Feltételes eloszlás, feltételes várható érték, lineáris regresszió

Szűk elméleti összefoglaló

- X diszkrét v.v. feltételes eloszlása tetszőleges A eseményre nézve:

$$P(X = x_i|A) = \frac{P(X = x_i, A)}{P(A)}$$

- Ha $A_1, A_2, \ldots$ teljes eseményrendszer, X pedig tetszőleges diszkrét v.v., akkor az $\left\{\{P\big(X = x_i|A_j\big), i = 1,2, \ldots\}, j = 1,2, \ldots\right\}$ eloszlásokat X-nek az A_j teljes eseményrendszerre vett feltételes eloszlásának hívjuk
- Az X diszkrét v.v. Y diszkrét v.v. adott y_j értékére vett feltételes várható értéke

$$E\big(X|Y = y_j\big) = \sum_{\forall x_i} x_i P\big(X = x_i|Y = y_j\big)$$

- Az X folytonos v.v. Y folytonos v.v.-ra vett feltételes sűrűségfüggvénye

$$f_{X|Y}(x|y) = \frac{\frac{\partial^2 F_{X,Y}(x,y)}{\partial y \partial x}}{f_Y(y)} = \frac{f_{X,Y}(x,y)}{f_Y(y)}$$

- Az X folytonos v.v. Y folytonos v.v.-ra vett feltételes eloszlásfüggvénye

$$F_{X|Y}(x|y) = \frac{\frac{\partial F_{X,Y}(x,y)}{\partial y}}{f_Y(y)}$$

- Bayes-formula:

$$f_{X|Y}(x|y) = \frac{f_{Y|X}(y|x) f_X(x)}{\int_{-\infty}^{\infty} f_{Y|X}(y|u)\, f_X(u) du}$$

- Az X folytonos v.v. Y folytonos v.v.-ra vett regressziója, feltételes várható értéke

$$E(X|Y = y) = \int_{-\infty}^{\infty} x f_{X|Y}(x|y) dx = \frac{\int_{-\infty}^{\infty} x f_{X,Y}(x,y) dx}{f_Y(y)}$$

- a feltételes várható érték tulajdonságai:
  - $E\big(E(X|Y)\big) = EX$
  - $E(h(Y) \cdot X|Y) = h(Y)E(X|Y)$
  - ha X és Y független: $E(X|Y) = EX$
  - $E(X|X) = X$
- Y v.v. X-re vonatkozó lineáris regressziója (azaz Y=aX+b):

$$a = R(X,Y)\frac{\sigma Y}{\sigma X} = \frac{cov(X,Y)}{\sigma^2 X}$$
$$b = EY - aEX$$

Feladatok

1. Példa

Legyen az X és Y v.v. együttes sűrűségfüggvénye:

$$f_{X,Y}(x,y) = \begin{cases} \frac{12}{5}(x^2 - xy + y^2), & 0 < x, y < 1 \\ 0, & egyébként \end{cases}$$

Számolja ki az $f_{X|Y}(x|y)$ feltételes sűrűségfüggvényt! Számolja ki a kovarianciamátrixot és az $E(X|Y = y)$ regressziós függvényt is!

Megoldás:

$$f_{X|Y}(x|y) = \frac{f_{X,Y}(x,y)}{f_Y(y)}$$

$$f_Y(y) = \int_{-\infty}^{\infty} f_{X,Y}(x,y)dx = \int_0^1 \frac{12}{5}(x^2 - xy + y^2)dx = \frac{12}{5}\left[\frac{x^3}{3} - \frac{yx^2}{2} + y^2 x\right]_0^1 = \frac{4}{5} - \frac{6y}{5} + \frac{12y^2}{5}, y \in (0,1)$$

$$f_{X|Y}(x|y) = \frac{12x^2 - 12xy + 12y^2}{4 - 6y + 6y^2}, x, y \in (0,1)$$

$$E(X|Y = y) = \int_{-\infty}^{\infty} x f_{X|Y}(x|y)dx = \int_0^1 x\frac{6x^2 - 6xy + 6y^2}{2 - 3y + 3y^2}dx$$

$$= \frac{1}{2 - 3y + 3y^2}\left[\frac{3x^4}{2} - 2x^3 y + 3x^2 y^2\right]_0^1 = \frac{3 - 4y + 6y^2}{4 - 6y + 12y^2}$$

$$cov(X,Y) = EXY - EX \cdot EY$$

$$EY = \int_{-\infty}^{\infty} y f_Y(y)dy = \frac{2}{5}\int_0^1 2y - 3y^2 + 6y^3\, dy = \frac{2}{5}[y^2 - y^3 + 1{,}5y^4]_0^1 = \frac{3}{5}$$

$$f_X(x) = \int_{-\infty}^{\infty} f_{X,Y}(x,y)dy = \int_0^1 \frac{12}{5}(x^2 - xy + y^2)dy = \frac{12}{5}\left[yx^2 - \frac{xy^2}{2} + \frac{y^3}{3}\right]_0^1 = \frac{12x^2 - 6x + 4}{5}$$

$EX = \frac{3}{5}$ (mivel gyakorlatilag ugyanaz a forma, mint f_Y, ugyanabban a tartományban integrálva)

$$EX^2 = \int_0^1 x^2 f_X(x)dx = \frac{2}{5}\int_0^1 2x^2 - 3x^3 + 6x^4 dx = \frac{2}{5}\left[\frac{2}{3}x^3 - \frac{3}{4}x^4 + \frac{6}{5}x^5\right]_0^1 = \frac{67}{150}$$

Hasonló indokkal mint előbb $EY^2 = EX^2$, és

$$\sigma^2 X = \sigma^2 Y = EX^2 - E^2X = \frac{67}{150} - \left(\frac{3}{5}\right)^2 = \frac{13}{150}$$

$$EXY = \int\int xyf_{X,Y}(x,y)dx\,dy = \frac{12}{5}\int_0^1\int_0^1 x^3y - x^2y^2 + xy^3 dxdy$$

$$= \frac{12}{5}\int_0^1 \left[\frac{x^4y}{4} - \frac{x^3y^2}{3} + \frac{x^2y^3}{2}\right]_0^1 dy = \frac{12}{5}\int_0^1 \frac{y}{4} - \frac{y^2}{3} + \frac{y^3}{2} dy = \frac{12}{5}\left[\frac{y^2}{8} - \frac{y^3}{9} + \frac{y^4}{8}\right]_0^1$$

$$= \frac{12}{5} \times \left(\frac{2}{8} - \frac{1}{9}\right) = \frac{1}{3}$$

$$cov(X,Y) = \frac{1}{3} - \left(\frac{3}{5}\right)\left(\frac{3}{5}\right) = -\frac{2}{75}$$

$$\Sigma = \begin{pmatrix} \frac{13}{150} & -\frac{2}{75} \\ -\frac{2}{75} & \frac{13}{150} \end{pmatrix}$$

2. Példa

Dobjunk n-szer egy szabályos kockával. Jelölje X a hatosok, Y a páros dobások számát. Adja meg az $E(Y|X)$ regressziót!

Megoldás:

$$X \in B\left(n, \frac{1}{6}\right), Y \in B\left(n, \frac{1}{2}\right)$$

$$P(Y = k|X = l) = \frac{P(Y = k, X = l)}{P(X = l)} = \frac{\frac{n!}{l!\,(k-l)!\,(n-k)!}\left(\frac{1}{6}\right)^l\left(\frac{2}{6}\right)^{k-l}\left(\frac{3}{6}\right)^{n-k}}{\binom{n}{l}\left(\frac{1}{6}\right)^l\left(\frac{5}{6}\right)^{n-l}}$$

$$= \frac{n!\,l!\,(n-l)!}{l!\,(k-l)!\,(n-k)!\,n!}\frac{\left(\frac{2}{6}\right)^{k-l}\left(\frac{3}{6}\right)^{n-k}}{\left(\frac{5}{6}\right)^{n-l}} = \binom{n-l}{k-l}\frac{\left(\frac{2^{k-l}}{6^{k-l}}\right)\left(\frac{3^{n-k}}{6^{n-k}}\right)}{\frac{5^{n-l}}{6^{n-l}}}$$

$$= \binom{n-l}{k-l}\frac{\frac{2^{k-l}3^{n-k}}{6^{n-l}}}{\frac{5^{n-l}}{6^{n-l}}} = \binom{n-l}{k-l}\frac{2^{k-l}3^{n-k}}{5^{n-l}} = \binom{n-l}{k-l}\left(\frac{2}{5}\right)^{k-l}\left(\frac{3}{5}\right)^{(n-l)-(k-l)}$$

Ez pedig nem más, mint egy $B\left(n-l, \frac{2}{5}\right)$, aminek a várható értéke $n \times p$ alakú, tehát

$$E(Y|X = l) = (n-l)\frac{2}{5}$$

$$E(Y|X) = (n-X)\frac{2}{5}$$

3. Példa

Az X és Y valószínűségi változók együttes sűrűségfüggvénye $f_{X,Y}(u,v) = \frac{4}{3}(u^2 - uv + 2v^2), u, v \in (0,1)$. Adja meg az $E(X|Y)$ regressziót!

Megoldás:

$$f_{X|Y}(u|v) = \frac{f_{X,Y}(u,v)}{f_Y(v)}$$

$$f_Y(v) = \int_0^1 \frac{4}{3}(u^2 - uv + 2v^2)du = \frac{4}{3}\left[\frac{u^3}{3} - \frac{u^2 v}{2} + 2uv^2\right]_0^1 = \frac{4}{9} - \frac{4}{6}v + \frac{8}{3}v^2$$

$$f_{X|Y}(u|v) = \frac{\frac{4}{3}(u^2 - uv + 2v^2)}{\frac{4}{3}\left(\frac{1}{3} - \frac{v}{2} + 2v^2\right)} = \frac{6u^2 - 6uv + 12v^2}{2 - 3v + 12v^2}$$

$$E(X|Y = v) = \int_{-\infty}^{\infty} u f_{X|Y}(u|v)du = \int_0^1 \frac{6u^3 - 6u^2 v + 12uv^2}{2 - 3v + 12v^2}du$$

$$= \frac{1}{2 - 3v + 12v^2}\int_0^1 6u^3 - 6u^2 v + 12uv^2 du = \frac{\frac{6}{4} - 2v + 6v^2}{2 - 3v + 12v^2}$$

$$E(X|Y) = \frac{\frac{6}{4} - 2Y + 6Y^2}{2 - 3Y + 12Y^2}$$

4. Példa

Legyenek $X, Y \in N(0,1)$ függetlenek! Számolja ki a $E(Z|X)$ regressziót, ha $Z = 3X + Y$!

Megoldás:

$$E(Z|X) = E(3X + Y|X) = E(3X|X) + E(Y|X) = 3X + EY = 3X$$

5. Példa

Legyenek $X, Y \in N(0,1)$ függetlenek! Számolja ki a $E(Z|X)$ regressziót, ha $Z = 3X + Y + 1$!

Megoldás:

$$E(Z|X) = E(3X + Y + 1|X) = 3E(X|X) + E(Y|X) + E(1|X) = 3X + EY + 1 = 3X + 1$$

6. Példa

Feldobunk 10 kockát. X a hatosok, Y a hárommal oszthatók száma. Adja meg az $E(Y|X)$ regressziót!

Megoldás:

Y ebben a formájában nem tartalmazza X-et, pedig az valószínűleg tudna segíteni. Vezessünk be egy Z változót, ami a 3-as dobások száma. Ekkor Y=X+Z.

$$E(Y|X) = E(X + Z|X) = X + E(Z|X)$$

A világot nem váltottuk meg, de még mindig könnyebb E(Z|X)-et kiszámolni, mint E(Y|X)-et.

$$P(Z = j|X = i) = \binom{10-i}{j}\left(\frac{1}{5}\right)^{j}\left(\frac{4}{5}\right)^{10-i-j}$$

Ezek szerint $Z|X$ egy $B\left(10-i, \frac{1}{5}\right)$ eloszlást követ, aminek a várható értéke

$$E(Z|X = i) = (10-i)\frac{1}{5}$$

$$E(Z|X) = (10-X)\frac{1}{5}$$

$$E(Y|X) = X + \frac{(10-X)}{5} = \frac{10+4X}{5}$$

7. Példa

Legyenek $X, Y \in N(0,1)$ függetlenek. $U = 3X + 2Y, V = 2X - Y$. Adja meg az $E(U|V)$ feltételes valószínűséget!

Megoldás:

Mivel normális változók között a regresszió lineáris, így

$$E(U|V) = \frac{cov(U,V)}{\sigma^2 V}(V - EV) + EU$$

$$cov(U,V) = cov(3X+2Y, 2X-Y) = E\big((3X+2Y)(2X-Y)\big) - E(3X-2Y)E(2X-Y)$$

$$E\big((3X+2Y)(2X-Y)\big) = E(6X^2 + 4XY - 3XY - 2Y^2) = 6E(X^2) + E(XY) - 2E(Y^2)$$

$$E(XY) \rightarrow függetlenek \rightarrow EX \cdot EY, = 0$$

$$EX^2 = EY^2 = \sigma^2 X + E^2 X = 1 + 0 = 1$$

$$E(3X - 2Y) = 3EX - 2EY = 0$$

$$E(2X - Y) = 2EX - EY = 0$$

$$cov(U,V) = (6 + 0 - 2) - 0 \times 0 = 4$$

$$\sigma^2 V = \sigma^2(2X - Y) = 4\sigma^2 X + \sigma^2 Y = 5$$

$$E(U|V) = \frac{4}{5}(V - 0) + 0 = \frac{4}{5}V$$