Домашнее задание 1

Юрасов Никита Андреевич

Обновлено 16 октября 2019 г.

Содержание

Эта работа представляет собой отчет к Домашнему Заданию №2. Все графики, которые представлены в этой работе, настроены так, что при перезапуске Jupyter Notebook ДЗ2.jpynb, они автоматически обновляются. К сожалению, с представлением выборок, вариационных рядов, квантилей по выборкам и т.д. так сделать не получается, поэтому они были скопированы с конкретными числами прямо в этот отчет, но в файле с кодом могут быть представлены уже другие значения.

1 Моделирования выбранных случайных величин

1.1 Распределение Пуассона

Алгоритмы для моделирования случайной величины были приведены еще в Домашнем Задании №1. Здесь будут приведены выборки для $n = 5$ и $n = 10$, которые можно найти в файле `ДЗ2.ipynb`:

```
Выборки для n = 5:

3, 2, 4, 5, 3
3, 2, 1, 0, 2
0, 0, 1, 0, 1
5, 2, 1, 2, 3
3, 5, 2, 1, 2

Выборки для n = 10:

5, 1, 2, 1, 2, 2, 0, 1, 2, 3
2, 3, 0, 5, 4, 0, 1, 2, 4, 5
3, 3, 0, 1, 4, 0, 5, 2, 5, 5
3, 0, 2, 1, 1, 4, 2, 5, 1, 0
0, 5, 3, 3, 1, 0, 3, 0, 3, 2
```

1.2 Распределение Эрланга

Ниже представлены 10 выборок для n=5 и n=10.

```
Выборка для n = 5 № 1
[19.1737165   6.05731135 17.04167437 29.69413063  7.01242978]
Выборка для n = 5 № 2
[20.94102033  6.48575996 16.88643751  3.67872403  6.82801986]
Выборка для n = 5 № 3
[10.54619228 17.4489759   8.25372007 17.6778771   4.9160208 ]
Выборка для n = 5 № 4
[ 8.58874631 15.31376918 12.88753409 36.23876433 13.14721791]
Выборка для n = 5 № 5
[ 9.47307096 13.56162222 11.03214536  8.11734598 26.75859288]
Выборка для n = 10 № 1
[34.22933019 12.21042258 11.1323889   5.43419976 19.62276674 11.09065638
  4.60553775  4.8749829  10.96165848  3.63172443]
Выборка для n = 10 № 2
[ 6.70257093 26.28899906 11.10276011 10.21477787  5.10451035 12.0788519
  3.79235536  1.01084478  5.48192401 14.51384613]
Выборка для n = 10 № 3
[ 9.46147936 23.59461559  4.07224105  1.54489137  0.27299634  6.8743321
 12.20886608  0.55305523  3.31145748  7.29781774]
Выборка для n = 10 № 4
[ 3.87097209 26.71823616 10.94367322 19.12355881  1.35076053  2.67453133
  4.25619674 12.80260401 12.32338934 11.46532159]
Выборка для n = 10 № 5
[ 4.02262128  4.58224671  7.37891429  6.93065105  0.98490678  4.81853169
  1.60210139 13.68201202  5.43227555 19.71421185]
```

2 Построение эмпирической функции распределения

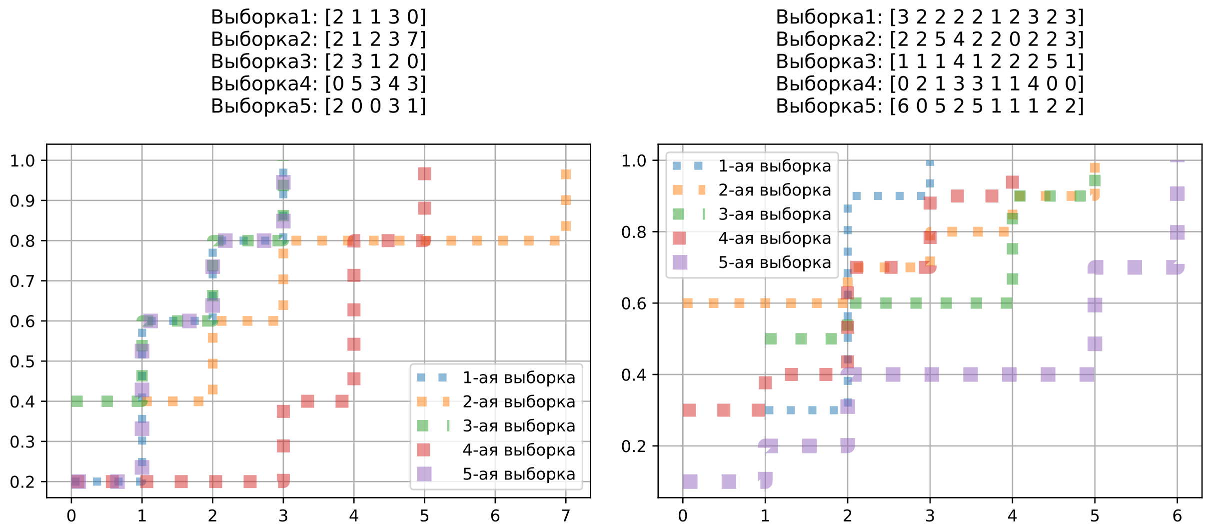
2.1 Распределение Пуассона

Определение 2.1. *Для произвольного $x \in \mathbb{R}$ рассмотрим случайную величину*

$$\mu_n(x) = \sum_{i=1}^{n} Ind(X_i \leq x)$$

равную числу элементов выборки меньше либо равных x.
Тогда функцию $\hat{F}_n(x) = \frac{\mu_n}{n}$ будем называть эмпирической функцией распределения.

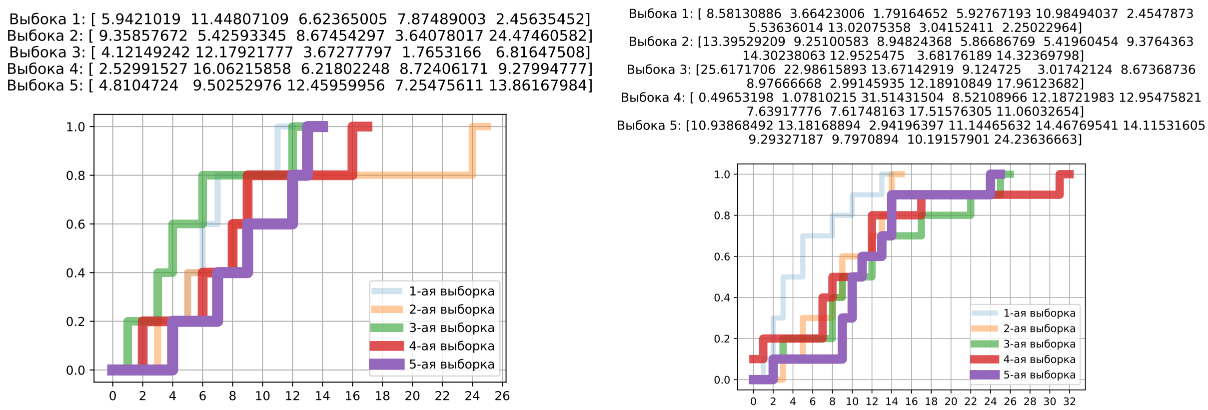
На рисунке 1 представлены графики эмпирический функций 10 выборок.



n=5 n=10

Рис. 1: Эмпирические функции

2.2 Распределение Эрланга



n=5 n=10

Рис. 2: Эмпирические функции

Для построения верхней границы разности двух эмпирических функций, были написаны два алгоритма (для каждого распределения), которые анализируют поточечно каждую эмпирическую функцию и находят максимальную разницу между ними. Воспользуемся теоремой Смирнова-Колмогорова, в которой определяется [?, Ивченко, Медведев]:

$$sup|\hat{F}(x) - F(x)|$$

Ниже будут приведены результаты, а сам алгоритм можно найти в `ДЗ3.jpynb`. Оценки для $n \in \{5, 10, 100, 1000, 100k\}$:

```
Poisson Distibution:
n=5 <-> n=5: 0.20000000000000007
n=5 <-> n=10: 0.39999999999999997
n=5 <-> n=100: 0.4100000000000001
n=5 <-> n=1000: 0.31600000000000006
n=5 <-> n=100k: 0.59141
Erlang Distribution:
n=5 <-> n=5: 0.4
n=5 <-> n=10: 0.5
n=5 <-> n=100: 0.2799999999999998
n=5 <-> n=1000: 0.33
n=5 <-> n=100k: 0.3245100000000001


Poisson Distibution:
n=10 <-> n=5: 0.2999999999999998
n=10 <-> n=10: 0.39999999999999997
n=10 <-> n=100: 0.3799999999999999
n=10 <-> n=1000: 0.31599999999999995
n=10 <-> n=100k: 0.5914099999999999
Erlang Distribution:
n=10 <-> n=5: 0.30000000000000004
n=10 <-> n=10: 0.4
n=10 <-> n=100: 0.3
n=10 <-> n=1000: 0.35799999999999993
n=10 <-> n=100k: 0.36229


Poisson Distibution:
n=100 <-> n=5: 0.18000000000000005
n=100 <-> n=10: 0.29
n=100 <-> n=100: 0.39000000000000007
n=100 <-> n=1000: 0.2460000000000001
n=100 <-> n=100k: 0.3414100000000001
Erlang Distribution:
n=100 <-> n=5: 0.21999999999999997
n=100 <-> n=10: 0.21999999999999975
n=100 <-> n=100: 0.09999999999999998
n=100 <-> n=1000: 0.03999999999999998
n=100 <-> n=100k: 0.04229000000000005


Poisson Distibution:
n=1000 <-> n=5: 0.17900000000000005
```

```
n=1000 <-> n=10: 0.318
n=1000 <-> n=100: 0.22199999999999998
n=1000 <-> n=1000: 0.13899999999999998
n=1000 <-> n=100k: 0.21806000000000003
Erlang Distribution:
n=1000 <-> n=5: 0.19899999999999995
n=1000 <-> n=10: 0.20499999999999974
n=1000 <-> n=100: 0.07299999999999995
n=1000 <-> n=1000: 0.04300000000000004
n=1000 <-> n=100k: 0.022100000000000036


Poisson Distibution:
n=100k <-> n=5: 0.26381
n=100k <-> n=10: 0.46380999999999994
n=100k <-> n=100: 0.21771000000000001
n=100k <-> n=1000: 0.08020000000000005
n=100k <-> n=100k: 0.30314
Erlang Distribution:
n=100k <-> n=5: 0.20226999999999995
n=100k <-> n=10: 0.19717999999999998
n=100k <-> n=100: 0.08894999999999997
n=100k <-> n=1000: 0.024510000000000032
n=100k <-> n=100k: 0.005320000000002134
```

3 Построение вариационного ряда выборки

Определение 3.1. *Пусть у нас имеется выборка* $X = (X_1, \cdots, X_n)$, *где* X_i *– независимая одинаково распределенная случайная величина из распределения* ξ.
Тогда ***вариационным рядом*** *выборки будем называть* $(X_{(1)}, \cdots, X_{(n)})$, *где* $X_{(1)} \leq \cdots \leq X_{(n)}$.

Другими словами можно сказать, что вариационный ряд является отсортированной выборкой по возрастанию. Сортировать выборки будем встроенными методами `Python`, а именно встроенной функцией `sorted()`, которая принимает в качестве аргумента массиво-подобный объект и меняет этот же объект.

К сожалению, сделать так, чтобы в отчете каждый раз обновлялись данные выборок автоматически, невозможно. Поэтому ниже представлены одни из возможных выборок и их вариационных рядов.

Код функций, которые генерирую выборки, а потом делают вариационные ряды из них, можно в файле `ДЗ2.jpynb`.

3.1 Распределение Пуассона

```
Вариационный ряд для выборки [3 2 4 5 3] --> [2, 3, 3, 4, 5]
Вариационный ряд для выборки [5 1 2 1 2 2 0 1 2 3] --> [0, 1, 1, 1, 2, 2, 2,
2, 3, 5]
```

Для поиска квантилей при заданном уровне была написана функция, которая на выходе дает вот такие результаты:

```
Квантиль для выборки [3 0 2 1 1 4 2 5 1 0] уровня 0.1 == 0.5
Квантиль для выборки [3 0 2 1 1 4 2 5 1 0] уровня 0.5 == 2.5
Квантиль для выборки [3 0 2 1 1 4 2 5 1 0] уровня 0.7 == 3
Квантиль для выборки [5 2 1 2 3] уровня 0.1 == 0
Квантиль для выборки [5 2 1 2 3] уровня 0.5 == 2
Квантиль для выборки [5 2 1 2 3] уровня 0.7 == 3
```

3.2 Распределение Эрланга

```
Вариационный ряд для выборки [19.1737165 6.05731135 17.04167437 29.69413063
7.01242978] --> [ 6.05731135 7.01242978 17.04167437 19.1737165 29.69413063]

Вариационный ряд для выборки [34.22933019 12.21042258 11.1323889 5.43419976
19.62276674 11.09065638 4.60553775 4.8749829 10.96165848 3.63172443] --> [
3.63172443 4.60553775 4.8749829 5.43419976 10.96165848 11.09065638 11.1323889
12.21042258 19.62276674 34.22933019]
```

Нахождение квантилей заданных трех уровней (0.1, 0.5, 0.7):

```
Квантиль для выборки [ 1.30148362 2.80649906 15.47504323 7.81896157 7.34531201]
уровня 0.1 = 3.0
```

```
Квантиль для выборки [ 1.30148362 2.80649906 15.47504323 7.81896157 7.34531201]
уровня 0.5 = 5.0
Квантиль для выборки [ 1.30148362 2.80649906 15.47504323 7.81896157 7.34531201]
уровня 0.7 = 15.0

Квантиль для выборки [ 9.02770023 28.94668457 3.10929781 13.83457346 1.22763611
11.26452997 8.32525049 14.16950039 5.91072393 8.57862944] уровня 0.1 = 1.5
Квантиль для выборки [ 9.02770023 28.94668457 3.10929781 13.83457346 1.22763611
11.26452997 8.32525049 14.16950039 5.91072393 8.57862944] уровня 0.5 = 8.0
Квантиль для выборки [ 9.02770023 28.94668457 3.10929781 13.83457346 1.22763611
11.26452997 8.32525049 14.16950039 5.91072393 8.57862944] уровня 0.7 = 11.5
```

4 Построение гистограммы и полигона частот

4.1 Распределение Пуассона

На рисунке 3 представлены 5 графиков, на каждом из которых изображена функция распределения и эмпирическая функция, для $n \in \{5, 10, 100, 1000, 10^5\}$ соответственно.

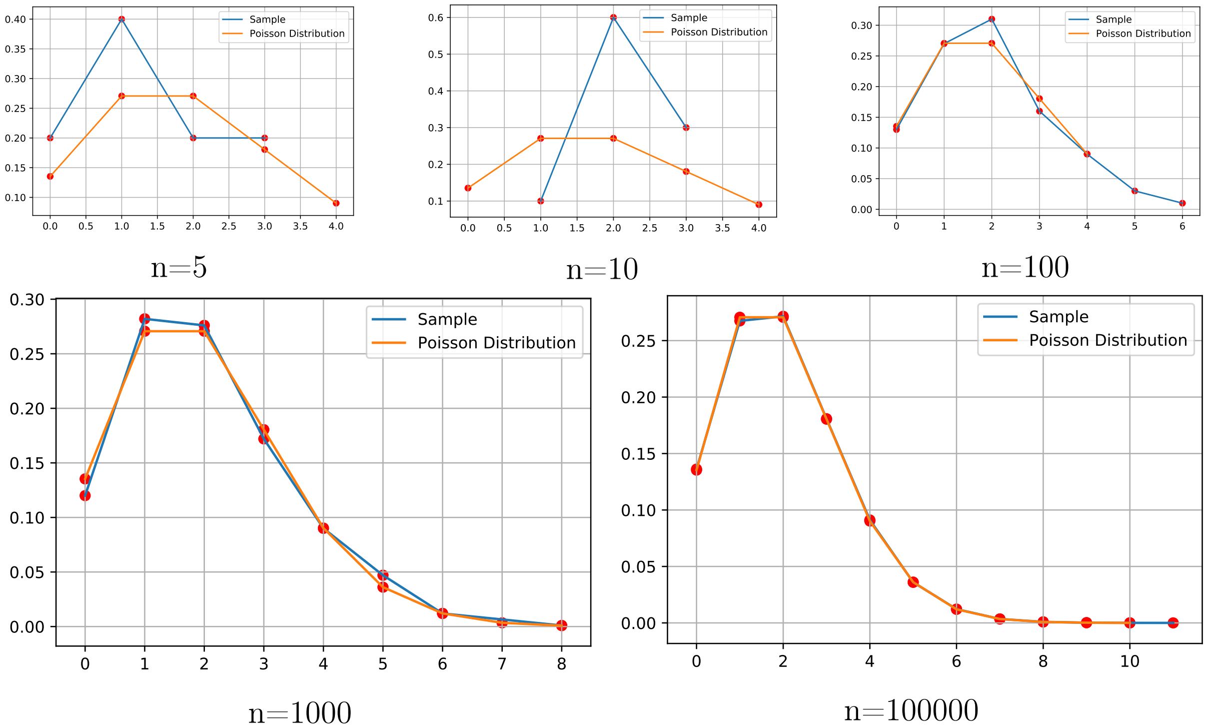


Рис. 3: Сравнение функции распределения и эмпирической функции

4.2 Распределение Эрланга

Для построения диаграмм был выбран шаг в 0.1, что можно заметить на графике. При уменьшении этого показателя можно добиться более «схожей» формы графика распределения и эмпирической функции. 5 графикой представлены на рисунке 4

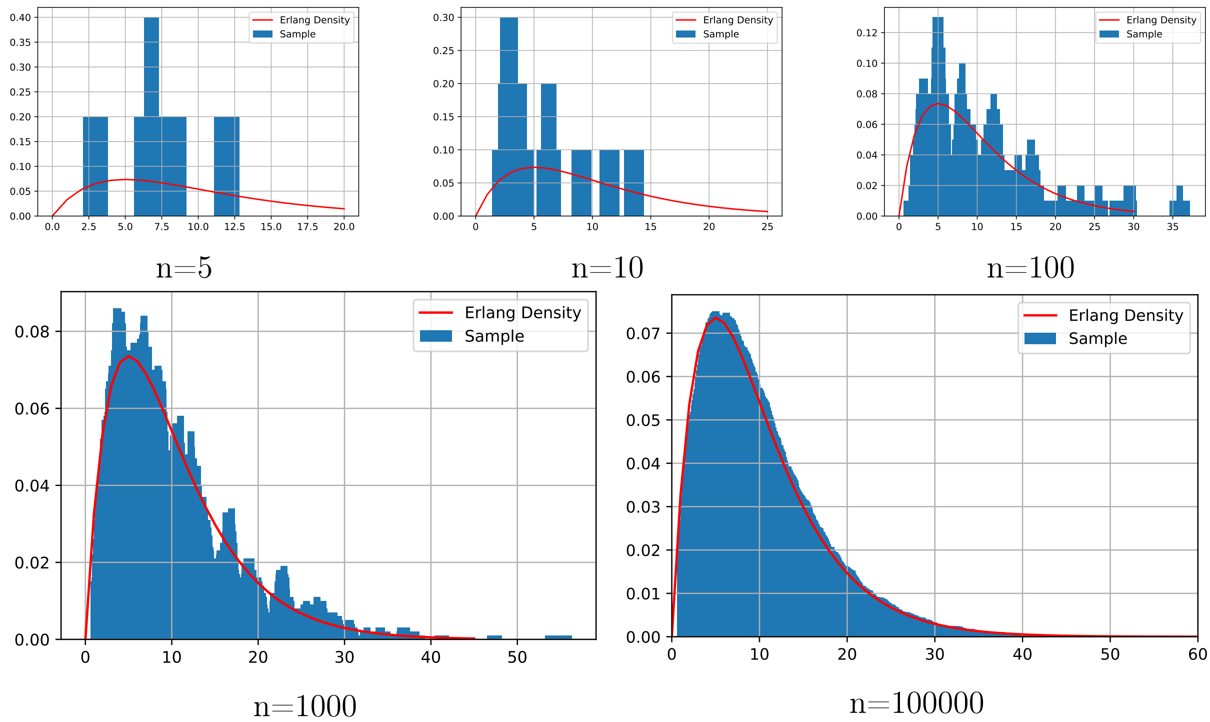


Рис. 4: Сравнение функции распределения и эмпирической функции