



# Data learning

Курс “Машинное обучение”  
Лабораторная работа



## Non-parametric one-dimensional density estimation

Сорокин С.В., М16-524  
Вариант 2-08

2017

# Исходные данные

**Дано:**

Вектор значений признака  $x$ .

**Размер выборки:** 300

**Задача:**

Оценить влияние вида и ширины окна на качество восстановления плотности распределения.

# Исходные данные

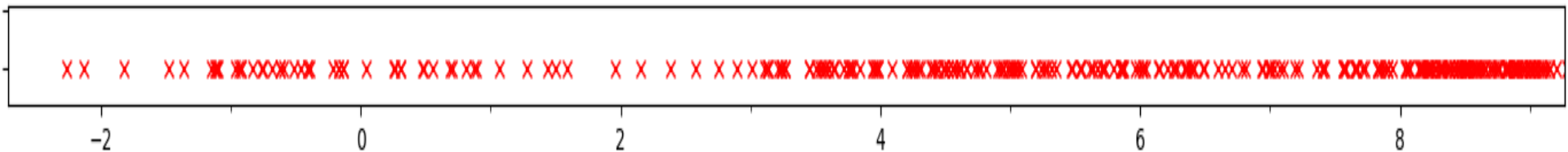


Рисунок 1. Исходное распределение данных

# Используемые методы и формулы

Восстановление плотности распределения

$$\tilde{f}(x) = \frac{1}{n * h} \sum_{i=1}^n K\left(\frac{x - x_i}{h}\right)$$

n – объем выборки

h – размер окна

K – Функция ядра

# Используемые методы и формулы

## Типы ядерных функций $K(u)$

1) Прямоугольное окно:  $K(u) = \begin{cases} \frac{1}{2}, & |u| < 1 \\ 0, & \text{иначе} \end{cases}$

2) Гауссово окно:  $K(u) = \frac{1}{\sqrt{2*\pi}} * e^{-\frac{u^2}{2}}$

3) Окно Епанечникова:  $K(u) = \begin{cases} \frac{3}{4} * (1 - u^2), & |u| < 1 \\ 0, & \text{иначе} \end{cases}$

4) Треугольное окно:  $K(u) = \begin{cases} 1 - |u|, & |u| < 1 \\ 0, & \text{иначе} \end{cases}$

# Используемые методы и формулы

Средняя интегральная ошибка восстановления плотности распределения

$$MISE(\tilde{f}) = \int_{-\infty}^{\infty} \varepsilon^2(x) dx = \int_{-\infty}^{\infty} Var(\tilde{f}(x)) dx + \int_{-\infty}^{\infty} Bias^2(x) dx$$

*Var* — дисперсия

*Bias* — смещение

# Результаты исследований

Задание 1. Восстановление плотности при фиксированной форме окна

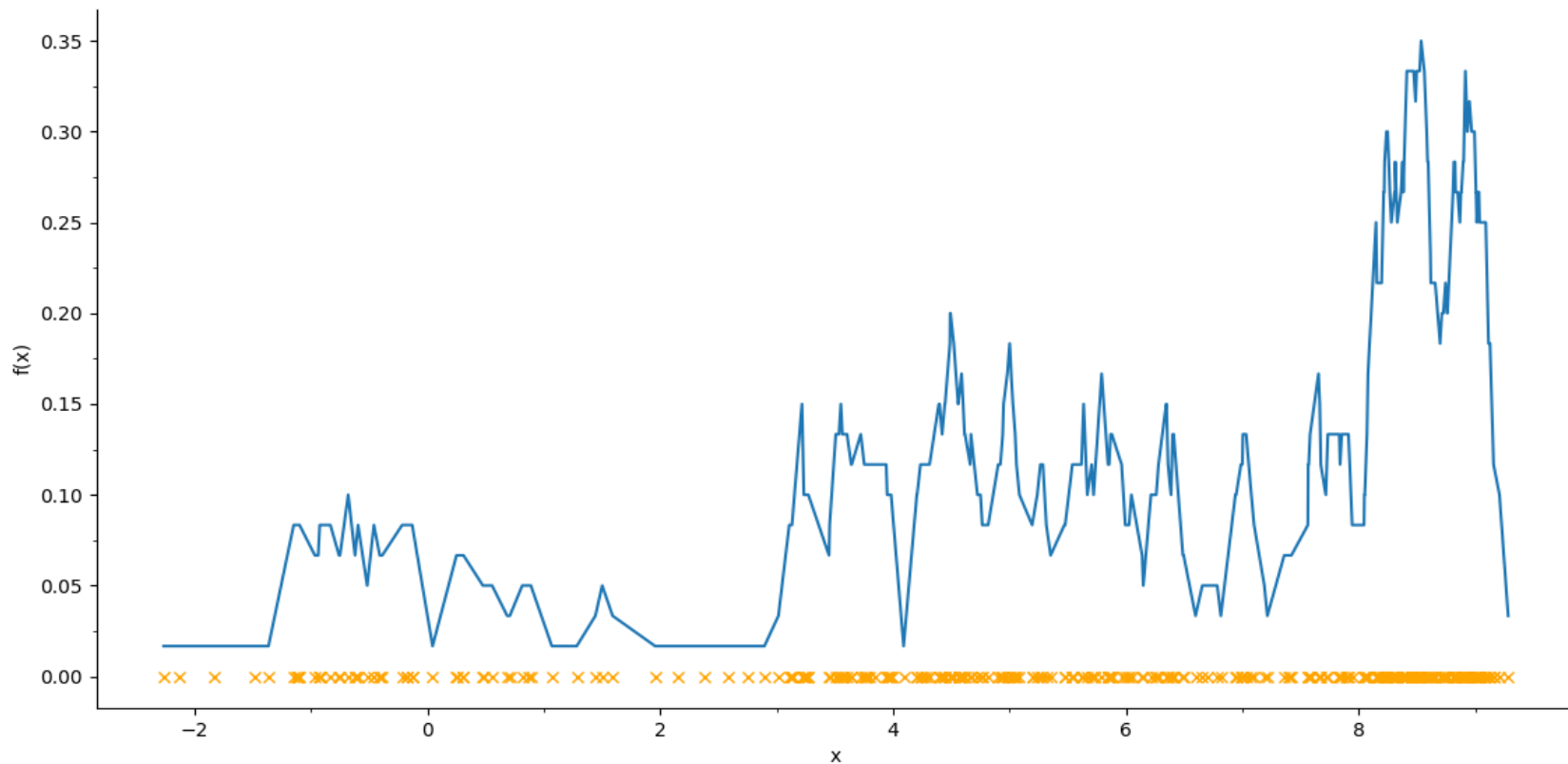


Рисунок 2. Восстановление плотности распределения с прямоугольным окном при маленькой ширине  $h = 0.1$

## Задание 1. Восстановление плотности при фиксированной форме окна

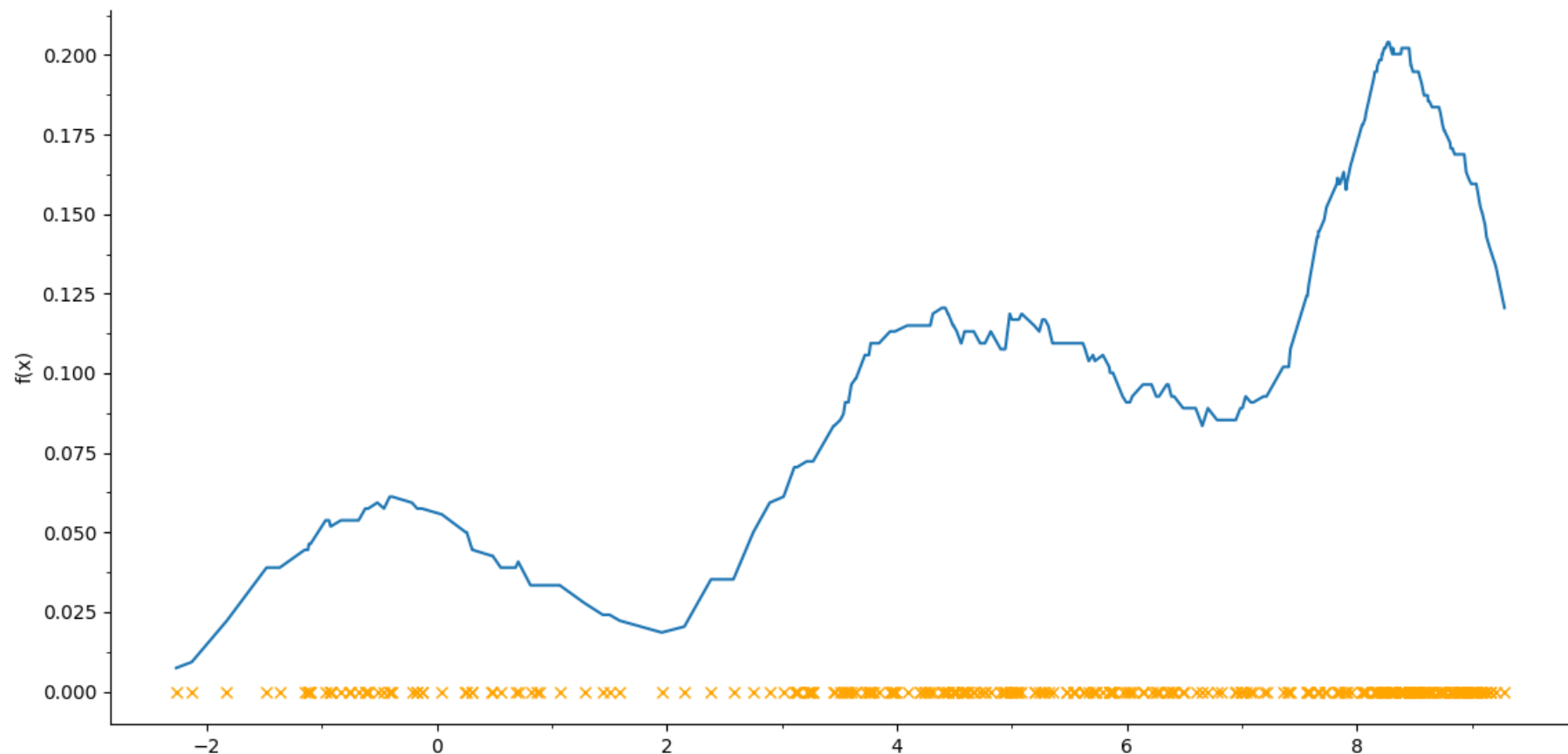


Рисунок 3. Восстановление плотности распределения с прямоугольным окном с шириной, рассчитанной по правилу Сильвермана  $h = 0.90$



## Задание 1. Восстановление плотности при фиксированной форме окна

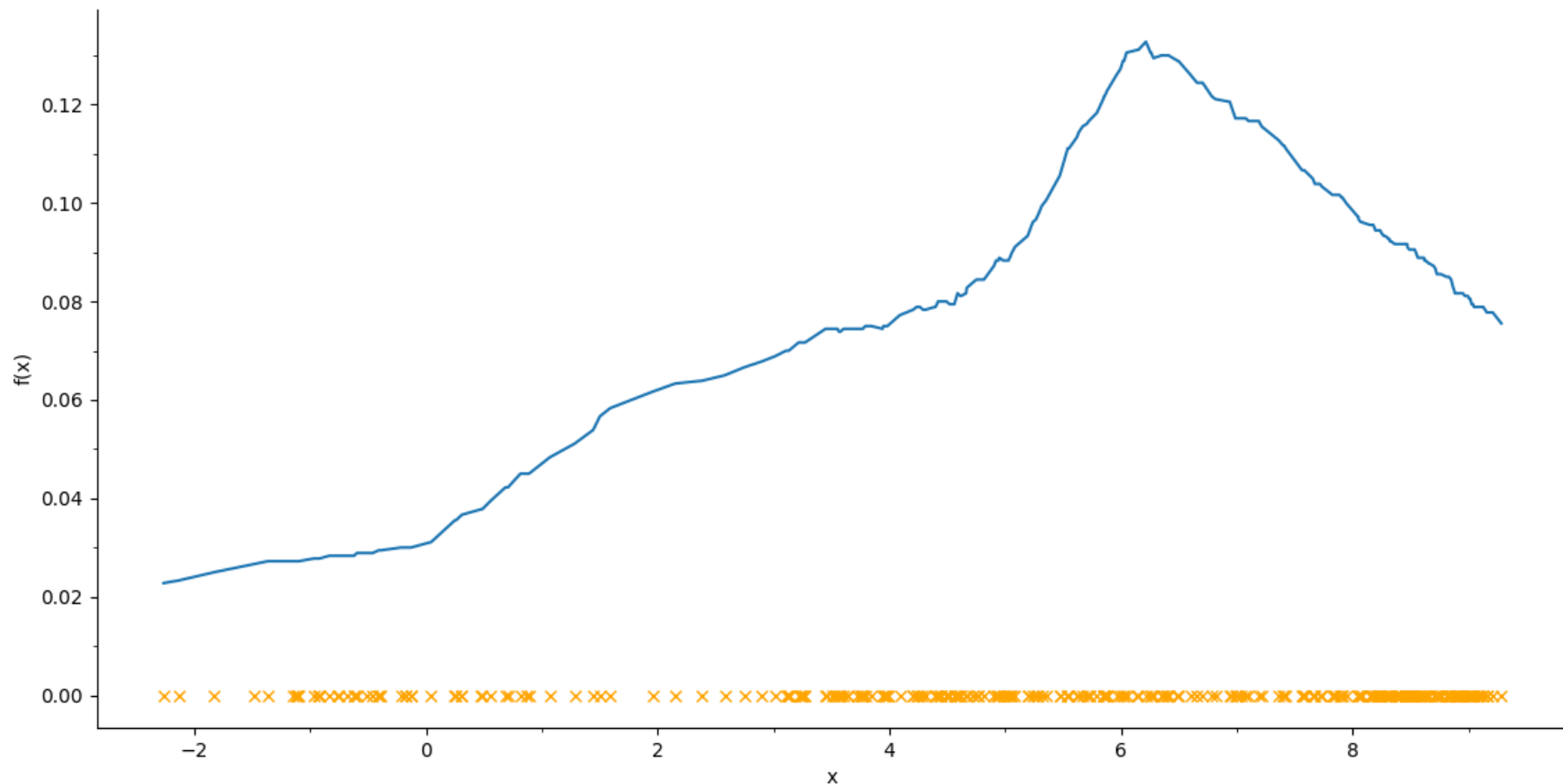


Рисунок 4. Восстановление плотности распределения с прямоугольным окном при большой ширине  $h = 3$

## Задание 2. Восстановление плотности при фиксированной ширине окна

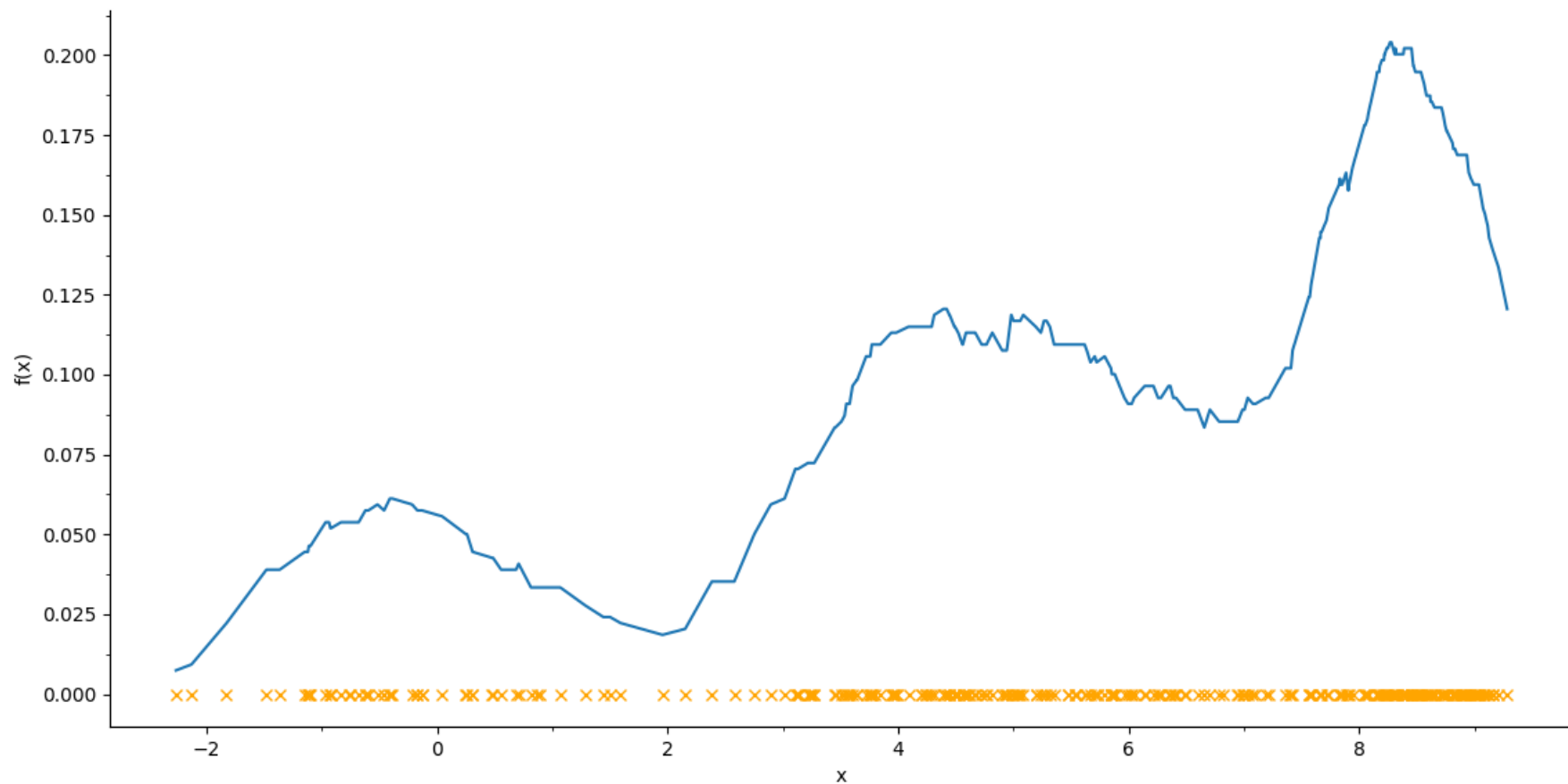


Рисунок 5. Восстановление плотности распределения с прямоугольным окном с шириной  $h = 0.9$

## Задание 2. Восстановление плотности при фиксированной ширине окна

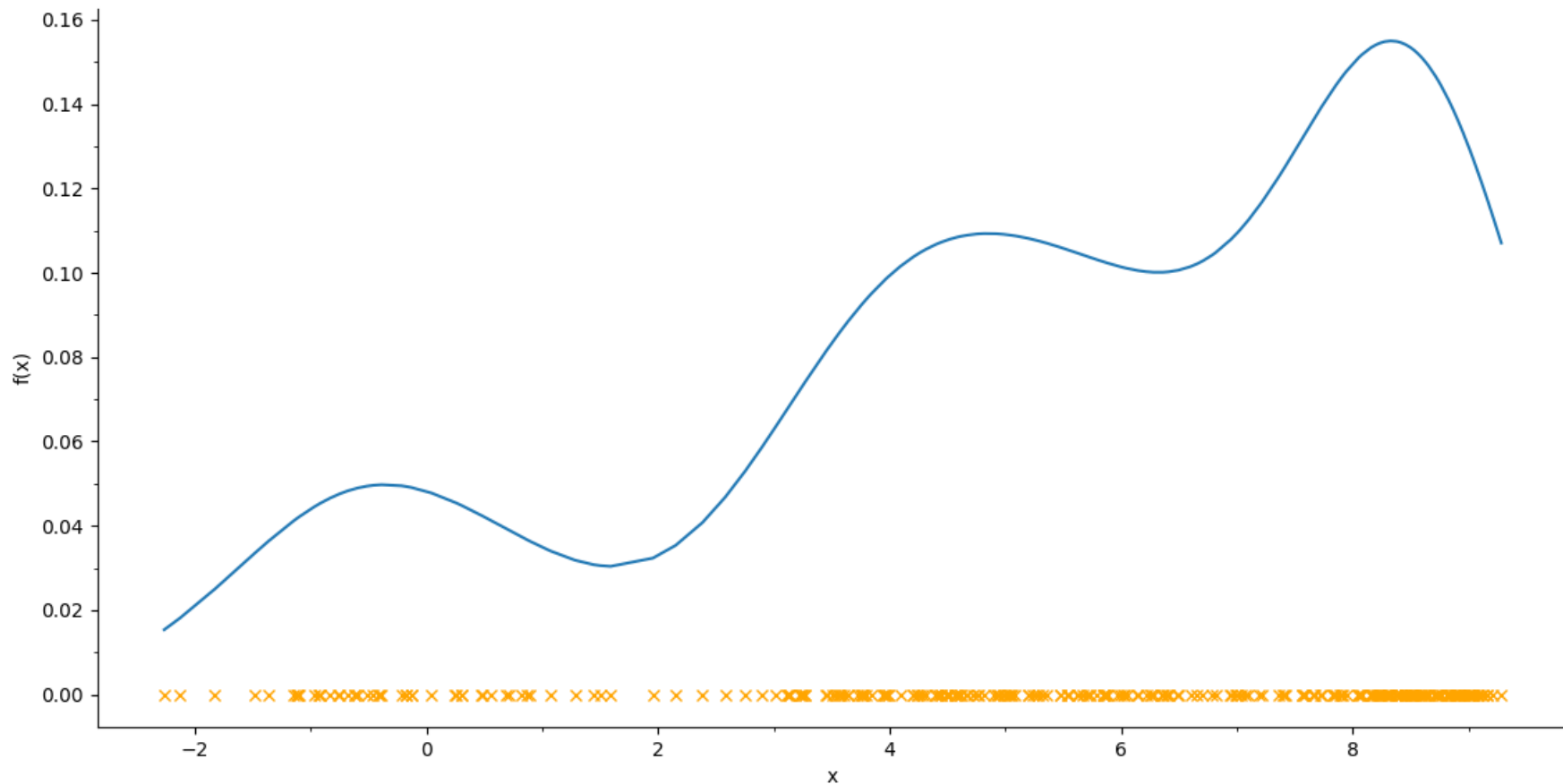


Рисунок 6. Восстановление плотности распределения с гауссовым окном с шириной  $h = 0.9$

## Задание 2. Восстановление плотности при фиксированной ширине окна

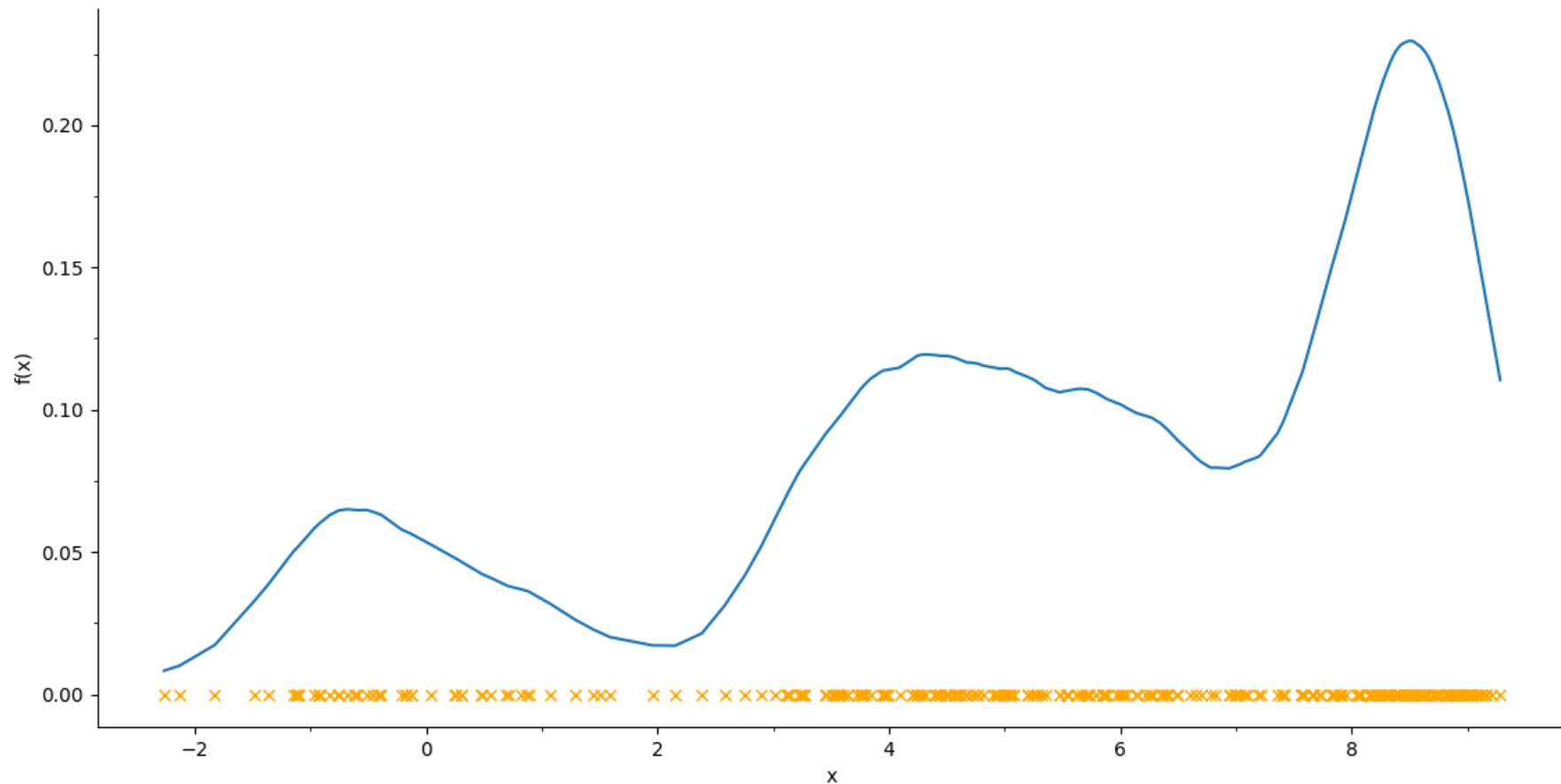


Рисунок 7. Восстановление плотности распределения с окном Епанечникова с шириной  $h = 0.9$

## Задание 2. Восстановление плотности при фиксированной ширине окна

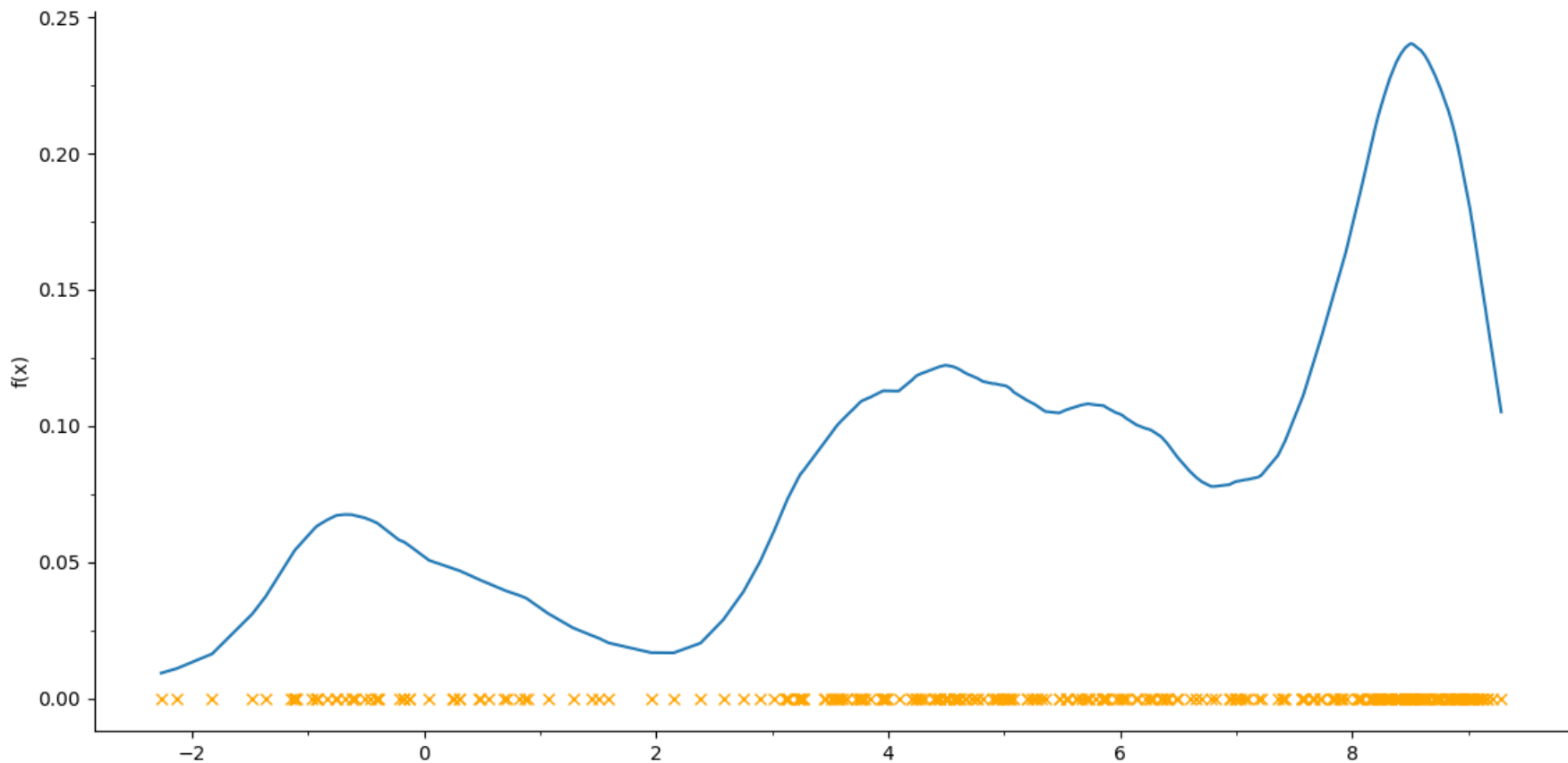


Рисунок 8. Восстановление плотности распределения с треугольным окном с шириной  $h = 0.9$

### Задание 3. Анализ дисперсии восстановленной плотности при фиксированной форме окна

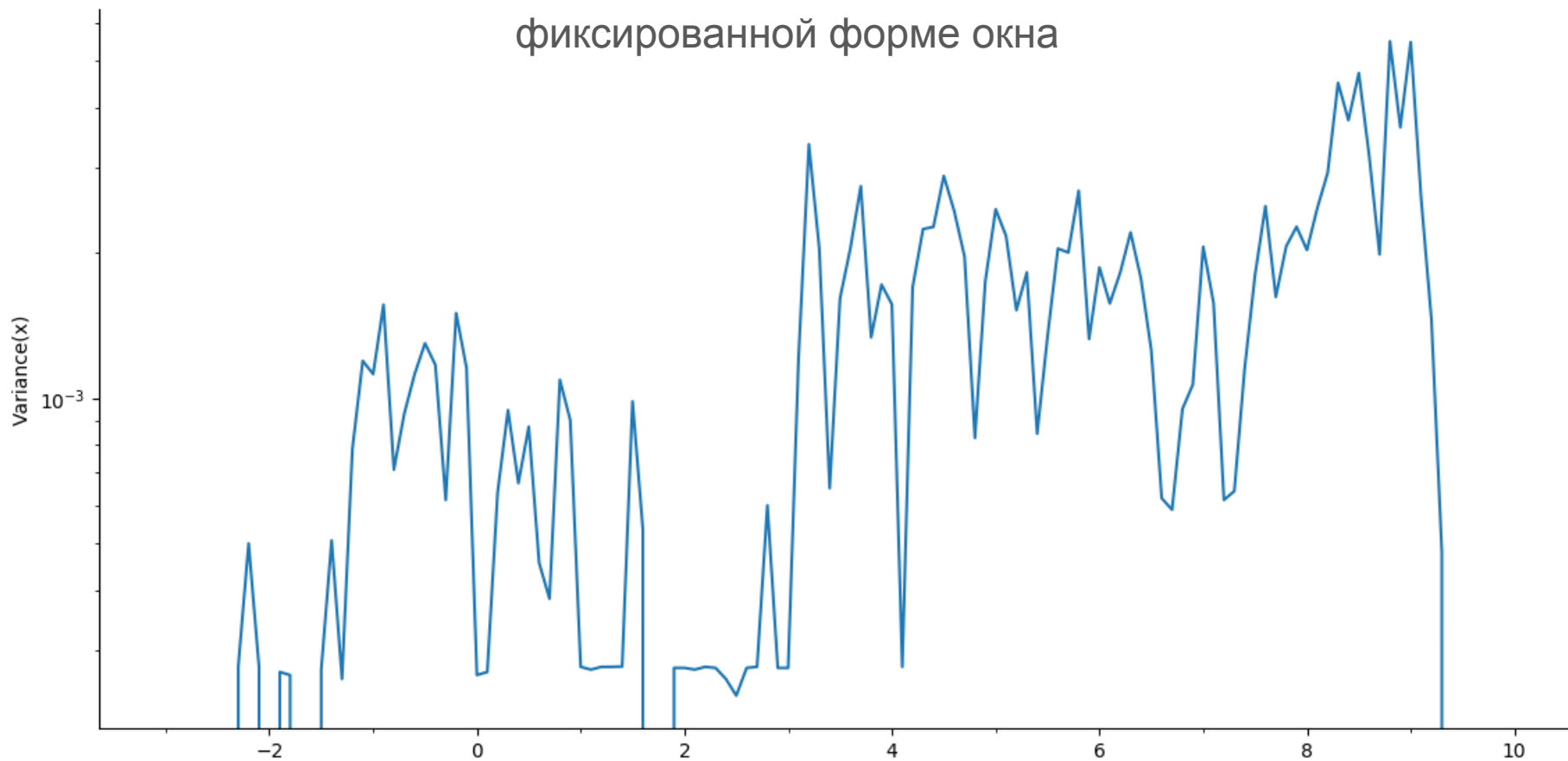


Рисунок 9. График зависимости дисперсий восстановленной плотности от  $x$  для маленькой ширины прямоугольного окна  $h = 0.1$

### Задание 3. Анализ дисперсии восстановленной плотности при фиксированной форме окна

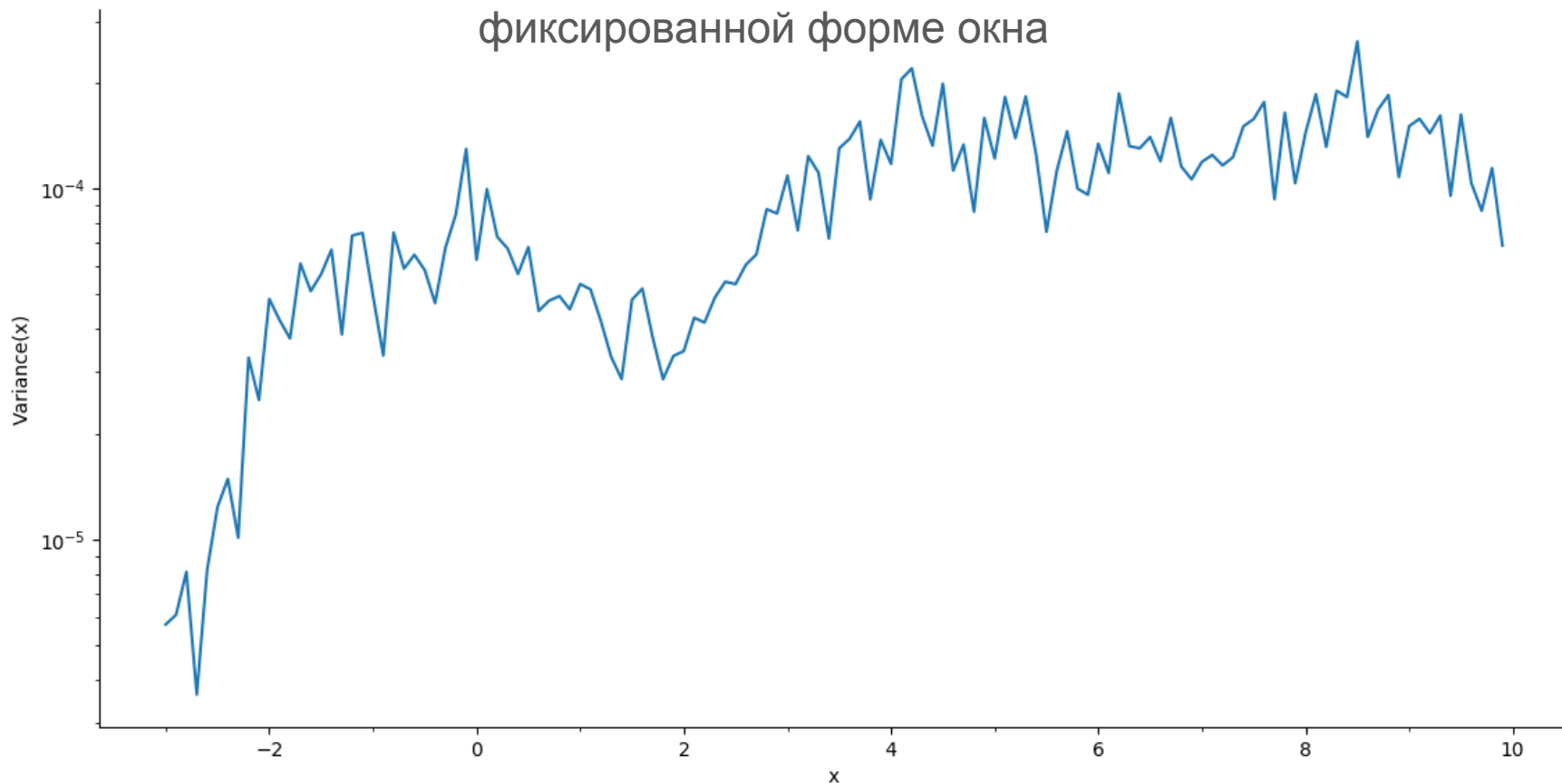


Рисунок 10. График зависимости дисперсии восстановленной плотности от  $x$  для ширины прямоугольного окна, рассчитанной по правилу Сильвермана  $h = 0.9$

### Задание 3. Анализ дисперсии восстановленной плотности при фиксированной форме окна

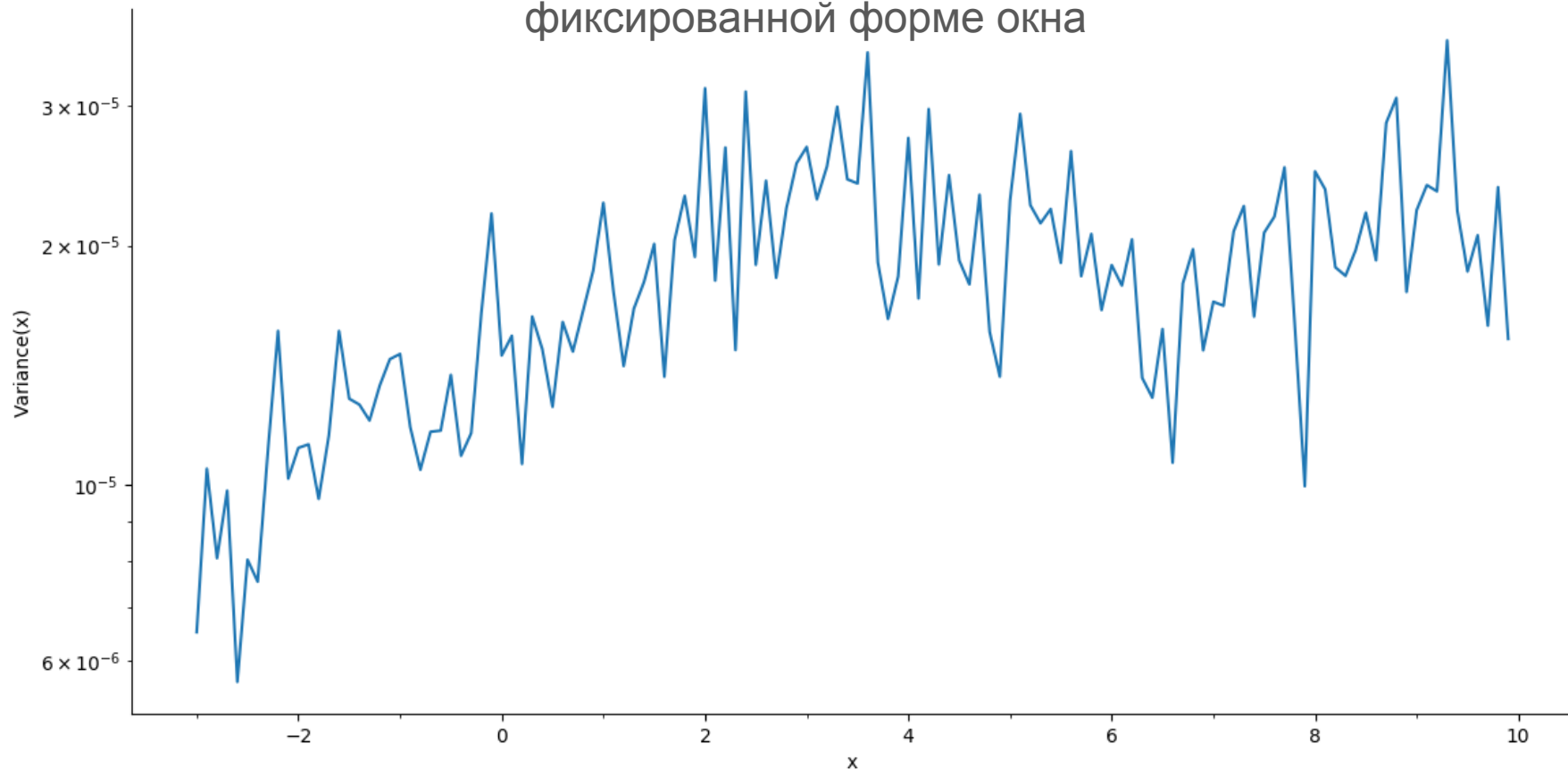


Рисунок 11. График зависимости дисперсии восстановленной плотности от  $x$  для большой ширины прямоугольного окна  $h = 3$



### Задание 3. Анализ смещения восстановленной плотности при фиксированной форме окна

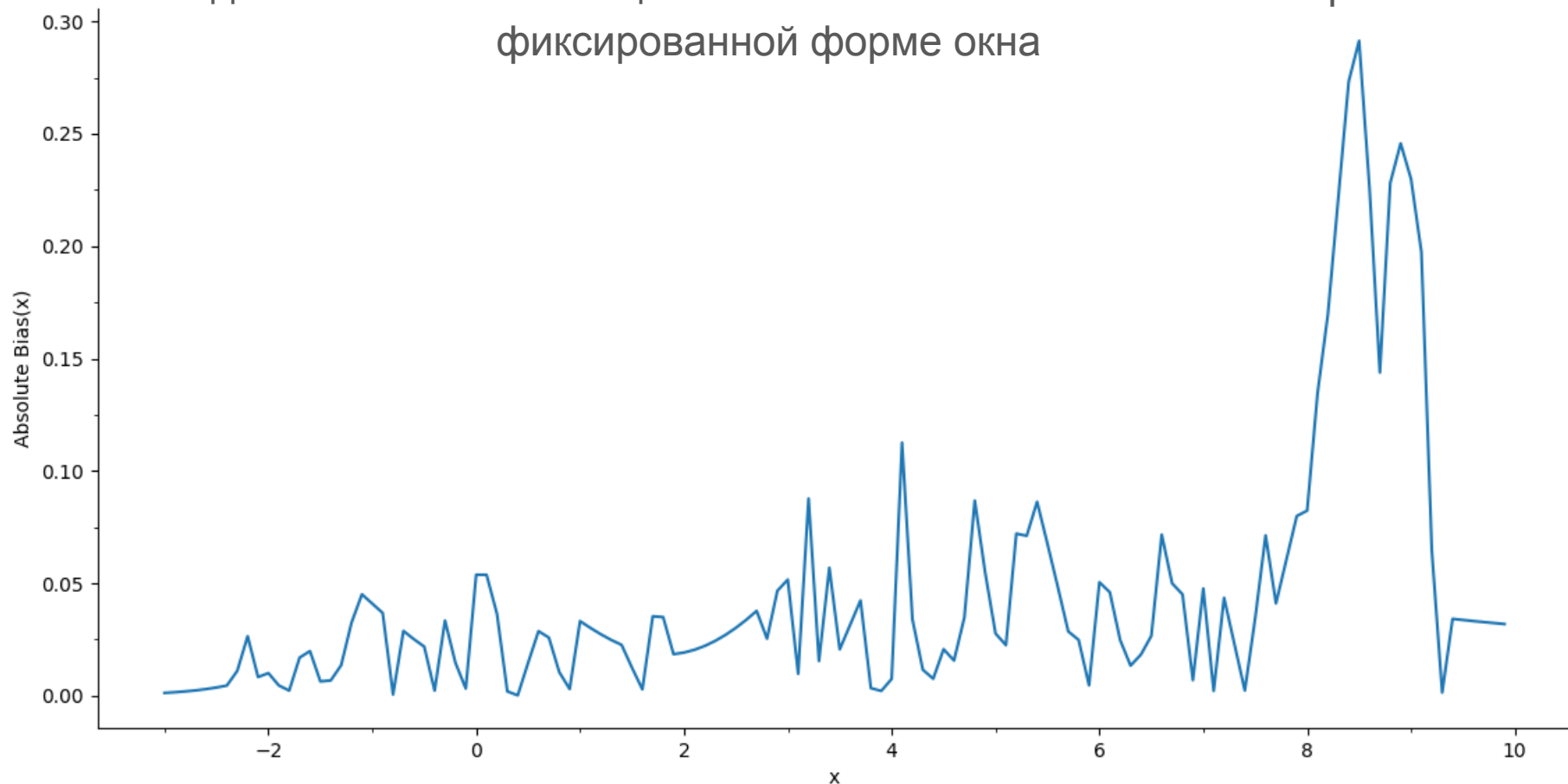


Рисунок 12. График зависимости смещения восстановленной плотности от  $x$  для маленькой ширины прямоугольного окна  $h = 0.1$

### Задание 3. Анализ смещения восстановленной плотности при фиксированной форме окна

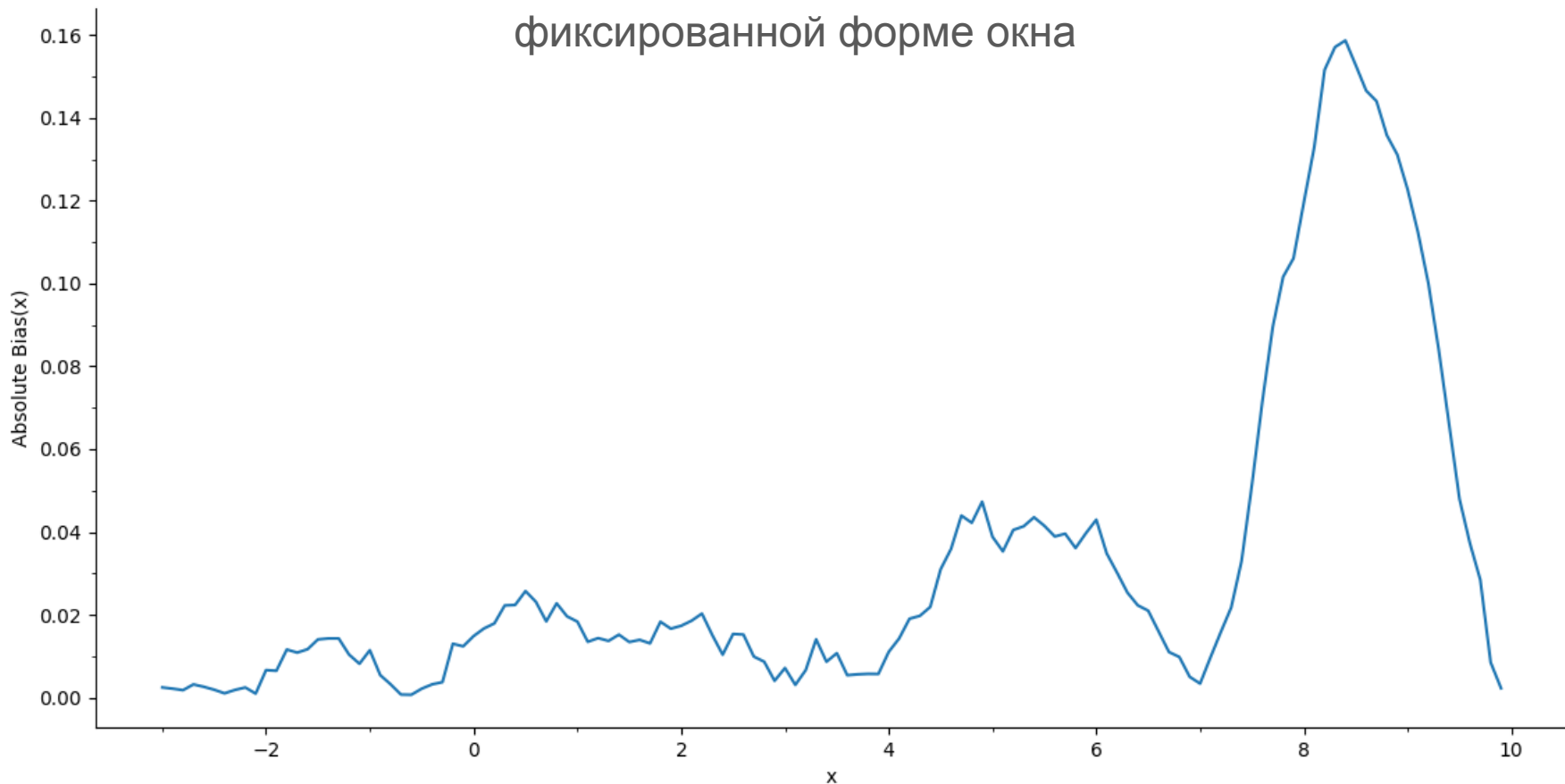


Рисунок 13. График зависимости смещения восстановленной плотности от  $x$  для ширины прямоугольного окна, рассчитанной по правилу Сильвермана  $h = 0.9$

### Задание 3. Анализ смещения восстановленной плотности при фиксированной форме окна

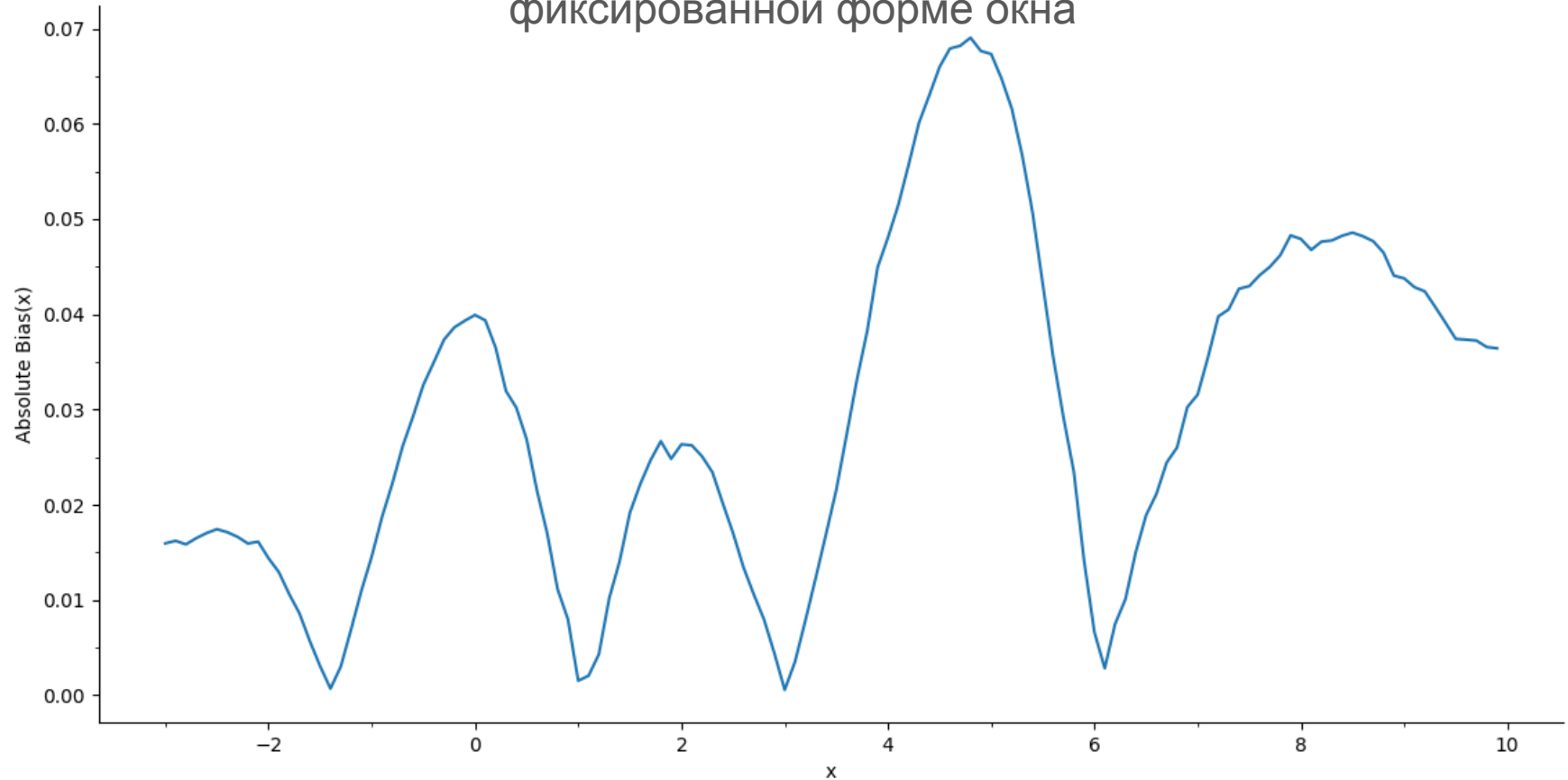


Рисунок 14. График зависимости смещения восстановленной плотности от  $x$  для большой ширины прямоугольного окна  $h = 3$

#### Задание 4. Анализ дисперсии восстановленной плотности при фиксированном значении $x$

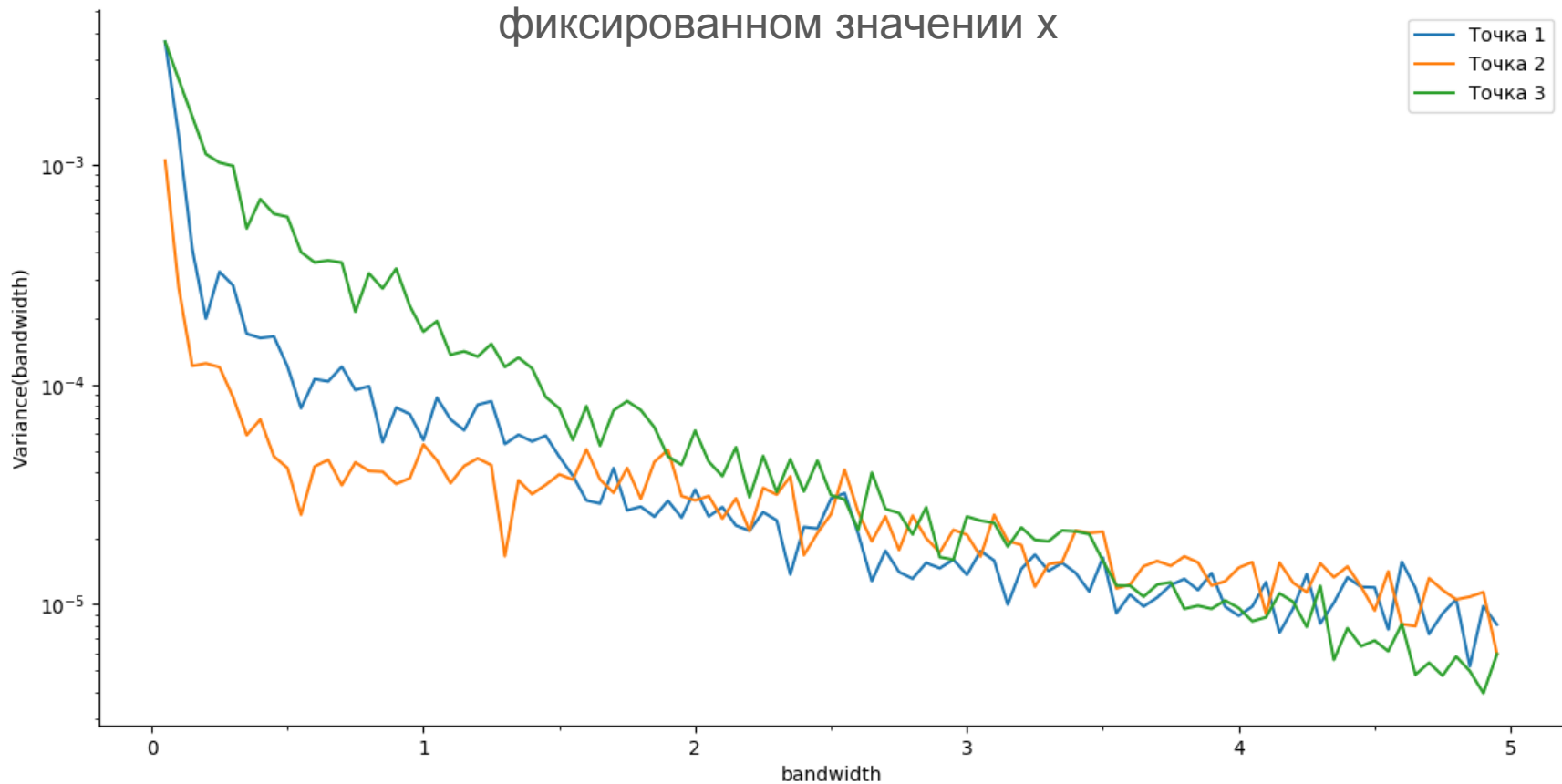


Рисунок 15. График зависимости дисперсии восстановленной плотности от ширины прямоугольного окна для трех фиксированных точек

#### Задание 4. Анализ дисперсии восстановленной плотности при фиксированном значении $x$

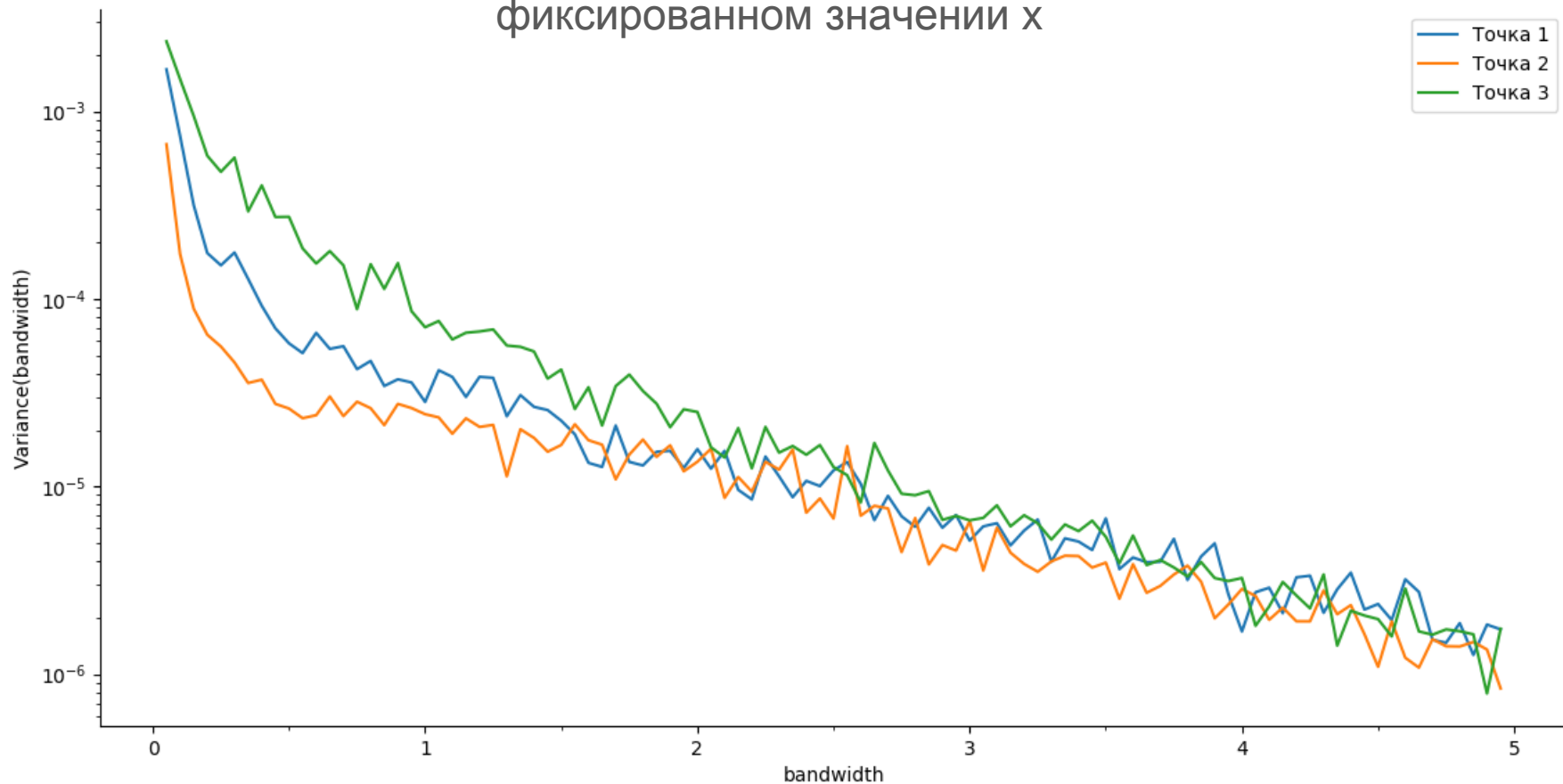


Рисунок 16. График зависимости дисперсии восстановленной плотности от ширины гауссова окна для трех фиксированных точек

#### Задание 4. Анализ дисперсии восстановленной плотности при фиксированном значении $x$

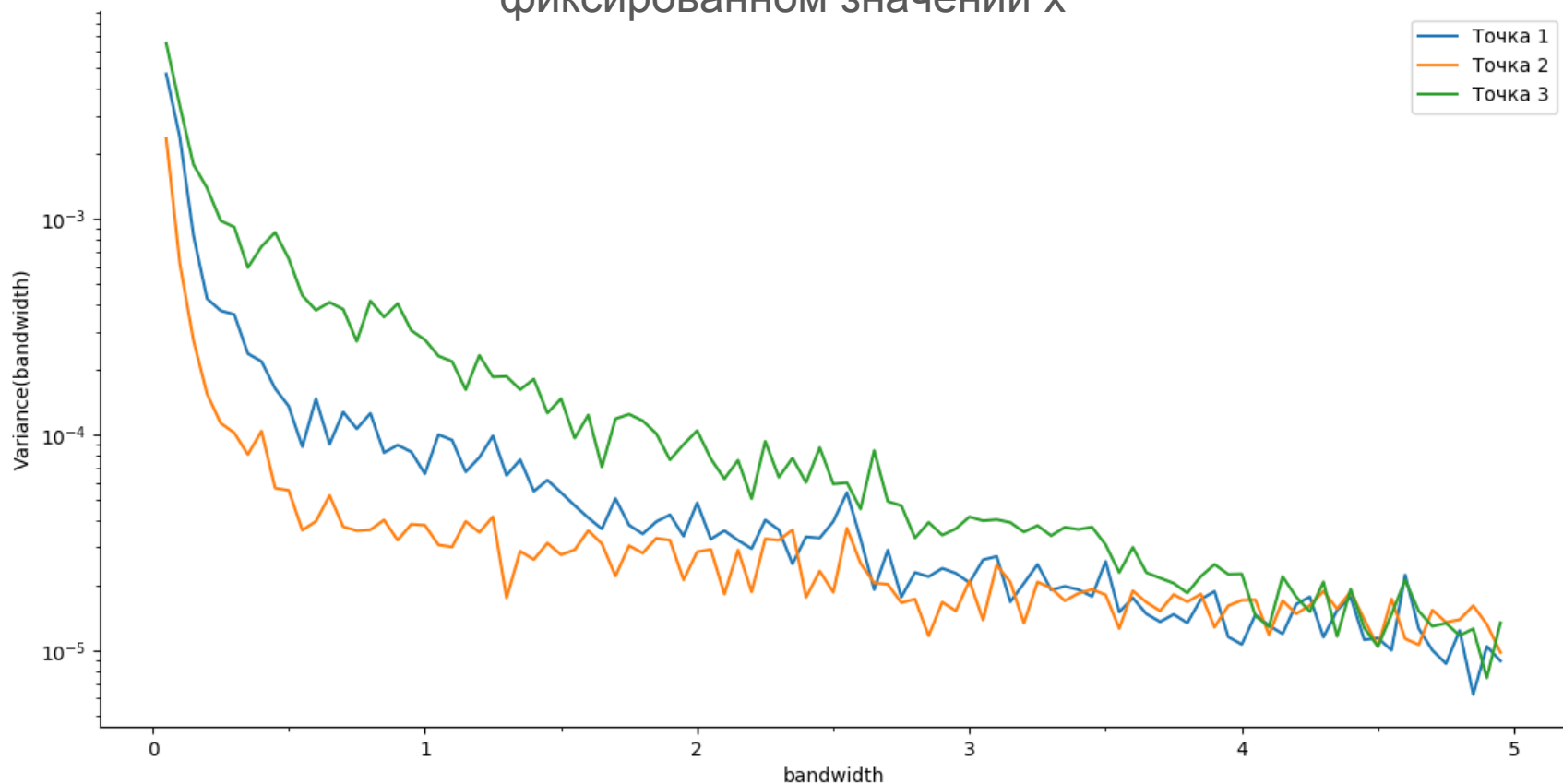


Рисунок 17. График зависимости дисперсии восстановленной плотности от ширины окна Епанечникова для трех фиксированных точек

# Задание 4. Анализ дисперсии восстановленной плотности при фиксированном значении $x$

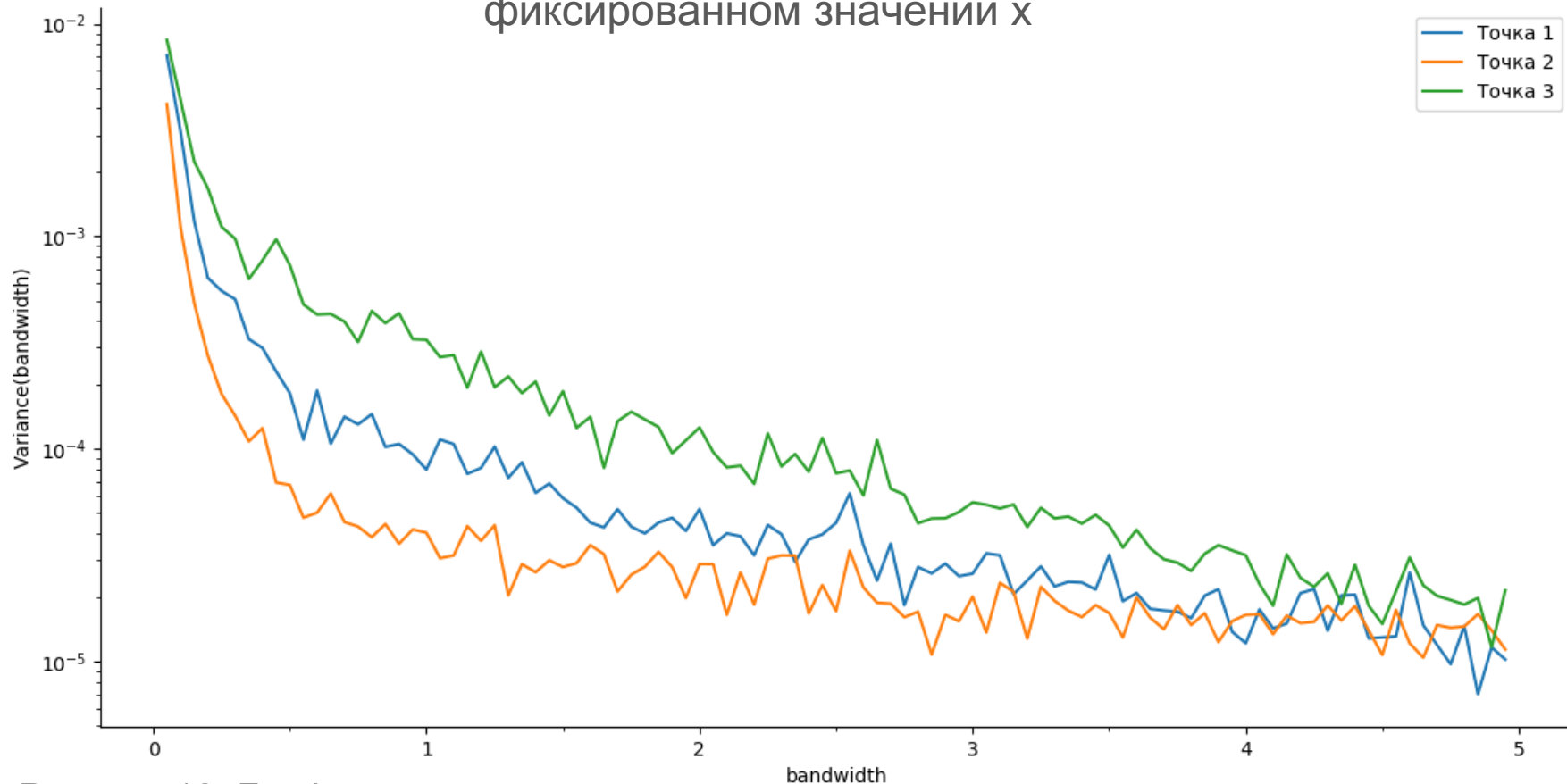


Рисунок 18. График зависимости дисперсии восстановленной плотности от ширины треугольного окна для трех фиксированных точек

## Задание 4. Анализ дисперсии восстановленной плотности при фиксированном значении $x$

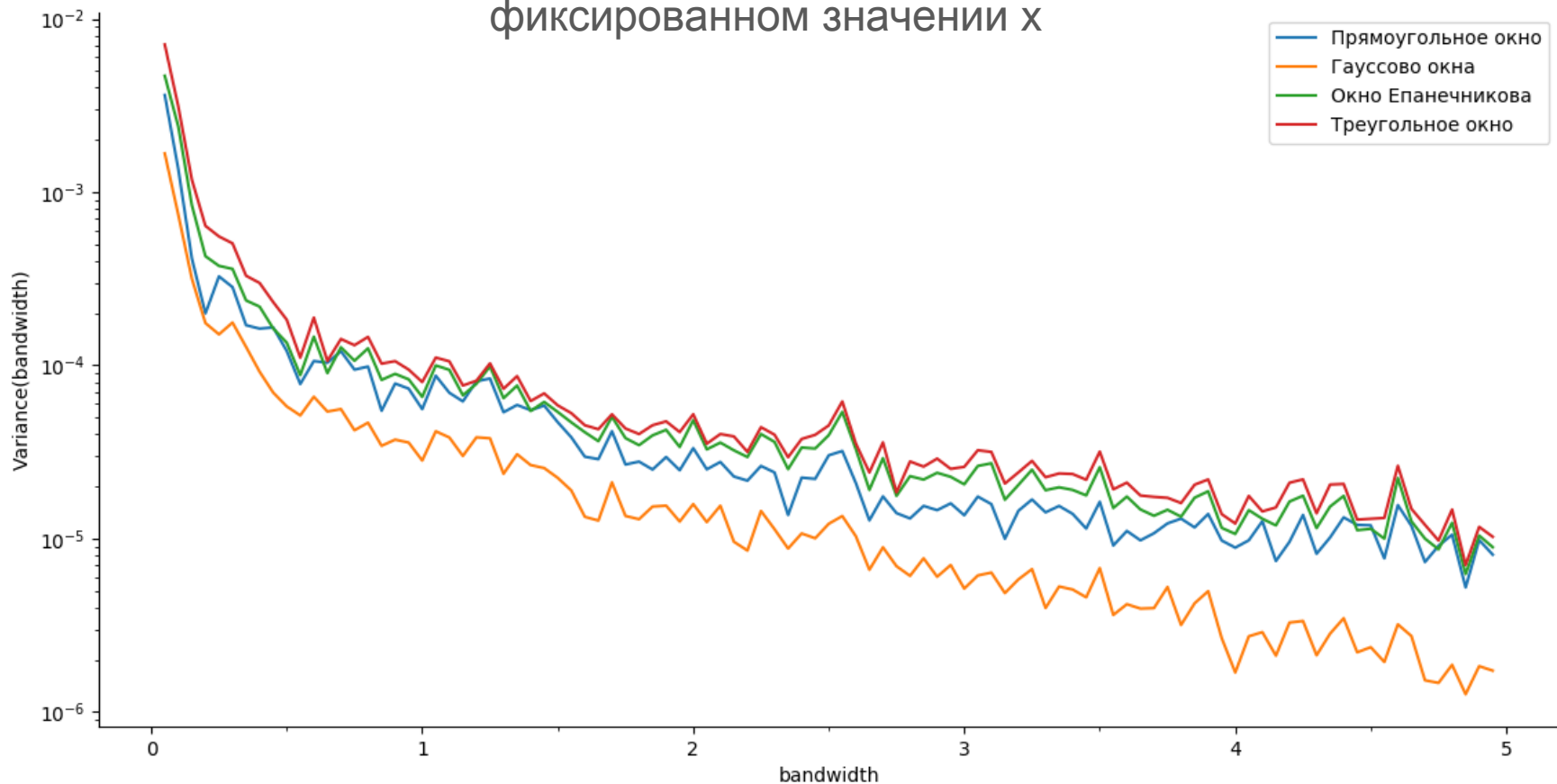


Рисунок 19. График зависимости дисперсии восстановленной плотности от ширины окна для первой фиксированной точки



## Задание 4. Анализ дисперсии восстановленной плотности при фиксированном значении $x$

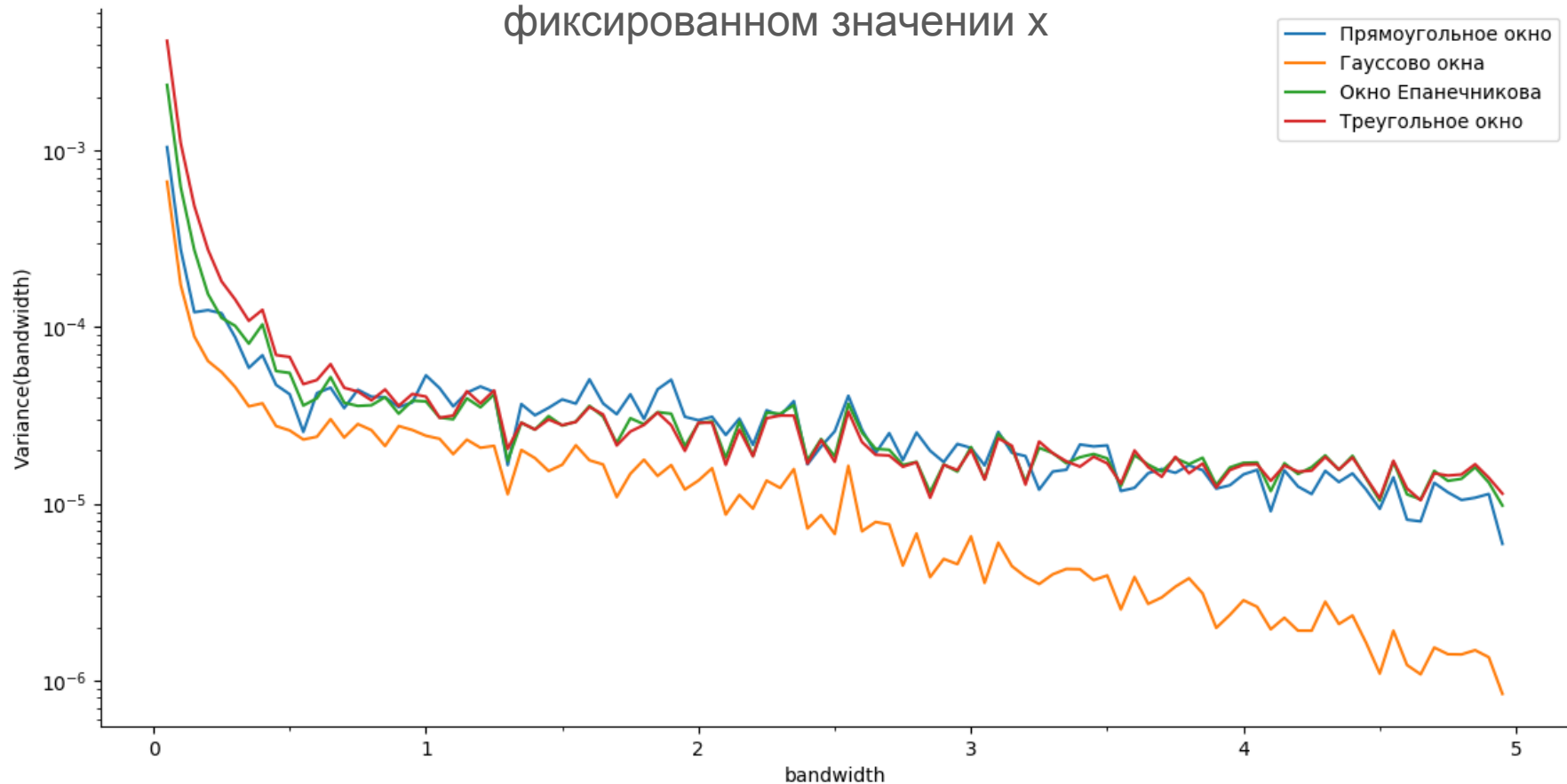


Рисунок 20. График зависимости дисперсии восстановленной плотности от ширины окна для второй фиксированной точки

# Задание 4. Анализ дисперсии восстановленной плотности при фиксированном значении $x$

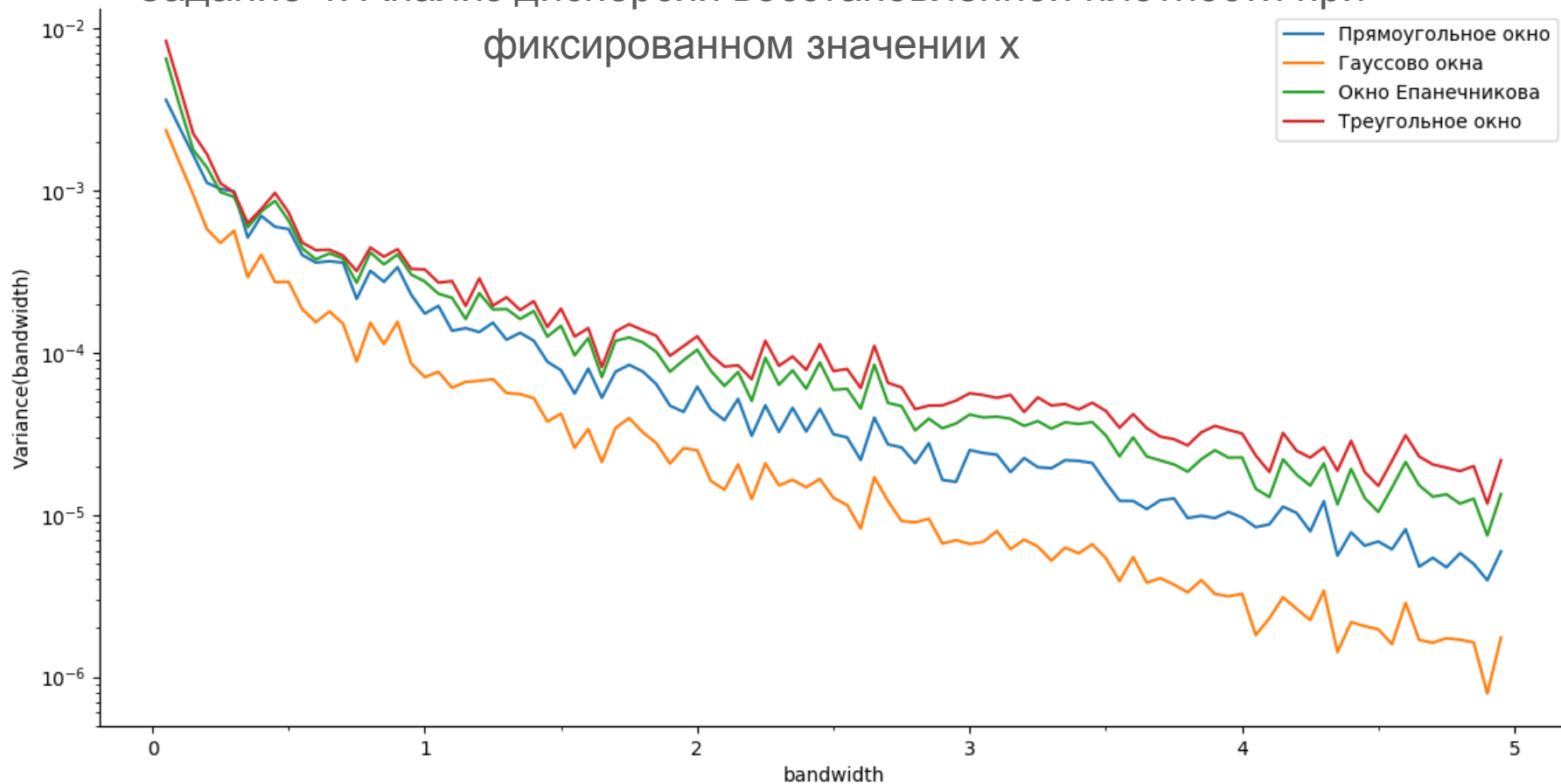


Рисунок 20. График зависимости дисперсии восстановленной плотности от ширины окна для третьей фиксированной точки

# Задание 4. Анализ смещения восстановленной плотности при фиксированном значении $x$

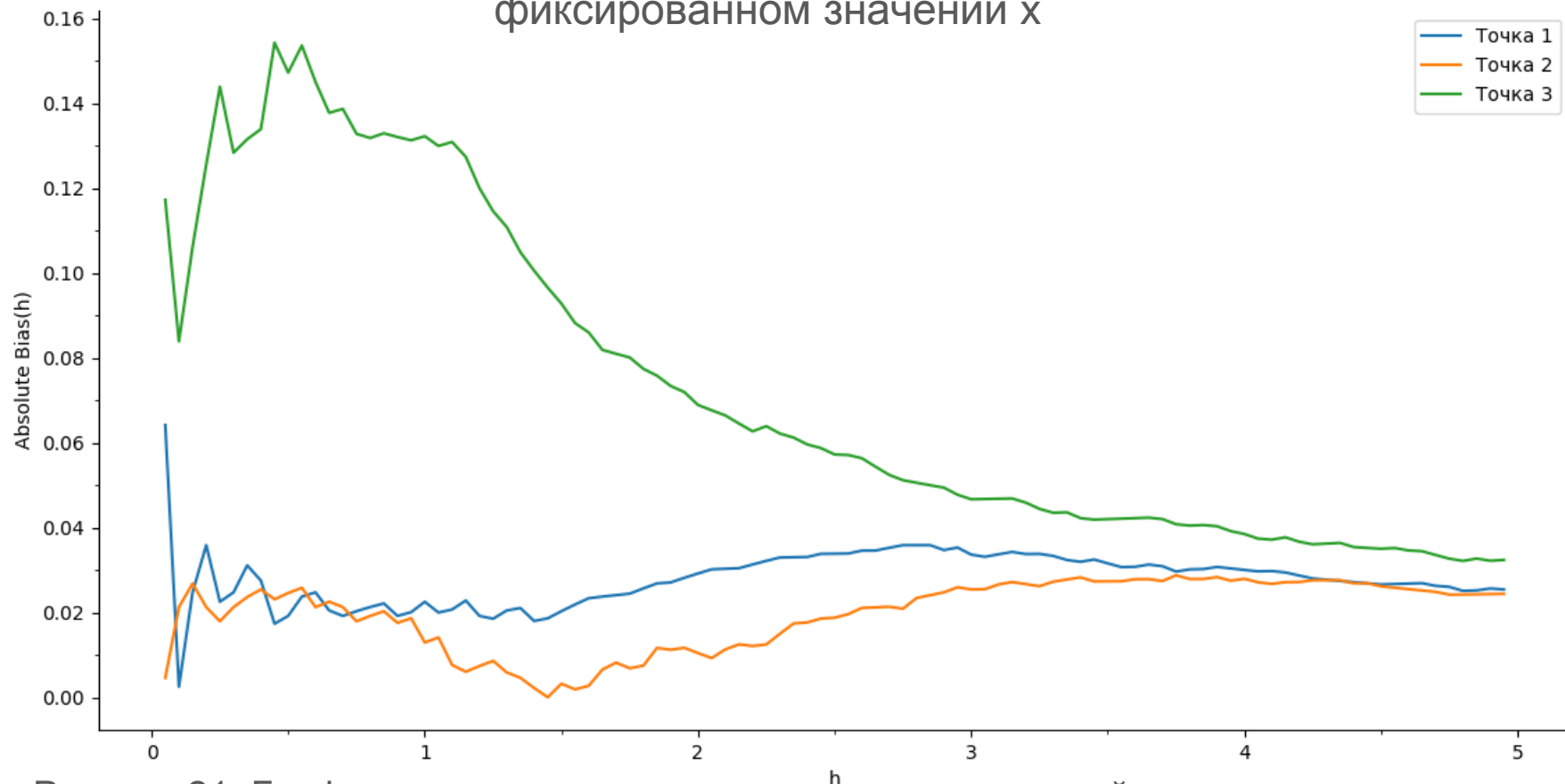


Рисунок 21. График зависимости смещения восстановленной плотности от ширины прямоугольного окна для трех фиксированных точек

#### Задание 4. Анализ смещения восстановленной плотности при фиксированном значении $x$

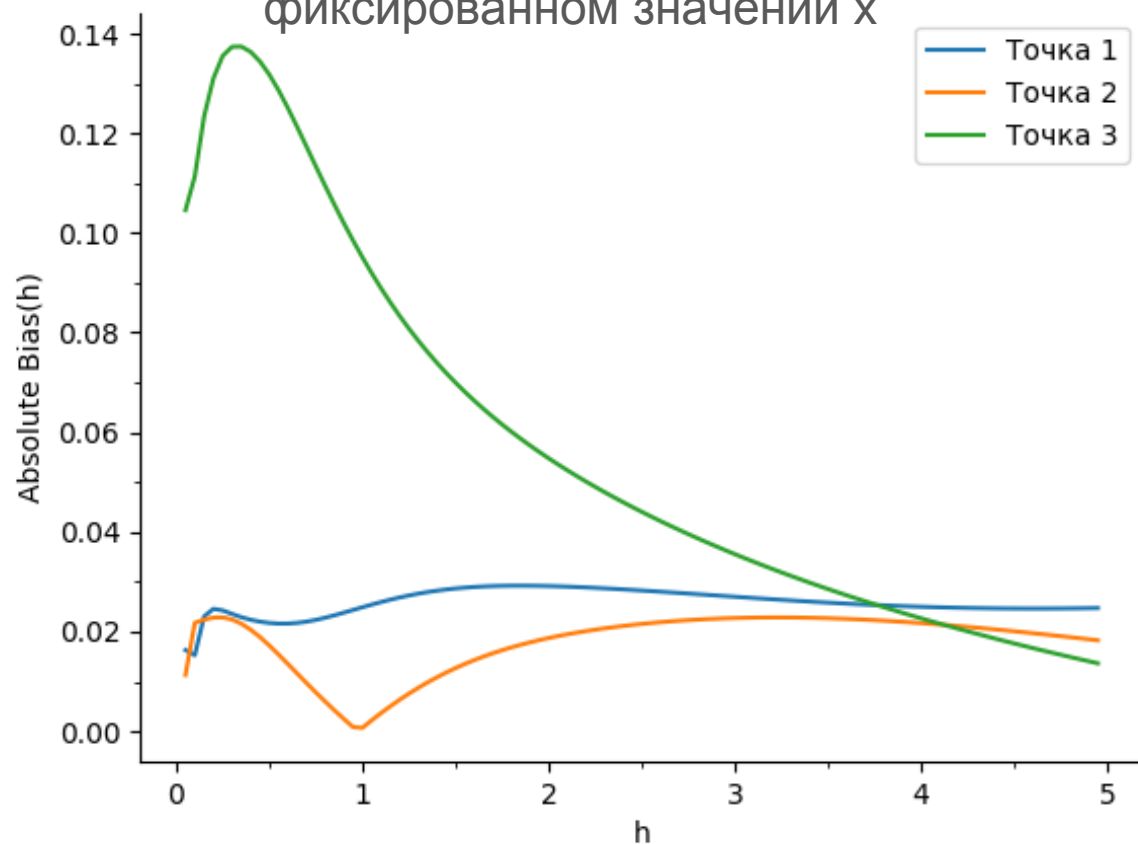


Рисунок 22. График зависимости смещения восстановленной плотности от ширины гауссова окна для трех фиксированных точек

#### Задание 4. Анализ смещения восстановленной плотности при фиксированном значении $x$

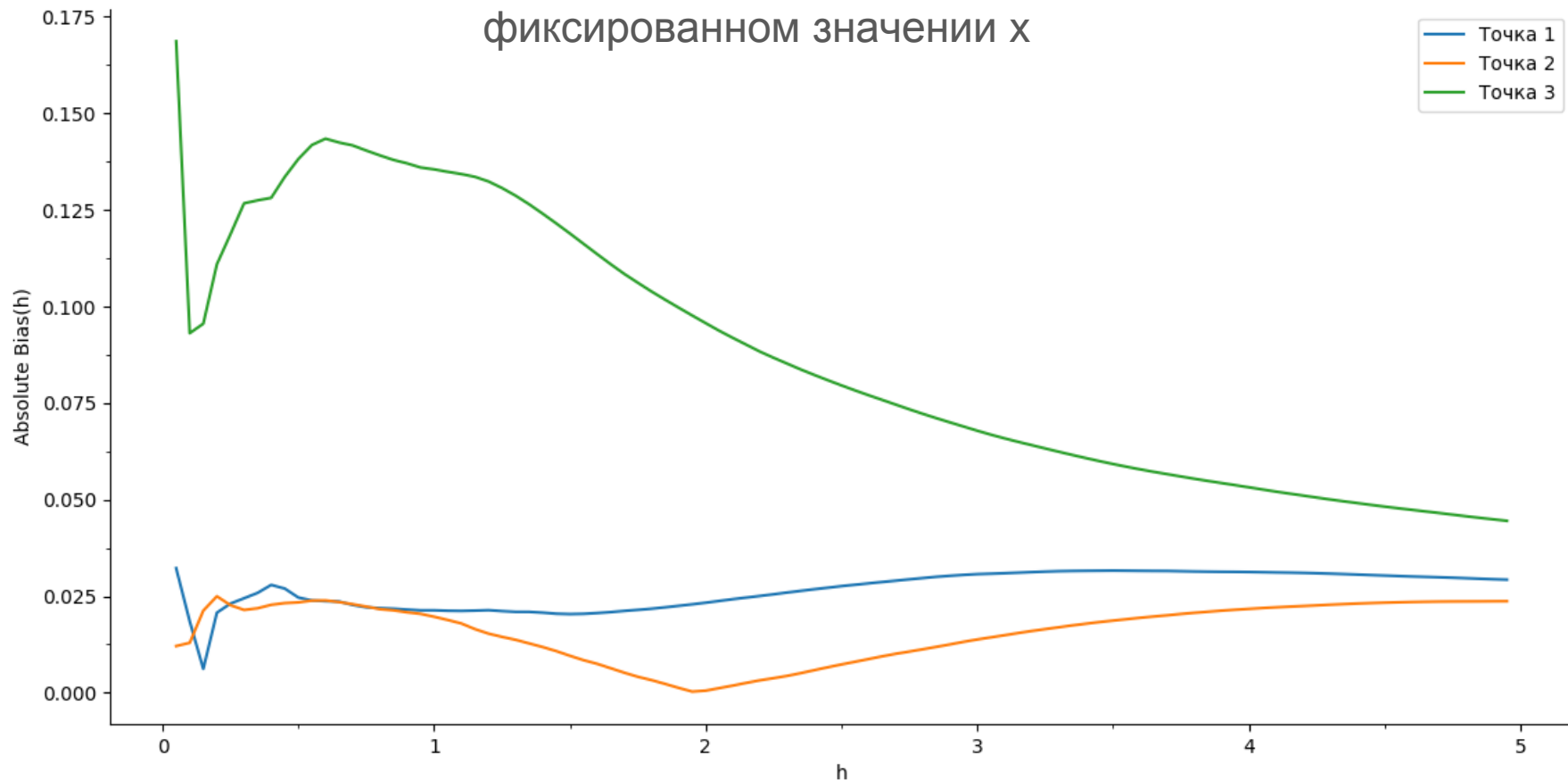


Рисунок 23. График зависимости смещения восстановленной плотности от ширины окна Епанечникова для трех фиксированных точек

#### Задание 4. Анализ смещения восстановленной плотности при фиксированном значении $x$

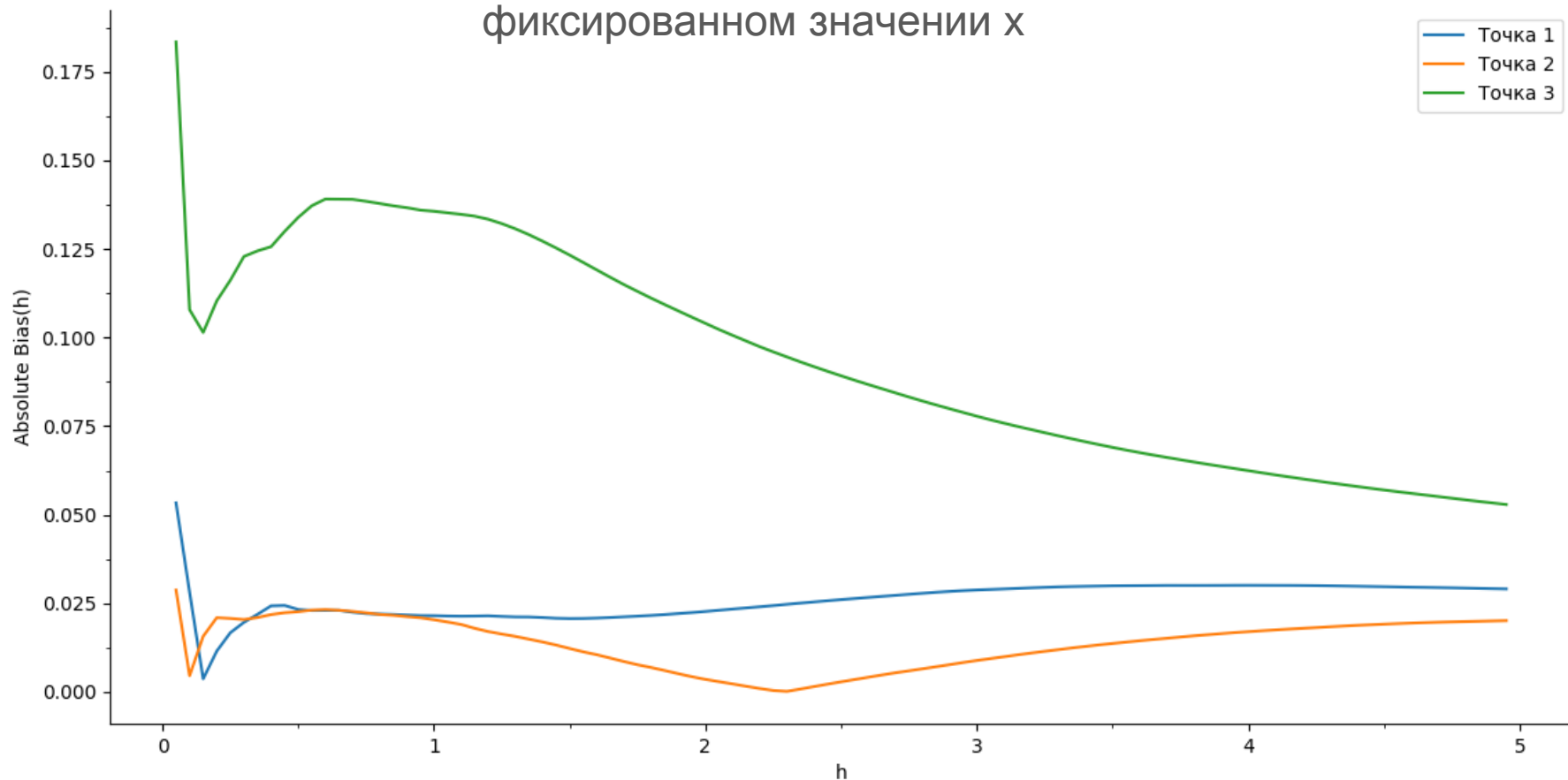


Рисунок 24. График зависимости смещения восстановленной плотности от ширины треугольного окна для трех фиксированных точек

## Задание 5. Анализ средней интегральной ошибки

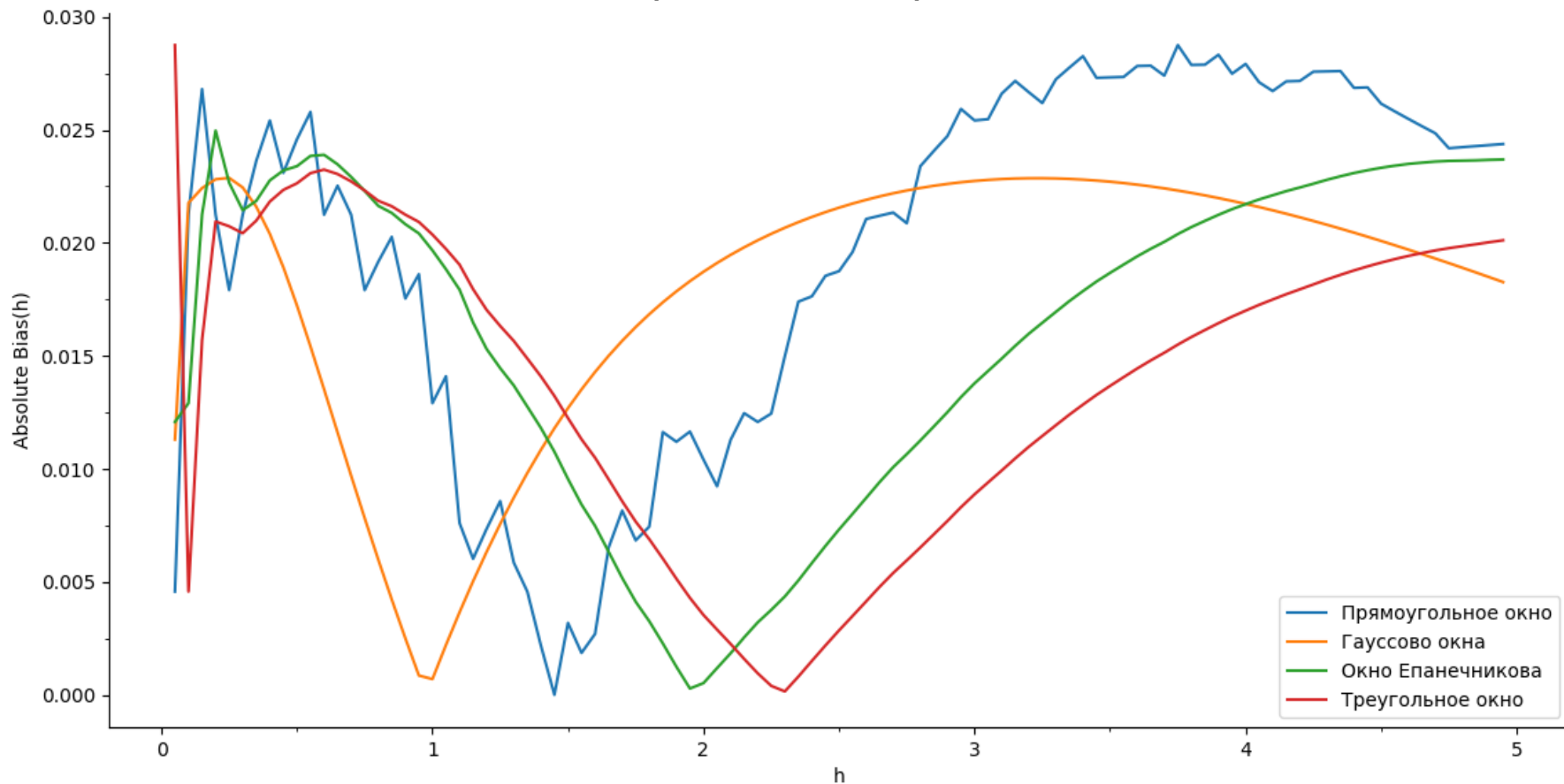


Рисунок 25. График зависимости средней интегральной ошибки от ширины окна

# Выводы

- . Дисперсия убывает с увеличением окна
- . Смещение незначительно убывает с увеличением окна
- . Интегральная ошибка убывает до ширины окна, рассчитанной по правилу Сильвермана, а затем возрастает
- . Оптимальная ширина окна находится в окрестности ширины по Сильверману
- . Наименьшая ошибка с окном Сильвермана получилась для окна Гаусса, он же и наиболее плавно описывает плотность распределения