

МИНОБРНАУКИ РОССИИ
САНКТ-ПЕТЕРБУРГСКИЙ ГОСУДАРСТВЕННЫЙ
ЭЛЕКТРОТЕХНИЧЕСКИЙ УНИВЕРСИТЕТ
«ЛЭТИ» ИМ. В.И. УЛЬЯНОВА (ЛЕНИНА)
Кафедра Алгоритмической математики

ОТЧЕТ
по лабораторной работе 2
по дисциплине «Статистический анализ»
Тема: Сравнение свойств статистических оценок.

Студент гр. 8374

Пихтовников К.С.

Преподаватель

Чирина А.В.

Санкт-Петербург

2020

Цель работы: сравнение статистических свойств различных статистических оценок параметра положения "a" распределения для различных распределений выборки методом статистического моделирования.

Задание:

Изучаются 3 оценки:

- a1--- Выборочное среднее;
- a2--- Выборочная медиана;
- a3--- Среднее между выборочными максимумом и минимумом.

Рассматриваются 3 распределения выборки :

- 1) нормальное распределение $N(a, s^2)$;
- 2) распределение Лапласа $L(a, u)$;
- 3) равномерное распределение $U(a-d/2, a+d/2)$.

Методом статистического моделирования оценивается среднеквадратический разброс оценок для разных объёмов выборки. Результаты сравниваются с теоретическими.

Методика выполнения работы:

а) моделирование "m" выборок длины "n" каждая (n=10, n=100, n=1000; m=1000) с заданным распределением (1--3) :

- 1) $n=10$; $m=100$; $X = s \cdot \text{randn}(n, m) + a$;
- 2) $n=10$; $m=100$; $y1 = \text{rand}(n, m)$; $y2 = \text{rand}(n, m)$; $X = (\log(y1) - \log(y2)) \cdot u + a$;
- 3) $n=10$; $m=100$; $A = a - d/2$; $B = a + d/2$; $X = (B - A) \cdot \text{rand}(n, m) + A$;

б) Расчёт "m" значений оценок (a1--a3):

$a1 = \text{mean}(X)$; $a2 = \text{median}(X)$; $a3 = (\max(X) + \min(X)) / 2$;

в) Расчёт характеристик разброса (СКО) оценок:

$r1 = \text{std}(a1)$; $r2 = \text{std}(a2)$; $r3 = \text{std}(a3)$; $r = [r1, r2, r3]$

г) Сравнение расчётных данных с теоретическими , анализ влияния "n" на точность оценивания.

Теоретические данные:

1) $N(a, s^2)$: $t_1 = s/\sqrt{n}$, $t_2 = s \cdot \sqrt{\pi/(2 \cdot n)}$, $t_3 = 0.65 \cdot s / \sqrt{\log(n)}$;

2) $L(a, u)$: $t_1 = u \cdot \sqrt{2/n}$, $t_2 = u/\sqrt{n}$, $t_3 = 0.97 \cdot u$;

3) $U(a-d/2, a+d/2)$: $t_1 = d/\sqrt{12 \cdot n}$, $t_2 = d/(2 \cdot \sqrt{n})$, $t_3 = d/\sqrt{(2 \cdot (n+1) \cdot (n+2))}$.

Вариант 15:

1. Равномерное распределение

$a=4$, $d=5$

2. Распределение Лапласа

$a=3$, $u=4$

1. Выполнение работы:

Текст основной программы (statistical_evaluation.m)

```
N1=10; N2=100; N3=1000;
M=1000;
%распределение равномерное
a1=4;a2=5;
%распределение Лапласа
b1=3;b2=4;
%шаг гистограммы
k=1+round(3.322*log10(M));
%генерация по равномерному распределению
m=M;
A=a1-a2/2;
B=a1+a2/2;
n=N1; r_1_1= (B-A)*rand(n,m)+A;
n=N2; r_1_2= (B-A)*rand(n,m)+A;
n=N3; r_1_3= (B-A)*rand(n,m)+A;
%генерация по распределению Лапласа
n=N1; y1=rand(n,m); y2=rand(n,m); r_2_1=(log(y1)-log(y2))*b2 +b1;
n=N2; y1=rand(n,m); y2=rand(n,m); r_2_2=(log(y1)-log(y2))*b2 +b1;
n=N3; y1=rand(n,m); y2=rand(n,m); r_2_3=(log(y1)-log(y2))*b2 +b1;
%оценки a1,a2,a3 и СКО оценок для равномерного распределения при N=10
a1_1_1=mean(r_1_1);
a2_1_1=median(r_1_1);
a3_1_1=(max(r_1_1)+min(r_1_1))/2;
r1_1_1=std(a1_1_1);
r2_1_1=std(a2_1_1);
r3_1_1=std(a3_1_1);
```

```

r1_1=[r1_1_1,r2_1_1,r3_1_1];
%теоретические данные для равномерного распределения при N=10
d=a2;          n=N1;          t1_1_1=d/sqrt(12*n);          t2_1_1=d/(2*sqrt(n));
t3_1_1=d/sqrt(2*(n+1)*(n+2));
t1_1=[t1_1_1,t2_1_1,t3_1_1];
%Вывод практических и теоретических характеристик разброса СКО оценок
disp('Равномерное распределение ');
disp('1_____r1_____r2_____r3__(10)');disp(r1_1);
disp('1_____t1_____t2_____t3__(10)');disp(t1_1);

%гистограммы для параметров а для равномерного распределения при N=10
figure
tiledlayout(3,1)
histogram_for_evaluation(a1_1_1,k)
histogram_for_evaluation(a2_1_1,k)
histogram_for_evaluation(a3_1_1,k)

%оценки a1,a2,a3 и СКО оценок для равномерного распределения при N=100
a1_1_2=mean(r_1_2);
a2_1_2=median(r_1_2);
a3_1_2=(max(r_1_2)+min(r_1_2))/2;
r1_1_2=std(a1_1_2);
r2_1_2=std(a2_1_2);
r3_1_2=std(a3_1_2);
r2_1=[r1_1_2,r2_1_2,r3_1_2];
%теоретические данные для равномерного распределения при N=100
n=N2;
t1_1_2=d/sqrt(12*n); t2_1_2=d/(2*sqrt(n)); t3_1_2=d/sqrt(2*(n+1)*(n+2));

```

```

t1_2=[t1_1_2,t2_1_2,t3_1_2];
%Вывод практических и теоретических характеристик разброса СКО оценок
disp('2_____r1_____r2_____r3____(100)');disp(r2_1);
disp('2_____t1_____t2_____t3____(100)');disp(t1_2);
%гистограммы для параметров а для равномерного распределения при N=100
figure
tiledlayout(3,1)
histogram_for_evaluation(a1_1_2,k)
histogram_for_evaluation(a2_1_2,k)
histogram_for_evaluation(a3_1_2,k)

%оценки a1,a2,a3 и СКО оценок для равномерного распределения при N=1000
a1_1_3=mean(r_1_3);
a2_1_3=median(r_1_3);
a3_1_3=(max(r_1_3)+min(r_1_3))/2;
r1_1_3=std(a1_1_3);
r2_1_3=std(a2_1_3);
r3_1_3=std(a3_1_3);
r3_1=[r1_1_3,r2_1_3,r3_1_3];
%теоретические данные для равномерного распределения при N=1000
n=N3;
t1_1_3=d/sqrt(12*n); t2_1_3=d/(2*sqrt(n)); t3_1_3=d/sqrt(2*(n+1)*(n+2));
t1_3=[t1_1_3,t2_1_3,t3_1_3];
%Вывод практических и теоретических характеристик разброса СКО оценок
disp('3_____r1_____r2_____r3____(1000)');disp(r3_1);
disp('3_____t1_____t2_____t3____(1000)');disp(t1_3);
%гистограммы для параметров а для равномерного распределения при
N=1000

```

```

figure
tiledlayout(3,1)
histogram_for_evaluation(a1_1_3,k)
histogram_for_evaluation(a2_1_3,k)
histogram_for_evaluation(a3_1_3,k)

%оценки a1,a2,a3 и СКО оценок для распределения Лапласа при N=10
a1_2_1=mean(r_2_1);
a2_2_1=median(r_2_1);
a3_2_1=(max(r_2_1)+min(r_2_1))/2;
r1_2_1=std(a1_2_1);
r2_2_1=std(a2_2_1);
r3_2_1=std(a3_2_1);
r1_2=[r1_2_1,r2_2_1,r3_2_1];

%теоретические данные для распределения Лапласа при N=10
u=b2; n=N1;
t1_2_1=u*sqrt(2/n); t2_2_1=u/sqrt(n); t3_2_1=0.97*u;
t2_1=[t1_2_1,t2_2_1,t3_2_1];

%Вывод практических и теоретических характеристик разброса СКО оценок
disp('Распределение Лапласа');
disp('4_____r1_____r2_____r3__(10)');disp(r1_2);
disp('4_____t1_____t2_____t3__(10)');disp(t2_1);

%гистограммы для параметров а для распределения Лапласа при N=10
figure
tiledlayout(3,1)
histogram_for_evaluation(a1_2_1,k)
histogram_for_evaluation(a2_2_1,k)
histogram_for_evaluation(a3_2_1,k)

```

```

%оценки a1,a2,a3 и СКО оценок для распределения Лапласа при N=100
a1_2_2=mean(r_2_2);
a2_2_2=median(r_2_2);
a3_2_2=(max(r_2_2)+min(r_2_2))/2;
r1_2_2=std(a1_2_2);
r2_2_2=std(a2_2_2);
r3_2_2=std(a3_2_2);
r2_2=[r1_2_2,r2_2_2,r3_2_2];

%теоретические данные для распределения Лапласа при N=100
n=N2;
t1_2_2=u*sqrt(2/n); t2_2_2=u/sqrt(n); t3_2_2=0.97*u;
t2_2=[t1_2_2,t2_2_2,t3_2_2];

%Вывод практических и теоретических характеристик разброса СКО оценок
disp('5_____r1_____r2_____r3____(100)');disp(r2_2);
disp('5_____t1_____t2_____t3____(100)');disp(t2_2);

%гистограммы для параметров а для распределения Лапласа при N=100
figure
tiledlayout(3,1)
histogram_for_evaluation(a1_2_2,k)
histogram_for_evaluation(a2_2_2,k)
histogram_for_evaluation(a3_2_2,k)

%оценки a1,a2,a3 и СКО оценок для распрделения Лапласа при N=1000
a1_2_3=mean(r_2_3);
a2_2_3=median(r_2_3);
a3_2_3=(max(r_2_3)+min(r_2_3))/2;
r1_2_3=std(a1_2_3);

```



```

r2_2_3=std(a2_2_3);
r3_2_3=std(a3_2_3);
r3_2=[r1_2_3,r2_2_3,r3_2_3];
%теоретические данные для распределения Лапласа при N=1000
n=N3;
t1_2_3=u*sqrt(2/n); t2_2_3=u/sqrt(n); t3_2_3=0.97*u;
t2_3=[t1_2_3,t2_2_3,t3_2_3];
%Вывод практических и теоретических характеристик разброса СКО оценок
disp('6_____r1_____r2_____r3__(1000)');disp(r3_2);
disp('6_____t1_____t2_____t3__(1000)');disp(t2_3);
%гистограммы для параметров а для распределения Лапласа при N=1000
figure
tiledlayout(3,1)
histogram_for_evaluation(a1_2_3,k)
histogram_for_evaluation(a2_2_3,k)
histogram_for_evaluation(a3_2_3,k)

```

**Текст функции, которая выводит гистограммы оценок (файл
histogram_for_evaluation.m)**

```

function histogram_for_evaluation(x,k)
nexttile
hold on;grid on
histogram(x,k)
[x,h]=hist(x,k);
plot(h,x,'--or')
hold off

```

Гистограммы для оценок a_1 -выборочное среднее, a_2 -выборочная медиана, a_3 -среднее между выборочными максимумом и минимумом:

Равномерное распределение:

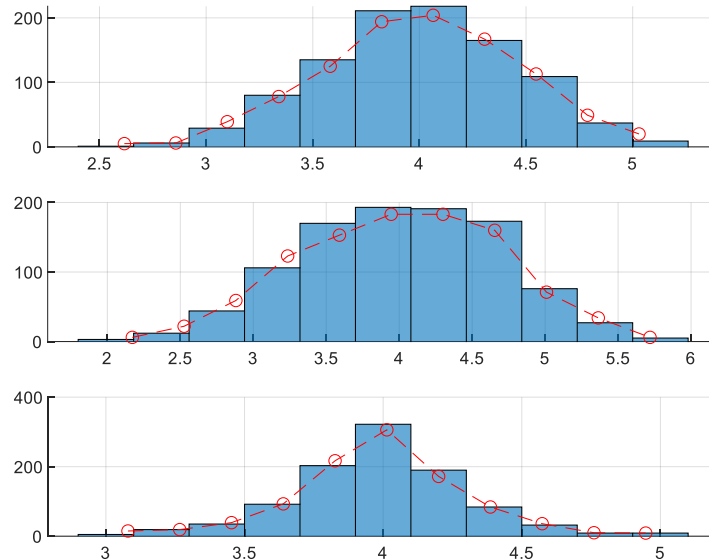


Рис 1. Гистограмма для оценок a_1, a_2, a_3 при $N=10$

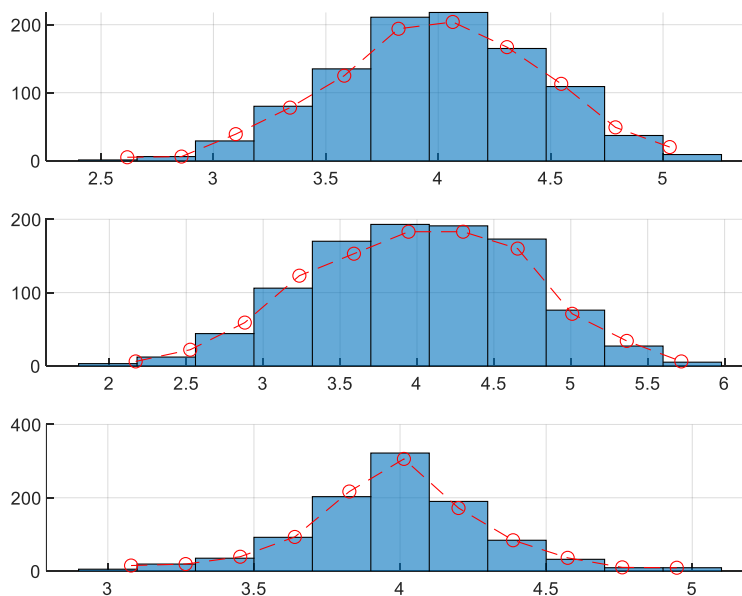


Рис 2. Гистограмма для оценок a_1, a_2, a_3 при $N=100$

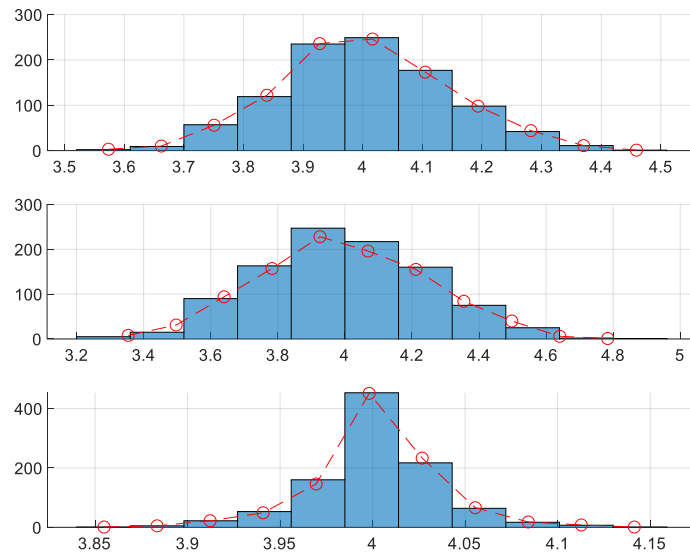


Рис 3. Гистограмма для оценок a_1, a_2, a_3 при $N=1000$

Распределение Лапласа:

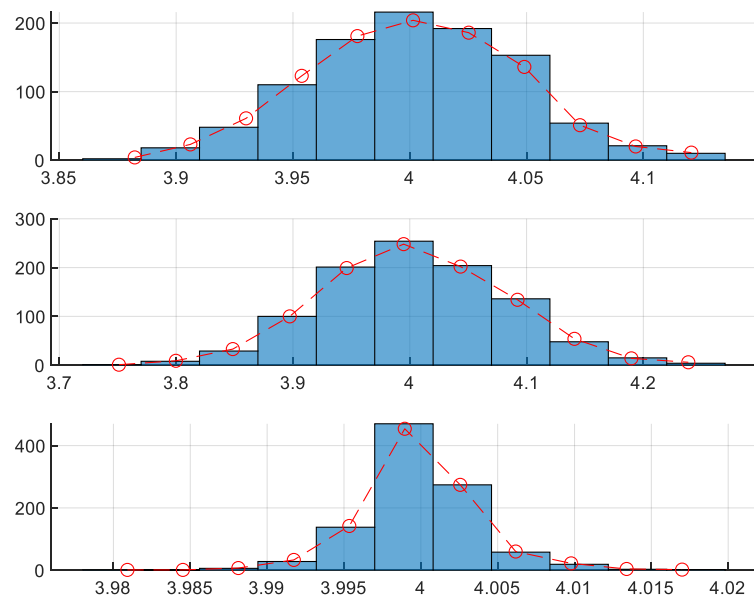


Рис 4. Гистограмма для оценок a_1, a_2, a_3 при $N=10$

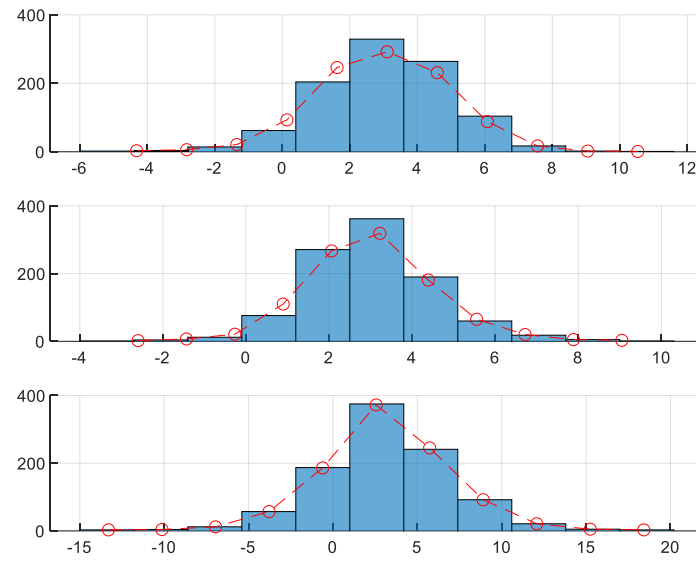


Рис 5. Гистограмма для оценок a_1, a_2, a_3 при $N=100$

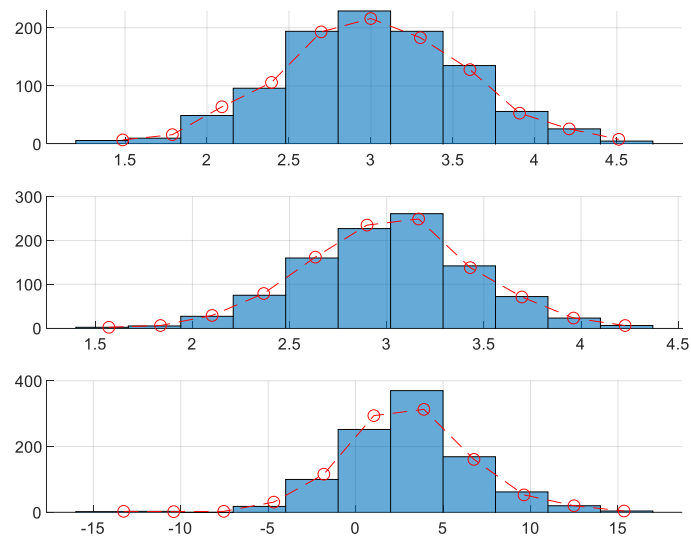


Рис 6. Гистограмма для оценок a_1, a_2, a_3 при $N=1000$

**Расчет практических характеристик разброса СКО оценок (r1,r2,r3) и
расчёт теоретических данных разброса СКО оценок (t1,t2,t3)**

Равномерное распределение:

N=10:

r1	r2	r3
0.4551	0.6829	0.3093
t1	t2	t3
0.4564	0.7906	0.3077

N=100:

r1	r2	r3
0.1421	0.2490	0.0337
t1	t2	t3
0.1443	0.2500	0.0348

N=1000:

r1	r2	r3
0.0441	0.0774	0.0037
t1	t2	t3
0.0456	0.0791	0.0035

Распределение Лапласа:

N=10:

r1	r2	r3
1.9208	1.5261	3.9320
t1	t2	t3
1.7889	1.2649	3.8800

N=100:

r1	r2	r3
0.5496	0.4254	3.6972
t1	t2	t3
0.5657	0.4000	3.8800

N=1000:

r1	r2	r3
0.1739	0.1280	3.6161
t1	t2	t3
0.1789	0.1265	3.8800

2. ответы на вопросы

а) Наилучшей оценкой равномерного распределения является среднее между выборочными максимумом и минимумом, так как мера разброса для оценки данного параметра наименьшая.

А наилучшей оценкой распределения Лапласа является выборочная медиана, так как мера разброса для оценки данного параметра наименьшая.

б) Разброс оценок - это дисперсия. Формула дисперсии равна:

$$\frac{\sum (x - \bar{x})^2}{n}$$

Отсюда следует, что дисперсия становится меньше, когда увеличивается объем выборки. То есть, с ростом n разброс оценок уменьшается.

в) относительная степень соответствия экспериментальных характеристик разброса (в %):

Равномерное распределение:

При n =10: 98.3%

При n =100: 98%

При $n = 1000$: 98.9%

Распределение Лапласа:

При $n = 10$: 88.9%

При $n = 100$: 91.6%

При $n = 1000$: 95.5%