Лабораторная работа 5

Борис Шапошников

Виктор Шарепо

Рамазан Рахматуллин

Обратная связь: bshaposhnikov01@gmail.com

Формулировка задачи

Для трех распределений $X \sim N(a, \sigma), X \sim U(a-\delta/2, a+\delta/2)$ и распределения Лапласа или двойного показательного — $L(a, u) = a+Exp_\lambda - Exp_\lambda, \lambda = \frac{1}{u}$ (суммируемые показательные распределения независимы). Сравнить следующие оценки параметра a – математического ожидания и медианы всех распределений, $\bar{X}_n$ - выборочного среднего, med_n — выборочной медианы и $\frac{x_{(1)}+x_{(n)}}{2}$ — полусуммы минимума и максимума вариационного ряда. Все оценки не смещены. Сравнивать оценки нужно с точки зрения квадратичного риска (т. е. для несмещенных оценок одномерного параметра – дисперсии оценки). При $n = 100$ – объем выборки, $m = 100$ – количество выборок, построить 100 оценок каждого вида и сравнить их выборочные среднеквадратичные отклонения, повторить при $n = 10000$, $m = 100$. Сравнить с теоретическими среднеквадратичными отклонениями. Результат – 6 таблиц и вывод о том какая из оценок с точки зрения квадратичного риска является наилучшей.

Теоретическая асимптотика дисперсий

	$X \sim N(a, \sigma)$	$L(a, u)$	$X \sim U(a - \delta/2, a + \delta/2)$
$\bar{X}_n$	$\frac{\sigma^2}{n}$	$\frac{2u^2}{n}$	$\frac{\delta^2}{12n}$
med_n	$\frac{\pi\sigma^2}{2n}$	$\frac{u^2}{n}$	$\frac{\delta^2}{4n}$
$\frac{x_{(1)}+x_{(n)}}{2}$	$\frac{c_1\sigma^2}{\ln n}$	c_2u^2	$\frac{\delta^2}{2n^2}$

$c_1 \approx 0.4$

$c_2 \approx 0.9$

Параметры распределений

- Нормальное распределение $X \sim N(a, \sigma)$

  $a = 0$

  $\sigma = 1$

- Равномерное распределение $X \sim U(a - \delta/2, a + \delta/2)$

  $a = 0$

  $\delta = 1$

- Распределение Лапласа $X \sim L(a, u) = a + Exp_\lambda - Exp_\lambda, \lambda = \frac{1}{u}$

  $a = 0$

  $u = 1$

Код и таблицы

- n = 100, m = 100
  - Нормальное распределение

```
clc;
clear;

pkg load statistics;

```

```
c1 = 0.4;
c2 = 0.9;

n = 100;
m = 100;
a = 0;
sigma = 1;

x = normrnd(a, sigma, n, m);

Davg = sigma ^ 2 / n;
Dmed = pi * sigma ^ 2 / (2 * n);
Dminmax = c1 * sigma ^ 2 / log(n);

Tsigma_avg = sqrt(Davg);
Tsigma_med = sqrt(Dmed);
Tsigma_minmax = sqrt(Dminmax);

printf("Theoretical sigma:\n\taverage: %f,\n\tmedian: %f,\n\tminmax: %f\n", Tsigma_avg, Tsigma_med, Tsigma_minmax);

Eavg = mean(x);
Emed = median(x);
Eminmax = (min(x) + max(x)) ./ 2;

sigma_avg = std(Eavg);
sigma_med = std(Emed);
sigma_minmax = std(Eminmax);

printf("Practical sigma:\n\taverage: %f,\n\tmedian: %f,\n\tminmax: %f\n", sigma_avg, sigma_med, sigma_minmax);
```

$X \sim N(a, \sigma), n = 100, m = 100$	$\bar{X}_n$	med_n	$\frac{x_{(1)}+x_{(n)}}{2}$
σ — теоретическое	0.100000	0.125331	0.294718
σ — практическое	0.117486	0.121579	0.299006

– Равномерное распределение

```
clc;
clear;

pkg load statistics;

```

```
c1 = 0.4;
c2 = 0.9;

n = 100;
m = 100;
a = 0;
delta = 1;

x = unifrnd(a - delta / 2, a + delta / 2, n, m);

Davg = delta ^ 2 / (12 * n);
Dmed = delta ^ 2 / (4 * n);
Dminmax = delta ^ 2 / (2 * n ^ 2);

Tsigma_avg = sqrt(Davg);
Tsigma_med = sqrt(Dmed);
Tsigma_minmax = sqrt(Dminmax);

printf("Theoretical sigma:\n\taverage: %f,\n\tmedian: %f,\n\tminmax
    : %f\n", Tsigma_avg, Tsigma_med, Tsigma_minmax);

Eavg = mean(x);
Emed = median(x);
Eminmax = (min(x) + max(x)) ./ 2;

sigma_avg = std(Eavg);
sigma_med = std(Emed);
sigma_minmax = std(Eminmax);

printf("Practical sigma:\n\taverage: %f,\n\tmedian: %f,\n\tminmax:
    %f\n", sigma_avg, sigma_med, sigma_minmax);
```

$X \sim U(a-\delta/2, a+\delta/2), n=100, m=100$	$\bar{X}_n$	med_n	$\frac{x_{(1)}+x_{(n)}}{2}$
σ — теоретическое	0.028868	0.050000	0.007071
σ — практическое	0.029562	0.047793	0.006985

– Распределение Лапласа

```
clc;
clear;

pkg load statistics;
```

```

c1 = 0.4;
c2 = 0.9;

n = 100;
m = 100;
a = 0;
u = 1;

x = a + exprnd(1 / u, n, m) - exprnd(1 / u, n, m);

Davg = 2 * u ^ 2 / n;
Dmed = u ^ 2 / n;
Dminmax = c2 * u ^ 2;

Tsigma_avg = sqrt(Davg);
Tsigma_med = sqrt(Dmed);
Tsigma_minmax = sqrt(Dminmax);

printf("Theoretical sigma:\n\taverage: %f,\n\tmedian: %f,\n\tminmax
    : %f\n", Tsigma_avg, Tsigma_med, Tsigma_minmax);

Eavg = mean(x);
Emed = median(x);
Eminmax = (min(x) + max(x)) ./ 2;

sigma_avg = std(Eavg);
sigma_med = std(Emed);
sigma_minmax = std(Eminmax);

printf("Practical sigma:\n\taverage: %f,\n\tmedian: %f,\n\tminmax:
    %f\n", sigma_avg, sigma_med, sigma_minmax);
```

$X \sim L(a,u), n = 100, m = 100$	$\bar{X}_n$	med_n	$\frac{x_{(1)}+x_{(n)}}{2}$
σ — теоретическое	0.141421	0.100000	0.948683
σ — практическое	0.121876	0.108329	1.006396

- n = 10000, m = 100
  - Нормальное распределение

```
clc;
clear;
```

```

pkg load statistics;

c1 = 0.4;
c2 = 0.9;

n = 10000;
m = 100;
a = 0;
sigma = 1;

x = normrnd(a, sigma, n, m);

Davg = sigma ^ 2 / n;
Dmed = pi * sigma ^ 2 / (2 * n);
Dminmax = c1 * sigma ^ 2 / log(n);

Tsigma_avg = sqrt(Davg);
Tsigma_med = sqrt(Dmed);
Tsigma_minmax = sqrt(Dminmax);

printf("Theoretical sigma:\n\taverage: %f,\n\tmedian: %f,\n\tminmax
    : %f\n", Tsigma_avg, Tsigma_med, Tsigma_minmax);

Eavg = mean(x);
Emed = median(x);
Eminmax = (min(x) + max(x)) ./ 2;

sigma_avg = std(Eavg);
sigma_med = std(Emed);
sigma_minmax = std(Eminmax);

printf("Practical sigma:\n\taverage: %f,\n\tmedian: %f,\n\tminmax:
    %f\n", sigma_avg, sigma_med, sigma_minmax);
```

$X \sim N(a, \sigma), n = 10000, m = 100$	$\bar{X}_n$	med_n	$\frac{x_{(1)}+x_{(n)}}{2}$
σ — теоретическое	0.010000	0.012533	0.208397
σ — практическое	0.010297	0.012542	0.207849

– Равномерное распределение

```
clc;
clear;
```

```

pkg load statistics;

c1 = 0.4;
c2 = 0.9;

n = 10000;
m = 100;
a = 0;
delta = 1;

x = unifrnd(a - delta / 2, a + delta / 2, n, m);

Davg = delta ^ 2 / (12 * n);
Dmed = delta ^ 2 / (4 * n);
Dminmax = delta ^ 2 / (2 * n ^ 2);

Tsigma_avg = sqrt(Davg);
Tsigma_med = sqrt(Dmed);
Tsigma_minmax = sqrt(Dminmax);

printf("Theoretical sigma:\n\taverage: %f,\n\tmedian: %f,\n\tminmax: %f\n", Tsigma_avg, Tsigma_med, Tsigma_minmax);

Eavg = mean(x);
Emed = median(x);
Eminmax = (min(x) + max(x)) ./ 2;

sigma_avg = std(Eavg);
sigma_med = std(Emed);
sigma_minmax = std(Eminmax);

printf("Practical sigma:\n\taverage: %f,\n\tmedian: %f,\n\tminmax: %f\n", sigma_avg, sigma_med, sigma_minmax);
```

$X \sim U(a-\delta/2, a+\delta/2), n = 10000, m = 100$	$\bar{X}_n$	med_n	$\frac{x_{(1)}+x_{(n)}}{2}$
σ — теоретическое	0.002887	0.005000	0.000071
σ — практическое	0.002674	0.004262	0.000058

– Распределение Лапласа

```
clc;
```

```
clear;

pkg load statistics;

c1 = 0.4;
c2 = 0.9;

n = 10000;
m = 100;
a = 0;
u = 1;

x = a + exprnd(1 / u, n, m) - exprnd(1 / u, n, m);

Davg = 2 * u ^ 2 / n;
Dmed = u ^ 2 / n;
Dminmax = c2 * u ^ 2;

Tsigma_avg = sqrt(Davg);
Tsigma_med = sqrt(Dmed);
Tsigma_minmax = sqrt(Dminmax);

printf("Theoretical sigma:\n\taverage: %f,\n\tmedian: %f,\n\tminmax
    : %f\n", Tsigma_avg, Tsigma_med, Tsigma_minmax);

Eavg = mean(x);
Emed = median(x);
Eminmax = (min(x) + max(x)) ./ 2;

sigma_avg = std(Eavg);
sigma_med = std(Emed);
sigma_minmax = std(Eminmax);

printf("Practical sigma:\n\taverage: %f,\n\tmedian: %f,\n\tminmax:
    %f\n", sigma_avg, sigma_med, sigma_minmax);
```

$X \sim L(a,u), n = 10000, m = 100$	$\bar{X}_n$	med_n	$\frac{x_{(1)}+x_{(n)}}{2}$
σ — теоретическое	0.014142	0.010000	0.948683
σ — практическое	0.013157	0.009899	0.848805

Вывод

В результате проведения работы практические и теоретические значения σ оказались близки (не отличаются по порядку). При увеличении n значения квадратичного риска приближаются к теоретическим для состоятельных оценок. Можно заметить что оценка MinMax не является состоятельной для распределения Лапласа. У каждого распределения своя оценка максимального правдоподобия имеет минимальную дисперсию (из-за несмещённости квадратичный риск = дисперсии).

Для $X \sim N(a, \sigma)$ — это $\bar{X}_n$,
для $X \sim U(a - \delta/2, a + \delta/2)$ — это $\frac{x_{(1)} + x_{(n)}}{2}$
и для $X \sim L(a, u)$ — это med_n