

# Eksploracja danych

## Sprawozdanie nr 2 z ćwiczeń laboratoryjnych

Tomasz Szkaradek  
Mateusz Niepokój  
Mariusz Biegański

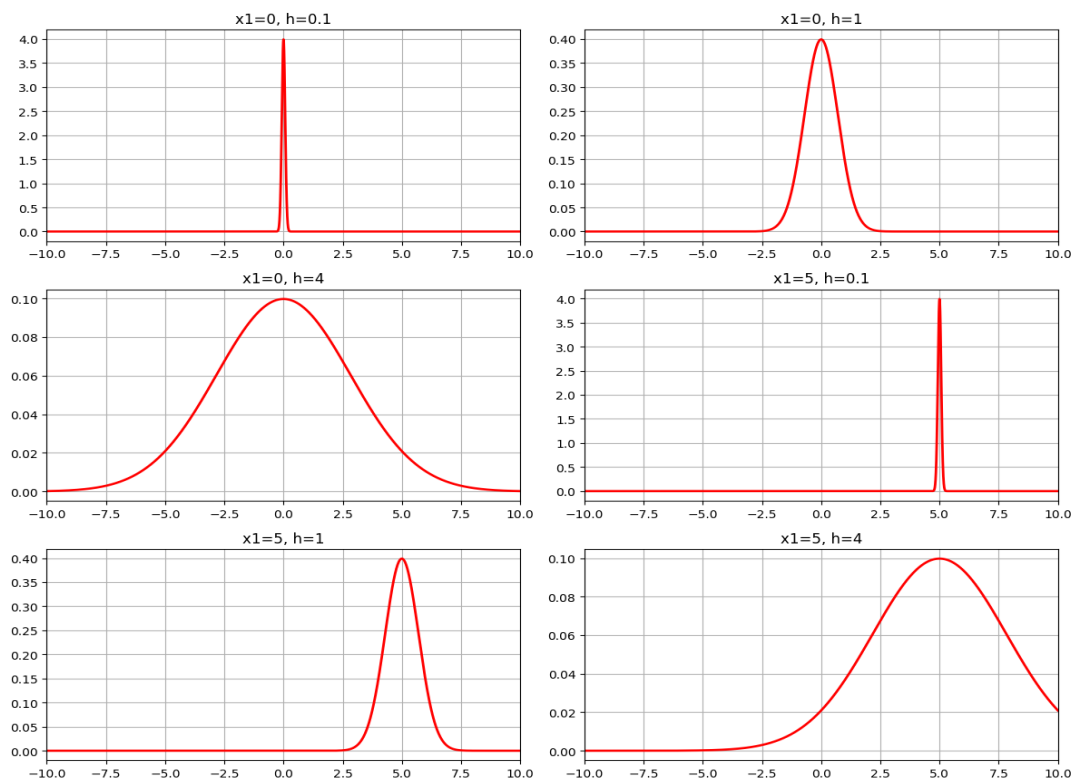
Wydział Fizyki i Informatyki Stosowanej

Akademia Górniczo-Hutnicza im. Stanisława Staszica w Krakowie

20 listopada 2023

### Zadanie 1

W pierwszym zadaniu skupiono się na analizie pojedynczej funkcji przypominającej jądro w estymatorze gęstości jądrowej. Sprawdzono, jaki kształt przyjmuje w zależności od parametrów. Dane wygenerowano i wyświetlono przy użyciu języka Python oraz biblioteki Matplotlib.

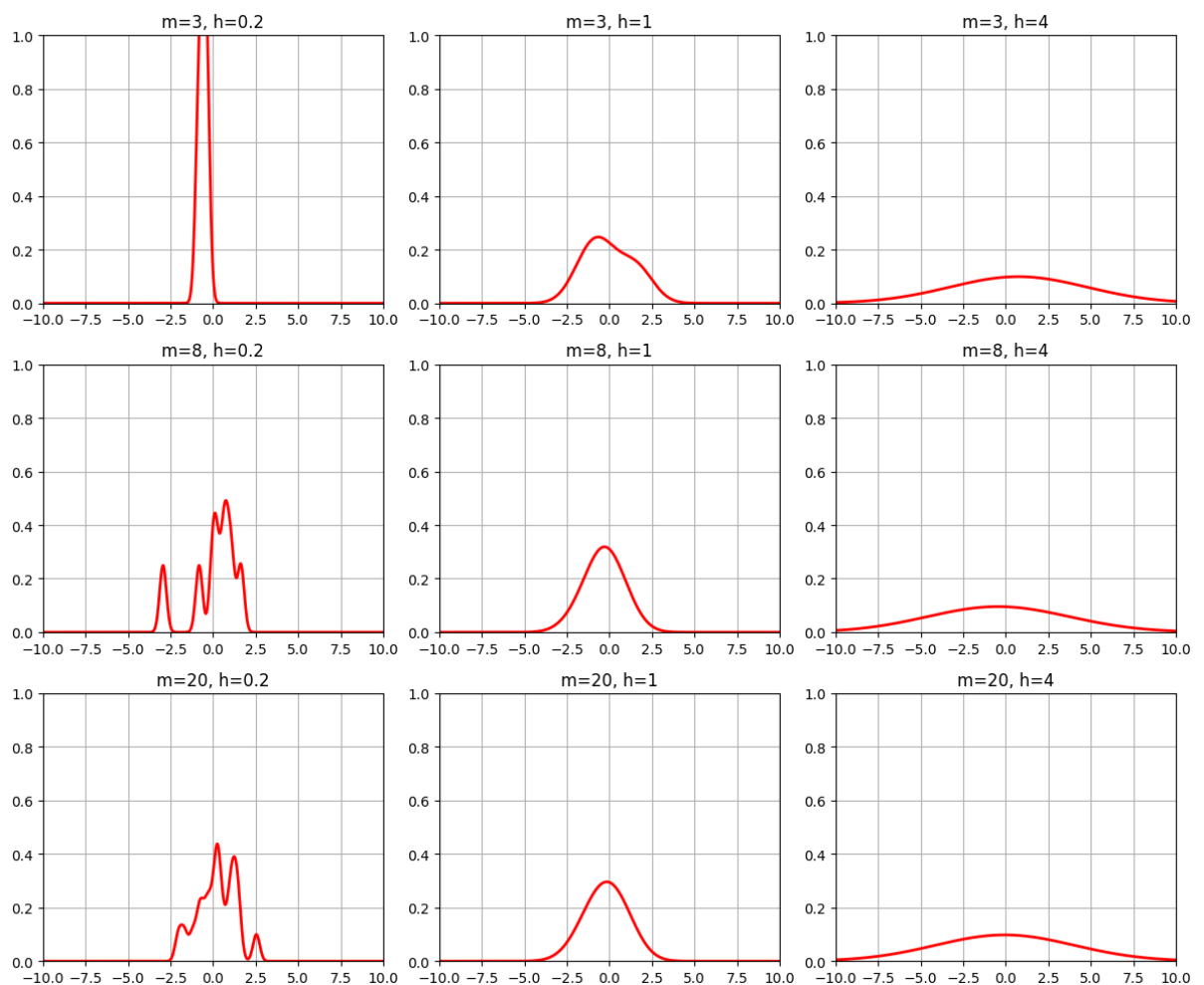


Rys. 1. Uzyskane wykresy w zależności od przyjętych parametrów  $x_1$  oraz " $h$ ".

Parametr "h" określa odchylenie standardowe rozkładu gaussowskiego. Im większa wartość parametru "h", tym szerszy jest rozkład gaussowski. Dlatego też wraz ze wzrostem wartości parametru "h", wykres uzyskany za pomocą funkcji rozkładu gaussowskiego będzie bardziej wygładzony.

## Zadanie 2

W zadaniu drugim zastąpiliśmy pojedynczą wartość  $x_i$  tablicą liczb z wcześniej przyjętego przedziału. Te wartości uzyskaliśmy w sposób losowy, zgodnie z rozkładem normalnym. Następnie sprawdziliśmy, w jaki sposób liczba wylosowanych  $x_i$  oraz parametr "h" wpływają na uzyskane wykresy. Poniższe wykresy zostały wygenerowane wielokrotnie dla różnych ziaren.



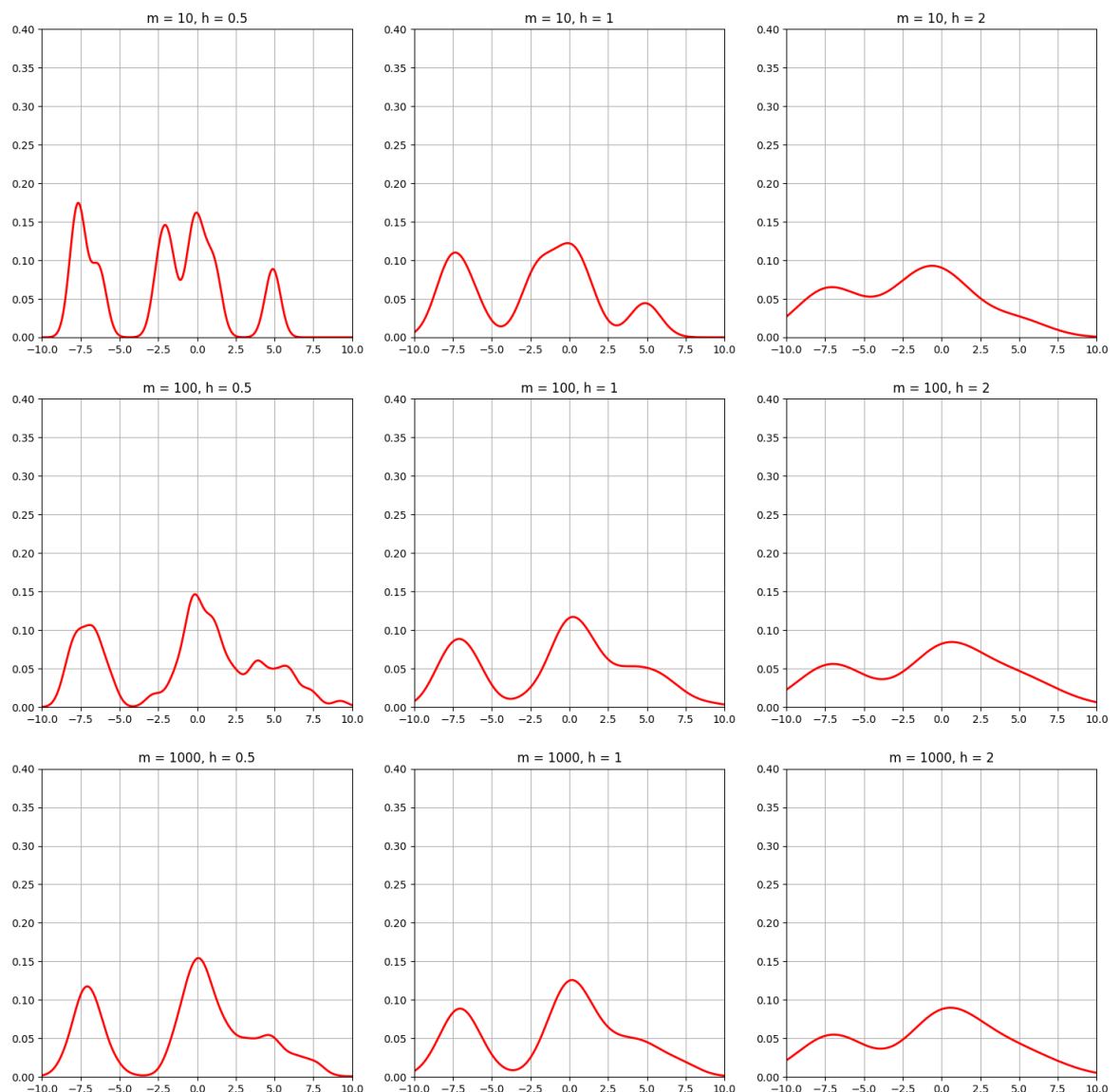
Rys. 2. Przykładowe uzyskane rezultaty w zależności od liczby wylosowanych wartości "m" oraz parametru "h".

Ciekawym przypadkiem wydaje się być wykres dla wartości  $m=3$  i  $h=0.2$ . Można zauważyć, że wszystkie punkty środkowe dość dużo się na nim nakładały. Warto

również dodać, że dla większych wartości parametru "m", funkcja staje się coraz bardziej wygładzona.

### Zadanie 3

Zadanie trzecie stanowiło rozwinięcie zadania drugiego. Wartości tablicy  $x_i$ , uzyskane wcześniej z rozkładu standardowego, zastąpiliśmy wartościami uzyskiwanymi z trzech rozkładów normalnych. Zgodnie z następującą charakterystyką: 30% z rozkładu  $N(-7, 1)$ , 40% z rozkładu  $(0, 1)$ , a pozostałe elementy pochodziły z rozkładu  $N(4, 2)$ . Dla powyższej receptury sprawdziliśmy wyniki w zależności od liczby elementów  $x_i$ , czyli "m", oraz parametru "h".

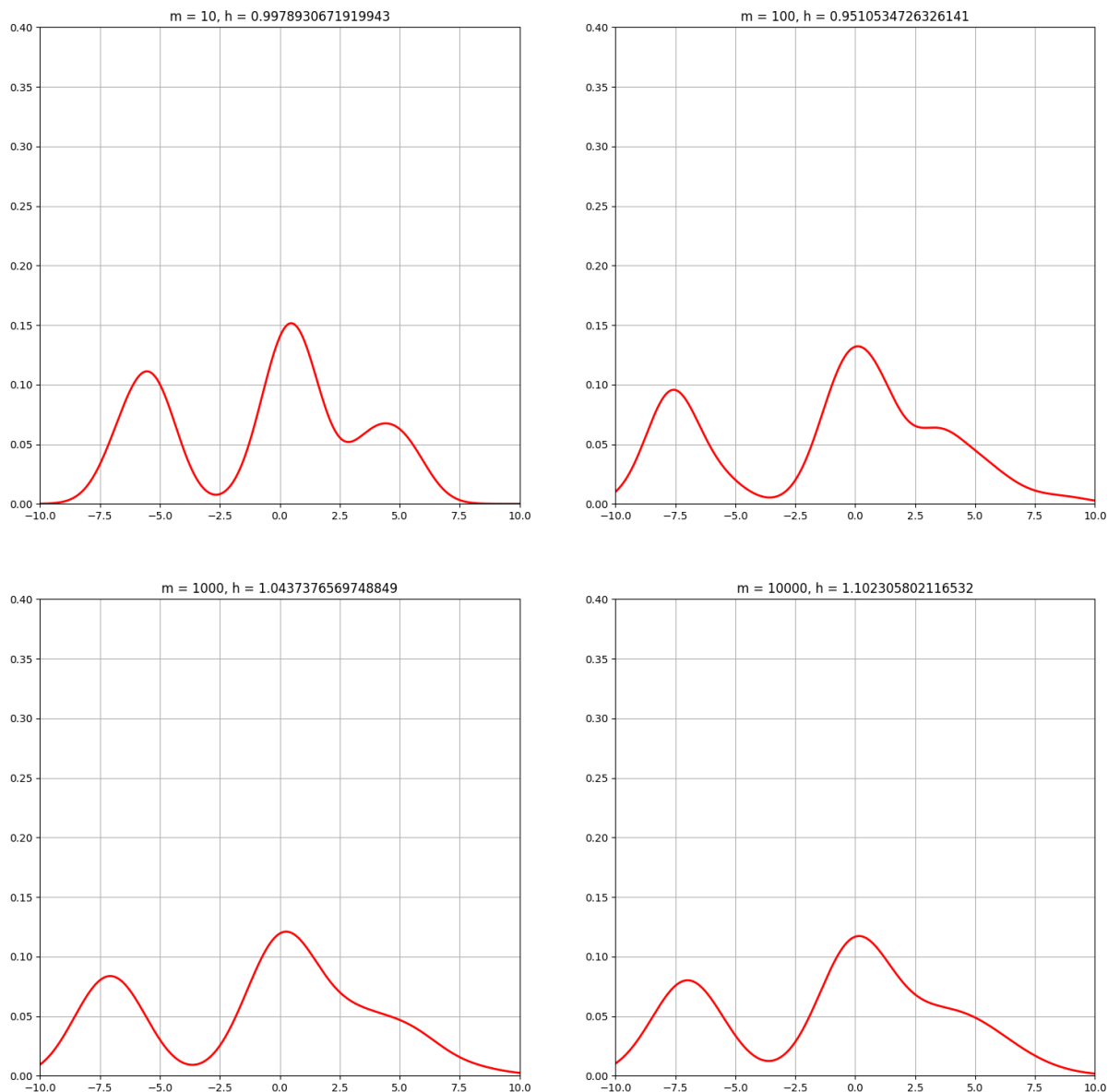


Rys. 3. Wykresy powstałe zgodnie z powyższą recepturą.

Zauważalne są wierzchołki w punktach będących centrami naszego rozkładu, czyli -7, 0 oraz 4. Interesujące wydaje się również to, że dla wartości "m" równego 1000 powyższe wykresy wyglądają tak samo.

#### Zadanie 4

W ostatnim zadaniu użyliśmy rozkładów z zadania 3-go i wyznaczyliśmy wartość parametru wygładzania "h" przy użyciu metody podstawień 2-go rzędu.



Rys. 4. Wykresy powstałe zgodnie z powyższą recepturą.

Na podstawie przedstawionych danych można stwierdzić, że wyniki dla parametru  $h$  są dobre. Wartości  $h$  są bliskie 1, co wskazuje, że rozkład danych jest zbliżony do rozkładu normalnego. Dodatkowo, wartości  $h$  nie są zbyt duże, co wskazuje, że dane nie są zbyt zaśmiecone. Te wyniki sugerują, że dane są dobrze ukształtowane.