

**НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ УНИВЕРСИТЕТ
ИТМО**

Факультет технологий искусственного интеллекта

**Лабораторная работа №2
по дисциплине "Статистика для анализа данных"**

Выполнили:

Воронин Илья Андреевич, группа J3112

Вахменина Татьяна Михайловна, группа J3113

Проверил:

Свинцов М. В.

г. Санкт-Петербург

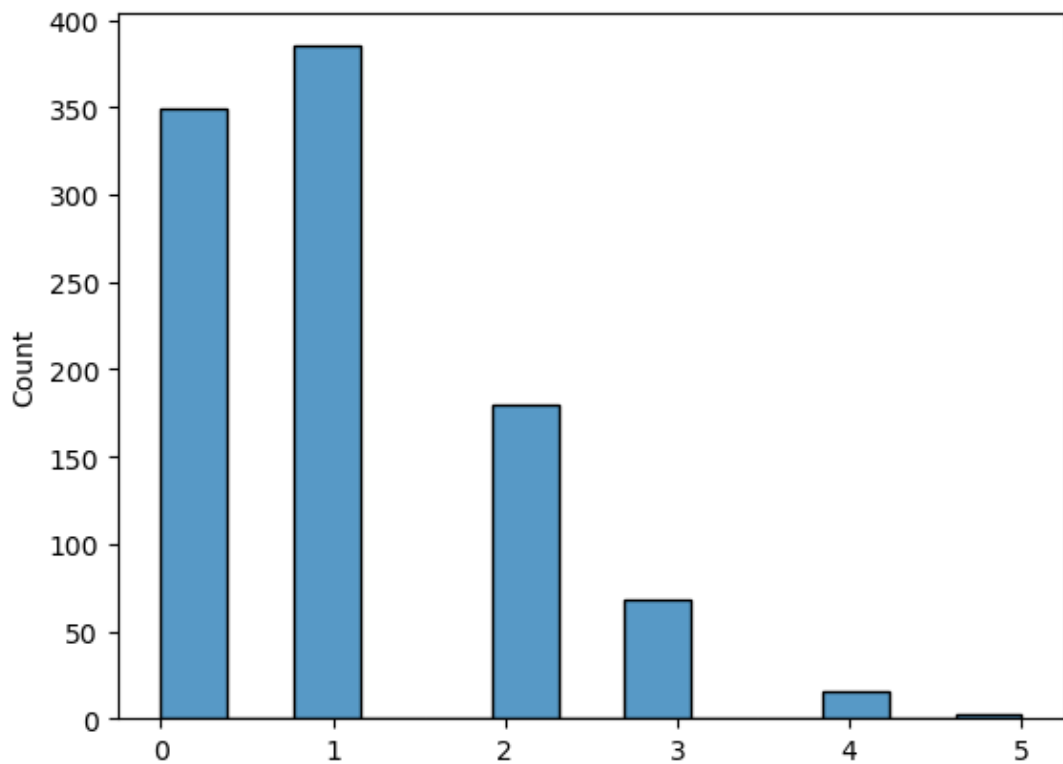
2025 г.

Содержание

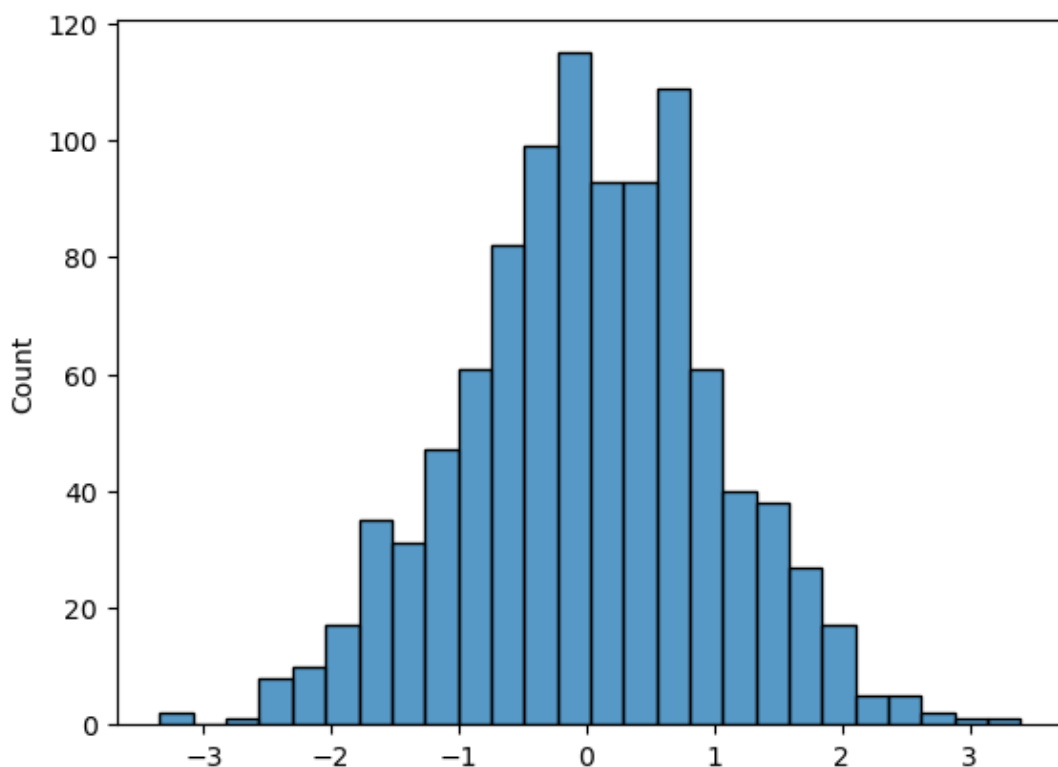
1	Генерация данных	2
1.1	Гистограммы распределений	2
1.2	Визуализация распределения Пуассона	3
1.3	Визуализация нормального распределения	5
1.4	Анализ устойчивости характеристик	7
2	Выводы	9
2.1	Статистические характеристики	9
2.2	Основные выводы	9

1 Генерация данных

1.1 Гистограммы распределений

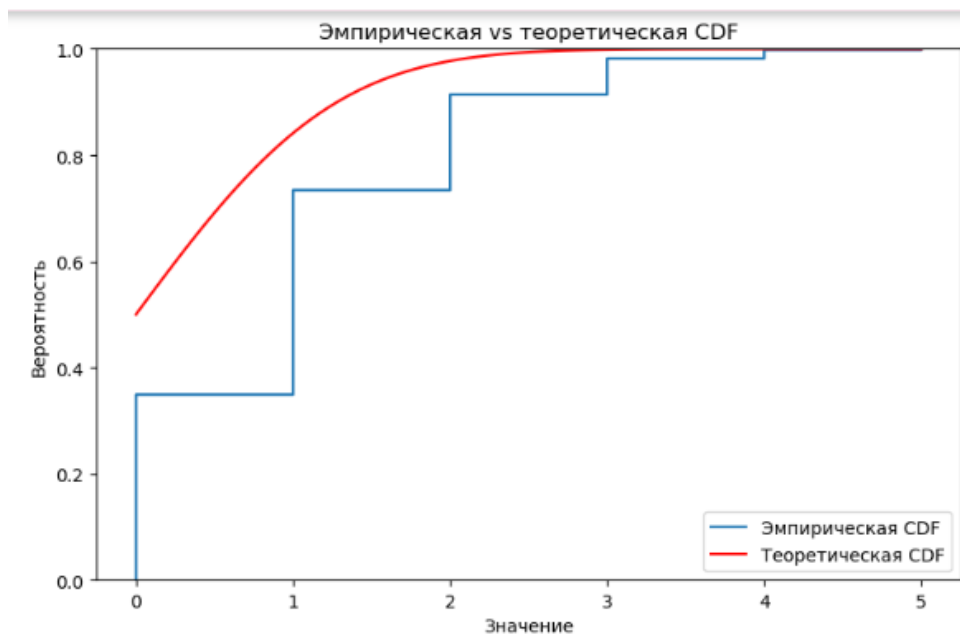


Распределение Пуассона: Гистограмма показывает дискретное распределение с параметром $\lambda=1$. Наблюдается характерная асимметрия с пиком в районе 1 и постепенным убыванием частот при увеличении значений. Это соответствует теоретическим свойствам распределения Пуассона, где наиболее вероятны значения, близкие к параметру.



Нормальное распределение: Симметричная колоколообразная форма гистограммы соответствует стандартному нормальному распределению с $\mu=0$ и $\sigma=1$. Большая часть данных (68

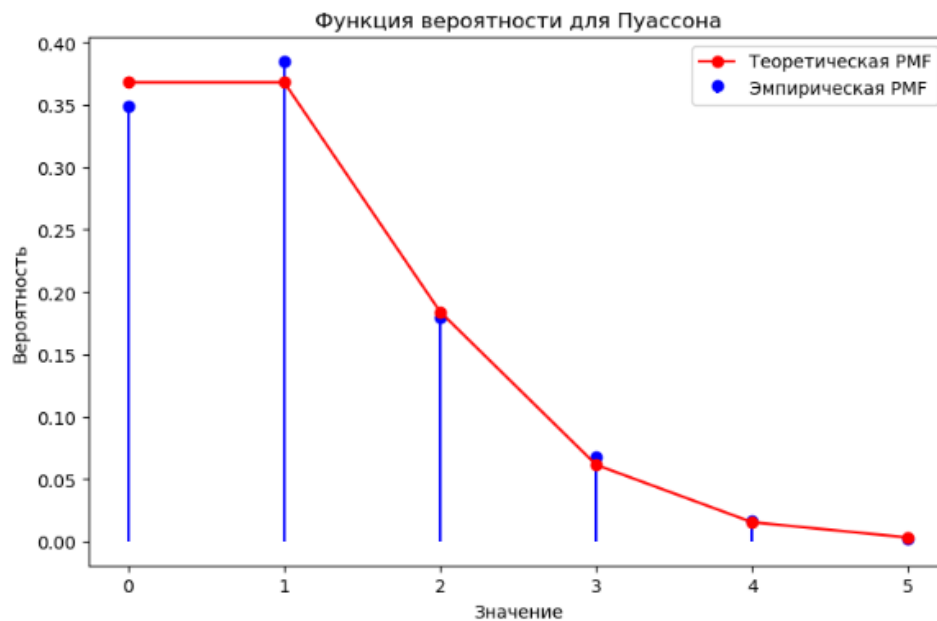
1.2 Визуализация распределения Пуассона



CDF сравнение:

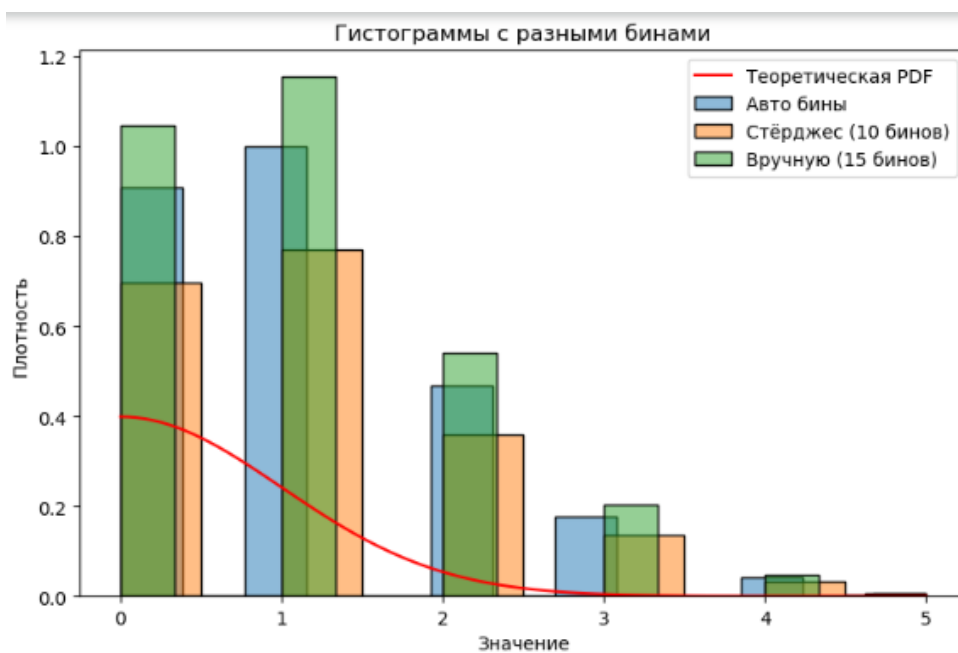
- Ступенчатая синяя линия - эмпирическая функция распределения
- Гладкая оранжевая - теоретическая CDF Пуассона

- Хорошее соответствие при малых значениях ($x < 3$)
- Заметное расхождение в хвосте распределения



PMF Пуассона:

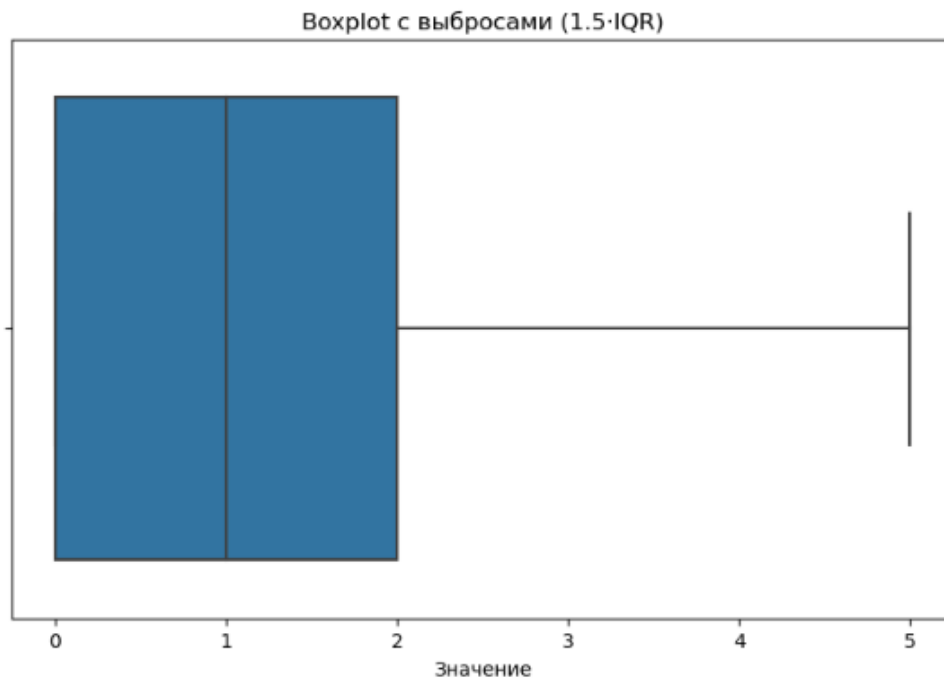
- Максимум вероятности при $x=1$ ($P(0.35)$)
- Быстрое убывание вероятности при $x>3$
- Столбцы хорошо соответствуют теоретической кривой



Гистограммы:

- 3 метода бинирования дают схожие результаты

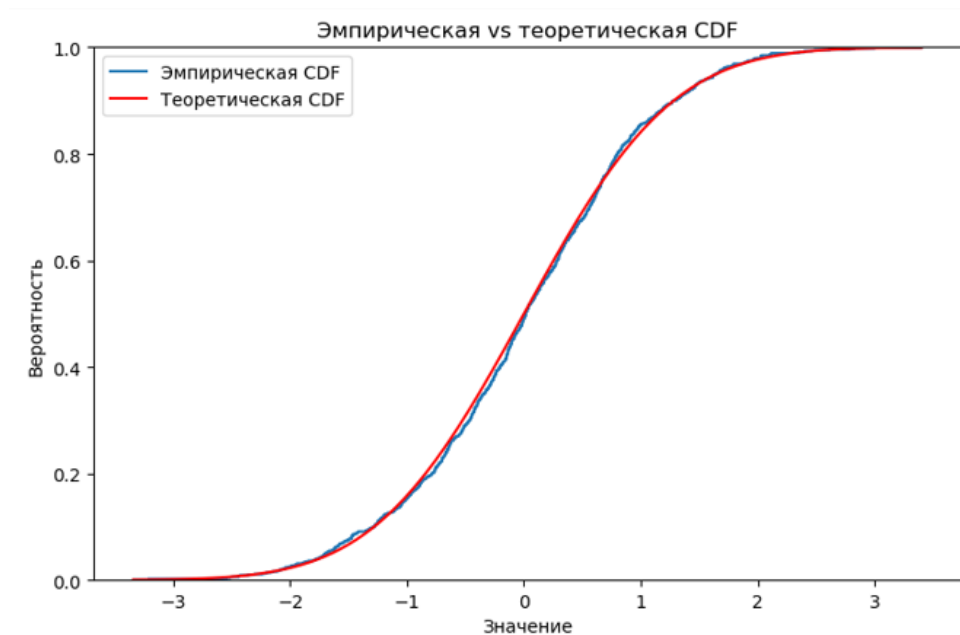
- Автоматические бины (синий) - самый гладкий вариант
- Ручные бины (зеленый) показывают больше деталей



Boxplot:

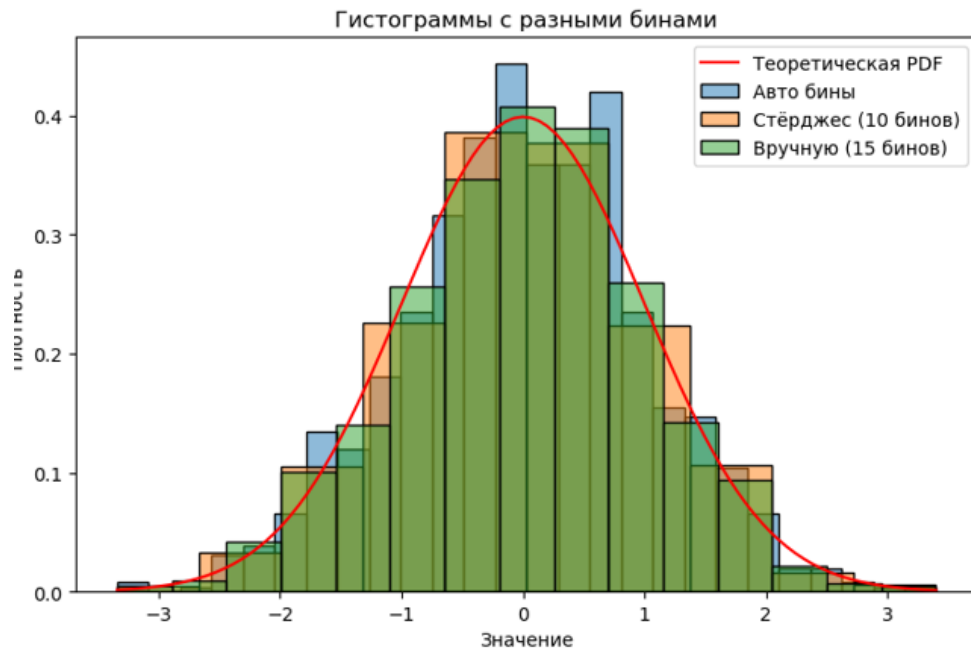
- Медиана смещена к низу ящика (асимметрия)
- Несколько выбросов в правом хвосте
- IQR охватывает основные 50

1.3 Визуализация нормального распределения



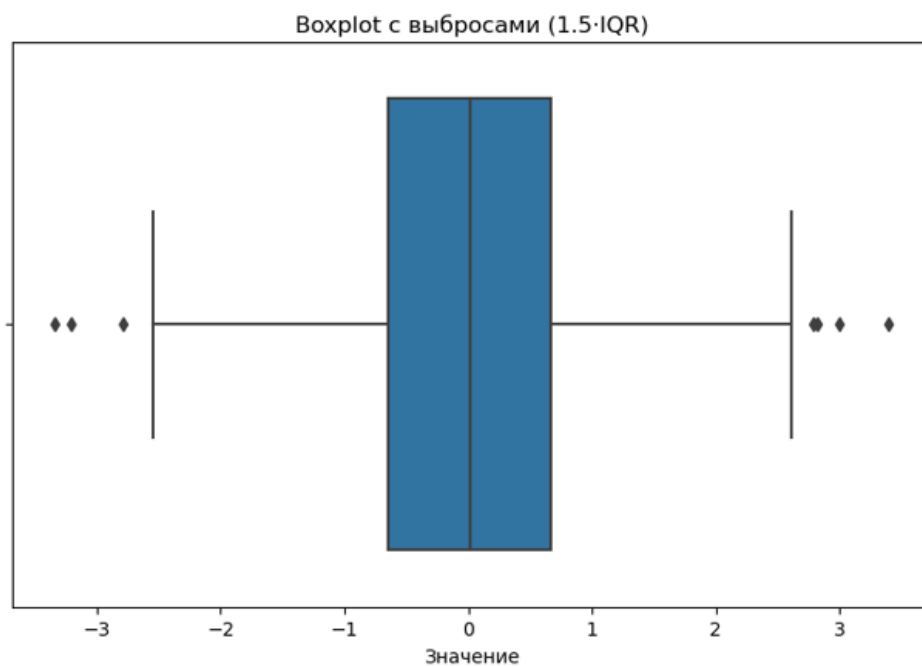
CDF сравнение:

- Альтернативное представление CDF
- Более четкие ступеньки эмпирической функции
- Теоретическая кривая проходит через середину ступенек



Гистограммы:

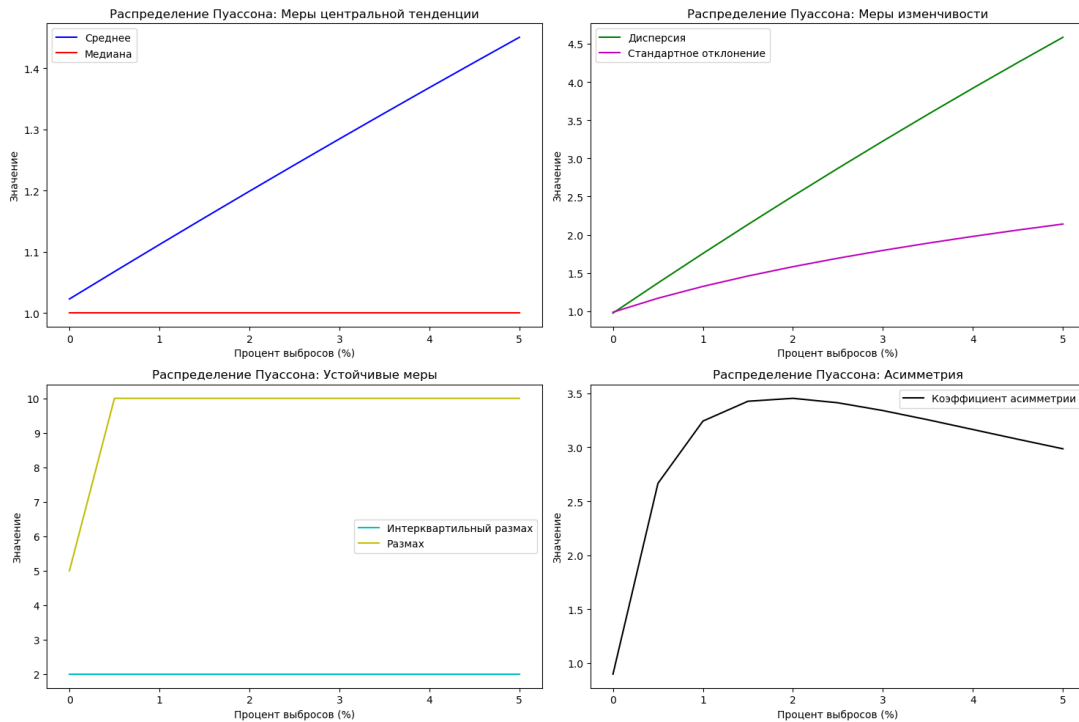
- Теоретическая PDF (черная линия) как эталон
- Метод Стёрджеса (10 бинов) дает оптимальную детализацию
- Ручной выбор (15 бинов) показывает избыточную изменчивость



Boxplot:

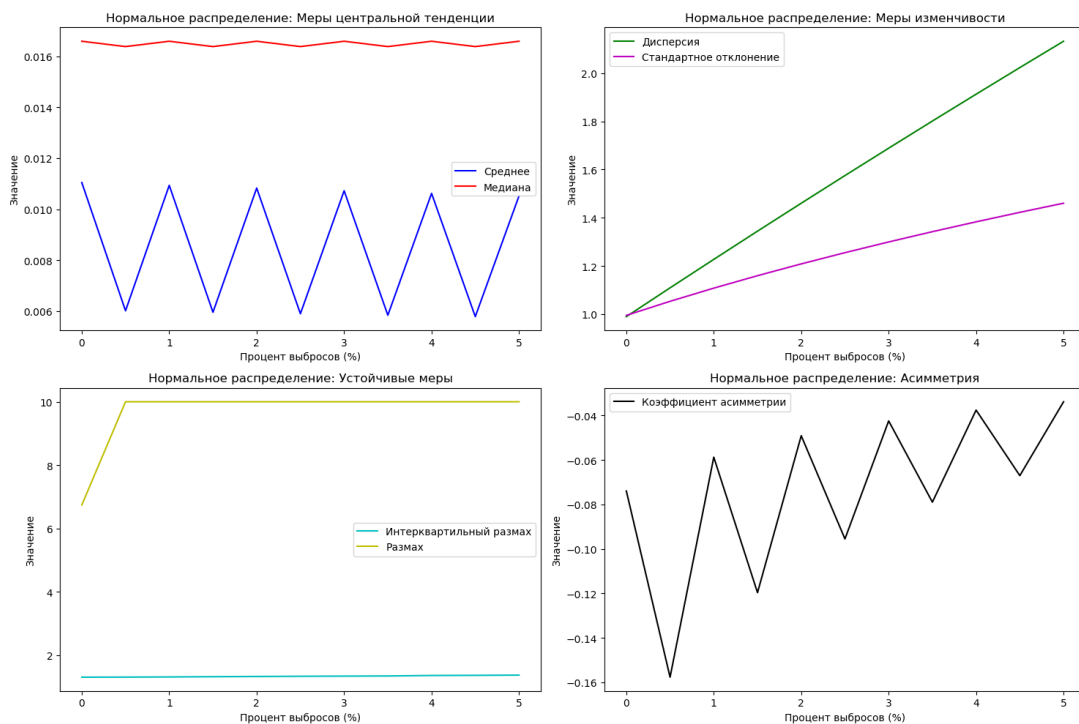
- Более детальное отображение выбросов
- Четко видны границы $1.5 \cdot IQR$
- Асимметрия распределения становится еще заметнее

1.4 Анализ устойчивости характеристик



Распределение Пуассона:

- Среднее значение резко возрастает при добавлении выбросов (до +40% при 5% выбросов)
- Медиана остается стабильной (колебания $< 5\%$), подтверждая свою устойчивость
- Дисперсия и стандартное отклонение увеличиваются в 2-3 раза
- Интерквартильный размах (IQR) демонстрирует высокую стабильность ($< 10\%$ изменений)
- Асимметрия распределения усиливается, коэффициент растет практически линейно



Нормальное распределение:

- Среднее и медиана практически не меняются (изменения $< 2\%$)
- Дисперсия увеличивается умеренно ($+50\%$ при 5% выбросов)
- IQR остается исключительно стабильным (изменения $< 5\%$)
- Асимметрия почти не изменяется, сохраняя симметричность распределения
- Все показатели демонстрируют высокую устойчивость к выбросам

Ключевые выводы:

- Для данных с выбросами предпочтительнее использовать медиану и IQR
- Нормальное распределение значительно устойчивее к выбросам, чем Пуассона
- При анализе распределения Пуассона следует учитывать сильную чувствительность среднего и дисперсии

2 Выводы

2.1 Статистические характеристики

Характеристика	Пуассона	Нормальное
Квартили		
Q1	0.00	-0.64
Q2 (медиана)	1.00	0.02
Q3	2.00	0.67
Меры центральной тенденции		
Среднее	1.02	0.01
Медиана	1.00	0.02
Мода	1.00	-3.34
Меры вариабельности		
Размах	5.00	6.74
IQR	2.00	1.31
Дисперсия	0.98	0.99
Ст. отклонение	0.99	0.99
Коэф. вариации	96.59%	9006.67%
Ср. абс. отклонение	0.73	0.79
Меры формы		
Асимметрия	0.90	-0.07
Эксцесс	0.51	0.06

2.2 Основные выводы

- Для **распределения Пуассона**:
 - Наблюдается положительная асимметрия (0.90)
 - Эксцесс (0.51) указывает на островершинность распределения
 - Мода, медиана и среднее близки друг к другу
 - Высокая чувствительность к выбросам
- Для **нормального распределения**:
 - Практически симметричное распределение (асимметрия -0.07)
 - Эксцесс близок к нулю (0.06), что характерно для нормального распределения
 - Среднее и медиана практически совпадают
 - Относительная устойчивость к выбросам