

САНКТ-ПЕТЕРБУРГСКИЙ ПОЛИТЕХНИЧЕСКИЙ
УНИВЕРСИТЕТ

Институт прикладной математики и механики

Высшая школа прикладной математики и
вычислительной физики

Отчет
по курсовой работе
по дисциплине

«Вычислительные комплексы»

Тема: Мода распределения интервальной выборки.
Мультимодальный случай.

Выполнил студент:
Смирнова Дарья
группа: 5030102/80201

Проверил:
к.ф.-м.н., доцент
Баженов Александр Николаевич

Санкт-Петербург
2022 г.

Содержание

	Страница
1 Постановка задачи	3
2 Теория	3
2.1 Мода интервальной выборки	3
2.2 Мультимодальный случай	3
2.3 Алгоритм нахождения моды в одномодальном случае .	4
2.4 Нахождение моды в многомодальном случае	4
3 Реализация	5
4 Результаты	5
5 Обсуждение	7

Список иллюстраций

	Страница
1 Интервальная выборка	5
2 Частоты пересечений интервалов C	6
3 Две моды интервальной выборки	7
4 Частоты пересечений интервалов с несколькими локаль- ными экстремумами рядом	8
5 Две моды интервальной выборки около значения $+5$. .	8

1 Постановка задачи

1. Сгенерировать интервальную выборку из 20 интервалов с двумя модами
2. Найти две моды выборки
3. Визуализировать результаты

2 Теория

2.1 Мода интервальной выборки

Распространим понятие моды и для работы с интервальными данными: модой интервальной выборки назовём совокупность интервалов пересечения наибольших совместных подвыборок выборки.

2.2 Мультимодальный случай

Мы сталкиваемся с мультимодальным случаем, когда в выборке находится несколько совместных подвыборок максимальной длины. В таком случае мода является мультиинтервалом.

При этом естественно говорить о наличии нескольких мод в случае, когда между двумя модами есть непустой промежуток с меньшей длиной совместной подвыборки (в точечном случае это означает, что мода отвечает локальному экстремуму)

2.3 Алгоритм нахождения моды в одномодальном случае

Алгоритм 1 Нахождение моды интервальной выборки

Условие: X - интервальная выборка

Результат: M - мода интервальной выборки μ - частота моды

$I \leftarrow \cup X$

Если $I \neq \emptyset$ **тогда**

$M, \mu \leftarrow I, |X|$

иначе

$C \leftarrow \emptyset$

Выполнить для всех $x \in X$ **выполнять**

$C \leftarrow C \cup \{\bar{x}, \underline{x}\}$

Конец цикла

$C \leftarrow \text{sort}(C)$

Цикл $i = 0, |C|-1$ **выполнять**

$\mu_i \leftarrow \text{intersectNum}([c_i, c_{i+1}], X)$

Конец цикла

$\mu \leftarrow \max \mu_i$

$M \leftarrow \cup([c_j, c_{j+1}], j = \text{argmax } \mu_i)$

Конец условия

Возвратить M, μ

2.4 Нахождение моды в многомодальном случае

Возьмем часть алгоритма для поиска одной моды: вычисляем частоты пересечения с исходной выборкой смоделированных интервалов $[c_i, c_{i+1}]$. Теперь мы имеем точечную выборку, у которой можно искать моду способами классической статистики: моде отвечает интервал с наибольшим числом μ_i .

Пусть пользователь исходя из природы данных и анализа указывает количество мод m , которое он предполагает увидеть. По выборке $\{\mu_i\}$ находим индексы локальных максимумов и сортируем их по убыванию значения экстремума. Выбираем m индексов из списка индексов отсортированных максимумов, интервалы, соответствующие этим индексам добавляем в результат.

3 Реализация

Лабораторная работа выполнена средствами языка python с использованием библиотек:

1. scipy
2. numpy
3. matplotlib

4 Результаты

Для генерации выборки были выбрана смесь двух нормальных распределений $N(-5, 1)$, $N(5, 1)$ с вероятностями по 0.5 для выбора центров интервалов выборки. Радиусы интервалов выборки получали как случайные величины, распределенные равномерно $U(0.3, 0.8)$. Полученная интервальная выборка проиллюстрирована ниже, где координата y выбрана произвольным образом для лучшего восприятия.

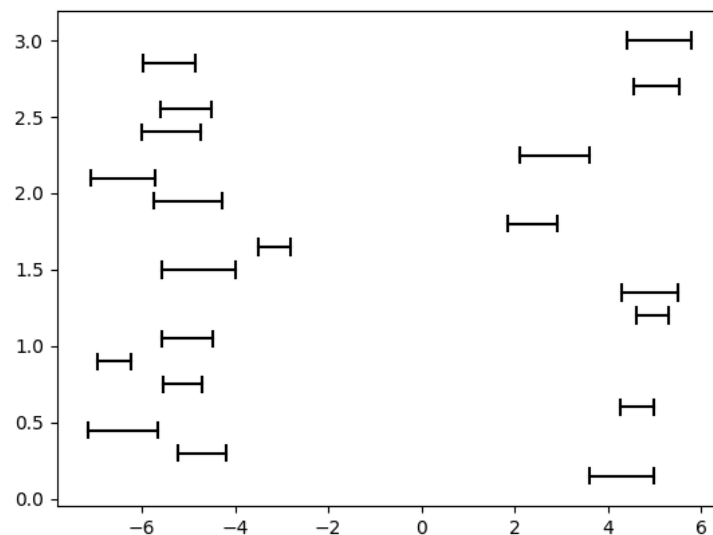


Рис. 1: Интервальная выборка

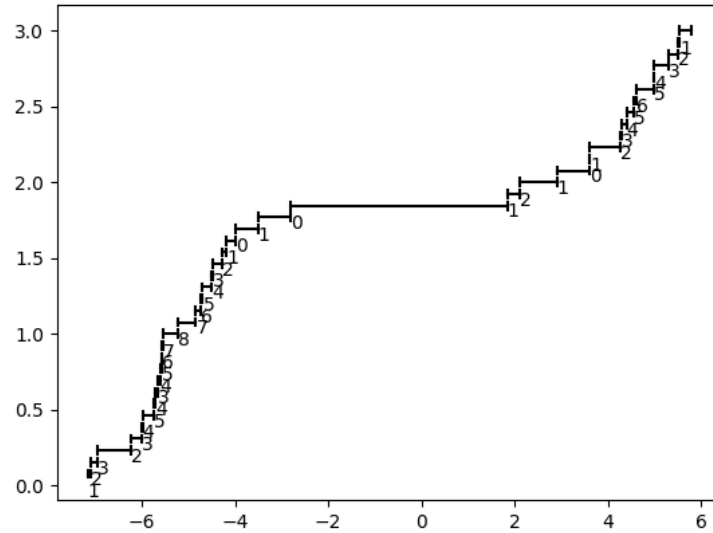


Рис. 2: Частоты пересечений интервалов S

После получения интервалов, соответствующих выборке S алгоритма рассчитаны соответствующие μ . При этом мы изначально "сглаживаем" интервалы из S : объединяются соседние интервалы с равными значениями μ . На графике изображены интервалы с подписанными внизу значениями частот μ_i .

По интервалам, сообщающим локальные максимумы и имеющим наибольшие значения экстремума, строим мультиинтервал, который является ответом, исходя из количества мод, которые указал пользователь (в нашем случае 2).

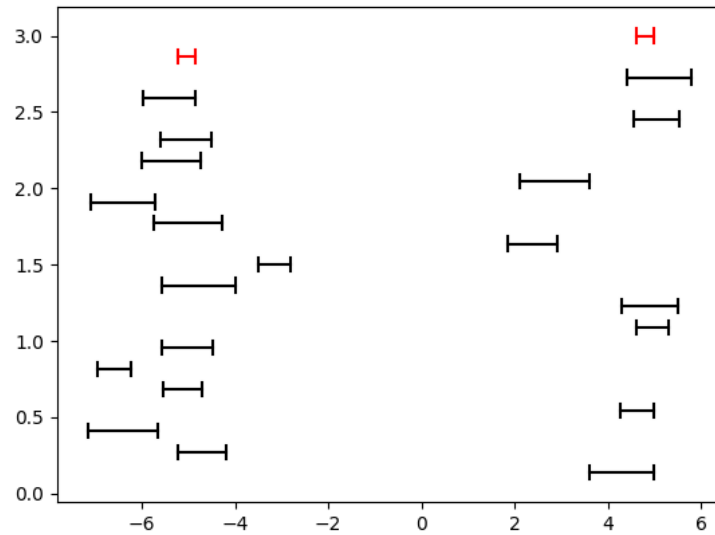


Рис. 3: Две моды интервальной выборки

5 Обсуждение

Данный алгоритм для одной моды применим всегда, для нескольких мод показывает хорошие результаты на выборках, у которых наибольшие локальные экстремумы частот пересечений (или в точечном случае плотности вероятности) соответствуют модам. Можно сгенерировать выборку, у которой вблизи одного пика будет несколько локальных экстремумов, величина которых будет больше экстремума частот, соответствующих второму явному пику выборки:

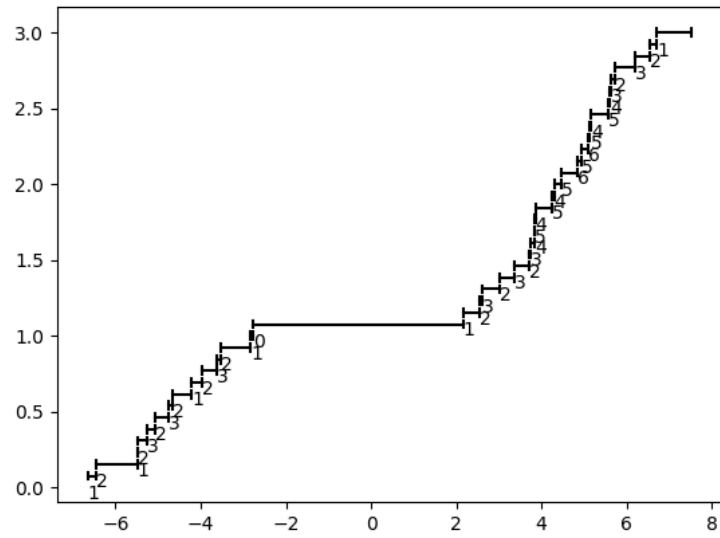


Рис. 4: Частоты пересечений интервалов с несколькими локальными экстремумами рядом

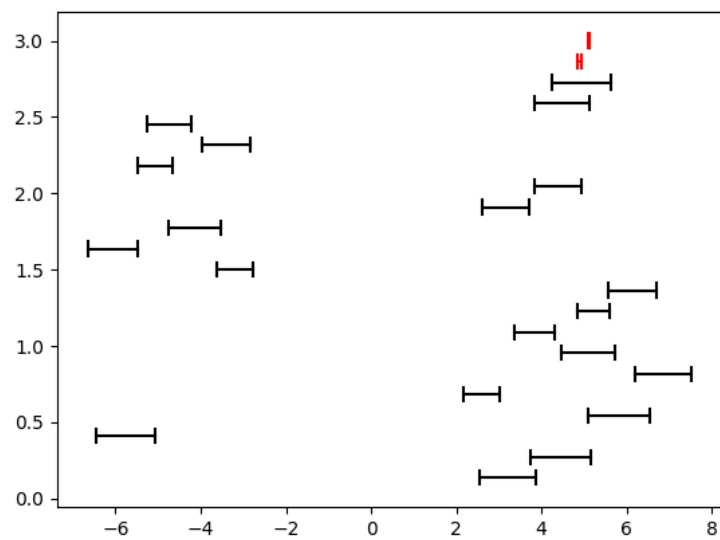


Рис. 5: Две моды интервальной выборки около значения +5

Примечание

С кодом работы можно ознакомиться по ссылке:
<https://github.com/DariaWelt/IntAnalysis>