

Histogram a jądrowy estymator gęstości (wydanie III, poprawione)

Przemysław Śliwiński

9 kwietnia 2018

1 Histogram vs estymator jądrowy

- Wygenerować ciąg liczb $\{X_1, \dots, X_N\}$, $N = 1024$, o rozkładzie Laplace'a¹ (albo Skellama^{2, 3}) o średniej $\mu = 0$ i wariancji $\sigma^2 = 32$.
- Zaimplementować (jak poprzednio):
 - histogram*, tj. **nieparametryczny estymator** funkcji gęstości prawdopodobieństwa o postaci:

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

z prostokątną funkcją jądra $K(x) = \mathbf{1}_{[-\frac{1}{2}, \frac{1}{2}]}(x)$

- nieparametryczny jądrowy estymator** funkcji gęstości prawdopodobieństwa:

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

- Wyznaczyć h tak, aby błąd estymatorów

$$\text{validation}(h) = \frac{1}{2Q} \sum_{q=-Q}^Q [f_h(x_q) - f(x_q)]^2, \text{ gdzie } f_h = \{\bar{f}_h, \hat{f}_h\},$$

był jak najmniejszy, odpowiednio, dla histogramu i dla estymatora jądrowego⁴.

¹Rozkład Laplace'a służy m.in. do modelowania wartości współczynników transformaty kosinusowej (stosowanej w standardach kompresji JPG, H.265).

²Rozkład Skellama o średniej $\mu = \mu_1 - \mu_2$ i wariancji $\sigma^2 = \mu_1 + \mu_2$ ma różnicę dwóch zmiennych losowych o rozkładzie Poissona o średnich i wariancjach, μ_1 i μ_2 . Może to być w praktyce np. różnica wartości zarejestrowanych przez dwie fotodiody.

³Warto zwrócić uwagę, że rozkład Skellama jest **dyskretny** - zachodzi więc pytanie, jak interpretować histogram (estymator **gęstości**, której ten rozkład nie posiada).

⁴Wyrazy ciągu $\{x_q\}$ mogą być zdefiniowane następująco: $x_q = 16qQ^{-1}$ (**dlaczego?!**).

4. Uzasadnić ewentualne rozbieżności pomiędzy parametrami h wyznaczonymi dla obu estymatorów.
5. Przedstawić wykresy obu estymatorów otrzymanych dla najlepszych parametrów h .
6. Podać potencjalne zastosowania obu typów estymatorów.