



Politechnika  
Wrocławska



---

MODELOWANIE I IDENTYFIKACJA  
SPRAWOZDANIE 2

---

*Autor:*

GRZEGORZ SĘK

*Prowadzący zajęcia:*

DR HAB. INŻ. PAWEŁ WACHEL

*Termin zajęć:*

Poniedziałek, 15:15

Semestr letni 2020/2021

1 lutego 2023

# Spis treści

<b>1</b>	<b>Laboratorium 3 – estymatory średniej oraz wariancji</b>	<b>2</b>
1.1	Zadanie 1 – implementacja estymatora średniej, obciążonego i nieobciążonego estymatora wariancji dla rozkładu normalnego $\mathcal{N}(\mu, \sigma^2)$	2
1.2	Zadanie 2 – implementacja i wykreślenie błędu empirycznego	2
1.3	Zadanie 3 – powtórzenie wcześniej przedstawionych eksperymentów dla rozkładu Cauchy’ego	4
1.4	Zadanie dodatkowe 1 – implementacja estymatora kowariancji i korelacji	6
1.5	Zadanie dodatkowe 2 – implementacja estymatora kowariancji i korelacji	7
<b>2</b>	<b>Laboratorium 4 – estymator dystrybuanty empirycznej</b>	<b>8</b>
2.1	Zadanie 2 – implementacja i wykreślenie dystrybuanty empirycznej dla ciągu liczb losowych z rozkładu:	8
2.2	Zadanie 3 – implementacja empirycznego odpowiednika wyrażenia:	9
2.3	Zadanie 4 – Znalezienie rozkładu dla podanych pomiarów	10
2.4	Zadanie 5 – Konstrukcja eksperymentu symulacyjnego, umożliwiającego estymację wariancji dystrybuanty empirycznej:	10
2.5	Zadanie dodatkowe 1 – Wyznaczanie błędu empirycznego:	11
<b>3</b>	<b>Laboratorium 5 – jądrowy estymator gęstości prawdopodobieństwa</b>	<b>13</b>
3.1	Zadanie 2 – implementacja estymatora jądrowego gęstości prawdopodobieństwa	13
3.2	Zadanie 3 – porównanie działania różnych funkcji jądra	14
3.2.1	Rozkład normalny	14
3.2.2	Rozkład trójkątny	15
3.3	Zadanie 4 – wyznaczenie błędu empirycznego gęstości prawdopodobieństwa	15
<b>4</b>	<b>Laboratorium 6 – jądrowy estymator funkcji regresji</b>	<b>17</b>
4.1	Zadanie 2 – Wykreślenie nieliniowej charakterystyki systemu	17
4.1.1	Zakłócenie z rozkładu normalnego	17
4.1.2	Zakłócenie z rozkładu Cauchy’ego	18
4.2	Zadanie 3 – implementacja jądrowego estymatora funkcji regresji:	18
4.2.1	Zakłócenie z rozkładu normalnego	18
4.2.2	Zakłócenie z rozkładu Cauchy’ego	19
4.3	Zadanie 4 – estymacja funkcji regresji dla różnych funkcji jądra	20
4.3.1	Zakłócenie z rozkładu normalnego	20
4.3.2	Zakłócenie z rozkładu Cauchy’ego	21
4.4	Zadanie 5 – minimalizacja błędu	22
4.4.1	Zakłócenie z rozkładu normalnego	22
4.4.2	Zakłócenie z rozkładu Cauchy’ego	22
4.5	Zadanie 6 – rysowanie wykresów z zadania 3 oraz 4 dla parametru wygładzania, dla którego uzyskano najmniejszy błąd	23
4.5.1	Zakłócenie z rozkładu normalnego	23
4.5.2	Zakłócenie z rozkładu Cauchy’ego	24
4.6	Zadanie dodatkowe 1 – błąd empiryczny w funkcji $h$	25
4.7	Zadanie dodatkowe 2 – Błąd empiryczny w funkcji $N$	26

# Rozdział 1

## Laboratorium 3 – estymatory średniej oraz wariancji

### 1.1 Zadanie 1 – implementacja estymatora średniej, obciążonego i nieobciążonego estymatora wariancji dla rozkładu normalnego $\mathcal{N}(\mu, \sigma^2)$

$$\hat{\mu}_N = \frac{1}{N} \sum_{n=1}^N X_N \quad (1.1)$$

$$\hat{s}_N^2 = \frac{1}{N} \sum_{n=1}^N (X_N - \hat{\mu}_N)^2 \quad (1.2)$$

$$\hat{S}_N^2 = \frac{1}{N-1} \sum_{n=1}^N (X_N - \hat{\mu}_N)^2 \quad (1.3)$$

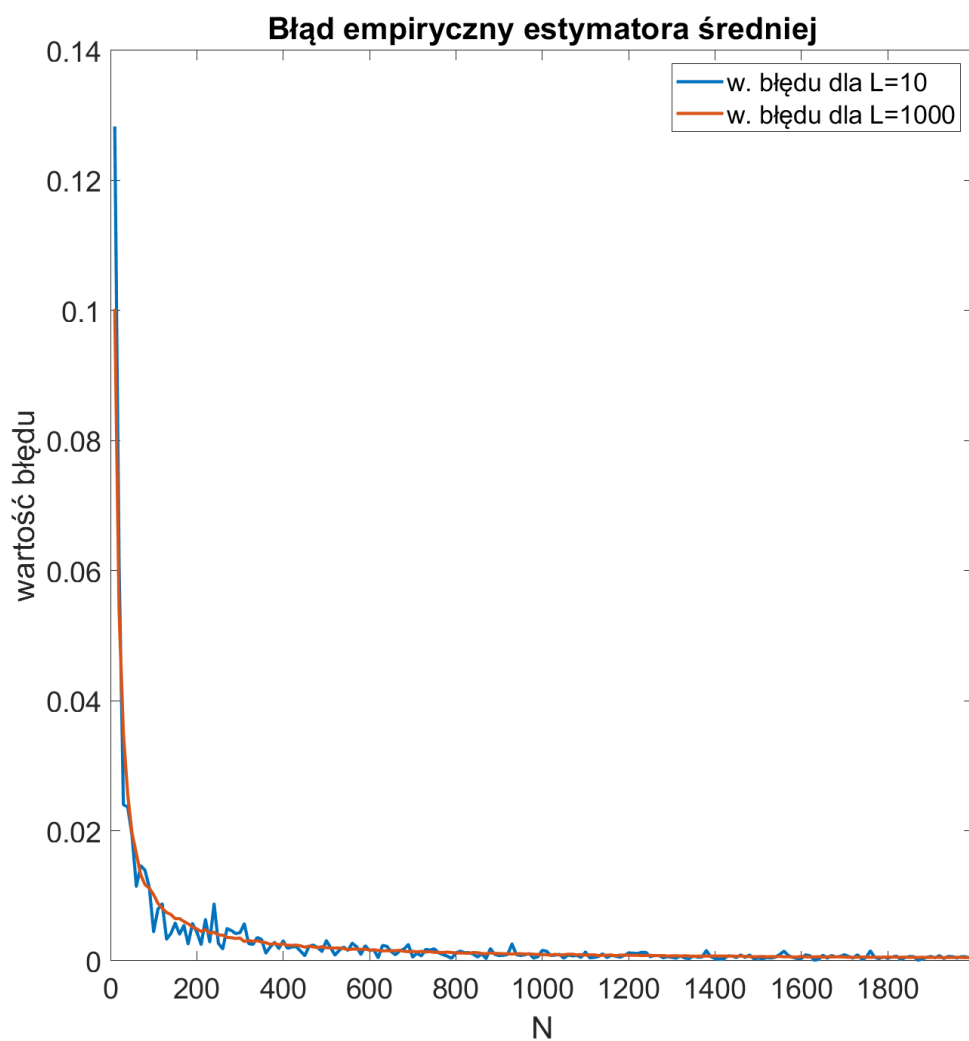
Przedstawione zadanie zostało wykonane dla rozkładu normalnego o parametrach  $\mathcal{N}(0, 1)$  oraz liczbie próbek  $N = 1000$ . Uzyskane wyniki to  $\hat{\mu}_N = -0.0395$ ,  $\hat{s}_N^2 = 0.9911$  oraz  $\hat{S}_N^2 = 0.9921$ , co pokrywa się z parametrami rozkładu normalnego.

### 1.2 Zadanie 2 – implementacja i wykreślenie błędu empirycznego

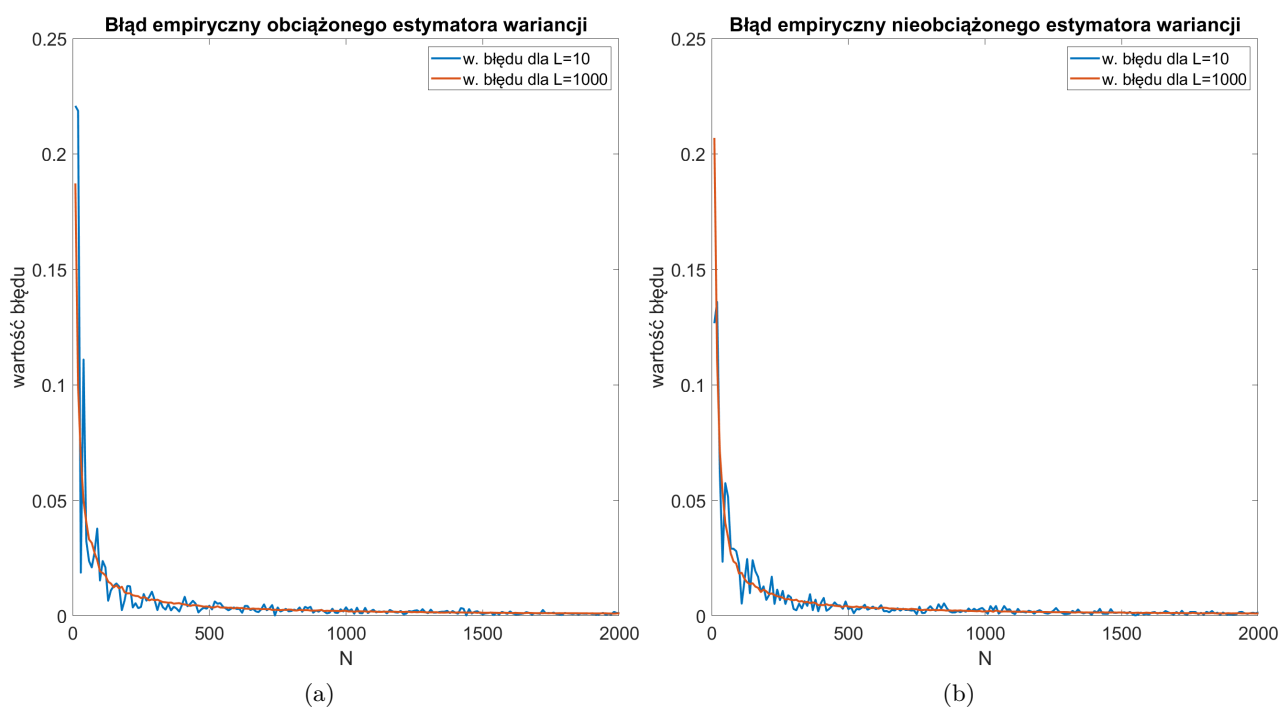
$$Err\{\hat{\mu}_N, \mu\} = \frac{1}{L} \sum_{l=1}^L [\hat{\mu}_N^{[l]} - \hat{\mu}_N]^2 \quad (1.4)$$

Zadanie zostało zrealizowane dla rozkładu normalnego  $\mathcal{N}(0, 1)$ , liczbie eksperymentów  $L = 10$ ,  $L = 1000$  oraz liczbie niezależnych próbek  $N = 2000$ .

Analizując rysunki (Rysunek 1.1 i Rysunek 1.2) można zauważyć, że stała  $L$  pozbywa wykres oscylacji, dzięki czemu wartość błędu dla danej liczby próbek jest dokładniejsza i bliższa wartości rzeczywistej. Można również stwierdzić, że wykorzystane estymatory są bardzo dobrymi narzędziami w estymowaniu średniej i wariancji, ponieważ wartości błędów dla każdego z estymatorów bardzo szybko zbiegają do zera.

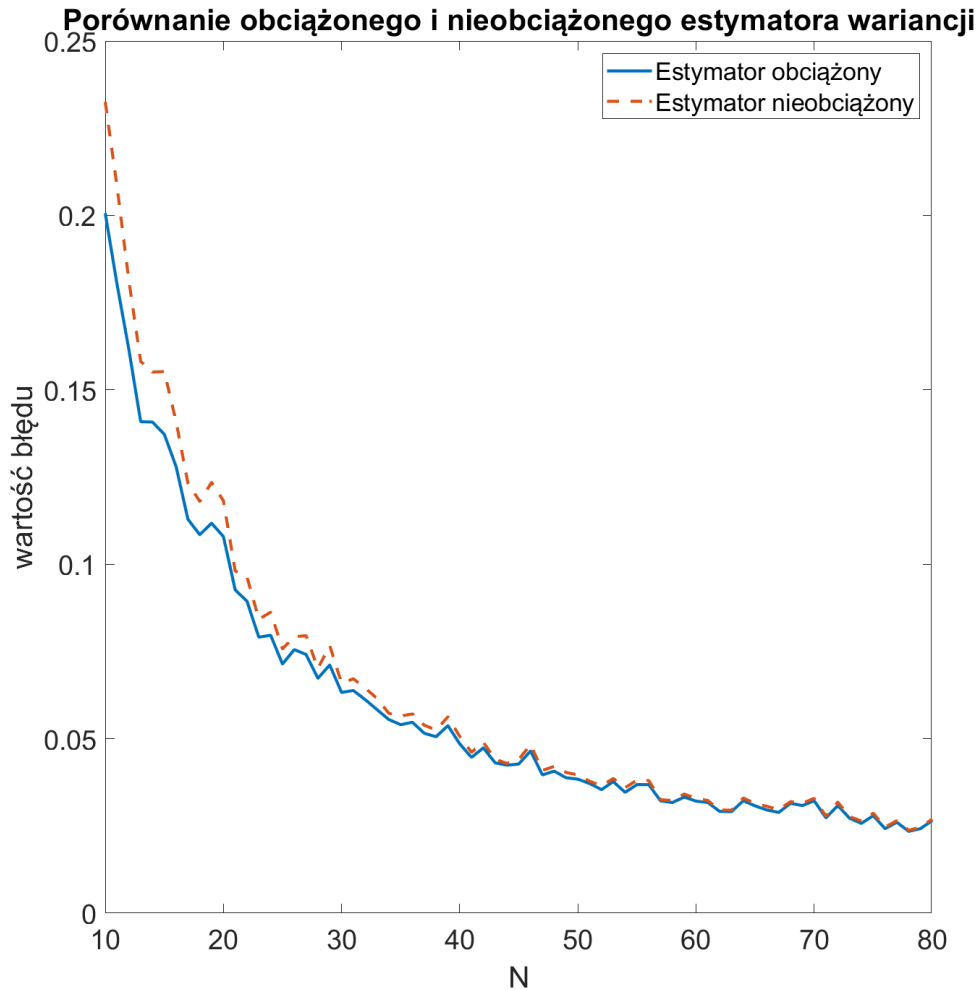


Rysunek 1.1. Błąd empiryczny estymatora średniej



Rysunek 1.2. Błędy empiryczne dla estymatora obciążonego i estymatora nieobciążonego

Porównując ze sobą dwa estymatory wariancji a mianowicie estymator obciążony i nieobciążony (Rysunek 1.3), można zauważyć, że różnica między tymi dwoma estymatorami jest jedynie zauważalna dla niewielkiej liczby obserwacji – co wynika bezpośrednio ze wzorów. Dodatkowo widać, że wartości błędów dla estymatora obciążonego są mniejsze niż dla estymatora nieobciążonego.

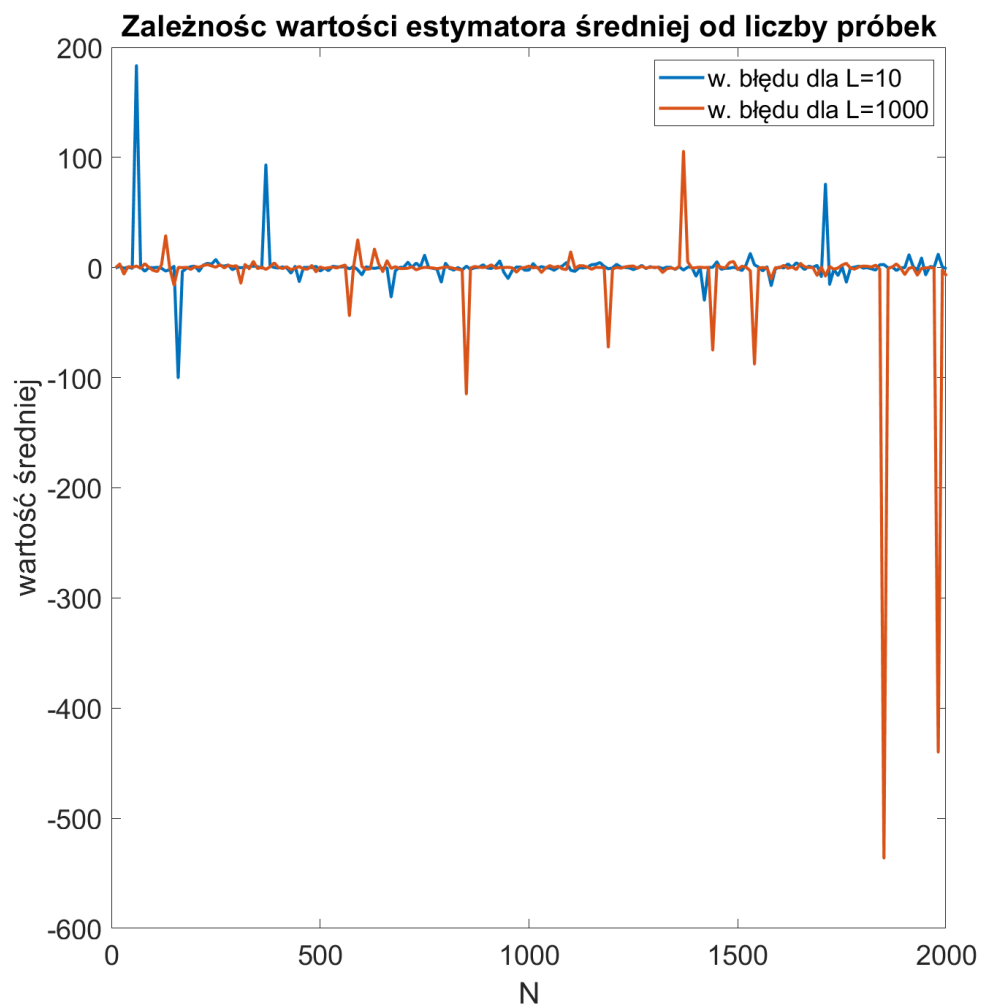


Rysunek 1.3. Porównanie obciążonego i nieobciążonego estymatora wariancji dla  $L = 1000$

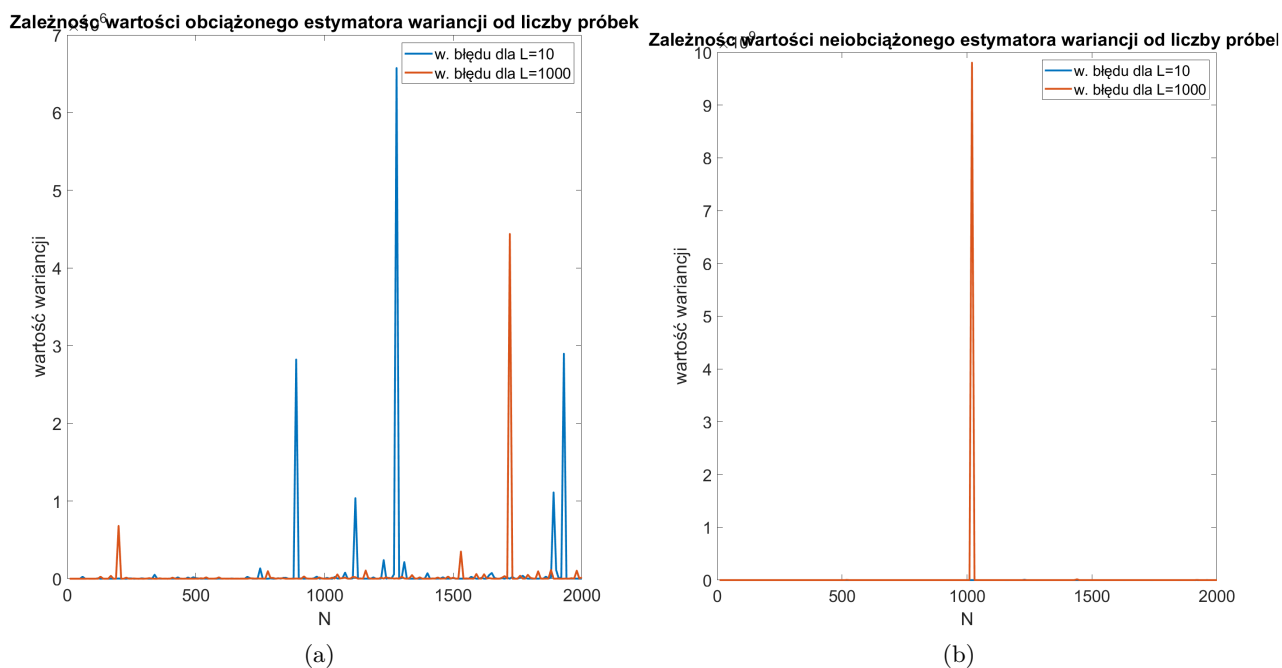
### 1.3 Zadanie 3 – powtórzenie wcześniej przedstawionych eksperymentów dla rozkładu Cauchy’ego

Powyższe zadanie zostało wykonane dla rozkładu Cauchy’ego o parametrach  $x_0 = 0$ ,  $\gamma = 1$ , maksymalnej liczbie obserwacji  $N = 2000$  i dwóch różnych wartości stałej  $L$  ( $L = 10$  oraz  $L = 100$ ).

Z rysunków (Rysunek 1.4 i Rysunek 1.5) można odczytać, że stała  $L$  nie ma żadnego wpływu na kształt wykresów. Podobnie jest w przypadku liczby próbek – one również nie mają żadnego wpływu na wartości błędów empirycznych każdego z estymatorów. Spowodowane jest to tym, że rozkład Cauchy’ego nie posiada określonej wartości oczekiwanej i wariancji.



Rysunek 1.4. Błąd empiryczny estymatora średniej dla rozkładu Cauchy'ego



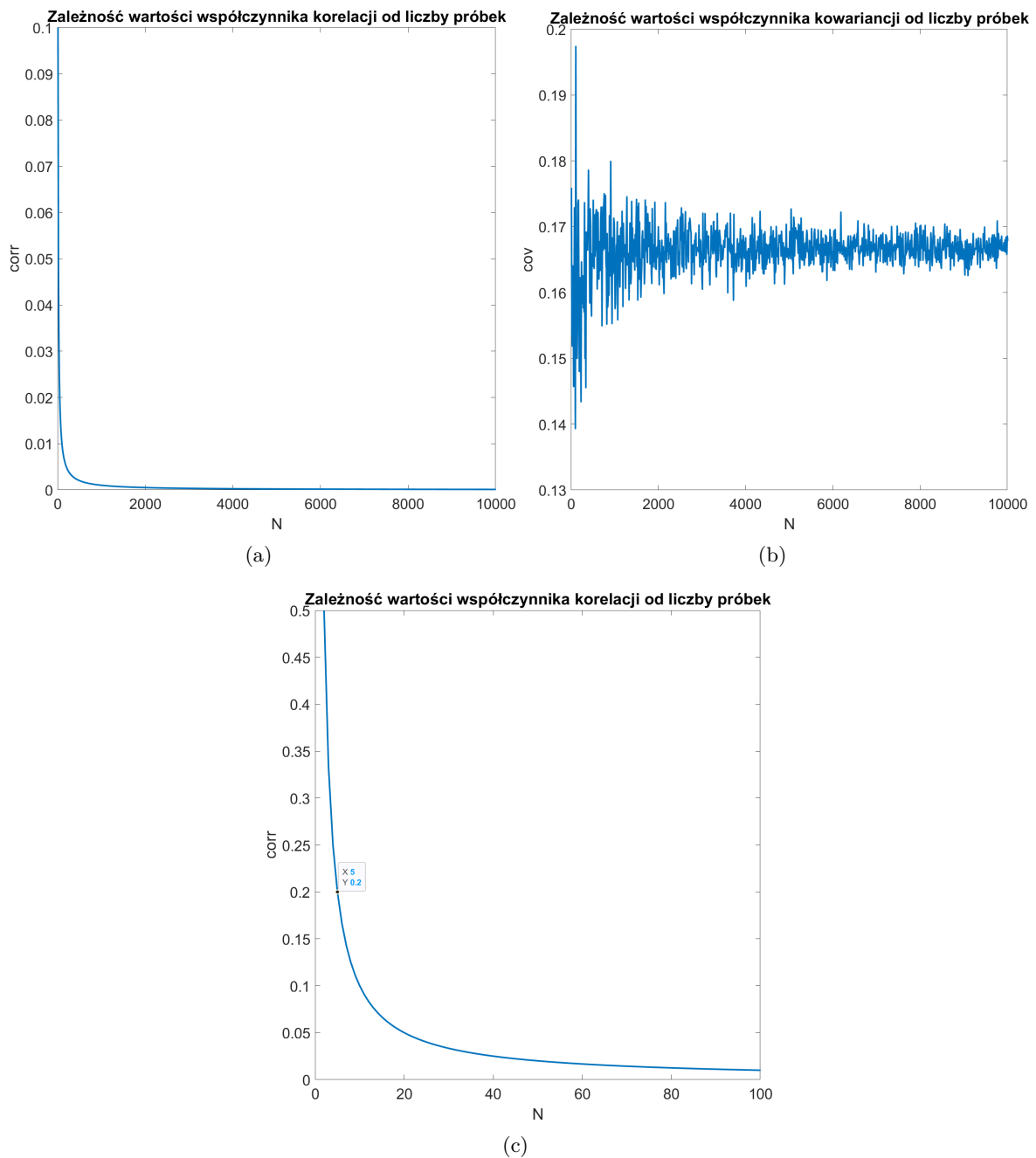
Rysunek 1.5. Błędy empiryczne dla estymatora obciążonego i estymatora nieobciążonego dla rozkładu Cauchy'ego

## 1.4 Zadanie dodatkowe 1 – implementacja estymatora kowariancji i korelacji

$$\hat{r}_N = \frac{\frac{1}{N} \sum_{n=1}^N (X_n - \bar{X}_N)(Y_n - \bar{Y}_N)}{\sqrt{\sum_{n=1}^N (X_n - \bar{X}_N)^2} \sqrt{\sum_{n=1}^N (Y_n - \bar{Y}_N)^2}} \quad (1.5)$$

Zadanie zostało wykonane dla parametrów  $a = 2$  oraz  $b = 3$ . Zmienne losowe zostały wygenerowane przy użyciu generatora stworzonego na zajęciach wprowadzających. Zmienna  $X_1 \sim U[0, 1]$  oraz  $X_2 = aX_1 + b$ . Następnie zostały wykreślone realizacje estymatora korelacji oraz kowariancji w funkcji liczebności próby  $N$ .

Z rysunku (Rysunek 1.6) wynika, że do pięciu próbek między zmiennymi występuje słaba dodatnia korelacja, natomiast powyżej pięciu próbek można dostrzec, że wartość współczynnika korelacji jest mniejsza od 0.2, zatem występuje brak korelacji między zmiennymi.



Rysunek 1.6. Realizacje estymatora korelacji i kowariancji w funkcji liczebności próby  $N$

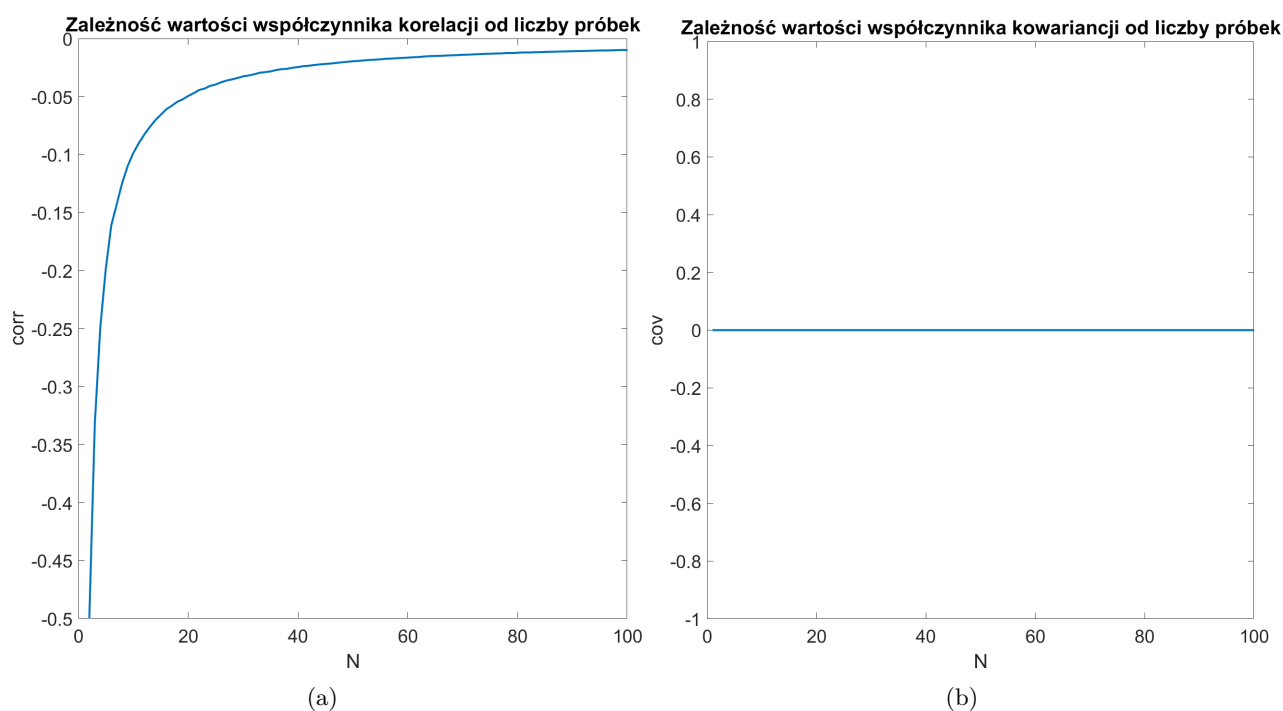
## 1.5 Zadanie dodatkowe 2 – implementacja estymatora kowariancji i korelacji

Zadanie zostało wykonane dla gęstości prawdopodobieństwa:

$$f(x) = \begin{cases} 2x & \text{dla } x \in [0, 1] \\ 0 & \text{dla } x \in (-\infty, 0) \cup (1, \infty) \end{cases} \quad (1.6)$$

Przyjęto, że zmienna  $X_1$  posiada gęstość prawdopodobieństwa taką, że  $f(x) = f(-x)$  dla każdego  $x$ . Dodatkowo zmienna  $X_2 = X_1^2$ .

Jak widać na rysunku (Rysunek 1.7) początkowo, do około pięciu próbek zmienne są ujemnie słabo skorelowane, następnie wartość współczynnika korelacji zaczyna dążyć do zera, co oznacza, że zmienne są nieskorelowane.



Rysunek 1.7. Realizacje estymatora korelacji i kowariancji w funkcji liczebności próby  $N$



## Rozdział 2

# Laboratorium 4 – estymator dystrybuanty empirycznej

### 2.1 Zadanie 2 – implementacja i wykreślenie dystrybuanty empirycznej dla ciągu liczb losowych z rozkładu:

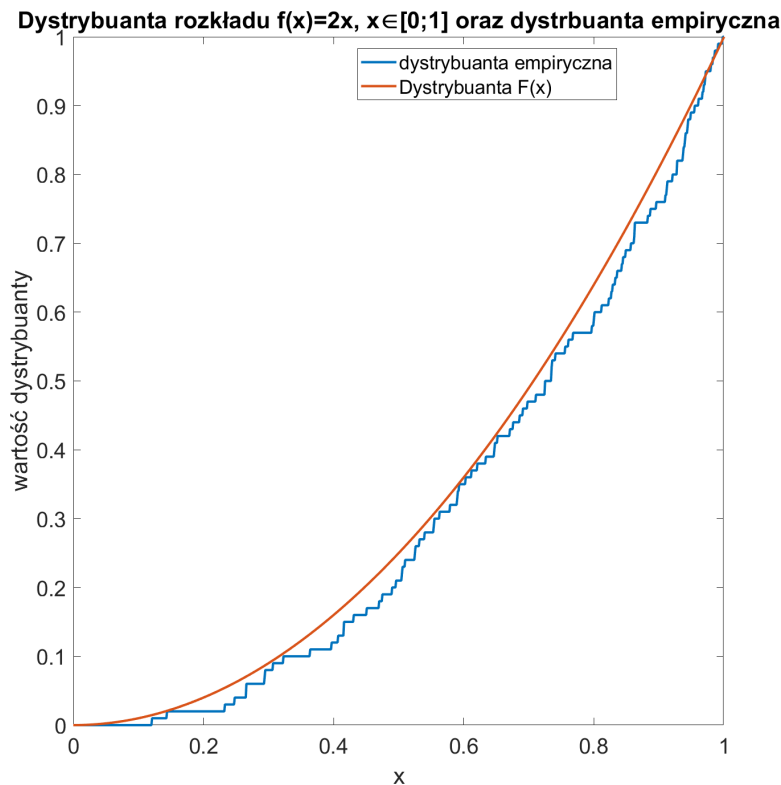
Zadanie zostało wykonane dla ciągu liczb z rozkładu 2.1 i liczbie próbek  $N = 100$ .

$$f(x) = \begin{cases} 2x & \text{dla } x \in [0, 1] \\ 0 & \text{dla } x \in (-\infty, 0) \cup (1, \infty) \end{cases} \quad (2.1)$$

Estymator dystrybuanty empirycznej:

$$\hat{F}_N = \frac{1}{N} \sum_{n=1}^N I(X_N \leq x), \quad (2.2)$$

gdzie  $I(a \leq b) = 1$ , gdy  $a \leq b$  i 0 w przeciwnym wypadku.

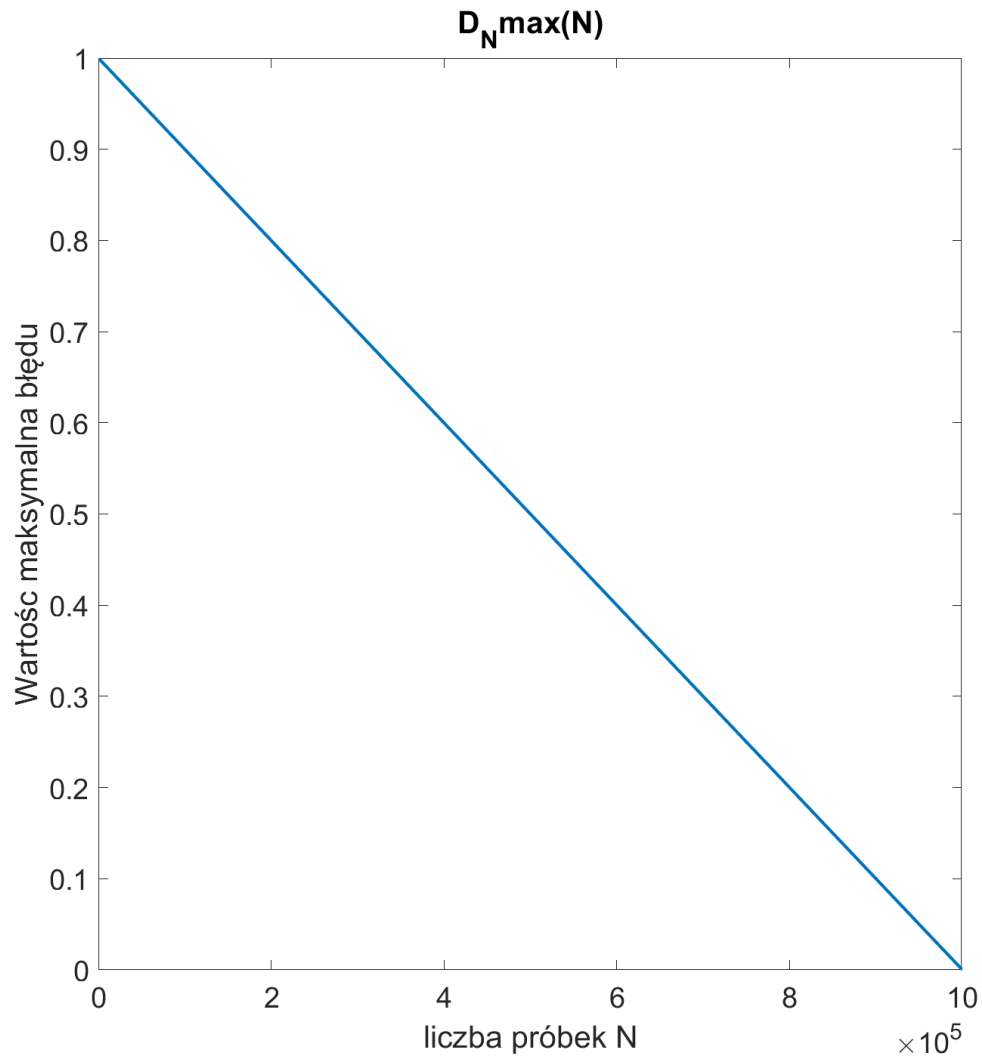


Rysunek 2.1. Wykreślona dystrybuanta empiryczna dla ciągu liczb losowych z rozkładu 2.1

## 2.2 Zadanie 3 – implementacja empirycznego odpowiednika wyrażenia:

$$D_N = \sup_x |\hat{F}_N - F(x)| \quad (2.3)$$

$D_N$  jest to moduł największej odległości między dystrybuantą empiryczną a rzeczywistą dystrybuantą.

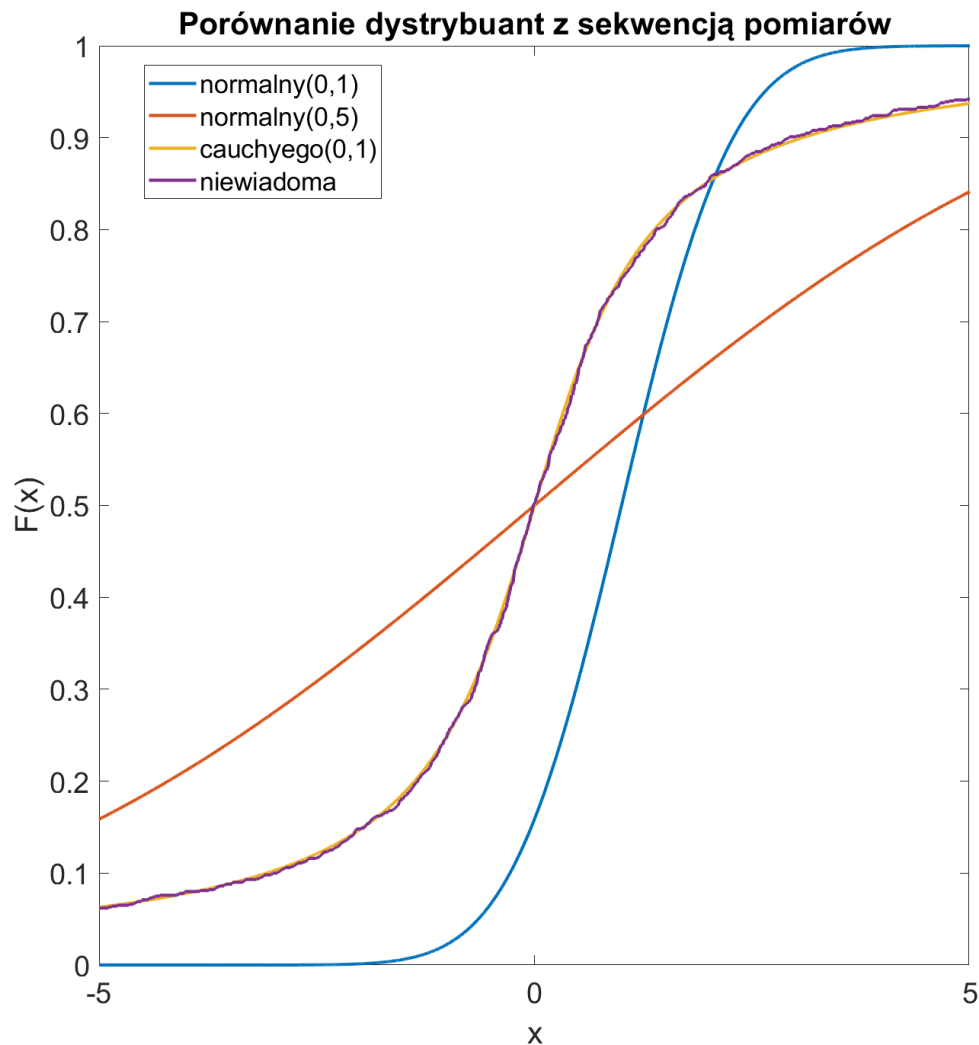


Rysunek 2.2. Największa odległość między dystrybuantą empiryczną a rzeczywistą dystrybuantą w zależności od liczby obserwacji

Analizując rysunek (Rysunek 2.2) można dojść do wniosku, że wraz ze wzrostem liczby obserwacji największa odległość między dystrybuantą empiryczną a dystrybuantą rzeczywistą maleje w sposób liniowy do zera.

## 2.3 Zadanie 4 – Znalezienie rozkładu dla podanych pomiarów

Dla wczytanych danych z pliku **ModelowanieLab4Data.txt** został wykreślony wykres (Rysunek 2.3) oraz trzy dystrybuanty a mianowicie: rozkład normalny  $\mathcal{N}(1, 1)$ , rozkład normalny  $\mathcal{N}(0, 5)$  oraz rozkład Cauchy’ego z parametrami  $x_0 = 0$  i  $\gamma = 1$ .



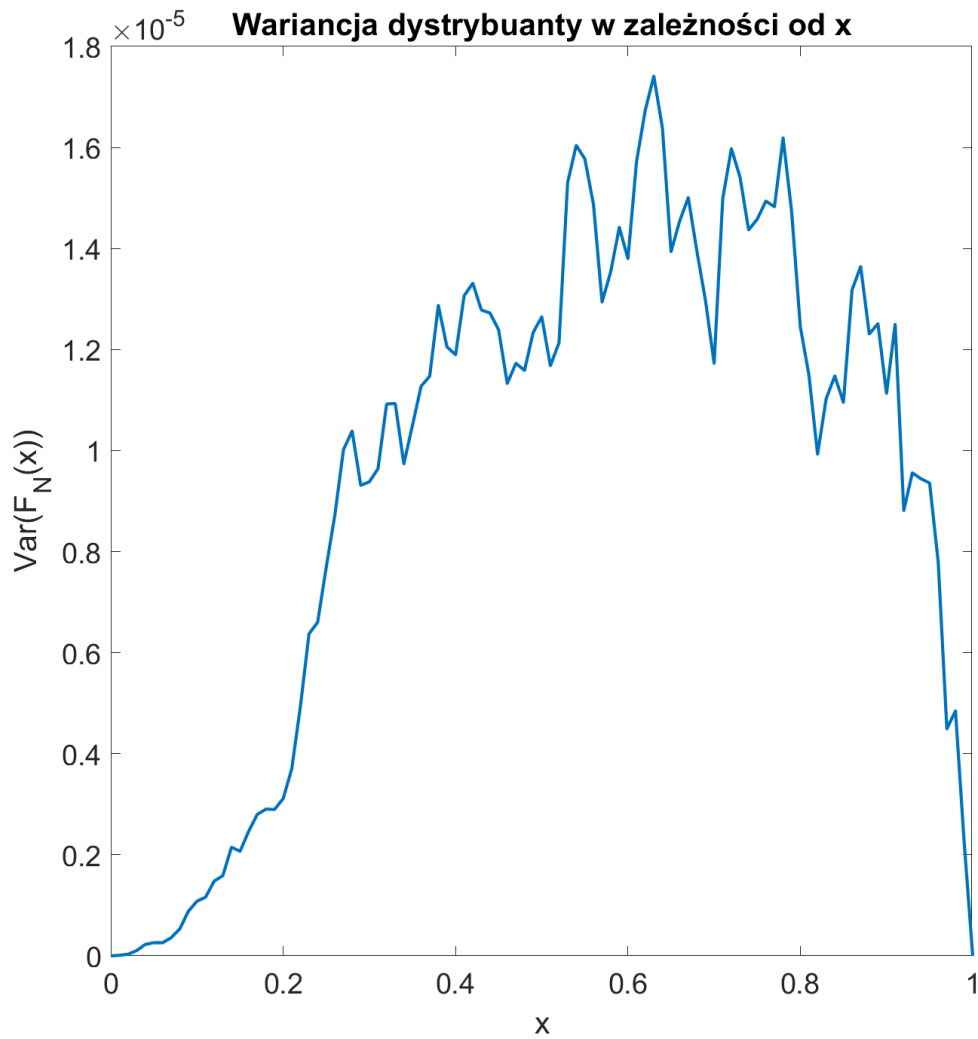
Rysunek 2.3. Porównanie danych pomiarowych do trzech różnych dystrybuant

## 2.4 Zadanie 5 – Konstrukcja eksperymentu symulacyjnego, umożliwiającego estymację wariancji dystrybuanty empirycznej:

$$\hat{\sigma}_F^2(x) = \text{Var}\{\hat{F}_N(x)\} \quad (2.4)$$

W celu wykonania tego zadania zostały wykorzystane wzory estymatorów średniej (1.1) i wariancji (1.2) z laboratorium nr 3. Następnie została wykreślona wartość oszacowania wariancji  $\hat{\sigma}_F^2(x)$  w funkcji  $x$ .

Jak widać na rysunku (Rysunek 2.4) początkowo wariancja dystrybuanty wzrasta, w okolicy  $x = 0.6$  osiąga swoje maksimum, a następnie z powrotem maleje do 0.



Rysunek 2.4. Porównanie danych pomiarowych do trzech różnych dystrybuant

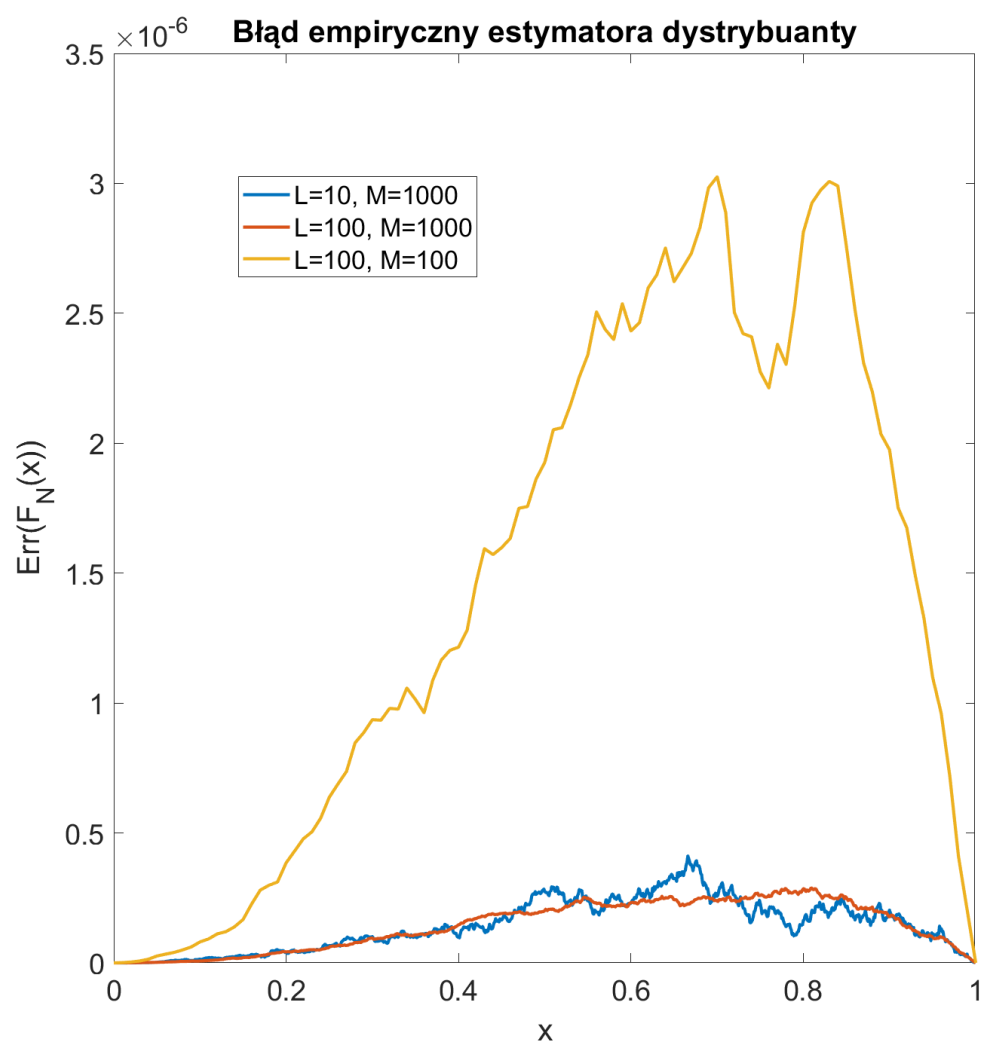
## 2.5 Zadanie dodatkowe 1 – Wyznaczanie błędu empirycznego:

$$Err\{\hat{F}_N\} = \frac{1}{LM} \sum_{l=1}^L \sum_{m=1}^M [\hat{F}_N^{[l]}(x_m) - F(x_m)]^2 \quad (2.5)$$

Zadanie zostało wykonane dla funkcji gęstości prawdopodobieństwa:

$$f(x) = \begin{cases} 2x & \text{dla } x \in [0, 1] \\ 0 & \text{dla } x \in (-\infty, 0) \cup (1, \infty) \end{cases} \quad (2.6)$$

Jak widać na rysunku (Rysunek 2.5 stała  $L$  ma wpływ na gładkość wykresu – pozbywa go oscylacji w pewnym stopniu. Dzięki większej stałej  $L$  uzyskujemy dokładniejszy błąd empiryczny dystrybuanty dla danego  $x$ . W przypadku stałej  $M$  można zauważyć, że jej wpływ na kształt wykresu jest bardzo duży. Wynika to bezpośrednio ze wzoru estymatora dystrybuanty empirycznej. Im większa jest wartość  $M$ , czyli liczba punktów  $x$ , tym uzyskany wynik dystrybuanty empirycznej będzie dokładniejszy, co przełoży się na mniejszy błąd empiryczny.



Rysunek 2.5. Błąd empiryczny estymatora dystrybuanty dla różnych wartości stałych  $M$  i  $L$

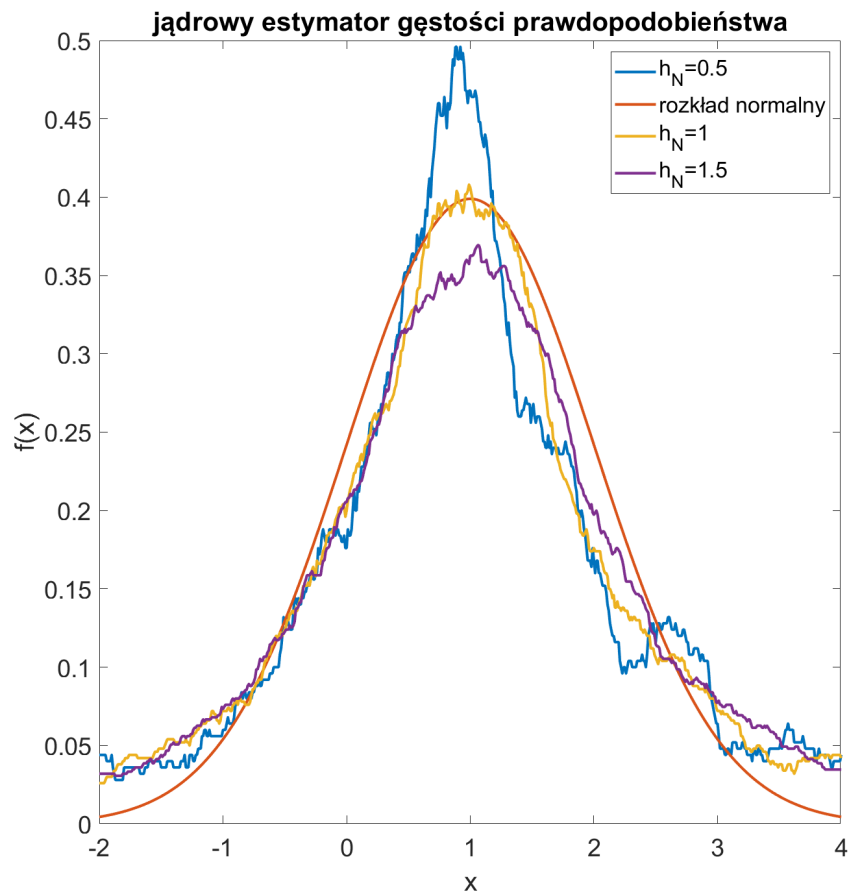
## Rozdział 3

# Laboratorium 5 – jądrowy estymator gęstości prawdopodobieństwa

### 3.1 Zadanie 2 – implementacja estymatora jądrowego gęstości prawdopodobieństwa

Zadanie to zostało wykonane dla ciągu liczb losowych o gęstości rozkładu  $\mathcal{N}(1, 1)$  o liczebności  $N = 500$  oraz jądra prostokątnego.

Analizując rysunek (Rysunek 3.1) można zauważyć, że wartość współczynnika wygładzania  $h_N$  ma wpływ na kształt uzyskanej gęstości prawdopodobieństwa. Zwiększanie  $h_N$  powoduje, że uzyskane wartości wokół wartości oczekiwanej się zwiększają, natomiast zmniejszanie wartości  $h_N$  powoduje odwrotny efekt. Zmiana parametru  $h_N$  ma niewielki wpływ na wykres w miejscach z dala od wartości oczekiwanej.



Rysunek 3.1. Porównanie gęstości prawdopodobieństwa dla różnych wartości  $h_n$  z rozkładem normalnym

## 3.2 Zadanie 3 – porównanie działania różnych funkcji jądra

Zadanie to zostało wykonane dla trzech różnych funkcji jądra, a mianowicie:  
jądra prostokątnego:

$$K(x) = \frac{1}{2}I(x) \quad (3.1)$$

jądra Gaussa:

$$K(x) = \frac{1}{\sqrt{2\pi}}e^{\frac{-x^2}{2}} \quad (3.2)$$

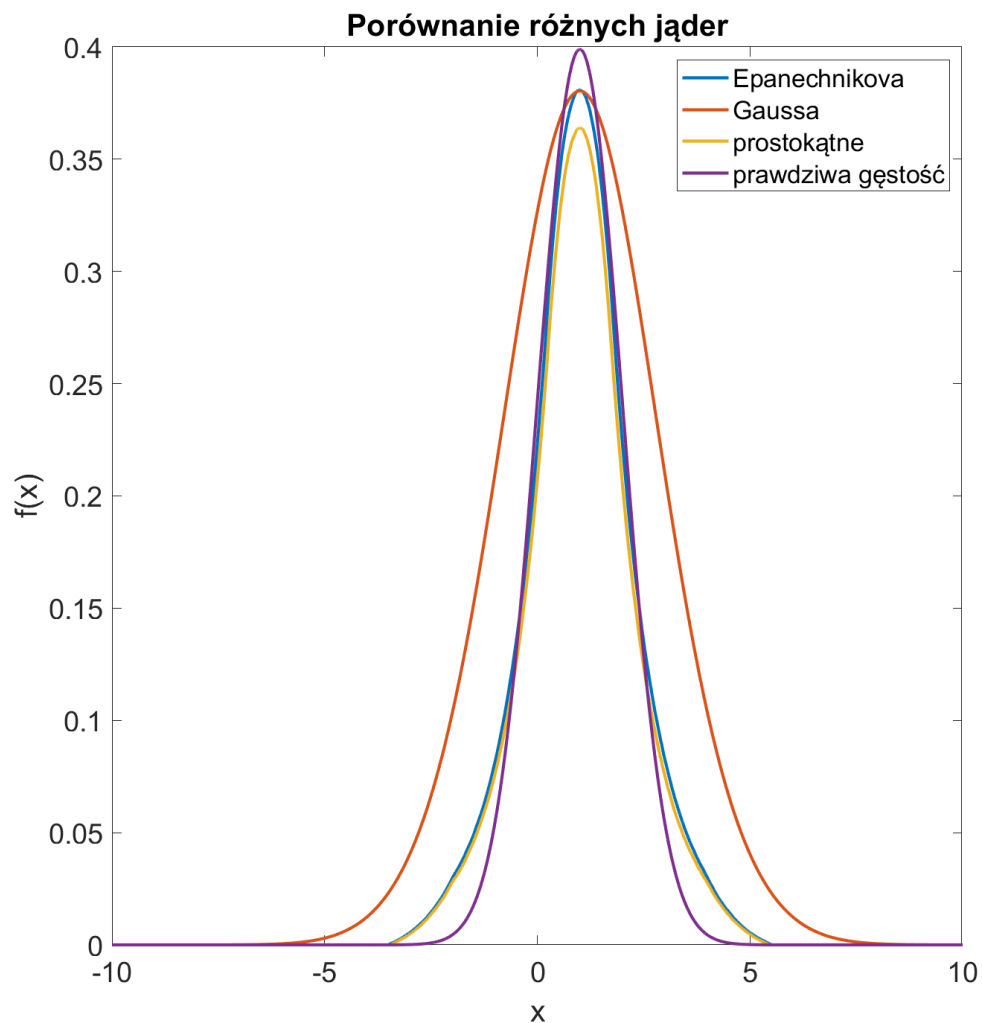
jądra Epanechnikova:

$$K(x) = \frac{3}{4}(1 - x^2)I(x) \quad (3.3)$$

### 3.2.1 Rozkład normalny

Przedstawiony wykres (Rysunek 3.2) został wykreślony dla liczby obserwacji  $N = 2 \cdot 10^5$  i wartości parametru wygładzania  $h_N = 0.8$ .

Z poniższego rysunku (Rysunek 3.2) wynika, że najlepszą funkcją jądra (z trzech użytych) jest funkcja jądra Epanechnikova. Dzięki wykorzystaniu tej funkcji jądra uzyskuje się najlepsze przybliżenie gęstości prawdopodobieństwa.

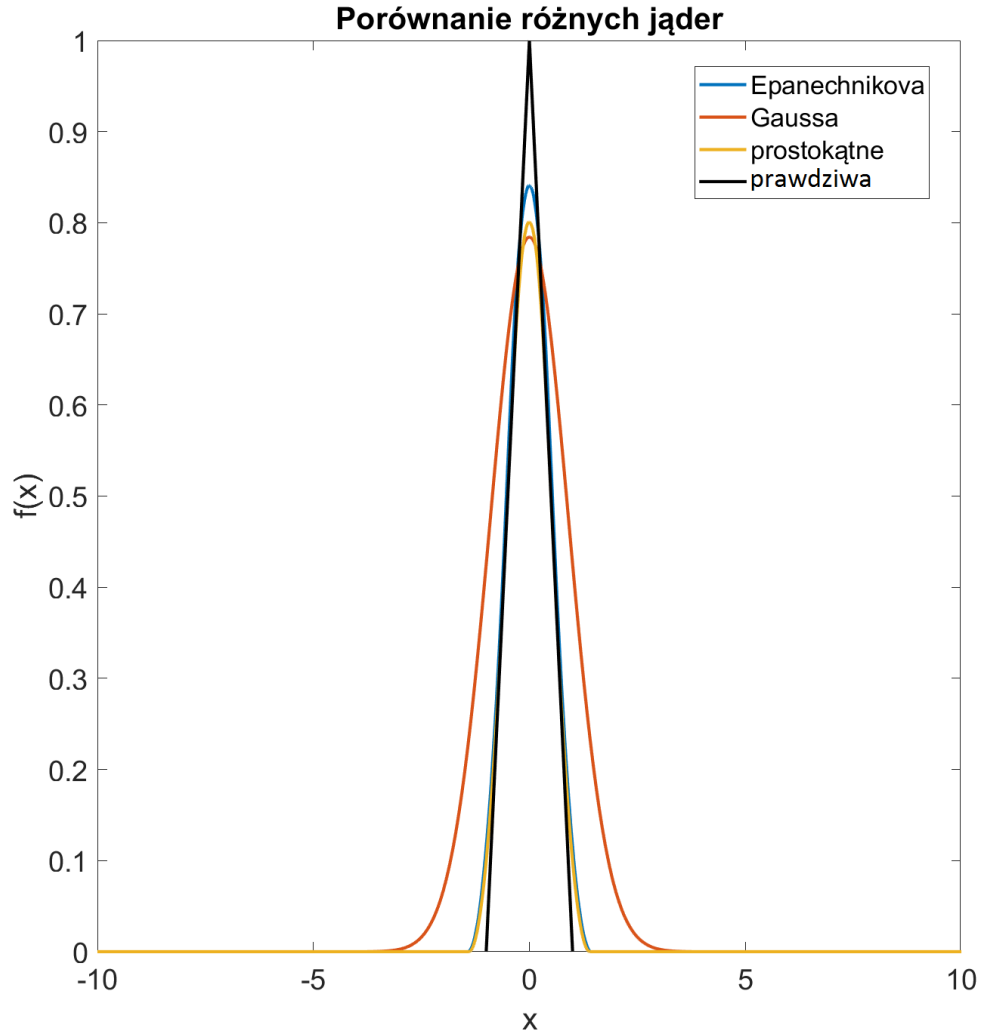


Rysunek 3.2. Porównanie działania różnych funkcji jądra dla rozkładu normalnego

### 3.2.2 Rozkład trójkątny

Przedstawiony wykres (Rysunek 3.3) został wykreślony dla liczby obserwacji  $N = 2 \cdot 10^5$  i wartości parametru wygładzania  $h_N = 0.8$ .

W przypadku tego zadania, dla rozkładu trójkątnego, uzyskany wynik jest identyczny jak poprzednio. Najlepszą funkcją jądra wykorzystaną do estymacji gęstości prawdopodobieństwa jest funkcja jądra Epanechnikova, dalej jądro prostokątne i jądro Gaussa.



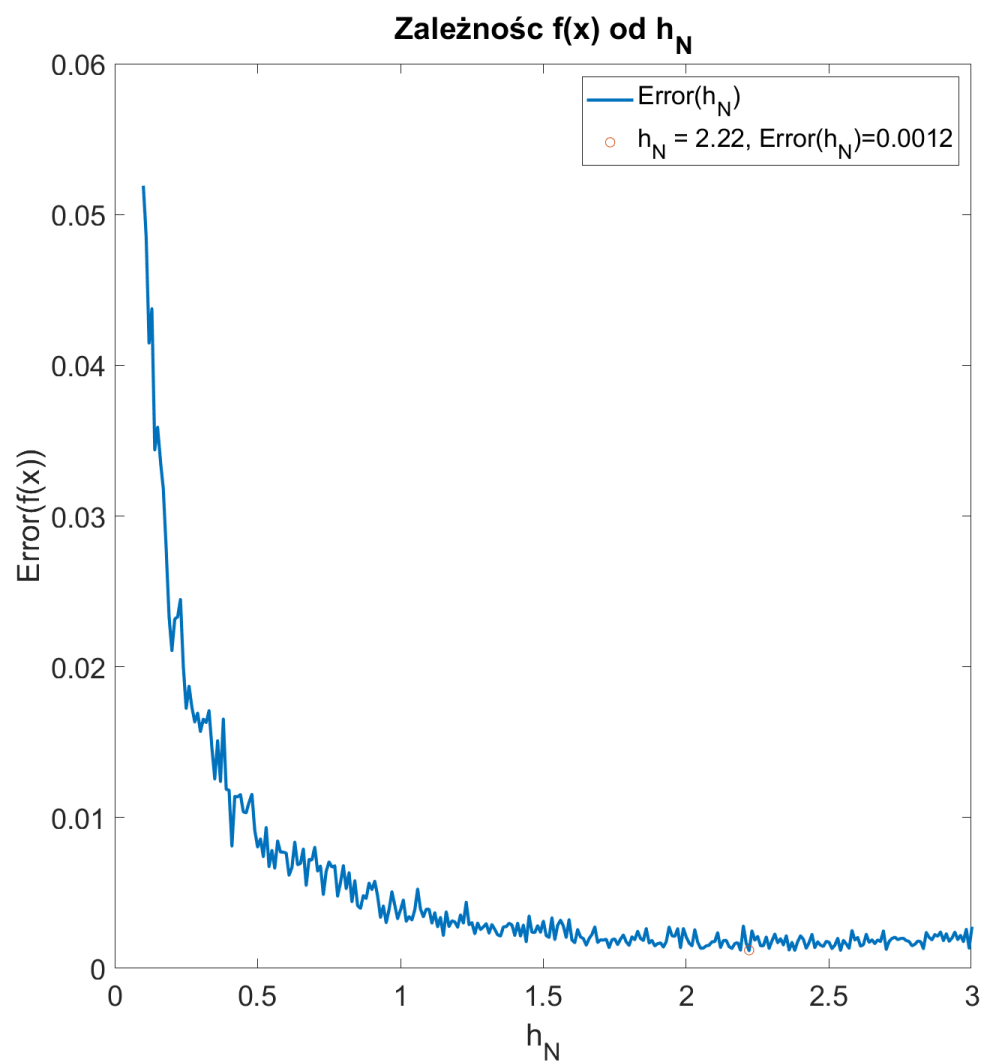
Rysunek 3.3. Porównanie działania różnych funkcji jądra dla rozkładu trójkątnego

### 3.3 Zadanie 4 – wyznaczenie błędu empirycznego gęstości prawdopodobieństwa

Zadanie zostało wykonane dla rozkładu normalnego o parametrach  $\mathcal{N}(1, 1)$ . Przyjęto  $M = 100$  oraz  $L = 10$ . Błąd  $Err\{\hat{f}_N\}$  został wykreślony w funkcji  $h_n$ .

Jak widać na rysunku (Rysunek 3.4) początkowo wartość błędu szybko maleje, aż do uzyskania minimum, następnie ponownie wzrasta, lecz tym razem znacznie wolniej. Najlepszym parametrem wygładzania  $h_N$  dla estymacji rozkładu normalnego  $\mathcal{N}(1, 1)$  okazuje się  $h_N = 2.22$ , dla którego błąd wynosi  $Err(h_n) = 0.0012$ .





Rysunek 3.4. Porównanie działania różnych funkcji jądra dla rozkładu trójkątnego

## Rozdział 4

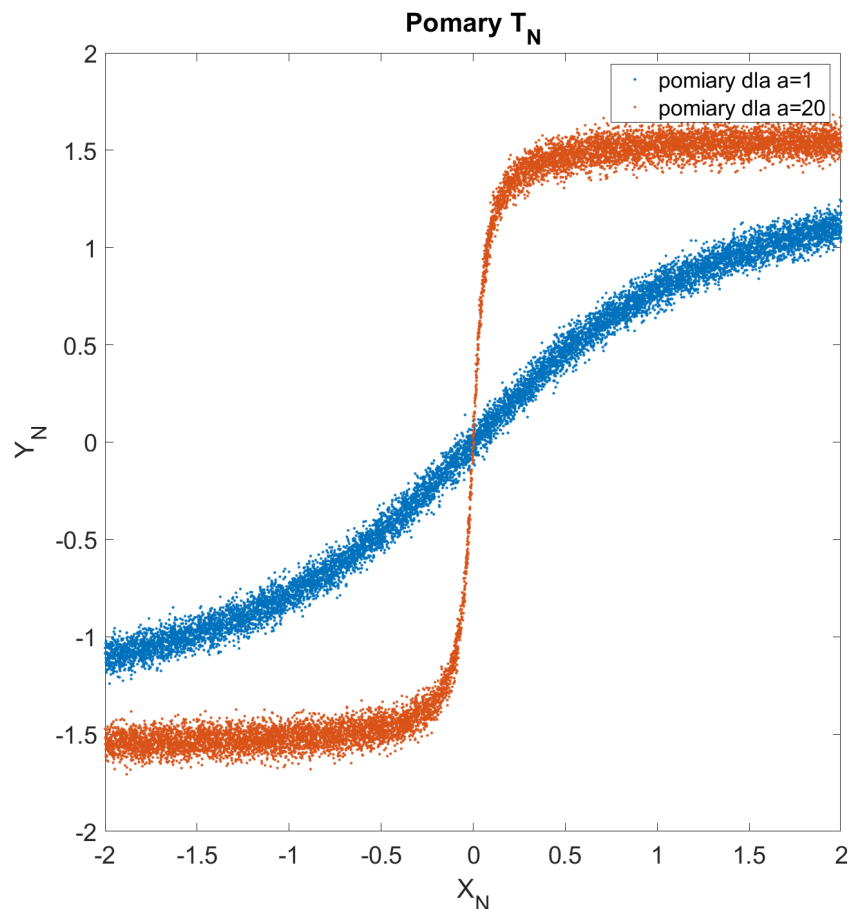
# Laboratorium 6 – jądrowy estymator funkcji regresji

### 4.1 Zadanie 2 – Wykreślenie nieliniowej charakterystyki systemu

#### 4.1.1 Zakłócenie z rozkładu normalnego

Zadanie to zostało wykonane dla statycznego systemu liniowego z charakterystyką  $m(x) = a \tan(a \cdot x)$ . Wykorzystany został również ciąg zmiennych losowych  $N = 110^5$  o rozkładzie jednostajnym  $U[-2, 2]$ . Dodatkowo wygenerowano sygnał zakłócający  $\{Z_n\}$  o rozkładzie normalnym  $\mathcal{N}(0, \sigma_Z^2)$ . Przyjęto wartość  $\sigma_Z^2 = 0.05$ , aby zakłócenie nie było większe od wartości sygnału podawanego na wejście.

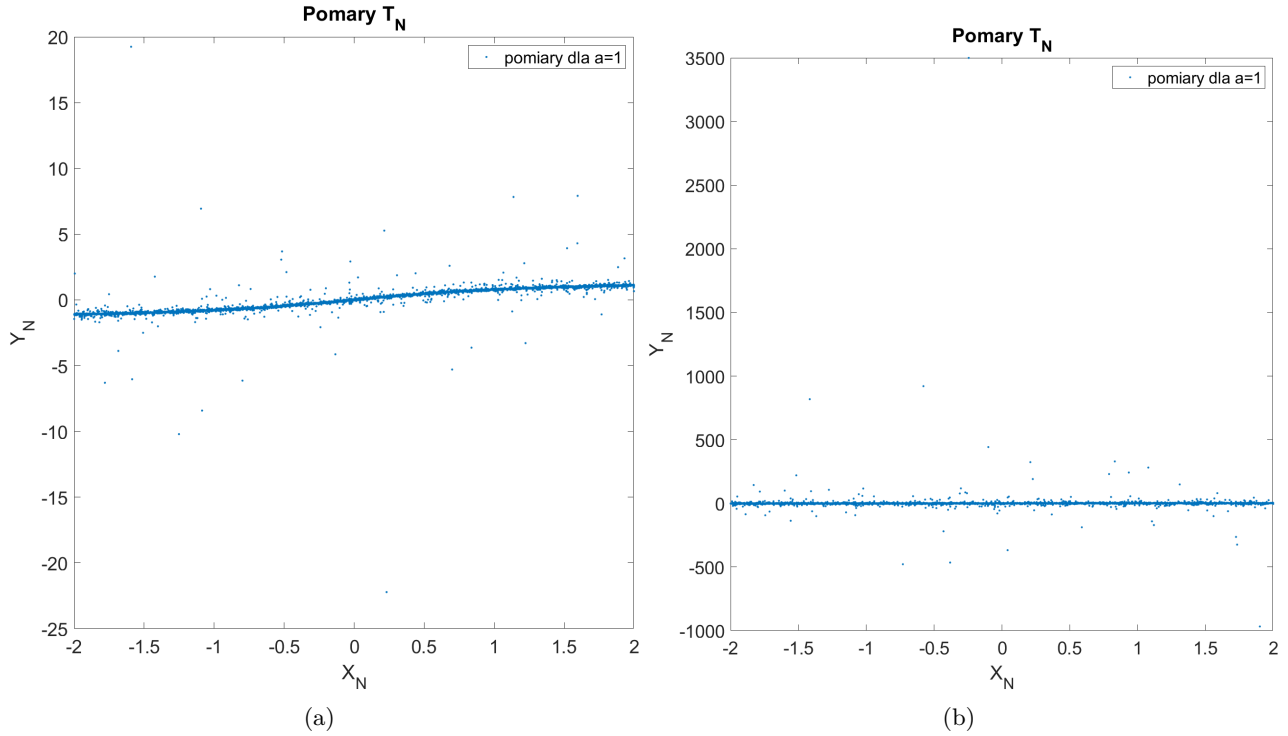
Jak widać na rysunku (Rysunek 4.1) zmniejszanie stałej  $a$  powoduje, że punkty zaczynają przypominać funkcję liniową, natomiast zwiększanie powoduje, że punkty układają się na kształt funkcji *arcustangens*.



Rysunek 4.1. Chmura pomiarów ze zbioru  $T_N$

### 4.1.2 Zakłócenie z rozkładu Cauchy’ego

Uzyskany wynik (Rysunek 4.2) przedstawia chmurę pomiarów ze zbioru  $T_N$ . Wykreślone punkty w większości znajdują się w obrębie funkcji  $\text{atan}(a \cdot x)$ , natomiast pojawiają się też punkty daleko oddalone od tej funkcji.



Rysunek 4.2. chmura pomiarów ze zbioru  $T_N$  a) dla  $\gamma = 0.01$  b) dla  $\gamma = 0.5$

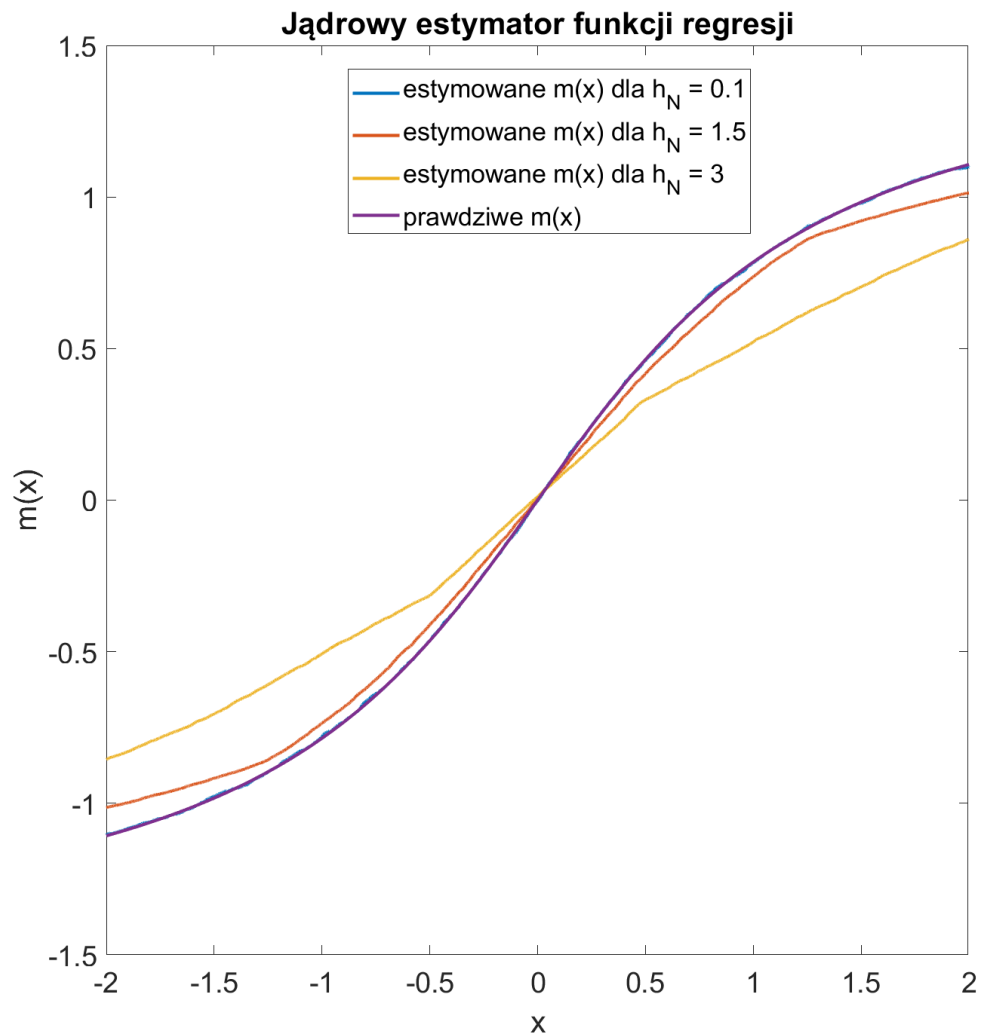
## 4.2 Zadanie 3 – implementacja jądrowego estymatora funkcji regresji:

$$\hat{m}_N = \frac{\sum_{n=1}^N Y_N K\left(\frac{X_n - x}{h_n}\right)}{\sum_{n=1}^N K\left(\frac{X_n - x}{h_n}\right)} \quad (4.1)$$

### 4.2.1 Zakłócenie z rozkładu normalnego

Zadanie to zostało wykonane dla liczby obserwacji  $N = 1 \cdot 10^5$  i wartości stałej  $a = 1$  oraz jądra prostokątnego.

Uzyskane wyniki zostały przedstawione na rysunku (Rysunek 4.3). Jak widać, dla parametru  $h_N = 0.1$  funkcja regresji najlepiej oddaje jej rzeczywistą postać. Zwiększanie  $h_N$  powoduje, że funkcja regresji dąży do charakterystyki liniowej.

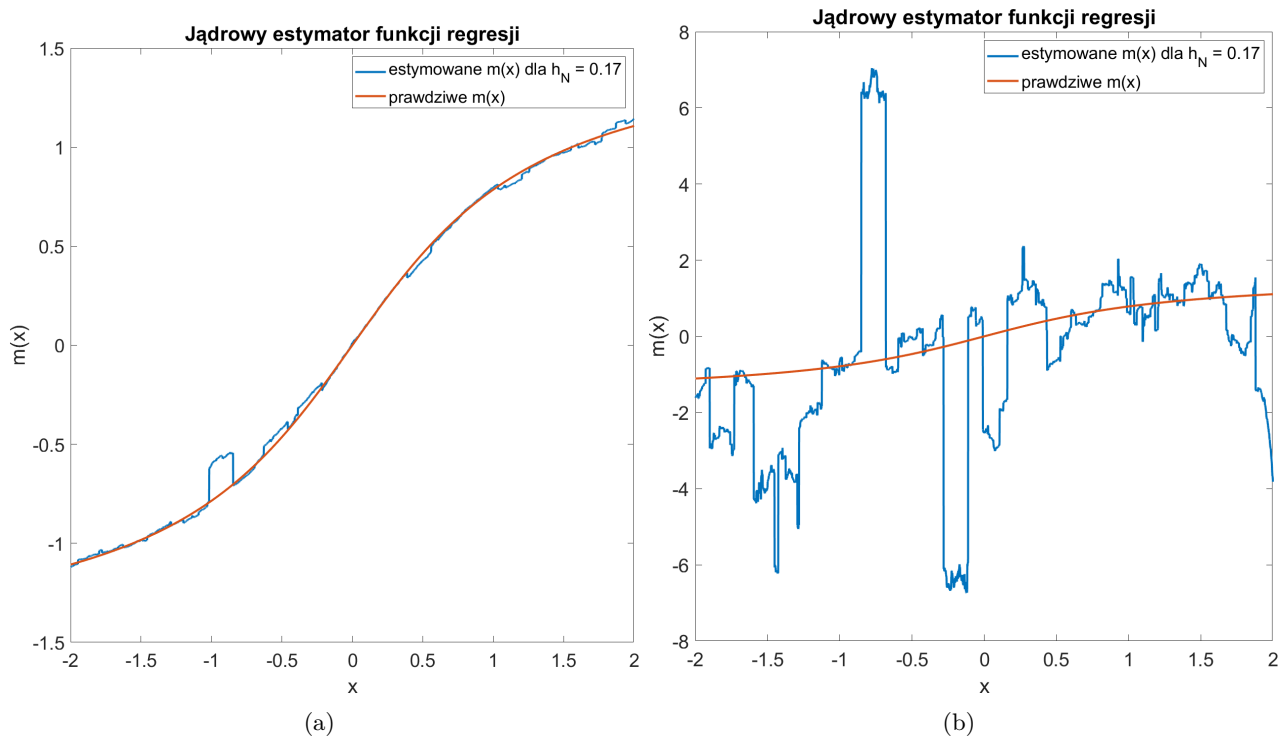


Rysunek 4.3. Estymacja funkcji regresji dla różnych wartości  $h_N$

#### 4.2.2 Zakłócenie z rozkładu Cauchy'ego

Jak widać na rysunku (Rysunek 4.4) estymowane  $\hat{m}(x)$  dla parametru  $h_N = 0.17$  oraz  $\gamma = 0.01$  całkiem dobrze sobie radzi. Czasami jednak można  $\hat{m}(x)$  odskakuje na bok. Dzieje się tak, ponieważ liczby generowane z rozkładu Cauchy'ego czasami przyjmują ogromne wartości, które mocno wpływają na uzyskaną średnią.

W przypadku wykresu dla  $\gamma = 0.5$  widać dobitniej cechy rozkładu Cauchy'ego. Dla tak podanej wartości  $\gamma$  losowane liczby często przyjmują większe wartości niż w przypadku  $\gamma = 0.01$ .



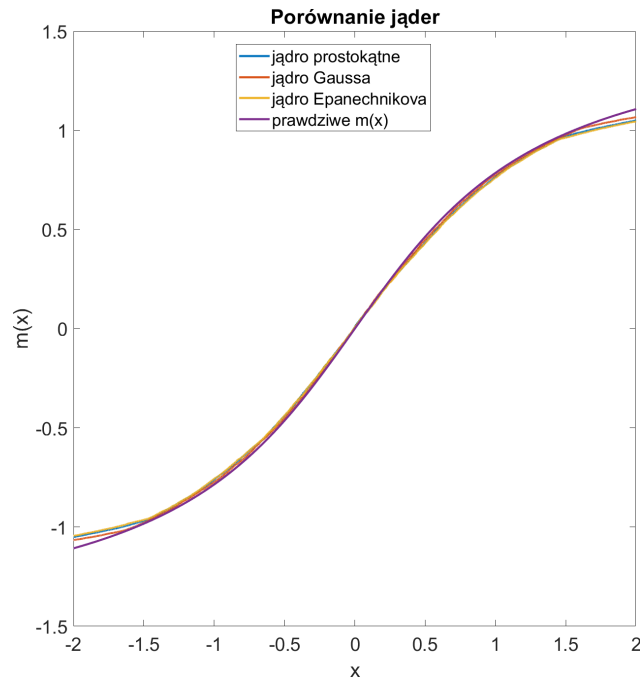
Rysunek 4.4. Funkcja regresji  $T_N$  a) dla  $\gamma = 0.01$  b) dla  $\gamma = 0.5$

## 4.3 Zadanie 4 – estymacja funkcji regresji dla różnych funkcji jądra

### 4.3.1 Zakłócenie z rozkładu normalnego

Zadanie to zostało wykonane dla liczby obserwacji  $N = 1 \cdot 10^5$ , stałej wartości  $h_N = 0.17$  dla trzech różnych funkcji jądra (prostokątnego 3.1, Gaussa 3.2 i Epanechnikova 3.3) oraz dwóch wartości stałej  $a$  ( $a = 1$  oraz  $a = 20$ ).

Jak widać na rysunku (Rysunek 4.5) dla parametru  $a = 1$  wszystkie jądra poradziły sobie z estymacją równie dobrze, natomiast dla parametru  $a = 20$  wykorzystane jądra nie poradziły sobie z odwzorowaniem prawdziwej funkcji regresji.

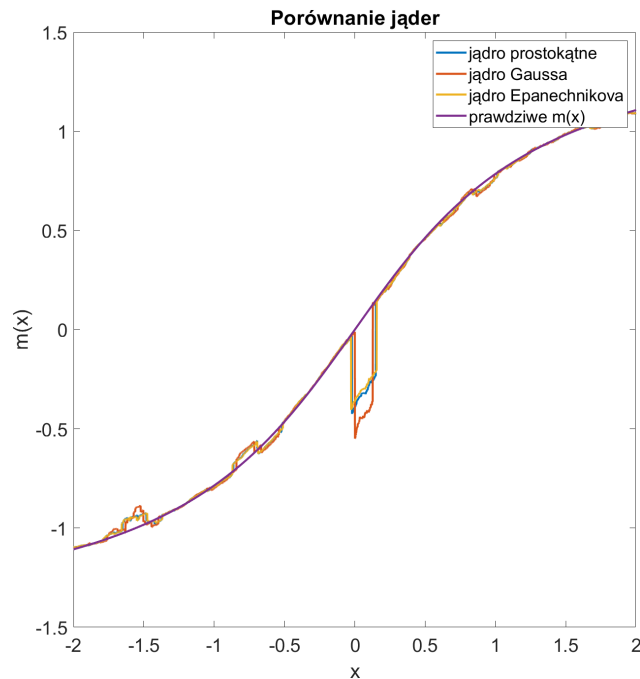


Rysunek 4.5. Porównanie funkcji jąder a) dla  $a = 1$  b) dla  $a = 20$

#### 4.3.2 Zakłócenie z rozkładu Cauchy’ego

Zadanie zostało wykonane dla wartości parametru  $h_N = 0.11$  (Uzyskanej podczas minimalizacji tego parametru) i liczbie obserwacji  $N = 1 \cdot 10^5$ .

Ponownie, jak poprzednio, najlepszą estymację udało się uzyskać dla jądra Epanechnikova, natomiast, na  $\hat{m}(x)$  wciąż są widoczne wpływy cech rozkładu Cauchy’ego.



Rysunek 4.6. Porównanie funkcji jąder dla  $a = 1$  i  $\gamma = 0.01$

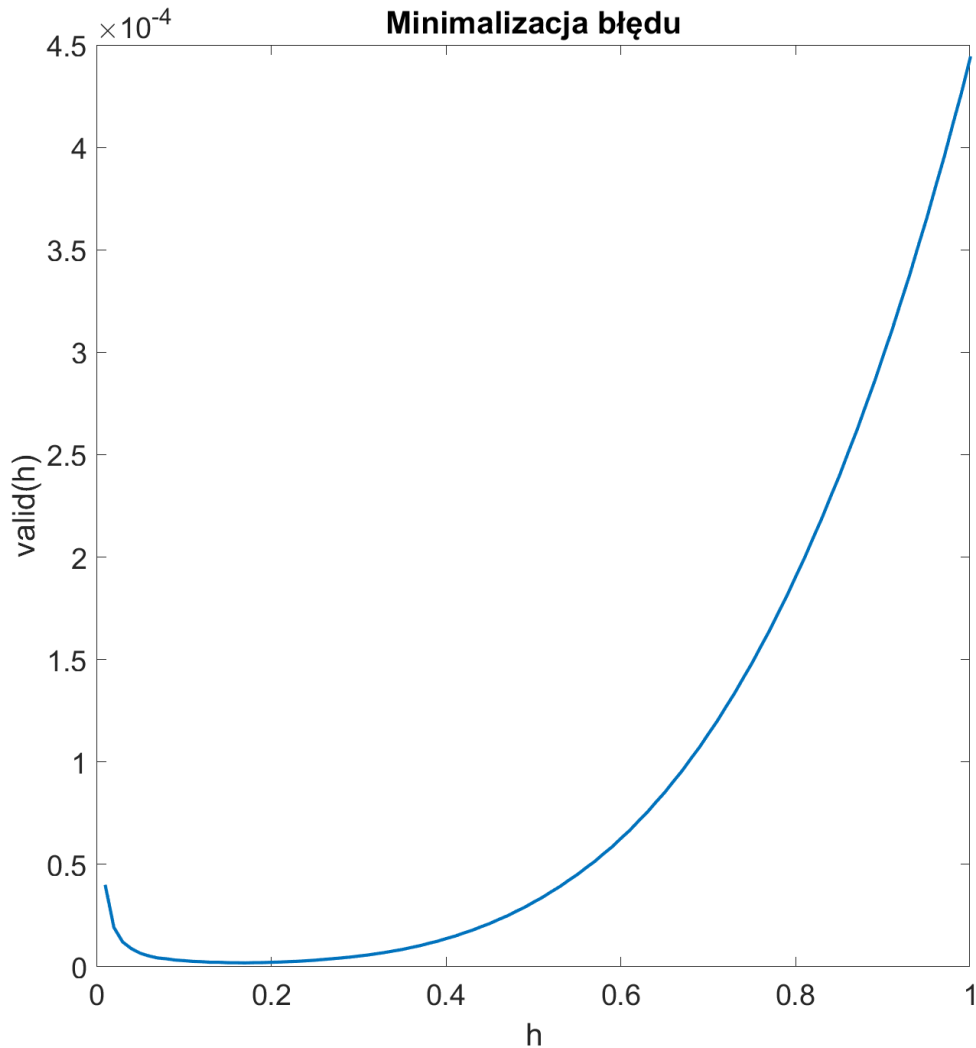
## 4.4 Zadanie 5 – minimalizacja błędu

### 4.4.1 Zakłócenie z rozkładu normalnego

W celu wykonania tego zadania zostało przyjęte  $Q = 100$ .

Jak widać na rysunku (Rysunek 4.7) błąd początkowo szybko maleje, następnie wykładniczo rośnie w kierunku nieskończoności. Najmniejszy błąd udało się uzyskać dla  $h = 0.17$  dla stałej  $a = 1$ , natomiast dla stałej  $a = 20$  uzyskano najmniejszy błąd dla  $h = 0.02$ .

$$valid(h) = \frac{1}{2Q} \sum_{i=-Q}^Q [\hat{m}_N(\frac{q}{Q}) - m(\frac{q}{Q})]^2 \quad (4.2)$$

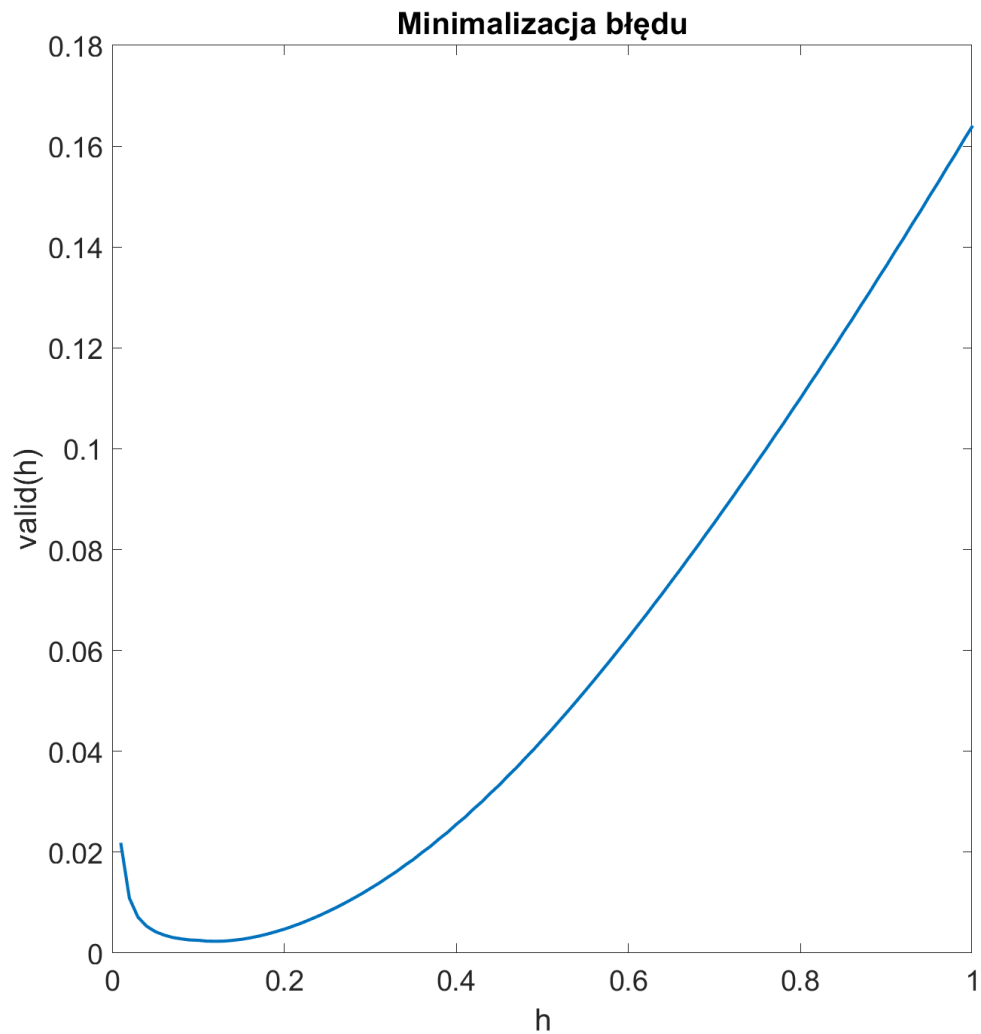


Rysunek 4.7. Zależność wartości błędu od parametru  $h$

### 4.4.2 Zakłócenie z rozkładu Cauchy'ego

Zadanie zostało wykonane z takimi samymi parametrami jak poprzednie, z tą różnicą, że zakłóceniem są liczby losowe wygenerowane z rozkładu Cauchy'ego o parametrach  $\gamma = 1$  i  $x_0 = 0$ .

Najmniejszy błąd udało się uzyskać dla parametru  $h = 0.11$ .



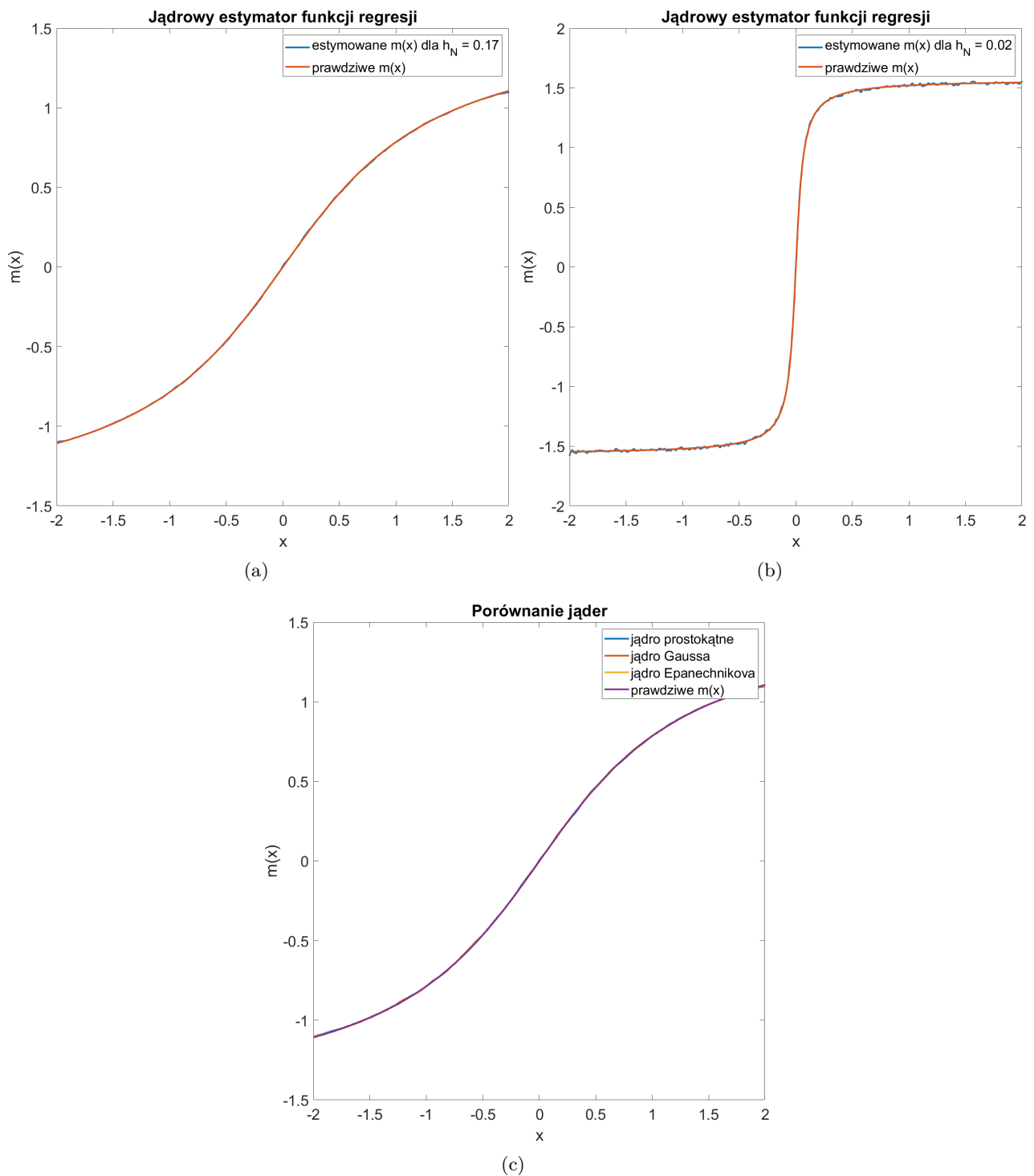
Rysunek 4.8. Zależność wartości błędu od parametru  $h$

## 4.5 Zadanie 6 – rysowanie wykresów z zadania 3 oraz 4 dla parametru wygładzania, dla którego uzyskano najmniejszy błąd

### 4.5.1 Zakłócenie z rozkładu normalnego

Zadanie zostało wykonane dla liczby próbek  $N = 1 \cdot 10^5$ . Jak widać na rysunku (Rysunek 4.9) zarówno dla stałej  $a = 1$  jak i  $a = 20$  estymowana funkcja regresji bardzo dobrze odwzorowuje prawdziwą funkcję regresji jeśli dobierze się odpowiednio wartość parametru wygładzania  $h$ .



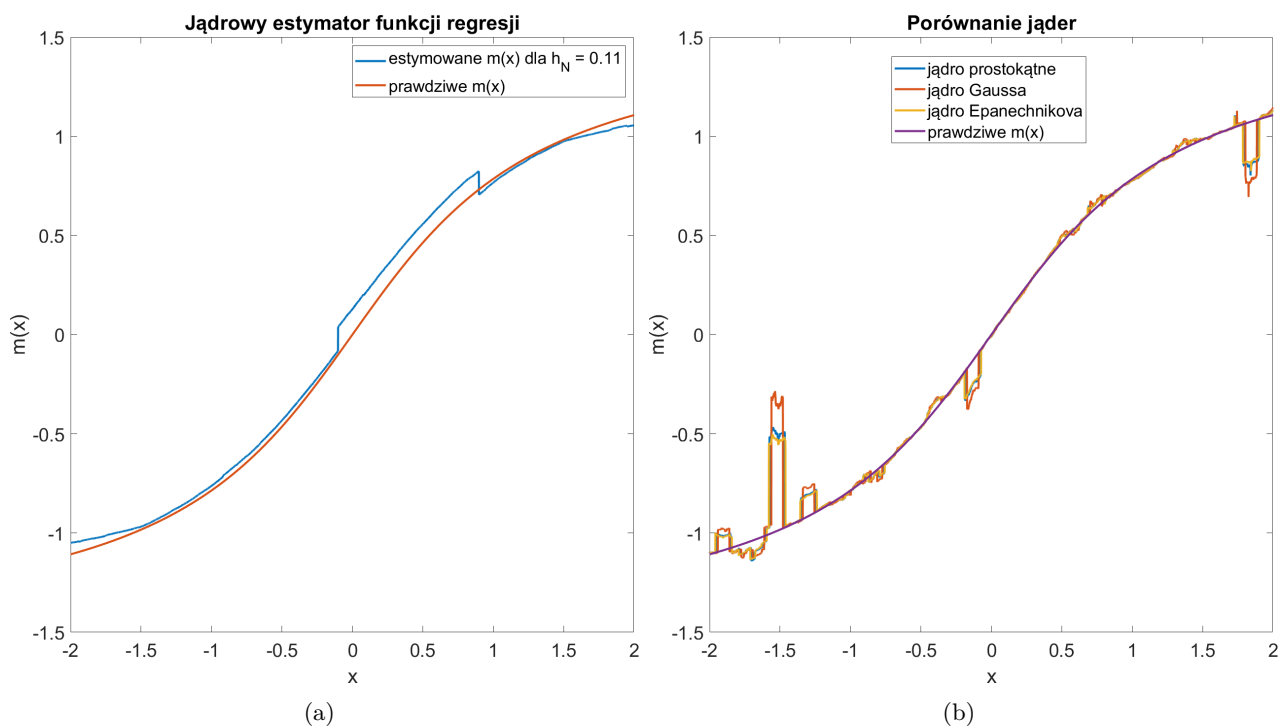


Rysunek 4.9. a) Zadanie 3 dla  $a = 1$  b) Zadanie 3 dla  $a = 20$  c) Zadanie 4 dla  $a = 1$

#### 4.5.2 Zakłócenie z rozkładu Cauchy'ego

Zadanie zostało wykonane dla  $N = 1 \cdot 10^5$  próbek, dla parametru  $h_N = 0.11$  (uzyskanego z minimalizacji błędu).

Jak widać na rysunku (Rysunek 4.10) wykreślone funkcje regresji dla najlepszego parametru  $h_N$  bardzo dobrze pokrywają się z prawdziwą funkcją regresji poza odskokami, spowodowanymi cechami rozkładu Cauchy'ego. Porównując bezpośrednio ze sobą oba wykresy z zdania 3 i 4 można zauważyć, że z każdym nowym ciągiem losowych liczb estymowana funkcja regresji będzie miała odskoki o różnych wartościach i w różnych miejscach.

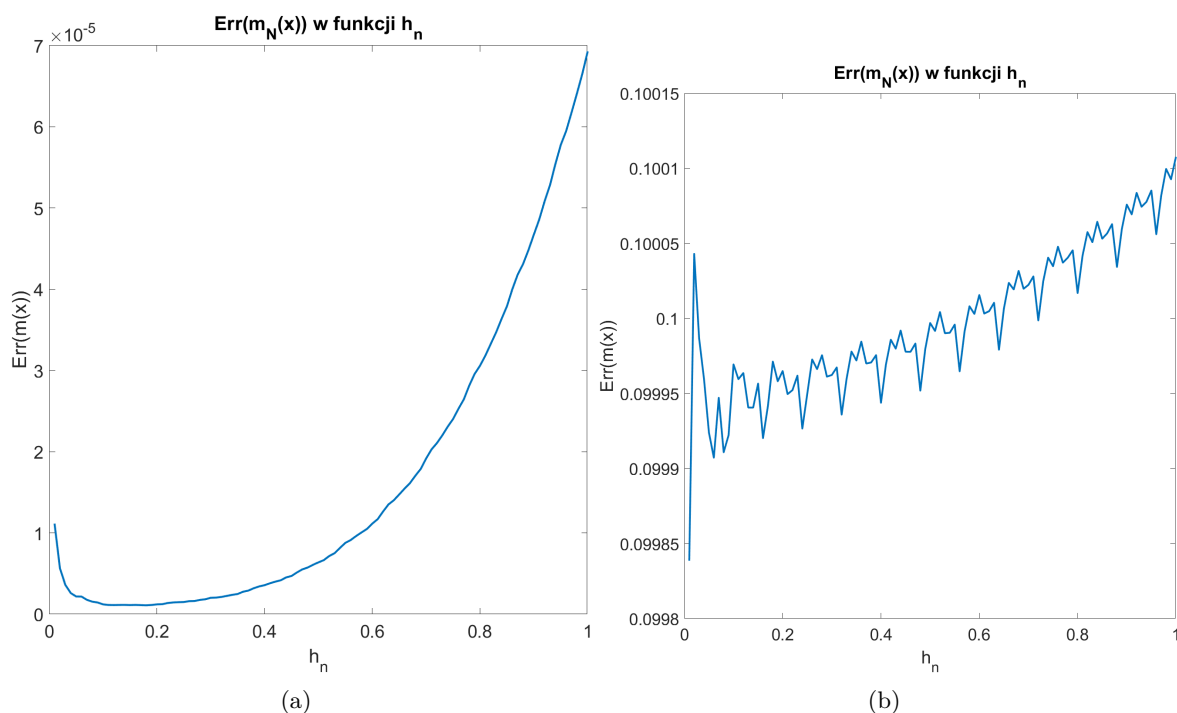


Rysunek 4.10. a) Zadanie 3 dla  $a = 1$  b) Zadanie 4 dla  $a = 1$

## 4.6 Zadanie dodatkowe 1 – błąd empiryczny w funkcji $h$

Zadanie to zostało wykonane dla funkcji regresji podanej w zadaniu pierwszym ( $\tan(a \cdot x)$ ) i zakłóceniu ( $\mathcal{N}(0, 1)$  i  $\mathcal{N}(1, 1)$ ). Przyjęto  $M = 100$  oraz  $L = 10$ , następnie wykreślono  $Err\{\hat{m}_N\}$  w funkcji  $h_N$ .

Dla wartości oczekiwanej zakłócenia  $E\{Z_n\} = 0$  uzyskany wynik (Rysunek 4.11) będzie bardzo zbliżony do tego uzyskanego w zadaniu nr 5, gdzie był minimalizowany błąd przy użyciu funkcji  $valid(h)$ .



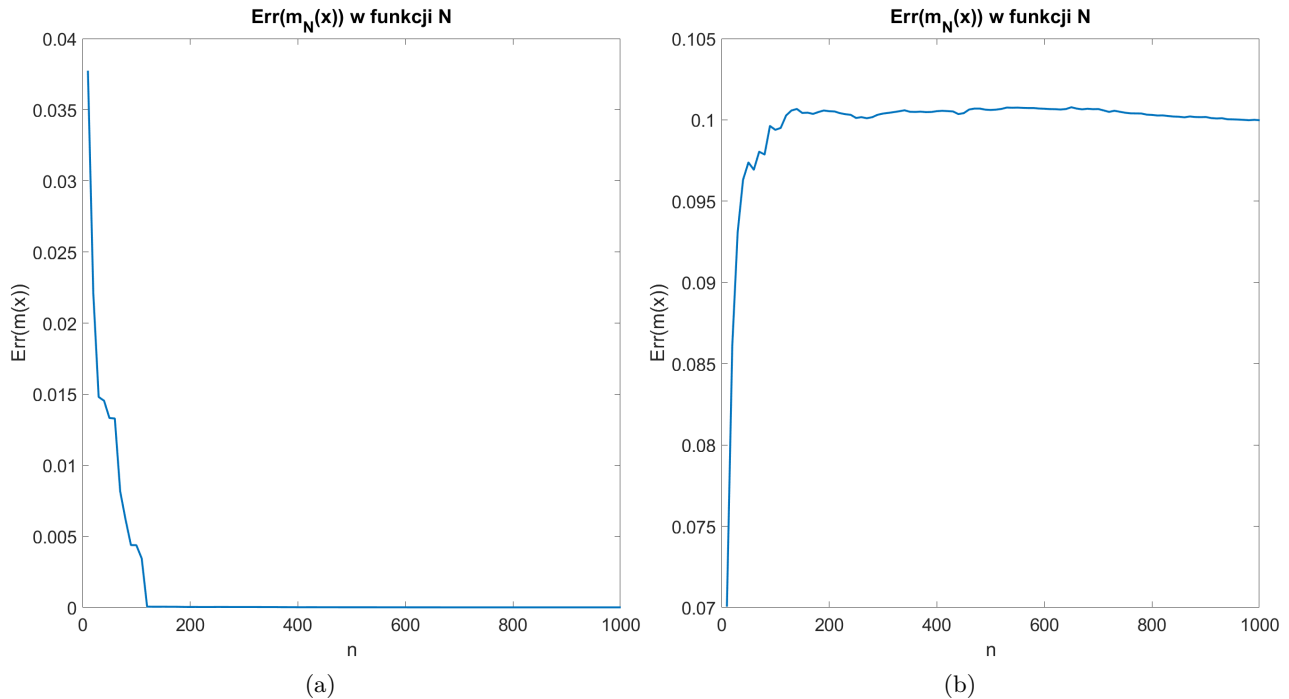
Rysunek 4.11. Błąd empiryczny a) dla  $E\{Z_n\} = 0$  b) dla  $E\{Z_n\} = 1$

## 4.7 Zadanie dodatkowe 2 – Błąd empiryczny w funkcji N

Parametry zadania identyczne jak w poprzednim.

Z rysunku (Rysunek 4.12) dla  $E\{Z_n\} = 0$  wynika, że przy błąd empiryczny bardzo szybko zbiega do zera. Przy niecałych 200 próbkach wartość błędu jest bliska zeru. Oznacza to, że przy niewielkiej liczbie próbek estymowanie funkcji regresji już jest skuteczne.

W przypadku  $E\{Z_n\} = 1$  można zauważyć również szybką zbieżność wartości błędu, lecz tym razem do wartości 0.1. Dzieje się tak, ponieważ zakłócenie posiada wartość oczekiwaną równą jeden. W wyniku tego estymowana funkcja regresji będzie obciążona błędem systematycznym.



Rysunek 4.12. Błąd empiryczny a) dla  $E\{Z_n\} = 0$  b) dla  $E\{Z_n\} = 1$