**ИДЗ №1  
Обработка одномерной выборки. Построение интервального ряда. Точечное и интервальное оценивание параметров распределения.**

**Вариант №7**

*Работу выполнили:*Батманов Даниил, P3207

Шорников Сергей, P3211

**Цель работы:**

1. Построить интервальный ряд исследуемого признака;
2. Полигон частот, выборочную функцию распределения и гистограмму;
3. Найти точечные оценки математического ожидания и дисперсии;
4. Построить доверительные интервалы для математического ожидания и дисперсии с доверительной вероятностью 0,95;
5. Проверить статистическую гипотезу о виде закона распределения.

**Исходные данные:**

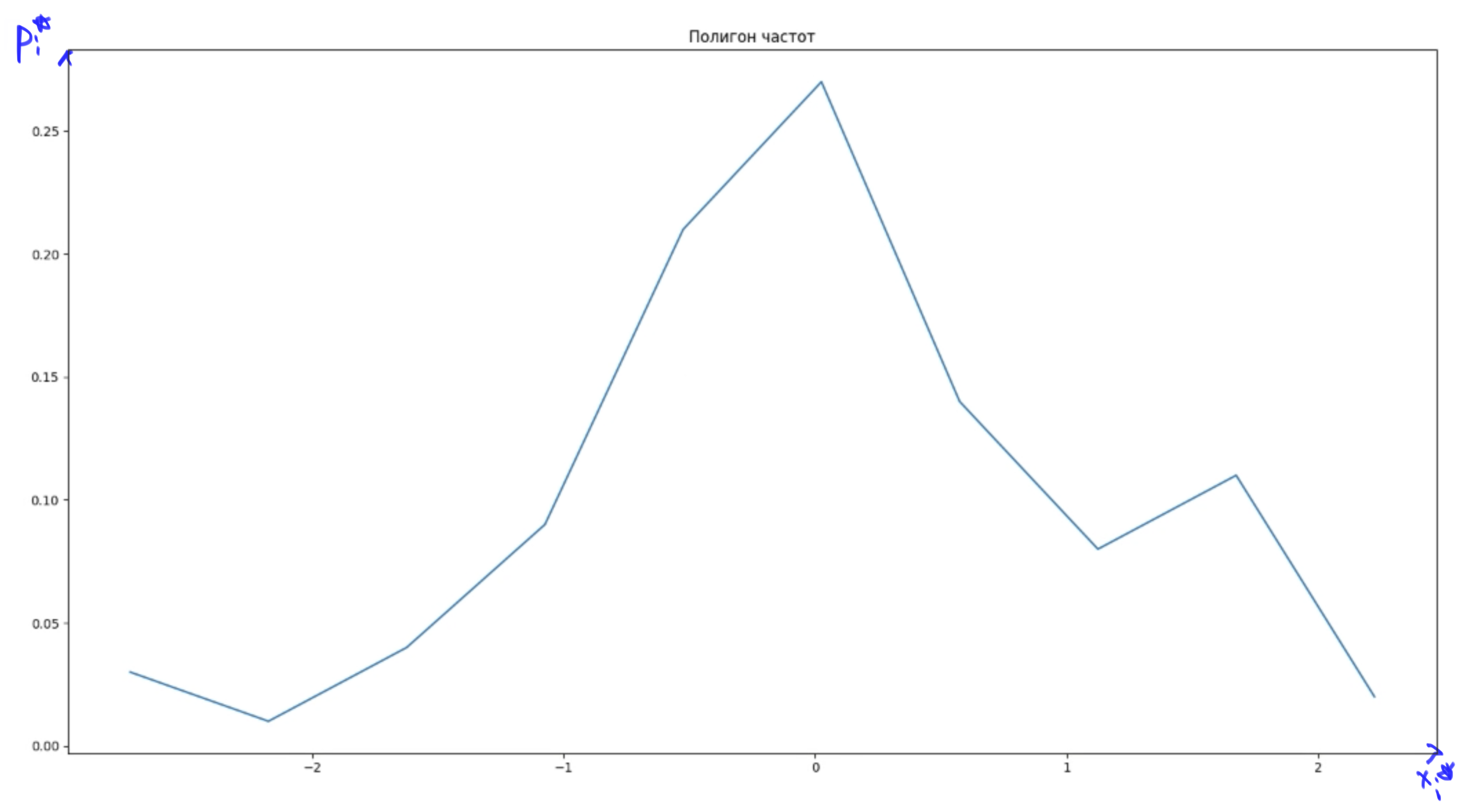
-0.499 1.683 2.247 1.444 -0.418 -2.977 -0.968 -0.308 -1.816 -0.446 1.627 1.555 0.310 -0.074 1.414 1.007 0.555 0.003 -2.789 0.005 -0.239 -1.050 1.991 -0.362 -0.847 0.884 0.759 -1.406 0.262 -0.206 -0.961 0.096 -0.119 -0.777 0.166 -0.405 -0.572 1.624 0.119 0.049 -0.152 0.251 -0.272 -0.250 -0.048 -2.619 1.158 0.139 0.332 0.926 0.350 0.033 0.478 0.637 -0.033 -0.319 0.570 -0.837 -0.413 -1.640 -0.795 -0.015 1.774 -1.568 0.302 -1.120 -0.917 -0.091 1.118 0.277 -0.622 -0.554 -0.470 0.700 -0.656 1.460 1.701 0.630 -0.700 -0.674 1.429 -1.163 -0.925 0.973 -0.052 0.409 -0.024 0.384 -0.350 0.203 -2.084 0.100 0.001 -0.070 0.773 1.132 -0.769 -0.609 1.816 1.307

**Ход работы:**

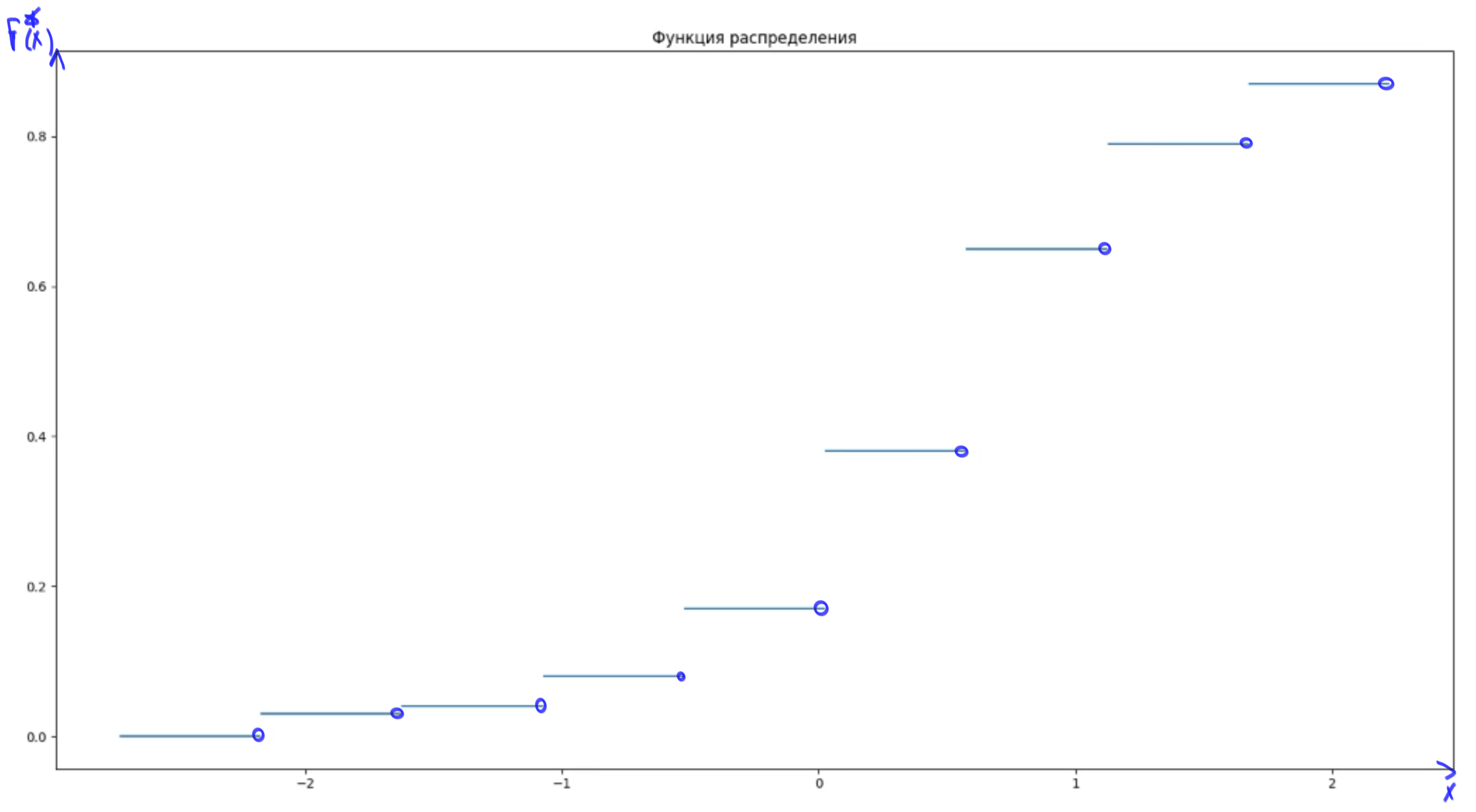
Таблица 1.

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Номер интервала  (x\_i-1, x\_i) | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
| Граница интервалов | [-3, -2.45] | [-2.45, -1.9] | [-1.9, -1.35] | [-1.35, -0.8] | [-0.8, -0.25] | [-0.25, 0.3] | [0.3, 0.85] | [0.85, 1.4] | [1.4, 1.95] | [1.95, 2.5] |
| x\*\_i | -2.725 | -2.175 | -1.625 | -1.075 | -0.525 | 0.025 | 0.575 | 1.125 | 1.675 | 2.225 |
| n\_i | 3 | 1 | 4 | 9 | 21 | 27 | 14 | 8 | 11 | 2 |
| p\*\_i | 0.03 | 0.01 | 0.04 | 0.09 | 0.21 | 0.27 | 0.14 | 0.08 | 0.11 | 0.02 |

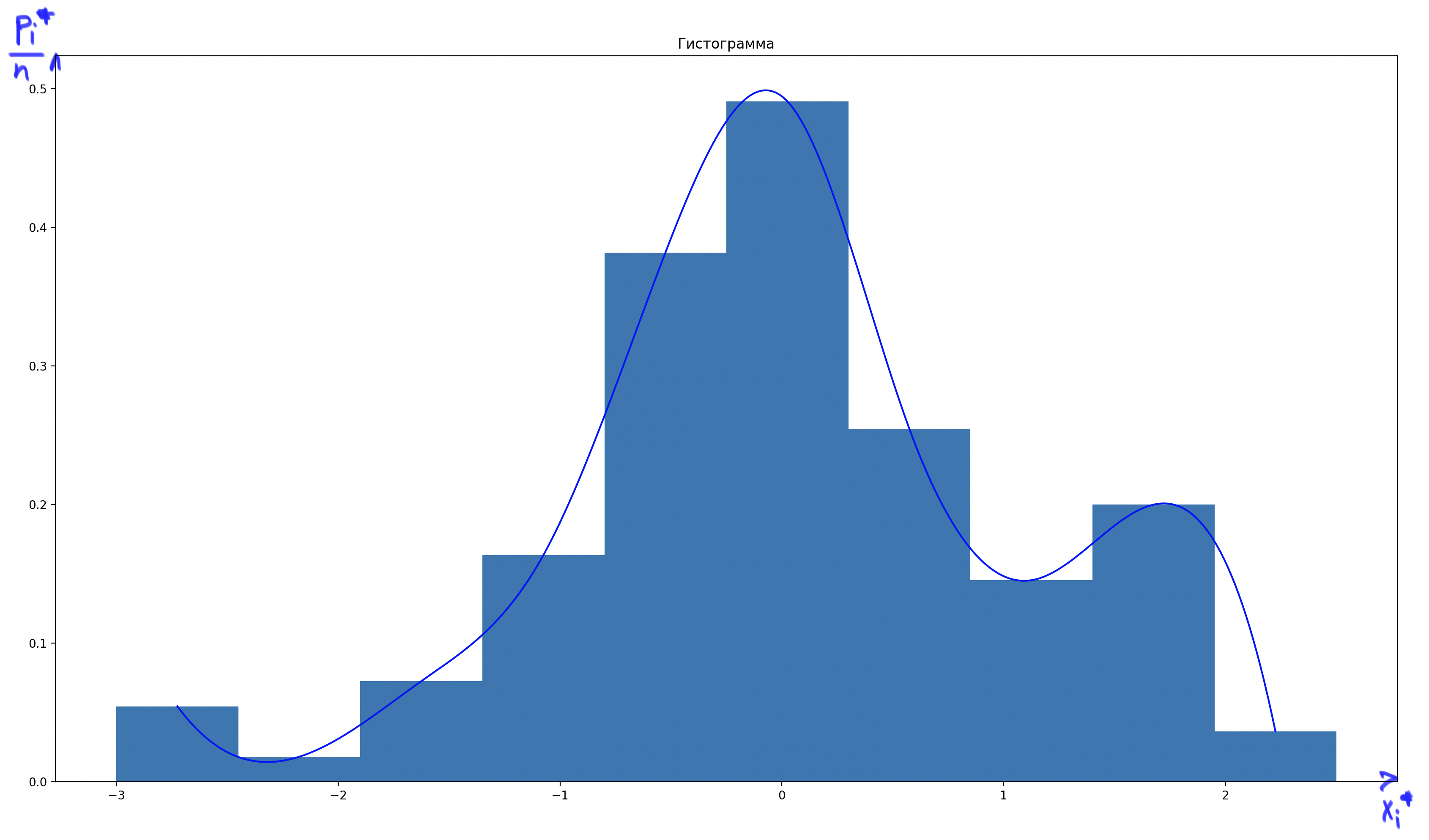
**Рисунок 1. Полигон частот.**



**Рисунок 2. Выборочная функция распределения.**



**Рисунок 3. Гистограмма.**



**Поиск точечных оценок для математического ожидания и дисперсии:**

Таблица 2.

|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Номер интервала | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | Некоторые результаты |
| x\*\_i | -2.725 | -2.175 | -1.625 | -1.075 | -0.525 | 0.025 | 0.575 | 1.125 | 1.675 | 2.225 |  |
| p\*\_i | 0.03 | 0.01 | 0.04 | 0.09 | 0.21 | 0.27 | 0.14 | 0.08 | 0.11 | 0.02 |  |
| x*ip*i | -0.08175 | -0.02175 | -0.065 | -0.09675 | -0.11025 | 0.00675 | 0.0805 | 0.09 | 0.18425 | 0.0445 | \_X =  0.0305 |
| x*2ip*i | 0.222769 | 0.047306 | 0.105625 | 0.104006 | 0.057881 | 0.000169 | 0.046288 | 0.10125 | 0.308619 | 0.099013 | m2 = 1,093 |

\_X = 0.0305

S^2 = m2 - X^2 = 0,999975

**Доверительные интервалы:**

Округляя полученные результаты, принимаем *X* = 0,03; *S^2=* 1

Для рассматриваемого примера будем иметь при гамма = 0,95, Ф(t ) = 0,975,

откуда *t* =1,95, поэтому в нашем примере имеем

Таким образом, доверительный интервал для математического ожидания

имеет вид

Доверительный интервал для дисперсии

**Проверить статистическую гипотезу о виде закона распределения**

Таблица 3.

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Номер интервала  (x\_i-1, x\_i) | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
| Граница интервалов | [-3, -1.35] | [-1.35, -0.8] | [-0.8, -0.25] | [-0.25, 0.3] | [0.3, 0.85] | [0.85, 1.4] | [1.4, 2.5] |
| x\*\_i | -2.175 | -1.075 | -0.525 | 0.025 | 0.575 | 1.125 | 1.95 |
| n\_i | 8 | 9 | 21 | 27 | 14 | 8 | 13 |
| p\*\_i | 0.08 | 0.09 | 0.21 | 0.27 | 0.14 | 0.08 | 0.13 |

|  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | | inf; | |  | |  | |  |  |  | ; inf | примечание |
| x | | -2.175 | | -1.075 | | -0.525 | | 0.025 | 0.575 | 1.125 | 1.95 |  |
| z | | -2.205 | | -1.105 | | -0.555 | | -0.005 | 0.545 | 1.095 | inf |  |
| Ф | | 0.0136 | | 0.1335 | | 0.2912 | | 0.5 | 0.5239 | 0.8686 | 1 |  |
| pi | | 0.0136 | | 0.1199 | | 0.1577 | | 0.2088 | 0.0239 | 0.3447 | 0.1314 | 1 |
| ni | | 8 | | 9 | | 21 | | 27 | 14 | 8 | 13 | 100 |
| ni^2 | | 64 | | 81 | | 441 | | 729 | 196 | 64 | 169 |  |
| npi | | 1.36 | | 11.99 | | 15.77 | | 20.88 | 2.39 | 34.47 | 13.14 | 100 |
|  | |  | |  | |  | |  |  |  |  |  |
| ni^2/npi | | 32.41882 | | 0.74563 | | 1.73449 | | 1.793793 | 56.39837 | 20.32669 | 0.001492 | 107.97 |
|  |  | |  | |  | |

**Вывод:**

Так как , значит гипотеза о нормальном распределении генеральной совокупности не отвергается.

*Исходный код:*

import math  
import numpy as np  
from scipy.interpolate import make\_interp\_spline  
import matplotlib.pyplot as plt  
from scipy.stats import norm  
  
def empirical\_func(k):  
 frequency\_count = 0  
 for j in range(intervals\_count):  
 if not (intervals[j][0] < k <= intervals[j][1]):  
 frequency\_count += freqs[j]  
 else:  
 break  
 return frequency\_count  
  
  
data = [  
 -0.499, 1.683, 2.247, 1.444, -0.418, -2.977, -0.968, -0.308, -1.816, -0.446,  
 1.627, 1.555, 0.310, -0.074, 1.414, 1.007, 0.555, 0.003, -2.789, 0.005,  
 -0.239, -1.050, 1.991, -0.362, -0.847, 0.884, 0.759, -1.406, 0.262, -0.206,  
 -0.961, 0.096, -0.119, -0.777, 0.166, -0.405, -0.572, 1.624, 0.119, 0.049,  
 -0.152, 0.251, -0.272, -0.250, -0.048, -2.619, 1.158, 0.139, 0.332, 0.926,  
 0.350, 0.033, 0.478, 0.637, -0.033, -0.319, 0.570, -0.837, -0.413, -1.640,  
 -0.795, -0.015, 1.774, -1.568, 0.302, -1.120, -0.917, -0.091, 1.118, 0.277,  
 -0.622, -0.554, -0.470, 0.700, -0.656, 1.460, 1.701, 0.630, -0.700, -0.674,  
 1.429, -1.163, -0.925, 0.973, -0.052, 0.409, -0.024, 0.384, -0.350, 0.203,  
 -2.084, 0.100, 0.001, -0.070, 0.773, 1.132, -0.769, -0.609, 1.816, 1.307  
]  
  
n = len(data)  
sorted\_data = sorted(data)  
intervals\_count = int(n \*\* 0.5)  
intervals = []  
bottom = -3  
top = 2.5  
interval\_length = (top - bottom) / intervals\_count  
  
print('Номер', end='\t')  
print(\*[i for i in range(1, intervals\_count + 1)], sep='\t')  
  
interval\_bounds = [bottom]  
  
for i in range(intervals\_count):  
 top = round(bottom + interval\_length, 5)  
 interval\_bounds.append(top)  
 intervals.append([bottom, top])  
 bottom = top  
  
print('Границы', end='\t')  
print(\*intervals, sep='\t')  
  
mid\_points = [round(i[0] + i[1], 5) / 2 for i in intervals]  
print('Середины', end='\t')  
print(\*mid\_points, sep='\t')  
  
m = []  
for interval in intervals:  
 count = sum(1 for x in data if interval[0] <= x < interval[1])  
 m.append(count)  
  
print('Попадания', end='\t')  
print(\*m, sep='\t')  
  
freqs = [x / n for x in m]  
print('Частота', end='\t')  
print(\*freqs, sep='\t')  
  
plt.plot(mid\_points, freqs)  
plt.title('Полигон частот')  
plt.show()  
  
empirical\_values = [empirical\_func(x) for x in mid\_points]  
for i in range(len(mid\_points) - 1):  
 plt.hlines(empirical\_values[i], mid\_points[i], mid\_points[i + 1])  
plt.title('Функция распределения')  
plt.show()  
  
h = [x / interval\_length for x in freqs]  
xnew = np.linspace(min(mid\_points), max(mid\_points), 300)  
spl = make\_interp\_spline(mid\_points, h, k=3)  
smooth = spl(xnew)  
plt.bar(mid\_points, h, interval\_length)  
plt.plot(xnew, smooth, color='blue')  
  
plt.title('Гистограмма')  
plt.show()  
  
print("Поиск точечных оценок для Мат ожидания и Дисперсии:\n")  
  
s2 = 0  
xLine = 0  
arrayXP = []  
arrayX2P = []  
for i in range(len(m)):  
 arrayXP.append(mid\_points[i] \* freqs[i])  
 arrayX2P.append(mid\_points[i]\*mid\_points[i] \* freqs[i])  
 xLine += mid\_points[i] \* freqs[i]  
 s2 += mid\_points[i]\*mid\_points[i] \* freqs[i]  
s2 -= xLine \* xLine  
  
print('x\*ip\*i: ', end='\t')  
print(\*arrayXP, sep='\t')  
print('x\*2ip\*i: ', end='\t')  
print(\*arrayX2P, sep='\t')  
  
print("Точная оценка Мат ожидания: ", xLine)  
print("Точная оценка Дисперсии: ", s2)  
  
print("Доверительные интервалы:\n")  
  
  
# t - величина из таблицы Ф(t) и y  
# o - СКО  
t = 1.83  
o = s2 \*\* 0.5  
print("Доверительный интервал для Мат ожидания: ", xLine - t \* o / math.sqrt(len(data)), ";", xLine + t \* o / math.sqrt(len(data)))  
print("Доверительный интервал для дисперсии: ", (len(mid\_points) - 1) \* s2 / 19, ";", (len(mid\_points) - 1) \* s2 / 2.7)  
print(len(mid\_points))  
  
print("Проверка гипотезы о виде закона распределения (H0 - нормальное распределение)")  
mean = xLine  
std = np.std(data, ddof=1)  
  
# Разбиваем данные на равные интервалы  
num\_intervals = int(np.sqrt(len(data)))  
  
# Считаем наблюдаемые частоты  
observed\_frequencies = np.histogram(data, bins=interval\_bounds)[0]  
  
  
  
# Вычисляем ожидаемые частоты  
expected\_frequencies = []  
for i in range(num\_intervals):  
 low = interval\_bounds[i]  
 high = interval\_bounds[i + 1]  
 expected\_freq = len(data) \* (norm.cdf(high, mean, std) - norm.cdf(low, mean, std))  
 expected\_frequencies.append(expected\_freq)  
  
# Объединяем интервалы, если ожидаемая частота меньше 5  
combined\_observed = []  
combined\_expected = []  
i = 0  
while i < len(expected\_frequencies):  
 current\_observed = observed\_frequencies[i]  
 current\_expected = expected\_frequencies[i]  
  
 while i < len(expected\_frequencies) - 1 and current\_expected < 7:  
 i += 1  
 current\_expected += expected\_frequencies[i]  
 current\_observed += observed\_frequencies[i]  
  
 combined\_observed.append(current\_observed)  
 combined\_expected.append(current\_expected)  
 i += 1  
  
  
  
# Проверяем последний интервал  
if combined\_expected[-1] < 7:  
 # Если последний интервал имеет ожидаемую частоту меньше 7, объединяем его с предыдущим  
 combined\_expected[-2] += combined\_expected[-1]  
 combined\_observed[-2] += combined\_observed[-1]  
 combined\_expected.pop()  
 combined\_observed.pop()  
  
  
# Обновляем количество интервалов после объединения  
num\_intervals = len(combined\_observed)  
  
# Вычисляем значение критерия хи-квадрат  
chi\_squared\_stat = sum((np.array(combined\_observed) - np.array(combined\_expected)) \*\* 2 / np.array(combined\_expected))  
degrees\_of\_freedom = num\_intervals - 3  
  
  
chi\_squared\_table = 9.488  
  
  
print("Количество степеней свободы: ", degrees\_of\_freedom)  
print("Хи квадрат наблюдаемое: ", chi\_squared\_stat)  
print("Хи квадрат табличное: ", chi\_squared\_table)  
  
print("Гипотеза H0 подтверждена" if chi\_squared\_stat < chi\_squared\_table else "Гипотеза H0 не подтверждена")

Вывод программы:

Номер 1 2 3 4 5 6 7 8 9 10

Границы [-3, -2.45] [-2.45, -1.9] [-1.9, -1.35] [-1.35, -0.8] [-0.8, -0.25] [-0.25, 0.3] [0.3, 0.85] [0.85, 1.4] [1.4, 1.95] [1.95, 2.5]

Середины -2.725 -2.175 -1.625 -1.075 -0.525 0.025 0.575 1.125 1.675 2.225

Попадания 3 1 4 9 21 27 14 8 11 2

Частота 0.03 0.01 0.04 0.09 0.21 0.27 0.14 0.08 0.11 0.02

Поиск точечных оценок для Мат ожидания и Дисперсии:

x\*ip\*i: -0.08175 -0.02175 -0.065 -0.09674999999999999 -0.11025 0.006750000000000001 0.0805 0.09 0.18425 0.044500000000000005

x\*2ip\*i: 0.22276875 0.047306249999999994 0.105625 0.10400624999999998 0.05788125 0.00016875000000000004 0.046287499999999995 0.10125 0.30861875 0.09901250000000002

Точная оценка Мат ожидания: 0.030499999999999992

Точная оценка Дисперсии: 1.09199475

Доверительные интервалы:

Доверительный интервал для Мат ожидания: -0.16073235129744654 ; 0.22173235129744653

Доверительный интервал для дисперсии: 0.5172606710526315 ; 3.6399824999999995

10

Проверка гипотезы о виде закона распределения (H0 - нормальное распределение)

Количество степеней свободы: 4

Хи квадрат наблюдаемое: 7.971859323974494

Хи квадрат табличное: 9.488

Гипотеза H0 подтверждена