

LABORATORIUM

MODELOWANIE I IDENTYFIKACJA

---

# Sprawozdanie

## Estymatory

### Laboratorium 3, 4, 5, 6

---

*Autor:*

Adam KRZYKAŁA, 235411

*Termin:* czwartek 9:15

*Prowadzący:*

dr inż. Paweł WACHEL, prof. ucz.

6 maja 2020

# Spis treści

<b>1</b>	<b>Laboratorium 3 - Podstawy estymacji</b>	<b>2</b>
1.1	Estymator wartości oczekiwanej i wariancji - teoria . . . . .	2
1.2	Zadanie 1 - Estymatory wartości oczekiwanej i wariancji . . . . .	2
1.3	Zadanie 2 - Błędy empiryczne estymatora - rozkład Gaussa . . . . .	3
1.4	Zadanie 3 - Błędy empiryczne estymatora - rozkład Cauchy'ego . . . . .	5
1.5	Zadanie 1* - Estymator kowariancji i korelacji Pearsona - zależność liniowa . . . . .	7
1.6	Zadanie 2* - Estymator kowariancji i korelacji Pearsona - zależność kwadratowa . . . . .	8
<b>2</b>	<b>Laboratorium 4 - Dystrybuanta empiryczna i jej własności</b>	<b>9</b>
2.1	Zadanie 1 - Generator liczb losowych o zadanej funkcji gęstości . . . . .	9
2.2	Zadanie 2 - Dystrybuanta empiryczna . . . . .	10
2.3	Zadanie 3 - Odległość Kołmogorowa - Smirnova estymacji dystrybuanty . . . . .	11
2.4	Zadanie 4 - Identyfikacja dystrybuanty . . . . .	12
2.5	Zadanie 5 - Wariancja dystrybuanty empirycznej . . . . .	12
2.6	Zadanie 1* - Błąd empiryczny szacowanej dystrybuanty . . . . .	13
2.7	Zadanie 2* - Zbieżność dystrybuanty empirycznej . . . . .	15
<b>3</b>	<b>Laboratorium 5 - Jądrowy estymator gęstości prawdopodobieństwa</b>	<b>16</b>
3.1	Zadanie 1 - Generator liczb z rozkładu normalnego . . . . .	16
3.2	Zadanie 2 - Implementacja estymatora jądrowego gęstości prawdopodobieństwa . . . . .	16
3.3	Zadanie 3 - Funkcje jądra - Boxcar, Gaussian, Epanechnikov, Tricube . . . . .	17
3.4	Zadanie 4 - Błąd empiryczny szacowanej funkcji gęstości metodą jądrową . . . . .	19
<b>4</b>	<b>Laboratorium 6 - Nieparametryczna identyfikacja statycznych systemów nieliniowych.</b>	
	<b>Jądrowy estymator funkcji regresji</b>	<b>22</b>
4.1	Zadanie 1 - Statyczny system nieliniowy z addytywnym zakłóceniem na wyjściu . . . . .	22
4.2	Zadanie 2 - Sygnał wyjściowy systemu a charakterystyka nieliniowa . . . . .	22
4.3	Zadanie 3 - Jądrowy estymator funkcji regresji . . . . .	23
4.4	Zadanie 4 - Porównanie typów jąder estymacji funkcji regresji . . . . .	25
4.5	Zadanie 5 - Optymalny parametr wygładzania i jądro estymacji . . . . .	27
4.6	Zadanie 6 - Najlepsze estymacje systemu nieliniowego - ocena . . . . .	28
4.7	Zadanie 7 - Estymator jądrowy regresji a zakłócenie Cauchy'ego . . . . .	29
4.8	Zadanie 1* - Estymacja regresji metodą jądrową a niezerowa wartość oczekiwana zakłócenia . . . . .	32
	<b>Bibilografia</b>	<b>35</b>

# 1 Laboratorium 3 - Podstawy estymacji

## 1.1 Estymator wartości oczekiwanej i wariancji - teoria

[2]

Najefektywniejszy estymator wartości oczekiwanej

$$\hat{\mu}_N = \frac{1}{N} \sum_{n=1}^N X_n \quad (1)$$

Estymator wartości oczekiwanej jest zgodny (zbieżny do parametru według prawdopodobieństwa) - prawo wielkich liczb Markowa i Kołmogorowa. Estymator najefektywniejszy wartości oczekiwanej spełnia własność:  $Var\{\hat{\mu}_N\} = \frac{1}{N}\sigma^2$ . Z nierówności Cramera-Rao wynika, że taka postać jest najmniejszą możliwą wariancją estymatora wartości oczekiwanej.

Obciążony estymator wariancji

$$\hat{s}_N^2 = \frac{1}{N} \sum_{n=1}^N [X_n - \hat{\mu}_N]^2 \quad (2)$$

Wartość oczekiwana tak skonstruowanego estymatora jest zależna od liczebności populacji N:

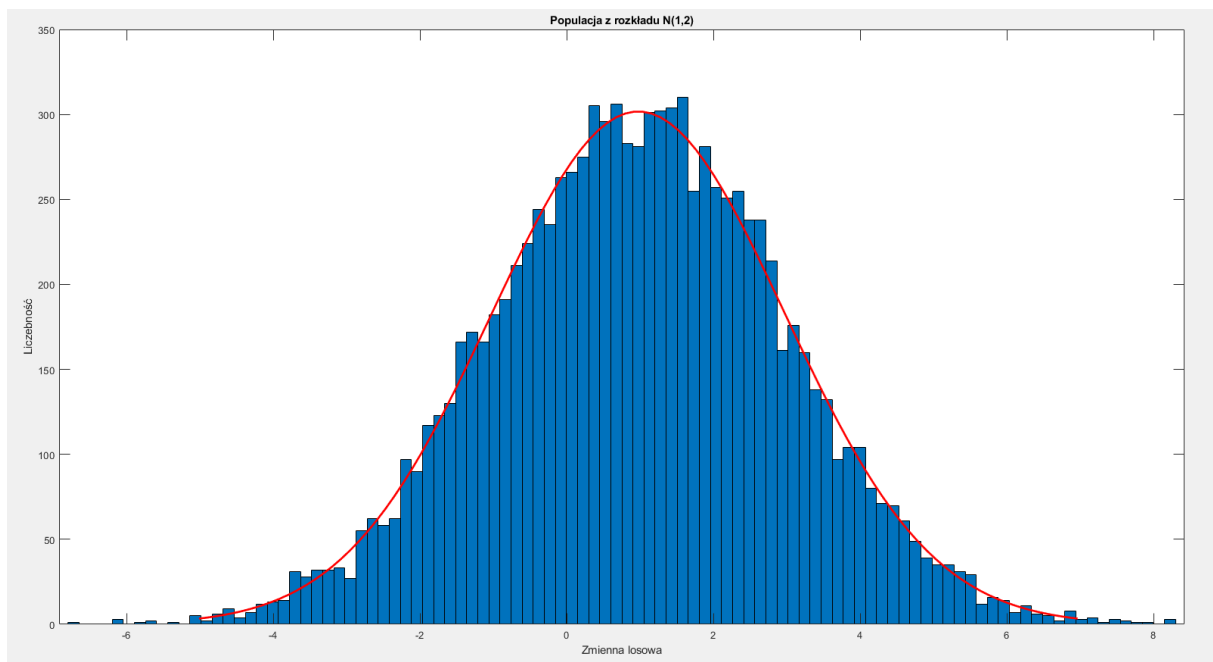
$$E\{\hat{s}_N^2\} = \frac{N-1}{N}\sigma^2$$

.

Nieobciążony estymator wariancji

$$\hat{S}_N^2 = \frac{1}{N-1} \sum_{n=1}^N [X_n - \hat{\mu}_N]^2 \quad (3)$$

## 1.2 Zadanie 1 - Estymatory wartości oczekiwanej i wariancji



Rysunek 1: Histogram wygenerowanej populacji w porównaniu do rozkładu  $N(1,2)$  -  $N = 10\,000$

Wartość oczekiwana:	$\mu = 1.0000$	$\hat{\mu}_N = 0.9748$
Wariancja:	$\sigma^2 = 4.0000$	$\hat{s}_N^2 = 3.9860$
		$\hat{S}_N^2 = 3.9864$

Tabela 1: Wartości wzorcowe wartości oczekiwanej i wariancji oraz ich estymatory

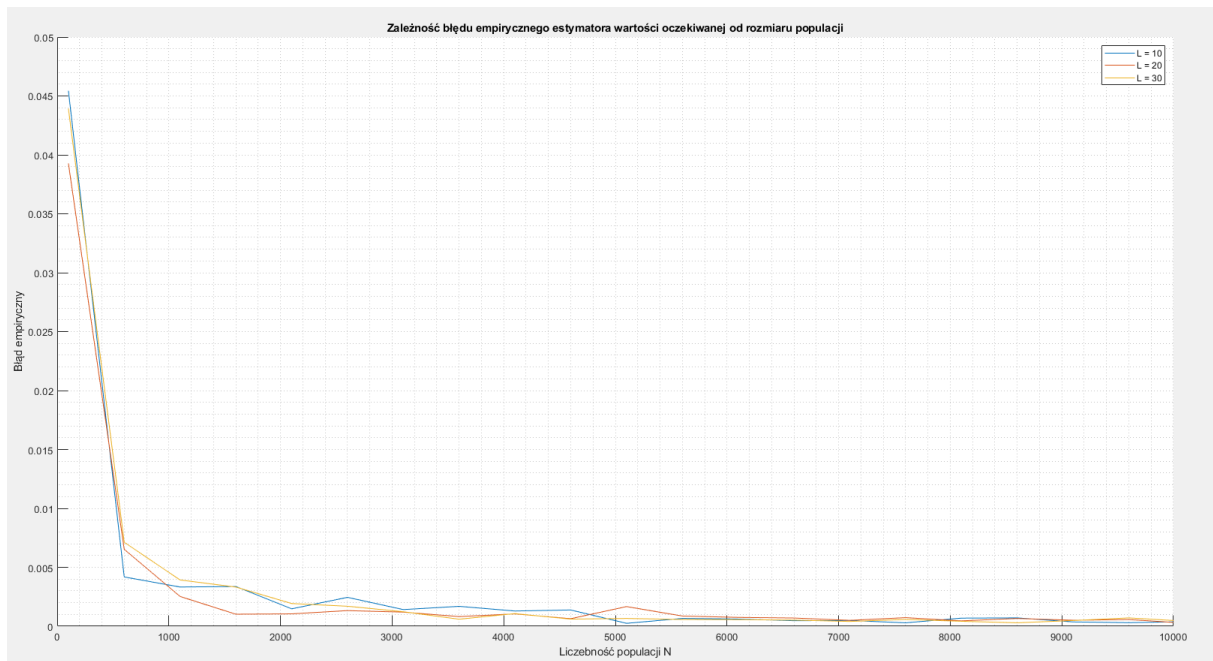
W zadaniu zaimplementowano generator  $N$  liczb z rozkładu  $\mathbf{N}(\mu, \sigma^2)$  oraz wyliczanie trzech estymatorów: wartości oczekiwanej, wariancji (obciążony), wariancji (nieobciążony). Rysunek (1) ukazuje poprawność generowanej populacji zgodnej z założeniami, natomiast tabela (1) przedstawia prawdziwą zadaną wartość oczekiwaną oraz wariancję, a także ich estymacje na podstawie zaimplementowanych estymatorów.

### 1.3 Zadanie 2 - Błędy empiryczne estymatora - rozkład Gaussa

**Cel:** Zadanie polega na wykreślaniu błęd empirycznych estymatorów dla różnych parametrów.

$$ERR\{\hat{\mu}_N; \mu\} = \frac{1}{L} \sum_{l=1}^L [\hat{\mu}_N^{[l]} - \mu]^2 \quad (4)$$

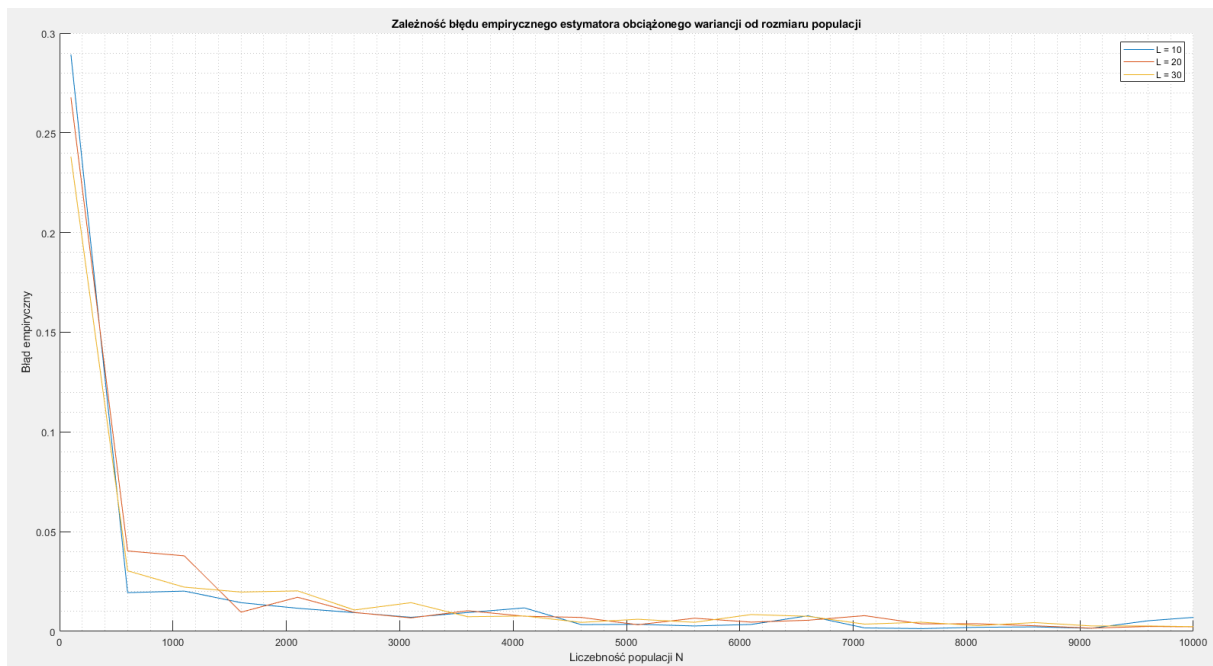
Wzór (4) określa błąd empiryczny dla estymatora wartości oczekiwanej.



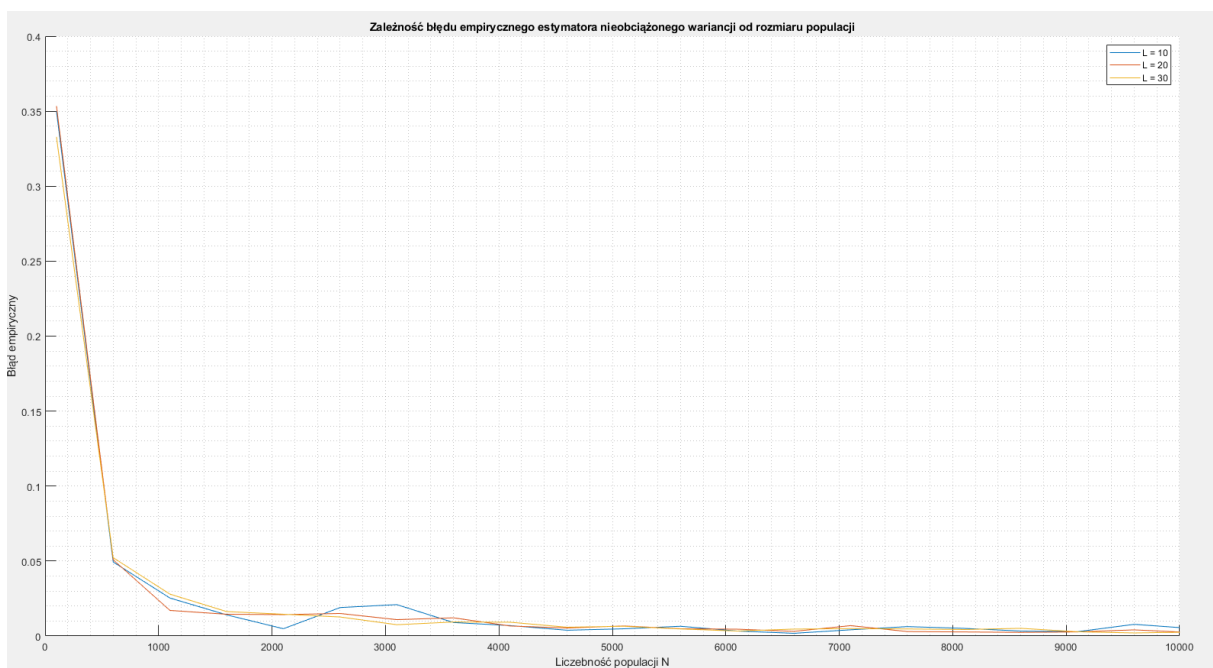
Rysunek 2: Zależność błęd empirycznych estymatorów wartości oczekiwanej od liczebności populacji dla trzech różnych wartości  $L$

**Komentarz** Rysunek (2) przedstawia zależność błęd empirycznych dla badanej populacji w zależności od jej liczebności. Wraz ze wzrostem liczby  $N$  błąd empiryczny maleje wykładniczo. Im większa liczba  $L$ , czyli im więcej prób wyznaczania błęd empirycznych, tym estymator błęd empirycznych zbliża się do wartości oczekiwanej błęd. Pozwala to uzyskiwać wykresy bliższe funkcji wykładniczej (bez losowych oscylacji).

Takie samo badanie przeprowadzono dla estymatorów wariancji: obciążonego i nieobciążonego.



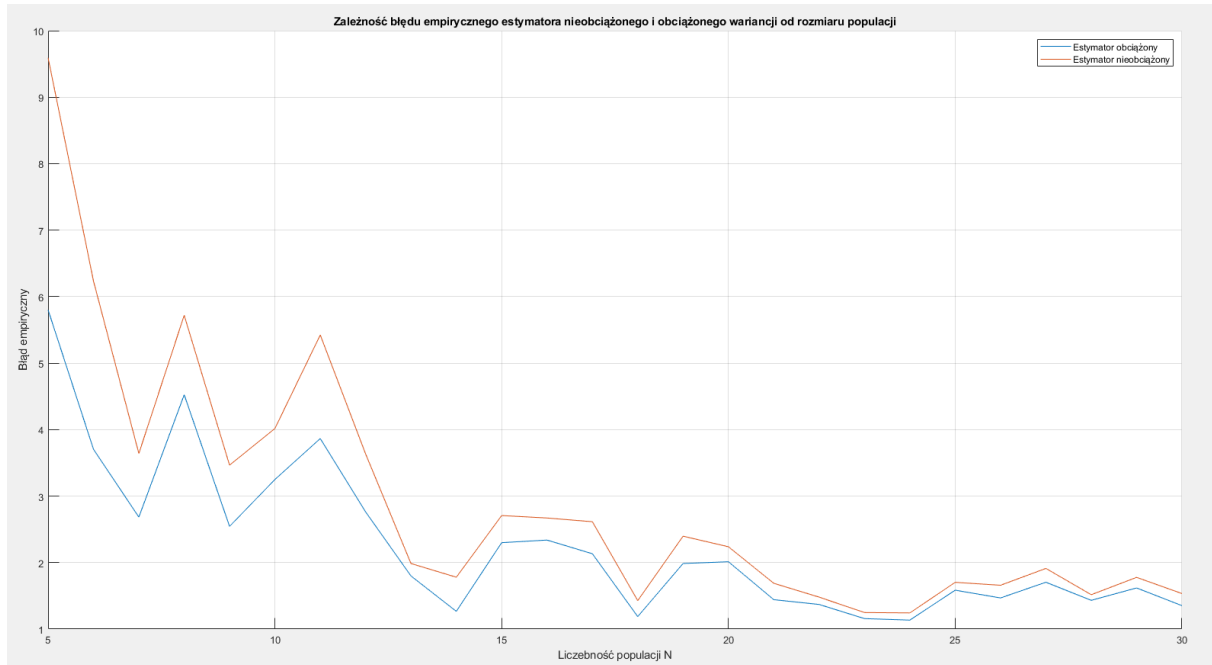
Rysunek 3: Zależność błędu empirycznego estymatora obciążonego wariancji od liczebności populacji dla trzech różnych wartości  $L$



Rysunek 4: Zależność błędu empirycznego estymatora nieobciążonego wariancji od liczebności populacji dla trzech różnych wartości  $L$

**Komentarz** Błędy empiryczne estymatorów wariancji również wraz ze wzrostem liczebności populacji maleją wykładniczo do zera. Im większa liczba  $L$ , tym linia spadku błędu dąży do funkcji wykładniczej malejącej (mniej oscylacji).

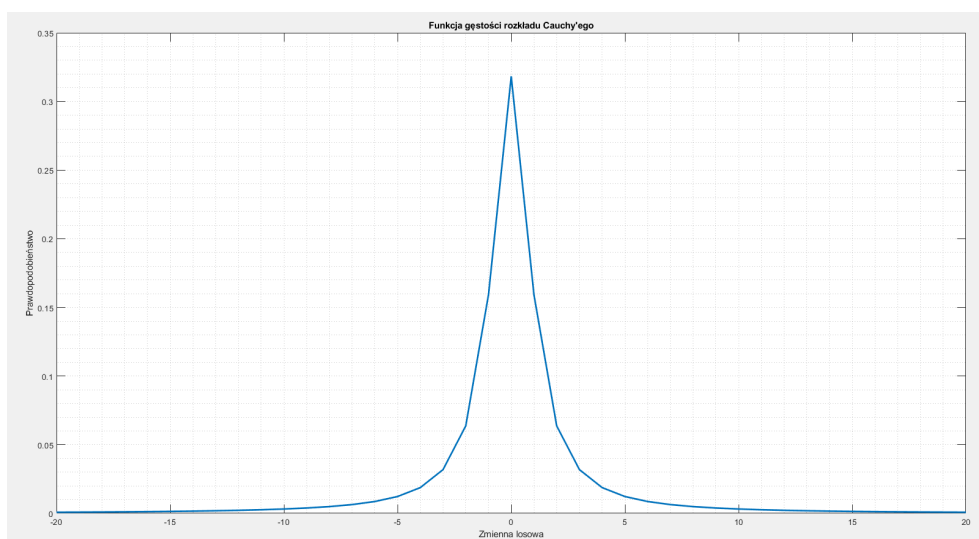
Zbadano zbieżność błędu empirycznego do zera dla estymatora wariancji obciążonego i nieobciążonego ( $L = 30$ ). Wartość oczekiwana estymatora obciążonego wariancji nie równa się szacowanej wariancji. Wartość oczekiwana zbiega do wartości szacowanej wraz ze zwiększaniem się liczby  $N$ , gdyż  $\frac{N-1}{N} \rightarrow 1$ . Estymator obciążony to estymator największej wiarygodności (maximum likelihood estimator), jednak jego obciążenie jest niezerowe. Estymator nieobciążony ma z kolei większą wariancję, a to powoduje, że średni błąd kwadratowy obciążonego estymatora jest nieco mniejszy niż nieobciążonego. Szczególnie widać to na rysunku (5).



Rysunek 5: Zależność błędu empirycznego estymatora nieobciążonego wariancji od liczebności populacji dla trzech różnych wartości  $L$

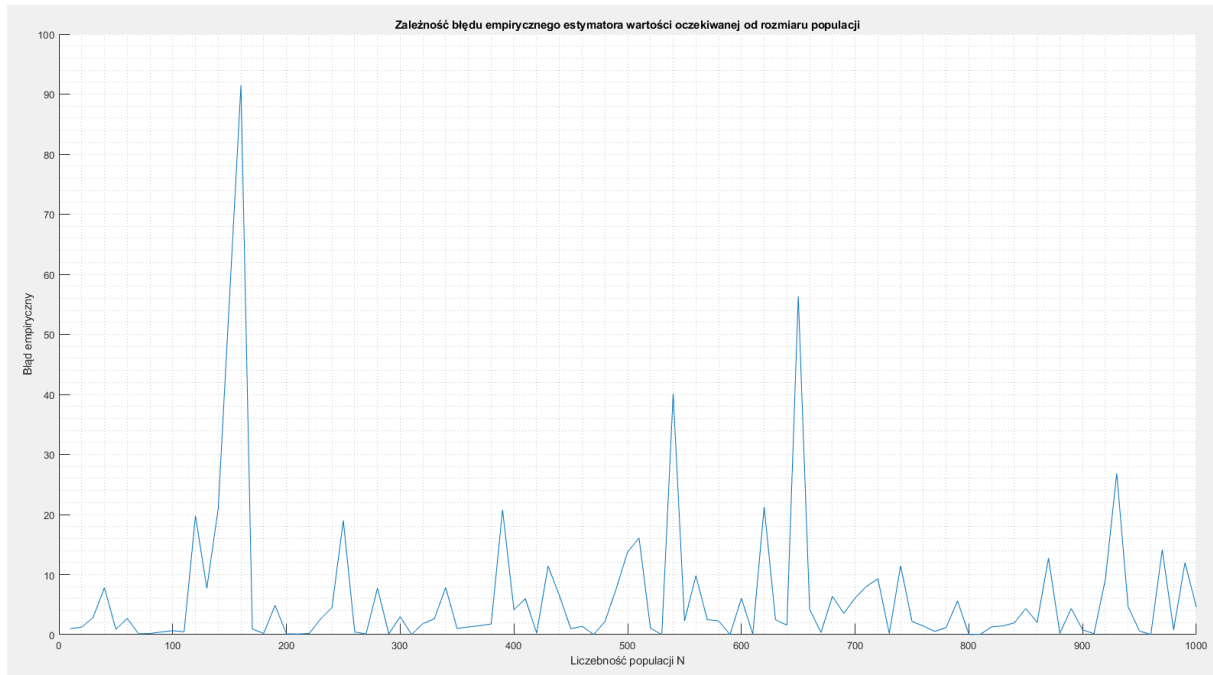
### 1.4 Zadanie 3 - Błędy empiryczne estymatora - rozkład Cauchy'ego

**Teoria** Rozkład Cauchy'ego jest szczególnym przypadkiem, dla którego niemożliwe jest wyznaczenie wartości oczekiwanej oraz wariancji. Zmienne z rozkładu Cauchy'ego mogą być ilorazami dwóch niezależnie wygenerowanych zmiennych z rozkładu normalnego.

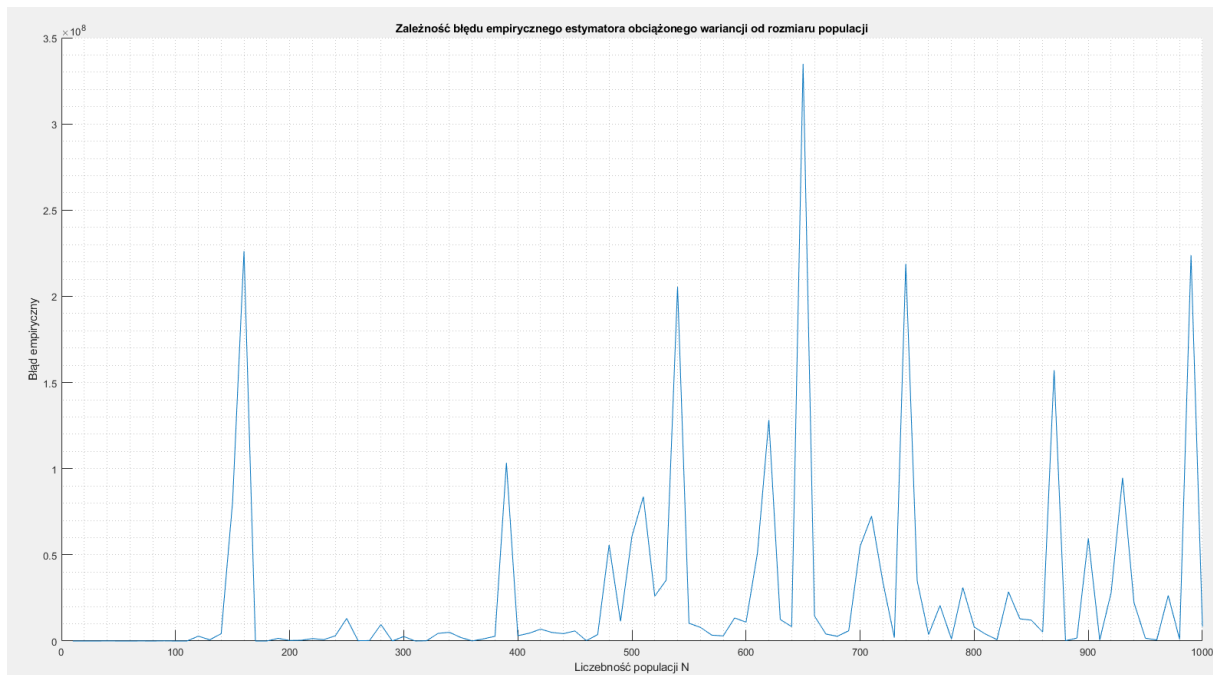


Rysunek 6: Funkcja gęstości rozkładu Cauchy'ego

Sprawdzono zachowanie błędów empirycznych estymatora wartości oczekiwanej i estymatora obciążonego wariancji. Przyjęto  $L = 30$ .



Rysunek 7: Błędy empiryczne estymatora wartości oczekiwanej w funkcji N



Rysunek 8: Błędy empiryczne estymatora obciążonego wariancji w funkcji N

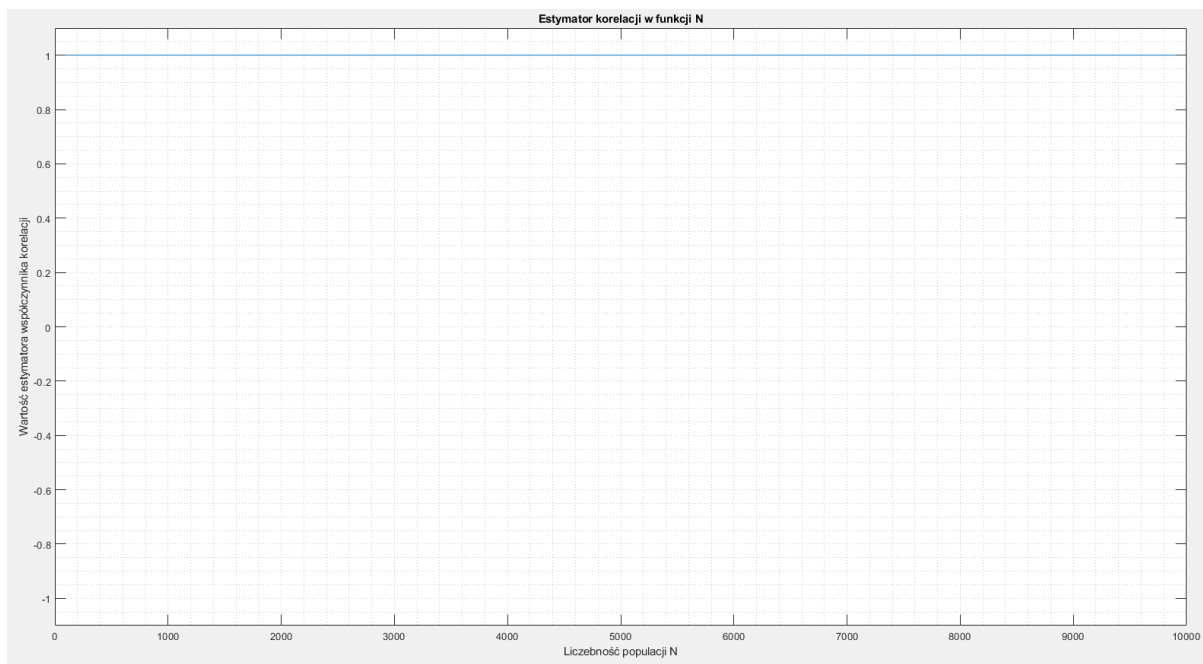
Potwierdzono na rys. 7 oraz na rys. 8, że nie jest możliwe wyznaczenie wartości oczekiwanej oraz wariancji dla rozkładu Cauchy'ego. Wraz ze wzrostem liczebności populacji ani wartość oczekiwana, ani wariancja nie zbiegają do żadnej wartości. Możliwe za pomocą mediany jest określenie skupienia układu, które w tym przypadku wynosi 0.

### 1.5 Zadanie 1\* - Estymator kowariancji i korelacji Pearsona - zależność liniowa

Zaimplementowano estymator korelacji Pearsona zadany wzorem [3]

$$r_{xy} = \frac{\frac{1}{N} \sum_{n=1}^N (X_n - \bar{X}_N)(Y_n - \bar{Y}_N)}{\sqrt{\left(\frac{1}{N} \sum_{n=1}^N (X_n^2 - \bar{X}_N^2)\right) \left(\frac{1}{N} \sum_{n=1}^N (Y_n^2 - \bar{Y}_N^2)\right)}} \quad (5)$$

Zadana korelacja Pearsona określa poziom zależności liniowej. W zadaniu przyjęto, że zmienna  $X_1$  jest z rozkładu jednostajnego  $(0, 1)$ , natomiast  $X_2 = 2X_1 + 2$ . Wykreślono realizacje estymatora korelacji w funkcji liczności próby  $N$ .

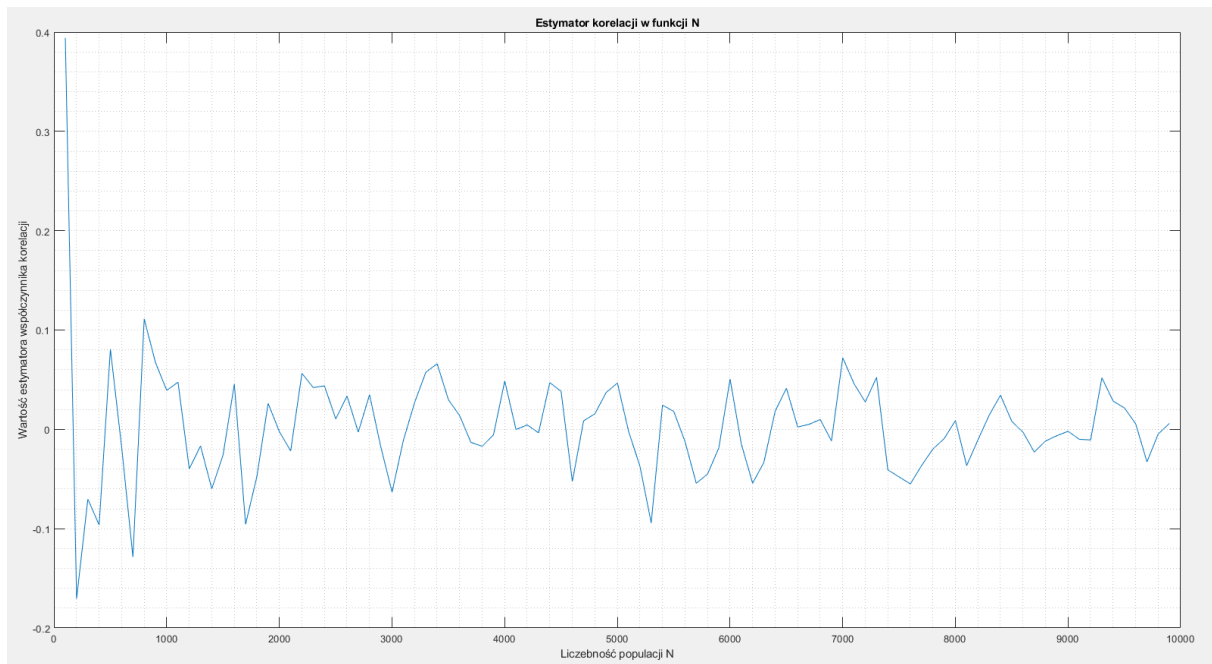


Rysunek 9: Błędy empiryczne estymatora obciążonego wariancji w funkcji  $N$



## 1.6 Zadanie 2\* - Estymator kowariancji i korelacji Pearsona - zależność kwadratowa

Zmienna losowa  $X_1$  wygenerowana została z rozkładu normalnego. Pomiedzy zmienną  $X_1$  a  $X_2$  występuje kwadratowa zależność taka że  $X_2 = X_1^2$ .



Rysunek 10: Błędy empiryczne estymatora obciążonego wariancji w funkcji N

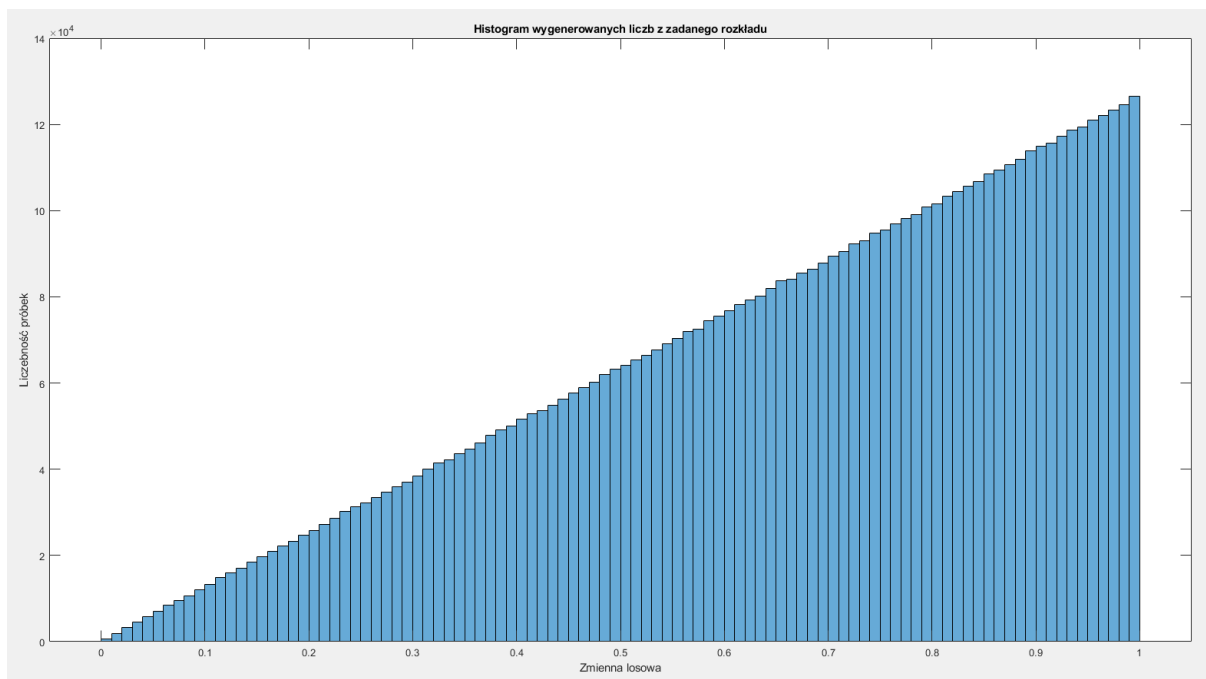
**Komentarz** Estymator prawidłowo wskazał korelację dla liniowej zależności. Bezużyteczny okazał się jednak, gdy występowała zależność kwadratowa, a estymator korelacji oscylował wokół zera.

## 2 Laboratorium 4 - Dystrybuanta empiryczna i jej własności

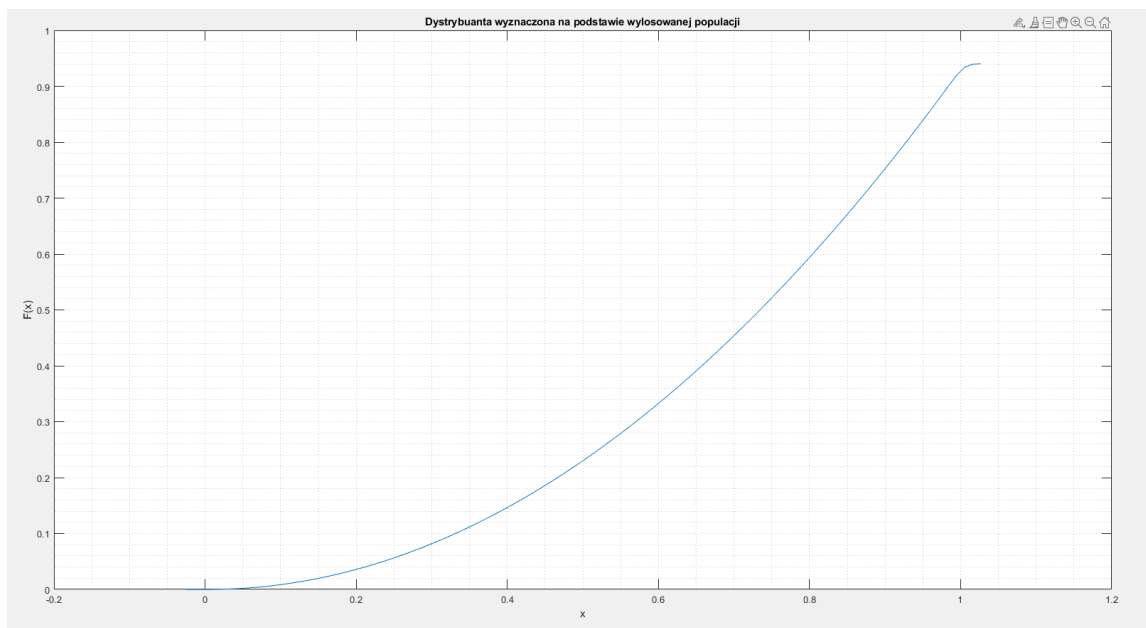
### 2.1 Zadanie 1 - Generator liczb losowych o zadanej funkcji gęstości

Zaimplementowano generator liczb losowych metodą odwrotnej dystrybuanty o zadanej funkcji gęstości

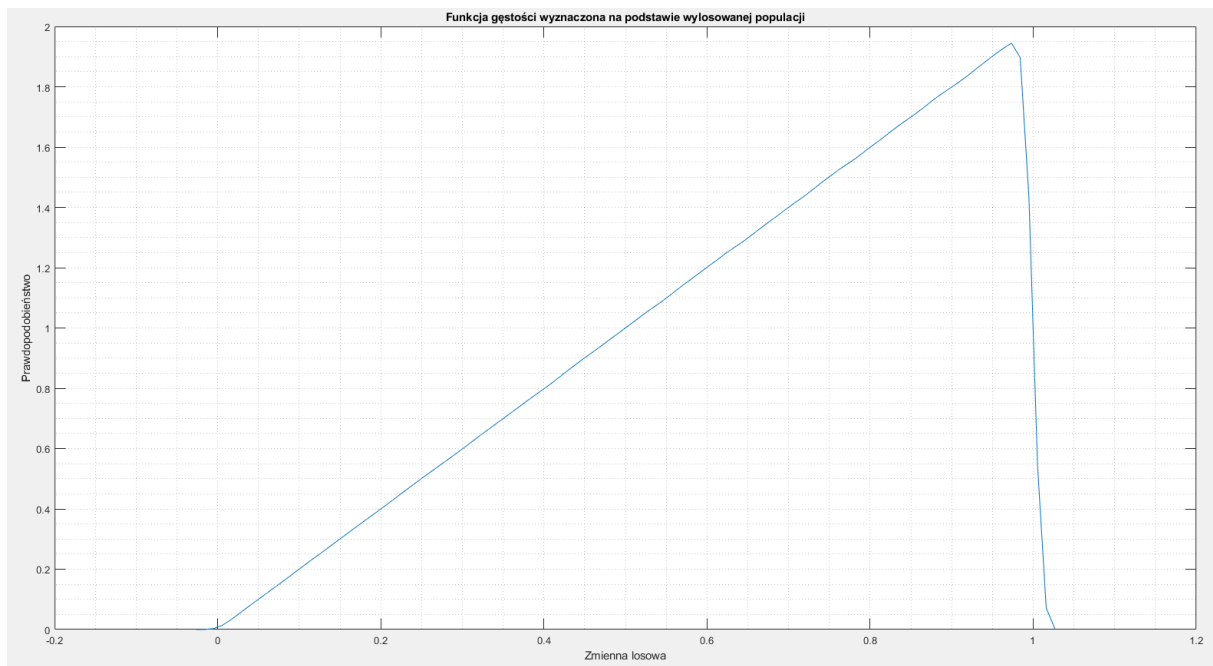
$$f(x) = \begin{cases} 2x & \text{dla } x \in [0, 1] \\ 0 & \text{dla } x \notin [0, 1] \end{cases} \quad (6)$$



Rysunek 11: Histogram 1 000 000 wygenerowanych próbek z rozkładu (6)



Rysunek 12: Wykres dystrybuanty wyznaczonej na podstawie wylosowanej populacji



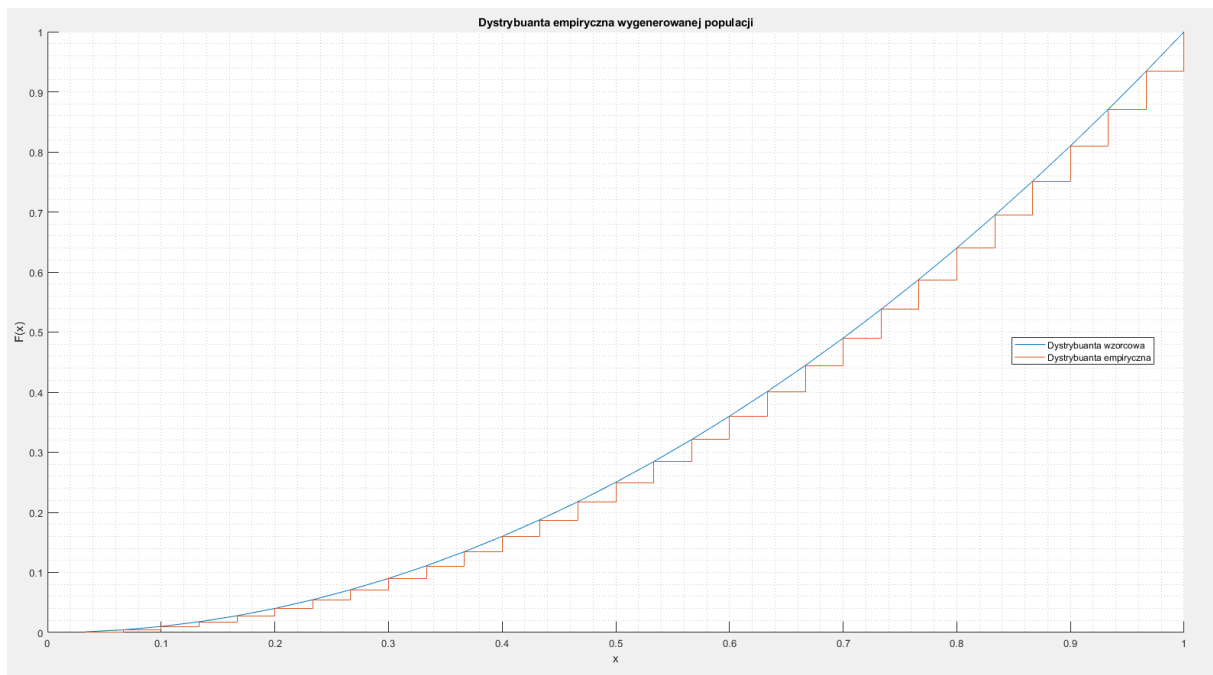
Rysunek 13: Wykres funkcji gęstości wyznaczonej na podstawie wylosowanej populacji

## 2.2 Zadanie 2 - Dystrybuanta empiryczna

Istnieje estymator dystrybuanty empirycznej zadany wzorem: [1]

$$\hat{F}_N(x) = \frac{1}{N} \sum_{n=1}^N \mathbb{I}(X_n \leq x), \quad -\infty < x < \infty, \quad (7)$$

gdzie  $\mathbb{I}$  jest równy 1, gdy  $X_n \leq x$ , lub równą 0, gdy  $X_n > x$ .



Rysunek 14: Wyznaczona dystrybuanta empiryczna i dystrybuanta wzorcowa

Dystrybuanta zadana jest wzorem analitycznym:

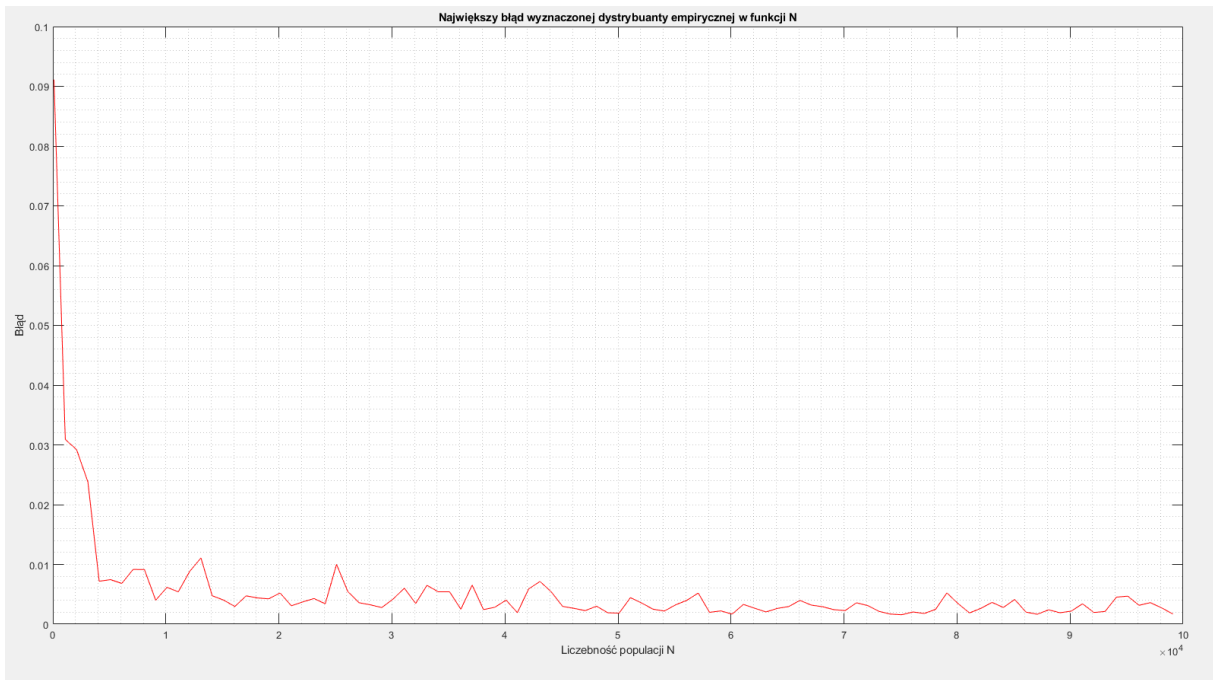
$$F(x) = \begin{cases} 0 & \text{dla } x < 0 \\ x^2 & \text{dla } x \in [0, 1] \\ 1 & \text{dla } x > 1 \end{cases} \quad (8)$$

**Komentarz** Dystrybuanta wzorcowa zadana wzorem (8) ukazana została na rysunku (14). Jej przybliżeniem jest dystrybuanta empiryczna uzyskana na drodze estymacji według wzoru (7).

### 2.3 Zadanie 3 - Odległość Kołmogorowa - Smirnova estymacji dystrybuanty

Dla estymatora dystrybuanty empirycznej wyznaczono błąd na podstawie zadanego wzoru odległości Kołmogorowa - Smirnova:

$$D_N = \sup_x |\hat{F}_N(x) - F(x)| \quad (9)$$



Rysunek 15: Największy błąd dystrybuanty empirycznej w zależności od N

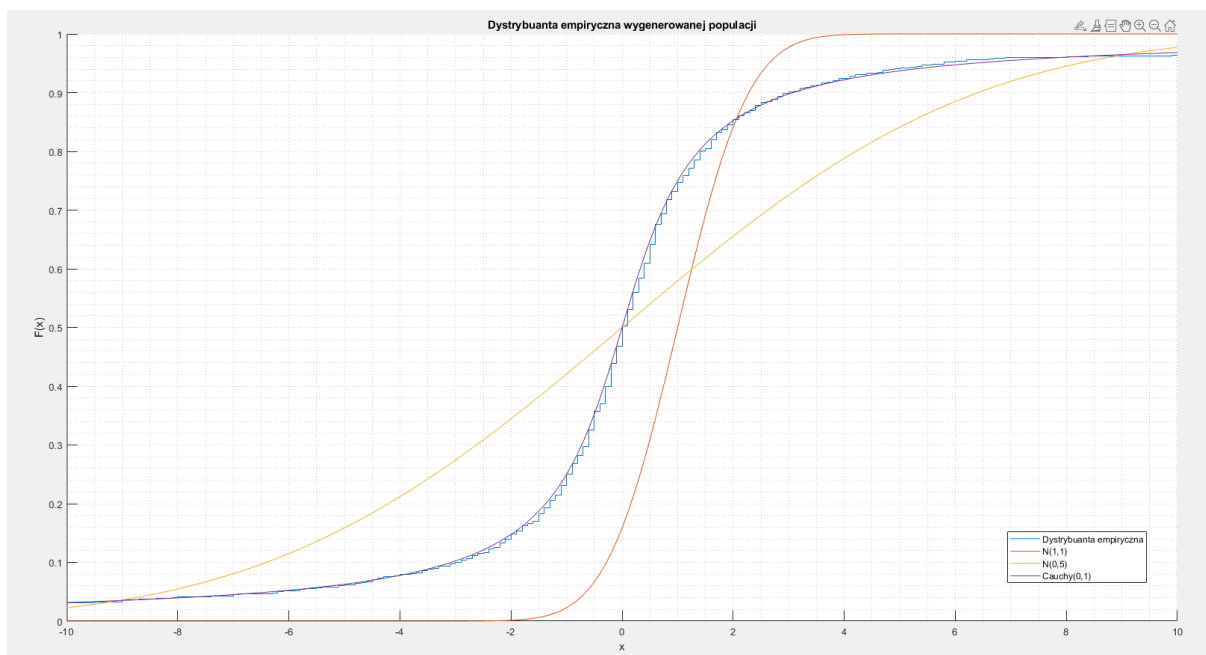
**Komentarz** Według twierdzenia Gliwienko- Cantelli gdy rozmiar populacji dąży do nieskończoności, wówczas dystrybuanta empiryczna zbiega do założonej dystrybuanty, ponieważ odległość Kołmogorowa - Smirnova dąży do zera. Wynika z tego jednoznacznie, że największy błąd dystrybuanty empirycznej dąży do 0.

## 2.4 Zadanie 4 - Identyfikacja dystrybuanty

**Histogram** Na podstawie danych z pliku ModelowanieData4.txt wyrysowano dystrybuantę empiryczną. Zbadano, z którego z poniższych rozkładów pochodzi populacja danych:

- Normalny (1,1)
- Normalny (0,5)
- Cauchy'ego (0,1)

Dystrybuanty wzorcowe dodano na wykresie.



Rysunek 16: Dystrybuanta empiryczna z danych oraz dystrybuanty porównawcze

**Komentarz** Metodą porównawczą stwierdzono, że dane z nieznanego rozkładu pochodzą z rozkładu Cauchy'ego o parametrach  $x_0 = 0, \lambda = 1$ . Niemożliwe jest określenie wartości oczekiwanej, wariancji oraz kurtozy.

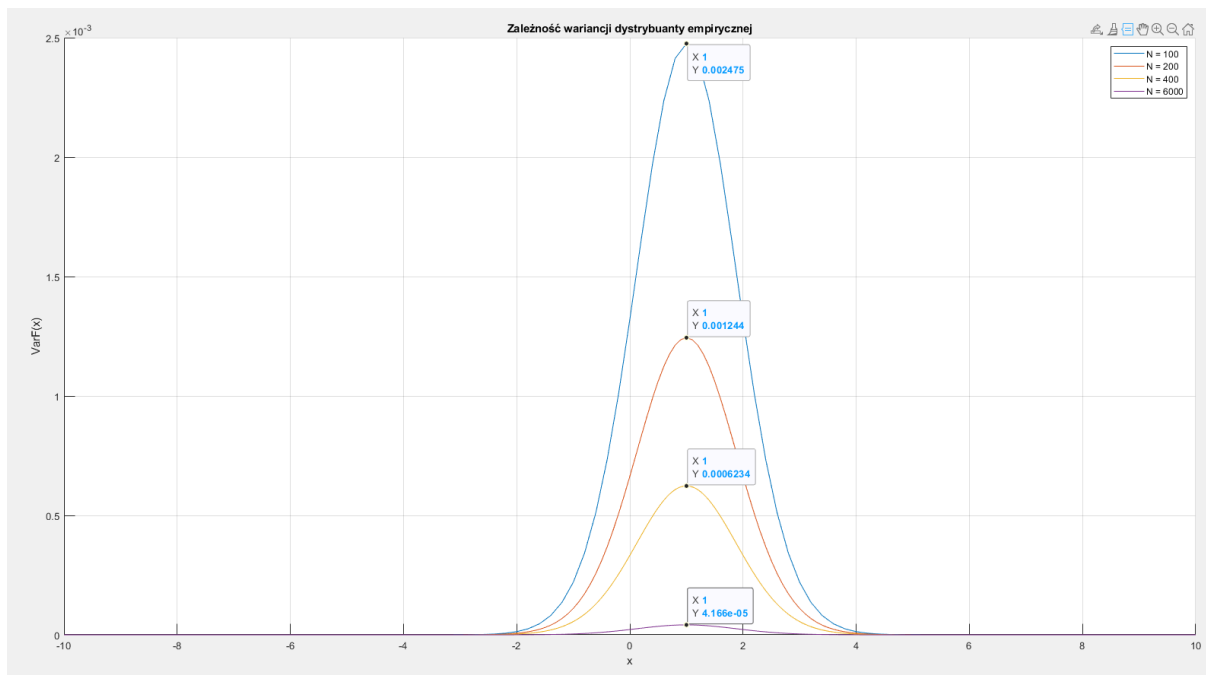
## 2.5 Zadanie 5 - Wariancja dystrybuanty empirycznej

Wariancja dystrybuanty empirycznej zdefiniowana jest za pomocą wzoru:

$$\text{Var}\{\hat{F}(x)\} = \frac{1}{N}F(x)(1 - F(x)) \quad (10)$$

Błąd średniokwadratowy MSE zbiega według prawdopodobieństwa do zera dla  $N \rightarrow \infty$ . Estymator dystrybuanty  $\hat{F}(x)$  jest mocno zgodny dla  $N \rightarrow \infty$  według prawa wielkich liczb Kołmogorowa. Zatem za dużych liczb dystrybuanta empiryczna zbiega do badanej dystrybuanty.

Zbadano zatem wpływ liczebności populacji na wariancję dystrybuanty.



Rysunek 17: Wariancja dystrybuanty empirycznej

**Komentarz** Rys. (17) wskazuje, że najwyższa wariancja dystrybuanty empirycznej rozkładu  $N(1, 1)$  występuje dla  $x = \mathbb{E}X(X) = 1$ . Wraz ze wzrostem liczby  $N$ , wariancja estymatora dystrybuanty w danym punkcie dąży do 0.

## 2.6 Zadanie 1\* - Błąd empiryczny szacowanej dystrybuanty

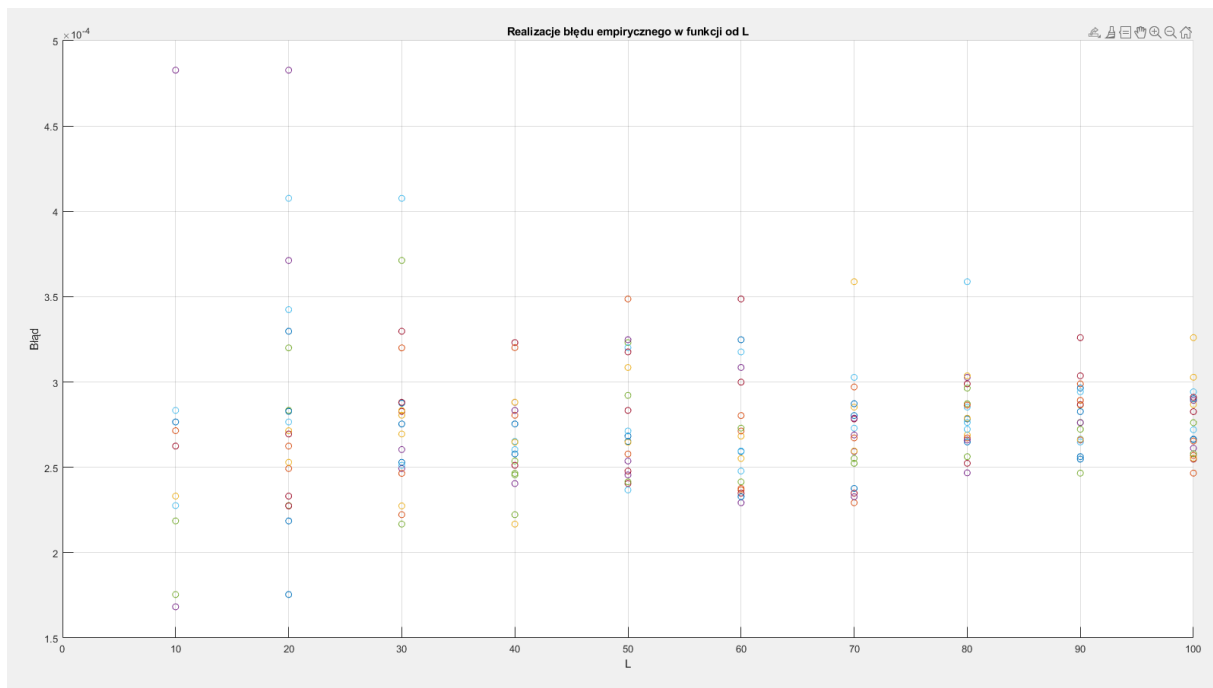
Błąd empiryczny dystrybuanty wyznaczono według wzoru:

$$Err\{\hat{F}_N\} = \frac{1}{LM} \sum_{l=1}^L \sum_{m=1}^M [\hat{F}_N^{[l]}(x_m) - F(x_m)]^2, \quad (11)$$

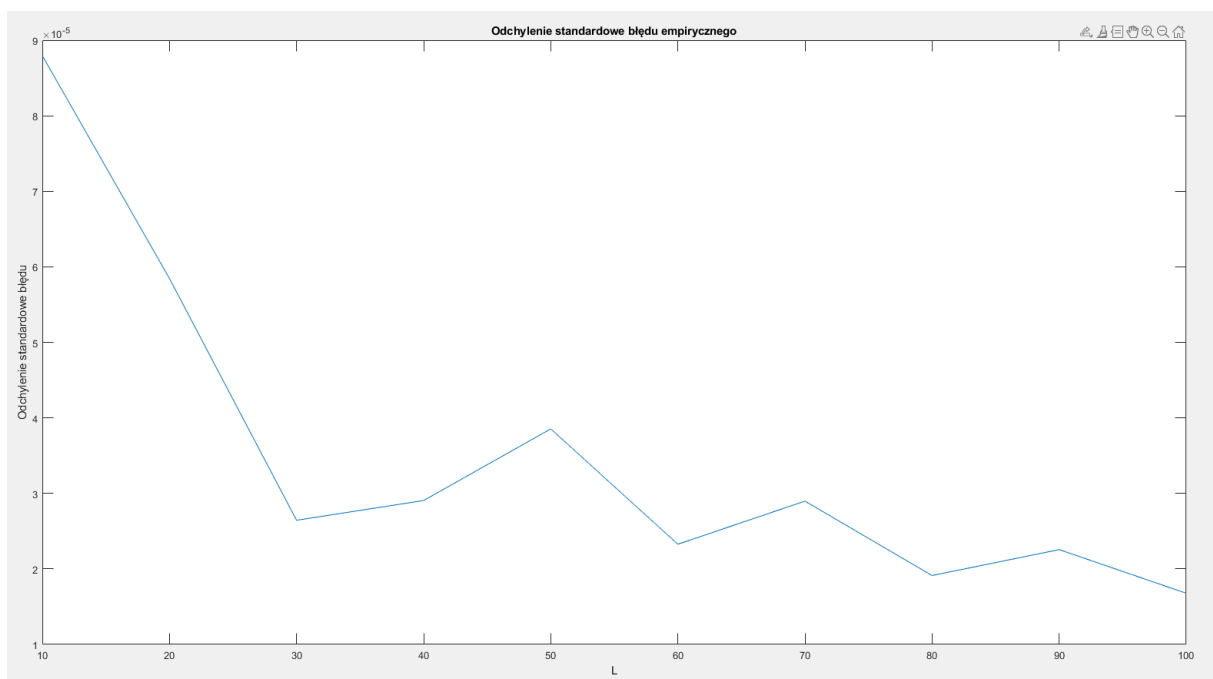
gdzie  $L, M$  są stałymi, które podlegają zmianie,  $\hat{F}_N^{[l]}(x_m)$  jest  $l$ -tą realizacją estymatora, a  $F(x_m)$  jest wartością w punkcie  $x_m$  rzeczywistej dystrybuanty. Błąd empiryczny dla szacowanej dystrybuanty rozkładu normalnego  $(1, 1)$  badano na odcinku  $(-10, 10)$ .

**Wpływ parametru  $L$**  Parametr  $L$  wpływa na wariancję otrzymywanego błędu empirycznego. Im większe  $L$  tym mniejszy rozrzut otrzymywanych błędów empirycznych. Wyniki badania zależności błędu od parametru  $L$  pokazano na rys. (18) i (20).

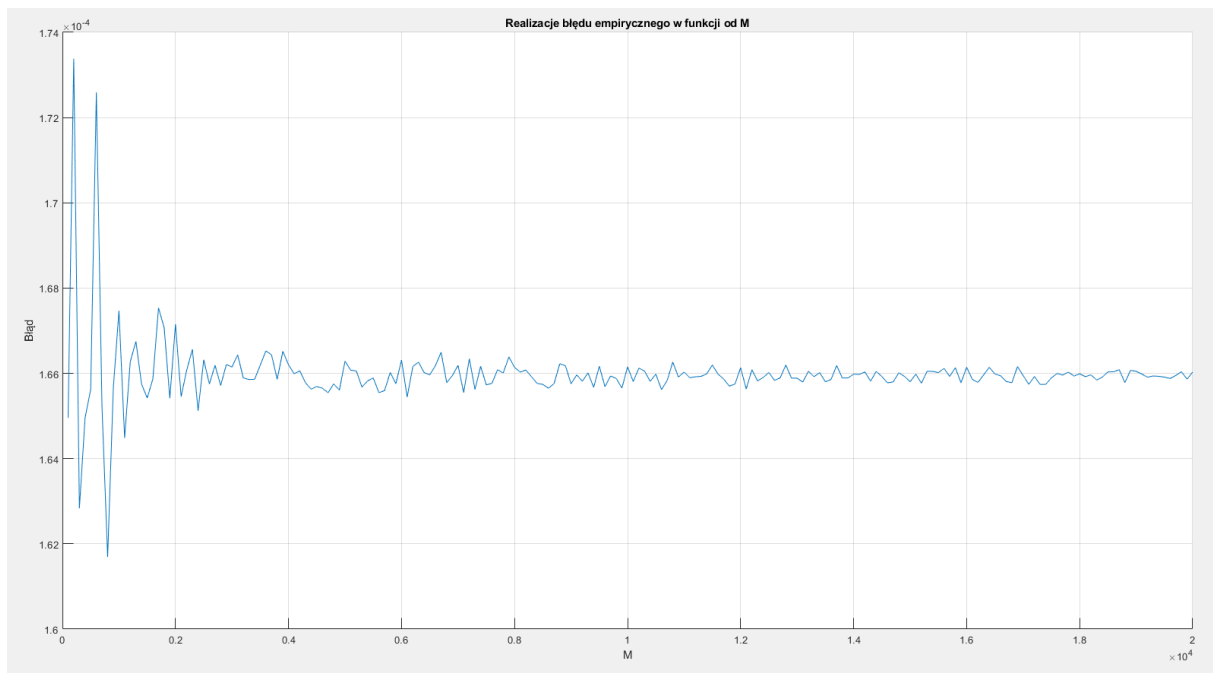
**Wpływ parametru  $M$**  Parametr  $M$  odpowiada za dokładność oszacowania błędu estymatora. Im większe  $M$ , tym błąd empiryczny estymatora zbiega do wartości oczekiwanej błędu (pole pomiędzy dystrybuantą empiryczną, a rzeczywistą dystrybuantą rozkładu).



Rysunek 18: Przykładowe realizacje błędu empirycznego estymatora w zależności od parametru L



Rysunek 19: Odchylenie standardowe błędu empirycznego estymatora w zależności od parametru L



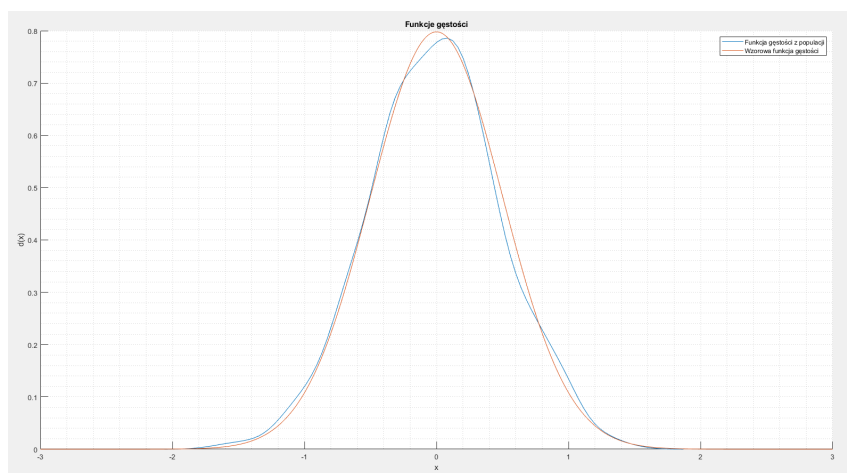
Rysunek 20: Błąd empiryczny w zależności od M

## 2.7 Zadanie 2\* - Zbieżność dystrybuanty empirycznej

Dystrybuanta empiryczna zbiega według rozkładu do rozkładu normalnego tzn.

$$\sqrt{(N)}[\hat{F}_N(x) - F(x)] \rightarrow N(0, (F(x)(1 - F(x))). \quad (12)$$

Eksperyment polega na wygenerowaniu  $N$  zmiennych losowych z rozkładu  $N(0, 1)$ . Następnie oblicza się  $\hat{F}_N(x)$  za pomocą funkcji określającej wartość dystrybuanty empirycznej w punkcie  $x = 0$ .  $F(x)$  to z kolei wartość rzeczywistej dystrybuanty rozkładu  $N(0, 1)$  w punkcie  $x = 0$ . Wartość  $\sqrt{(N)}[\hat{F}_N(x) - F(x)]$  podejrzana jest o pochodzenie z rozkładu  $N(0, (F(x)(1 - F(x)))$ . W celu potwierdzenia tej tezy wszystkie czynności powtarzane są 1000 razy, dzięki czemu uzyskuje się 1000 liczb. Następnie za pomocą funkcji "ksdensity" w pakiecie Matlab generowana jest funkcja gęstości z wszystkich liczb. Na tym samym wykresie przedstawia się wzorowy wykres funkcji gęstości rozkładu  $N(0, (F(x)(1 - F(x)))$ .



Rysunek 21: Funkcje gęstości rozkładu normalnego i z populacji ( $N = 10000$ )

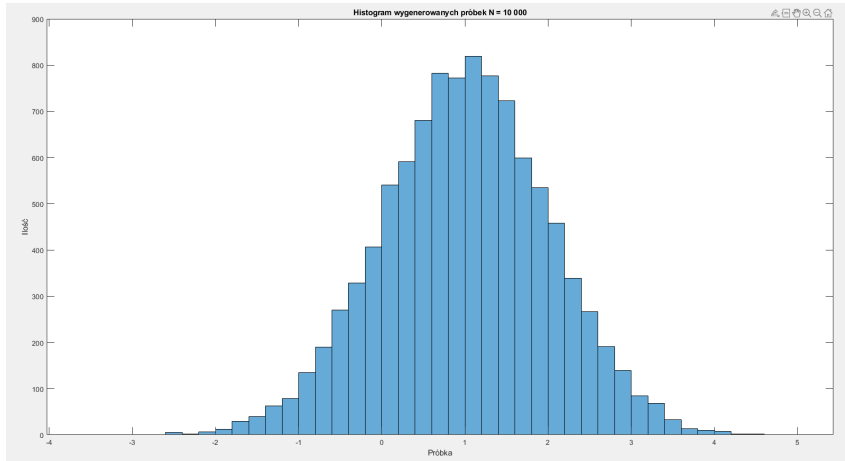
**Komentarz** Eksperyment potwierdza zbieżność dystrybuanty empirycznej według rozkładu do rozkładu normalnego według wzoru 12. Wykresy się pokrywają na rys. (21).



### 3 Laboratorium 5 - Jądrowy estymator gęstości prawdopodobieństwa

#### 3.1 Zadanie 1 - Generator liczb z rozkładu normalnego

Wygenerowano 10 000 próbek z rozkładu  $N(1, 1)$ . Histogram przedstawiono na rys. (22). Średnia populacji to 1.0053, natomiast odchylenie standardowe to 1.0027.



Rysunek 22: Histogram wygenerowanej populacji

#### 3.2 Zadanie 2 - Implementacja estymatora jądrowego gęstości prawdopodobieństwa

Jeżeli funkcja gęstości prawdopodobieństwa jest ciągła w punkcie  $x$ ,  $h_N \rightarrow 0$ ,  $Nh_N \rightarrow \infty$ , podczas gdy  $N \rightarrow \infty$ , a także  $K(\cdot)$  spełnia warunek

$$\int_{-\infty}^{\infty} K(x) dx = 1, \quad \int_{-\infty}^{\infty} x^2 K(x) dx = \sigma_K^2 \quad (13)$$

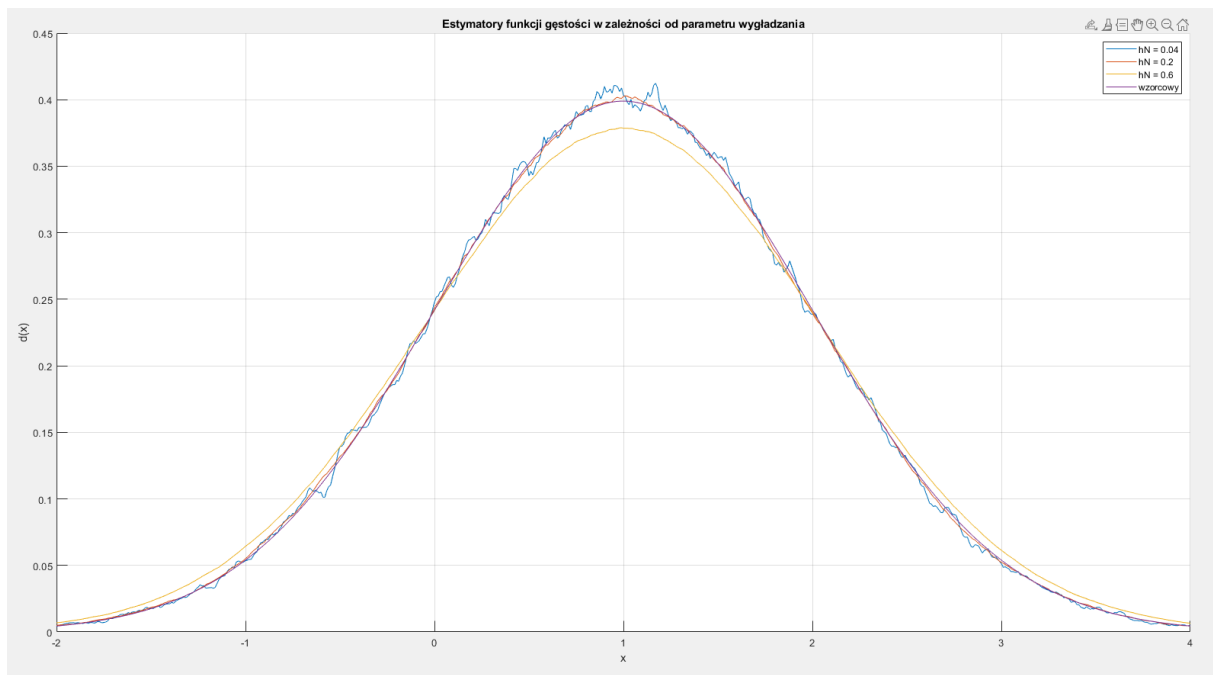
to jądrowy estymator gęstości prawdopodobieństwa (Rosenblatt - Parzen) [2]

$$\hat{f}_N(x) = \frac{1}{Nh_N} \sum_{n=1}^N K\left(\frac{X_n - x}{h_N}\right) \quad (14)$$

zbiega do rzeczywistej funkcji gęstości według prawdopodobieństwa.

Zaimplementowano jądrowy estymator gęstości prawdopodobieństwa, bazujący na jądrze Boxcar. Wygenerowano 100 000 próbek z rozkładu  $N(1, 1)$ . Następnie wyznaczono estymator funkcji gęstości odpowiednio z parametrami  $h_N$  równymi 0.04, 0.2, 0.6.

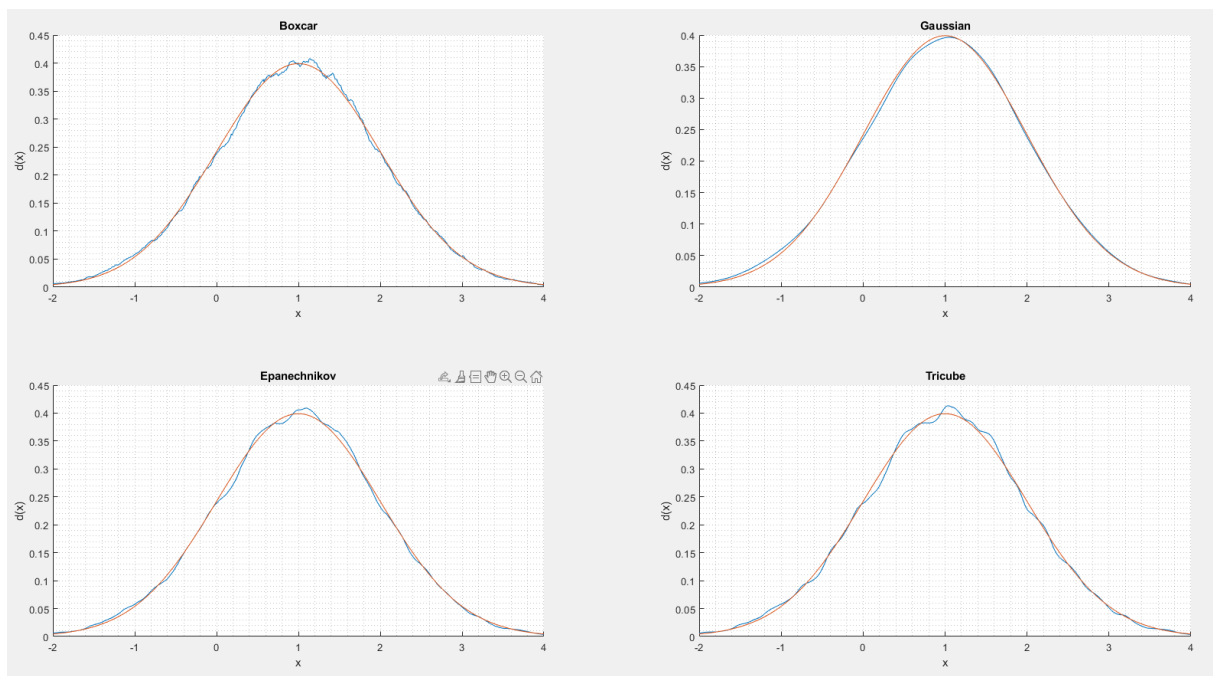
**Komentarz** Według rys. (23) najlepszym parametrem wygładzania okazał się  $h_N = 0.2$  - najbardziej zbliżona funkcja gęstości do oryginalnej. Parametr wygładzania wpływa na obecność oscylacji - im większy tym zmiany wartości funkcji mniejsze. Za duży parametr z kolei nie pozwala już poprawnie estymować oryginalnej funkcji.



Rysunek 23: Funkcje gęstości w zależności od parametru wygładzania

### 3.3 Zadanie 3 - Funkcje jądra - Boxcar, Gaussian, Epanechnikov, Tricube

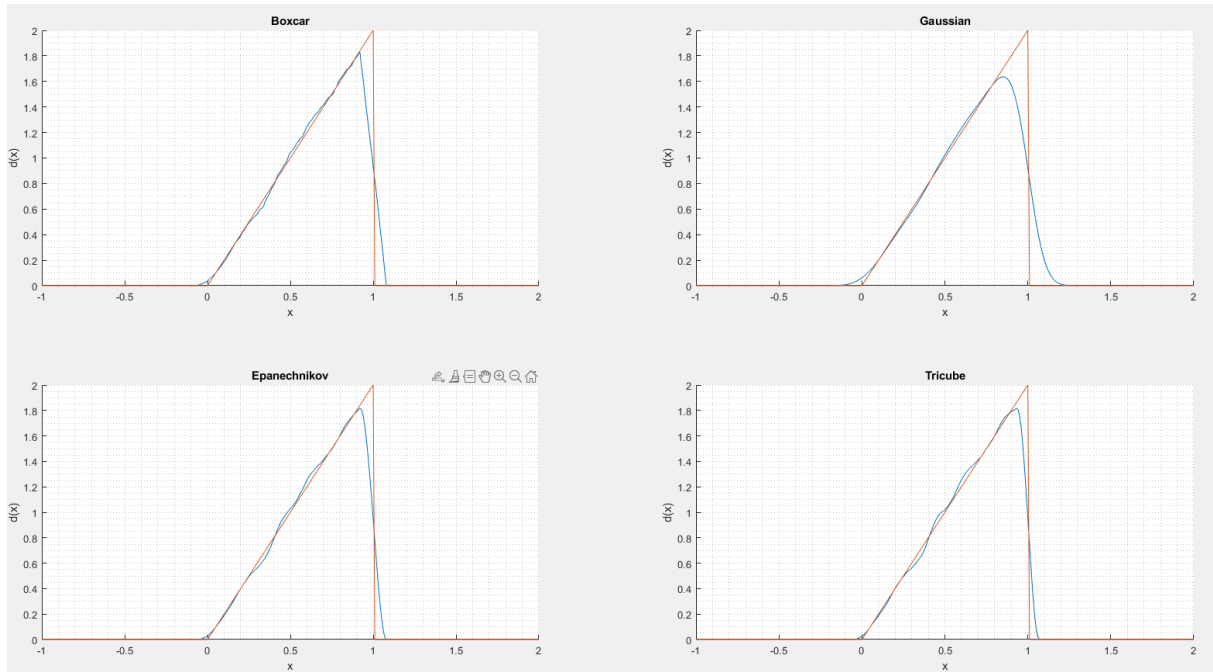
Zaimplementowano dodatkowe funkcje jądra - Boxcar, Epanechnikov, Tricube. Porównano efekty ich działania. Wygenerowano 10 000 próbek z rozkładu  $N(1, 1)$ .



Rysunek 24: Funkcje gęstości w zależności od funkcji jądra - rozkład normalny -  $h_N = 0.2$

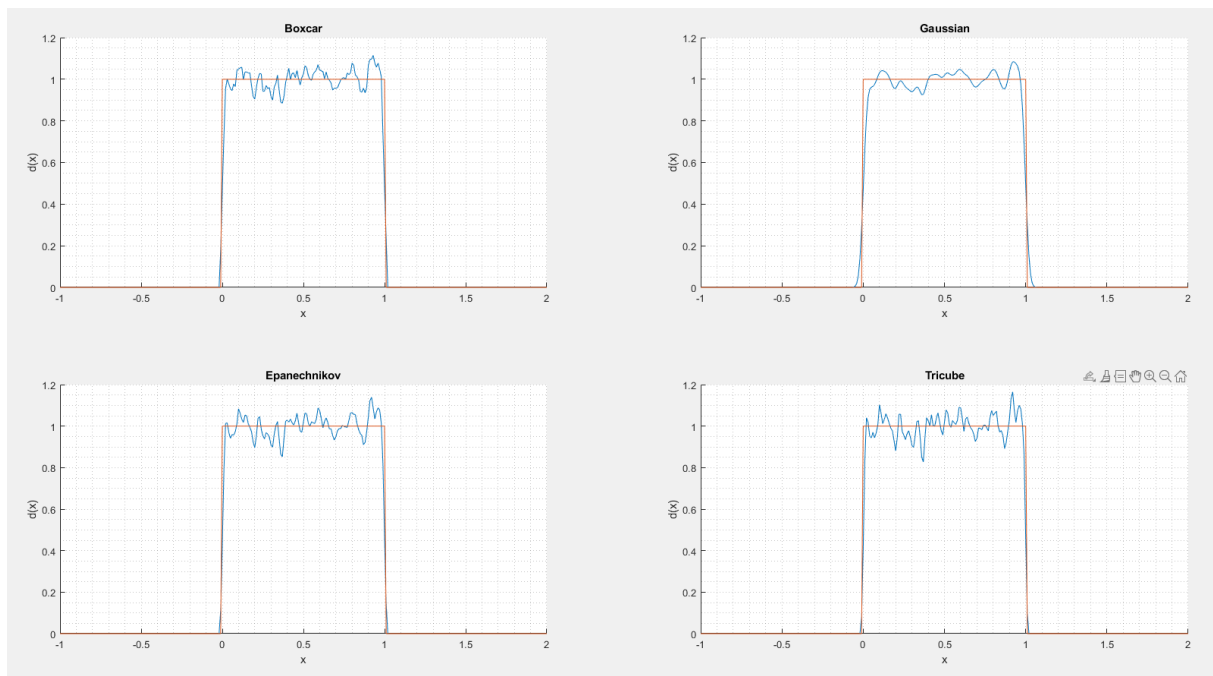
**Komentarz** Dla populacji wygenerowanej z rozkładu normalnego najlepiej poradziła sobie funkcja jądra gaussowskiego. Drugi rezultat uzyskała funkcja jądra Epanechnikova. Jądro boxcar wyróżniło się największymi oscylacjami, natomiast jądro tricube najgorszą estymacją gęstości prawdopodobieństwa.

Identyczne badanie przeprowadzono na próbkach wygenerowanych z rozkładu trójkątnego i jednostajnego.



Rysunek 25: Funkcje gęstości w zależności od funkcji jądra - rozkład trójkątny -  $h_N = 0.08$

**Komentarz** W przypadku rozkładu trójkątnego gaussowska funkcja jądra wypadła najgorzej. Jądra boxcar, epanechnikova i tricube pozwoliły na uzyskanie podobnych rezultatów.



Rysunek 26: Funkcje gęstości w zależności od funkcji jądra - rozkład jednostajny -  $h_N = 0.02$

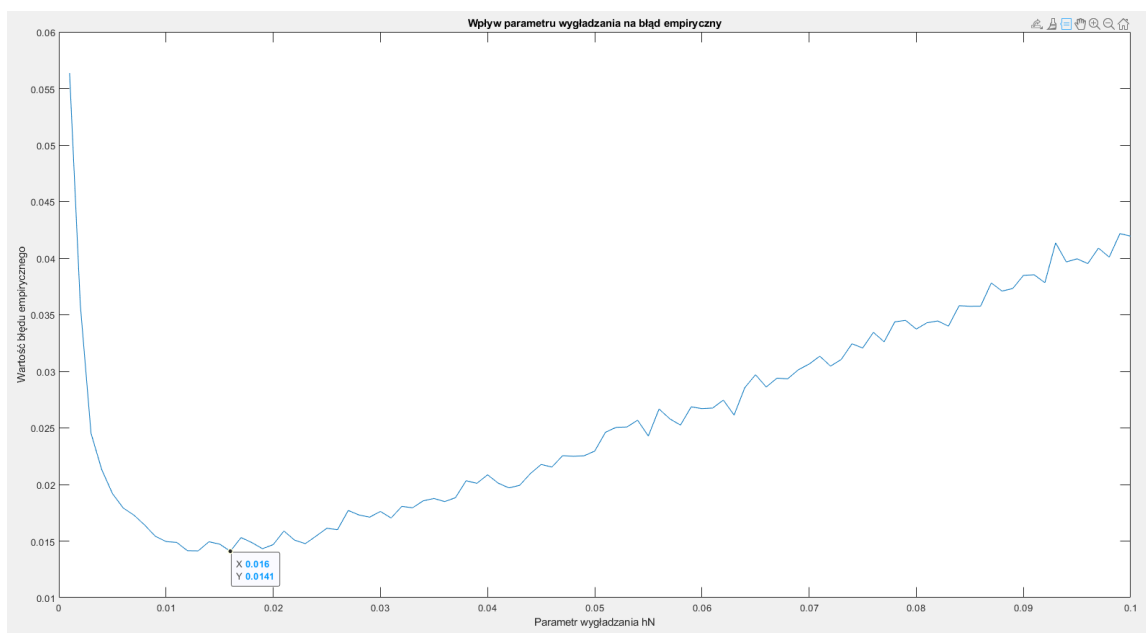
**Komentarz** W przypadku rozkładu jednostajnego funkcje jądra boxcar, epanechnikova i tricube poradzily sobie podobnie i lepiej niż gaussowskie. Najlepszym wyborem było natomiast jądro boxcar.

### 3.4 Zadanie 4 - Błąd empiryczny szacowanej funkcji gęstości metodą jądrową

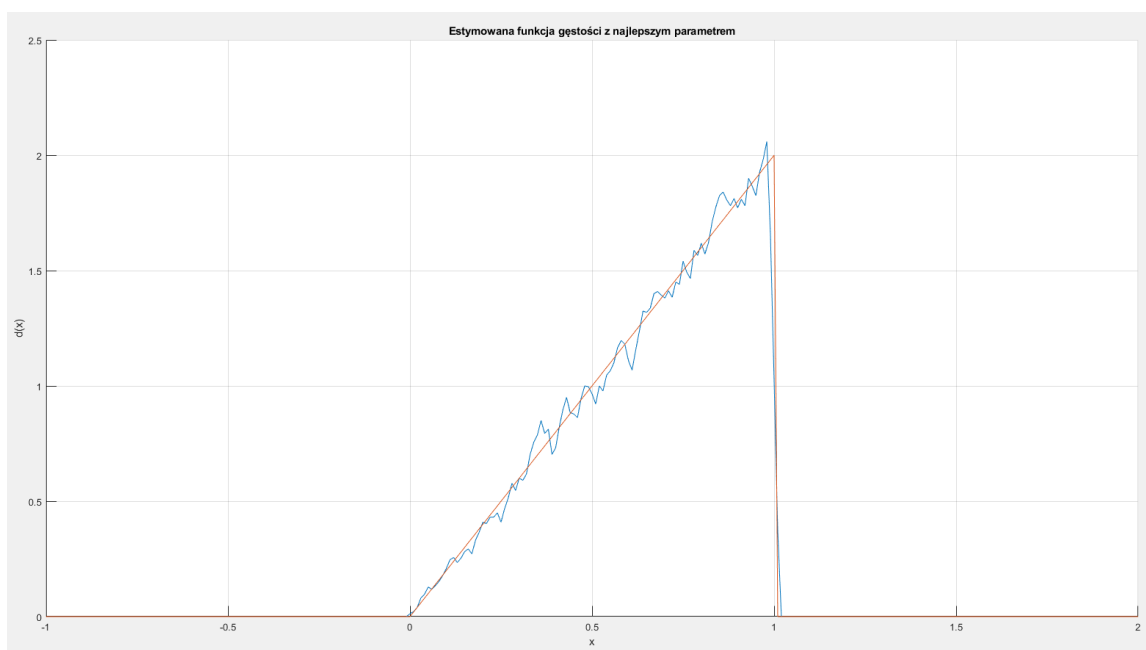
W celu wyboru najlepszego parametru  $h_N$  zaimplementowano możliwość wyliczania błędu empirycznego estymacji funkcji gęstości metodą jądrową za pomocą wzoru

$$Err\{\hat{f}_N\} = \frac{1}{LM} \sum_{l=1}^L \sum_{m=1}^M [\hat{f}_N^{[l]}(x_m) - f(x_m)]^2. \quad (15)$$

Przyjęto  $L = 10, M = 100$ . Pierwszy badany przypadek opiera się na funkcji boxcar jądra i rozkładzie trójkątnym. Za każdym razem generowano 10 000 próbek.



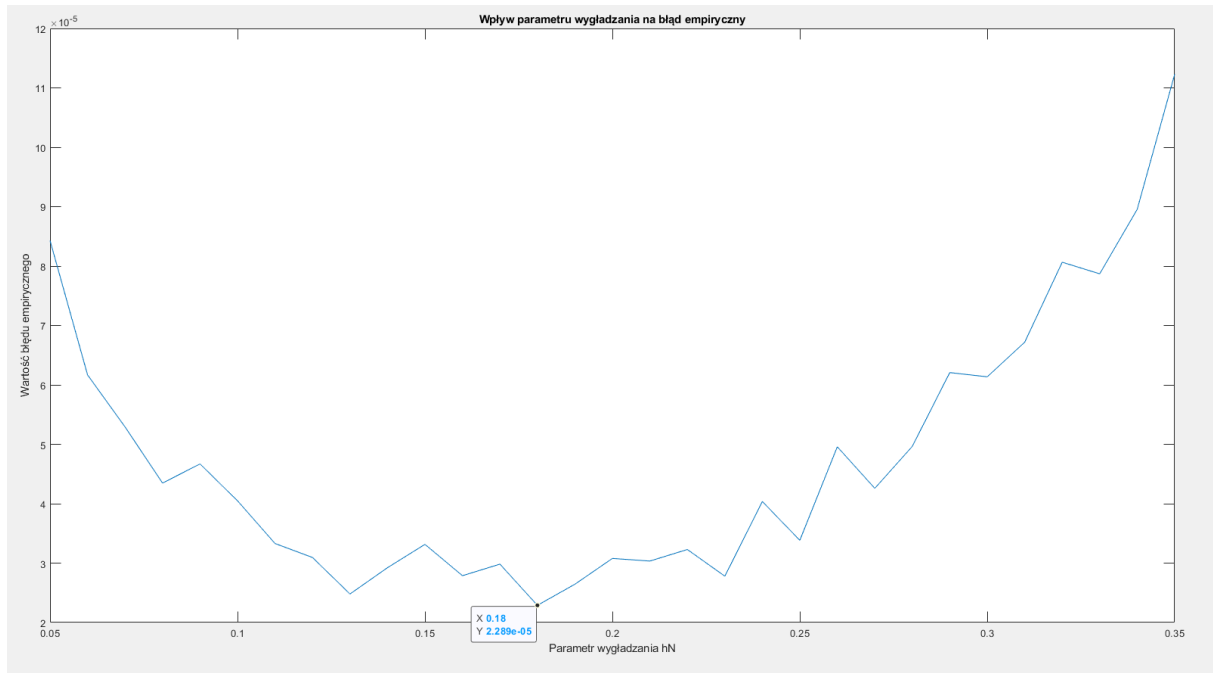
Rysunek 27: Błąd empiryczny w zależności od parametru  $h_N$



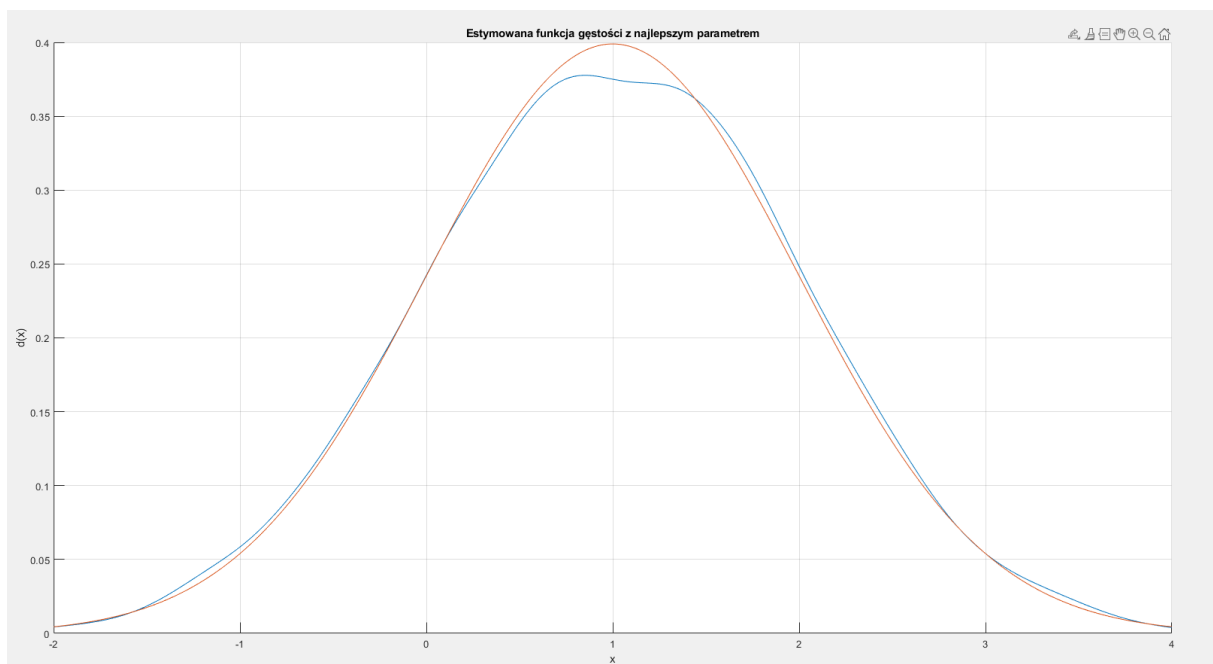
Rysunek 28: Estymata funkcji gęstości dla  $h_N = 0.016$

**Komentarz** Dla powyższej symulacji najlepszym parametrem okazał się  $h_N = 0.016$ .

Drugi badany przypadek opiera się na funkcji gaussowskiej jądra i rozkładzie normalnym  $\mathcal{N}(1, 1)$ . Za każdym razem generowano 10 000 próbek.



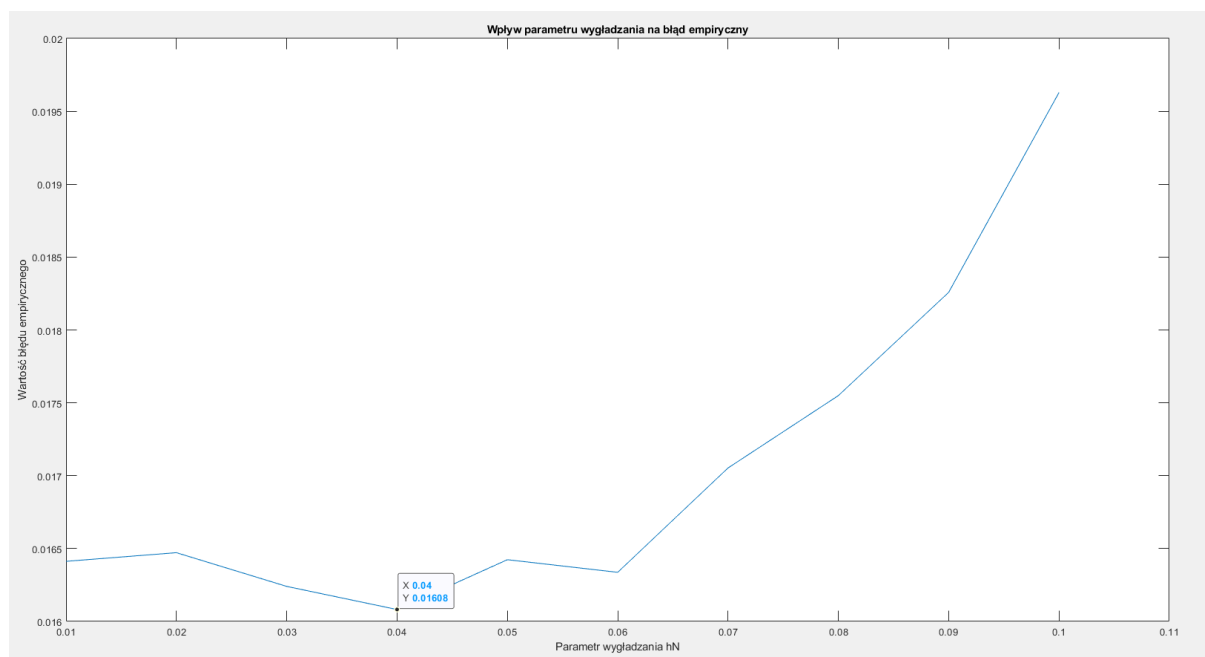
Rysunek 29: Błąd empiryczny w zależności od parametru  $h_N$



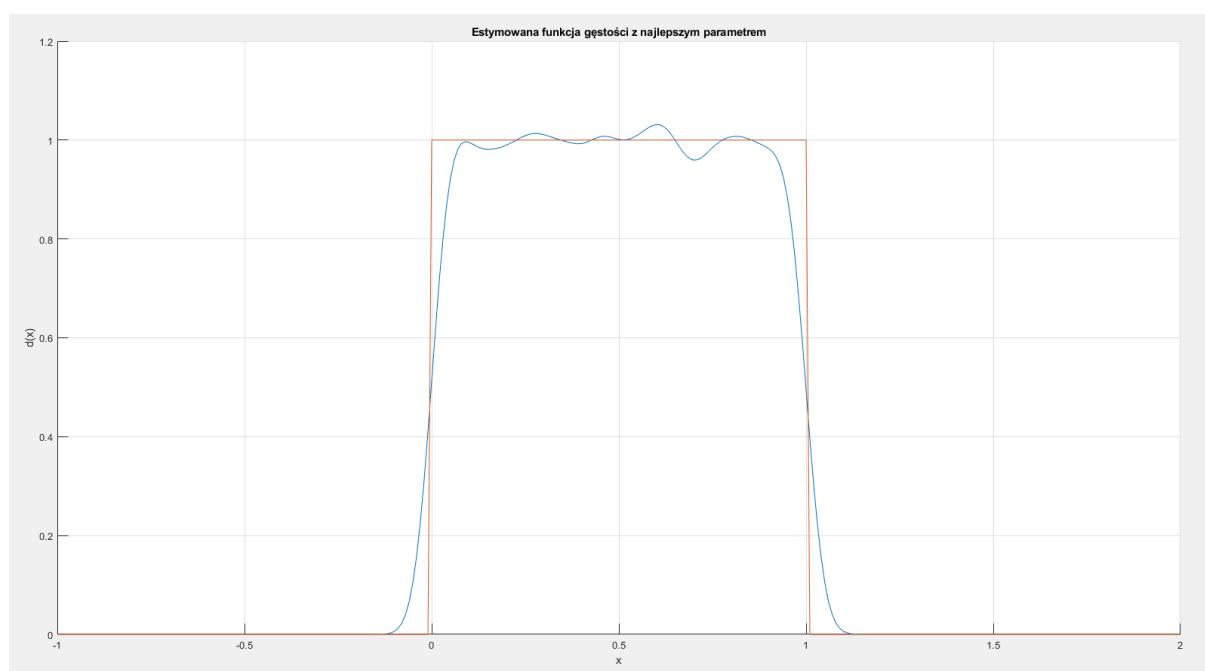
Rysunek 30: Estymata funkcji gęstości dla  $h_N = 0.18$

**Komentarz** Dla powyższej symulacji najlepszym parametrem okazał się  $h_N = 0.18$ .

Trzeci badany przypadek opiera się na funkcji gaussowskiej jądra i rozkładzie jednostajnym na odcinku  $(0,1)$ . Za każdym razem generowano 10 000 próbek.



Rysunek 31: Błąd empiryczny w zależności od parametru  $h_N$



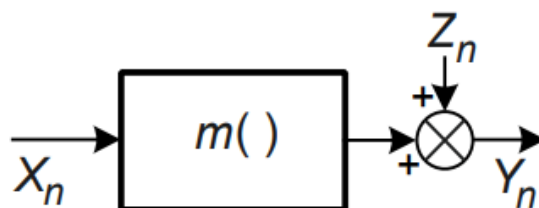
Rysunek 32: Estymata funkcji gęstości dla  $h_N = 0.04$

**Komentarz** Dla powyższej symulacji najlepszym parametrem okazał się  $h_N = 0.04$ .

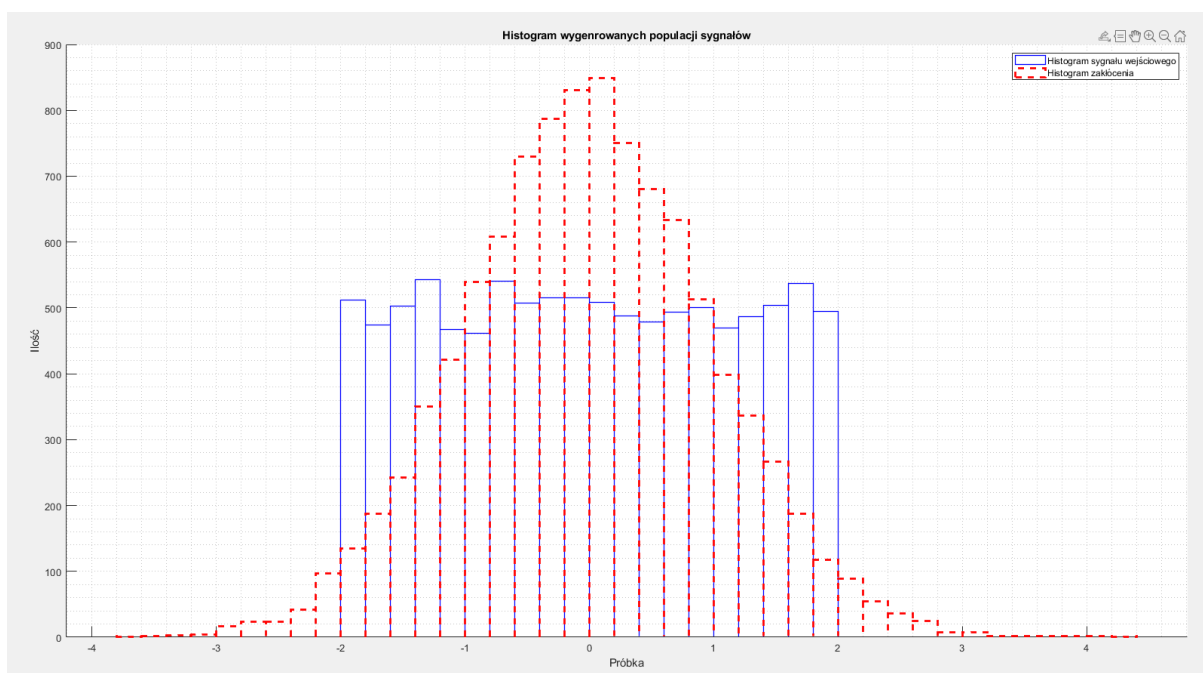
## 4 Laboratorium 6 - Nieparametryczna identyfikacja statycznych systemów nieliniowych. Jądrowy estymator funkcji regresji

### 4.1 Zadanie 1 - Statyczny system nieliniowy z addytywnym zakłóceniem na wyjściu

Na potrzeby dalszych rozważań wygenerowano 10 000 liczb z rozkładu  $\mathbb{U}(-2, 2)$  składających się na sygnał wejściowy  $X$ . Ponadto wygenerowano 10 000 liczb z rozkładu  $\mathcal{N}(0, 1)$  (biały szum), składających się na sygnał zakłócający wyjście (pomiar wyjścia obiektu).



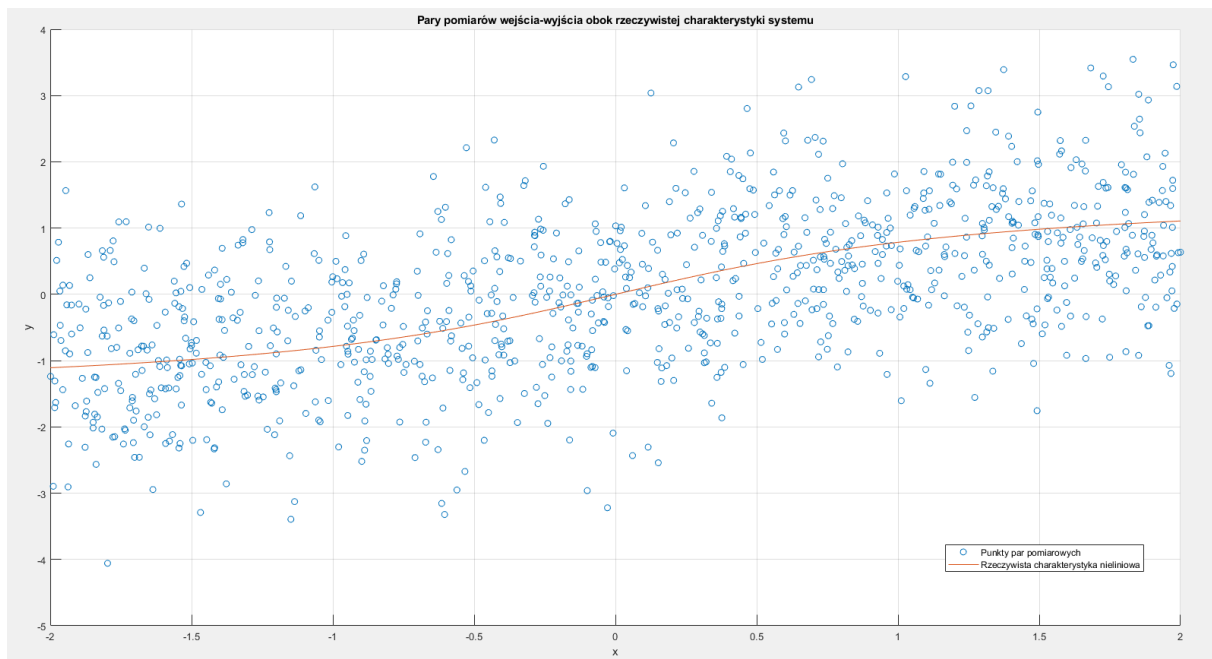
Rysunek 33: Statyczny system nieliniowy z addytywnym zakłóceniem na wyjściu



Rysunek 34: Histogramy wygenerowanych sygnałów

### 4.2 Zadanie 2 - Sygnał wyjściowy systemu a charakterystyka nieliniowa

Dla wygenerowanych sygnałów wyznaczono wyjście systemu (sekwencja pomiarów wyjścia w zależności od sygnału wejściowego i zakłócającego). Pary liczb  $\{(X_i, Y_i)\}$  wykreślono na wykresie obok wzorcowej charakterystyki nieliniowej systemu, co przedstawia rys. (35).



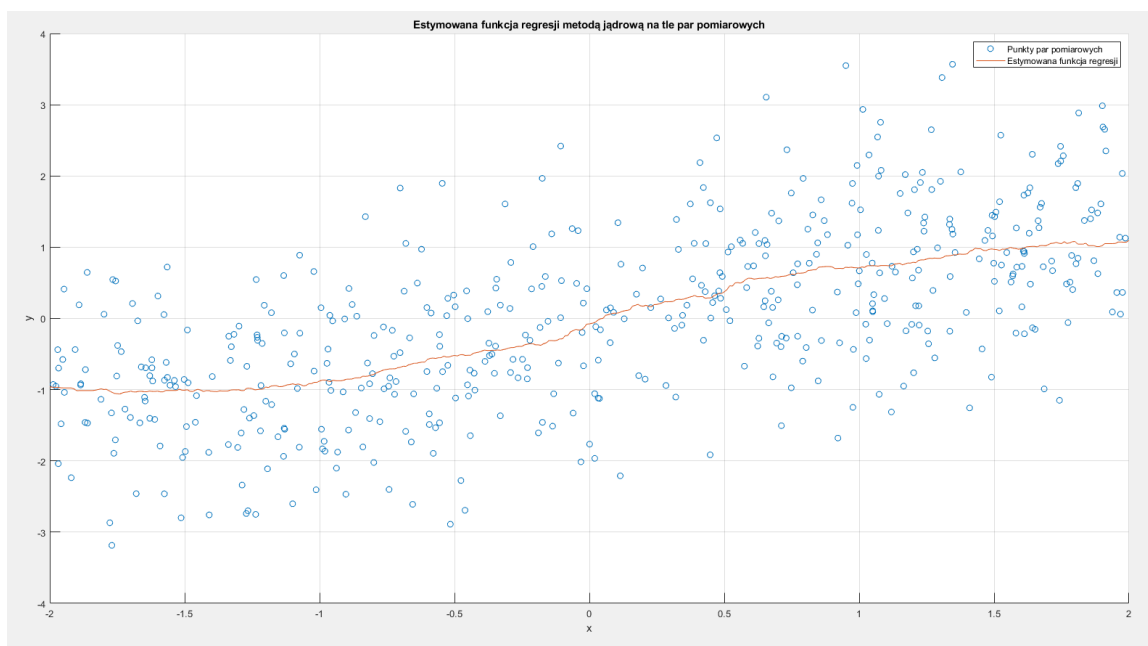
Rysunek 35: Histogramy wygenerowanych sygnałów

### 4.3 Zadanie 3 - Jądrowy estymator funkcji regresji

Zaimplementowano jądrowy estymator funkcji regresji według wzoru [1]

$$\hat{m}_N(x) = \frac{\sum_{n=1}^N Y_n K\left(\frac{X_n - x}{h_N}\right)}{\sum_{n=1}^N K\left(\frac{X_n - x}{h_N}\right)}, \quad (16)$$

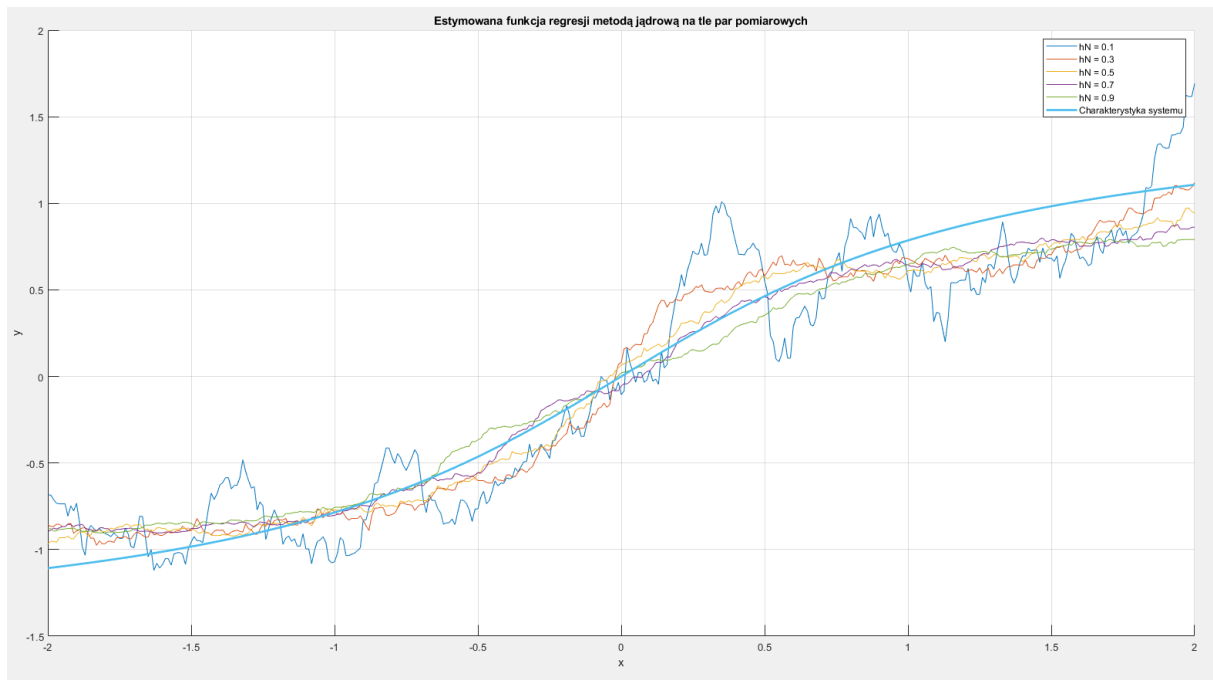
gdzie  $K(\cdot)$  jest funkcją jądra Boxcar, Gaussian, Epanechnikov, Tricube, a  $h_N$  parametrem wygładzania dobieranym symulacyjnie. Dla  $N = 500$  estymowano funkcję regresji - jądro Boxcar prostokątne (rys. (36)).



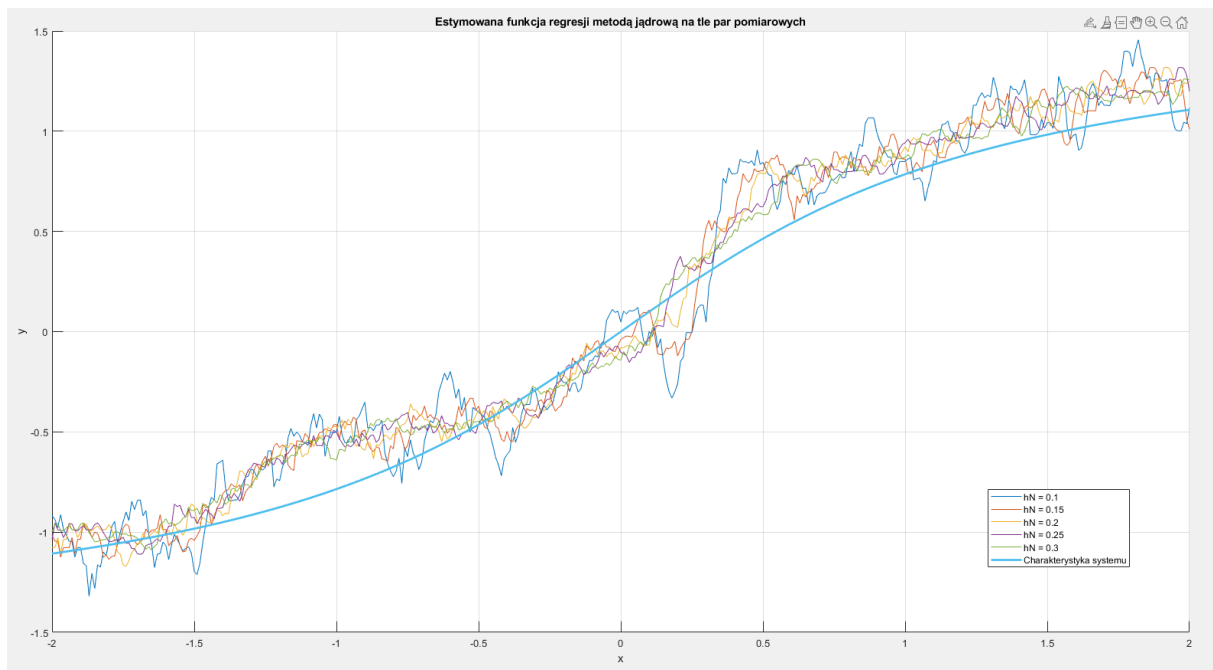
Rysunek 36: Estymowana funkcja regresji na tle punktów pomiarowych  $h_N = 0.5$



Następnie wykreślono kilka estymacji funkcji regresji w zależności od parametru  $h_N$ .



Rysunek 37: Estymowana funkcja regresji w zależności od  $h_N \in (0, 1)$

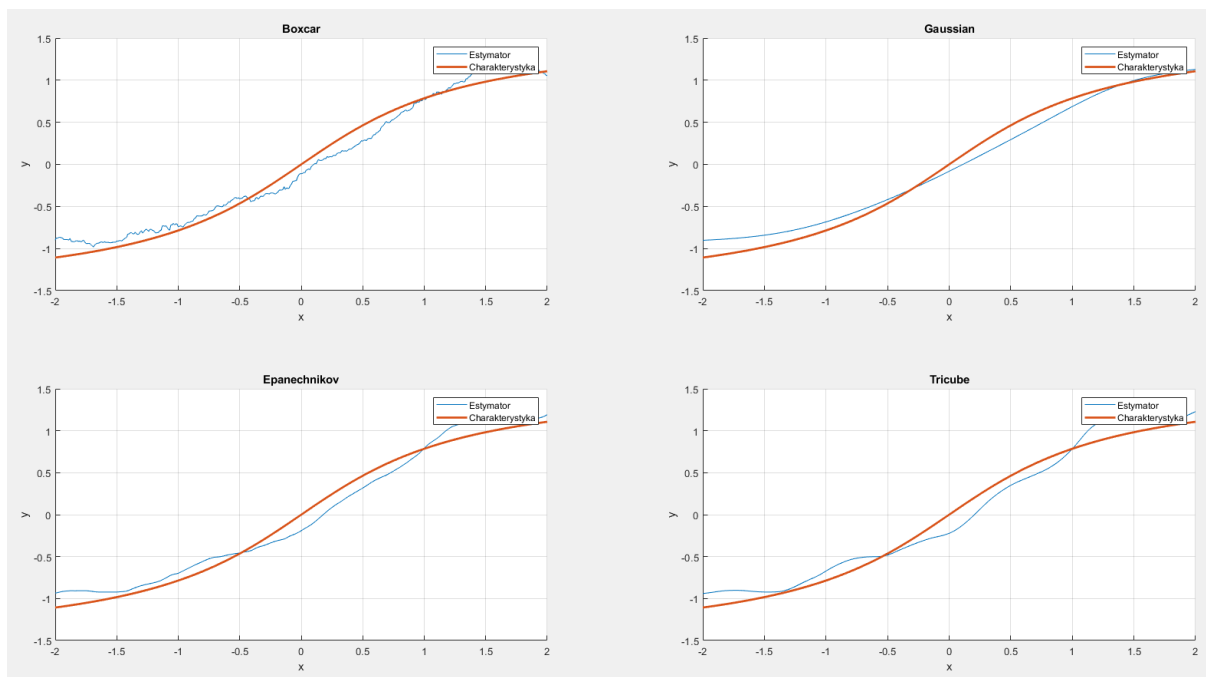


Rysunek 38: Estymowana funkcja regresji w zależności od  $h_N \in [0.1, 0.3]$

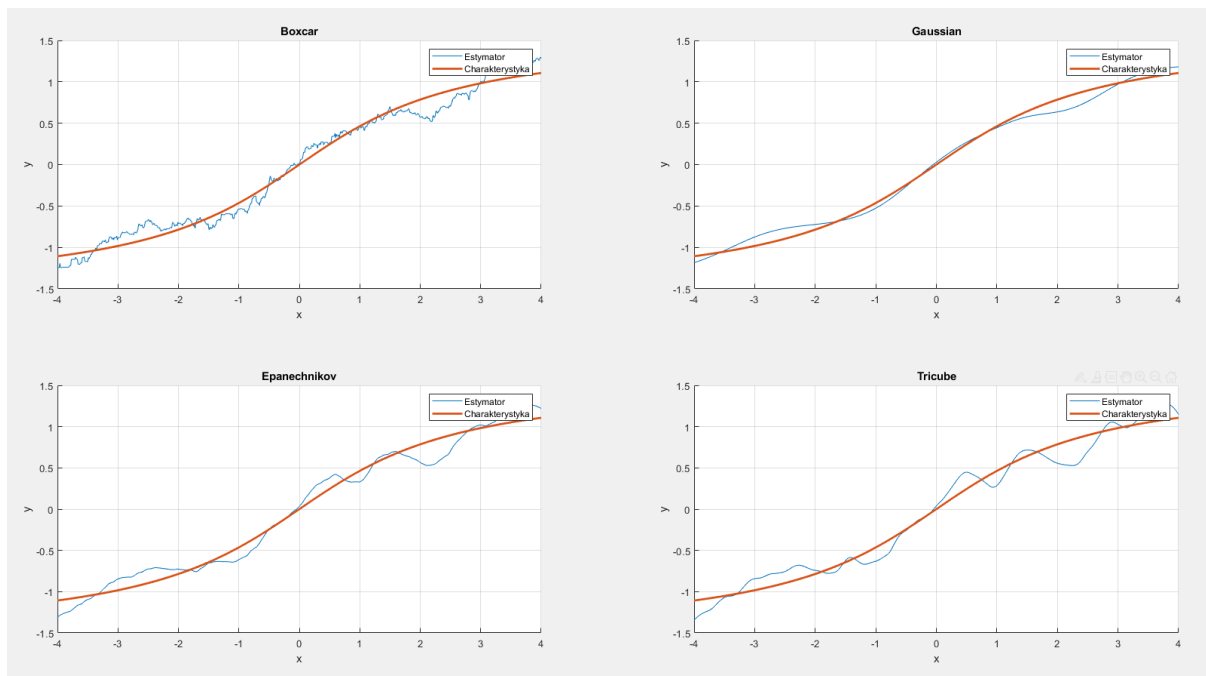
**Komentarz** Dla zbyt małych parametrów  $h_N$  zgodnie z rys. (37) i (38) estymowana funkcja regresji przyjmuje zbyt duże oscylacje. Dla zbyt wysokich  $h_N$  estymator dąży do funkcji liniowej, zwiększając błąd całkowity. Jądro prostokątne Boxcar nie jest najlepszym narzędziem do estymacji niniejszej charakterystyki - zbyt duże szybkozmienne oscylacje.

#### 4.4 Zadanie 4 - Porównanie typów jąder estymacji funkcji regresji

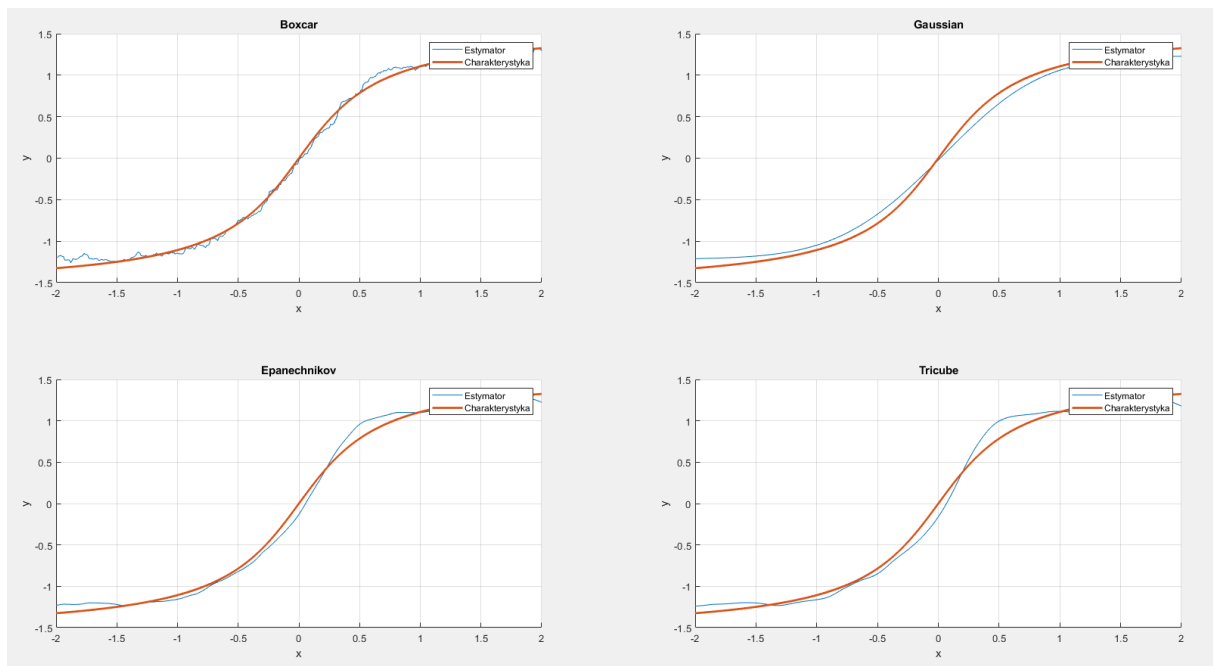
Ustalono parametr wygładzania jako  $h_N = 0.5$ . Zbadano estymator dla czterech typów funkcji jądra: Boxcar, Gaussian, Epanechnikov, Tricube.



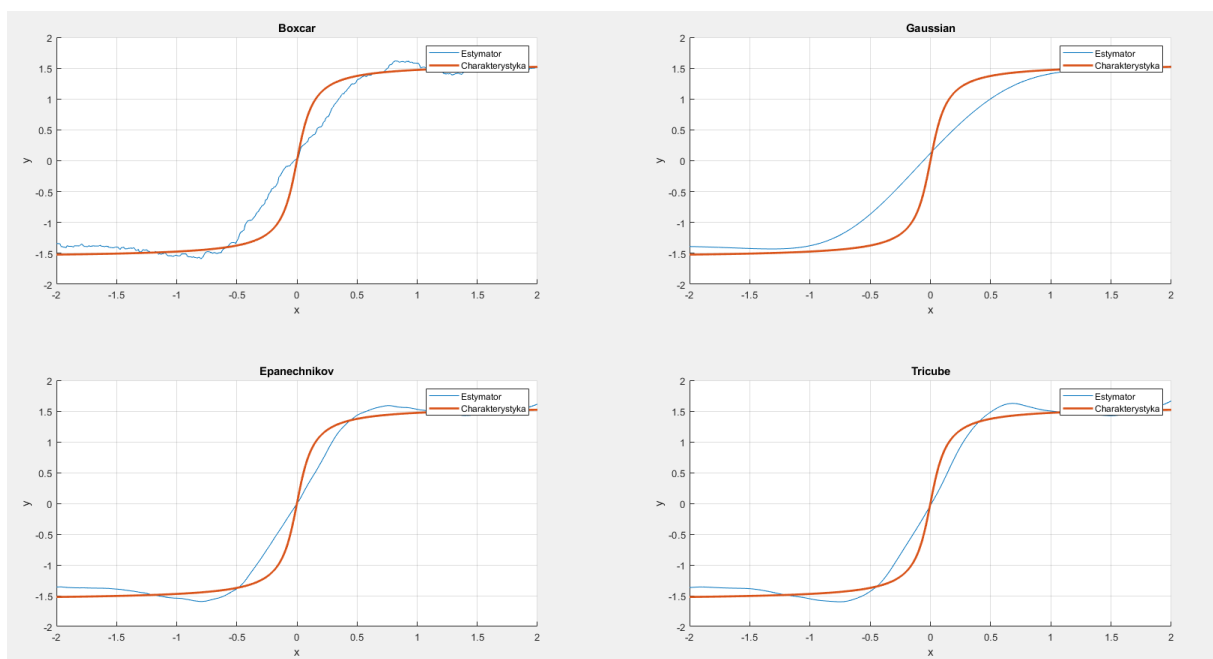
Rysunek 39: Estymowana funkcja regresji w zależności od typu jądra  $h_N = 0.5, a = 1$



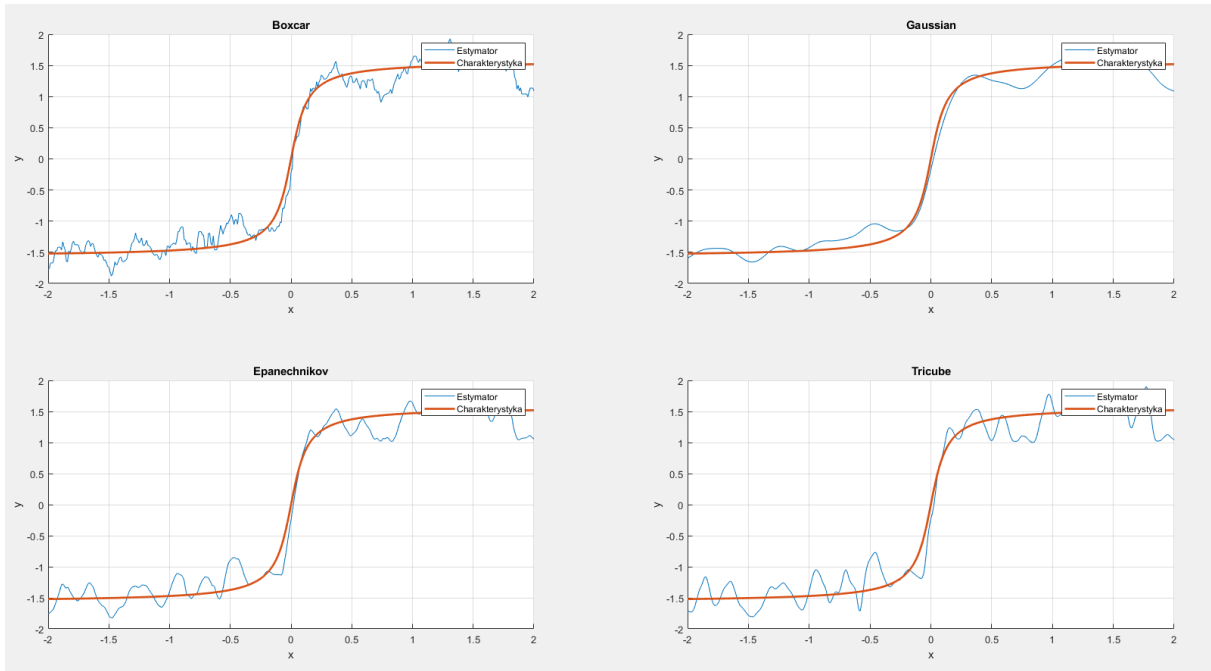
Rysunek 40: Estymowana funkcja regresji w zależności od typu jądra  $h_N = 0.5, a = 0.5$



Rysunek 41: Estymowana funkcja regresji w zależności od typu jądra  $h_N = 0.5, a = 2$



Rysunek 42: Estymowana funkcja regresji w zależności od typu jądra  $h_N = 0.5, a = 10$



Rysunek 43: Estymowana funkcja regresji w zależności od typu jądra  $h_N = 0.1, a = 10$

**Komentarz** Na podstawie rys. (39), (40), (41), (42), (43) wyciągnięto następujące wnioski:

- parametr  $a$  charakterystyki wpływa na nachylenie funkcji w okolicach zera - przy znacznym zwiększeniu parametru  $a$  i stałym parametrze  $h_N$  zaobserwowano trudności w estymacji funkcji regresji,
- znaczne zwiększenie parametru  $a$  powodowało potrzebę zmniejszenia parametru wygładzania - większe zmiany na wykresie spowodowały konieczność wprowadzenia większej wrażliwości estymatora, aby poprawnie oszacować charakterystykę w okolicach punktu zero kosztem zwiększenia oscylacji na mniej nachylonych częściach wykresu.
- jądro Boxcar wprowadzało największe szybkozmienne oscylacje,
- najmniejsze oscylacje wprowadzało jądro gaussowskie dla wszystkich wariantów,
- dla standardowej sytuacji  $a = 1, h_N = 0.5$  najlepszym rozwiązaniem było wprowadzenie jądra gaussowskiego,
- niemożliwe jest stwierdzenie, które jądro jest najbardziej optymalnym wyborem przy estymacji danej charakterystyki - potrzeba wyznaczenia optymalnego parametru  $h_N$  dla każdego jądra.

Rozwiązaniem problemu zwiększenia poprawności estymacji na niektórych odcinkach kosztem wprowadzenia oscylacji na innych jest zmienny zależny od nachylenia wykresu w punkcie parametr wygładzania.

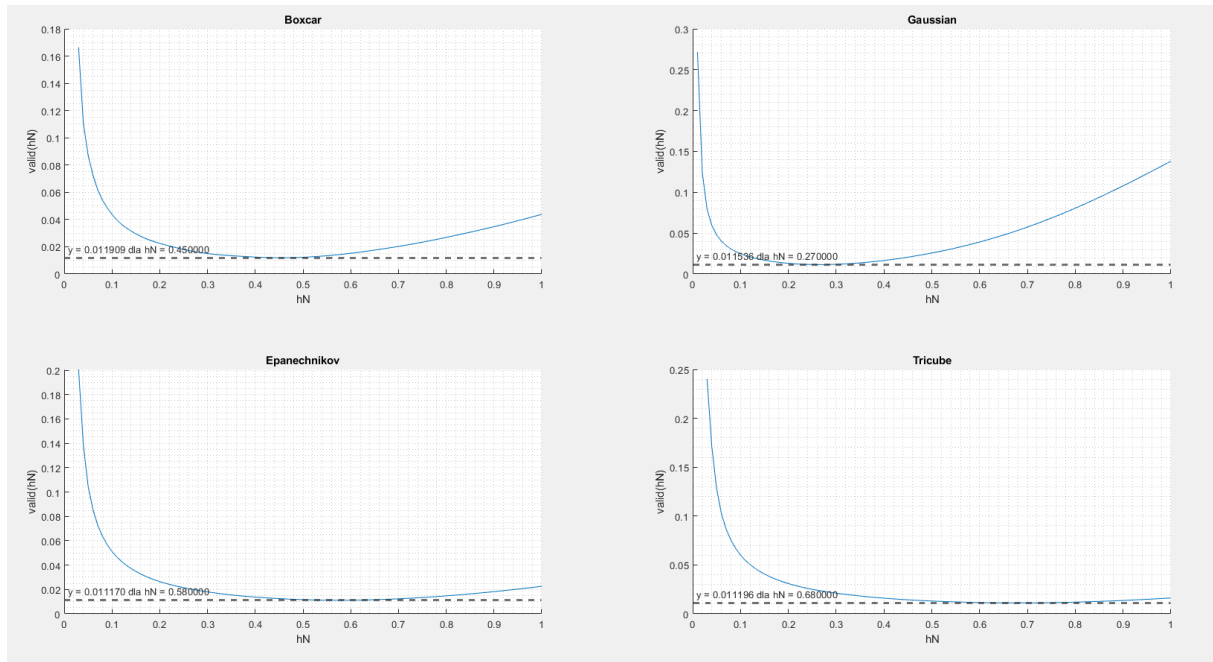
**Uwaga** Stwierdzenie, które jądro jest najbardziej optymalnym narzędziem do estymacji badanej charakterystyki, jest możliwe po wyznaczeniu optymalnych parametrów  $h_N$ , co zrobiono w kolejnym zadaniu.

#### 4.5 Zadanie 5 - Optymalny parametr wygładzania i jądro estymacji

W celu wyznaczenia optymalnej nastawy parametru wygładzania minimalizowano odpowiednik błędu całkowitego jako wskaźnika jakości na odcinku  $(-1,1)$

$$valid(h) = \frac{1}{2Q} \sum_{q=-Q}^{q=Q} \left[ \hat{m}_N \left( \frac{q}{Q} \right) - m \left( \frac{q}{Q} \right) \right]^2. \quad (17)$$

$N = 500$ , a parametr  $a = 2$ . Dodatkowo w celu osądzenia, które jądro estymatora okazało się najlepszym wyborem, wygenerowano 10 populacji par pomiarowych, a wyniki wskaźnika jakości uśredniono (Wprowadzono parametr  $L = 10$ ).



Rysunek 44: Zadanie optymalizacji parametru wygładzania dla różnych jąder estymacji

Jądro	$h_N$	$\text{valid}(h)$
Boxcar	0.45	0.01191
Gaussian	0.27	0.01154
<b>Epanechnikov</b>	<b>0.58</b>	<b>0.01117</b>
Tricube	0.68	0.01120

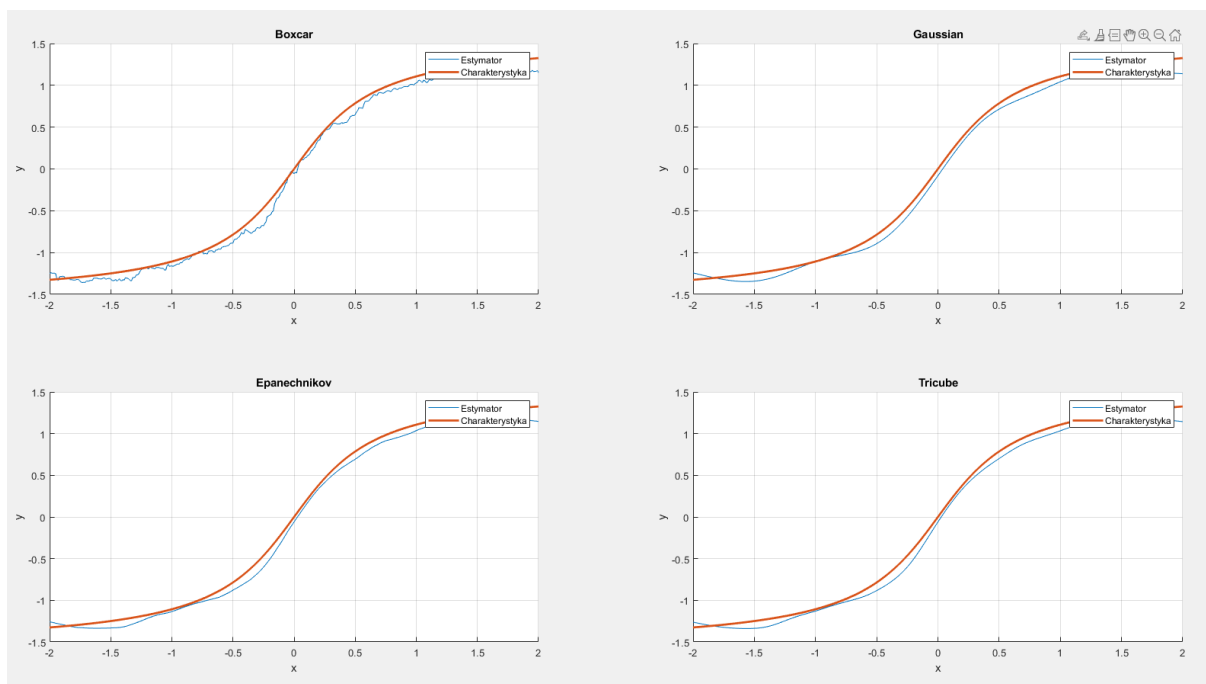
Tabela 2: Wyniki zadania optymalizacji - minimalizacja błędu całkowego

**Komentarz** Różnice w minimalnych wartościach wskaźnika jakości są zbyt małe przy założonej dokładności badania parametru wygładzania równej 0.01, aby stwierdzić, która estymacja poradziła sobie najlepiej. Wyniki potwierdzają, że podejrzanym o bycie najlepszym rozwiązaniem dla badanego przypadku jest estymacja za pomocą jądra Epanechnikov. Teza ta jednak nie może być jednoznacznie potwierdzona tym eksperymentem.

Znaleziono natomiast jednoznacznie najlepsze parametry wygładzania dla zadanych jąder estymacji systemu nieliniowego na odcinku  $(-1,1)$ .

#### 4.6 Zadanie 6 - Najlepsze estymacje systemu nieliniowego - ocena

Dla wymienionych w zadaniu 5 parametrów wygładzania wykreślono estymacje zadanego systemu nieliniowego.



Rysunek 45: Zadanie optymalizacji parametru wygładzania dla różnych jąder estymacji

**Komentarz** Dla jąder gaussowskiego, Epanechnikov i Tricube estymatory wykazały się dużym podobieństwem oszacowania systemu nieliniowego. Odmienny okazał się estymator bazujący na jądrze prostokątnym (największe szybkozmienne oscylacje).

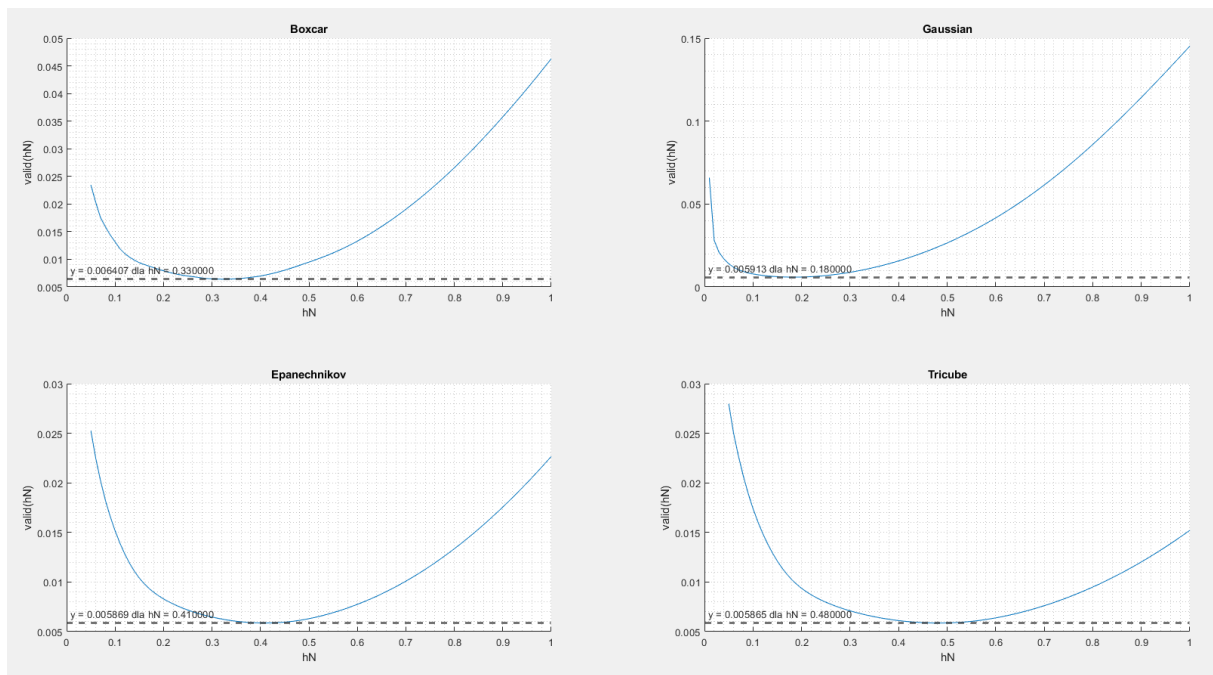
#### 4.7 Zadanie 7 - Estymator jądrowy regresji a zakłócenie Cauchy'ego

Zbadano estymację tego samego systemu, gdzie  $a = 2$  na podstawie 10 populacji par pomiarowych, liczących po 500 osobników. Minimalizacja wskaźnika jakości również odbyła się na odcinku  $(-1,1)$ , a zakres badanych parametrów wygładzania to  $(0,1]$  ze skokiem (dokładnością) 0.01. Zmianie uległa tylko natura zakłócenia, gdyż pochodzi ono z rozkładu Cauchy'ego  $(0,0.01)$ .

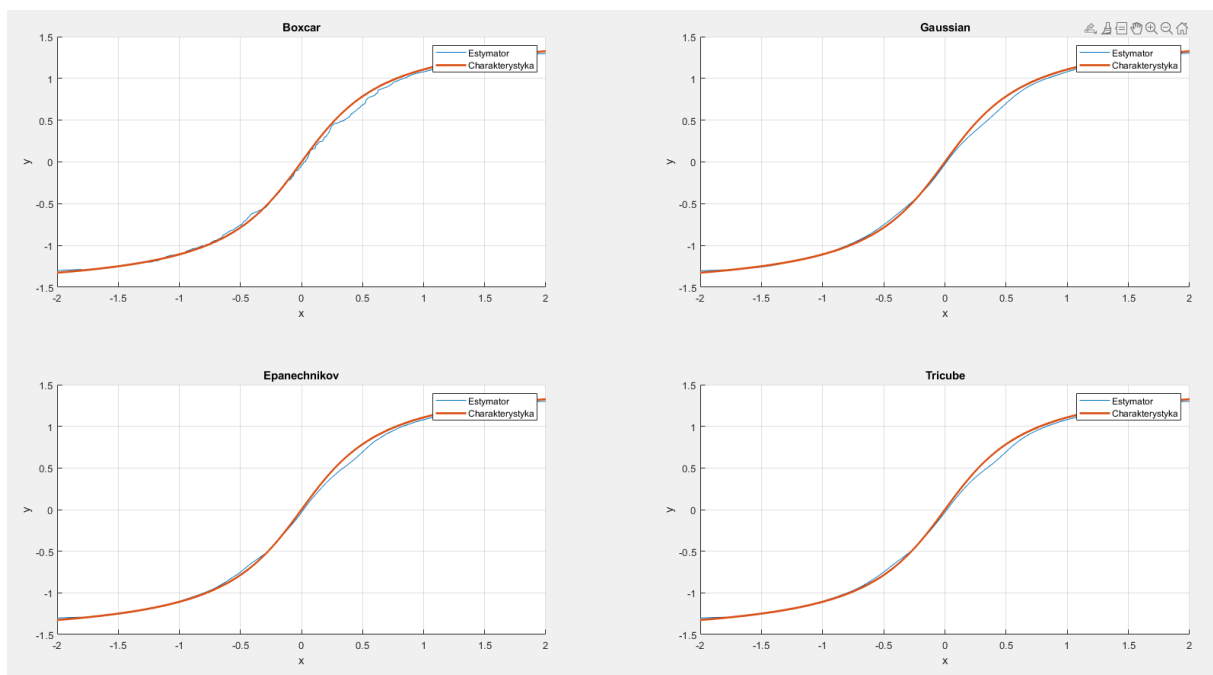
Jądro	$h_N$	valid(h)
Boxcar	0.33	0.00641
Gaussian	0.18	0.00591
Epanechnikov	0.41	0.00587
<b>Tricube</b>	<b>0.48</b>	<b>0.00586</b>

Tabela 3: Wyniki zadania optymalizacji - minimalizacja błędu całkowego

Szum o charakterze rozkładu Cauchy'ego  $(0,0.01)$  jest nieznaczny, co pozwalało z dużą dokładnością estymować charakterystykę obiektu (rys.(46),(47)).



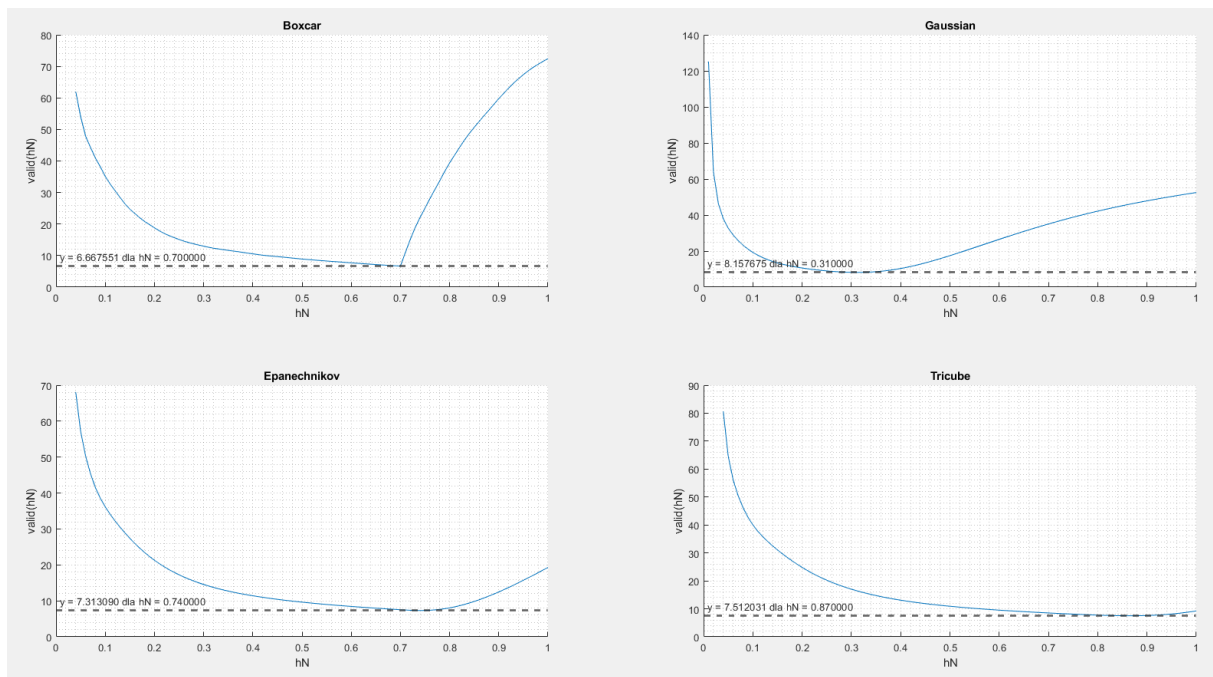
Rysunek 46: Zadanie optymalizacji parametru wygładzania dla różnych jąder estymacji



Rysunek 47: Zadanie optymalizacji parametru wygładzania dla różnych jąder estymacji

Szum Cauchy'ego o małym parametrze  $\lambda = 0.01$  miał mniej destrukcyjny wpływ na estymację charakterystyki niż biały szum.

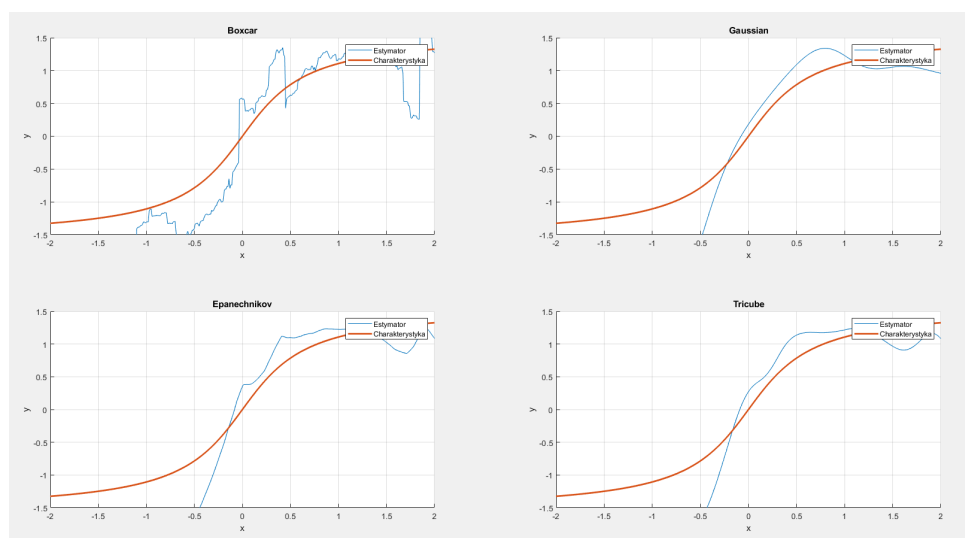
Przeprowadzono próbę estymacji charakterystyki przy obecności szumu z rozkładu Cauchy'ego (0,1).



Rysunek 48: Zadanie optymalizacji parametru wygładzania dla różnych jąder estymacji

Jądro	$h_N$	valid(h)
<b>Boxcar</b>	<b>0.70</b>	<b>6.68</b>
Gaussian	0.31	8.16
Epanechnikov	0.74	7.31
Tricube	0.87	7.51

Tabela 4: Wyniki zadania optymalizacji - minimalizacja błędu całkowego



Rysunek 49: Zadanie optymalizacji parametru wygładzania dla różnych jąder estymacji

Próba estymacji charakterystyki przy szumie Cauchy'ego (0,1) zakończyła się niepowodzeniem nawet dla najbardziej optymalnych parametrów wygładzania.

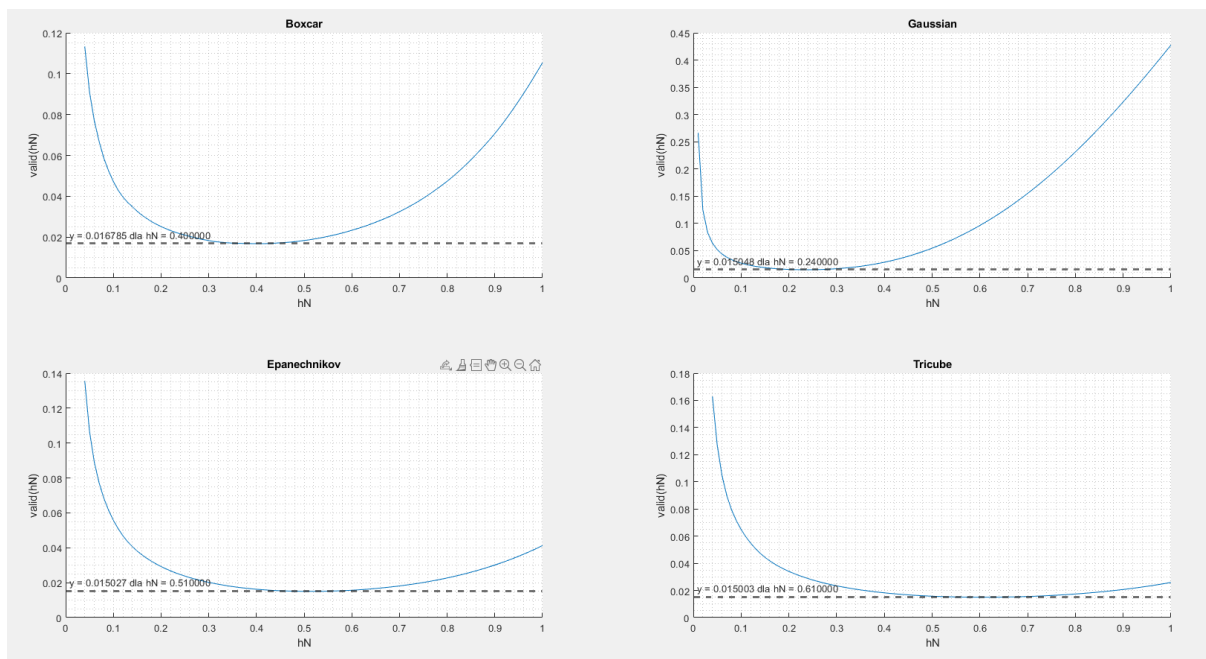


## 4.8 Zadanie 1\* - Estymacja regresji metodą jądrową a niezerowa wartość oczekiwana zakłócenia

Zbadano możliwość estymacji systemu o charakterystyce kwadratowej

$$m(x) = (x - 1)^2. \quad (18)$$

Początkowe zakłócenie wyjścia pochodzi z rozkładu  $\mathcal{N}(0, 1)$ . Zadanie optymalizacji odbyło się na odcinku  $(-1, 1)$ .

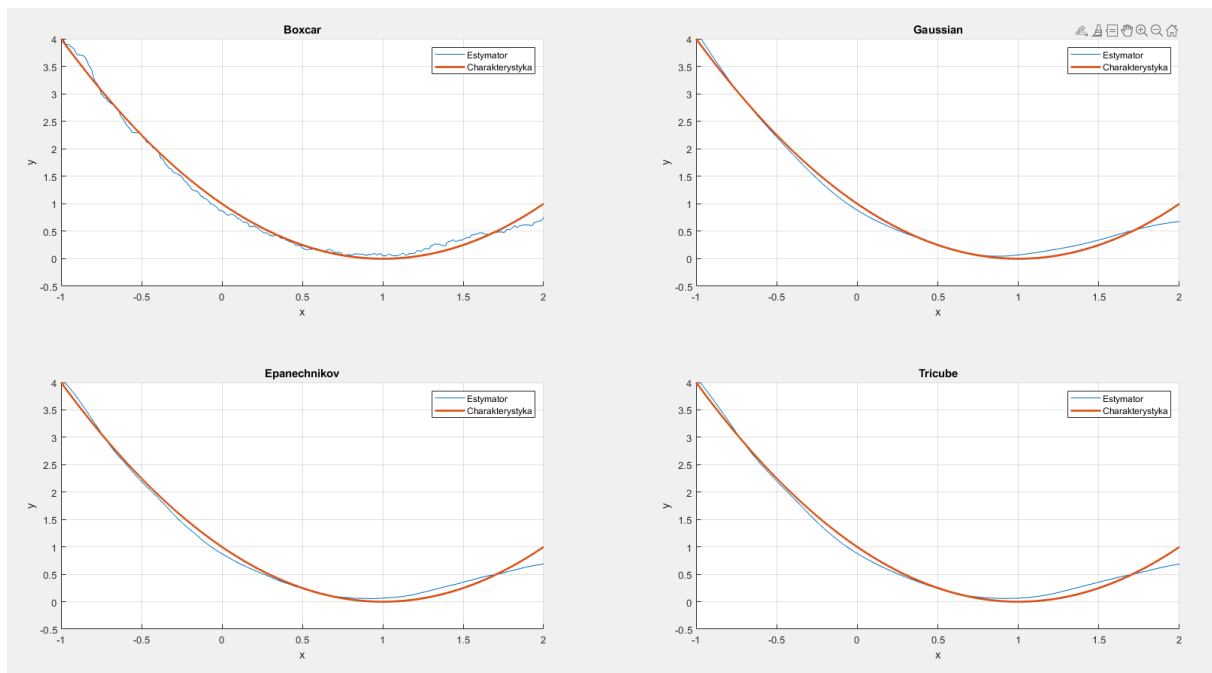


Rysunek 50: Zadanie optymalizacji parametru wygładzania dla różnych jąder estymacji

Jądro	$h_N$	valid(h)
Boxcar	0.40	0.016785
Gaussian	0.24	0.015048
Epanechnikov	0.51	0.015027
<b>Tricube</b>	<b>0.61</b>	<b>0.015003</b>

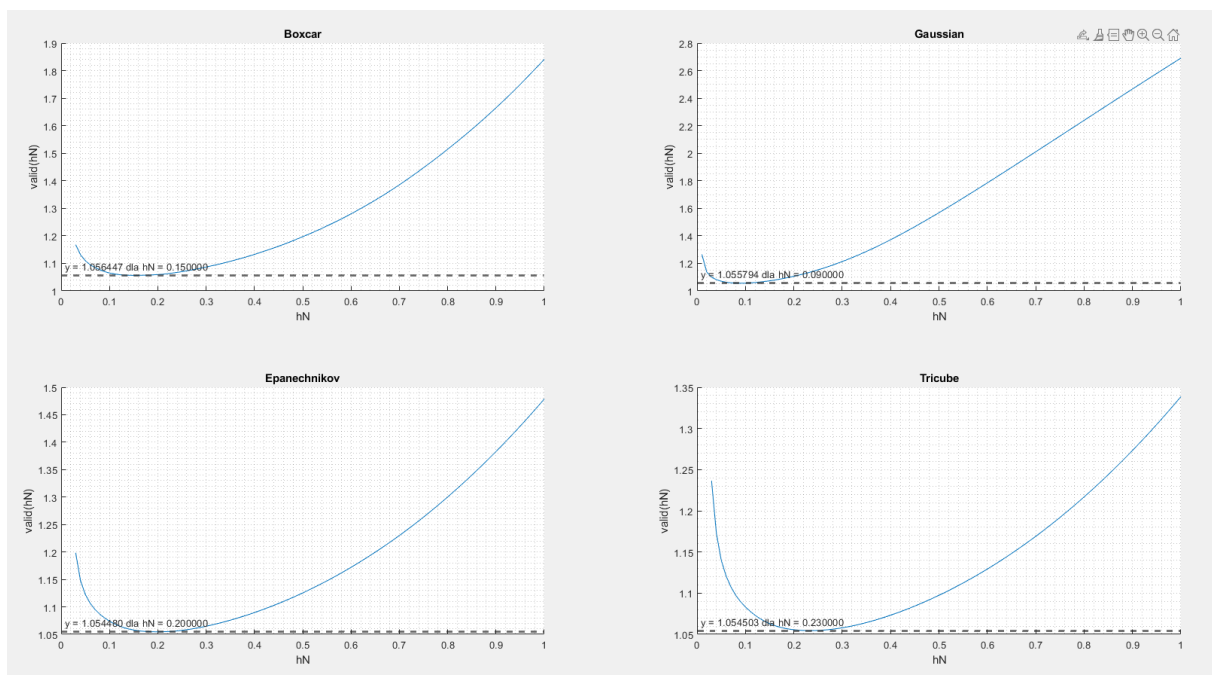
Tabela 5: Wyniki zadania optymalizacji - minimalizacja błędu całkowego

**Komentarz** Uzyskane minimalne błędy pozwalają z taką dokładnością estymować charakterystkę obiektu na odcinku  $(-1, 1)$ . Została ona przedstawiona na rys. (51).



Rysunek 51: Zadanie optymalizacji parametru wygładzania dla różnych jąder estymacji

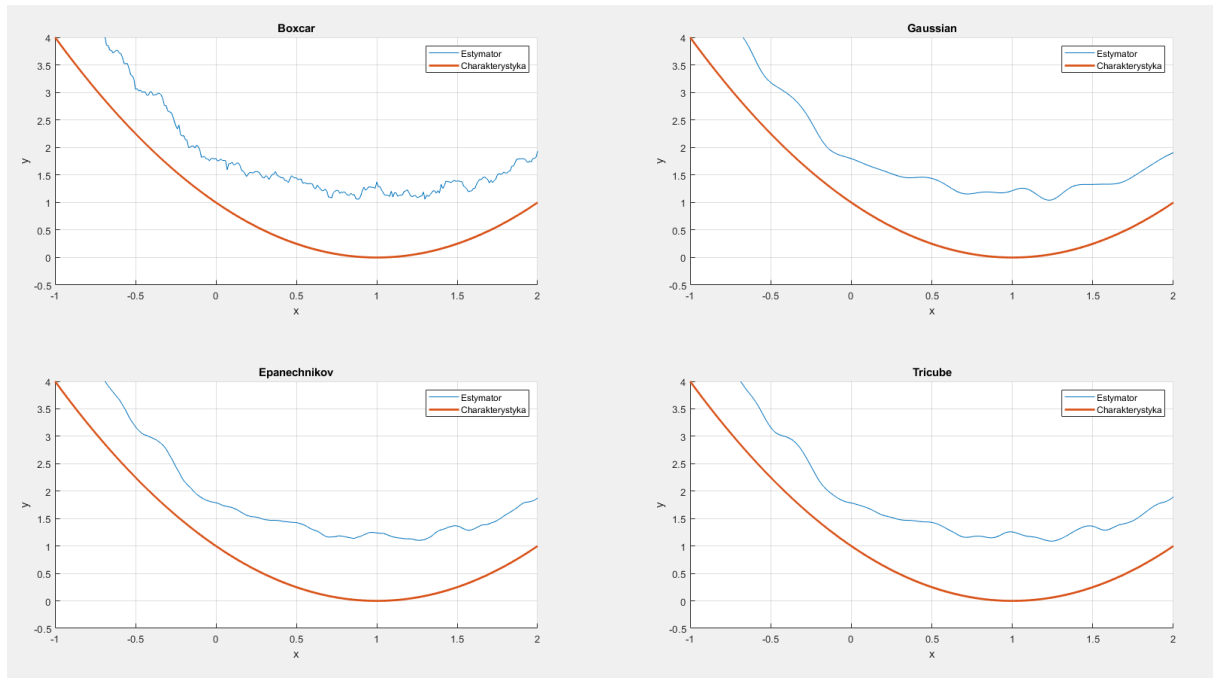
Zbadano możliwość estymacji systemu o charakterystyce kwadratowej jednak z obecnym zakłóceniem na wyjściu o niezerowej wartości oczekiwanej. Zakłócenie pochodziło z rozkładu  $N(1, 1)$ .



Rysunek 52: Zadanie optymalizacji parametru wygładzania dla różnych jąder estymacji

Jądro	$h_N$	valid(h)
Boxcar	0.15	1.05645
Gaussian	0.09	1.05579
<b>Epanechnikov</b>	<b>0.2</b>	<b>1.05448</b>
Tricube	0.23	1.05450

Tabela 6: Wyniki zadania optymalizacji - minimalizacja błędu całkowego



Rysunek 53: Zadanie optymalizacji parametru wygładzania dla różnych jąder estymacji

**Komentarz** Zakłócenie w systemie w każdej symulacji miało addytywny charakter. Gdy wartość oczekiwana addytywnego zakłócenia jest niezerowa (w tym przypadku równa 1), estymacja charakterystyki jest przesunięta wzdłuż osi y o wartość oczekiwaną zakłócenia (estymacja jest obciążona). Zakłócenia o niezerowej wartości oczekiwanej wprowadza obciążenie estymacji charakterystyki systemu.

## Literatura

- [1] Kałuszk Marek Gajek Lesław. *Wnioskowanie statystyczne: modele i metody*. Wydawnictwo Naukowo - Techniczne, 1993.
- [2] Sztencel Rafał Jakubowski Jacek. *Wstęp do teorii prawdopodobieństwa*. Script, 2001.
- [3] Edmund Pluciński Plucińska, Agnieszka. *Probabilistyka: rachunek prawdopodobieństwa, statystyka matematyczna, procesy stochastyczne*. Wydawnictwo Naukowo-Techniczne, 2000.