Distribuciones Condicionales y Esperanza Condicional

Pablo L. De Nápoli

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

Probabilidades y Estadística para Matemática
Segundo cuatrimestre de 2021

Probabilidad Condicional (clase 3)

Definición

Sea $(\Omega, \mathcal{E}, P)$ *un espacio de probabilidad. La probabilidad condicional* $P(A/B)$ *de un evento* A *suponiendo que ocurre el evento* B *se define por:*

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

siempre que $P(B) > 0$.

Fórmula de la probabilidad total

Proposición

Consideramos una partición del espacio muestral Ω en eventos disjuntos $B_1, B_2, \ldots B_n, \ldots$ (finita o infinita numerable)

$$\Omega = \bigcup_{k \in \mathbb{N}} B_k, \quad B_i \cap B_j = \emptyset \text{ si } i \neq j$$

con $P(B_k) > 0$ para todo k, entonces

$$P(A) = \sum_{k=1}^{\infty} P(A/B_k) \cdot P(B_k)$$

Parte I

Esperanza comdicional dado un evento

Esperanza condicional dado un evento: caso discreto

Sea $X : \Omega \to \overline{\mathbb{R}}$ una variable aleatoria discreta. Recordamos que la esperanza de X se define como la serie

$$E[X] = \sum_i x_i \cdot P\{X = x_i\}$$

donde $\mathrm{Im} = \{x_i\}$ es por hipótesis a lo sumo numerable; siempre que dicha serie sea absolutamente convergente.
En consecuencia, resulta natural definir la esperanza de X dado que ocurre el evento A de probabilidad positiva, por:

$$E[X/A] = \sum_i x_i \cdot P\{X = x_i/A\}$$

Esperanza condicional dado un evento: caso discreto

Teniendo en cuenta la definición de probabilidad condicional esto es equivalente a:

$$E[X/A] = \sum_i x_i \cdot \frac{P(\{X = x_i\} \cap A\}}{P(A)} = \frac{1}{P(A)} \sum_i x_i \cdot I_A(x_i) P\{X = x_i\}$$

Es decir que:

Otra manera de escribir la esperanza condicionada a un evento

$$E[X/A] = \frac{1}{P(A)} E[I_A X]$$

Notemos que esta fórmula puede adoptarse como definición de la esperanza condicional respecto de un evento para cualquier variable aleatoria (sea discreta o no) mientras tenga esperanza finita, y el evento A tenga probabilidad positiva.

Notamos que $E[X/A]$ es lineal como función de X,

$$E[\lambda_1 X_1 + \lambda_2 X_2/A] = \lambda_1 \cdot E[X_1/A] + \lambda_2 \cdot E[X_2/A]$$

Un ejemplo con una variable discreta (1)

Supongamos que $X \sim \mathcal{P}(\lambda)$ donde $\lambda > 0$. Recordamos que su distribución puntual viene dada por

$$p_k = P\{X = k\} = e^{-\lambda} \cdot \frac{\lambda^k}{k!} \quad k \in \mathbb{N}_0$$

y que $E[X] = \lambda$. Pero supongamos que ahora sabemos que $X \geq 1$. Entonces nuestra estimación de las probabilidades cambiará Notamos que

$$P\{X = 0\} = p_0 = e^{-\lambda} \Rightarrow P(A) = 1 - e^{-\lambda}$$

Tendremos la distribución condicional

$$P\{X = k/A\} = \begin{cases} 0 & \text{si} \quad k = 0 \\ \frac{e^{-\lambda}}{1-e^{-\lambda}} \cdot \frac{\lambda^k}{k!} & \text{si} \quad k \geq 1 \end{cases}$$

Estamos interesados en calcular $E[X/A]$ siendo $A = \{X \geq 1\}$.

Un ejemplo con una variable discreta (2)

$$E[X/A] = \frac{1}{P(A)} \sum_k x_k \cdot I_A(x_k) P\{X = x_k\}$$
$$= \frac{e^{-\lambda}}{1 - e^{-\lambda}} \sum_{k=1}^{\infty} k \frac{\lambda^k}{k!}$$

Pero haciendo un cambio de índice $j = k - 1$:

$$\sum_{k=1}^{\infty} k \frac{\lambda^k}{k!} = \sum_{k=1}^{\infty} \frac{\lambda^k}{(k-1)!} = \sum_{j=0}^{\infty} \frac{\lambda^{j+1}}{j!} = \lambda \sum_{j=0}^{\infty} \frac{\lambda^j}{j!} = \lambda e^{\lambda}$$

(esta cuenta es la misma que para calcular $E[X]$!). Nos queda:

$$E[X/A] = \frac{\lambda}{1 - e^{-\lambda}}$$

Un ejemplo con una variable continua

Supongamos que $X \sim N(0,1)$. Entonces $E[X] = 0$.
Pero supongamos que además sabemos que $X > 0$. Entonces nuestra estimación de las probabilidades cambia, y ahora estamos interesados en calcular $E[X/A]$ siendo $A = \{X > 0\}$. Notamos que

$$P(A) = \int_0^\infty \phi(x)\, dx = \frac{1}{2} \text{ donde } \phi(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$$

$$E[X/A] = \frac{1}{P(A)} E[I_A X] = \frac{1}{P(A)} E[X^+] = \frac{1}{1/2} \int_{-\infty}^\infty x^+ \cdot \phi(x)\, dx$$
$$= \frac{2}{\sqrt{2\pi}} \int_0^\infty x \cdot e^{-x^2/2}\, dx \text{ donde } x^+ = \begin{cases} x & \text{si} \quad x \geq 0 \\ 0 & \text{si} \quad x < 0 \end{cases}$$

Haciendo el cambio de variable $y = x^2/2$ vemos que:

$$E[X/A] = \sqrt{\frac{2}{\pi}} \int_0^\infty e^{-y}\, dy = \sqrt{\frac{2}{\pi}}$$

Generalización

Si X es una variable continua con densidad $f(x)$, $U \subset \mathbb{R}$ abierto y $A = \{X \in U\}$.

$$E[X/A] = \frac{1}{P(A)} E[I_A X] = \frac{1}{P(A)} E[g_U(X)]$$

$$= \frac{1}{P(A)} \int_{-\infty}^{\infty} g_U(x) f(x)\, dx = \frac{1}{P(A)} \int_U f(x)\, dx$$

donde $g_U(x) = x$ si $x \in U$ y 0 si $x \notin U$.

Parte II

Esperanza comdicional de una variable con respecto a otra: caso discreto

Esperanzas condicionales en el caso discreto

Ahora consideremos dos variables discretas $X, Y : \Omega \to \mathbb{R}$. Nos proponemos definir el concepto de esperanza condicional $E[X/Y]$ de X dada Y. Supondremos que X tiene esperanza finita.
Sean $\{y_j\}$ los distintos valores que toma la variable Y, y notemos que los eventos $A_j = \{\omega \in \Omega : Y(\omega) = y_j\}$ forman una partición del espacio muestral Ω.
Si $P\{Y = y_j\} > 0$, podemos definir

$$E[X/Y = y_j] = E[X/A_j]$$

utilizando la definición introducida anteriormente.
Más explícitamente:

$$E[X/Y = y_j] = \sum_i x_i \cdot P\{X = x_i / Y = y_j\}$$

Las probabilidades $P\{X = x_i / Y = y_j\}$ que aparecen en esta definición se llaman la distribución condicional de probabilidades de X dada Y.

Esperanzas condicionales en el caso discreto (2)

Notemos que depende del valor y_j de la variable Y. En consecuencia, $E[X/Y]$ puede considerarse como una nueva variable aleatoria. Más explícitamente, definimos $E[X/Y] : \Omega \to \overline{\mathbb{R}}$ por:

$$E[X/Y](\omega) = E[X/Y = Y(\omega)]$$

Notamos que esta variable aleatoria será constante en cada uno de los conjuntos que forma la partición A_j. En otra palabras:

$$E[X/Y] = \sum_j E[X/Y = y_j] \cdot I_{A_j}$$

Esperanza condicional e independencia

Si X e Y son variables discretas independientes, entonces

$$P\{X = x_i / Y = y_j\} = P\{X = x_i\}$$

Luego

$$E[X/Y = y_j] = \sum_i x_i \cdot P\{X = x_i/Y = y_j\} = \sum_i x_i \cdot P\{X = x_i\} = E[X]$$

En consecuencia, con lo que $E[X/Y] = E[X]$ (una variable aleatoria constante), en este caso.

Esperanzas condicionales en el caso discreto (3)

En el otro extremo, ¿qué pasa cuando $Y = f(X)$ siendo $f : \mathbb{R} \to \mathbb{R}$?

$$P\{Y = y_j / X = x_i\} = \begin{cases} 1 & \text{si} \quad y_j = f(x_i) \\ 0 & \text{si} \quad y_j \neq f(x_i) \end{cases}$$

Entonces:

$$E[Y/X = x_i] = \sum_j y_j \cdot P\{Y = y_j / X = x_i\} = f(x_i)$$

Es decir:

$$E[f(X)/X] = f(X)$$

En particular:

$$E[X/X] = X$$

Esperanzas condicionales en el caso discreto (4)

Otras propiedades útiles son:

- Linealidad: si $\lambda_1, \lambda_2 \in \mathbb{R}$,

$$E[\lambda_1 X_1 + \lambda_2 X_2 / Y] = \lambda_1 \cdot E[X_1/Y] + \lambda_2 \cdot E[X_2/Y]$$

- Más generalmente, podemos sacar afuera de la esperanza condicional funciones de la variable con respecto a la que estamos condicionando:

$$E[f(Y)X/Y] = f(Y)E[X/Y]$$

porque:

$$\begin{aligned} E[f(Y)X/Y = y_j] &= \sum_i f(y_j \cdot) \cdot x_i \cdot P\{X = x_i/Y = y_j\} \\ &= f(y_j) \sum_i x_i \cdot P\{X = x_i/Y = y_j\} = f(y_j) \cdot E[X/Y = y_j] \end{aligned}$$

Un ejemplo

En la clase 5, consideramos el siguiente ejemplo. Tiramos dos dados en forma sucesiva. Nuestro espacio muestral es:

$$\Omega = \{\omega = (\omega_1, \omega_2) : \omega_i \in D\}$$

donde $D = \{1, 2, 3, 4, 5, 6\}$. Consideramos la suma S de los puntos obbtenidos. $S : \Omega \to \mathbb{R}$ Tenemos que $S = X_1 + X_2$ donde $X_1(\omega) = \omega_1, X_2(\Omega) = \omega_2$.
Tenemos que

$$E[S] = E[X_1] + E[X_2] = 3,5 + 3,5 = 7$$

Pero si sabemos cuánto salió en la primera tirada (o sea, cuándo vale X_1), nuestra estimación de las probabilidades para S cambia.

$$E[S/X_1] = E[X_1/X_1] + E[X_2/X_1] = X_1 + E[X_2] = X_1 + 3,5$$

Fórmula de la probabilidad total

$E[X/Y]$ es una nueva variable aleatoria. ¿Qué pasa si calculamos su esperanza? Recordamos que $A_j = P\{Y = y_j\}$ es una partición de Ω.

$$
\begin{aligned}
E[E[X/Y]] &= \sum_j E[X/Y = y_j] \cdot P(A_j) \\
&= \sum_j \frac{1}{P(A_j)} E[XI_{A_j}] \cdot P(A_j) \\
&= \sum_j E[XI_{A_j}] = \\
&= E\left[X\left(\sum_j I_{A_j}\right)\right] = E[X]
\end{aligned}
$$

Fórmula de la probabilidad total

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

Esperanza condicional de una variable continua respecto de una discreta

La definición anterior de $E[X/Y]$,

$$E[X/Y] = \sum_j E[X/Y = y_j] \cdot I_{A_j} \text{ donde } A_j = \{Y = y_j\}$$

también se puede aplicar si X es una variable aleatoria continua, e Y una variable discreta (siempre que $P(A_j) > 0$).

Un ejemplo

Ejerrcicio del parcial

Se tira un dado equilibrado de tres caras (o sea: se elege un número del 1 al 3 con idénticas probabilidades). Sea I el número obtenido en el dado. A continuación se define $Z = \sum_{j=1}^{I} X_j$ donde las variables aleatorias X_j tienen distribución exponencial de parámetro 1, y son todas independientes entre sí y del lanzamiento del dado.

i) Encuentre una expresión explícita para la densidad de probabilidad de Z.

ii) Utilizando dicha expresión, calcule $E[Z]$.

iii) Calcular $P(Z > 3)$.

Solución del item i)

Si conociéramos el valor i de I, tendríamos la variable

$$Z_i = \sum_{j=1}^{i} X_j$$

Sabemos que $Z_i \sim \Gamma(i, 1)$ por ser suma de suma de i variables aleatorias indepedendientes con distribución $\text{Exp}(1) = \Gamma(1, 1)$. [por un resultado que vimos en la clase 11]

Esta es una distribución condicional. ¡ Pero I es aleatoria! La verdadera distribución de Z se encuentra mezclando estas distribuciones condicionales, pesándolas de acuerdo a la distribución de probabilidades de I, como vimos en la clase 9.

$$\begin{aligned} f_Z(z) &= \sum_{i=1}^{3} f_{Z_i}(z) \cdot P\{I = i\} \\ &= \frac{1}{3} \left[\sum_{i=1}^{3} \frac{z^{i-1}}{(i-1)!} \right] \cdot I_{(0+\infty}(z) \end{aligned}$$

Solución del item ii)

Una vez determinada la distribución de Z su esperanza se encuentra mediante la fórmula de siempre.

$$E[Z] = \int_{-\infty}^{\infty} z \cdot f_Z(z)\, dz$$

Pero ahora podríamos pensar esta cuenta de otra manera

$$E[Z/I = i] = E[Z_i] = i$$

dado que ya calculamos la esperanza de una variable con distribución $\Gamma(i, 1)$ (en la clase 8). Entonces

$$E[Z] = E[E[Z/I]] = \sum_{i=1}^{3} E[Z/I = i] \cdot P\{I = i\} = \frac{1}{3} \sum_{i=1}^{3} i = \frac{1+2+3}{3} = 2$$

Parte III

Esperanza condicional de variables continuas

Problema: ¿Cómo lo generalizamos a variables continuas?

La definición anterior tiene un serio problema si queremos generalizar el concepto de esperanza condicional $E[X/Y]$ cuando la variable aleatoria Y es continua: en general

$$P\{Y = y_0\}$$

puede ser cero, por lo que las probabilidades condicionales:

$$P\{X \in I/Y = y_0\}$$

donde I es un intervalo, no va a estar definida.
Vamos a investigar primero el caso en que X e Y admiten una densidad conjunta $f_{XY}(x, y)$ continua. Recordamos que en esta situación Y se distribuye según la densidad marginal

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

Distribuciones condicionales

Consideramos un pequeño intervalo $J = [yo, y_0 + h]$, entonces:

$$P\{X \in I / Y \in J\} = \frac{P\{X \in I, Y \in J\}}{P\{Y \in J\}} = \frac{\int_I \int_J f_{XY}(x,y)\, dxdy}{\int_J f_Y(y)\, dy}$$

Entonces elegimos $I = (-\infty, x]$ y dividimos arriba y abajo por h

$$P\{X \leq x / Y \in J\} = \frac{\frac{1}{h} \int_{-\infty}^{x} \int_{y_0}^{y_0+h} f_{XY}(x,y)\, dxdy}{\frac{1}{h} \int_{y_0}^{y_0+h} f_Y(y)\, dy}$$

Cuando $h \to 0$ esta expresión converge a

$$F_{X/Y=y_0}(x) = \frac{\int_{-\infty}^{x} f_{XY}(x, y_0)\, dx}{f_Y(y_0)}$$

por el teorema fundamental del cálculo. Esta expresión se llama función de distribución condicional de X dada Y. Esta cuenta tiene sentido sólo si $f_Y(y) > 0$.

De donde obtenemos la densidad densidad condicional de X dada Y dada por

$$f_{X/Y=y_0}(x) = \frac{f_{XY}(x, y_0)}{f_Y(y_0)}$$

que podemos pensar como una versión infinitesimal de la definición de probabilidad condicional.
Entonces podemos definir la esperanza condicional en este caso, integrando la densidad condicional:

$$E[X/Y = y_0] = \int_{\infty}^{\infty} x \, dF_{X/Y=y_0}(x) = \int_{-\infty}^{\infty} x \, f_{X/Y=y_0}(x) \, dx$$

Todas las propiedades anteriores van a seguir valiendo con esta definición.

Un ejemplo

En el ejercicio 25 de la práctica 7 y la clase 13 consideramos la distribución normal bivariada. Recordamos que esta distribución es la distribución de un vector aleatorio

$$\begin{pmatrix} X \\ Y \end{pmatrix} = A \cdot \begin{pmatrix} Z_1 \\ Z_2 \end{pmatrix} + \mu$$

donde $Z_1, Z_2 \sim N(0,1)$ son independientes,

$$\mu = \begin{pmatrix} \mu_X \\ \mu_Y \end{pmatrix} \in \mathbb{R}^2$$

y $A \in \mathbb{R}^{2\times 2}$ es una matriz no singular. Encontramos que su densidad conjunta es

$$f_{XY}(x,y) = \frac{1}{2\pi\sigma_X\sigma_Y\sqrt{1-\rho^2}} e^{\left\{-\frac{1}{2(1-\rho^2)}\left[\left(\frac{x-\mu_X}{\sigma_X}\right)^2 + \left(\frac{y-\mu_Y}{\sigma_Y}\right)^2 - 2\rho\left(\frac{x-\mu_X}{\sigma_X}\right)\left(\frac{y-\mu_Y}{\sigma_Y}\right)\right]\right\}}$$

donde

$$\Sigma = A \cdot A^t = \begin{pmatrix} \sigma_X^2 & \rho\sigma_X\sigma_Y \\ \rho\sigma_X\sigma_Y & \sigma_Y^2 \end{pmatrix}.$$

es la matriz de covariancias, y ρ es el coeficiente de correlación entre X e Y.

La esperanza condicional en la densidad normal bivariada

Proposición

Si el vector (X, Y) se distribuye según la densidad normal bivariada $N(\mu, \Sigma)$, entonces

$$E[Y|X] = \mu_Y + \rho\frac{\sigma_Y}{\sigma_X}(X - \mu_X)$$

Esto dice, que en este caso la esperanza condicional está dada por la recta de regresión lineal (de la que hablamos en la clase 12).

Corolario

Por simetría, en eesta situación

$$E[X|Y] = \mu_X + \rho\frac{\sigma_X}{\sigma_Y}(Y - \mu_Y)$$

Demostración (1)

Buscamos la descomposición de Cholesky de la matriz de covariancia. Es decir buscamos $A = \text{Chol}(Z)$ triangular tal que

$$A \cdot A^t = \begin{pmatrix} a & 0 \\ b & c \end{pmatrix} \cdot \begin{pmatrix} a & b \\ 0 & c \end{pmatrix} = \begin{pmatrix} a^2 & ab \\ ab & b^2 + c^2 \end{pmatrix} = \Sigma = \begin{pmatrix} \sigma_X^2 & \rho\sigma_X\sigma_Y \\ \rho\sigma_X\sigma_Y & \sigma_Y^2 \end{pmatrix}$$

Nos quedan tres ecuaciones con tres incógnitas

$$a^2 = \sigma_x^2, \quad ab = \rho\sigma_X\sigma_Y, \quad b^2 + c^2 = \sigma_Y^2$$

Entonces

$$\begin{aligned} a &= \sigma_X \\ b &= \rho\sigma_X\sigma_Y / a = \rho\sigma_Y \\ c &= \sqrt{\sigma_Y^2 - b^2} = \sigma_Y(1 - \rho^2)^{1/2} \end{aligned}$$

Demostración (2)

Usando el resultado del ejercicio que mencionamos antes podemos escribir:

$$\begin{pmatrix} X \\ Y \end{pmatrix} = \mu + \mathrm{Chol}(\Sigma) Z$$

donde Z es un vector con distribución normal bivariada estándar, es decir con componentes Z_1, Z_2 que son $N(0,1)$ independientes.

$$\begin{pmatrix} X \\ Y \end{pmatrix} = \begin{pmatrix} \mu_X \\ \mu_Y \end{pmatrix} + \begin{pmatrix} \sigma_X & 0 \\ \rho\sigma_Y & \sigma_Y(1-\rho^2)^{1/2} \end{pmatrix} \cdot \begin{pmatrix} Z_1 \\ Z_2 \end{pmatrix}$$

de donde

$$\begin{aligned} X &= \mu_X + \sigma_X Z_1 \\ Y &= \mu_Y + \sigma_Y[\rho Z_1 + (1-\rho^2)^{1/2} Z_2] \end{aligned}$$

Demostración (3)

$$
\begin{aligned}
E[Y/X] &= E\left[\mu_Y + \sigma_Y[\rho Z_1 + (1-\rho^2)^{1/2} Z_2/X\right] \\
&= E\left[\mu_Y + \sigma_Y\left(\rho\frac{X-\mu_X}{\sigma_X}\right) + (1-\rho^2)^{1/2} Z_2/X\right] \\
&= E\left[\mu_Y/X\right] + E\left[\sigma_Y\left(\rho\frac{X-\mu_X}{\sigma_X}\right)/X\right] + E\left[(1-\rho^2)^{1/2} Z_2/X\right] = \\
&= \mu_Y + \sigma_Y\left(\rho\frac{X-\mu_X}{\sigma_X}\right) + (1-\rho^2)^{1/2} E\left[Z_2/X\right]
\end{aligned}
$$

Ahora X es una función de Z_1, y Z_1, Z_2 eran independientes. Se deduce que X es independiente de Z_2. Y la relación de independencia entre las variables es simétrica. Luego:

$$E[Z_2/X] = E[Z_2] = 0$$

Se deduce que:

$$E[Y/X] = \mu_Y + \sigma_Y\left(\rho\frac{X-\mu_X}{\sigma_X}\right)$$

Parte IV

Definición axiomática de la esperanza condicional

Propiedades que caracerizan a la esperanza condicional

Lema

Sean $X, Y : \Omega \to \overline{\mathbb{R}}$ *variables aleatorias discretas, donde* $E[|X|] < \infty$*, y* $P\{Y = y_j\} > 0$ *para todo* j*. La variable aleatoria* $h(Y) = E[X/Y]$ *tiene las siguientes propiedades:*

- *Tiene esperanza finita.*
- *Para cualquier función* $f : \mathbb{R} \to \mathbb{R}$ *acotada, se verifica que:*

$$E[f(Y)h(Y)] = E[f(Y)X]$$

Más aún: la esperanza condicional $E[X/Y]$ *está caracterizada por estas dos propiedades. en el siguiente sentido: si* $h_1, h_2 : \mathbb{R} \to \mathbb{R}$ *son dos funciones que verifican estas dos propiedades, entonces*

$$P\{h_1(Y) = h_2(Y)\} = 1$$

La segunda propiedad dice que:

$$E[f(Y)E[X/Y]] = E[f(Y)X]$$

Demostración

Para probar que $h(Y)$ tiene esperanza finita, debemos mostrar que la serie

$$\sum_j h(y_j) P\{Y = y_j\}$$

donde (y_j) recorre los posibles valores que la variable Y toma con probabilidad positiva, es absolutamente convergente.

$$\sum_j |h(y_j)| P\{Y = y_j\} = \sum_j \left| \sum_i x_i P\{X = x_i / Y = y_j\} \right| P\{Y = y_j\}$$

$$\leq \sum_i \sum_j |x_i| P\{X = x_i, Y = y_j\} = E(|X|) < +\infty$$

Esta cuenta dice en particular que:

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

Demostración (2)

Para probar la segunda afirmación calculamos:

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

$$= \sum_i f(y_j)P\{Y = y_j\} \sum_i x_i P\{X = x_i / Y = y_j\}$$

$$= \sum_i \sum_j f(y_j)x_i P\{X = x_i, Y = y_j\} = E[f(Y)X]$$

donde el reordenamiento de la serie se justifica utilizando que dicha serie converge absolutamente (dado que f es acotada).

Observación: En particular eligiendo $f \equiv 1$ en esta propiedad, vemos que

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

que es otra versión de la fórmula de probabilidad total.

Demostración (3)

Ahora probaremos la unicidad: supongamos que $h_1, h_2 : \mathbb{R} \to \mathbb{R}$ son funciones que verifican las propiedades anteriores. Entonces para cualquier función $f : \mathbb{R} \to \mathbb{R}$ acotada, tenemos que:

$$E[f(Y)h_1(Y)] = E[f(Y)h_2(Y)] = E[f(Y)X]$$

En consecuencia, si llamamos $h = h_1 - h_2$ por la linealidad de la esperanza:

$$E[f(Y)h(Y)] = 0$$

Eligiendo $f(t) = I_{\{y_j\}}(t)$ deducimos que:

$$h(y_j)P\{Y = y_j\} = 0$$

Por lo tanto si $h(y_j) \neq 0$, $P\{Y = y_j\} = 0$. En consecuencia:

$$P\{h(Y) \neq 0\} = \sum_{y_j : h(y_j) \neq 0} P\{Y = y_j\} = 0$$

Es decir que: $P\{h_1(Y) = h_2(Y)\} = 1$.

Un detalle muy técnico

Recordamos que la σ-álgebra de Borel se define como la σ-álgebra de subconjuntos de $\mathbb{R}$ generada por los intervalos (abiertos).

Definición

Una función $h : \mathbb{R} \to \mathbb{R}$ se dice boreliana si es medible respecto a la σ-álgebra de Borel, o sea que $h^{-1}(I)$ es un conjunto boreliano para todo intervalo abierto $I \subset \mathbb{R}$.

Nota: Esta definición garantiza que si $X : \Omega \to \mathbb{R}$ es una variable aleatoria, $h(X) = h \circ X : \Omega \to \mathbb{R}$ también lo es. pues

$$(h \circ X)^{-1}(I) = X^{-1}(h^{-1}(I))$$

Entonces si I es un intervalo, $h^{-1}(I)$ es un conjunto boreliano y entonces $X^{-1}(h^{-1}(I))$ es un evento (le podemos asignar una probabilidad).
Notemos que si $h : \mathbb{R} \to \mathbb{R}$ es continua, es boreliana.

Definición axiomática de la esperanza condicional

Definición

Sean $X, Y : \Omega \to \overline{\mathbb{R}}$ *variables aleatorias. Decimos que una variable aleatoria* $Z = h(Y)$ *es una versión de la esperanza condicional* $E[X/Y]$ *si donde* $h : \mathbb{R} \to \mathbb{R}$ *es una función boreliana, si se verifican las siguiente propiedades:*

1. $h(Y)$ *tiene esperanza finita.*
2. *Para cualquier función boreliana acotada* $f : \mathbb{R} \to \mathbb{R}$ *se verifica que:*

$$E[f(Y)h(Y)] = E[f(Y)X]$$

El caso continuo

Haciendo las mismas cuentas de antes, pero con integrales en lugar de sumas, y densidades en lugar de distribuciones puntuales, se prueba:

Teorema

Si el vector (X, Y) *se distribuye según la densidad de probabilidad conjunta* f_{XY} *y* $E(|X|) < \infty$. *Supongamos además que*

$$f_Y(y) > 0 \ \forall y \in \mathbb{R}$$

entonces

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

donde

$$f_{X/Y=y_0}(x) = \frac{f_{XY}(x, y_0)}{f_Y(y_0)}$$

es la densidad condicional, proposciona una versión de la esperanza condicional $E[X/Y]$.

Teorema de existencia

El siguiente teorema afirma que siempre existe una versión de la esperanza condicional, aunque no proporciona ninguna fórmula para calcularla. No demostraremos este teorema ya que su demostración depende de un teorema de análisis real (el teorema de Radon-Nikodym)

Teorema

Si $X, Y : \Omega \to \overline{\mathbb{R}}$ son variables aleatorias, siempre existe una versión de la esperanza condicional $E[X/Y]$. Además si $h_1(Y)$, $h_2(Y)$ son dos versiones de la esperanza condicional $E[X/Y]$, entonces

$$P\{h_1(Y) = h_2(Y)\} = 1$$