$, entonces el lema 1 asegura que $E\left[X\mid\mathcal{G}\right]\in\mathcal{S}$ y la proposición 11 afirma que $X-E\left[X\mid\mathcal{G}\right]\in\mathcal{S}^\perp$, de manera que, como $X=E\left[X\mid\mathcal{G}\right]+(X-E\left[X\mid\mathcal{G}\right])$, $E\left[X\mid\mathcal{G}\right]$ es la proyección ortogonal de X sobre $\mathcal{S}$.