RPIS

Lista 12 FULL

22.05.24

[1-3] Dane są obserwacje $x_1, \dots, x_n$, pochodzące z wymienionych rozkładów. Znaleźć estymator dla parametrów (metoda MLE)

1) F-ja wiarygodności $L(p) = \prod_{i=1}^{n} f(x_i|p)$
2) $l(p) = \log(L(p))$
3) $\frac{\partial l(p)}{\partial p} = 0 \rightarrow$ szukamy p
4) $\frac{\partial^2 l(p)}{\partial p^2} \leq 0 \rightarrow p$ - maximum

zad 1 Geom(p), parametr p

1) $f(x_i|p) = (1-p)^{x_i} \cdot p$ gdzie $x_i \in \{ ,1,2,\dots\}$

$$L(p) = \prod_{i=1}^{n} f(x_i|p) = \prod_{i=1}^{n} (1-p)^{x_i} \cdot p =$$

$$= \prod_{i=1}^{n} (1-p)^{x_i} \cdot \prod_{i=1}^{n} p =$$

$$= (1-p)^{\sum_{i=1}^{n} x_i} \cdot p^n$$

Zauważmy, że $\prod_{i=1}^{n} x^m = x^{\sum_{i=1}^{n} m}$

2) $$l(p) = \log L(p) = \log\left[(1-p)^{\sum_{i=1}^{n} x_i} \cdot p^n\right] =$$

$$= \sum_{i=1}^{n} x_i \cdot \log(1-p) + n \cdot \log p$$

3) $$\frac{\partial l(p)}{\partial (p)} = \frac{1}{1-p} \cdot (-1) \cdot \underbrace{\sum_{i=1}^{n} x_i}_{\text{stała}} + \frac{1}{p} \cdot n$$

$$\left.\frac{\partial l(p)}{\partial p}\right|_{p=\hat{p}} = 0 \iff -\frac{1}{1-\hat{p}} \cdot \sum_{i=1}^{n} x_i + \frac{n}{\hat{p}} = 0$$

$$\frac{n}{\hat{p}} = \frac{\sum_{i=1}^{n} x_i}{1-\hat{p}}$$

$$\frac{1-\hat{p}}{\hat{p}} = \frac{\sum_{i=1}^{n} x_i}{n}$$

$$\frac{1}{\hat{p}} = \frac{\sum_{i=1}^{n} x_i}{n} + \frac{n}{1} \iff \frac{\sum_{i=1}^{n} x_i + n}{n} = \frac{1}{\hat{p}}$$

wiemy że to jest średnia czyli $\overline{x}$

$$\overline{x} = \frac{1}{\hat{p}}$$

$$\hat{p} = \frac{1}{\overline{x}}$$

4) $$\frac{\partial^2 l(p)}{\partial p^2} = -\frac{1}{(1-p)^2} \sum_{i=1}^{n} x_i - \frac{n}{p^2}$$

$$\Updownarrow$$

$$-\left[\underbrace{\frac{1}{(1-p)^2}}_{>0} \underbrace{\sum_{i=1}^{n} x_i}_{x>0} + \underbrace{\frac{n}{p^2}}_{>0}\right] < 0$$

bo $(1-p)^2 > 0$ czyli >0

zad 2

Rozkład Pareto, $f(x,a,k) = \frac{k\,a^k}{x^{k+1}}$, $x \in [a, \infty)$, k - znane,

parametr a

1) $L(x;a,k) = \prod_{i=1}^{n} \frac{k \cdot a^k}{x_i^{k+1}} = k^n \cdot a^{kn} \cdot \prod_{i=1}^{n} \frac{1}{x_i^{k+1}}$

2) $l(x;a,k) = n \cdot \log(k) + nk \cdot \log(a) - (k+1) \sum_{i=1}^{n} \log x_i$

3) $\frac{\partial l(x;a,k)}{\partial(a)} = 0 + \frac{nk}{a} - 0 = 0$

Ale tutaj dochodzimy do sprzeczności ↯

bo $n > 0$, $k > 0$

To znaczy, że f-ja nie ma maximum

⇓

Ale z zał. $x \in [a, \infty)$, czyli $x \geq a$,

więc $a = \min\{x_i\}$.

zad 8 Prosta regresji dla danych:

x_k	1	3	4	6	8	9	11	14
y_k	1	2	4	4	5	7	8	9

$\hat{y}_i = a + b \cdot x_i$

$a = \bar{y} - b \cdot \bar{x}$

$$b = \frac{n \cdot \sum_{i=1}^{n} x_i y_i - \sum_{i=1}^{n} x_i \sum_{i=1}^{n} y_i}{n \cdot \sum_{i=1}^{n} x_i^2 - (\sum x_i)^2} \qquad n = 8$$

Wtedy:
$\sum x_i y_i = 364$
$\sum x_i^2 = 524$
$\sum x_i = 56 \quad \sum y_i = 40$
$(\sum x_i)^2 = 3136$

$$b = \frac{8 \cdot 364 - 56 \cdot 40}{8 \cdot 524 - 3136} = \frac{672}{1056} = \frac{7}{11}$$

$\bar{y} = \frac{\sum y_i}{n} = \frac{40}{8} = 5 \qquad \bar{x} = \frac{\sum x_i}{n} = \frac{56}{8} = 7$

$a = 5 - \frac{7}{11} \cdot 7 = 5 - \frac{49}{11} = \frac{6}{11}$

Prosta regresji: $\hat{y}_i = \frac{6}{11} + \frac{7}{11} \cdot x_i$

$\boxed{\text{zad 4}}$ Dane punkty $(x_1, y_1), \dots, (x_n, y_n)$.
Szukamy krzywej regresji $y = a + bx + cx^2$
Uzasadnić, że a, b, c są rozwiązaniem układu:

$$\begin{bmatrix} n & \sum x_i & \sum x_i^2 \\ \sum x_i & \sum x_i^2 & \sum x_i^3 \\ \sum x_i^2 & \sum x_i^3 & \sum x_i^4 \end{bmatrix} \cdot \begin{bmatrix} a \\ b \\ c \end{bmatrix} = \begin{bmatrix} \sum y_i \\ \sum x_i y_i \\ \sum x_i^2 y_i \end{bmatrix}$$

Regresja kwadratowa (metoda najmniejszych kwadratów)
Cel: znaleźć takie a, b, c, które minimalizują sumę kwadratów błędów:

$$S(a,b,c) = \sum_{i=1}^{n} (y_i - (a + bx_i + cx_i^2))^2 \quad \text{funkcja kosztu}$$

Aby znaleźć a, b, c, obliczamy pochodne cząstkowe i przyrównujemy do 0:

$$\frac{\partial S}{\partial a} = -\cancel{2} \sum_{i=1}^{n} 1 (\underbrace{y_i - a - bx_i - cx_i^2}) = 0$$

$$\frac{\partial S}{\partial b} = -\cancel{2} \cdot \sum_{i=1}^{n} x_i \cdot (\underbrace{y_i - a - bx_i - cx_i^2}) = 0$$

$$\frac{\partial S}{\partial c} = -\cancel{2} \cdot \sum_{i=1}^{n} x_i^2 \cdot (\underbrace{y_i - a - bx_i - cx_i^2}) = 0$$

$$\begin{cases} na + \sum x_i b + \sum x_i^2 c = \sum y_i \\ \sum x_i a + \sum x_i^2 b + \sum x_i^3 \cdot c = \sum x_i y_i \\ \sum x_i^2 a + \sum x_i^3 b + \sum x_i^4 c = \sum x_i^2 y_i \end{cases}$$

a to jest dokładnie to samo, co wyżej

$\boxed{\text{zad 5}}$ Punkty $(x_1, y_1, z_1), \dots, (x_n, y_n, z_n)$
Równanie regresji $z = a + bx + cy$. Uzasadnić, że a, b, c są rozwiązaniem

$$\begin{bmatrix} n & \sum x_i & \sum y_i \\ \sum x_i & \sum x_i^2 & \sum x_i y_i \\ \sum y_i & \sum x_i y_i & \sum y_i^2 \end{bmatrix} \begin{bmatrix} a \\ b \\ c \end{bmatrix} = \begin{bmatrix} \sum z_i \\ \sum x_i z_i \\ \sum y_i z_i \end{bmatrix}$$

$$S(a,b,c) = \sum (z_i - \hat{z}_i)^2 = \sum (z_i - (a + bx_i + cy_i))^2$$

$$\frac{\partial S}{\partial a} = -\cancel{2} \sum (\underbrace{z_i - (a + bx_i + cy_i)}) = 0$$

$$\frac{\partial S}{\partial b} = -\cancel{2} \sum x_i (\underbrace{z_i - (a + bx_i + cy_i)}) = 0$$

$$\frac{\partial S}{\partial c} = -\cancel{2} \sum y_i (\underbrace{z_i - (a + bx_i + cy_i)}) = 0$$

znowu jak w $\boxed{\text{zad 4}}$

Scanned with ACE Scanner

zad 2 Rozkład Pareto, $f(x,a,k) = \frac{k a^k}{x^{k+1}}$, $x \in [a, \infty)$, k - znane, parametr a

1) $L(x;a,k) = \prod_{i=1}^{n} \frac{k \cdot a^k}{x_i^{k+1}} = k^n \cdot a^{kn} \cdot \prod_{i=1}^{n} \frac{1}{x_i^{k+1}}$

2) $l(x;a,k) = n \cdot \log(k) + nk \cdot \log(a) - (k+1) \sum_{i=1}^{n} \log x_i$

3) $\frac{\partial l(x;a,k)}{\partial (a)} = 0 + \frac{nk}{a} - 0 = 0$

Ale tutaj dochodzimy do sprzeczności ↯
bo $n > 0$, $k > 0$

To znaczy, że f-ja nie ma maximum

↓

Ale z zał. $x \in [a, \infty)$, czyli $x \geq a$,
więc $a = \min\{x_i\}$.

zad 8 Prosta regresji dla danych:

x_k	1	3	4	6	8	9	11	14
y_k	1	2	4	4	5	7	8	9

$\hat{y}_i = a + b \cdot x_i$

$a = \bar{y} - b \cdot \bar{x}$

$b = \frac{n \cdot \sum_{i=1}^{n} x_i y_i - \sum_{i=1}^{n} x_i \sum_{i=1}^{n} y_i}{n \cdot \sum_{i=1}^{n} x_i^2 - (\sum x_i)^2}$ $n = 8$

Wtedy: $\sum x_i y_i = 364$
$\sum x_i^2 = 524$
$\sum x_i = 56$ $\sum y_i = 40$
$(\sum x_i)^2 = 3136$

$b = \frac{8 \cdot 364 - 56 \cdot 40}{8 \cdot 524 - 3136} = \frac{672}{1056} = \frac{7}{11}$

$\bar{y} = \frac{\sum y_i}{n} = \frac{40}{8} = 5$ $\bar{x} = \frac{\sum x_i}{n} = \frac{56}{8} = 7$

$a = 5 - \frac{7}{11} \cdot 7 = 5 - \frac{49}{11} = \frac{6}{11}$

Prosta regresji: $\hat{y}_i = \frac{6}{11} + \frac{7}{11} \cdot x_i$