**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 "а" распределения для различных распределений выборки методом статистического моделирования.

**Задание:**

Изучаются 3 оценки:

а1--- Выборочное среднее;

а2--- Выборочная медиана;

а3--- 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\*randn(n,m)+a;

2) n=10; m=100; y1=rand(n,m); y2=rand(n,m); X=(log(y1)-log(y2))\*u +a;

3) n=10; m=100; A=a-d/2;B=a+d/2;X=(B-A)\*rand(n,m)+A;

б) Расчёт "m" значений оценок (а1--а3):

a1=mean(X); a2=median(X); a3=(max(X)+min(X))/2;

в) Расчёт характеристик разброса (СКО) оценок:

r1=std(a1); r2=std(a2); r3=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): t1=s/sqrt(n), t2=s\*sqrt(pi/(2\*n)), t3=0.65\*s\*/sqrt(log(n));

2) L(a,u): t1=u\*sqrt(2/n), t2=u/sqrt(n), t3=0.97\*u;

3)U(a-d/2,a+d/2):t1=d/sqrt(12\*n),t2=d/(2\*sqrt(n)),t3=d/sqrt((2\*(n+1)\*(n+2))).

**Вариант 15:**

1. Равномерное распределение

a=4, d=5

1. 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 a 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 a 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 a 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 a 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 a 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 a 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 a1-выборочное среднее, a2-выборочная медиана, a3-среднее между выборочными максимумом и минимумом:**

***Равномерное распределение:***

****

*Рис 1. Гистограмма для оценок a1,a2,a3 при N=10*

****

*Рис 2. Гистограмма для оценок a1,a2,a3 при N=100*

****

*Рис 3. Гистограмма для оценок a1,a2,a3 при N=1000*

***Распределение Лапласа:***

****

*Рис 4. Гистограмма для оценок a1,a2,a3 при N=10*

****

*Рис 5. Гистограмма для оценок a1,a2,a3 при N=100*

****

*Рис 6. Гистограмма для оценок a1,a2,a3 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:

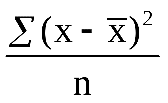
|  |  |  |
| --- | --- | --- |
| r1 | r2 | r3 |
| 0.1739 | 0.1280 | 3.6161 |
| t1 | t2 | t3 |
| 0.1789 | 0.1265 | 3.8800 |

1. **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:



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%