МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ РОССИЙСКОЙ ФЕДЕРАЦИИ

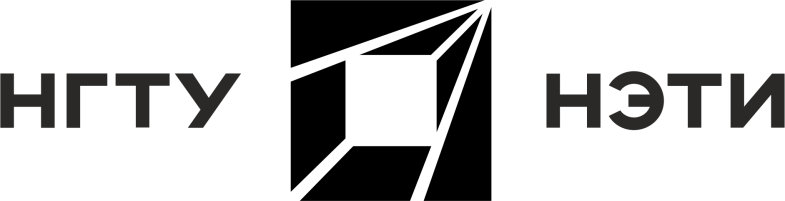
ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ БЮДЖЕТНОЕ

ОБРАЗОВАТЕЛЬНОЕ УЧРЕЖДЕНИЕ

ВЫСШЕГО ОБРАЗОВАНИЯ

«НОВОСИБИРСКИЙ ГОСУДАРСТВЕННЫЙ ТЕХНИЧЕСКИЙ УНИВЕРСИТЕТ»

Кафедра теоретической и прикладной информатики



**ЛАБОРАТОРНАЯ РАБОТА №1**

по дисциплине: Компьютерные технологии моделирования и анализа данных

на тему: Исследование свойств оценок параметров распределений вероятностей по эмпирическим данным

  Факультет: ФПМИ

Группа: ПММ-21

Выполнили: Сухих А.С.

Вариант №2

Проверил: д.т.н., профессор Лемешко Б.Ю.

Дата выполнения: 12.10.22

Отметка о защите:

Новосибирск 2022

**Цель работы.** Вычисление оценок параметров распределений вероятностей по эмпирическим данным различными методами. Исследование асимптотических свойств оценок методом Монте-Карло.

|  |  |
| --- | --- |
| 2 | Распределение Лапласа, параметр масштаба. |

**Ход выполнения**

В Pos\_KTAD примерно на 80 странице описано, как формировать засоренное распределение через файл .ini

1. Определение количества экспериментов *N.*

Количество экспериментов было определено, исходя из требуемой точности ε*=0.01* согласно формуле

2. Моделирование выборок.

С использованием программы ISW были смоделированы выборки, состоящие из оценок:

- максимального правдоподобия –,

- минимального расстояния Колмогорова – ,

- минимального расстояния  Крамера–Мизеса–Смирнова –,

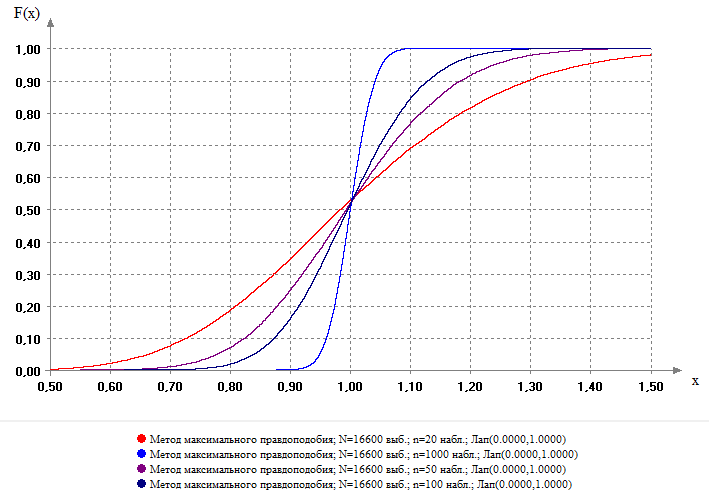
- минимального расстояния  Андерсона–Дарлинга,

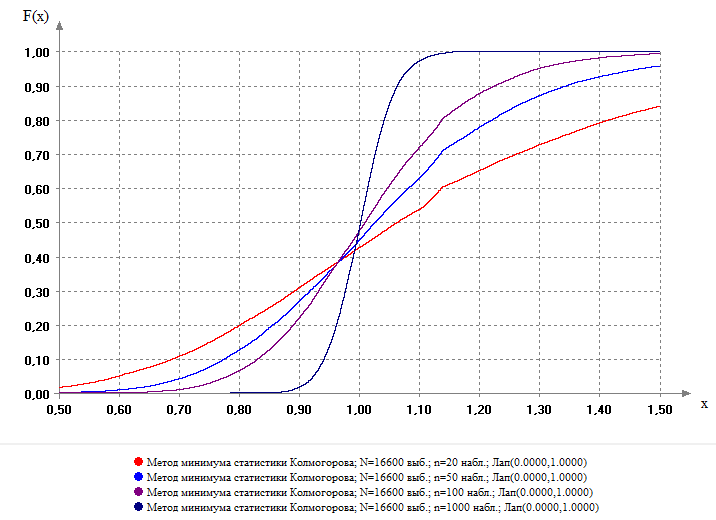
- по порядковым статистикам (*L*-оценки)

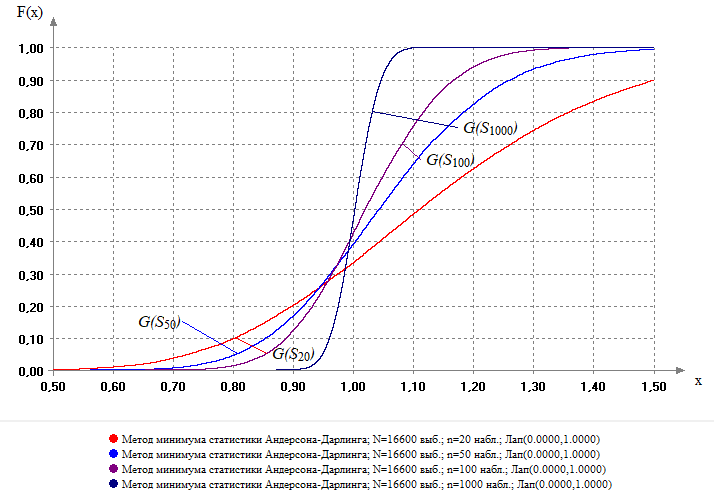
при различном объеме наблюдений *n* = 20,50,100,1000.

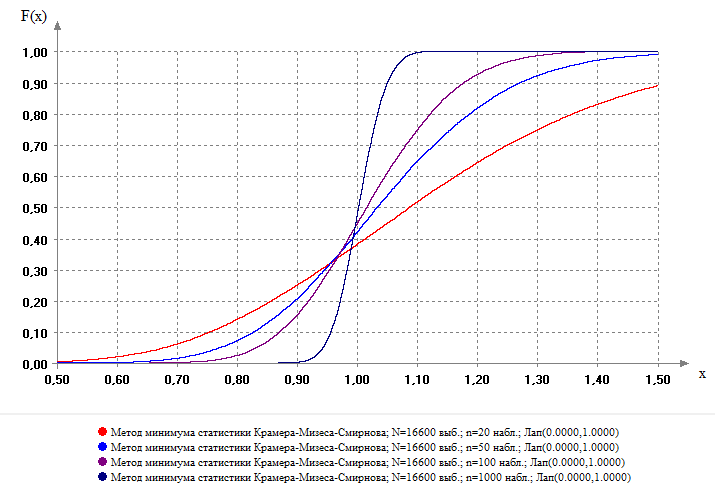
3. Построение графиков функций распределения смоделированных выборок

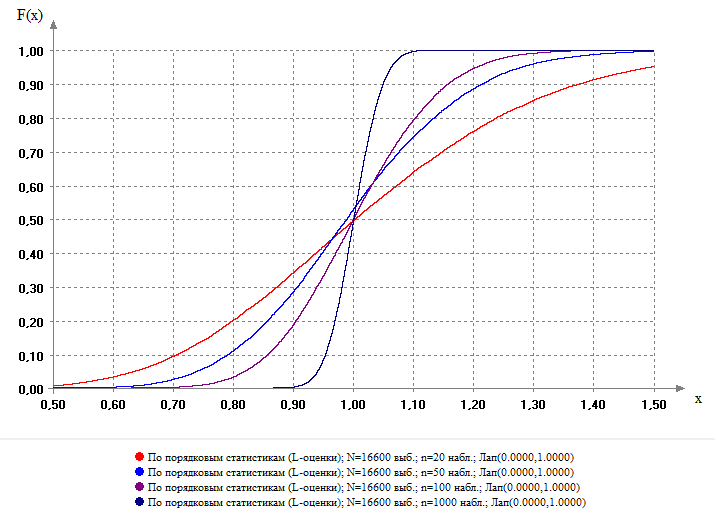
Построенные графики приведены на рисунках 1-5 для различных оценок и при различном объеме наблюдений *n*. Как видно на рисунках, по мере увеличения объема наблюдений функции распределения принимают форму, близкую к функции нормального распределения.

Рисунок 1: Сравнение графиков распределения ОМП при различных количествах наблюдений n

Рисунок 2: Сравнение графиков распределения MD-оценки Колмогорова при различных количествах наблюдений n

Рисунок 3: Сравнение графиков распределения MD-оценки Андерсона-Дарлинга при различных количествах наблюдений n

Рисунок 4: Сравнение графиков распределения MD-оценки Крамера-Мизеса-Смирнова при различных количествах наблюдений n

Рисунок 5: Сравнение графиков распределения L-оценки при различных количествах наблюдений n

4. Исследование на нормальность

Проверка гипотезы о согласии оценки нормальному закону распределения была произведена с помощью критерия Хи-квадрат.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Набл.  Оценка | n = 20 | n = 50 | n = 100 | n = 1000 |
| ОМП | P = 3,4766\*10-71,  отвергается | P = 1,1093\*10-27,  отвергается | P = 1,3289\*10-14,  отвергается | P = 0,06,  не отвергается |
| Колмогорова | P = 0,  отвергается | P = 0,  отвергается | P = 0,  отвергается | P = 1,3166\*10-15,  отвергается |
| К-М-С, ɷ2 | P = 0,  отвергается | P = 2,6823\*10-46,  отвергается | P = 4,6383\*10-27,  отвергается | P = 0,0451,  не отвергается |
| Андерсона-Дарлинга | P = 0,  отвергается | P = 6,7715\*10-50,  отвергается | P = 2,715\*10-21,  отвергается | P = 0,0519,  не отвергается |
| L | P = 0,  отвергается | P = 2,9521\*10-39,  отвергается | P = 1,0528\*10-20,  отвергается | P = 0,0041  отвергается |

По полученным в таблице данным видно, что наиболее эффективным методом оценивания оказалась оценка максимального правдоподобия. Смоделированные выборки соответствовали нормальному закону распределения лишь в 3 методах из 5 при количестве наблюдений *n* =1000.

5. Исследование свойств оценок

При помощи программы ISW были определены свойства несмещенности, состоятельности и эффективности для оценок при *n* = 20,50,100,1000 методами:

- максимального правдоподобия –,

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | *n =20* | *n =50* | *n =100* | *n =1000* |
|  | 0,0025 | 0,0004 | -0,0005 | 0,0003 |
|  | 65,75% | 48,06% | 31,77% | 0% |
|  | -0,0002 | 2,225\*10-5 | 8,016\*10-5 | 0,489\*10-5 |

- минимального расстояния Колмогорова – ,

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | *n =20* | *n =50* | *n =100* | *n =1000* |
|  | 0,0863 | 0,0334 | 0,0151 | 0,0018 |
|  | 78,52% | 59,92% | 52,48% | 3,86% |
|  | 0,0883 | 0,0296 | 0,0139 | 1,352\*10-3 |

- минимального расстояния Крамера–Мизеса–Смирнова –,

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | *n =20* | *n =50* | *n =100* | *n =1000* |
|  | 0,0977 | 0,0375 | 0,0178 | 0,0023 |
|  | 74,59% | 57,46% | 40,89% | 0,83% |
|  | 0,0372 | 9,86\*10-3 | 4,472\*10-3 | 3,6161\*10-4 |

- минимального расстояния Андерсона–Дарлинга,

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | *n =20* | *n =50* | *n =100* | *n =1000* |
|  | 0,1399 | 0,0486 | 0,0230 | 0,0029 |
|  | 75,69% | 53,87% | 37,84% | 0,55% |
|  | 0,0305 | 5,313\*10-3 | 1,99\*10-3 | 1,2896\*10-4 |

- по порядковым статистикам (*L*-оценки)

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | *n =20* | *n =50* | *n =100* | *n =1000* |
|  | 0,0176 | -0,0056 | 0,0033 | 0,0011 |
|  | 70,72% | 52,49% | 38,95 | 1,1% |
|  | 0,0195 | 0,0071 | 0,0038 | 4,0625\*10-4 |

**Вывод**

По результатам, полученным в ходе выполнения лабораторной работы было установлено, что с ростом объема выборки улучшаются такие свойства оценок как несмещенность, состоятельность и эффективность, также существенно повышается достигаемый уровень значимости.

Наиболее эффективным методом можно назвать оценку максимального правдоподобия, обеспечивающую наибольший рост свойств оценки и уровня значимости. В свою очередь выборки, полученные методами MD-оценки Колмогорова и L-оценки при всех исследуемых объемах выборки отвергали гипотезу о согласии нормальному закону распределения, что говорит об их малой эффективности применительно к моделированию и оценке параметра масштаба распределения Лапласа.