Clase Práctica 4

Maite Angel

2022-05-03

Independencia de variables

Definición: Las variables aleatorias X e Y son independientes si y sólo si para todo $a < b$, $c < d$ se cumple que:

$$P(a < X < b, c < Y < d) = P(a < X < b)P(c < Y < d)$$

- En el caso de que (X, Y) sea discreto se tiene que X e Y son independientes si y sólo si

$$p_{X,Y}(x, y) = p_X(x)p_Y(y) \qquad \forall (x, y) \in \mathbb{R}^2$$

Luego para probar que no son independientes es suficiente mostrar un punto (x_0, y_0) donde $p_{X,Y}(x_0, y_0) \neq p_X(x_0)p_Y(y_0)$.

- En el caso de que (X, Y) sea continuo se tiene que X e Y son independientes si

$$f_{X,Y}(x, y) = f_X(x)f_Y(y) \qquad \forall (x, y) \in \mathbb{R}^2$$

Para probar que no son independientes debemos mostrar un conjunto $(a, b)x(c, d)$ en el que no se satisfaga $f_{X,Y}(x, y) = f_X(x)f_Y(y)$ (no vale un punto solo).

Propiedades:

- Si X, Y son independientes entonces $E(XY) = E(X)E(Y)$.

Ejercicio 1

Sea X, Y v.a. tal que

- $f_X(x) = \frac{1}{10-4}1_{[4,10]}(x)$
- $f_{Y|X=x}(y) = \frac{1}{8-3}1_{[3,8]}(y)$

es decir $X \sim U(4, 10)$ y $Y|X \sim U(3, 8)$.

Item 1: ¿X, Y son independientes? Calcular $E(XY).

Item 2: Si ahora tenemos dos v.a.:

- Z: vale 1 cuando $X \geq 5$ si no 0.
- S: vale 1 cuando $X \leq 6$ e $Y \geq 5$ si no 0.

¿Z, S son independientes? Calcular $E(ZS)$

Solución

Item 1 Usando la propiedad de que $f_{Y|X=x}(y) = \frac{f_{X,Y}(x,y)}{f_X(x)}$ obtenemos:

- $f_{X,Y}(x,y) = f_{Y|X=x}(y) f_X(x) = \frac{1}{8-3} 1_{[3,8]}(y) \frac{1}{10-4} 1_{[4,10]}(x)$

Además $f_Y(y) = \int_{-\infty}^{+\infty} f_{X,Y}(x,y)dx$ luego,

- $f_Y(y) = \int_{-\infty}^{+\infty} \frac{1}{8-3} 1_{[3,8]}(y) \frac{1}{10-4} 1_{[4,10]}(x)dx = \frac{1}{8-3} 1_{[3,8]}(y) \frac{1}{10-4} \int_{-\infty}^{+\infty} 1_{[4,10]}(x) = \frac{1}{8-3} 1_{[3,8]}(y) \frac{1}{10-4}(10 - 4) = \frac{1}{8-3} 1_{[3,8]}(y)$

Luego $f_{X,Y}(x,y) == \frac{1}{8-3} 1_{[3,8]}(y) \frac{1}{10-4} 1_{[4,10]}(x) = f_Y(x) f_X(y)$ por lo que son independientes.
Usando la propiedad que recién escribimos:
$E(XY) = E(X)E(Y) = \frac{8+3}{2} \frac{10+4}{2}$

Item 2 Ahora tenemos un vector discreto (Z, S), $R_Z = R_S = \{0, 1\}$. Veamos la proba puntual

$$P(Z = 1, S = 1) = P(5 \leq X \leq 6, Y \geq 5) = \int_5^6 \int_5^8 \frac{1}{6}\frac{1}{5} dxdy = \frac{1}{6}\frac{1}{5}(8-5)(6-5) = 0.1$$

$$P(Z = 1) = \int_5^{10} \frac{1}{6} dx = \frac{5}{6}$$

$$P(S = 1) = P(X \leq 6, Y \geq 5) = \int_4^6 \int_5^8 \frac{1}{6}\frac{1}{5} dxdy = \frac{1}{6}\frac{1}{5}(8-5)(6-4) = 0.2$$

Como $P(Z = 1)P(S = 1) = \frac{5}{6} * 0.2 \neq 0.1 = P(Z = 1, S = 1)$ no son independientes.
Observación: Si tenemos X,Y no independientes tenemos que ir por la clásica usando que la proba de $E(h(X,Y)) = \sum_{R_X} \sum_{R_Y} h(x,y) p_{X,Y}(x,y)$ en el caso discreto, $E(h(X,Y)) = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} h(x,y) f_{X,Y}(x,y)dxdy$ en el caso continuo.

$$E(ZS) = \sum_{\{0,1\}} \sum_{\{0,1\}} zs p_{Z,S}(z,s) = 1 * 1 * p_{Z,S}(1,1) = P(Z = 1, S = 1) = 0.1$$

Covarianza y correlación

$$Cov(X,Y) = E((X - E(X))(Y - E(Y))) = E(XY) - E(X)E(Y)$$

Buscamos tener una medida de la variación conjunta de X,Y

$$\rho(X,Y) = \frac{Cov(X,Y)}{\sqrt{V(X)}\sqrt{V(Y)}}$$

Buscamos entender cómo se correlacionan, qué pasa con la otra cuando una crece/decrece.

Propiedades

Sean X,Y v.a.:

- Si X, Y son independientes entonces $Cov(X,Y) = 0$ (ojo! no vale la vuelta).
- $V(aX + bY) = a^2V(X) + b^2V(Y) + 2abCov(X,Y)$. (Esta era la prop que necesitábamos hace algunas clases!). Si X, Y son independientes entonces $V(X + Y) = V(X) + V(Y)$.
- Si a es una constante entonces $Cov(X, a) = 0$.
- $Cov(X,Y) = Cov(Y,X)$
- $Cov(aX + bZ, Y) = aCov(X,Y) + bCov(Z,Y)$
- $Cov(X,X) = V(X)$ es 0 sii X es constante.
- $|\rho(X,Y)| \leq 1$ y es 1 sii $P(Y = aX + b) = 1$ para algunas a, b constantes.
- Interpretación ρ:
  - Si $|\rho| = 1$ hay correlación perfecta entre X e Y. Si es -1 cuando una crece la otra decrece proporcionalmente, si es 1 ambas crecen/decrecen proporcionalmente.
  - Si $\rho = 0$ no hay correlación.
  - Si $0 < \rho < 1$ hay correlación positiva, cuando una crece/decrece la otra también crece/decrece pero no proporcionalmente
  - Si $-1 < \rho < 0$ hay correlación negativa, cuando una crece la otra decrece pero no proporcionalmente.

Ejercicio 2

Ana está entrenando para competir en una maratón, el tiempo en minutos que Ana tarda en completar el circuito es una v.a. A con varianza 5.

María está cocinando un bizcochuelo, el tiempo en minutos que tarda Maria es una v.a. con varianza 1.

Juan está viendo tele, el tiempo en minutos que tarda Juan en perder la concentración es una v.a. J que se describe como $J = 2A - 2M$.

Se sabe que A y M son independientes.

Item 1: ¿Cuál es la covarianza entre A y J ¿Son independientes?

Item 2 ¿Cómo se correlacionan los tiempos de Ana y Juan? ¿Y los de María y Juan? ¿Qué significa?

Solución

Item 1

$Cov(A, J) = Cov(A, 2A-2M) = Cov(2A-2M, A) = 2Cov(A, A)-2Cov(M, A) = 2V(A)-2*0 = 2*5 = 10$

No son independientes, si lo fueran tendríamos que entonces su covarianza es 0.

Item 2 Correlación Ana y Juan:

$$\rho(A,J) = \frac{Cov(A,J)}{\sqrt{V(A)}\sqrt{V(J)}} = \frac{10}{\sqrt{5}\sqrt{V(J)}}$$

Calculemos entonces $V(J)$:

$$V(J) = V(2A-2M) = 2^2V(A)+(-2)^2V(M)+2*2*(-2)Cov(A,M) = 4V(A)+4V(M) = 4*5+4*1 = 24$$

Luego $\rho(A,J) = \frac{10}{\sqrt{5}\sqrt{24}} \approx 0.9129$.

Significado: A y J tiene correlación positiva por lo que cuando una crece la otra también no proporcionalmente.

Correlación María y Juan:

Muy análogo al anterior, calculemos primero lo que ahora ya sabemos que necesitamos $Cov(M,J)$ (y la $V(J)$ pero la agarramos del punto anterior).

$$Cov(M,J) = Cov(M,2A-2M) = Cov(2A-2M,M) = 2Cov(A,M)-2Cov(M,M) = 2*0-2V(M) = -2*1 = -2$$

$$\rho(M,J) = \frac{Cov(M,J)}{\sqrt{V(M)}\sqrt{V(J)}} = \frac{-2}{\sqrt{1}\sqrt{24}} \approx -0.4082$$

Significado: A y J tiene correlación negativa por lo que cuando una crece la otra decrece no proporcionalmente.

Pero para, lo más lógico de pensar es que mis variables no tenían nada que ver!

Moraleja: **Correlación no implica causalidad**: es decir que la variables estén correlacionadas y su comportamiento se parezca no implica que alguna cause a la otra, la relación podría ser totalmente espuria (super importante a tener en cuenta en las épocas de pandemia).

Bonus track:

Entrar a la página https://www.tylervigen.com/spurious-correlations para ver correlaciones locas.