МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ РОССИЙСКОЙ ФЕДЕРАЦИИ

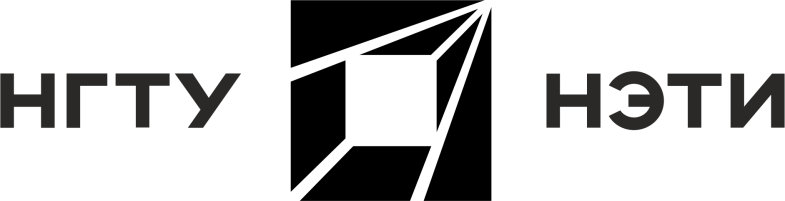
ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ БЮДЖЕТНОЕ

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.** Исследование устойчивости оценок на наличие в выборке аномальных наблюдений. Исследование эффективности параметрической процедуры исключения аномальных наблюдений при использовании робастных оценок. Построение функций влияния Хампеля для ОМП. Исследование распределений статистик типа Граббса, предназначенных для анализа на аномальность сразу нескольких наблюдений.

|  |  |  |
| --- | --- | --- |
| № |  |  |
| 2 | Нормальное | Нормальное с масштабом 5 |

**Ход выполнения**

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

1. Исследовать робастность оценок максимального правдоподобия (ОМП); ОМП по группированным данным; MD-оценок, минимизирующих расстояния, задаваемые статистиками Колмогорова, и  Мизеса; -оптимальных оценок по выборочным квантилям. Для этого:

* моделируются выборки с засорением (2.1), где параметр ;
* оцениваются параметры распределения  всеми методами;
* сравниваются значения оценок при разных значениях .

Были смоделированы выборки с засорением



с различной долей засорения v=0, 0.1, 0.2. Объем выборки — 1000.

Оценим параметры нормального распределения требуемыми методами:

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Доля v  Оценка | v = 0 | | v = 0.1 | | v = 0.2 | |
| μ | σ | μ | σ | μ | σ |
| ОМП | 0.0025 | 0.9706 | 0.0368 | 1.6871 | 0.0556 | 2.2367 |
| ОМП по групп. | -0.0026 | 0.9862 | -0.0009 | 1.1264 | 0.0041 | 1.3357 |
| Колмогорова | -0.0065 | 0.9789 | -0.019 | 1.1572 | -0.0243 | 1.4416 |
| К-М-С, ɷ2 | -0.0116 | 0.976 | -0.0136 | 1.0969 | -0.0169 | 1.247 |
| L | -0.0013 | 0.9733 | -0.0000 | 1.0972 | 0.0028 | 1.2674 |

Группированная выборка составлена с помощью равночастотного группирования с числом интервалов 10.

2. Отбраковка аномальных наблюдений. Используя параметрическую процедуру отбраковки аномальных наблюдений очистить полученные в п.1 выборки от аномальных наблюдений и проверить, как изменились результаты применения критериев для проверки согласия эмпирического распределения с распределением  после удаления части наблюдений.

Для выборки с v=0 аномальные наблюдения не были обнаружены. У выборки v=0.1 было удалено 17 наблюдений. У выборки v=0.2 было удалено 58 наблюдений.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Доля v  Оценка | | v = 0.1 | | v = 0.2 | |
| μ | σ | μ | σ |
| ОМП | исходная | 0.0368 | 1.6871 | 0.0556 | 2.2367 |
| чистая | 0.0044 | 1.2494 | 0.0012 | 1.4283 |
| ОМП по групп. | исходная | -0.0009 | 1.1264 | 0.0041 | 1.3357 |
| чистая | -0.0084 | 1.0794 | -0.0146 | 1.1754 |
| Колмогорова | исходная | -0.019 | 1.1572 | -0.0243 | 1.4416 |
| чистая | -0.0236 | 1.1015 | -0.0388 | 1.243 |
| К-М-С, ɷ2 | исходная | -0.0136 | 1.0969 | -0.0169 | 1.247 |
| чистая | -0.017 | 1.0672 | -0.0260 | 1.1548 |
| L | исходная | -0.0000 | 1.0972 | 0.0028 | 1.2674 |
| чистая | -0.0072 | 1.0683 | -0.0213 | 1.1601 |

3. Построить функции влияния Хампеля для ОМП и ОМП по группированным данным.

4. Исследовать распределения статистик критериев типа Граббса, предназначенных для анализа на аномальность сразу нескольких наблюдений (*по заданию преподавателя*) в предположении о принадлежности выборки нормальному закону. Построить эмпирические распределения для статистик критериев типа Граббса, найти приближенные значения процентных точек. Для вариантов 1-5 применить критерий Граббса для отбраковки аномальных наблюдений по выборкам с засорением, полученным в п.1.

**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.