

Histogram a estymator jądrowy gęstości (wydanie poprawione)

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

26 kwietnia 2016

1 Histogram

1. Wygenerować ciąg liczb $\{X_1, \dots, X_N\}$, $N = 1024$, o gęstości będącej dowolną funkcją prostą (np. schodkową z czterema różnymi schodkami)¹. Przekazać go osobie po prawicy **wraz ze wzorem na gęstość**.
2. Zaimplementować (jak poprzednio):

(a) *histogram*, tj. estymator funkcji gęstości o postaci:

$$\bar{f}_h(x) = \frac{1}{N} \sum_{n=1}^N \frac{1}{h} K\left(\left\lfloor \frac{x}{h} \right\rfloor - \frac{x_n}{h}\right),$$

z prostokątną funkcją jądra $K(x) = \mathbf{1}_{[0,1)}(x)$

(b) nieparametryczny jądrowy estymator gęstości:

$$\hat{f}_h(x) = \frac{1}{N} \sum_{n=1}^N \frac{1}{h} K\left(\frac{x - x_n}{h}\right)$$

3. Wyznaczyć h tak, aby błąd estymatorów

$$\text{validation}(h) = \sum_{q=-Q}^Q [f_h(x_q) - f(x_q)]^2, \text{ gdzie } f_h = \{\hat{f}_h, \bar{f}_h\},$$

był jak najmniejszy, odpowiednio, dla histogramu i dla estymatora jądrowego.

4. Uzasadnić ewentualne rozbieżności pomiędzy h dla obu estymatorów.

¹Proszę pamiętać (przy doborze wysokości schodków), że funkcja – aby była gęstością – musi być nieujemna i całkować się do jedności.