МИНИСТЕРСТВО ОБРАЗОВАНИЯ И НАУКИ РОССИЙСКОЙ ФЕДЕРАЦИИ
Федеральное государственное бюджетное образовательное учреждение
высшего образования
«МИРЭА – Российский технологический университет»

ИНСТИТУТ КИБЕРНЕТИКИ

КАФЕДРА ВЫСШЕЙ МАТЕМАТИКИ

Лабораторная работа 1

по курсу «**Теория вероятностей и математическая статистика, часть 2**»

Тема: _______**Первичная обработка выборки из**______________

_______________**дискретной генеральной совокупности**__________

Выполнил:
Студент 3-го курса
Захаров А.М.

Группа: КМБО-01-17

МОСКВА 2020

Задание

Задание 1. Получить выборку, сгенерировав 100 псевдослучайных чисел распределенных по биномиальному закону с параметрами n и p .

Задание 2. Получить выборку, сгенерировав 100 псевдослучайных чисел распределенных по геометрическому закону с параметром p .

Задание 3. Получить выборку, сгенерировав 100 псевдослучайных чисел распределенных по закону Пуассона с параметром λ .

Следуя указаниям для всех выборок построить:

1) статистический ряд;

2) полигон относительных частот;

3) эмпирическую функцию распределения.

Найти:

1) выборочное среднее;

2) выборочную дисперсию;

3) выборочное среднее квадратическое отклонение;

4) моду;

5) медиану;

6) выборочный коэффициент асимметрии;

7) выборочный коэффициент эксцесса.

Все вычисления проводить с точностью до 0,00001 .

Краткие теоретические сведения

Биномиальное распределение:

- ряд распределения: $P(X=k)=C_n^k p^k q^{n-k}$
- математическое ожидание (среднее значение): *np*
- дисперсия: *npq*
- среднее квадратичное отклонение: $\sqrt{npq}$
- мода: *[(n+1)p]*, если *(n+1)p* – дробное; *(n+1)p–0,5* , если *(n+1)p* – целое
- медиана: *Round(np)*
- коэффициент асимметрии: $\frac{q-p}{\sqrt{npq}}$
- коэффициент эксцесса: $\frac{1-6pq}{npq}$

Геометрическое распределение:

- ряд распределения: $P(X=n)=q^n p$
- математическое ожидание (среднее значение): $\frac{q}{p}$, где $q=1-p$
- дисперсия: $\frac{q}{p^2}$, где $q=1-p$
- среднее квадратичное отклонение: $\sqrt{\frac{q}{p^2}}$
- мода: *0*
- медиана: $[\frac{-\ln 2}{\ln q}$¿, если $\frac{\ln 2}{\ln q}$ – дробное; $\frac{-\ln 2}{\ln q}-\frac{1}{2}$, если $\frac{\ln 2}{\ln q}$ – целое
- коэффициент асимметрии: $\frac{2-p}{\sqrt{q}}$
- коэффициент эксцесса: $6+\frac{p^2}{q}$

Распределение Пуассона:

- ряд распределения: $P(X=k)=\frac{\lambda^k}{k!}e^{-\lambda}$
- математическое ожидание (среднее значение): λ
- дисперсия: λ

- среднее квадратичное отклонение: $\sqrt{\lambda}$
- мода: $[\lambda]$
- медиана: $[\lambda + \frac{1}{3} - \frac{0.02}{\lambda}]$
- коэффициент асимметрии: $\frac{1}{\sqrt{\lambda}}$
- коэффициент эксцесса: $\frac{1}{\lambda}$

Средства языка Python 3

В программе используются встроенные средства языка Python 3, а также функции из библиотек NumPy и matplotlib.

- Numpy.random.binomial(n, p, size) — генерирует список длины size случайных значений из биномиального распределения с параметрами n и p, где n — количество испытаний, а p — вероятность успеха.
- Numpy.random.geometric(p, size) — генерирует список длины size случайных значений из геометрического распределения с параметром p.
- Numpy.random.poisson(lam, size) — генерирует список длины size случайных значений из распределения Пуассона с параметром lam.
- List.sort() — сортирует содержимое списка.
- List.count(x) — подсчитывает количество элементов, равных x, содержащихся в списке.
- sum(list) — суммирует элементы списка.
- max(list) — возвращает максимальный элемент из списка.
- Math.comb(n, k) — возвращает количество сочетаний из n по k.
- Matplotlib.pyplot.plot() - строит график.

Результаты расчетов

Задание 1 (биномиальное распределение)

n = 12, p = 0,335

Неупорядоченная выборка (200 чисел):

2	2	5	5	6	3	5	1	3	4
4	2	4	5	4	4	5	6	4	4
3	5	4	6	4	5	6	4	4	2
3	3	3	5	6	5	7	4	5	2
5	7	3	4	6	5	3	1	4	3
4	4	7	4	4	4	5	0	3	4
5	4	2	5	4	3	7	2	4	2
3	3	3	4	3	4	4	4	2	6
6	2	6	1	2	2	4	3	5	3
7	7	6	5	5	4	1	3	5	4
4	4	2	5	6	5	7	2	2	4
2	4	5	2	6	5	5	4	2	6
4	4	4	5	3	2	5	3	7	5
3	3	3	5	1	3	4	3	1	6
4	6	4	5	3	3	4	3	4	2
7	5	5	4	4	4	4	3	8	4
6	5	6	3	4	3	4	2	3	6
6	4	3	2	7	6	5	4	3	5
6	7	2	7	4	5	5	4	3	3
3	5	2	6	5	4	6	3	5	3

Упорядоченная выборка (200 чисел):

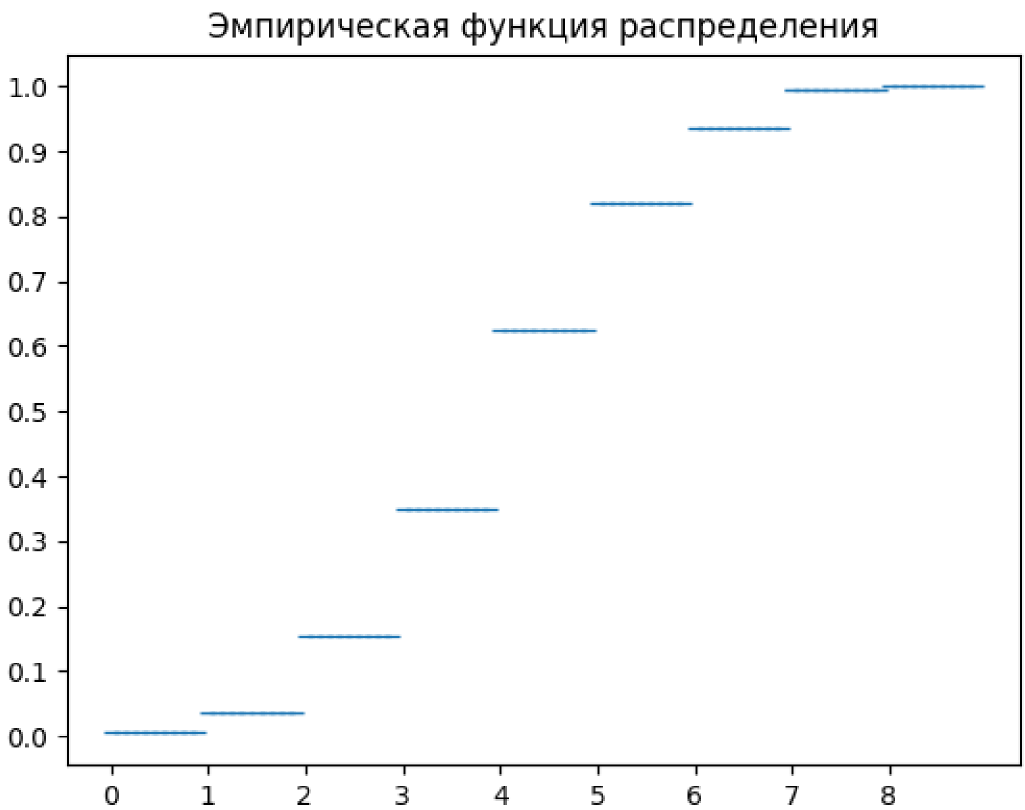
0	1	1	1	1	1	1	2	2	2
2	2	2	2	2	2	2	2	2	2
2	2	2	2	2	2	2	2	2	2
2	3	3	3	3	3	3	3	3	3
3	3	3	3	3	3	3	3	3	3
3	3	3	3	3	3	3	3	3	3
3	3	3	3	3	3	3	3	3	3
4	4	4	4	4	4	4	4	4	4
4	4	4	4	4	4	4	4	4	4
4	4	4	4	4	4	4	4	4	4
4	4	4	4	4	4	4	4	4	4
4	4	4	4	4	4	4	4	4	4
4	4	4	4	4	5	5	5	5	5
5	5	5	5	5	5	5	5	5	5
5	5	5	5	5	5	5	5	5	5
5	5	5	5	5	5	5	5	5	5
5	5	5	5	6	6	6	6	6	6
6	6	6	6	6	6	6	6	6	6
6	6	6	6	6	6	6	7	7	7
7	7	7	7	7	7	7	7	7	8

Статистический ряд:

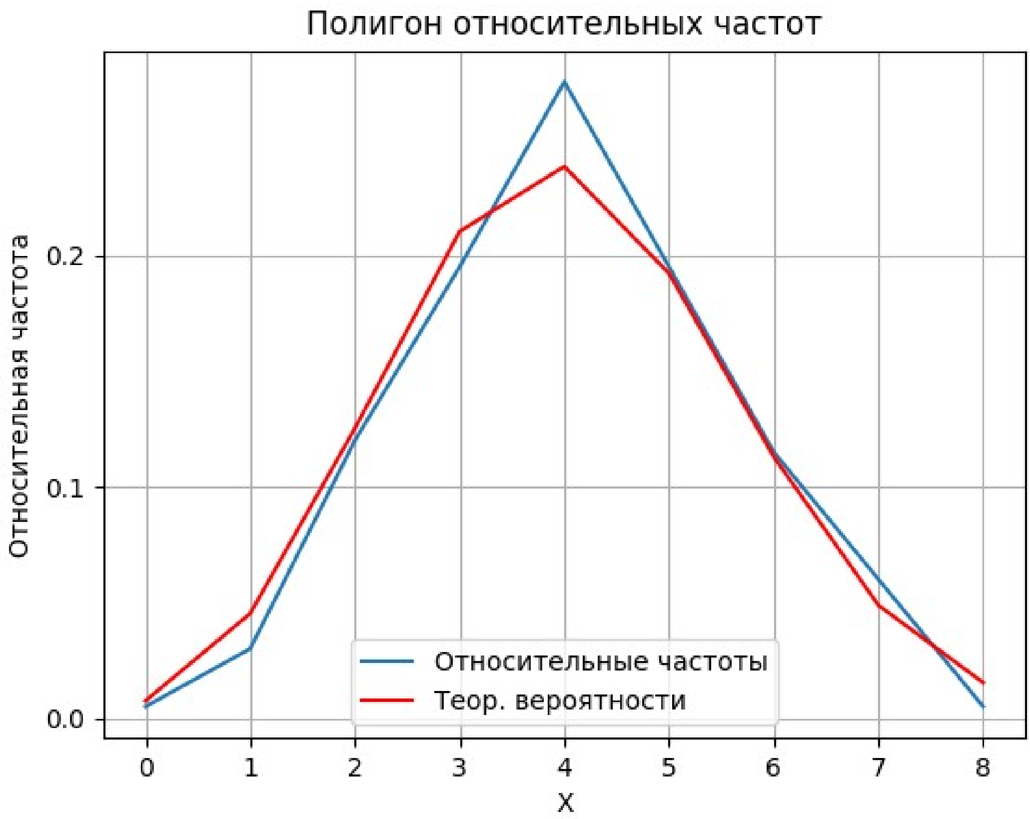
x_k	0	1	2	3	4	5	6	7	8
n_k	1	6	24	39	55	39	23	12	1
w_k	0.005	0.03	0.12	0.195	0.275	0.195	0.115	0.06	0.005
s_k	0.005	0.035	0.155	0.35	0.625	0.82	0.935	0.995	1

Эмпирическая функция распределения и ее график:

$$F_{200}^{Э}(x) = \sum_{xi \le x} w_i = \begin{cases} 0, x < 0 \\ 0.005, 0 \le x < 1 \\ 0.035, 1 \le x < 2 \\ 0.155, 2 \le x < 3 \\ 0.35, 3 \le x < 4 \\ 0.625, 4 \le x < 5 \\ 0.82, 5 \le x < 6 \\ 0.935, 6 \le x \le 7 \\ 0.995, 7 \le x \le 8 \\ 1, x \ge 8 \end{cases}$$



Полигон относительных частот:



Выборочное среднее: 4.08

Выборочная дисперсия: 2.2936

Выборочное среднее квадратическое отклонение: 1.5144636

Выборочная мода: 4

Выборочная медиана: 4

Выборочный коэффициент асимметрии: 0.0630543

Выборочный коэффициент эксцесса: -0.36878

Задание 2 (геометрическое распределение)

p = 0,335

Неупорядоченная выборка (200 чисел):

3	5	2	7	1	1	7	1	1	1
2	1	4	2	8	1	1	7	3	6
1	1	15	1	6	6	2	3	12	1
2	2	4	2	2	5	3	1	1	3
1	3	8	1	2	2	3	3	3	4
6	2	1	1	4	9	1	3	2	3
2	1	3	1	1	5	1	2	2	5
3	3	1	5	4	2	2	2	6	1
3	1	4	5	3	1	2	4	4	3
1	4	5	3	2	1	2	2	1	3
2	1	2	7	6	4	1	6	1	1
2	4	1	1	1	1	1	1	18	3
1	1	1	3	5	2	5	1	2	2
5	1	1	3	1	1	5	1	4	1
3	2	1	2	2	1	5	3	1	1
2	2	2	1	1	2	7	5	11	1
1	3	3	9	9	3	1	2	5	3
2	1	4	1	3	5	4	1	2	2
10	1	2	4	15	2	1	3	3	2
1	5	1	3	8	3	1	3	2	1

Упорядоченная выборка (200 чисел):

1	1	1	1	1	1	1	1	1	1
1	1	1	1	1	1	1	1	1	1
1	1	1	1	1	1	1	1	1	1
1	1	1	1	1	1	1	1	1	1
1	1	1	1	1	1	1	1	1	1
1	1	1	1	1	1	1	1	1	1
1	1	1	1	1	1	1	1	2	2
2	2	2	2	2	2	2	2	2	2
2	2	2	2	2	2	2	2	2	2
2	2	2	2	2	2	2	2	2	2
2	2	2	2	2	2	2	2	2	2
2	3	3	3	3	3	3	3	3	3
3	3	3	3	3	3	3	3	3	3
3	3	3	3	3	3	3	3	3	3
3	3	3	3	3	4	4	4	4	4
4	4	4	4	4	4	4	4	4	4
5	5	5	5	5	5	5	5	5	5
5	5	5	5	5	5	6	6	6	6
6	6	6	7	7	7	7	7	8	8
8	9	9	9	10	11	12	15	15	18

Статистический ряд:

x_k	1	2	3	4	5	6	7
n_k	68	43	34	15	16	7	5
w_k	0.34	0.215	0.17	0.075	0.08	0.035	0.025

s_k	0.34	0.555	0.725	0.8	0.88	0.915	0.94
x_k	8	9	10	11	12	15	18
n_k	3	3	1	1	1	2	1
w_k	0.015	0.015	0.005	0.005	0.005	0.01	0.005
s_k	0.955	0.97	0.975	0.98	0.985	0.995	1

Эмпирическая функция распределения и ее график:

$$F_{200}^{Э}(x)=\sum_{xi\le x} w_i=\begin{cases} 0, x<1 \\ 0.34, 1\le x<2 \\ 0.555, 2\le x<3 \\ 0.725, 3\le x<4 \\ 0.8, 4\le x<5 \\ 0.88, 5\le x<6 \\ 0.915, 6\le x<7 \\ 0.94, 7\le x<8 \\ 0.955, 8\le x<9 \\ 0.97, 9\le x<10 \\ 0.975, 10\le x<11 \\ 0.98, 11\le x<12 \\ 0.985, 12\le x<15 \\ 0.995, 15\le x<18 \\ 1, x\ge 18 \end{cases}$$

Полигон относительных частот:

Выборочное среднее: 3.025

Выборочная дисперсия: 7.13437

Выборочное среднее квадратическое отклонение: 2.67102

Выборочная мода: 1

Выборочная медиана: 2

Выборочный коэффициент асимметрии: 2.465348

Выборочный коэффициент эксцесса: 8.12216

Задание 3 (распределение Пуассона)

λ=0,57

Неупорядоченная выборка (200 чисел):

0	0	0	1	2	1	0	2	1	0
0	0	1	1	0	0	0	0	2	0
1	1	0	1	1	1	3	0	0	0
4	1	0	0	2	1	0	1	1	0
0	0	0	0	0	1	2	0	0	0
1	0	0	0	1	1	0	0	2	0
0	0	0	0	2	0	1	1	3	0
3	0	0	0	2	0	0	1	0	0
0	1	0	0	2	0	1	0	1	0
0	0	0	0	0	0	1	1	0	0
1	1	1	0	1	0	0	0	2	2
0	0	2	0	0	0	1	0	2	0
0	1	1	0	1	0	1	0	0	1
0	1	0	2	0	0	0	0	0	0
3	0	0	0	0	1	0	1	0	1
0	0	0	0	0	1	3	1	3	1
0	0	1	2	0	0	0	0	0	0
0	2	0	0	0	0	1	1	0	1
2	1	0	0	1	0	0	0	1	2
1	1	0	2	0	0	1	1	1	0

Упорядоченная выборка (200 чисел):

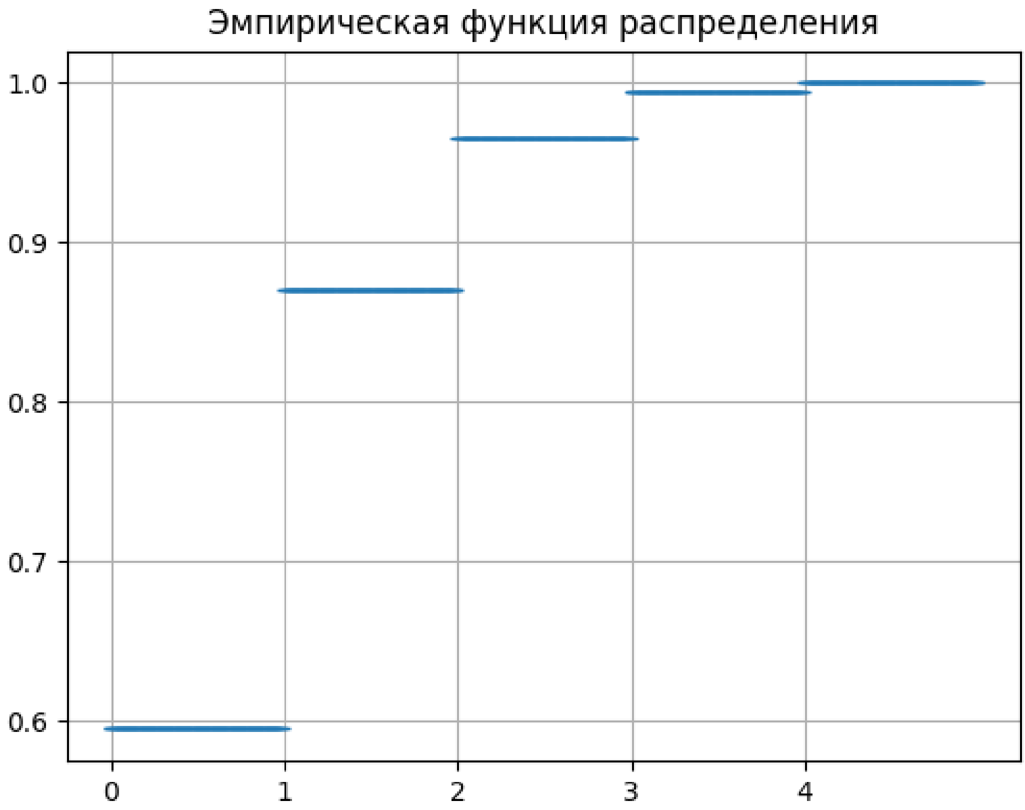
0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	1
1	1	1	1	1	1	1	1	1	1
1	1	1	1	1	1	1	1	1	1
1	1	1	1	1	1	1	1	1	1
1	1	1	1	1	1	1	1	1	1
1	1	1	1	1	1	1	1	1	1
1	1	1	1	2	2	2	2	2	2
2	2	2	2	2	2	2	2	2	2
2	2	2	3	3	3	3	3	3	4

Статистический ряд:

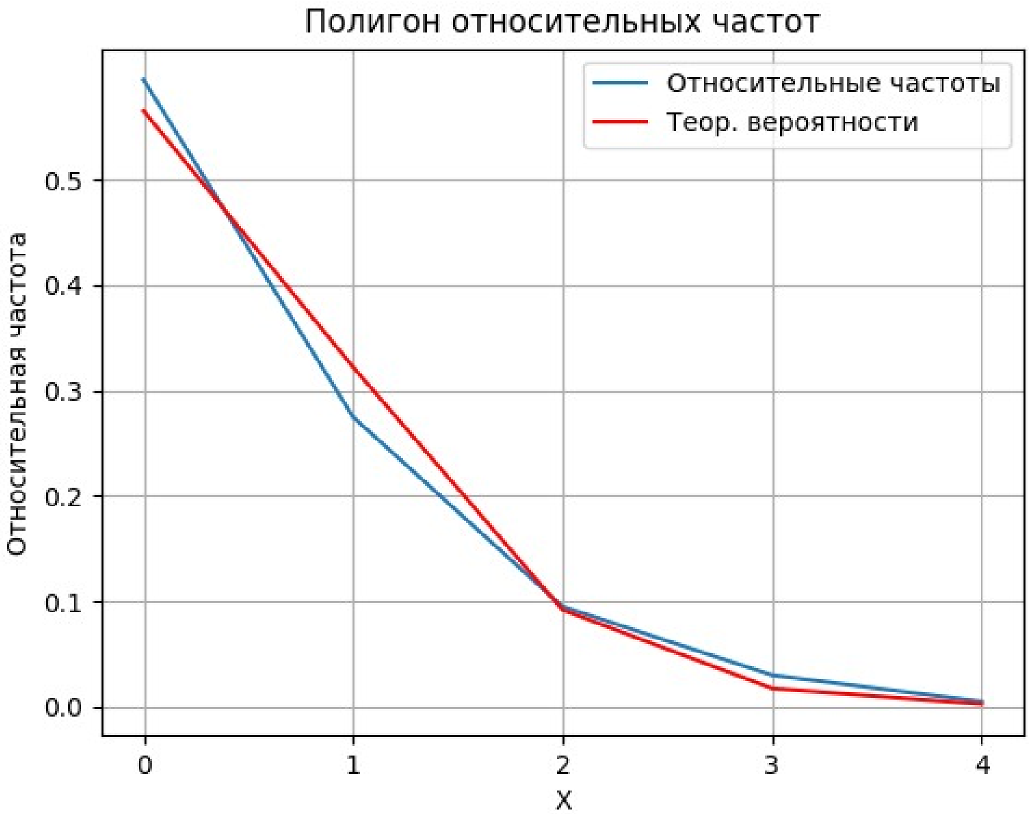
x_k	0	1	2	3	4
n_k	119	55	19	6	1
w_k	0.595	0.275	0.095	0.03	0.005
s_k	0.595	0.87	0.965	0.995	1

Эмпирическая функция распределения и ее график:

$$F_{200}^{Э}(x) = \sum_{xi \le x} w_i = \begin{cases} 0, x < 0 \\ 0.595, 0 \le x < 1 \\ 0.87, 1 \le x < 2 \\ 0.965, 2 \le x < 3 \\ 0.995, 3 \le x < 4 \\ 1, x \ge 4 \end{cases}$$



Полигон относительных частот:



Выборочное среднее: 0.575

Выборочная дисперсия: 0.674375

Выборочное среднее квадратическое отклонение: 0.8212

Выборочная мода: 0

Выборочная медиана: 0

Выборочный коэффициент асимметрии: 1.4655

Выборочный коэффициент эксцесса: 1.8182

Анализ результатов

Задание 1 (биномиальное распределение)

n = 12, p = 0,335

Таблица сравнения относительных частот и теоретических вероятностей:

x_j	w_j	P_j	$\lvert w_j - P_j \rvert$
0	0.005	0.00747	0.00247
1	0.03	0.04521	0.01521
2	0.12	0.12527	0.00527
3	0.195	0.21035	0.01535
4	0.275	0.23842	0.03658
5	0.195	0.19217	0.00283
6	0.115	0.11294	0.00206
7	0.06	0.04876	0.01124
8	0.005	0.01535	0.01035

Таблица сравнения рассчитанных характеристик с теоретическими значениями:

Название показателя	Экспериментальное значение	Теоретическое значение	Абсолютное отклонение	Относительное отклонение
Выборочное среднее	4.08	4.02	0.06	1.4925%
Выборочная дисперсия	2.2936	2.6733	0.3797	14.20341%
Выборочное среднее квадратичное отклонение	1.51446	1.63502	0.12056	7.3736%
Выборочная мода	4	4	0	0%
Выборочная медиана	4	4	0	0%
Выборочный коэффициент асимметрии	0.06305	0.20183	0.13878	68.7608%
Выборочный коэффициент эксцесса	-0.36878	-0.12593	0.24285	192.8452%

Задание 2 (геометрическое распределение)

p = 0,335

Таблица сравнения относительных частот и теоретических вероятностей:

x_j	w_j	P_j	$\lvert w_j - P_j \rvert$
1	0.34	0.50375	0.16375
2	0.215	0.335	0.12
3	0.17	0.22277	0.05277
4	0.075	0.14814	0.07314
5	0.08	0.09851	0.01851
6	0.035	0.06551	0.03351
7	0.025	0.04356	0.01856
8	0.015	0.02897	0.2747
9	0.015	0.01926	0.00426
10	0.005	0.01281	0.00781
11	0.005	0.00852	0.00352
12	0.005	0.00566	0.00066

15	0.01	0.00166	0.00834
18	0.005	0.00048	0.00452

Таблица сравнения рассчитанных характеристик с теоретическими значениями:

Название показателя	Экспериментальное значение	Теоретическое значение	Абсолютное отклонение	Относительное отклонение
Выборочное среднее	3.025	1.98507	1.03993	52.3875%
Выборочная дисперсия	7.13437	5.92559	1.20878	20.3993%
Выборочное среднее квадратичное отклонение	2.67102	2.43425	0.23677	9.7266%
Выборочная мода	1	0	1	∞
Выборочная медиана	2	1	1	100%
Выборочный коэффициент асимметрии	2.46534	2.04175	0.42358	20.7459%
Выб. коэфф. эксцесса	8.12216	6.16875	1.95341	31.6662%

Задание 3 (распределение Пуассона)

λ=0,57

Таблица сравнения относительных частот и теоретических вероятностей:

x_j	w_j	P_j	$\lvert w_j - P_j \rvert$
0	0.595	0.56552	0.02948
1	0.275	0.32234	0.05033
2	0.095	0.09186	0.00314
3	0.03	0.01745	0.01254
4	0.005	0.00248	0.00252

Таблица сравнения рассчитанных характеристик с теоретическими значениями:

Название показателя	Экспериментальное значение	Теоретическое значение	Абсолютное отклонение	Относительное отклонение
Выборочное среднее	0.575	0.57	0.005	0.8771%
Выборочная дисперсия	0.67437	0.57	0.10437	18.3105%
Выборочное среднее квадратичное	0.8212	0.75498	0.06622	8.77109%

отклонение				
Выборочная мода	0	0	0	0%
Выборочная медиана	0	0	0	0%
Выборочный коэффициент асимметрии	1.4655	1.32453	0.14097	10.64302%
Выб. коэфф. эксцесса	1.8182	1.75438	0.06382	3.6377%

Вывод

В ходе лабораторной работы выяснилось, что полученные экспериментальным путем данные соответствуют заданным распределениям, если принимать в расчет отклонения от теоретического значения.

Экспериментальная оценка выборочных показателей может сильно отличаться от теоретического значения, в силу того, что выборки из 200 элементов недостаточно для проведения точных расчетов. С увеличением выборки точность будет улучшаться.

Список литературы

1. Математическая статистика [Электронный ресурс]: метод. указания по выполнению лаб. работ / А.А. Лобузов — М.: МИРЭА, 2017.
2. Боровков А. А. Математическая статистика. − СПб.: Лань, 2010. − 704 с.
3. Гмурман В.Е. Теория вероятностей и математическая статистика. − М.: Юрайт, 2013. − 479 с.
4. Wes McKinney Python for Data Analysis − O'Reilly Media, Inc., 2012

Приложение

generate.py:

```
import numpy as np

V = 7
N = 5 + V % 16
P = 0.3 + 0.005*V
L = 0.5 + 0.01*V

with open("data.txt", 'w') as file:
    a = np.random.binomial(N, P, 200)
    for i in a: file.write(str(i) + " ")
    file.write("\n")
    a = np.random.geometric(P, 200)
    for i in a: file.write(str(i) + " ")
    file.write("\n")
    a = np.random.poisson(L, 200)
    for i in a: file.write(str(i) + " ")
    file.write("\n")
```

lab1.py:

```
import numpy as np
import matplotlib.pyplot as plt
import math

V = 7
N = 5 + V % 16
P = 0.3 + 0.005*V
L = 0.5 + 0.01*V

def make_stat_row(seq):
    seq.sort()
    seq_no_dup = list(set(seq))
    seq_no_dup.sort()
    abs_freq = [seq.count(x) for x in seq_no_dup]
    rel_freq = [x / len(seq) for x in abs_freq]
    rel_freq_sums = [sum(rel_freq[0:i+1]) for i in range(len(rel_freq))]
    return (seq_no_dup, abs_freq, rel_freq, rel_freq_sums)


def plot_rel_freq_poly(seq_no_dup, rel_freq, teor):
    rel_freq_poly = [rel_freq[seq_no_dup.index(x)] if x in seq_no_dup else 0
                     for x in range(seq_no_dup[-1]+1)]

    x = range(seq_no_dup[0], seq_no_dup[-1]+1)
    t = [teor[i] for i in x]
    y = np.arange(0.0, max(max(rel_freq_poly), max(t)), 0.1)
    plt.title("Полигон относительных частот")
    plt.xlabel("X")
    plt.ylabel("Относительная частота")
    plt.xticks(x)
    plt.yticks(y)
    plt.grid()
    plt.plot(x, rel_freq_poly, x, t, "r")
    plt.legend(["Относительные частоты", "Теор. вероятности"])
    plt.show()
```

```python


def sample_mean(seq_no_dup, rel_freq):
    return selective_moment(seq_no_dup, rel_freq, 1)


def selective_moment(seq_no_dup, rel_freq, deg):
    return sum([pow(seq_no_dup[i], deg)*rel_freq[i]
           for i in range(len(rel_freq))])


def dispersion(seq_no_dup, rel_freq):
    return (selective_moment(seq_no_dup, rel_freq, 2) -
           pow(sample_mean(seq_no_dup, rel_freq), 2))


def central_selective_moment(seq_no_dup, rel_freq, deg):
    sm = sample_mean(seq_no_dup, rel_freq)
    return sum([pow(seq_no_dup[i]-sm, deg)*rel_freq[i]
           for i in range(len(rel_freq))])

def standard_deviation(seq_no_dup, rel_freq):
    return np.sqrt(dispersion(seq_no_dup, rel_freq))

def mode(seq_no_dup, abs_freq):
    max_n = max(abs_freq)
    cnt = abs_freq.count(max_n)
    if cnt == 1:
        return seq_no_dup[abs_freq.index(max_n)]
    for i in range(abs_freq.index(max_n)+1, len(abs_freq)):
        if abs_freq[i] != max_n:
            if max_n in abs_freq[i:]:
                return math.nan
            return (seq_no_dup[abs_freq.index(max_n)] + seq_no_dup[i-1])/2

def assym(seq_no_dup, rel_freq):
    return (central_selective_moment(seq_no_dup, rel_freq, 3) /
           (standard_deviation(seq_no_dup, rel_freq)**3))

def exc(seq_no_dup, rel_freq):
    return (central_selective_moment(seq_no_dup, rel_freq, 4) /
           (standard_deviation(seq_no_dup, rel_freq)**4)) - 3

def plot_empiric(seq_no_dup, sum_freq):
    x = range(seq_no_dup[-1]+1)
    plt.xticks(x)
    plt.yticks(np.arange(0, 1.1, 0.1))
    xsp = np.arange(0, x[-1]+1, 0.01)
    ysp = [empiric(i, seq_no_dup, sum_freq) for i in xsp]
    plt.title("Эмпирическая функция распределения")
    plt.grid()
    plt.plot(xsp, ysp, "_")
    plt.show()

def empiric(x, seq_no_dup, sum_freq):
    ind = 0
    for i in range(len(seq_no_dup)):
        if seq_no_dup[i] > x:
            break
        ind += 1
    if ind < 1: return 0
    return sum_freq[ind-1]
```

```python

def teor_geometric(p, seq_no_dup):
    return [p * (1-p)**(i-1) for i in range(seq_no_dup[-1]+1)]

def teor_binomial(n, p):
    return [math.comb(n, i) * p**i * (1-p)**(n-i) for i in range(n+1)]

def teor_poisson(l, seq_no_dup):
    return [math.exp(-l) * l**i / math.factorial(i) for i in seq_no_dup]

with open("data.txt") as file:
    bin, geom, pois = [[int(x) for x in l.split()] for l in file]
```