

Regressogram*

(wersja II)

dr hab. inż. Przemysław Śliwiński, prof. PWr.

10 maja 2016

1 Regressogram

1. Wygenerować ciąg liczb $\{X_1, \dots, X_N\}$, $N = 1024$, o dowolnej gęstości w przedziale $[-1, 1]$.
2. Dla uzyskanego ciągu wygenerować wartości

$$Y_n = m(X_n) + Z_n, \quad n = 1, \dots, N$$

gdzie

$$m(x) = \arctan(2\pi x) \quad \text{oraz} \quad Z_n \sim N(0, 0.05)$$

3. Zaimplementować *regressogram*¹ o postaci:

$$\begin{aligned} \bar{m}_h(x) &= \frac{\sum_{n=1}^N \frac{1}{hN} Y_n \cdot K\left(\lfloor \frac{x}{h} \rfloor - \frac{x_n}{h}\right)}{\sum_{n=1}^N \frac{1}{hN} K\left(\lfloor \frac{x}{h} \rfloor - \frac{x_n}{h}\right)} = \frac{\sum_{n=1}^N Y_n \cdot K\left(\lfloor \frac{x}{h} \rfloor - \frac{x_n}{h}\right)}{\sum_{n=1}^N K\left(\lfloor \frac{x}{h} \rfloor - \frac{x_n}{h}\right)} \\ &= \frac{\sum_{n \in \kappa_x} Y_n}{\#\kappa_x}, \end{aligned}$$

gdzie

$$\kappa_x = \left\{ n = 1, \dots, 1024 : x_n \in \text{supp } K\left(\lfloor \frac{x}{h} \rfloor\right) \right\}$$

i gdzie $K(x)$ jest prostokątną funkcją jądra

$$K(x) = \mathbf{1}_{[0,1)}(x)$$

przyjmując

$$h = 1/2, 1/4, \dots, 1/128 \quad (= 2^{-1}, 2^{-2}, \dots, 2^{-7})$$

*A portmanteau of *regression* and *histogram*.

¹Nieparametryczny estymator funkcji regresji o kształcie wykresu przypominającym histogram.

4. Wyznaczyć h takie, aby błąd *regressogramu*

$$validation(h) = \sum_{q=-Q}^Q \left[\bar{m}_h \left(\frac{q}{Q} \right) - m \left(\frac{q}{Q} \right) \right]^2$$

był jak najmniejszy.

5. Zastąpić zakłócenie normalne Z_n , zakłóceniem o rozkładzie Cauchy'ego $C(0, 0.05)$ i powtórzyć eksperyment.
6. Wyciągnąć wnioski.