ESPERANZA CONDICIONAL

Informalmente, la idea es analizar como afecta a una variable aleatoria el conocimiento de otra. En particular este concepto de esperanza condicional será muy importante desde el punto de vista estadístico para predecir el valor de una variable aleatoria.

Motivación en el caso discreto: Sea (X, Y) un vector aleatorio discreto, con función de frecuencia $p_{XY}(x, y)$ y $p_X(x)$ la distribución marginal de X.

Dado x, con $p_X(x) > 0$, tenemos que

$$P(Y = y|X = x) = p_{XY}(x, y)/p_X(x),$$

y fijado x, al variar y tenemos una distribución de probabilidad, pues $\sum_y p_{XY}(x, y)/p_X(x) = 1$. Luego podemos calcular su esperanza,

$$E(Y|X = x) = \sum_y y\ p_{XY}(x, y)/p_X(x).$$

Podemos entonces además definir la variable aleatoria $E(Y|X)$ (que resulta una función de X) como la variable aleatoria que toma para cada x, $E(Y|X = x)$.

Queremos definir la noción de esperanza condicional en un marco más general, que no se restrinja a variables discretas, y que coincida en el caso discreto con lo anterior. Aqui vamos.

Sea Y una variable aleatoria definida en un espacio de probabilidad $(\Omega, \mathcal{A}, \mathbb{P})$, y $A \in \mathcal{A}$, con $P(A) > 0$. Definimos

$$\mathbb{E}(Y|A) = \frac{\mathbb{E}(Y\mathcal{I}_A)}{P(A)},$$

siendo $\mathcal{I}_A$ la función indicatriz del conjunto A. Observar que $\mathcal{I}_A$ tiene distribución Bernoulli(p), con $p = P(A)$. Observemos que $\forall B \in \mathcal{A}$, tenemos que

$$P(B|A) = \frac{\mathbb{P}(B \cap A)}{P(A)} = \frac{\mathbb{E}(\mathcal{I}_B\mathcal{I}_A)}{P(A)} = \mathbb{E}(\mathcal{I}_B|A), .$$

- Sea ahora X una variable aleatoria discreta, que toma valores

$$\{x_n : n \geq 1\},$$

con probabilidad positiva. Si consideramos los sucesos

$$A_n = \{\omega \in \Omega : X(\omega) = x_n\},$$

podemos expresar la variable aleatoria X como

$$X(\omega) = \sum_{j=1}^{\infty} x_j \mathcal{I}_{A_j}(\omega).$$

- Para cada A_n tenemos definido

$$\mathbb{E}(Y|A_n) = \mathbb{E}(Y|X = x_n) = \frac{\mathbb{E}(Y\mathcal{I}_{A_n})}{P(A_n)}.$$

- Luego, si X es discreta, definimos $\mathbb{E}(Y|X)$ **como la variable aleatoria** que toma valores $\mathbb{E}(Y|A_n)$ en los conjuntos $A_n = \{X = x_n\}$, o sea que

$$\mathbb{E}(Y|X) = \sum_{n=1}^{\infty} \mathbb{E}(Y|A_n)\mathcal{I}_{A_n} = \sum_{n=1}^{\infty} a_n \mathcal{I}_{A_n},$$

con $a_n = \mathbb{E}(Y\mathcal{I}_{A_n})/P(A_n)$, que es una función de X ($\sigma(X)$–medible).

1. Esperanza condicional: propiedades importantes

Se verifican las siguientes propiedades

(1) $\mathbb{E}(Y) = \mathbb{E}(\mathbb{E}(Y|X))$

(2) Más generalmente,

$$\mathbb{E}(Yh(X)) = \mathbb{E}(h(X)\mathbb{E}(Y|X)),$$

para toda h integrable.

(3) En particular, si B es $\sigma(X)$–medible,

$$\mathbb{E}(\mathcal{I}_B Y) = \mathbb{E}(\mathcal{I}_B \mathbb{E}(Y|X)).$$

- Demostración de (1).

$$\mathbb{E}(\mathbb{E}(Y|X)) = \mathbb{E}(\sum_{n=1}^{\infty} \mathbb{E}(Y\mathcal{I}_{A_n})\mathcal{I}_{A_n})/P(A_n)) =$$

$$\sum_{n=1}^{\infty} \mathbb{E}(Y\mathcal{I}_{A_n})P(A_n)/P(A_n) = \sum_{n=1}^{\infty} \mathbb{E}(Y\mathcal{I}_{A_n}) = \mathbb{E}(Y),$$

ya que A_n es una partición de Ω.

- Demostración de (2).

$$\mathbb{E}(h(X)\mathbb{E}(Y|X)) = \mathbb{E}(h(X)\sum_{n=1}^{\infty}\mathbb{E}(Y\mathcal{I}_{A_n})\mathcal{I}_{A_n})/P(A_n)) =$$

$$\sum_{n=1}^{\infty}\frac{1}{P(A_n)}\mathbb{E}(\mathbb{E}(Y\mathcal{I}_{A_n})\ h(X)\mathcal{I}_{A_n}) = \sum_{n=1}^{\infty}\frac{1}{P(A_n)}\mathbb{E}(\mathbb{E}(Y\mathcal{I}_{A_n}h(X))\mathcal{I}_{A_n})$$

ya que $h(X)$ es constante en A_n $(= h(x_n))$.

- Además como $\mathbb{E}(Y\mathcal{I}_{A_n}h(X))$ es una constante nos queda

$$\sum_{n=1}^{\infty}\frac{1}{P(A_n)}\mathbb{E}(h(X)Y\mathcal{I}_{A_n})P(A_n) = \mathbb{E}(h(X)Y),$$

pues $\{A_n : n \geq 1\}$ es una partición de Ω.

2. Caso general

Si X no es discreta, no podemos usar la definición anterior. Definimos $\mathbb{E}(Y|X)$ como aquella variable aleatoria que verifica:

Definition 2.1. (1) $\mathbb{E}(Y|X)$ es $\sigma(X)$ medible (o sea $\mathbb{E}(Y|X) = g(X)$)
(2) $\forall h$ integrable

$$\mathbb{E}(Yh(X)) = \mathbb{E}(h(X)\mathbb{E}(Y|X)). \tag{1}$$

Proof. Existencia: Teorema de Radon–Nykodim.

Unicidad. Si $g_1(X)$ y $g_2(X)$ verifican (1), entonces

$$P(g_1(X) = g_2(X)) = 1.$$

En efecto, tomemos $h(X) = g_1(X) - g_2(X)$.
$\mathbb{E}(Y[g_1(X)-g_2(X)]) = \mathbb{E}([g_1(X)-g_2(X)]g_1(X)) = E([g_1(X)-g_2(X)]g_2(X))$.
Restando $0 = \mathbb{E}([g_1(X) - g_2(X)]^2)$, y por tanto $P(g_1(X) = g_2(X)) = 1$. (Aquí usamos que X es una variable aleatoria, que verifica $P(X \geq 0) = 1$ y $E(X) = 0$ entonces $P(X = 0) = 1$.) $\square$

- **La probabilidad condicional se define de manera natural como la esperanza condicional de la función indicatriz del conjunto.**

$$\mathbb{P}^{\beta}(A) = \mathbb{P}(A|X) = \mathbb{E}(\mathcal{I}_A|X),$$

donde β es la σ–algebra generada por X.

- Definición equivalente

$$\int_C \mathbb{P}^\beta(A)dP = P(A\cap C).\ \ \forall\, C\in\beta.$$

Ver en la práctica.

3. Ejercicios

(1) Probar que
- Si X e Y son independientes $\mathbb{E}(Y|X)=\mathbb{E}(Y)$.
- $\mathbb{E}(g(X)Y|X)=g(X)\mathbb{E}(Y|X)$.
- Sean X e Y v.a. tales que $\mathbb{E}(X^2)<\infty$, $\mathbb{E}(Y^2)<\infty$. Supongamos que observamos X y buscamos la función $g(X)$ que mejor aproxime a Y en el sentido que minimize el error de predicción dado por

$$\mathbb{E}((Y-g(X))^2).$$

  A esa función llamaremos el mejor predictor de Y basado en X. Probar que la solución al problema es $g_0(X)=\mathbb{E}(Y|X)$. Dar una interpretación geométrica.

(2) casos discreto y absolutamente continuo.
Probar usando la definición que
- Si X e Y son v.a. discretas con distribución conjunta $p_{XY}(x,y)$, y la marginal de X es $p_X(x)$

$$\mathbb{E}(Y|X=x)=\sum_y y\frac{p_{XY}(x,y)}{p_X(x)}=\sum_y y\ P(Y=y|X=x).$$

- Sea (X,Y) un vector absolutamente continuo con densidad $f_{XY}(x,y)$. Entonces

$$\mathbb{E}(Y|X=x)=\int y\frac{f_{XY}(x,y)}{f_X(x)}dy.$$

(3) Probar que si X es una variable aleatoria continua, que verifica $P(X\geq 0)=1$ y $E(X)=0$ entonces $P(X=0)=1$.