

# Практические задачи по курсу: Методы анализа распределений данных с тяжёлыми хвостами. Вариант: 7

Васильев Михаил Владимирович  
Студент 5 курса факультета ФРКТ

(Московский физико-технический институт)  
(Dated: 16 декабря 2023 г.)

## I. DATA

Данные взяты из массив данных моделирования случайных графов, полученных с помощью моделей предпочтительного присоединения и кластерного присоединения с разными параметрами для анализа. Данные представляют из себя столбец Excel размером 500 значений.

## II. EXERCISE 1

Постановка задачи:

Сгенерировать распределение Фреше с параметром  $\gamma = 1.5$ .

$$F(x) = \exp(-(\gamma x)^{-1/\gamma} 1(x > 0)) \quad (1)$$

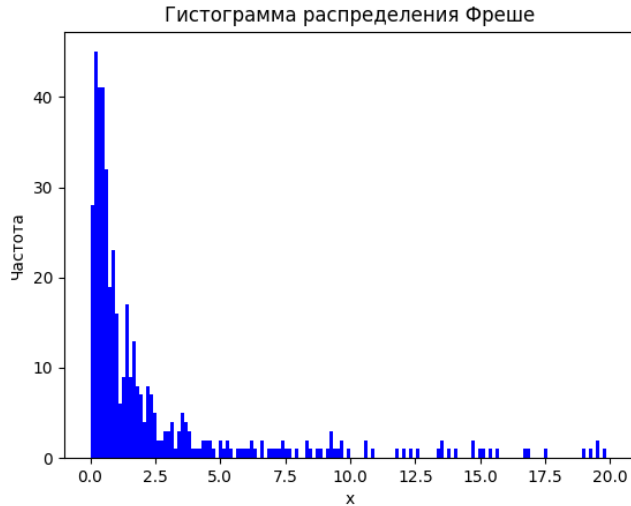


Рис. 1: Распределение Фреше

## III. EXERCISE 2

Постановка задачи: Рассчитать следующую статистику:

$$R_n(p) = \frac{M_n(p)}{S_n(p)}, n \geq 1, p > 0 \quad (2)$$

$$M_n(p) = \max(|X_1|^p, \dots, |X_n|^p), \quad (3)$$

$$S_n(p) = |X_1|^p + \dots + |X_n|^p \quad (4)$$

sample:  $X^n = X_1, \dots, X_n$  для  $n = 1, 2, \dots$

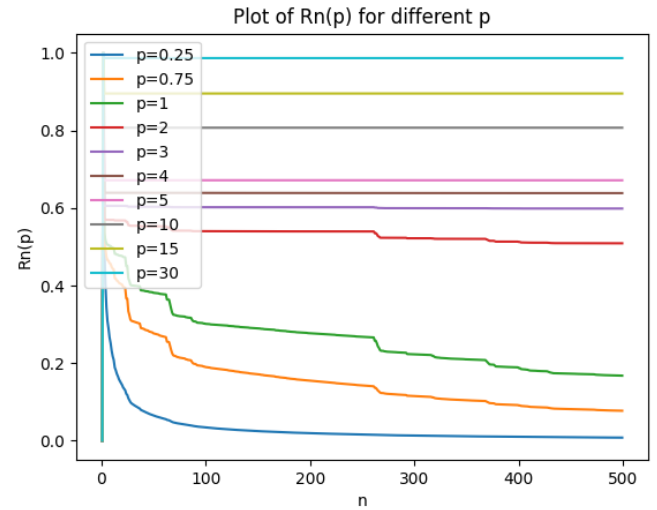


Рис. 2: Зависимость отношения максимума к сумме элементов для различных значений хвостового индекса

По итогам исследования можно сказать, Для  $p \in \{0.25, 0.75, 1\}$   $R_n(p)$  по всей видимости стремиться к нулю при возрастании  $n$ . Для  $p \in \{2, 3, 4, 5, 10, 15, 30\}$   $R_n(p)$  по всей видимости стремиться к положительной константе при возрастании  $n$ .

Вывод:  $E|X|^p < \infty$  для  $p \leq 1$  только,  $E|X|^p = \infty$  для  $p > 1$ .

## IV. EXERCISE 3

Постановка задачи: Построить QQ-plot для упорядоченных данных для распределений: нормального, логнормального, экспоненциального и Парето общего.

На графиках представлен QQ-plot, где красная линия  $y = x$

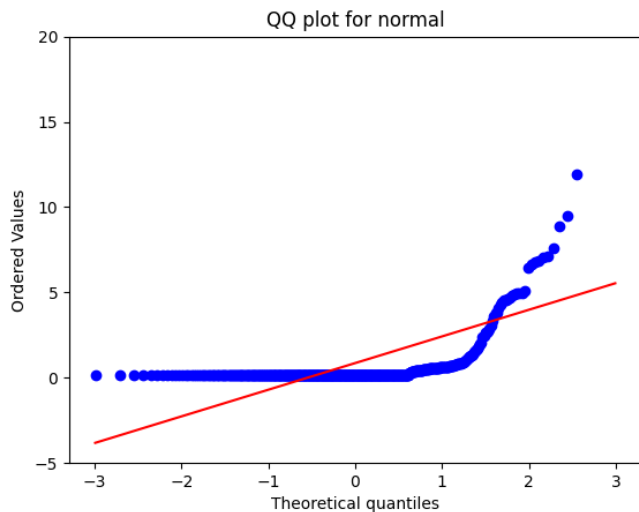


Рис. 3: QQ-plot для нормального распределения

Как видно нормальное распределение не подходит для выборки.

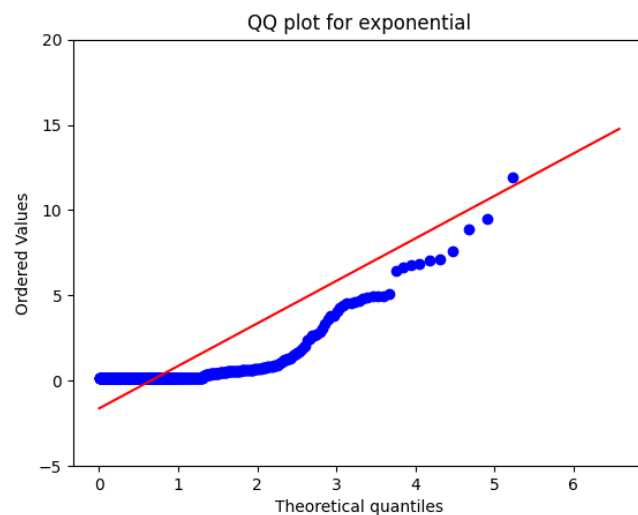


Рис. 5: QQ-plot для экспоненциального распределения

Как видно экспоненциальное распределение не подходит для выборки.

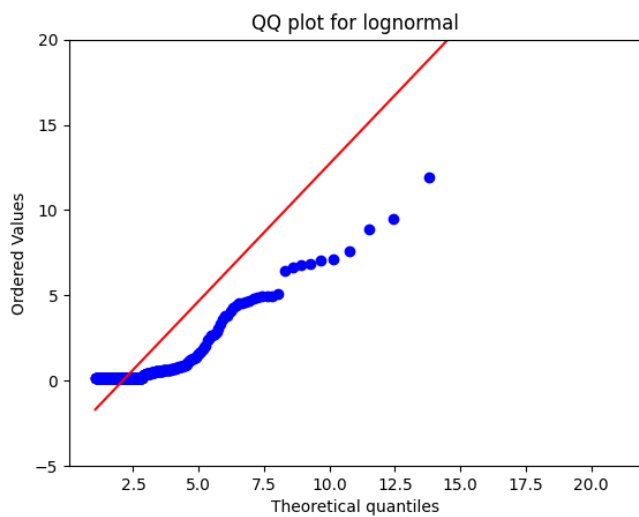


Рис. 4: QQ-plot для логнормального распределения

Как видно логнормальное распределение не подходит для выборки.

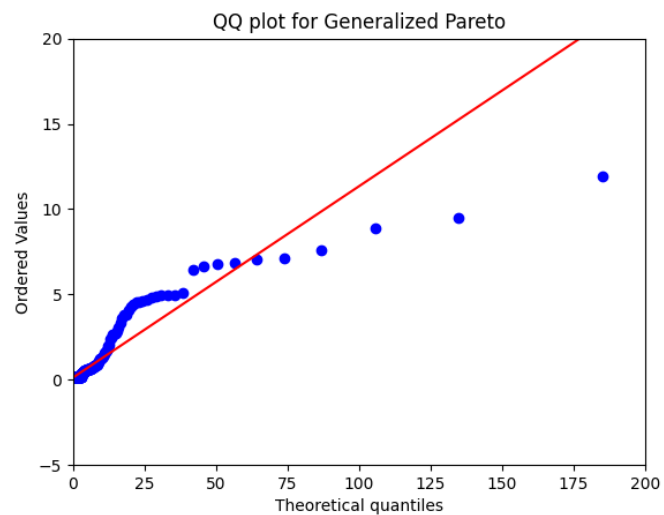


Рис. 6: QQ-plot для Парето обобщённого

Как видно Парето обобщённое распределение не подходит для выборки.

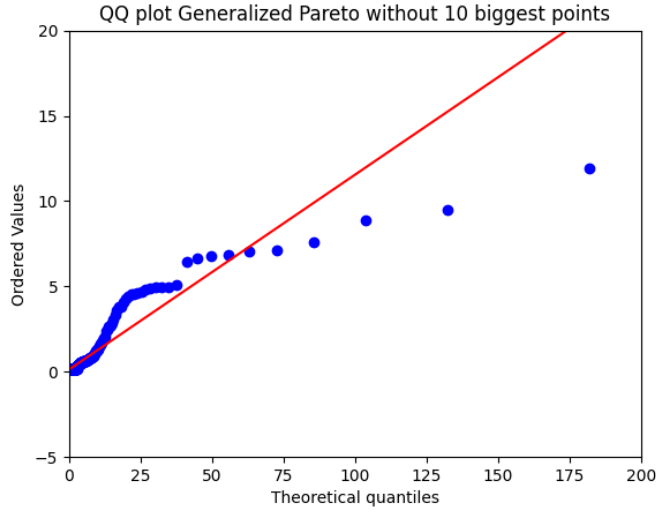


Рис. 7: QQ-plot для Парето обобщённого распределения без наибольших 10 точек

Как видно убирание наибольших значений из выборки слабо влияет на QQ-plot.

Вывод: ни одно из распределений не подходит для описания выборки, убирание наибольших значений выборки не имеет смысла.

#### V. EXERCISE 4

Постановка задачи: Построить график mean excess function.

$$e_n(u) = \frac{\sum_{i=1}^n (X_i - u)1(X_i > u)}{\sum_{i=1}^n 1(X_i > u)} \quad (5)$$

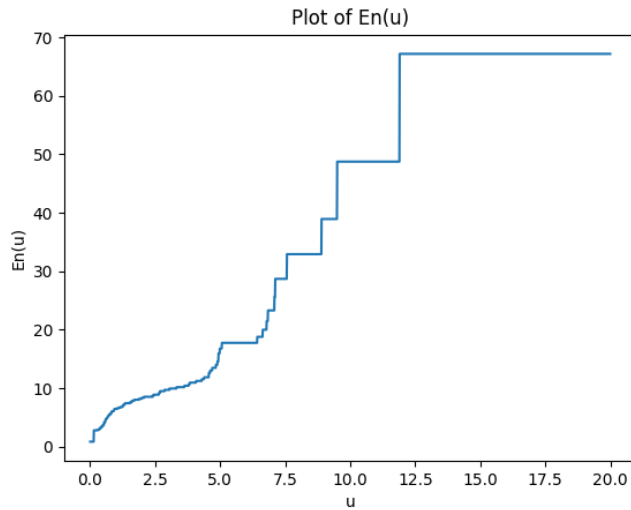


Рис. 8: График для mean excess function

Видно что для больших  $u$  график  $e_n(u)$  стремится к бесконечности, что означает, что распределение принадлежит к классу распределений с тяжёлыми хвостами.

Вывод: распределение имеет тяжёлые хвосты.

#### VI. EXERCISE 5

Постановка задачи: Having the empirical or generated data  $X_n = \{X_1, \dots, X_n\}$  reorder the data as  $X(1) \leq X(2) \leq \dots \leq X(n)$ . Calculate and compare the following estimates of the tail index of your data. Investigate the sign of an estimate and make conclusion regarding the heavy tails.

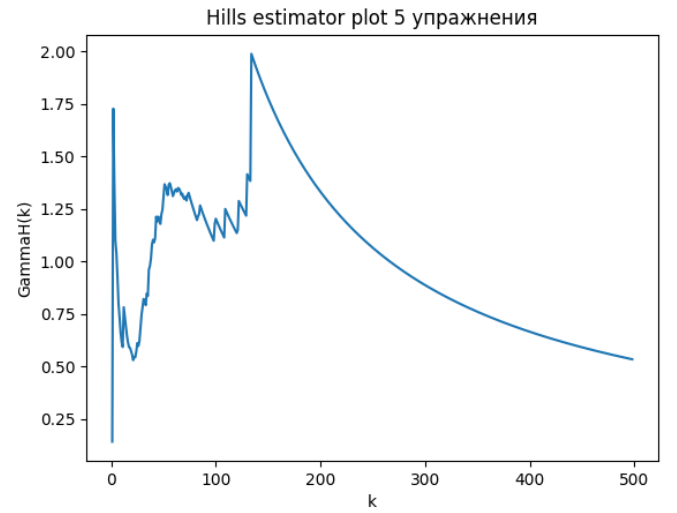


Рис. 9: Hills estimator plot

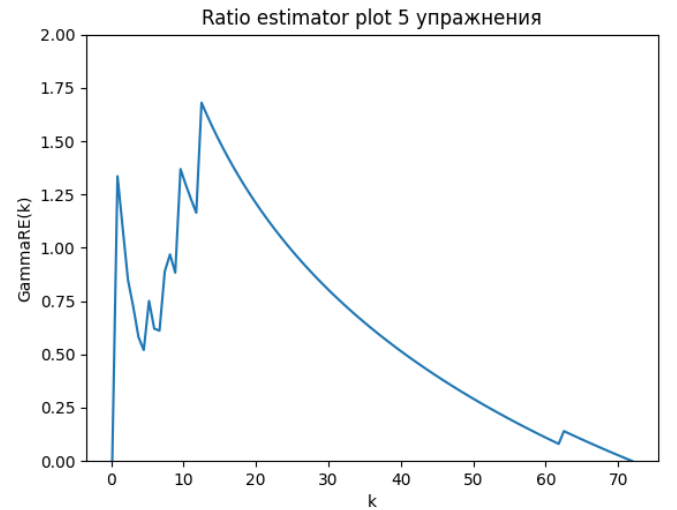


Рис. 10: Ratio estimator plot

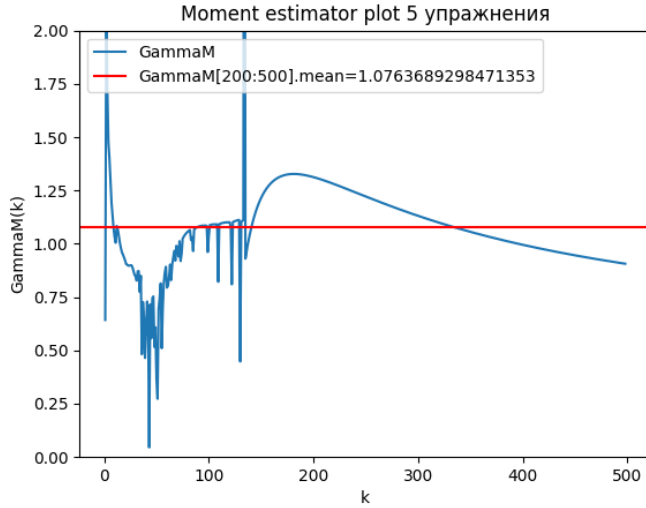


Рис. 11: Moment estimator plot

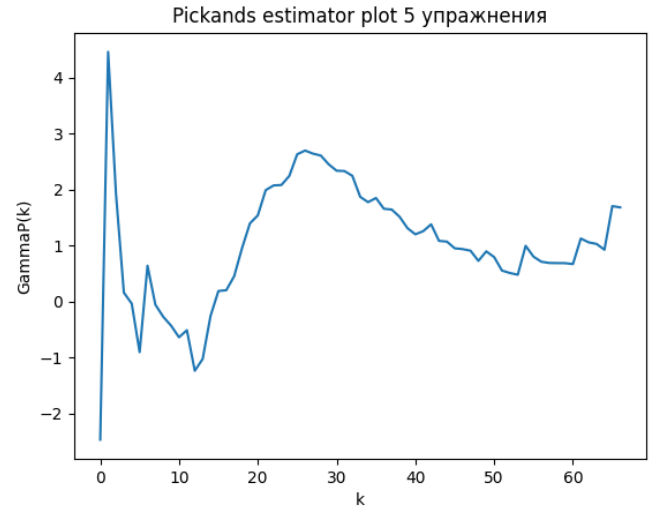


Рис. 13: Pickands estimator plot

Из графиков видно, что оценки Hill, Ratio и Pickands на предоставленных данных работают плохо. Тогда как оценки Moment и UH стабилизируются между 200 и 500 значением выборки. При взятии среднего по этим оценкам между 200 и 500 элементом выборки получаются значения 1.0763 и 1.131 соответственно.

Вывод: по всей видимости хвостовой индекс можно положить примерно равным  $\gamma \approx 1.1$ .

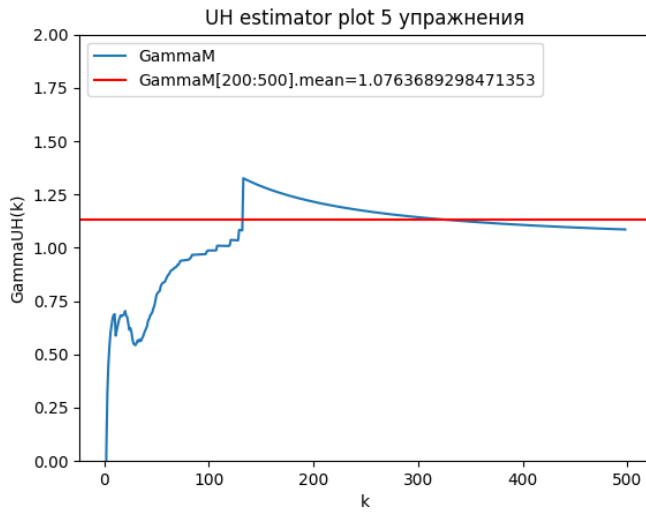


Рис. 12: UH estimator plot

## VII. EXERCISE 6

Постановка задачи: Вывести эстиматор Хилла с доверительным интервалом.

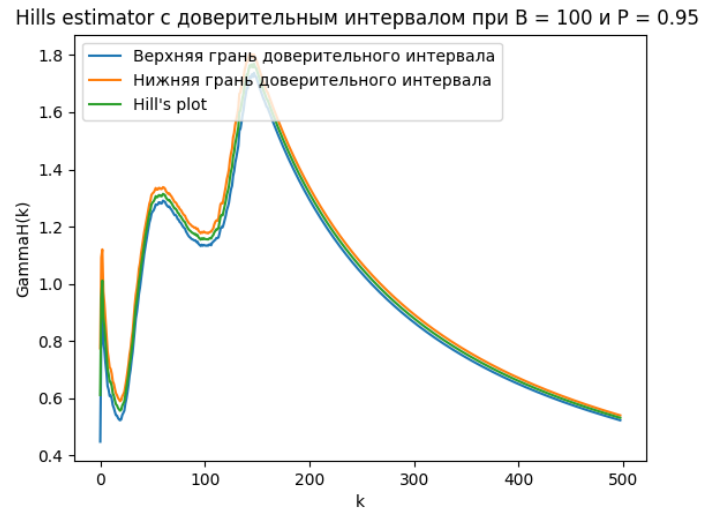


Рис. 14: Hills estimator with confidence interval