Miscelánea Matemática **39** (2004) 17–30 SMM

¿Qué es la esperanza condicional?

Luis Rincón
Departamento de Matemáticas
Facultad de Ciencias, UNAM
Circuito Exterior de CU
04510 México D.F.
México
lars@fciencias.unam.mx

Resumen

Partiendo de conceptos elementales de probabilidad y desde un punto de vista constructivo, presentamos el concepto de esperanza condicional de una variable aleatoria respecto de otra variable aleatoria. Esto nos da pauta para la definición más general de esperanza condicional respecto de una σ-álgebra. Mencionamos algunas propiedades generales y algunos de sus usos e interpretaciones.

Hace algunos años cuando por fortuna asistí de manera informal a un curso de probabilidad impartido de manera altruista y generosa por el profesor Juan Ruiz de Chávez, encontré la definición formal del concepto de esperanza condicional. Debo confesar que tal definición no me era muy clara tomando como base los conceptos elementales de probabilidad. ¿Es la esperanza condicional una esperanza? ¿o es en realidad una probabilidad condicional? A raíz de estas preguntas es que surge el presente artículo en donde intentamos presentar de una manera más natural el concepto de esperanza condicional así como algunas de sus propiedades básicas. Iniciamos explicando la situación en el caso de variables aleatorias discretas, después damos la definición general junto con algunas de sus propiedades y finalmente mencionamos algunos contextos en los cuales surge o es usada esta esperanza.

1. Algunos conceptos básicos

El modelo matemático que se ha creado para estudiar de una manera científica los tan comunes y cotidianos fenómenos azarosos de la naturaleza y a los que nos vemos expuestos todos los días, es el así llamado espacio de probabilidad. Éste consta de una terna denotada regularmente por $(\Omega, \mathcal{A}, P)$, en donde Ω es el conjunto de todos los posibles resultados del experimento aleatorio en cuestión y $\mathcal{A}$ es una clase no vacía de subconjuntos de Ω que es cerrada bajo las operaciones de tomar complementos y obtener uniones numerables. En la colección $\mathcal{A}$ se colocan todos los eventos de interés del experimento aleatorio y a quienes deseamos calcular su probabilidad. Finalmente P es una medida de probabilidad, es decir, una función real con dominio la colección $\mathcal{A}$, y que cumple con ser no negativa, le asigna el valor uno al espacio total Ω y finalmente se le pide ser σ-aditiva, es decir, la probabilidad de cualquier unión numerable y ajena de eventos dos a dos es la suma de las probabilidades de todos los eventos.

El modelo anterior es por supuesto una simplificación extrema de una situación o sistema de cosas mucho mas complejo como lo es un fenómeno aleatorio, pero ha sido bien utilizado desde el primer tercio del siglo XX cuando se creó la teoría axiomática de la probabilidad. No es nuestra intención cuestionar este estado de cosas sino recordar brevemente algunos conceptos elementales de la teoría establecida.

Podemos por ejemplo modelar matemáticamente el experimento aleatorio de lanzar un dado y observar la cara superior una vez que el dado cae. Naturalmente el conjunto de todos los posibles resultados es el conjunto $\Omega = \{1, 2, 3, 4, 5, 6\}$. En casos simples como este podemos tomar a $\mathcal{A}$ como el conjunto potencia de Ω, es decir 2^{Ω}, y definir la función P de la forma clásica. Para cualquier $A \subseteq \Omega$, definimos $P(A) = \#A/\#\Omega$. De este modo podemos calcular la probabilidad de ocurrencia de cualquier evento $A \subseteq \Omega$. Pero a menudo existen situaciones en donde tenemos información adicional del experimento aleatorio y esta información afecta el cálculo de las probabilidades.

Imaginemos por ejemplo la situación de desear adivinar el resultado que se obtiene al lanzar un dado equilibrado. Podríamos apostar a que se obtendrá, por ejemplo, el número "2". La probabilidad de ganar la apuesta es obviamente $1/6$. Ahora, supongamos que un "informante" nos provee de la información de que un número par será obtenido. Claramente nuestras probabilidades de ganar, apostando por el número "2" nuevamente, serán ahora de $1/3$. Este tipo de situaciones son capturadas por el concepto de probabilidad condicional de un evento A

dado otro evento B, denotada y definida como sigue

$$P(A|B) = \frac{P(A \cap B)}{P(B)}.$$

La probabilidad condicional de un evento dado otro evento es entonces un concepto que nos permite interaccionar con la realidad del fenómeno azaroso al incorporar información adicional del mismo y actualizar el cálculo de probabilidades.

Otro de los conceptos clave en el lenguaje moderno de la teoría de la probabilidad es el de variable aleatoria. Una variable aleatoria es simplemente una función $X : \Omega \to \mathbb{R}$ tal que para cualquier intervalo abierto (a, b) de $\mathbb{R}$, la imagen inversa $X^{-1}(a, b)$ es un elemento de la σ-álgebra $\mathcal{A}$, dominio de definición de la medida de probabilidad P. La medida de probabilidad P se traslada entonces a los intervalos abiertos de $\mathbb{R}$ mediante la siguiente regla. La probabilidad asignada al intervalo (a, b) es simplemente el número $P(X^{-1}(a, b))$.

Ahora, dada una variable aleatoria X, uno puede calcular ciertos números asociados a X. Estos números son los llamados momentos que se denotan y calculan como sigue. Para cada número natural n,

$$E(X^n) = \int_\Omega X^n(\omega) dP(\omega),$$

suponiendo por supuesto que la correspondiente integral existe. Estos números son mediciones de ciertas características de la función X y bajo ciertas condiciones la identifican de manera única. Véase el artículo [1] para un bosquejo histórico de algunos conceptos básicos de probabilidad. Con estos elementos preliminares estamos en posición de dar una primera aproximación a la esperanza condicional.

2. Caso discreto

Consideremos primeramente el caso sencillo en donde tenemos dos variables aleatorias discretas X y Y con posibles valores $x_1, \ldots, x_n$, y $y_1, \ldots, y_m$ respectivamente. Comúnmente tenemos información de ambas variables aleatorias a un mismo tiempo a través de la función de densidad conjunta

$$f_{X,Y}(x_i, y_j) = P(X = x_i, Y = y_j).$$

Definición. La esperanza condicional de la variable aleatoria X dado el evento $(Y = y_j)$ se calcula como sigue

$$E(X|Y = y_j) = \sum_{i=1}^{n} x_i P(X = x_i|Y = y_j).$$

Este es un número que depende del evento $(Y = y_j)$. Podemos considerar entonces que este valor está en función del número y_j, o bien considerar que tenemos una función del espacio muestral Ω de la siguiente forma. Sea ω en Ω tal que $Y(\omega) = y_j$. Definimos entonces

$$E(X|Y)(\omega) = E(X|Y = y_j).$$

De este modo hemos construido una función $E(X|Y) : \Omega \to \mathbb{R}$ y nos preguntamos si tal función es en realidad una variable aleatoria.

Medibilidad. La respuesta a la pregunta anterior es afirmativa y explicamos a continuación las razones. A través de sus m posibles valores la variable aleatoria Y secciona el espacio muestral Ω en m eventos disjuntos, a saber, $(Y = y_1), \ldots, (Y = y_m)$. Sobre cada uno de estos eventos la función $E(X|Y)$ es constante. Por ejemplo, sobre el evento $(Y = y_j)$ el valor constante es

$$z_j = E(X|Y = y_j) = \sum_{i=1}^{n} x_i P(X = x_i|Y = y_j).$$

Globalmente podemos escribir $E(X|Y)$ en términos de funciones indicadoras como sigue

$$E(X|Y)(\omega) = \sum_{j=1}^{m} E(X|Y = y_j) 1_{(Y=y_j)}(\omega).$$

La mínima σ-álgebra respecto de la cual la función Y es variable aleatoria es aquella generada por la partición $\{(Y = y_1), \ldots, (Y = y_m)\}$ y denotada por $\sigma(Y)$. Esta σ-álgebra consiste de 2^m eventos distintos y como Y es variable aleatoria entonces $\sigma(Y) \subseteq \mathcal{A}$. Por ejemplo, en el caso en el que Y es una variable aleatoria constante $Y = c$, se tiene que $\sigma(Y) = \{\emptyset, \Omega\}$.

Como $E(X|Y)$ es constante en cada elemento de la partición resulta que la función $E(X|Y)$ es $\sigma(Y)$-medible y en consecuencia es verdaderamente una variable aleatoria. Observe que $E(X|Y)$ toma a lo sumo tantos valores distintos como lo hace Y.

Propiedad fundamental. Dadas las consideraciones anteriores demostraremos a continuación una primera y fundamental propiedad que satisface la esperanza condicional: Para cualquier evento F de la σ-álgebra $\sigma(Y)$ se cumple la igualdad

$$\int_F E(X|Y)(\omega)dP(\omega) = \int_F X(\omega)dP(\omega). \tag{1}$$

Como cada elemento de $\sigma(Y)$ es una unión ajena de elementos de la forma $(Y = y_j)$ entonces por propiedades de la integral es suficiente demostrar (1) para estos eventos simples. Tenemos entonces que

$$\begin{aligned}
\int_{(Y=y_j)} E(X|Y)(\omega)dP(\omega) &= z_j P(Y = y_j) \\
&= \sum_{i=1}^{n} x_i P(X = x_i | Y = y_j) P(Y = y_j) \\
&= \sum_{i=1}^{n} x_i P(X = x_i, Y = y_j) \\
&= \int_{(Y=y_j)} X(\omega)dP(\omega)
\end{aligned}$$

Comentarios. De esta forma hemos definido la esperanza condicional $E(X|Y)$ para dos variables aleatorias discretas y hemos visto que es una variable aleatoria $\sigma(Y)$-medible con esperanza finita y que cumple la igualdad (1). A la variable aleatoria $E(X|Y)$ se le denota también por $E(X|\sigma(Y))$. El argumento anterior es también válido cuando tanto X como Y toman una cantidad numerable de valores. Observemos la diferencia entre $E(X|Y = y_j)$ y $E(X|Y)$. El primero término es un posible valor numérico del segundo término que es una variable aleatoria, sin embargo a ambas expresiones se les llama esperanza condicional.

Ejemplo. Veamos un caso particular y revelador. Sean A y B dos eventos cualesquiera con $0 < P(B) < 1$. Supongamos que la variable aleatoria X es la función indicadora del evento A y $\mathcal{F}$ es la σ-álgebra $\{\emptyset, \Omega, B, B^c\}$. En otras palabras, $\mathcal{F} = \sigma(Y)$ con Y la función indicadora del evento B. Calcularemos $E(X|\mathcal{F})$ que también podemos escribir como $E(1_A|1_B)$. Primeramente recordemos que $E(X|\mathcal{F})$ es $\mathcal{F}$-medible. Por lo tanto $E(X|\mathcal{F})$ es constante tanto en B como en B^c. Además $E(X|\mathcal{F})$ debe satisfacer la igualdad (1) para cualquier F en $\mathcal{F}$. Tomando el caso particular $F = B$ obtenemos

$$\int_B E(X|\mathcal{F})(\omega)dP(\omega) = \int_B I_A(\omega)dP(\omega) = P(A \cap B).$$

Como $E(X|\mathcal{F})$ es constante sobre B tenemos que para cualquier ω en B,

$$E(X|\mathcal{F})(\omega) = \frac{P(A \cap B)}{P(B)} = P(A|B).$$

Análogamente, tomando el caso $F = B^c$ obtenemos para cualquier ω en B^c,

$$E(X|\mathcal{F})(\omega) = \frac{P(A \cap B^c)}{P(B^c)} = P(A|B^c).$$

De modo que haciendo uso de las funciones indicadoras podemos escribir

$$E(X|\mathcal{F}) = P(A|B)1_B + P(A|B^c)1_{B^c}.$$

Lo anterior sugiere que la esperanza condicional puede verse como una generalización del concepto básico de probabilidad condicional. También puede considerarse como una generalización del concepto de esperanza pues dada la discusión anterior, no es difícil verificar que para cualquier variable aleatoria X y para cualquier evento A,

1. $E(X \mid \{\emptyset, \Omega\}) = E(X)$.
2. $E(1_A \mid \{\emptyset, \Omega\}) = P(A)$.

El hecho es que la σ-álgebra $\mathcal{F} = \{\emptyset, \Omega\}$ realmente no proporciona ninguna información adicional del experimento aleatorio y por lo tanto la esperanza se calcula directamente sobre la variable aleatoria. A menudo se usa el término $P(A|\mathcal{F})$ para indicar la esperanza $E(1_A|\mathcal{F})$.

Podemos ahora retomar el ejemplo consistente en apostar en adivinar el resultado de lanzar un dado equilibrado. Supongamos que insistimos en apostar a que se obtendrá el número "2". Definamos entonces los eventos $A = \{2\}$ y $B = \{2, 4, 6\}$. Sea $X = 1_A$ y tomemos $\mathcal{F} = \{\emptyset, \Omega, B, B^c\}$. Esta σ-álgebra puede "distinguir" los resultados "Cae número par", evento B, y "Cae número impar", evento B^c, de modo que nuestro "informante" nos reportará alguno de estos dos resultados después de haber sido lanzado el dado. Entonces

$$\begin{aligned} E(X|\mathcal{F}) &= P(A|B)1_B(\omega) + P(A|B^c)1_{B^c}(\omega) \\ &= \frac{1}{3} \cdot 1_B(\omega) + 0 \cdot 1_{B^c}(\omega) \\ &= \begin{cases} \frac{1}{3} & \text{si } \omega \in B, \\ 0 & \text{si } \omega \notin B. \end{cases} \end{aligned}$$

De esta forma $E(X|\mathcal{F})$ es una función que reporta nuestras probabilidades de ganar apostando por el número "2" en cada una de las dos situaciones que la σ-álgebra $\mathcal{F}$ o nuestro "informante" distinguen: Resultado par o impar.

En cambio, si tomamos $\mathcal{F} = 2^{\Omega}$ con $\Omega = \{1, 2, 3, 4, 5, 6\}$, es decir, la σ-álgebra "distingue" totalmente el resultado del lanzamiento del dado y nuestro "informante" nos dice exactamente el resultado del experimento, entonces definiendo los eventos $B_i = \{i\}$, para $i = 1, \ldots, 6$, tenemos

$$E(X|\mathcal{F}) = \sum_{i=1}^{6} P(A|B_i)1_{B_i} = P(A|A)1_A = 1_A = X,$$

es decir, no hay sorpresas, cuando sabemos con precisión el resultado del experimento, conocemos obviamente la probabilidad de ganar apostando por el numero "2", ésta es 0 ó 1.

3. Definición general

Hemos definido entonces el término $E(X|Y)$ cuando X y Y son variables aleatorias que toman un número finito de valores. Ahora consideraremos cualquier variable aleatoria X, con la única condición de que sea integrable, y en lugar de Y o $\sigma(Y)$, consideraremos una sub σ-álgebra cualquiera $\mathcal{F} \subseteq \mathcal{A}$. A través del siguiente resultado se define el término más general $E(X|\mathcal{F})$.

Teorema 1. *Sea $(\Omega, \mathcal{A}, P)$ un espacio de probabilidad y sea X una variable aleatoria con esperanza finita. Sea $\mathcal{F}$ una sub σ-álgebra de $\mathcal{A}$. Entonces existe una variable aleatoria denotada por $E(X|\mathcal{F}) : \Omega \to \mathbb{R}$ tal que*

1. *$E(X|\mathcal{F})$ es $\mathcal{F}$-medible.*
2. *$E(X|\mathcal{F})$ tiene esperanza finita.*
3. *Para cada F en $\mathcal{F}$,*

$$\int_F E(X|Y)(\omega)dP(\omega) = \int_F X(\omega)dP(\omega). \tag{2}$$

Además, la variable aleatoria $E(X|\mathcal{F})$ es única en el siguiente sentido. Si existe otra variable aleatoria $\tilde{X}$ con las tres propiedades anteriores, entonces $\tilde{X} = E(X|\mathcal{F})$ casi seguramente, es decir, $P(\tilde{X} = E(X|\mathcal{F})) = 1$.

No presentaremos aquí la demostración del teorema anterior, que en realidad es consecuencia del teorema de Radon-Nikodym. Fue demostrado por Kolmogorov en 1933 y es la forma en la que aparece como definición en casi cualquier libro de probabilidad que trate el tema.

Después de una primera lectura, la definición puede parecer un tanto oscura pues ésta no nos provee de una expresión explícita para esta esperanza, como en el caso discreto mencionado antes. Únicamente conocemos parte de su comportamiento a través de las tres propiedades que la definen. Sin embargo, estas tres condiciones no deben parecernos ahora arbitrarias o extrañas tomando en consideración nuevamente el caso discreto.

El problema planteado al inicio de nuestro trabajo y que se resume en la pregunta ¿Cómo concebir intuitivamente este concepto llamado "esperanza" cuando en realidad es una variable aleatoria?, fue parcialmente resuelto pues pudimos darnos cuenta, en el caso discreto, que la variable aleatoria $E(X|\mathcal{F})$ puede, en algunos casos, tomar los valores $E(X)$ y $P(A|B)$. Para comprender más profundamente a la esperanza condicional nos propondremos ahora como objetivo descubrir algunas otras de sus propiedades. Para ello asumiremos hipótesis adicionales sobre X o sobre $\mathcal{F}$ y veremos cómo cambia la esperanza. Más aún, en la última sección de este trabajo veremos algunos usos e interpretaciones interesantes de este concepto.

4. Primeras propiedades elementales

Estudiaremos ahora algunas propiedades elementales y bien conocidas de la esperanza condicional. La mayoría de estas propiedades se desprenden de las propiedades de la integral de Lebesgue y de la importante propiedad (2).

a) Primeramente observemos que X y $E(X|\mathcal{F})$ son dos variables aleatorias con la misma esperanza, es decir, $E(E(X|\mathcal{F})) = E(X)$. Esta propiedad se sigue directamente de la igualdad (2) tomando el caso particular cuando el conjunto F es la totalidad Ω.

b) Cuando X es $\mathcal{F}$-medible entonces X mismo cumple con las tres condiciones que caracterizan a $E(X|\mathcal{F})$. Por la unicidad tenemos entonces que casi seguramente $E(X|\mathcal{F}) = X$.

c) Debido a la propiedad lineal de las integrales tenemos la propiedad análoga
$$E(aX + Y|\mathcal{F}) = aE(X|\mathcal{F}) + E(Y|\mathcal{F}).$$

d) La esperanza condicional preserva la no negatividad, es decir, si $X \geq 0$ entonces $E(X|\mathcal{F}) \geq 0$. Para demostrar esta propiedad procedemos por contradicción. Supongamos entonces que $P(E(X|\mathcal{F}) < 0) > 0$ y definamos el evento $A_n = (E(X|\mathcal{F}) < -1/n)$. Entonces forzosamente $P(A_n) > 0$ para algún número natural n. Por lo tanto

$$\int_{A_n} X dP = \int_{A_n} E(X|\mathcal{F}) dP < -\frac{1}{n} P(A_n) < 0.$$

Esto contradice la hipótesis $X \geq 0$.

Éstas y otras propiedades aparecen resumidas más adelante. Veamos ahora algunas otras propiedades un poco más avanzadas.

5. Otras propiedades

Mencionamos a continuación algunas otras propiedades y omitimos la demostración correspondiente. Nuestro objetivo es únicamente ilustrar el comportamiento de la esperanza condicional.

e) Primeramente mencionaremos la desigualdad de Jensen. Esta desigualdad establece que si h es una función convexa tal que la variable aleatoria $h(X)$ es integrable entonces casi seguramente

$$h[E(X|\mathcal{F})] \leq E(h(X)|\mathcal{F}).$$

Así por ejemplo, para la función convexa $h(x) = e^x$ obtenemos la desigualdad $e^{E(X|\mathcal{F})} \leq E(e^X|\mathcal{F})$. Otro caso particularmente interesante se obtiene cuando se toma la función convexa $h(x) = |x|^n$ con n un número natural. La desigualdad de Jensen establece entonces que $|E(X|\mathcal{F})|^n \leq E(|X|^n|\mathcal{F})$. Tomando esperanzas de ambos lados obtenemos la desigualdad $E(|E(X|\mathcal{F})|^n) \leq E(E(|X|^n|\mathcal{F})) = E(|X|^n)$. En palabras, lo anterior dice que la variable aleatoria $E(X|\mathcal{F})$ tiene momentos menores o iguales a los de X.

f) Otra propiedad interesante es aquella que se obtiene cuando suponemos que tenemos dos sub σ-álgebras $\mathcal{F}_1$ y $\mathcal{F}_2$ que guardan la relación $\mathcal{F}_1 \subseteq \mathcal{F}_2$. Bajo estas condiciones $E(X|\mathcal{F}_1)$ es entonces $\mathcal{F}_2$-medible y por lo tanto

$$E(E(X|\mathcal{F}_1)|\mathcal{F}_2) = E(X|\mathcal{F}_1) \quad \text{c.s.}$$

g) Si Y es una variable aleatoria $\mathcal{F}$-medible y acotada entonces

$$E(XY|\mathcal{F}) = YE(X|\mathcal{F}) \text{ c.s.}$$

h) Finalmente, si la variable aleatoria X es independiente de la σ-álgebra $\mathcal{F}$ entonces

$$E(X|\mathcal{F}) = E(X) \text{ c.s.}$$

A manera de resumen presentamos a continuación una lista parcial de propiedades de la esperanza condicional.

Algunas propiedades de la esperanza condicional $E(X\|\mathcal{F})$
1. Existe y es única c.s.
2. Es $\mathcal{F}$-medible e integrable .
3. Para cualquier $F \in \mathcal{F}$, $\int_F E(X\|\mathcal{F})dP = \int_F XdP$.
4. $E(1_A\|1_B) = P(A\|B)1_B + P(A\|B^c)1_{B^c}$
5. $\mathcal{F} = \{\emptyset, \Omega\} \Rightarrow E(X\|\mathcal{F}) = E(X)$.
6. $E(E(X\|\mathcal{F})) = E(X)$
7. X es $\mathcal{F}$-medible $\Rightarrow E(X\|\mathcal{F}) = X$ c.s.
8. $E(aX + Y\|\mathcal{F}) = aE(X\|\mathcal{F}) + E(Y\|\mathcal{F})$ c.s.
9. $X \geq 0 \Rightarrow E(X\|\mathcal{F}) \geq 0$ c.s.
10. Jensen: h convexa $\Rightarrow h[E(X\|\mathcal{F})] \leq E(h(X)\|\mathcal{F})$ c.s.
11. $\mathcal{F}_1 \subseteq \mathcal{F}_2 \Rightarrow E(E(X\|\mathcal{F}_1)\|\mathcal{F}_2) = E(X\|\mathcal{F}_1)$ c.s.
12. Y es $\mathcal{F}$-medible y acotada $\Rightarrow E(XY\|\mathcal{F}) = YE(X\|\mathcal{F})$ c.s.
13. X independiente de $\mathcal{F} \Rightarrow E(X\|\mathcal{F}) = E(X)$ c.s.

Éstas y otras propiedades interesantes pueden encontrarse en [3] o [5]. Por ejemplo, puede demostrarse que la esperanza condicional es "continua" cuando tomamos operaciones límite tanto en X como en $\mathcal{F}$, bajo algunas condiciones usuales del límite, por ejemplo monotonía.

6. Algunos usos de la esperanza condicional

En esta sección final mencionaremos algunas interpretaciones y contextos en donde surge, a veces de manera inesperada, la esperanza condicional. El material siguiente es ligeramente avanzado y el lector requerirá necesariamente de conocimientos más allá de los elementales de probabilidad básica. Ningún lector debe sentirse amendrentado por este comentario, mejor aún, se le invita a profundizar en el tema siguiendo alguna de las referencias que se proporcionan en cada sección.

6.1. Esperanza condicional y predicción

Comúnmente existen situaciones en donde a través de una variable X, o una transformación de ella, deseamos predecir o explicar el comportamiento de otra variable Y. Supongamos que $g(X)$ es la función utilizada para predecir Y. Entonces uno desea encontrar la expresión de g tal que el error cuadrático medio definido por el número $E[(Y - g(X))^2]$ sea mínimo.

Se puede demostrar que la esperanza condicional cumple con estas condiciones, es decir, de entre todas las posibles expresiones de $g(X)$ para predecir Y, la mejor en el sentido de tener el error cuadrático medio más pequeño es la función $E(Y|X)$, es decir,

$$E[(Y - g(X))^2] \geq E[(Y - E(Y|X))^2].$$

La demostración de este hecho puede ser encontrada en el capítulo 8 de [3].

6.2. La esperanza condicional es una proyección

Podemos considerar a la esperanza condicional como un operador $X \mapsto E(X|\mathcal{F})$ que toma elementos del espacio $L^1(\Omega, \mathcal{A}, P)$, constituido por todas las variables aleatorias integrables, y las lleva a elementos del mismo espacio. Se puede demostrar que este operador es continuo, lineal y tiene norma uno.

Más específicamente, nos restringiremos al subespacio lineal y cerrado $L^2(\Omega, \mathcal{A}, P) \subset L^1(\Omega, \mathcal{A}, P)$, conformado por todas aquellas variables aleatorias cuadrado integrables, es decir, que tienen segundo momento finito. En realidad este subespacio es un espacio de Hilbert con producto interior dado por la siguiente expresión

$$\langle X, Y \rangle = \int_\Omega X(\omega) Y(\omega)\, dP(\omega).$$

Este producto proporciona una estructura geométrica al espacio y nos permite hablar de ortogonalidad entre variables aleatorias y en particular de proyecciones sobre subespacios.

Denotemos por $X \mapsto \Pi_{\mathcal{F}}(X)$ a la proyección del espacio $L^2(\Omega, \mathcal{A}, P)$ al subespacio $L^2(\Omega, \mathcal{F}, P)$ en donde $\mathcal{F} \subseteq \mathcal{A}$. Observe que la única diferencia entre los dos espacios es la σ-álgebra usada. ¿Cómo es este mapeo proyección? Se puede demostrar, y aqui viene la sorpresa, que tal proyección adquiere la forma de la esperanza condicional, es decir,

$$\Pi_{\mathcal{F}}(X) = E(X|\mathcal{F}).$$

Véase [3] o [5] para un recuento más detallado de esta forma de interpretar la esperanza condicional en el espacio de Hilbert $L^2(\Omega, \mathcal{A}, P)$.

6.3. Procesos estocásticos

En esta pequeña sección veremos cómo el término de esperanza condicional aparece en el lenguaje de los procesos estocásticos. Un proceso estocástico es simplemente una colección de variables aleatorias $\{X_t : t \geq 0\}$ indexadas por el parámetro $t \geq 0$ usualmente interpretado como el tiempo. Este es un modelo a través del cual se pretende representar fenómenos aleatorios que se desarrollan en el tiempo. Para cada $t \geq 0$ la variable aleatoria X_t representa el valor de la variable de interés al tiempo t. Los distintos tipos de procesos se obtienen al establecer los grados de dependencia o independencia de las variables aleatorias que conforman el proceso estocástico. Véase por ejemplo [2] para profundizar sobre este tema. Antes de presentar dos ejemplos de procesos estocásticos particulares, definamos para cada $t \geq 0$ la σ-álgebra

$$\mathcal{F}_t = \sigma\{X_s : 0 \leq s \leq t\}.$$

La colección $\mathcal{F}_t$ comúnmente se interpreta como la historia del proceso hasta el tiempo t. Es decir, en $\mathcal{F}_t$ están contenidos todos los posibles eventos o sucesos generados por el proceso en el intervalo $[0, t]$.

Procesos de Markov. Decimos que el proceso $\{X_t : t \geq 0\}$ es un proceso de Markov si para cada par de tiempos s y t tales que $0 \leq s < t$, se verifica la igualdad

$$P(X_t \in A \mid \mathcal{F}_s) = P(X_t \in A \mid X_s).$$

Esta condición puede parafrasearse del siguiente modo. El comportamiento futuro (tiempo t) del proceso dado la historia del mismo hasta el tiempo s no depende del pasado remoto (tiempos antes de s) sino únicamente del pasado inmediato (tiempo s). Los procesos de Markov constituyen una clase importante de procesos estocásticos por su amplia aplicación.

Martingalas. Otro tipo de proceso estocástico que puede definirse usando el concepto de esperanza condicional son las martingalas. Decimos que el proceso estocástico $\{X_t : t \geq 0\}$ es una martingala si para cualesquiera tiempos s y t con $0 \leq s < t$ se cumple la igualdad

$$E(X_t | \mathcal{F}_s) = X_s \quad \text{c.s.}$$

Nuevamente podemos dar una interpretación de esta ecuación del siguiente modo. El valor promedio del proceso al tiempo t dada la historia del mismo en el intervalo de tiempo $[0, s]$ es el valor del proceso en el último momento observado, es decir, X_s. Una de las interpretaciones más claras del concepto de martingala se logra en términos de juegos de apuestas justas. Si X_t representa el capital de un apostador al tiempo t en una sucesión continua de apuestas justas, entonces en promedio el capital del jugador al tiempo futuro t será el capital que tuvo al tiempo s. Es decir, en promedio el jugador no gana ni pierde y en este sentido el juego es justo.

6.4. Teorema ergódico de Birkhoff

Para concluir este trabajo mencionaremos un contexto mas en el que aparece la esperanza condicional.

Consideremos un espacio de probabilidad $(\Omega, \mathcal{A}, P)$ junto con una transformación $T : \Omega \to \Omega$ que es $\mathcal{A}$-medible y preservadora de la medida P. Esto último quiere decir que para cualquier conjunto A de la clase $\mathcal{A}$ se cumple la igualdad $P(T^{-1}A) = P(A)$.

Supongamos que tenemos una variable aleatoria $X : \Omega \to \mathbb{R}$ que es integrable. Entonces el teorema ergódico individual (o puntual) de Birkhoff de 1931 establece que existe una variable aleatoria $X^* : \Omega \to \mathbb{R}$ tal que cuando n tiende a infinito

$$\frac{1}{n} \sum_{k=0}^{n} X(T^k \omega) \to X^*(\omega), \tag{3}$$

para casi todo punto ω de Ω. Adicionalmente la variable aleatoria X^* resulta ser también integrable, de hecho $E|X^*| \leq E|X|$. Es además T-invariante, es decir, se cumple que $X^*(T\omega) = X^*(\omega)$ para casi todo punto ω.

Pero ¿quién es esta misteriosa variable aleatoria X^*? Pues resulta que si definimos la σ-álgebra $\mathcal{F}$ que consta de todos aquellos conjuntos F que son T-invariantes, es decir, que satisfacen la igualdad $T^{-1}F = F$, entonces ¿Puede el lector adivinar la respuesta? La misteriosa variable aleatoria es nuevamente la esperanza condicional, es decir,

$$X^* = E(X|\mathcal{F}).$$

En particular, cuando la medida de probabilidad P es ergódica, es decir, cuando todo subconjunto invariante bajo T tiene probabilidad cero o uno (alternativamente se dice que T es ergódica), entonces toda variable aleatoria X es independiente de $\mathcal{F}$ y por lo tanto $E(X|\mathcal{F})$ es constante $E(X)$. En este caso particular tenemos entonces que el promedio temporal (3) es igual al promedio espacial $E(X)$. Esta igualdad se conoce con el nombre de "hipótesis ergódica" y anteriormente se pensaba que los sistemas dinámicos como el que hemos mencionado satisfacían esta hipótesis. El lector interesado puede consultar [4] para profundizar sobre este tema.

7. Conclusiones

Hemos intentado presentar el concepto de esperanza condicional de una manera simple con la intención de que no parezca un objeto demasiado alejado de las ideas elementales de probabilidad. Esto es de particular importancia en el aspecto didáctico pues no es difícil encontrar situaciones en donde los términos $E(X|Y=y)$ y $E(X|Y)$ son confundidos. Adicionalmente y con el fin de hacer más asequible este concepto hemos ilustrado brevemente algunos ambientes en los cuales surge o es usada la esperanza condicional.

Referencias

[1] Luis Gorostiza, *La probabilidad en el siglo XX*, Miscelánea Matemática, Sociedad Matemática Mexicana **33**, Junio 2001.

[2] Samuel Karlin y Howard M. Taylor, A first course in stochastic processes. Academic Press, 1975.

[3] Alan F. Karr, Probability, Springer-Verlag, 1993.

[4] Karl Petersen, Ergodic theory, Cambridge University Press, 1983.

[5] David Williams, Probability with martingales, Cambridge University Press, 1991.