TEORÍA DE LA PROBABILIDAD

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 axiomática 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. La primera referencia a un problema de este tipo se encuentra en un artículo de Thomas Bayes, publicado en el año 1763 y que lleva por título *An essay towards solving a problem in the doctrine of chances*. En ese artículo Bayes buscaba estimar la probabilidad de ocurrencia de un evento a partir de la frecuencia con la que el evento había ocurrido; es decir, se trataba del mismo problema que había estudiado Jacques Bernoulli, cuya solución condujo a la Ley Débil de los Grandes Números. Sin embargo, Bayes abordó el problema desde otro punto de vista, basándose en la idea de que cuando no se tiene ninguna información acerca de la ocurrencia o no ocurrencia de un evento, su probabilidad de ocurrencia se puede pensar como un número que se elige al azar en el intervalo $[0, 1]$. Después de contar con información acerca de la ocurrencia del evento, esa probabilidad se modifica. Es la idea central de lo que ahora se conoce como Estadística Bayesiana.

El problema que planteó Bayes es el siguiente:

Dado el número de veces en el cual un evento desconocido ha ocurrido y fallado, encontrar la probabilidad de que su probabilidad de ocurrencia en un ensayo esté comprendida entre dos valores dados.

Denotemos por p a la probabilidad de ocurrencia del evento en consideración. De acuerdo con la idea de Bayes, inicialmente p es una variable aleatoria con distribución uniforme en el intervalo $[0,1]$, de manera que el problema se puede plantear de la siguiente manera.

Denotemos por X al número de veces en que el evento ocurre en n observaciones y denotemos por Y a la probabilidad de ocurrencia del evento. Entonces, dado un intervalo $[a,b]$ contenido en el intervalo $[0,1]$ y $k \in \{0,1,2,\ldots,n\}$, se trata de calcular la siguiente probabilidad condicional:

$P\left[a < Y < b \mid X = k\right]$

Lo que se conoce es lo siguiente:

$P\left[X = k \mid Y = y\right] = \binom{n}{k} y^k (1-y)^{n-k}$

Esto, no por la definición de probabilidad condicional, sino simplemente porque el valor de Y es la probabilidad de ocurrencia del evento en consideración.

Para resolver el problema planteado, Bayes demostró que:

$P(X = k, a < Y < b) = \int_a^b P(X = k \mid Y = y)dy = \int_a^b \binom{n}{k} y^k (1-y)^{n-k} dy$

De aquí se sigue que:

$P(X = k) = \int_0^1 \binom{n}{k} y^k (1-y)^{n-k} dy = \binom{n}{k} \int_0^1 y^k (1-y)^{n-k} dy = \frac{1}{n+1}$

Y entonces:

$P\left[a < Y < b \mid X = k\right] = \frac{P(X=k, a<Y<b)}{P[X=k]} = (n+1)\binom{n}{k} \int_a^b y^k (1-y)^{n-k} dy$

Si definimos:

$f_{X|Y}(k \mid y) = P\left[X = k \mid Y = y\right]$

y denotamos por $f_{Y|X}$ a la función de densidad de Y dado X, entonces el resultado de Bayes puede escribirse en la siguiente forma:

$f_{Y|X}(y \mid k) = \frac{f_{X|Y}(k|y) f_Y(y)}{f_X(k)}$

Así que, para $k \in \{0,1,2,\ldots,n\}$, se tiene:

$$f_{Y|X}(y \mid k) = \begin{cases} \frac{1}{f_X(k)}\binom{n}{k} y^k (1-y)^{n-k} & \text{si } y \in [0,1] \\ 0 & \text{en otro caso} \end{cases}$$

$$= \begin{cases} (n+1)\binom{n}{k} y^k (1-y)^{n-k} & \text{si } y \in [0,1] \\ 0 & \text{en otro caso} \end{cases}$$

Obsérvese que al escribir $f_{Y|X}(y \mid k)$, estamos considerando que k es fija; la que varía es y. Además, como función de y, $f_{Y|X}(y \mid k)$ es una función de densidad de una variable aleatoria absolutamente continua; es decir, es no negativa y se tiene:

$$\int_0^1 f_{Y|X}(y \mid k) dy = 1$$

La distribución que corresponde a esta función de densidad es conocida como distribución beta; en este caso, con parámetros $k+1$ y $n-k+1$.

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. $E[X \mid Y_1, Y_2, \ldots, Y_n]$ 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 $\mathcal{F}$ 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 $\mathcal{F}$. Veremos también que 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 con 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. Veremos que la definición general incluye este caso ya que se define la esperanza de una variable aleatoria dada una σ-álgebra.

Definición de la esperanza condicional

Sea $(\Omega, \Im, P)$ un espacio de probabilidad completo y X una variable aleatoria de esperanza finita.

Para motivar la definición general, vamos a definir la esperanza condicional de X dada la ocurrencia de un evento A de probabilidad positiva. Para esto observemos que la función $P_A : \Im \to \mathbb{R}$ definida por $P_A(B) = P(B \mid A)$ es una medida de probabilidad, lo cual motiva la siguiente definición:

Definición 1. *Sea A un evento de probabilidad positiva y X una variable aleatoria de esperanza finita. Se define la esperanza condicional de X dada la ocurrencia del evento A, $E[X \mid A]$, mediante la relación $E[X \mid A] = \int_\Omega X dP_A$.*

Obsérvese que $\int_\Omega X dP_A = \frac{1}{P(A)} \int_A X dP$, de manera que $E[X \mid A]$ puede interpretarse como el promedio de los valores de X en el conjunto A.

Con esta definición, $E[X \mid A]$ tiene como valor un número real; es simplemente una extensión natural de la probabilidad condicional. El concepto que queremos definir es mucho más general; podríamos continuar con una definición de $E[X \mid Y]$ donde Y es una variable aleatoria. Esto ya tiene su dificultad para el caso de una variable aleatoria continua ya que, en ese caso, tendríamos $P[Y = y] = 0$ para cualquier $y \in \mathbb{R}$, de manera que no podríamos utilizar la definición de probabilidad condicional para definir $E[X \mid Y = y]$. Vamos a proceder extendiendo la definición anterior al caso de una familia finita de eventos mutuamente excluyentes y de probabilidad positiva (esto equivale a definir $E[X \mid Y]$ para el caso en que Y es una variable aleatoria discreta cuyo conjunto de posibles valores es finito). Veremos entonces que la esperanza condicional queda caracterizada por determinadas propiedades, las cuales permitirán definir el concepto general.

Proposición 1. *Sea X una variable aleatoria de esperanza finita y $A_1, A_2, \ldots, A_n$ eventos de probabilidad positiva, mutuamente excluyentes y tales que $\bigcup_{k=1}^n A_k = \Omega$. Definamos la función $Z : \Omega \to \mathbb{R}$ mediante la relación:*

$$Z(\omega) = E[X \mid A_k] \text{ si } \omega \in A_k$$

Entonces:

i) Z es una variable aleatoria medible con respecto a la σ-álgebra generada por los eventos $A_1, A_2, \ldots, A_n$.

ii) Z tiene esperanza finita.

iii) $\int_B Z dP = \int_B X dP$ para cualquier evento $B \in \sigma(A_1, A_2, \ldots, A_n)$

iv) Z es la única variable aleatoria con las propiedades *i*, *ii* y *iii*.

Demostración

La propiedad i es inmediata ya que $Z = \sum_{k=1}^{n} E[X \mid A_k] I_{A_k}$.

$ii)$ $E[|Z|] = E[|\sum_{k=1}^{n} E[X \mid A_k] I_{A_k}|] = E\left[\left|\sum_{k=1}^{n} \frac{1}{P(A_k)} \left(\int_{A_k} XdP\right) I_{A_k}\right|\right]$

1. $\leq \sum_{k=1}^{n} E\left[\left|\frac{1}{P(A_k)} \left(\int_{A_k} XdP\right) I_{A_k}\right|\right] \leq \sum_{k=1}^{n} E\left[\frac{1}{P(A_k)} \left|\int_{A_k} XdP\right| I_{A_k}\right]$
$= \sum_{k=1}^{n} E\left[\left|\int_{A_k} XdP\right|\right] \leq \sum_{k=1}^{n} E\left[\int_{A_k} |X|\, dP\right] = \int_{\Omega} |X|\, dP = E[|X|] < \infty$

$iii)$ Sea $B \in \sigma(A_1, A_2, \ldots, A_n)$, entonces o bien B es el conjunto vacío o bien se puede expresar como una unión finita de uno o más de los eventos $A_1, A_2, \ldots, A_n$. Si B es el conjunto vacío entonces la igualdad $\int_B ZdP = \int_B XdP$ es evidente. Supongamos ahora que $B = \bigcup_{j=1}^{m} A_{k_j}$, donde $k_j \in \{1, 2, \ldots, n\}$ y los eventos $A_{k_1}, A_{k_2}, \ldots, A_{k_m}$ son ajenos. Entonces,

1. $\int_B ZdP = \sum_{j=1}^{m} \int_{A_{k_j}} ZdP = \sum_{j=1}^{m} \int_{A_{k_j}} E\left[X \mid A_{k_j}\right] dP$
$= \sum_{j=1}^{m} E\left[X \mid A_{k_j}\right] P(A_{k_j}) = \sum_{j=1}^{m} \int_{A_{k_j}} XdP = \int_B XdP$

$iv)$ Sea Y una variable aleatoria con las propiedades i, ii y iii. Por la propiedad i, para cada $k \in \{1, 2, \ldots, n\}$, Y toma un valor constante, α_k, sobre A_k, así que, por la propiedad iii:

$$\alpha_k P(A_k) = \int_{A_k} YdP = \int_{A_k} XdP$$

Por lo tanto, si $\omega \in A_k$, se tiene

$Y(\omega) = \alpha_k = \frac{1}{P(A_k)} \int_{A_k} XdP = E[X \mid A_k] = Z(\omega)$

■

La variable aleatoria Z es llamada la esperanza condicional de X dada la $\sigma-$álgebra generada por los conjuntos $A_1, A_2, \ldots, A_n$ y se denota por:

$$E[X \mid \sigma(A_1, A_2, \ldots, A_n)]$$

El resultado anterior nos dice que esta variable aleatoria queda caracterizada por las propiedades i, ii y iii.

Ahora vamos a utilizar el teorema de Radon-Nikodym para extender la idea anterior.

Se trata de resolver el siguiente problema:

Dado un espacio de probabilidad completo $(\Omega, \Im, P)$, una variable aleatoria X, de esperanza finita, definida en ese espacio y una $\sigma-$álgebra $\mathcal{G}$ contenida en F, encontrar una variable aleatoria Z, de esperanza finita, definida en el espacio de probabilidad $(\Omega, \mathcal{G}, P)$, que satisfaga las siguientes propiedades:

1. Z **es una variable aleatoria medible con respecto a la** σ**-álgebra** $\mathcal{G}$**.**

2. Z **tiene esperanza finita.**

3. $\int_B ZdP = \int_B XdP$ **para cualquier conjunto** $B \in \mathcal{G}$**.**

Como lo mencionamos antes, **la existencia de una variable aleatoria Z con esas propiedades se demuestra utilizando el teorema de Radon-Nykodim; pero cabe aclarar que lo que se demuestra es únicamente la existencia de esa variable aleatoria; el teorema no nos dice cómo se puede dar Z explicitamente**. Ésta es una de las dificultades que hay al tratar con la esperanza condicional; queda caracterizada únicamente por las propiedades que tiene.

En algunos casos particulares se puede dar explicitamente la esperanza condicional e incluso la distribución condicional de una variable aleatoria. Esto ocurre, por ejemplo, cuando se calcula la esperanza condicional de una variable aleatoria dado que se conoce el valor de otra variable aleatoria, cuando existe una función de densidad conjunta de las dos variables aleatorias.

Pero **el principal uso que tiene la esperanza condicional es en la teoría de los procesos estocásticos, para definirlos y demostrar algunas de sus propiedades**.

Definición 2. *Sean μ y ν dos medidas. 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 $\mu(E) = 0$.*

Por ejemplo, si μ es una medida definida sobre un espacio de medida $(\mathbb{F}, \mathcal{F})$ y f una función medible no negativa, entonces la función $\nu : \Im \to \overline{\mathbb{R}}$ definida por $\nu(E) = \int_E f d\mu$ es una medida absolutamente continua con respecto a μ. El teorema de Radon Nykodim afirma que el inverso de esta propiedad es válido. Es un resultado que parece sorprendente por su simplicidad; sin embargo no olvidemos el trabajo previo de mucha gente que permitió llegar a este resultado.

La demostración del teorema de Radon-Nykodim es algo laboriosa y requiere de la introducción de varios conceptos y resultados. Aquí únicamente lo vamos a enunciar para aplicarlo después al estudio de la esperanza condicional.

Teorema 1 (**Radon-Nikodym**). *Sean μ y ν dos medidas definidas sobre el mismo espacio medible $(\mathbb{F}, \mathcal{F})$. Supongamos que μ es σ-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, definida sobre $(\Omega, \Im, P)$, y $\mathcal{G}$ una sub σ-álgebra de $\Im$, existe entonces una variable aleatoria Z, medible con respecto a $\mathcal{G}$, de esperanza finita y tal que $\int_B ZdP = \int_B XdP$ para cualquier conjunto $B \in \mathcal{G}$. Además, si Y es otra variable aleatoria con las mismas propiedades que Z, entonces $Y = Z$ c.s.*

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, 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 $\int_B YdP = \int_B ZdP$ para cualquier evento $B \in \mathcal{G}$, así que $Y = Z$ c.s.

■

Obsérvese que la propiedad $\int_B ZdP = \int_B XdP$ 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 $B \in \mathcal{G}$. De aquí se sigue que si $f : (\Omega, \mathcal{G}) \to \overline{\mathbb{R}}$ es una función medible y acotada, entonces $E[fZ] = E[fX]$.

Con el objetivo de familiarizarse con el concepto de esperanza condicional, vamos a ver cómo se obtienen esperanzas condicionales de una variable aleatoria dado que se conoce el valor de otra cuando existe una función de densidad conjunta. Esta parte es relativamente simple, únicamente requiere manejar bien las funciones de densidad conjuntas, las series dobles y las integrales dobles. Las propiedades más importantes de la esperanza condicional las veremos más adelante.

Esperanzas condicionales en el caso discreto

Proposición 2. *Sean X y Y dos variables aleatorias discretas y $g : \mathbb{R}^2 \mapsto \mathbb{R}$ una función boreliana tal que $g(X, Y)$ tiene esperanza finita. Definamos la función $h : \mathbb{R} \to \mathbb{R}$ de la siguiente manera:*

$$h(y) = \begin{cases} \sum_k g(x_k, y) \frac{f_{X,Y}(x,y)}{f_Y(y)} & \text{si } f_Y(y) > 0 \\ 0 & \text{en otro caso} \end{cases}$$

donde $x_1, x_2, \ldots$ son los posibles valores de X. Entonces $h(Y)$ es una versión de la esperanza condicional $E[g(X,Y) \mid Y]$.

Demostración

Demostremos primero que h está bien definida. En efecto, si $y_1, y_2, \ldots$ son los posibles valores de Y, se tiene:

$\sum_j \sum_k |g(x_k, y_j)| P[X = x_k, Y = y_j] = E[|g(X,Y)|] < \infty$

Por lo tanto, $\sum_k |g(x_k, y)| P[X = x_k, Y = y] < \infty$ para cualquier $y \in \mathbb{R}$.

De manera que, si $P[Y = y] > 0$:

$\sum_k |g(x_k, y)| \frac{P[X=x_k, Y=y]}{P[Y=y]} = \frac{1}{P[Y=y]} \sum_k |g(x_k, y)| P[X = x_k, Y = y] < \infty$

Demostremos ahora que la variable aleatoria $h(Y)$ tiene esperanza finita. En efecto, se tiene:

$E[|h(Y)|] = \sum_j |h(y_j)| P[Y = y_j] \leq \sum_j \sum_k |g(x_k, y_j)| \frac{P[X=x_k, Y=y_j]}{P[Y=y_j]} P[Y = y_j]$

$= \sum_j \sum_k |g(x_k, y_j)| P[X = x_k, Y = y_j] = E[|g(X,Y)|] < \infty$

Sea ahora $f : \mathbb{R} \mapsto \mathbb{R}$ cualquier función acotada. Se tiene entonces:

$E[f(Y)h(Y)] = \sum_j f(y_j) h(y_j) P[Y = y_j]$

$= \sum_j f(y_j) \sum_k g(x_k, y_j) P[X = x_k, Y = y_j]$

$= \sum_{j,k} f(y_j) g(x_k, y_j) P[X = x_k, Y = y_j] = E[f(Y)g(X,Y)]$

■

Ejemplo 1. *Se eligen, al azar y sin reemplazo, dos tarjetas de una urna que contiene N tarjetas numeradas del 1 al N, donde N es un número entero mayor que 1. Sean X y Y el menor y mayor, respectivamente, de los números de las tarjetas seleccionadas. Encuentra:*

a) $E[X \mid Y]$

b) $E[Y \mid X]$

Solución

La función de densidad conjunta del vector aleatorio (X, Y) está dada por:

$$f_{X,Y}(x,y) = \begin{cases} \frac{2}{N(N-1)} & \text{si } x < y, \ x, y \in \{1, \ldots, N\} \\ 0 & \text{en otro caso} \end{cases}$$

a. Para $y \in \{2, \ldots, N\}$, se tiene:

$f_Y(y) = \sum_{x=1}^{N} f_{X,Y}(x,y) = \sum_{x=1}^{y-1} \frac{2}{N(N-1)} = \frac{2}{N(N-1)}(y-1)$

Así que:

$h\left(y\right) = \sum_{x=1}^{N} x \frac{f_{X,Y}(x,y)}{f_Y(y)} = \frac{1}{y-1} \sum_{x=1}^{y-1} x = \frac{1}{2} y$

Por lo tanto:

$E\left[X \mid Y\right] = \frac{1}{2} Y.$

b. Para $x \in \{1, \ldots, N-1\}$, se tiene:

$f_X(x) = \sum_{y=1}^{N} f_{X,Y}(x,y) = \sum_{y=x+1}^{N} \frac{2}{N(N-1)} = \frac{2}{N(N-1)}(N-x)$

Así que:

$h\left(x\right) \sum_{y=1}^{N} y \frac{f_{X,Y}(x,y)}{f_X(x)} = \frac{1}{N-x} \sum_{y=x+1}^{N} y = \frac{1}{2} x + \frac{1}{2}(N+1)$

Por lo tanto:

$E\left[Y \mid X\right] = \frac{1}{2} X + \frac{1}{2}(N+1).$

Ejemplo 2. *Sean X y Y dos variables aleatorias independientes, ambas con distribución geométrica de parámetro p. Encuentra $E\left[X \mid \text{mín}(X,Y)\right]$.*

Solución

Para $z \in \{0, 1, \ldots\}$, se tiene:

$P\left[\text{mín}(X,Y) = z\right] = 2P\left[X > z, Y = z\right] + P\left[X = z, Y = z\right]$

$= 2(1-p)^{z+1} p(1-p)^z + p^2(1-p)^{2z} = p(1-p)^{2z}(2-p)$

Así que:

$E\left[X \mid \text{mín}(X,Y) = z\right] = \sum_{x=0}^{\infty} x \frac{P[X=x, \text{mín}(X,Y)=z]}{P[\text{mín}(X,Y)=z]}$

$= \frac{1}{P[\text{mín}(X,Y)=z]} \left[\sum_{x=z+1}^{\infty} x P\left[X = x, Y = z\right] + z P\left[X = z, Y \geq z\right]\right]$

$= \frac{1}{p(1-p)^{2z}(2-p)} \left[p^2(1-p)^z \sum_{x=z+1}^{\infty} x(1-p)^x + zp(1-p)^{2z}\right]$

$= \frac{1}{p(1-p)^{2z}(2-p)} \left[p^2(1-p)^z \, (1-p)^{z+1} \, \frac{pz+1}{p^2} + zp(1-p)^{2z}\right]$

$= \frac{p(2-p)z+(1-p)}{p(2-p)} = z + \frac{1-p}{p(2-p)}$.

Por lo tanto, $E\left[X \mid \text{mín}(X,Y)\right] = \text{mín}(X,Y) + \frac{(1-p)}{p(2-p)}$.

Esperanzas condicionales en el caso absolutamente continuo

Proposición 3. *Sea (X,Y) un vector aleatorio absolutamente continuo con función de densidad conjunta $f_{X,Y}$ y $g : \mathbb{R}^2 \mapsto \mathbb{R}$ una función boreliana tal que $g(X,Y)$ tiene esperanza finita. Definamos la función $h : \mathbb{R} \mapsto \mathbb{R}$ de la siguiente manera:*

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

Entonces $h(Y)$ es una versión de la esperanza condicional $E\left[g(X,Y) \mid Y\right]$.

Demostración

Recordemos que, siendo la función $f_{X,Y}$ integrable, por el teorema de Fubini, el conjunto de puntos $y \in \mathbb{R}$ para los cuales la función $x \mapsto f_{X,Y}(x,y)$ no es integrable, tiene medida cero. Además:

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

Definamos los siguientes conjuntos.

$A = \{y \in \mathbb{R} : f_Y(y) > 0\}$

$B = \left\{y \in \mathbb{R} : \int_{-\infty}^{\infty} |g(x,y)| \, f_{X,Y}(x,y)dx < \infty\right\}$

$C = \left\{y \in \mathbb{R} : \int_{-\infty}^{\infty} f_{X,Y}(x,y)dx < \infty\right\}$

$D_y = \{x \in \mathbb{R} : f_{X,Y}(x,y) > 0\}$

Como mencionamos antes, C^c tiene medida cero. Además:

$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} |g(x,y)| \, f_{X,Y}(x,y)dxdy < \infty$

Así que, por el teorema de Fubini, B^c tiene medida cero.

Por otra parte, si $y \in A^c \cap C$, entonces $\int_{-\infty}^{\infty} f_{X,Y}(x,y)dx = 0$, así que D_y tiene medida cero.

También por el teorema de Fubini, h es una función boreliana, así que $h(Y)$ es una variable aleatoria.

Demostremos que $h(Y)$ tiene esperanza finita. En efecto, se tiene:

$\int_{-\infty}^{\infty} |h(y)|\, f_Y(y)dy = \int_{A\cap B\cap C} |h(y)|\, f_Y(y)dy$

$\leq \int_{A\cap B\cap C} \int_{-\infty}^{\infty} |g(x,y)|\, f_{X,Y}(x,y)dxdy$

$\leq \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} |g(x,y)|\, f_{X,Y}(x,y)dxdy < \infty$

Sea $f : \mathbb{R} \mapsto \mathbb{R}$ cualquier función boreliana acotada. Entonces:

$E\left[f(Y)h(Y)\right] = \int_{-\infty}^{\infty} f(y)h(y)f_Y(y)dy$

$= \int_{A\cap B\cap C} f(y) \int_{-\infty}^{\infty} g(x,y)f_{X,Y}(x,y)dxdy$

$= \int_{A\cap C} \int_{-\infty}^{\infty} f(y)g(x,y)f_{X,Y}(x,y)dxdy$

$= \int_{A\cap C} \int_{D_y} f(y)g(x,y)f_{X,Y}(x,y)dxdy$

$= \int_{C} \int_{D_y} f(y)g(x,y)f_{X,Y}(x,y)dxdy$

$= \int_{C} \int_{-\infty}^{\infty} f(y)g(x,y)f_{X,Y}(x,y)dxdy$

$= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(y)g(x,y)f_{X,Y}(x,y)dxdy = E\left[f(Y)g(X,Y)\right]$

Así que $h(Y)$ es una versión de la esperanza condicional $E\left[g(X,Y) \mid Y\right]$.

∎

Ejemplo 3. *Sean X y Y dos variables aleatorias con función de densidad conjunta dada por:*

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

Encuentra $E\left[X+Y \mid Y-X\right]$.

Solución

$f_{X+Y,Y-X}(u,v) = \frac{1}{2} f\left(\frac{u-v}{2}, \frac{u+v}{2}\right) = \begin{cases} \frac{1}{2}\lambda^2 e^{-\frac{1}{2}\lambda(u+v)} & \text{si } 0 < v < u \\ 0 & \text{en otro caso} \end{cases}$

$f_{Y-X}(v) = \int_{v}^{\infty} \frac{1}{2}\lambda^2 e^{-\frac{1}{2}\lambda(u+v)}du = \lambda e^{-\lambda v}$

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

$E\left[X+Y \mid Y-X=v\right] = \int_v^\infty \frac{1}{2}\lambda u e^{-\frac{1}{2}\lambda(u-v)}du = v + \frac{2}{\lambda}$

Así que, $E\left[X+Y \mid Y-X\right] = Y - X + \frac{2}{\lambda}$.

Ejemplo 4. *Sean X y Y dos variables aleatorias independientes, ambas con distribución normal estándar. Encuentra $E\left[X^n \mid X^2+Y^2\right]$ para cualquier $n \in \mathbb{N}$.*

Solución

Busquemos una función h tal que $E\left[h(X^2+Y^2)f(X^2+Y^2)\right] = E\left[f(X^2+Y^2)X^n\right]$ para cualquier función boreliana f acotada.

Como X^2+Y^2 tiene distribución exponencial con parámetro $\lambda = \frac{1}{2}$, se tiene:

$E\left[h(X^2+Y^2)f(X^2+Y^2)\right] = \frac{1}{2}\int_0^\infty h(z)f(z)e^{-\frac{z}{2}}dz.$

$E\left[f(X^2+Y^2)X^n\right] = \frac{1}{2\pi}\int_{-\infty}^\infty \int_{-\infty}^\infty f(x^2+y^2)x^n e^{-\frac{1}{2}(x^2+y^2)}dxdy$

$= \frac{1}{2\pi}\int_0^\infty \int_0^{2\pi} f(r^2)r^n \cos^n\theta e^{-\frac{1}{2}r^2} r d\theta dr = \frac{1}{2\pi}\int_0^\infty \int_0^{2\pi} r^{n+1}f(r^2)e^{-\frac{1}{2}r^2}\cos^n\theta d\theta dr$

$= C\int_0^\infty r^{n+1}f(r^2)e^{-\frac{1}{2}r^2}dr = \frac{C}{2}\int_0^\infty z^{\frac{1}{2}n}f(z)e^{-\frac{1}{2}z}dz$

donde $C = \frac{1}{2\pi}\int_0^{2\pi}\cos^n\theta d\theta$.

Por lo tanto, $h(z) = Cz^{\frac{1}{2}n}$. Es decir:

$E\left[X^n \mid X^2+Y^2\right] = C\left(X^2+Y^2\right)^{\frac{n}{2}}$

Para n impar, se tiene $\int_0^{2\pi}\cos^n\theta d\theta = 0$.

Para n par, se tiene $\int \cos^n\theta d\theta = \frac{1}{n}\cos^{n-1}\theta \operatorname{sen}\theta + \frac{n-1}{n}\int\cos^{n-2}\theta d\theta$. Así que:

$\int_0^{2\pi}\cos^n\theta d\theta = \frac{n-1}{n}\int_0^{2\pi}\cos^{n-2}\theta d\theta = \cdots = \frac{1\cdot 3\cdots(n-1)}{2\cdot 4\cdots n}2\pi = \frac{n!}{2^n\left[\left(\frac{n}{2}\right)!\right]^2}2\pi$

Por lo tanto:

$$E\left[X^n \mid X^2+Y^2\right] = \begin{cases} \frac{n!}{2^n\left[\left(\frac{n}{2}\right)!\right]^2}\left(X^2+Y^2\right)^{\frac{n}{2}} & \textit{si } n \textit{ es par} \\ 0 & \textit{si } n \textit{ es impar} \end{cases}$$

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 4 (**Propiedades de la esperanza condicional**). *Sean X y Y dos variables aleatorias de esperanza finita, $\mathcal{G}$ y $\mathcal{H}$ dos subσ-álgebras de $\Im$ 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) i 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, *ii*, *iii* y *iv* son inmediatas de la definición de esperanza condicional.

v también se sigue inmediatamente de la definición de esperanza condicional; en efecto, sabemos que $\int_B E[X \mid \mathcal{G}]\, dP = \int_B X dP$ para cualquier conjunto $B \in \mathcal{G}$, así que, en particular, se tiene:

$$E[E[X \mid \mathcal{G}]] = \int_\Omega E[X \mid \mathcal{G}]\, dP = \int_\Omega X dP = E[X]$$

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:

$$\int_B Y dP = \int_B Z dP = \int_B X dP$$

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:

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

Así que $E[X]$ 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 $E[|X| \mid \mathcal{G}] \geq E[X \mid \mathcal{G}]$ y $E[|X| \mid \mathcal{G}] \geq -[X \mid \mathcal{G}]$; por lo tanto, $E[|X| \mid \mathcal{G}] \geq |E[X \mid \mathcal{G}]|$.

$\blacksquare$

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

Proposición 5. *Sea X es una variable aleatoria de esperanza finita, $\mathcal{G}$ una subσ-álgebra de $\Im$ 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

Observese primero que, en esta demostración, X y Z son variables aleatorias fijas. Lo que queremos demostrar es que la relación $E[WZ] = E[WX]$ se cumple para cualquier variable aleatoria acotada W $\mathcal{G}$-medible.

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

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

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

Así que se cumple la relación para φ.

Si W es una variable aleatoria acotada no negativa $\mathcal{G}$-medible, existe una sucesión no decreciente de variables aleatorias simples no negativas $(\varphi_n)_{n\in\mathbb{N}}$ tal que la sucesión $(\varphi_n(\omega))_{n\in\mathbb{N}}$ converge a $W(\omega)$ para cualquier $\omega \in \Omega$. Por lo tanto, las sucesiones $(\varphi_n Z)_{n\in\mathbb{N}}$ y $(\varphi_n X)_{n\in\mathbb{N}}$ son también no decrecientes y convergen puntualmente a WZ y WX, respectivamente. Así que, por el teorema de la convergencia monótona, se tiene:

$E[WZ] = E[\lim_{n\to\infty} \varphi_n Z] = \lim_{n\to\infty} E[\varphi_n Z] = \lim_{n\to\infty} E[\varphi_n X]$

$= E[\lim_{n\to\infty} \varphi_n X] = E[WX]$

Finalmente, si W es una variable aleatoria acotada $\mathcal{G}$-medible cualquiera, W^+ y W^- son también variables aleatorias acotadas $\mathcal{G}$-medibles. Así que, siendo no negativas, se tiene:

$$E\left[W^+Z\right] = E\left[W^+X\right]$$

$$E\left[W^-Z\right] = E\left[W^-X\right]$$

Por lo tanto:

$$E\left[WZ\right] = E\left[W^+Z\right] - E\left[W^-Z\right] = E\left[W^+X\right] - E\left[W^-X\right] = E\left[WX\right]$$

■

Proposición 6. *Sea $\mathcal{G}$ una subσ-álgebra de $\Im$, X una variable aleatoria de esperanza finita y Y una variable aleatoria $\mathcal{G}$-medible tal que YX tiene esperanza finita, entonces $YE\left[X \mid \mathcal{G}\right]$ tiene esperanza finita y $E\left[YX \mid \mathcal{G}\right] = YE\left[X \mid \mathcal{G}\right]$.*

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 $E\left[WZ \mid \mathcal{G}\right] = E\left[WX\right]$ 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 no decreciente, $(W_n)_{n\in\mathbb{N}}$ de variables aleatorias simples no negaivas $\mathcal{G}$-medibles, así que, para cada $n \in \mathbb{N}$ y $B \in \mathcal{G}$, se tiene:

$$\int_B W_nZdP = \int_B W_nXdP$$

Uilizando el eorema de la convergencia monótona, 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.

■

Proposición 7 (Teorema de la convergencia monótona). *Sea $\mathcal{G}$ una sub σ-álgebra de $\Im$ y $\{X_n\}_{n\in\mathbb{N}}$ una sucesión monótona creciente de variables aleatorias de esperanza finita tales que $X = \lim_{n\to\infty} X_n$ tiene esperanza finita. Entonces:*

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

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 $Z^* = \lim_{n\to\infty} Z_n$ 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. ∎

Corolario 1. *Sea $\mathcal{G}$ una sub σ-álgebra de $\Im$ 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 8 (Lema de Fatou). *Sea $\mathcal{G}$ una sub σ-álgebra de $\Im$ y $\{X_n\}_{n\in\mathbb{N}}$ una sucesión de variables aleatorias no negativas de esperanza finita tales que* $\liminf X_n$ *tiene esperanza finita. Entonces:*

$$E\left[\liminf X_n \mid \mathcal{G}\right] \leq \liminf E\left[X_n \mid \mathcal{G}\right]$$

Demostración

La sucesión $Y_n = \inf\{X_j : j \geq n\}$ es monótona creciente y $\liminf X_n = \lim_{n\to\infty} Y_n$, así que, por la proposición 7, $E\left[\liminf X_n \mid \mathcal{G}\right] = \lim_{n\to\infty} E\left[Y_n \mid \mathcal{G}\right]$.

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,

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

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

Proposición 9. *Sea* $\mathcal{G}$ *una sub* σ*-álgebra de* $\Im$*,* 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* $\lim_{n\to\infty} E\left[|X_n - X| \mid \mathcal{G}\right] = 0$*, 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 8, se tiene

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

$\leq \liminf E\left[Z_n \mid \mathcal{G}\right] = 2E\left[Y \mid \mathcal{G}\right] - \limsup E\left[|X_n - X| \mid \mathcal{G}\right]$.

Así que,

$$\limsup E\left[|X_n - X| \mid \mathcal{G}\right] = 0$$

∎

Corolario 2 (**Teorema de la convergencia dominada**)**.** *Sea* $\mathcal{G}$ *una sub* σ*-álgebra de* $\Im$*,* 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:*

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