# Modelowanie i identyfikacja – laboratorium 4.
# Dystrybuanta empiryczna i jej własności

Paweł Wachel

**Wymagania wstępne:**

1. Wymagania wstępne z poprzednich zajęć[1].
2. Znajomość konstrukcji estymatora dystrybuanty empirycznej.
3. Pojęcie obciążenia i wariancji estymatora – definicja i interpretacja.

**Zadania do wykonania:**

1. Wykorzystując opracowane na wcześniejszych zajęciach generatory, wygenerować $N$-elementowy ciąg liczb losowych $\{X_1, X_2, \ldots, X_N\}$ o gęstości prawdopodobieństwa:
$$f(x) = \begin{cases} 2x & \text{dla} \quad x \in [0,1] \\ 0 & \text{dla} \quad x \in (-\infty, 0) \cup (1, \infty) \end{cases} .$$
2. Wykreślić dystrybuantę powyższego rozkładu (oznaczaną dalej przez $F(x)$), zaimplementować i wykreślić dystrybuantę empiryczną:
$$\hat{F}_N(x) = \frac{1}{N} \sum_{n=1}^{N} \mathbf{I}(X_n \leq x),$$
gdzie $\mathbf{I}(a \leq b) = 1$, gdy $a \leq b$ i 0 w przeciwnym przypadku.
3. Zaimplementować empiryczny odpowiednik wyrażenia
$$D_N = \sup_x \left| \hat{F}_N(x) - F(x) \right|$$
i wykreślić realizację $D_N$ w funkcji $N$. Zinterpretować uzyskane rezultaty.
4. Plik `ModelowanieLab4Data.txt` zawiera pewną sekwencję pomiarów. Wykreślić dystrybuantę empiryczną $\hat{F}_N(x)$ dla różnych wartości $N$. Z jakiego rozkładu pochodzą dane[2]? Rozważyć rozkłady: normalny[3] $\mathcal{N}(1, 1)$, normalny $\mathcal{N}(0, 5)$ oraz Cauchy'ego z parametrami $x_0 = 0$ i $\gamma = 1$?
5. Skonstruować eksperyment symulacyjny, umożliwiający estymację wariancji dystrybuanty empirycznej ($\sigma_F^2(x) = Var\{\hat{F}_N(x)\}$). Wykreślić wartość oszacowania wariancji $\hat{\sigma}_F^2(x)$ w funkcji $x$.

[1] Całkujemy wiedzę... przynajmniej do wakacji.

[2] Można zaproponować własne kryterium ułatwiające określenie rozkładu lub skorzystać z wybranego testu statystycznego.

[3] Do pracy z dytrybuantą rozkładu normalnego można wykorzystać funkcję `erf(x)`.

**Zadania dodatkowe:**

1. Dla wybranego rozkładu prawdopodobieńtwa wygenerować $L$ niezależnych, $N$-elementowych sekwencji pomiarowych (prób) i wyznaczyć błąd empiryczny

$$Err\left\{\hat{F}_N\right\} = \frac{1}{LM}\sum_{l=1}^{L}\sum_{m=1}^{M}\left[\hat{F}_N^{[l]}\left(x_m\right) - F\left(x_m\right)\right]^2,$$

w którym $\{x_1, x_2, \ldots, x_M\}$ jest sekwencją równoodlegych punktów z pewnego odcinka $[a, b]$, a $\hat{F}_N^{[l]}$ jest $l$-tą realizacją estymatora. Badania rozpocząć od $M = 100$ i $L = 10$. Jaki wpływ na uzyskiwane wyniki mają stałe $L$ i $M$?

2. Zaproponować eksperyment numeryczny obrazujący (punktową) zbieżność według rozkładu dystrybuanty empirycznej do rozkładu normalnego, tzn.

$$\sqrt{N}\left[\hat{F}_N\left(x\right) - F\left(x\right)\right] \xrightarrow{d} \mathcal{N}\left(0, F\left(x\right)\left(1 - F\left(x\right)\right)\right),$$

dla dowolnego (ustalonego) $x \in \mathbb{R}$. Dokonać interpretacji uzyskanych wyników. W jaki sposób asymptotyczny rozkład normalny zależy od argumentu $x$?

**Literatura:**


1. Wasserman, Larry. All of statistics: a concise course in statistical inference. Springer Science & Business Media, 2013.

2. Plucińska Agnieszka, Pluciński Edmund. Probabilistyka: rachunek prawdopodobieństwa, statystyka matematyczna, procesy stochastyczne. Wydawnictwa Naukowo-Techniczne, 2000.

3. Gajek Lesław, Kałuszka Marek. Wnioskowanie statystyczne: modele i metody. Wydawnictwa Naukowo-Techniczne, 1993.

4. Notatki z wykładu.