Modelowanie i identyfikacja
Laboratorium nr 6

1. Zaimplementować estymator gęstości w oparciu o histogram:
   (a) Czy jest to estymator obciążony?
   (b) Jaka jest jego wariancja?
   (c) Jaki wpływ ma na działanie estymatora wartość współczynnika wygładzania h? Jego wartość można wyznaczyć ze wzoru:
   $$h = N^{-\alpha}, \tag{1}$$
   gdzie $\alpha \in (0, 1)$. Ustalić stałą liczbę klas np: $N = \sqrt{n}$ (n to liczba elementów w wygenerowanym ciągu liczb losowych).
   (d) Jaki wpływ na działanie estymatora ma liczba klas N?
   (e) Przy ustalonym h i N na podstawie poprzednich punktów, zbadać wpływ liczby próbek na jakość estymacji.

   Punkty a) i b) wyznaczyć również analitycznie.

2. Dokonać estymacji gęstości estymatorem jądrowym z użyciem różnych funkcji jądra (minimum 3 różne):
   (a) wykreślić estymowaną funkcję gęstości dla różnych funkcji jądra,
   (b) zbadać błąd empiryczny estymatora dla L niezależnych, N-elementowych pomiarów.

3. Powyższe zadania wykonać dla zmiennych losowych pochodzących z rozkładu:
   (a) jednostajnego,
   (b) trójkątnego,
   (c) normalnego.

4. Niezbędna teoria:

   (a) U podstaw estymacji gęstości rozkładu za pomocą histogramu leży teoretyczna zależność pomiędzy dystrybuantą a gęstością:

$$\hat{f}(x) = \frac{dF(x)}{dx} = \lim_{h \to 0} \frac{F(x+h) - F(x)}{h}$$

   Korzystając ze wzoru na empiryczną dystrybuantę oraz wykonując kilka prostych przekształceń otrzymujemy wzór na gęstość empiryczną (proszę zawrzeć wyprowadzenie w sprawozdaniu):

$$\begin{aligned} \hat{f}_N(x) &= \frac{\#\left\{x < X_k \leqslant x + h\right\}}{Nh} \\ &= \frac{1}{Nh} \sum_{k=1}^{N} I\left\{x < X_k \leqslant x + h\right\}. \end{aligned}$$

   (b) Estymator jądrowy wyraża się wzorem:

$$\hat{f}_N(x) = \frac{1}{Nh_N} \sum_{k=1}^{N} K\left(\frac{X_k - x}{h_N}\right),$$

   gdzie $K(\cdot)$ jest funkcją jądra, h_N parametrem wygładzania, N liczbą elementów w wygenerowanym ciągu liczb losowych.