Министерство науки и высшего образования
Российской Федерации

Федеральное государственное бюджетное образовательное
учреждение высшего образования
«НОВОСИБИРСКИЙ ГОСУДАРСТВЕННЫЙ ТЕХНИЧЕСКИЙ УНИВЕРСИТЕТ»

Кафедра прикладной математики

Лабораторная работа

Устойчивые методы оценивания параметров статистических моделей (логистическое распределение)

Факультет: ПМИ

Группа: ПММ-03

Студент: Москалев Дмитрий

Новосибирск

2020

1. Генерация выборки размером 1000 и 10000 элементов, где ε = 0.25, x - случайные значения в диапазоне (0;1), $y = \mu + s * \ln(\frac{x}{1-x})$, $y_{2,\dots} = \mu_{2,\dots} + s_{2,\dots} * \ln(\frac{x}{1-x})$, $y_{2\_z,\dots} = (1-\varepsilon) * f(x) + \varepsilon * h(x)$, где $f(x) = \frac{e^{-x}}{[1+e^{-x}]^2}$, μ – параметр сдвига, s – параметр масштаба, где $0 < \varepsilon < 0.5, f(x), h(x)$ – плотности чистого и засоряющего распределений. Параметры μ, s = 0, 1 соответственно генерируют стандартные величины, при этом засоренные случайные величины генерируется с μ, s = 2, 1; μ, s = 0, 2 соответственно.

Выборка 1000 значений:

μ, s = 2, 1

№	x	y	y_2	y_{2_z}
1	0.25	-1.12	0.88	0.31
2	0.08	-2.43	-0.43	-0.02
3	0.6	0.41	2.41	0.72

...

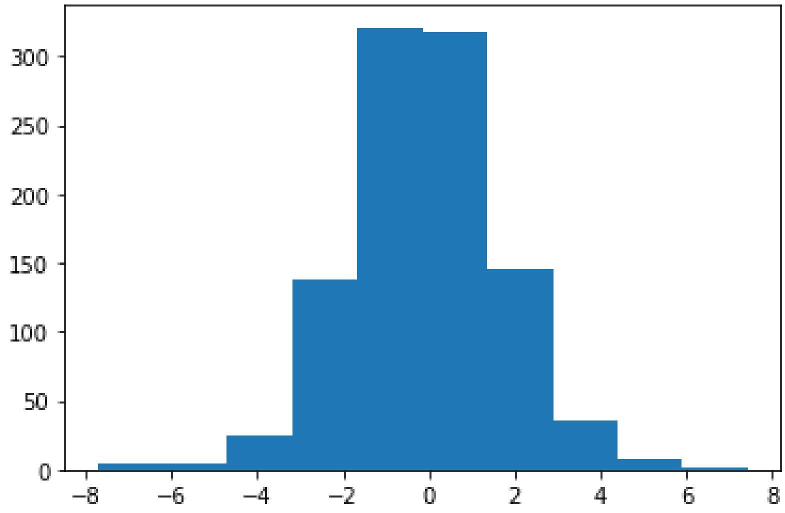
998	0.8	1.41	3.41	0.99
999	0.06	-2.67	-0.67	-0.09
1000	0.25	-1.08	0.92	0.32

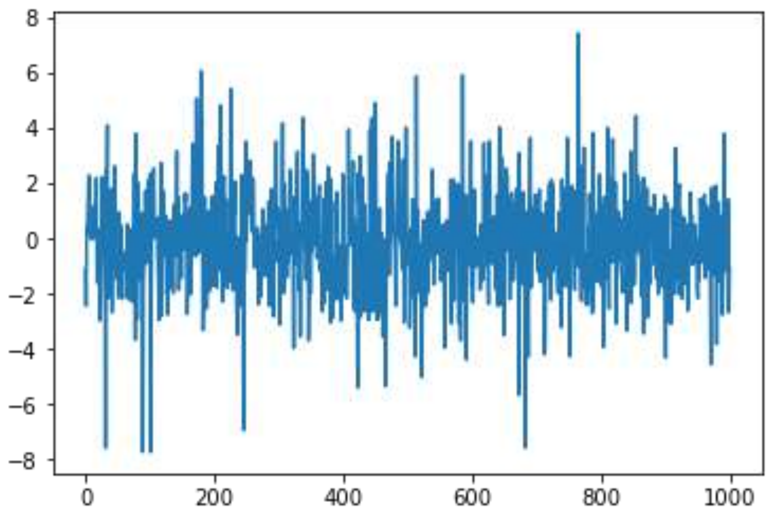
μ, s = 0, 2

№	x	y	y_3	y_{3_z}
1	0.25	-1.12	-2.25	-0.47
2	0.08	-2.43	-4.86	-1.12
3	0.6	0.41	0.83	0.3

...

998	0.8	1.41	2.83	0.8
999	0.06	-2.67	-5.35	-1.24
1000	0.25	-1.08	-2.16	-0.45





Выборка 10000 значений:

μ, s = 2, 1

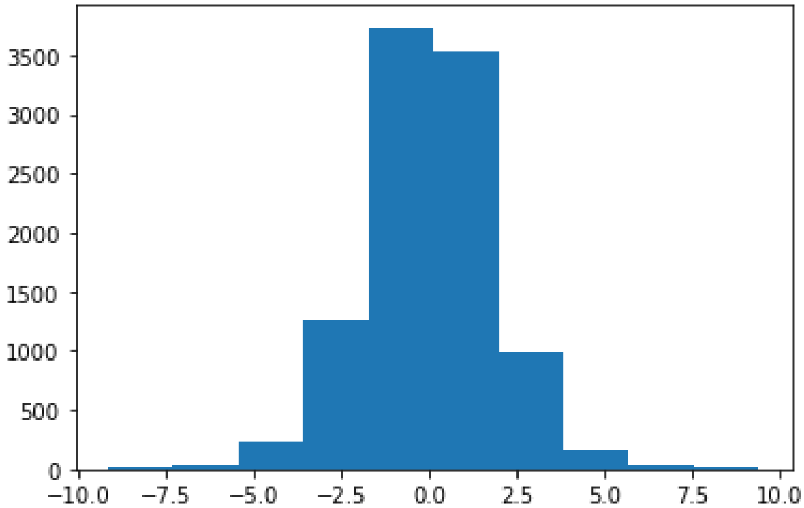
№	x	y	y_2	y_{2_z}
1	0.87	1.94	3.94	1.12
2	0.98	4.11	6.11	1.67
3	0.94	2.83	4.83	1.35

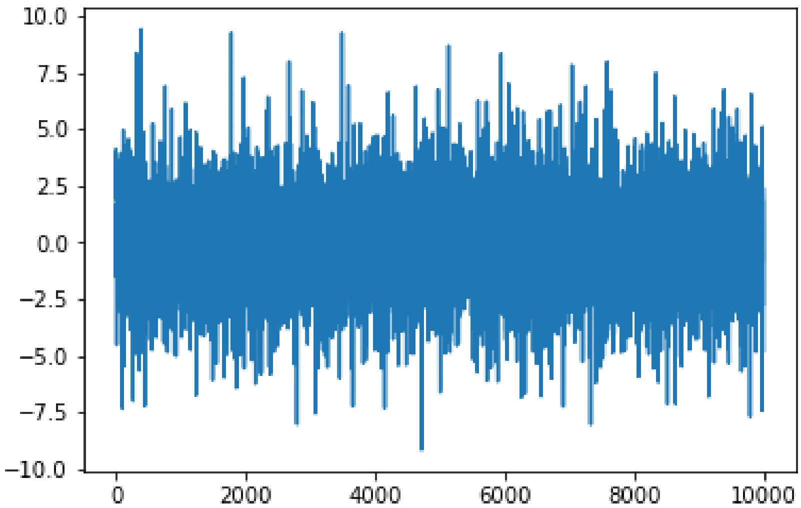
…

9998	0.6	0.4	2.4	0.72
9999	0.89	2.04	4.04	1.15
10000	0.06	-2.73	-0.73	-0.1

μ, s = 0, 2

№	x	y	y_3	y_{3_z}
1	0.87	1.94	3.88	1.06
2	0.98	4.11	8.22	2.14
3	0.94	2.83	5.66	1.5
…				
9998	0.6	0.4	0.8	0.29
9999	0.89	2.04	4.08	1.11
10000	0.06	-2.73	-5.46	-1.27





2. Основные характеристики логистического распределения:

Характеристика	Значение
медиана	μ
дисперсия	$\frac{(\pi^2 * s^2)}{3}$
коэффициенты асимметрии	0
коэффициенты эксцесса	4.2

По условию задания случайные величины в выборке (x) распределены равномерно на (0,1).

Вычисление значений из таблицы производились с помощью MS Excel, с помощью функций:

=СРЗНАЧ();

=МЕДИАНА();

=ДИСП.В();

=СКОС();

=ЭКСЦЕСС();

=СТАНДОТКЛОН.В();

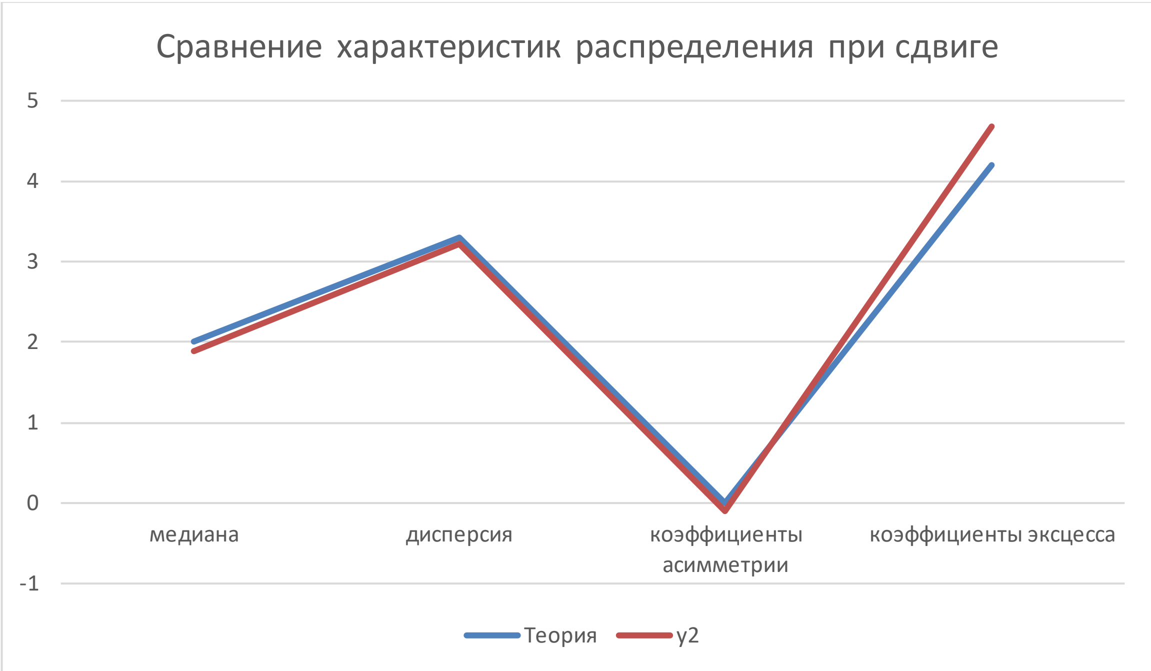
Выборка 1000 значений:

μ, s = 2, 1

Характеристика	Значение
медиана	μ=2
дисперсия	$\frac{(\pi^2 * s^2)}{3} = [s = 1] = \frac{\pi^2}{3} \sim 3.289868133696453$
коэффициенты асимметрии	0
коэффициенты эксцесса	4.2

Характеристика	x	y	y_2	y2_z
среднее арифметическое	0,48	-0,09	1,91	0,59
медиана	0,47	-0,12	1,89	0,58

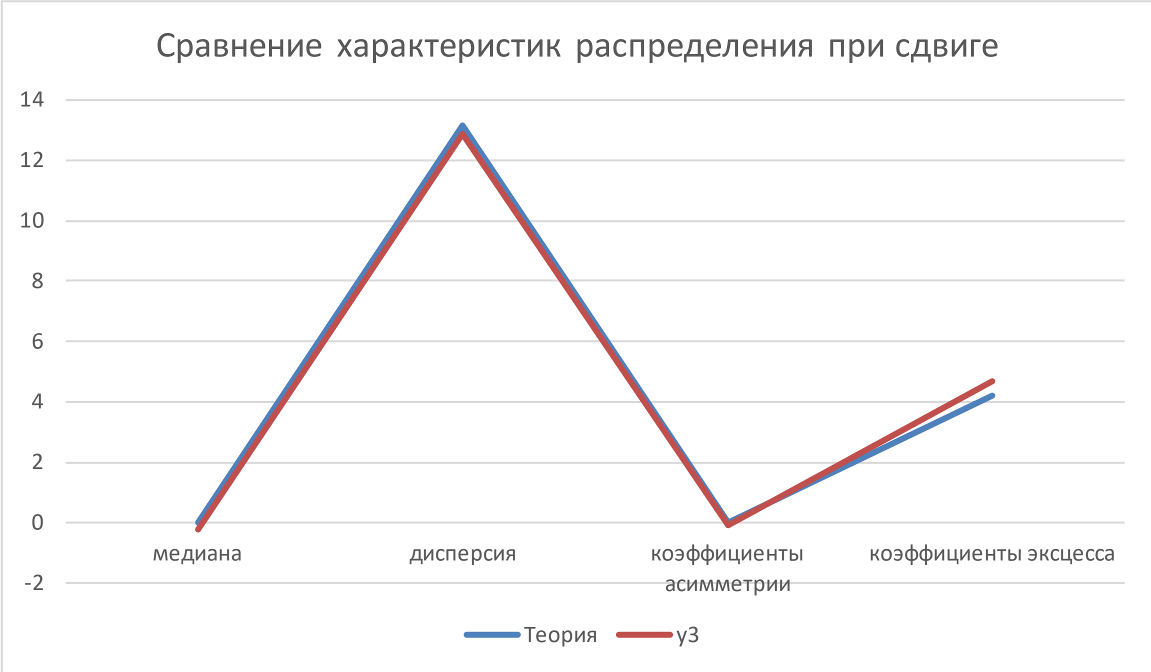
дисперсия		3,22	3,22	0,22
коэффициенты асимметрии		-0,10	-0,10	-0,08
коэффициенты эксцесса		4,68	4,68	4,44
стандартное отклонение		1,79	1,79	0,47



μ, s = 0, 2

Характеристика	Значение
медиана	$\mu = 0$
дисперсия	$\frac{(\pi^2 * s^2)}{3} = [s = 2] = \frac{4 * \pi^2}{3} \sim 13.159472534785811$
коэффициенты асимметрии	0
коэффициенты эксцесса	4.2

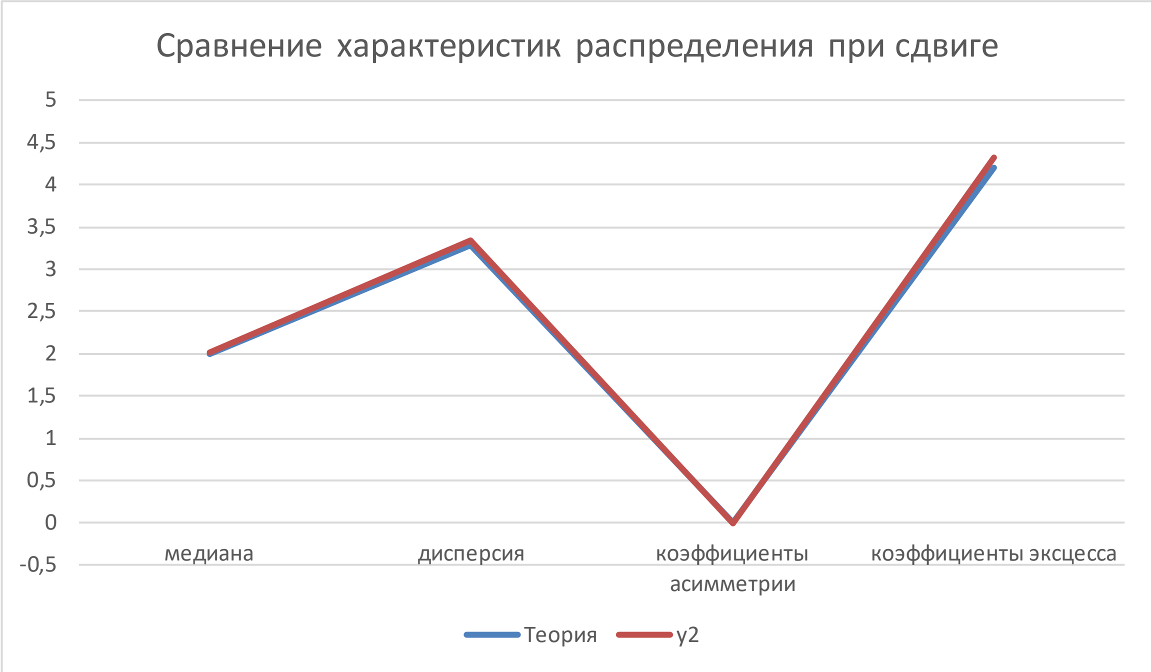
Характеристика	x	y	y3	y3_z
среднее арифметическое	0,48	-0,09	-0,18	0,05
медиана	0,47	-0,12	-0,23	0,04
дисперсия		3,22	12,88	0,80
коэффициенты асимметрии		-0,10	-0,10	-0,10
коэффициенты эксцесса		4,68	4,68	4,69
стандартное отклонение		1,79	3,59	0,90



Выборка 10000 значений:

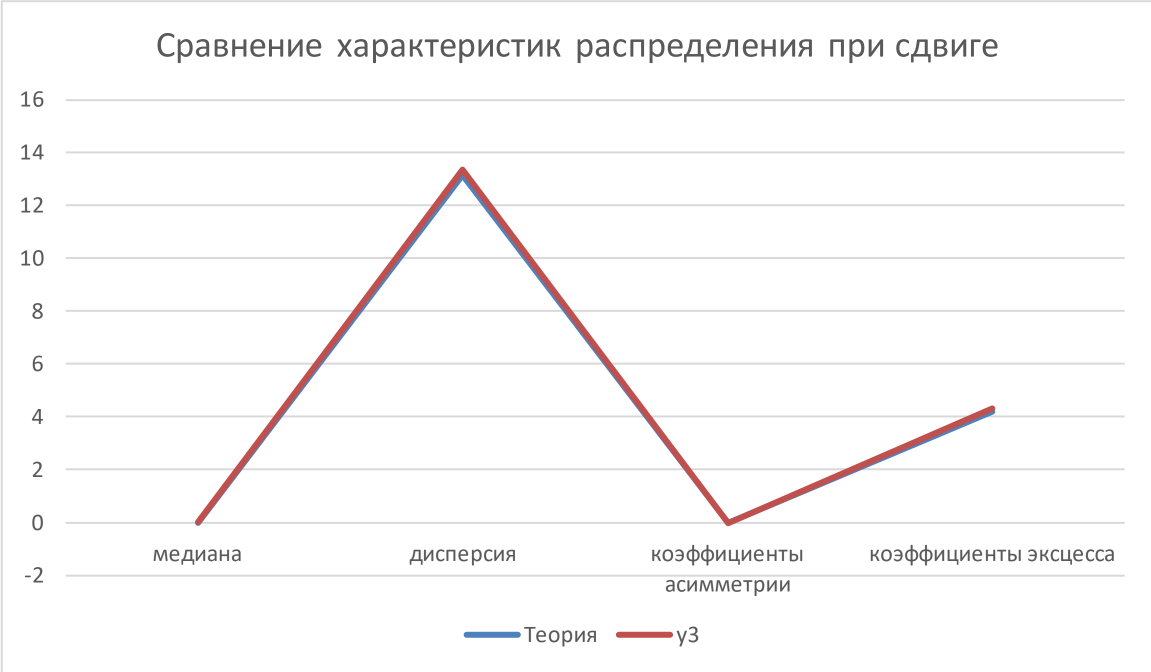
μ, s = 2, 1

Характеристика	x	y	y_2	y_{2_z}
среднее арифметическое	0,50	-0,02	1,98	0,61
медиана	0,50	0,01	2,01	0,61
дисперсия		3,34	3,34	0,23
коэффициенты асимметрии		-0,01	-0,01	-0,01
коэффициенты эксцесса		4,32	4,32	4,12
стандартное отклонение		1,83	1,83	0,48



μ, s = 0, 2

Характеристика	x	y	y_3	y_{3_z}
среднее арифметическое	0,50	-0,02	-0,04	0,08
медиана	0,50	0,01	0,02	0,10
дисперсия		3,34	13,37	0,83
коэффициенты асимметрии		-0,01	-0,01	-0,01
коэффициенты эксцесса		4,32	4,32	4,33
стандартное отклонение		1,83	3,66	0,91



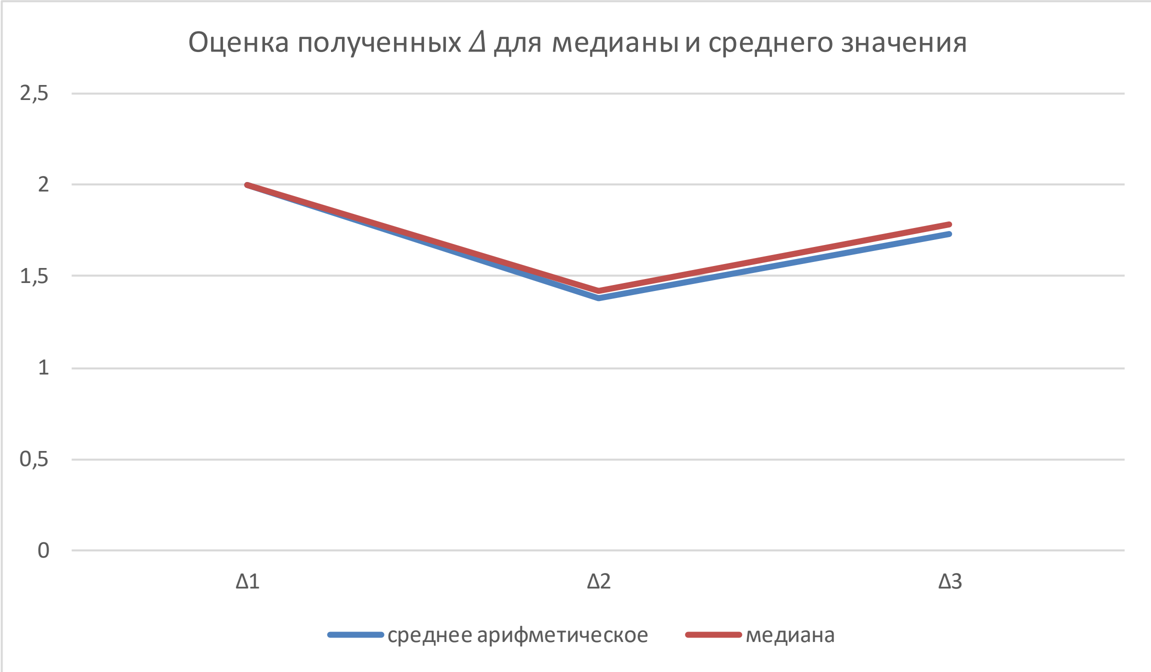
3.

Выборка 1000 значений:

μ, s = 2, 1

Характеристика	Значение
медиана	μ=2
дисперсия	$\frac{(\pi^2 * s^2)}{3} = [s = 1] = \frac{\pi^2}{3} \sim 3.289868133696453$
коэффициенты асимметрии	0
коэффициенты эксцесса	4.2

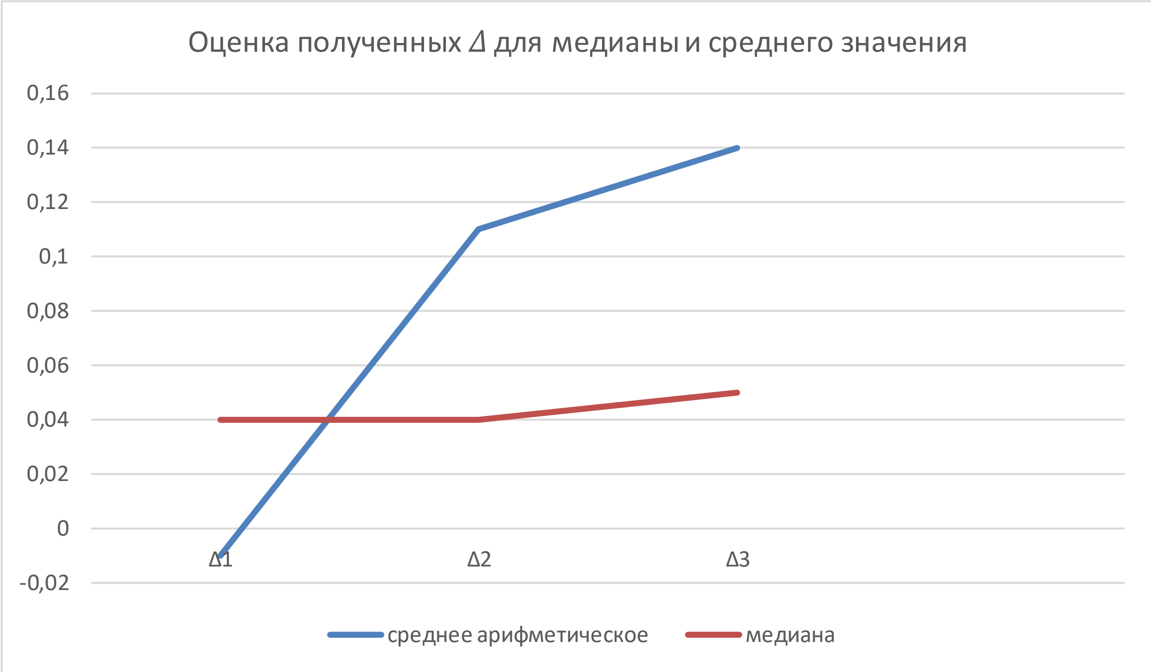
Характеристика	x	y	y2	y2_z	$\lvert\Delta(y_2 - y)\rvert$	$\lvert\Delta(y_{2\_z} - y_2)\rvert$	$\lvert\Delta(y_{2\_z} - y_2)\rvert * (1 + \varepsilon)$
среднее арифметическое	0,50	-0,01	1,99	0,61	2,00	1,38	1,73
медиана	0,51	0,04	2,04	0,62	2,00	1,42	1,78



μ, s = 0, 2

Характеристика	Значение
медиана	$\mu = 0$
дисперсия	$\frac{(\pi^2 * s^2)}{3} = [s = 2] = \frac{4 * \pi^2}{3} \sim 13.159472534785811$
коэффициенты асимметрии	0
коэффициенты эксцесса	4.2

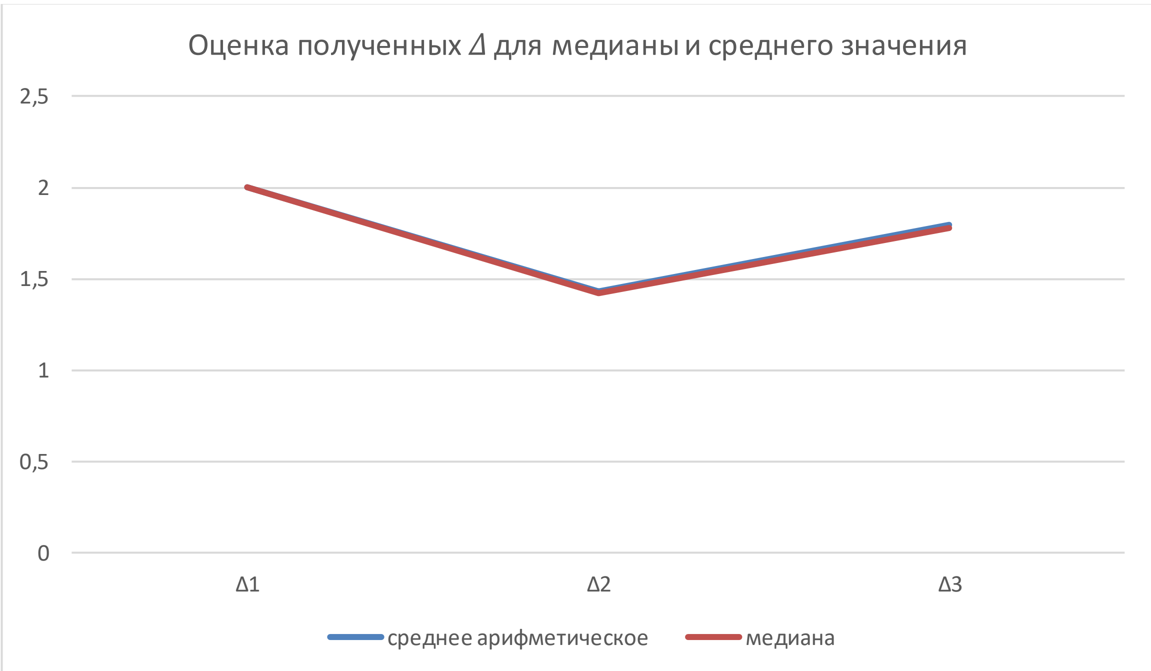
Характеристика	x	y	y3	y3_z	$\lvert\Delta(y_3 - y)\rvert$	$\lvert\Delta(y_{3\_z} - y_3)\rvert$	$\lvert\Delta(y_{3\_z} - y_3)\rvert * (1 + \varepsilon)$
среднее арифметическое	0,50	-0,01	-0,03	0,08	-0,01	0,11	0,14
медиана	0,51	0,04	0,08	0,11	0,04	0,04	0,05



Выборка 10000 значений:

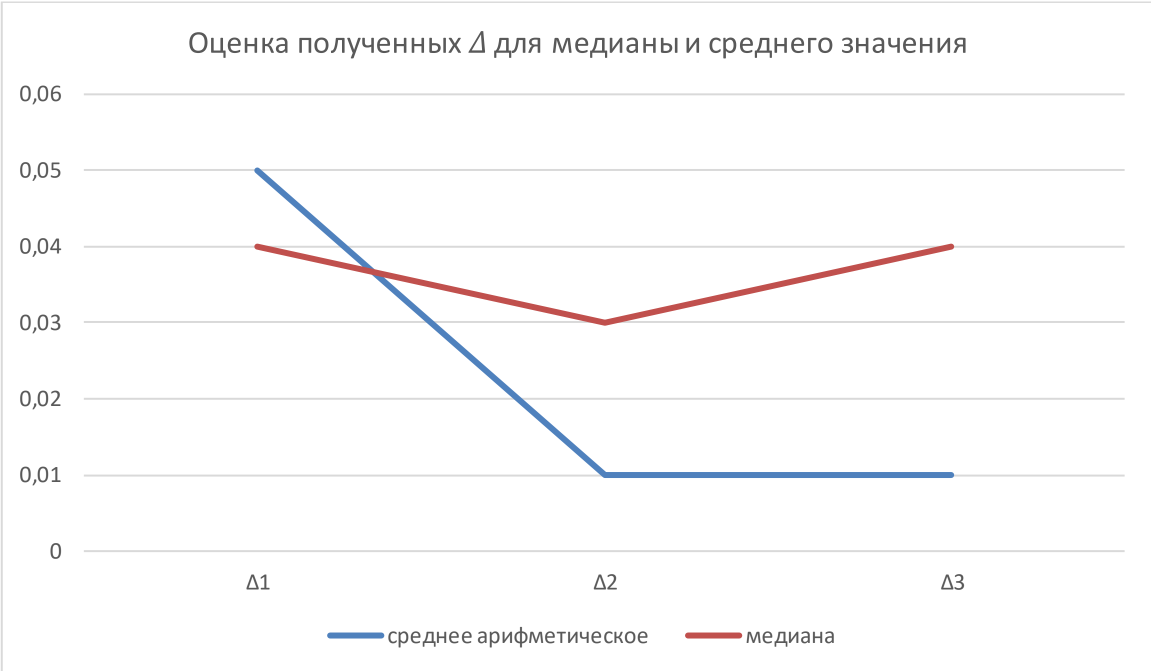
μ, s = 2, 1

Характеристика	x	y	y₂	y_{2_z}	$\lvert\Delta(y_2 - y)\rvert$	$\lvert\Delta(y_{2\_z} - y_2)\rvert$	$\lvert\Delta(y_{2\_z} - y_2)\rvert * (1+\varepsilon)$
среднее арифметическое	0,51	0,05	2,05	0,63	2,00	1,43	1,79
медиана	0,51	0,04	2,04	0,62	2,00	1,42	1,78



μ, s = 0, 2

Характеристика	x	y	y_3	y_{3_z}	$\lvert\Delta(y_3 - y)\rvert$	$\lvert\Delta(y_{3\_z} - y_3)\rvert$	$\lvert\Delta(y_{3\_z} - y_3)\rvert * (1+\varepsilon)$
среднее арифметическое	0,51	0,05	0,11	0,12	0,05	0,01	0,01
медиана	0,51	0,04	0,09	0,11	0,04	0,03	0,04



Исходя из полученных значений видно, что значения почти совпадают с характеристическими. Небольшие расхождения значений могут быть также из-за округления при вычислении значений. При симметрическом засорении графики влияния оценок менее схожи, в отличии от асимметрического засорения.

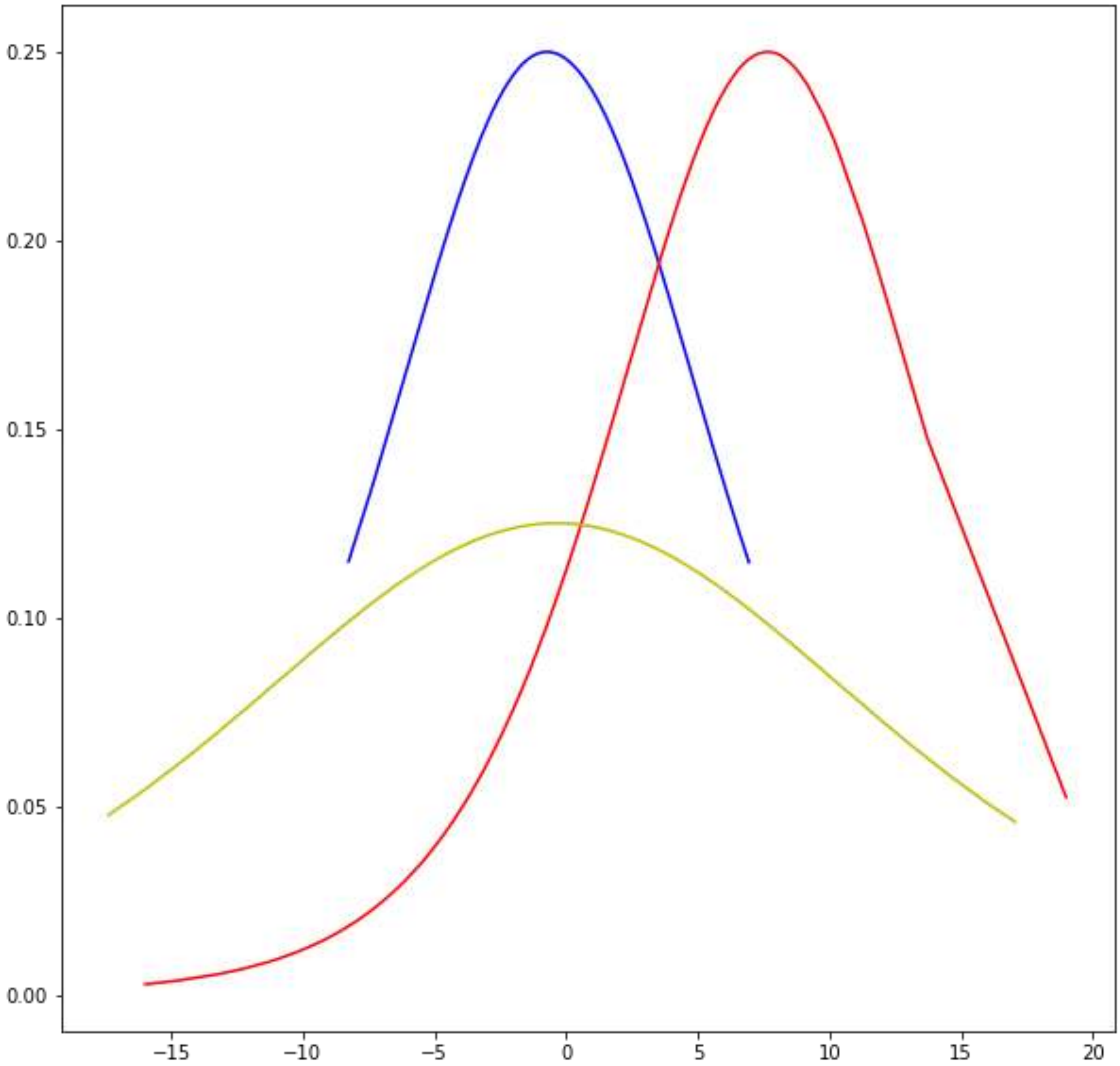
4. Значения по оси OY нормированы относится отрезка [0,1].

Графики плотностей распределений (μ, s = 0, 2).

Чистое распределение,

Засоренное распределение (1000 значений),

Засоренное распределение (10000 значений).

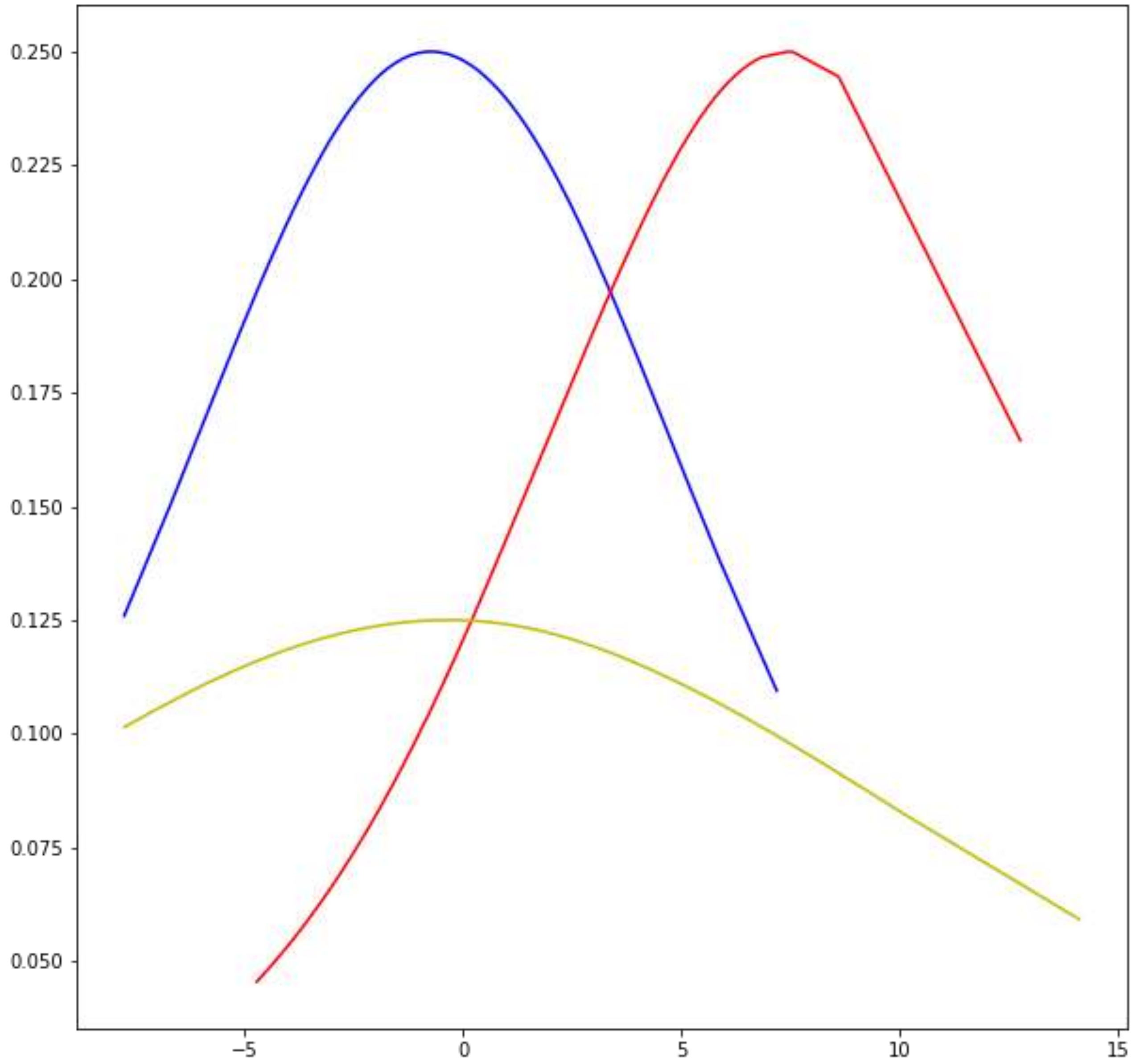


Графики плотностей распределений (μ, s = 2, 1).

Чистое распределение,

Засоренное распределение (1000 значений),

Засоренное распределение (10000 значений).



Вывод:

Исходя из графиков выборок видно, что сдвиг и масштаб, в основном, влияют только на медиану и дисперсию значений логистического распределения соответственно. Сдвиг влияет на вершину параболы плотности логистистического распределения, масштаб на растяжение плотности логистического распределения по осям OX, OY.

Текст программы:

```
from random import random as rnd
import math
import pylab as plb
import matplotlib.pyplot as plt
import numpy as np
from scipy.stats import logistic

loc, scale = 0, 1 # Стандартное распределение.
loc2, scale2 = 2, 1 # Засоренное распределение с асимметричным засорением
µ=2, s=1.
loc3, scale3 = 0, 2 # Засоренное распределение с симметричным засорением
µ=0, s=2.

eps = 0.25

randvars = []

file = open('file1.txt','w')
file2 = open('file2.txt','w')
for i in range(1000):
    x = rnd()
    # math.log(x / (1-x)) - моделирование случайной величины log=ln,
https://docs.python.org/2/library/math.html
    y = loc + scale * math.log(x / (1-x)) # Чистое распределение
    y2 = loc2 + scale2 * math.log(x / (1-x))
    y3 = loc3 + scale3 * math.log(x / (1-x))
    y2_z=((1-eps)*(logistic.pdf(x, loc=2, scale=1)))+eps*y2 # Засоряющая
функция y2
    y3_z=((1-eps)*(logistic.pdf(x, loc=0, scale=2)))+eps*y3 # Засоряющая
функция y3
    print (i+1, round(x,2), round(y,2), round(y2,2), round(y2_z,2),
round((y2-y), 2), round((y2_z-y),2), sep='\t', file=file)
    print (i+1, round(x,2), round(y,2), round(y3,2), round(y3_z,2),
round((y3-y), 2), round((y3_z-y),2), sep='\t', file=file2)
    randvars.append(y)
file.close()
file2.close()
plb.hist(randvars)
plb.show()
plt.plot(randvars)
plt.show()
sigma = (pow(math.pi,2)/3) #Дисперсия
print (sigma)
loc, scale = 0, 1 # Стандартное распределение.
loc2, scale2 = 2, 1 # Засоренное распределение с асимметричным засорением
µ=2, s=1.
loc3, scale3 = 0, 2 # Засоренное распределение с симметричным засорением
µ=0, s=2.

randvars = []

file = open(file3.txt','w')
file2 = open(file4.txt','w')
for i in range(10000):
    x = rnd()
```

```
    # math.log(x / (1-x)) - моделирование случайной величины log=ln,
https://docs.python.org/2/library/math.html
    y = loc + scale * math.log(x / (1-x)) # Чистое распределение
    y2 = loc2 + scale2 * math.log(x / (1-x))
    y3 = loc3 + scale3 * math.log(x / (1-x))
    y2_z=((1-eps)*(logistic.pdf(x, loc=2, scale=1)))+eps*y2 # y2
    y3_z=((1-eps)*(logistic.pdf(x, loc=0, scale=2)))+eps*y3 # y3
    print (i+1, round(x,2), round(y,2), round(y2,2), round(y2_z,2),
round((y2-y), 2), round((y2_z-y),2), sep='\t', file=file)
    print (i+1, round(x,2), round(y,2), round(y3,2), round(y3_z,2),
round((y3-y), 2), round((y3_z-y),2), sep='\t', file=file2)
    randvars.append(y)
file.close()
file2.close()
plb.hist(randvars)
plb.show()
plt.plot(randvars)
plt.show()
sigma = (4*pow(math.pi,2)/3) #Дисперсия
print (sigma)
#μ=2, s=1.
loc, scale = 0, 1 # Стандартное распределение.

randvars1 = []
randvars2 = []

for i in range(1000):
    x = rnd()
    y = loc + scale * math.log(x / (1-x))
    y_z=((1-eps)*(logistic.pdf(x))+eps*y) # y
    #y2_z=((1-eps)*(logistic.pdf(x))+eps*logistic.pdf(y2, loc=2, scale=1)) #
y2
    randvars1.append(y)
    randvars2.append(y_z)
randvars1.sort()
randvars2.sort()
randvars3 = []
randvars4 = []

for i in range(1000):
    x = rnd()
    y2 = loc2 + scale2 * math.log(x / (1-x))
    y2_z=((1-eps)*(logistic.pdf(x))+eps*y2) # y2
    #y2_z=((1-eps)*(logistic.pdf(x))+eps*logistic.pdf(y2, loc=2, scale=1)) #
y2
    randvars3.append(y2)
    randvars4.append(y2_z)
randvars3.sort()
randvars4.sort()
randvars5 = []
randvars6 = []

for i in range(10000):
    x = rnd()
    y2 = loc2 + scale2 * math.log(x / (1-x))
    y2_z=((1-eps)*(logistic.pdf(x, loc=2, scale=1)))+eps*y2 # y2
    randvars5.append(y2)
    randvars6.append(y2_z)
randvars5.sort()
randvars6.sort()
#μ=0, s=2.
```

```
loc, scale = 0, 1 # Стандартное распределение.

randvars7 = []
randvars8 = []

for i in range(1000):
    x = rnd()
    y = loc + scale * math.log(x / (1-x))
    y_z=((1-eps)*(logistic.pdf(x))+eps*y) # y
    #y2_z=((1-eps)*(logistic.pdf(x))+eps*logistic.pdf(y2, loc=2, scale=1)) #
y2
    randvars7.append(y)
    randvars8.append(y_z)
randvars7.sort()
randvars8.sort()
randvars9 = []
randvars10 = []

for i in range(1000):
    x = rnd()
    y3 = loc3 + scale3 * math.log(x / (1-x))
    y3_z=((1-eps)*(logistic.pdf(x, loc=0, scale=2)))+eps*y3 # y3
    randvars9.append(y3)
    randvars10.append(y3_z)
randvars9.sort()
randvars10.sort()
randvars11 = []
randvars12 = []

for i in range(10000):
    x = rnd()
    y3 = loc3 + scale3 * math.log(x / (1-x))
    y3_z=((1-eps)*(logistic.pdf(x, loc=0, scale=2)))+eps*y3 # y3
    randvars11.append(y3)
    randvars12.append(y3_z)
randvars11.sort()
randvars12.sort()
figure, location = plt.subplots(figsize=(10, 10))
plt.plot(randvars7, logistic.pdf(randvars8, loc=0, scale=1),'b', randvars9,
logistic.pdf(randvars10, loc=2, scale=1), 'r', randvars11,
logistic.pdf(randvars12, loc=0, scale=2),'y')
plt.savefig('line_plot1.png')

figure, location = plt.subplots(figsize=(10, 10))
plt.plot(randvars1, logistic.pdf(randvars2, loc=0, scale=1),'b', randvars3,
logistic.pdf(randvars4, loc=2, scale=1), 'r', randvars5,
logistic.pdf(randvars6, loc=0, scale=2),'y')
plt.savefig('line_plot2.png')
print(np.mean(randvars1),np.mean(randvars2),np.mean(randvars3),np.mean(randv
ars4),np.mean(randvars5),np.mean(randvars6),np.mean(randvars7),np.mean(randv
ars8),np.mean(randvars9),np.mean(randvars10),np.mean(randvars11),np.mean(ran
dvars12), sep='\t')
print(np.median(randvars1),np.median(randvars2),np.median(randvars3),np.medi
an(randvars4),np.median(randvars5),np.median(randvars6),np.median(randvars7)
,np.median(randvars8),np.median(randvars9),np.median(randvars10),np.median(r
andvars11),np.median(randvars12), sep='\t')
```