TEMA 4. Esperanza condicionada: regresión y correlación

Contenidos

- Esperanza condicionada: definición y propiedades.
- Momentos condicionados.
- Regresión mínimo cuadrática bidimensional:
- Curvas de regresión.
- Razones de correlación.
- Regresión mínimo cuadrática bidimensional:
- Rectas de regresión.
- Coeficiente de determinación lineal y coeficiente de correlación lineal.

Esperanza condicionada

Sean X e Y variables aleatorias sobre el mismo espacio de probabilidad. se consideran las distribuciones condicionada de Y a X y de X a Y. Supongamos que ambas distribuciones poseen momento no centrado de orden uno finito. (En particular, si las distribuciones marginales tienen momentos no centrados de orden uno finitos, también existen los momentos no centrados de orden uno de las condicionadas). La esperanza condicionada de X a Y, denotada por $E[X/Y]$, y la esperanza condicionada de Y a X, denotada por $E[Y/X]$, se definen, como sigue, en el caso discreto y continuo:

Caso discreto

$$E[Y/X = x] = \sum_{y \in Supp(p_{Y/X=x})} y p_{Y/X=x}(y), \quad \forall x \in E_X.$$

Caso continuo

$$E[Y/X = x] = \int_{Supp(f_{Y/X=x})} y f_{Y/X=x}(y) dy, \quad \forall x \in E_X,$$

donde $Supp(f)$ denota el soporte de la función f.

De la misma forma, se define la esperanza condicionada de una función $g_1 : E_X \longrightarrow \mathbb{R}$, medible, de la v.a. X, condicionada a un valor concreto y de Y, denotada por $E[g_1(X)/Y = y]$. Similarmente, se define, para una función medible $g_2 : E_Y \longrightarrow \mathbb{R}$, la esperanza condicionada de $g_2(Y)$ a un valor concreto x de X, denotada por $E[g_2(Y)/X = x]$. Más concretamente, por ejemplo, para la esperanza condicionada de $g_2(Y)$, condicionada a un valor concreto x de X, se tiene

Caso discreto

$$E[g_2(Y)/X = x] = \sum_{y \in Supp(p_{Y/X=x})} g_2(y) p_{Y/X=x}(y), \quad \forall x \in E_X.$$

Caso continuo

$$E[g_2(Y)/X = x] = \int_{Supp(f_{Y/X=x})} g_2(y) f_{Y/X=x}(y) dy, \quad \forall x \in E_X.$$

Propiedades de la Esperanza Condicionada

Se notará, por simplicidad, en lo que sigue, por $E[X/Y]$ y $E[Y/X]$ a las esperanzas condiconadas de X a Y, y de Y a X, respectivamente.

- Si $X \geq 0$ c.s., entonces si existe $E[X/Y]$, se tiene que $E[X/Y] \geq 0$, y $E[X/Y] = 0$, si y sólo si $P(X = 0) = 1$.

- Si existe la $E[X/Y]$, entonces $|E[X/Y]| \leq E[|X|/Y]$.

- Linealidad: $\forall i = 1, \ldots, n$, $\exists E[X_i/Y]$, entonces $\exists E[\sum_{i=1}^n a_i X_i + b/Y] = \sum_{i=1}^n a_i E[X_i/Y] + b$, $\forall (a_1, \ldots, a_n) \in \mathbb{R}^n$, $b \in \mathbb{R}$.

- *Conservación del orden* Si $\exists E[X_i/Y]$, $i = 1, 2$, y $X_1 \leq X_2$, c.s., entonces $E[X_1/Y] \leq E[X_2/Y]$.

- Si existe $E[g(X)/Y]$, entonces $E[E[g(X)/Y]] = E[g(X)]$, para cualquier función medible g, y para cualesquiera variables aleatorias, X e Y, sobre el mismo espacio probabilístico base.

- Si X e Y son independientes, para cualquier función medible g, tal que existe $E[g(X)]$, se tiene $E[g(X)/Y] = E[g(X)]$, puesto que las distribuciones condicionadas coinciden con las marginales.
- $E[Xg(Y)/Y] = g(Y)E[X/Y]$, para cualquier función medible g.
- $E[X/Y, g(Y)] = E[X/Y]$, para cualquier función medible g.
- Para cualquier función h medible $E[h(X)/X] = h(X)$.
- $E[X/Y] = E[c/Y] = c$, para cualquier variable degenerada X sobre el mismo espacio de probabilidad de Y, con $P(X = c) = 1$.

Las propiedades enunciadas anteriormente se obtienen de forma directa, a partir de las definiciones proporcionadas sobre esperanza condicionada.

Esperanza condicionada de vectores aleatorios

Sean $\mathbf{X} = (X_1, \ldots, X_n)$ y $\mathbf{Y} = (Y_1, \ldots, Y_m)$ dos vectores aleatorios sobre el mismo espacio de probabilidad. Se define, suponiendo que las sumas e integrales, correspondientes al caso discreto y continuo, respectivamente, son absolutamente convergentes,

$$E[\mathbf{X}/\mathbf{Y}] = (E[X_1/\mathbf{Y}], \ldots, E[X_n/\mathbf{Y}]),$$

donde, para el caso discreto, es decir, para $(X_i, \mathbf{Y})$ un vector aleatorio discreto:

$$E[X_i/\mathbf{Y} = \mathbf{y}] = \sum_{x_i \in Supp(p_{X_i/\mathbf{Y}=\mathbf{y}})} x_i p_{X_i/\mathbf{Y}=\mathbf{y}}(x_i), \quad \forall \mathbf{y} \in E_{\mathbf{Y}}, \quad i = 1, \ldots n.$$

Para el caso continuo, es decir, si $(X_i, \mathbf{Y})$ es un vector aleatorio continuo,

$$E[X_i/\mathbf{Y} = \mathbf{y}] = \int_{Supp(f_{X_i/\mathbf{Y}=\mathbf{y}})} x_i f_{X_i/\mathbf{Y}=\mathbf{y}}(x_i) dx_i, \quad \forall \mathbf{y} \in E_{\mathbf{Y}} \quad i = 1, \ldots n.$$

Esperanza condicionada de una función unidimensional medible de un vector aleatorio

Para $E_{\mathbf{X}}$ tal que $P(\mathbf{X} \in E_{\mathbf{X}}) = 1$, y $E_{\mathbf{Y}}$ tal que $P(\mathbf{Y} \in E_{\mathbf{Y}}) = 1$, y $g : E_{\mathbf{X}} \longrightarrow \mathbb{R}$, una función medible,

$$E[g(\mathbf{X})/\mathbf{Y} = \mathbf{y}] = \sum_{\mathbf{x} \in Supp(p_{\mathbf{X}/\mathbf{Y}=\mathbf{y}})} g(\mathbf{x}) P(\mathbf{X} = \mathbf{x}/\mathbf{Y} = \mathbf{y}), \quad \forall \mathbf{y} \in E_{\mathbf{Y}},$$

donde $(\mathbf{X}, \mathbf{Y})$ es un vector aleatorio discreto.

Para $(\mathbf{X}, \mathbf{Y})$ un vector aleatorio continuo, y $g : E_{\mathbf{X}} \longrightarrow \mathbb{R}$, medible,

$$E[g(\mathbf{X})/\mathbf{Y} = \mathbf{y}] = \int_{Supp(f_{\mathbf{X}/\mathbf{Y}=\mathbf{y}})} g(\mathbf{x}) f_{\mathbf{X}/\mathbf{Y}=\mathbf{y}}(\mathbf{x}) d\mathbf{x}, \quad \forall \mathbf{y} \in E_{\mathbf{Y}}.$$

Se supone, como en el apartado anterior, que las sumatorias e integrales anteriores son absolutamente convergentes.

Caso Bidimensional. Momentos condicionados

Se introduce ahora la expresión de los momentos condicionados, no centrados y centrados, en el caso discreto y continuo, para vectores bidimensionales. Nótese que las expresiones que a continuación se formulan, se obtienen de forma directa del apartado anterior considerando $n = m = 1$, $g(X) = X^k$, y $g(X) = [X - E(X/Y)]^k$, $k \in \mathbb{N}$..

Momentos condicionados no centrados de orden k

Para $k \in \mathbb{N}$,

Caso discreto $$E[X^k/Y = y] = \sum_{x \in Supp(p_{X/Y=y})} x^k p_{X/Y=y}(x) \quad \forall y \in E_Y$$

Caso continuo $$E[X^k/Y = y] = \int_{Supp(f_{X/Y=y})} x^k f_{X/Y=y}(x) dx \quad \forall y \in E_Y.$$

Momentos condicionados centrados, respecto a la media, de orden k

Para $k \in \mathbb{N}$,

Caso discreto $$E\left[(X - E[X/Y])^k / Y = y\right] = \sum_{x \in Supp(p_{X/Y=y})} (x - E[X/Y])^k p_{X/Y=y}(x) \quad \forall y \in E_Y$$

Caso continuo $$E\left[(X - E[X/Y])^k / Y = y\right] = \int_{Supp(f_{X/Y=y})} (x - E[X/Y])^k f_{X/Y=y}(x) dx \quad \forall y \in E_Y.$$

Caso especial: Varianza condicionada

Supongamos que X, tal que $\exists E[X^2]$, e Y son dos variables aleatorias sobre el mismo espacio de probabilidad, entonces:

- $\exists \mathrm{Var}(X/Y)$, y $\mathrm{Var}(X/Y) \geq 0$.
- $\mathrm{Var}(X/Y) = E[X^2/Y] - (E[X/Y])^2$.
- *Descomposición de la varianza*: Si existen $\mathrm{Var}\left(E[X/Y]\right)$, y $E\left[\mathrm{Var}(X/Y)\right]$, entonces
$$\mathrm{Var}\left(X\right) = \mathrm{Var}\left(E[X/Y]\right) + E\left[\mathrm{Var}(X/Y)\right].$$

  Demostracin.

$$\begin{aligned}
\mathrm{Var}\left(X\right) &= E\left[(X - E[X])^2\right] = E\left[E\left[(X - E[X])^2/Y\right]\right] \\
&= E\left[E\left[(X^2 + (E[X])^2 - 2XE[X])/Y\right]\right] \\
&= E\left[E[X^2/Y] + (E[X])^2 - 2E[X]E[X/Y]\right] \\
&= E\left[E[X^2/Y] - (E[X/Y])^2 + (E[X/Y])^2 + (E[X])^2 - 2E[X]E[X/Y]\right] \\
&= E\left[\mathrm{Var}(X/Y)\right] + E\left[(E[X/Y] - E[X])^2\right] \\
&= E\left[\mathrm{Var}(X/Y)\right] + \mathrm{Var}\left(E[X/Y]\right).
\end{aligned}$$

Regresión mínimo cuadrática bidimensional

El problema de regresión consiste en aproximar una v.a. Y mediante una función de otra variable X, es decir,

$$Y \simeq \varphi(X),$$

donde Y es la variable dependiente, explicada o endógena; X es la variable independiente, explicativa o exógena y φ es la función de regresión. Nos centraremos en la aproximimación óptima de Y mediante la v.a. $\varphi(X)$, que proporciona la solución al problema de optimización, definido a partir de la función de pérdida $E[(Y - \varphi(X))^2]$, conocida como el error cuadrático medio (E.C.M) asociado a la estimación $\widehat{Y} = \varphi(X)$ de la v.a. Y. Es decir, se considera φ_{opt}, donde

$$\varphi_{\mathrm{opt}}(X) = \min_{\varphi} E[(Y - \varphi(X))^2].$$

La función φ_{opt}, que minimiza el E.C.M., en la aproximación de Y mediante una función de X viene dada por:

$$\varphi_{\mathrm{opt}}(X) = E[Y/X].$$

Por tanto,

$$\begin{aligned}
\text{E.C.M.}(\varphi(X)) &= E\left[(Y - E[Y/X])^2\right] \\
&= E\left[E\left[(Y - E[Y/X])^2/X\right]\right] = E\left[\mathrm{Var}(Y/X)\right].
\end{aligned}$$

Calculando

$$\begin{aligned}
E\left[(Y - E[Y/X])^2\right] &= E[Y^2] + E[(E[Y/X])^2] - 2E[YE[Y/X]] \\
&= E[Y^2] + E[(E[Y/X])^2] - 2E[E[YE[Y/X]/X]] \\
&= E[Y^2] + E[(E[Y/X])^2] - 2E[(E[Y/X])^2] \\
&= E[Y^2] - E\left[(E(Y/X))^2\right]
\end{aligned}$$

Alternativamente,

$$\begin{aligned}
\text{E.C.M.}(\varphi(X)) &= E\left[\mathrm{Var}(Y/X)\right] \\
&= E[E[Y^2/X] - (E(Y/X))^2]] = E[Y^2] - E\left[(E(Y/X))^2\right]
\end{aligned} \tag{1}$$

La ecuación (1) puede ser utilizada en el cálculo del E.C.M.($\varphi(X)$).

Según lo que se vio, en el apartado anterior, considerando la fórmula de descomposición de la varianza, se tiene

$$\mathrm{Var}(Y) = E\left[\mathrm{Var}(Y/X)\right] + \mathrm{Var}(E[Y/X]) = \text{E.C.M.}(\varphi(X)) + \mathrm{Var}(\varphi_{\text{opt}}(X)).$$

Curvas de regresión mínimo-cuadráticas

Se consideran, para diferentes valores x de la variable X, los correspondientes valores estimados de la v.a. Y, mediante la función φ_{opt}, es decir,

$$y = \varphi_{\text{opt}}(x) = E[Y/X = x], \quad \forall x \in E_X; \quad P(X \in E_X) = 1,$$

que recibe el nombre de curva de regresión de Y sobre X. El E.C.M asociado a la curva de regresión de Y sobre X es $E[\mathrm{Var}(Y/X)]$. Para un valor concreto, se tiene que el E.C.M. asociado a $E[Y/X = x]$ es $\mathrm{Var}(Y/X = x)$.

De forma análoga, para diferentes valores y de la variable Y, los correspondientes valores estimados de la v.a. X, mediante la función φ_{opt}, es decir,

$$x = \varphi_{\text{opt}}(y) = E[X/Y = y], \quad \forall y \in E_Y; \quad P(Y \in E_Y) = 1,$$

que recibe el nombre de curva de regresión de X sobre Y. El E.C.M asociado a la curva de regresión de X sobre Y es $E[\mathrm{Var}(X/Y)]$. Para un valor concreto, se tiene que el E.C.M. asociado a $E[X/Y = y]$ es $\mathrm{Var}(X/Y = y)$.

- Si Y depende funcionalmente de X, es decir, $Y = f(X)$, la curva de regresión de Y sobre X, $E[Y/X = x]$, $x \in E_X$, coincide con la curva de dependencia $y = f(x)$, $x \in E_X$.

- Si hay dependencia funcional recíproca entre X e Y, es decir, $Y = f(X)$ y $X = f^{-1}(Y)$, ambas curvas de regresión coinciden con las curvas de dependencia: $y = f(x)$, $x \in E_X$, y $x = f^{-1}(y)$, $y \in E_Y$.

- Si X e Y son independientes, las curvas de regresión son rectas paralelas a los ejes: $y = E[Y]$ y $x = E[X]$.

Razón de correlación

Se define la razón de correlación de X sobre Y, $\eta^2_{X/Y}$ y de Y sobre X, $\eta^2_{Y/X}$, como sigue:

$$\begin{aligned}
\eta^2_{X/Y} &= \frac{\mathrm{Var}(E[X/Y])}{\mathrm{Var}(X)} = 1 - \frac{E[\mathrm{Var}(X/Y)]}{\mathrm{Var}(X)} \\
\eta^2_{Y/X} &= \frac{\mathrm{Var}(E[Y/X])}{\mathrm{Var}(Y)} = 1 - \frac{E[\mathrm{Var}(Y/X)]}{\mathrm{Var}(Y)}
\end{aligned}$$

Propiedades

Para cualquier $a > 0$, se tiene

$$\begin{aligned}
\eta^2_{aX/Y} &= \eta^2_{X/Y} \\
\eta^2_{aY/X} &= \eta^2_{Y/X}
\end{aligned}$$

$$0 \leq \eta^2_{X/Y} \leq 1 \qquad 0 \leq \eta^2_{Y/X} \leq 1$$

Si $\eta^2_{X/Y} = 0$ y $\eta^2_{Y/X} = 0$, entonces las curvas de regresión de X/Y e Y/X coinciden respectivamente con las rectas $x = E[X]$ e $y = E[Y]$, y viceversa, si las curvas de regresión de X/Y e Y/X coinciden con las rectas $x = E[X]$ e $y = E[Y]$, respectivamente, entonces $\eta^2_{X/Y} = 0$ y $\eta^2_{Y/X} = 0$.

$\eta^2_{X/Y} = 1 \Leftrightarrow X$ dependen funcionalmente de Y.
$\eta^2_{Y/X} = 1 \Leftrightarrow Y$ dependen funcionalmente de X.
$\eta^2_{Y/X} = \eta^2_{X/Y} = 1 \Leftrightarrow$ hay dependencia funcional recíproca entre X e Y.

Regresión lineal mínimo-cuadrática

Se considera el problema de regresión mínimo-cuadrática, cuando la función φ^L es una recta. Es decir,

$$\varphi^L_{\text{opt}}(X) = \min_{a,b} E[(Y - aX - b)^2] = E[Y] + \frac{\text{Cov}(X,Y)}{\text{Var}(X)}(X - E[X]).$$

El E.C.M asociado viene dado por:

$$\begin{aligned} E.C.M.(\varphi^L_{\text{opt}}(X)) &= \text{Var}(Y) - \frac{[\text{Cov}(X,Y)]^2}{\text{Var}(X)} \\ \text{Var}(Y) &= \text{Var}(\varphi^L_{\text{opt}}(X)) + E.C.M.(\varphi^L_{\text{opt}}(X)) \end{aligned}$$

De forma similar, se calcula $\varphi^L_{\text{opt}}(Y)$ en la regresión de X sobre Y. Es decir,

$$\varphi^L_{\text{opt}}(Y) = \min_{a,b} E[(X - aY - b)^2] = E[X] + \frac{\text{Cov}(X,Y)}{\text{Var}(Y)}(Y - E[Y]).$$

El E.C.M asociado viene dado por:

$$\begin{aligned} E.C.M.(\varphi^L_{\text{opt}}(Y)) &= \text{Var}(X) - \frac{[\text{Cov}(X,Y)]^2}{\text{Var}(Y)} \\ \text{Var}(X) &= \text{Var}(\varphi^L_{\text{opt}}(Y)) + E.C.M.(\varphi^L_{\text{opt}}(Y)) \end{aligned}$$

Los coeficientes de regresión de Y/X, $\gamma_{Y/X}$, y de X/Y, $\gamma_{X/Y}$, se definen como sigue:

$$\begin{aligned} \gamma_{Y/X} &= \frac{\text{Cov}(X,Y)}{\text{Var}(X)} \\ \gamma_{X/Y} &= \frac{\text{Cov}(X,Y)}{\text{Var}(Y)} \end{aligned}$$

Propiedades

- Si $\text{Cov}(X,Y) = 0$, entonces las rectas de regresión de Y/X y de X/Y vienen respectivamente dadas por $y = E[Y]$ y $x = E[X]$.
- Si la curva de regresión es una recta, coincide con la correspondiente recta de regresión.
- Si X e Y están linealmente relacionadas, las rectas de regresión coinciden con la recta de dependencia.

Coeficiente de determinación lineal

En este apartado, se considera el coeficiente de determinación lineal $\rho^2_{X,Y}$, dado por:

$$\begin{aligned} \rho^2_{X,Y} &= \frac{[\mathrm{Cov}(X,Y)]^2}{\mathrm{Var}(X)\mathrm{Var}(Y)} \\ &= \frac{\mathrm{Var}(\varphi^L_{\mathrm{opt}}(X))}{\mathrm{Var}(Y)} = \frac{\mathrm{Var}(\varphi^L_{\mathrm{opt}}(Y))}{\mathrm{Var}(X)}. \end{aligned}$$

Propiedades

- $\rho^2_{aX+b,cY+d} = \rho^2_{X,Y}$
- $\rho^2_{X,Y} = \gamma_{X/Y}\gamma_{Y/X}$
- $0 \leq \rho^2_{X,Y} \leq 1$
- $\rho^2_{X,Y} = 0 \;\Leftrightarrow$ Las rectas de regresión de Y/X y de X/Y son respectivamente $y = E[Y]$, y $x = E[X]$.
- $\rho^2_{X,Y} = 1 \;\Leftrightarrow$ Existe dependencia funcional lineal entre las variables X e Y.
- $\rho^2_{X,Y} \leq \eta^2_{Y/X}$ y $\rho^2_{X,Y} \leq \eta^2_{X/Y}$. Las desigualdades anteriores se convierten en igualdades si y sólo si las curvas de regresión coinciden con las correspondientes rectas de regresión.

Coeficiente de correlación lineal

$$\rho_{X,Y} = \frac{\mathrm{Cov}(X,Y)}{\sqrt{\mathrm{Var}(X)\mathrm{Var}(Y)}}$$

Propiedades

- $\rho_{aX+b,cY+d} = \pm\rho_{X,Y}$
- $-1 \leq \rho_{X,Y} \leq 1$
- $\rho_{X,Y} = 0 \Leftrightarrow$ Las rectas de regresión de Y/X y de X/Y son respectivamente $y = E[Y]$, y $x = E[X]$.
- $|\rho^2_{X,Y}| = 1 \;\Leftrightarrow$ Existe dependencia funcional lineal (positiva o negativa, en función del signo) entre X e Y.