Histogram a estymator jądrowy gęstości (wydanie poprawione)

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

20 kwietnia 2017

1 Histogram vs estymator jądrowy

1. Wygenerować ciąg liczb $\{X_1, \ldots, X_N\}$, $N = 1024$, o gęstości Skellama[1] o średniej $\mu = 0$ i wariancji $\sigma^2 = 32$.

2. Zaimplementować (jak poprzednio):

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

   (b) **nieparametryczny** jądrowy estymator funkcji gęstości prawdopodobieństwa:

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

3. Wyznaczyć h tak, aby błąd estymatorów

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

   był jak najmniejszy, odpowiednio, dla histogramu i dla estymatora jądrowego[2].

4. Uzasadnić ewentualne rozbieżności pomiędzy parametrami h wyznaczonymi dla obu estymatorów.

[1] Rozkład Skellama o średniej $\mu = \mu_1 - \mu_2$ i wariancji $\sigma^2 = \mu_1 + \mu_2$ ma różnica 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.

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

5. Przedstawić wykresy obu estymatorów otrzymanych dla najlepszych parametrów h.

6. Podać potencjalne zastosowania obu typów estymatorów.