1. Министерство образования и науки Российской Федерации
2. Санкт-Петербургский Политехнический Университет Петра Великого
3. —
4. Институт кибербезопасности и защиты информации

**КУРСОВАЯ РАБОТА**

1. **«Исследование распределение Рэлея»**
2. по дисциплине «Теория вероятностей и математическая статистика»
3. Выполнили
4. студенты гр. 4851003/10001 Васильев О. С.

<*подпись*>

1. Преподаватель
2. асс. преподавателя Штыркина А. А.

<*подпись*>

Санкт-Петербург

1. 2023

Содержание

[Содержание 2](#_Toc138408100)

[1 Цель работы 3](#_Toc138408101)

[2 Постановка задач 3](#_Toc138408102)

[3 Исследование распредедения 3](#_Toc138408103)

[4 Стандартные статистики 3](#_Toc138408104)

[5 Понятие эмпирической функции распределения 4](#_Toc138408105)

[6 Понятие гистограммы 8](#_Toc138408106)

[7 Описание параметров распределения 10](#_Toc138408107)

[8 Понятие точечных оценок 13](#_Toc138408108)

[9 Понятие интервальных оценок 21](#_Toc138408109)

[10 Понятие статистических критериев 25](#_Toc138408110)

[11 Выводы 34](#_Toc138408111)

[Список используемых источников 35](#_Toc138408112)

# Цель работы

Исследовать распределения Рэлея, и применить теоретические знания, полученные из курса теории вероятностей и математической статистики на практике.

# Постановка задач

1. Провести исследования распределения;
2. Найти стандартные статистики для выборки;
3. Построить эмпирические функции распределения для подвыборок;
4. Построить гистограммы для подвыборок;
5. Описать параметры распределения;
6. Построить точечные оценки;
7. Построить интервальные оценки;
8. Проверить гипотезу о параметрах распределения;
9. Проверить гипотезу о виде распределения;
10. Проверить гипотезу об однородности выборок.

# Исследование распредедения

История: Лорд Рэлей ввел распределение Рэлея в 1880 году в связи с проблемой в области акустики. В природе физические явления во многих областях науки распределения амплитуд могут быть охарактеризованы функцией плотности Рэлея или некоторой функцией, которая может быть получена из функции плотности Рэлея.[[1]](#пепв)

Данное распределение применяется во многих сферах жизни. Один из вариантов в физической океанографии распределение высоты волн приблизительно соответствует распределению Рэлея.[[2]](#втор) Также Распределение Рэлея используется в области питания для установления связи между уровнем питательных веществ в рационе и реакцией организма человека и животных.[[3]](#трет)Параметры, определяющие кривую, дают основу для выведения взаимосвязей между питательным веществом и физиологическими реакциями и объясняют питательные вещества и реакции.

# Стандартные статистики

Первоначально была реализовано считывание выборки с файла. Далее была написана функция, в которой вычисляются следующие статистики:

1. сумма элементов выборки = 1207.295484
2. выборочное среднее = 4.024318
3. Медиана = 4.790104
4. Мода = 3.873812
5. Размах = 3.326221
6. Смещенная дисперсия = 0.447881
7. Несмещенная дисперсия = 0.449378
8. выборочный начальный момент k-ого порядка = 70.776661
9. выборочный центральный момент k-ого порядка = 0.19504. В данном случае k = 3.

# Понятие эмпирической функции распределения

1. Слово эмпирическая означает, что функция вычисляется по данным опыта (эмпирическим данным), то есть, по выборке.

Определение эмпирической функции распределения:

где

1. Алгоритм построения:
2. Сортируем подвыборку, составляя вариационный ряд.
3. Подсчитываем количество элементов меньших x, где x каждый элемент подвыборки.
4. Строим эмпирическую функцию по полученным данным.
5. Получившиеся эмпирические функции распределения представлены на рисунках 1-3:

Изображение выглядит как текст, снимок экрана, диаграмма, Параллельный

Автоматически созданное описание

Рис. 1. Для 10 элементов.

Изображение выглядит как текст, диаграмма, линия, снимок экрана

Автоматически созданное описание

Рис. 2. Для 100 элементов.

Изображение выглядит как текст, диаграмма, линия, График

Автоматически созданное описание

Рис. 3. Для 200 элементов.

1. Получившиеся эмпирическая функция распределения похожа на теоретическую функцию распределения, особенно это прослеживается при увеличении элементов подвыборки. Вначале эмпирическая функции имеет обратный выгиб в отличии от середины, что также подтверждает о факте совпадения с теоретической функцией распределения.

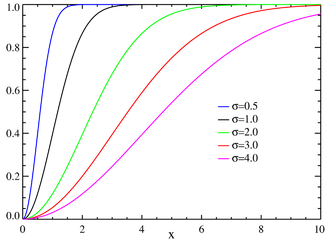


Рис. 4. Теоретическая функция распределения Рэлея.

Сравнивая эмпирические функции распределения и теоретическую, можно предположить, что параметр масштаба приблизительно равен одному или двум, а параметр смещения меньше 3.

# Понятие гистограммы

1. Гистограмма визуально представляет распределение случайной величины, при котором измеряется количество появления в наборе данных сходных значений. По оси х откладываются числовые значения, которые разбиты на диапазоны или интервалы. Каждому диапазону соответствует столбец гистограммы.
2. Алгоритм построения:
3. Разбиваем подвыборку на диапазоны.
4. Подсчитываем количество попаданий элементов из выборки в каждый диапазон
5. Строим гистограмму, где каждому диапазону соответствует столбец, высота столбца зависит от количества элементов, лежащих в этом диапазоне.
6. Далее была написана функция, которая делает подвыборки случайным образом размера 10, 100, 200. После по этим выборкам строятся гистограммы, представленные на рисунках 5-7:

Изображение выглядит как текст, снимок экрана, График, диаграмма

Автоматически созданное описание

Рис. 5. Для 10 элементов.

Изображение выглядит как текст, снимок экрана, диаграмма, График

Автоматически созданное описание

Рис. 6. Для 100 элементов

Изображение выглядит как текст, снимок экрана, диаграмма, График

Автоматически созданное описание

Рис. 7. Для 200 элементов

1. По полученным диаграммам можно сделать вывод, что данные подвыборки имеет плотность распределения похожую на плотность распределения Рэлея. Но по полученным данным сложно судить о параметрах распределения.

# Описание параметров распределения

1. Теоретической функции распределения Рэлея:

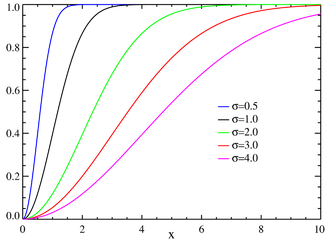


Рис. 8. Теоретическая функция распределения Рэлея.

1. Формула для вычисления теоретической функции распределения:
2. В распределение Рэлея два параметр .Первый параметр называется параметром масштаба. Этот параметр влияет на растяжение графика вдоль оси x. А второй параметр – параметр смещения (loc), который отвечает за смещения функции вдоль оси x.
3. Далее была написана функция на языке Python, которая строит теоретическую функцию распределения с заданными входными параметрами. С помощью этой функции получены графики, которые представлены на рисунках 9 - 10:

Изображение выглядит как текст, снимок экрана, линия, График

Автоматически созданное описание

Рис. 9. Теоретическая функция распределения с параметрами равными 2, 2

Изображение выглядит как текст, снимок экрана, линия, График

Автоматически созданное описание

Рис. 10. Теоретическая функция распределения с параметрами равными 5, 5

Видно, что при увеличении значения параметра масштаба, график растягивается вдоль оси Х. А при увеличении параметра смещения, график смещается вправо.

# Понятие точечных оценок

1. Оценим параметры распределения с помощью метода моментов. Исходя из условия, что сигма равна 1, получим следующую функцию плотности:

Найдем теоретический момент первого порядка. Для этого проинтегрируем плотность умноженную на x:

Посчитаем выборочный момент первого порядка:

выборочное среднее было посчитано ранее.

Далее найдем параметр

1. Оценим параметр методом максимального правдоподобия:

Приведем выражение под один знак суммы и к общему знаменателю:

Для решения данного уравнения должна выполняться система:

Распишем первое уравнение подробнее:

1. **Несмещенность**

Оценка \widehat\theta_n параметра \theta называется несмещенной, если ее математическое ожидание равно истинному значению оцениваемого параметра:

\mathbb{E}\widehat\theta_n=\theta.

Метод моментов:

Следовательно, оценка, полученная методом моментов, несмещенная.

Метод максимального правдоподобия:

Следовательно, оценка, полученная методом максимального подобия, смещенная.

1. **Состоятельность**

Состоятельная оценка в математической статистике — это точечная оценка, сходящаяся по вероятности к оцениваемому параметру.



Также для того, чтобы оценка была состоятельной необходимо:

* оценка была несмещенной;
* дисперсия оценки стремилась к 0.

Метод моментов:

Следовательно оценка, полученная методом моментов, состоятельная.

Метод максимального правдоподобия:

Значение дисперсии обозначили как c (константа).

Следовательно, оценка, полученная методом моментов, не состоятельная.

1. **Эффективность**

Эффективной называют статистическую оценку, которая при заданном объеме выборки имеет наименьшую возможную дисперсию. Если – несмещённая оценка параметра , построенная по выборке , то дисперсия оценки ограничена снизу следующим образом: . Отмечается также, что должны соблюдаться условия регулярности семейства .

Эффективность несмещённой оценки : . По Рао-Крамеру, несмещённая оценка является эффективной, если . Соответственно, оценку, для которой в неравенстве Рао – Крамера достигается равенство, называют R-эффективной оценкой.

Также необходимое и достаточное условие существования эффективной оценки:

где h(x),s(x) зависят только от x, A(loc), B(loc) зависят только от loc.

Если посмотреть на плотность распределения Рэлея

то можно увидеть, что в данной случае не существует эффективной оценки, так как h(x)= – зависит не только от x. Следовательно не существует эффективной оценки.

На рисунке 11 приведены графики теоретической функции распределения с полученными параметрами и эмпирической функции распределения.

Изображение выглядит как текст, диаграмма, снимок экрана, линия

Автоматически созданное описание

Рис. 11. Теоретическая функция распределения с полученными параметрами и эмпирическая функция распределения.

# Понятие интервальных оценок

Пусть

Тогда возрастающая последовательность, подчиняющаяся экспоненциальному распределению. Исходя из этого результата, можно получить можно получить следующее:

которые являются независимыми случайными величинами из стандартного экспоненциального распределения.

подчиняется распределению хи-квадрат с двумя степенями свободы.

также подчиняется распределению хи-квадрат с 2(n-1) степенями свободы.

Рассмотрим следующую функцию:

Данная функция имеет F распределение с 2(n − 1) и 2 степенями свободы.

Тогда доверительный интервал для loc будет следующим:

* Рассчитаем интервал с уровнем доверия :

Получился интервал:

()

Изображение выглядит как текст, линия, диаграмма, Шрифт

Автоматически созданное описание

Рис. 12. Близость значений интервальных и точечных оценок.

* Рассчитаем интервал с уровнем доверия :

()

Изображение выглядит как линия, текст, диаграмма, антенна

Автоматически созданное описание

Рис. 13. Близость значений интервальных и точечных оценок.

# Понятие статистических критериев

1. **Гипотезы о параметрах распределениях**

Задача: построить наиболее мощный критерий для проверки нулевой гипотезы вида для каждого параметра распределения . Здесь оценка параметра распределения , полученная в результате применения метода моментов (), метода максимального правдоподобия (). В качестве уровня значимости принять .

Для построения наиболее мощного критерия по теореме Неймана-Пирсона - критерий отношения правдоподобия.

1. **Для метода моментов.**

* Функция максимального правдоподобия для :
* Аналогичным образом будет рассчитана функция максимального правдоподобия для :

Отношение правдоподобия:

Критерий гласит: если , то принимаем основную гипотезу, иначе отвергаем ее в пользу альтернативной.

Согласно асимптотическому свойству метода отношения правдоподобия, для выборок большого объема эта величина имеет распределение хи-квадрат, а значит величину C можно найти, как квантиль распределения хи-квадрат.

Имеем распределение хи-квадрат с 1 степенью свободы, уровень значимости в нашем случае 0.05, тогда

Значит верна гипотеза .

1. **Для метода максимального правдоподобия**
2. Функция максимального правдоподобия для :

**Вывод**

Следовательно, нужно принять гипотезу .

1. **Ошибка первого рода**

Уровень значимости – это ошибка первого рода, поэтому

1. **Гипотезы о виде распределения**

Данный критерий позволяет определить является ли выборка из распределения Рэлея.

1. **Критерий Колмагорова.**

Критерий Колмогорова приемлем при n >= 20 для проверки гипотезы, подчиняется ли случайная величина некоторому теоретическому закону распределения, если его [параметры](http://www.arhiuch.ru/lab1.html) предполагаются известными.

Результаты располагают в вариационном ряду. Находят верхнюю и нижнюю границы соответствующего отклонения:

рисунок 11.2

рисунок 11.1

Здесь F(xi) – значения теоретической функции распределения.

Выбирают максимальную из границ отклонений:

рисунок 11.3

Статистику критерия можно рассчитать по формуле

рисунок 11.4

Для вычисления был разработан скрипт на ЯП Python.

Для параметра, полученным методом моментов,

При уровне значимости равном , . От сюда следует, что для параметра, полученным методом моментов, выборка принадлежит распределению из варианта, так как

**Вывод**

Для параметра, полученного методом максимального правдоподобия,

Однако, для параметра, полученным методом максимального правдоподобия, выборка не принадлежит распределению из варианта, так как

1. **Критерий**

K – количество непересекающихся интервалов

] - количество наблюдений в j-м интервале

=F()-F() - вероятность попадания наблюдения в j-ый интервал

=n ожидаемое число попаданий в j-ый интервал

=

Для этого разобьем выборку на 10 отрезков. Также был написан скрипт, который считает количество и вероятность попаданий в отрезки. Далее в зависимости от параметра считается значение критерия. Полученные значения:

* Метод моментов: 6.407608
* Метод максимального правдоподобия: 181.109775

Табличное значения хи-квадрат для 9 степеней свободы и уровню значимости 0.05 равно 16.919.

**Вывод**

Следовательно, с помощью этого критерия мы принимаем гипотезу о том, что параметр равен значению, которое было получено с помощью метода моментов. А параметр, полученный с помощью максимального правдоподобия, отвергаем в пользу альтернативного.

1. **Гипотезы об однородности выборок**

Было выбрано экспоненциальное распределение.

Гипотезы:

Подберем параметр графически. При параметре получилась функция, которая больше всего похожа на распределения Рэлея.

Изображение выглядит как линия, График, снимок экрана, диаграмма

Автоматически созданное описание

Рис. 14. Распределение с параметром 0.64

Найдем обратную функцию экспоненциального распределения:

С помощью обратной функции была сгенерирована выборка для двух параметров. Далее был выбран критерий Смирнова. Получены следующие значения критерия:

* Метод моментов: = 2.327015
* Метод максимального правдоподобия: = 4.123308

**Вывод**

Получается, что для двух параметров, критерий не сработал, так как , где d = 1.36. Значит, принимаем гипотезу

Для вычисления мощности критерия составим обратную функцию для распределения Рэлея:

Теперь будем генерировать выборки с экспоненциальным распределением и распределением Рэлея. Далее по критерию Смирного будем определять выполняется данный критерий или нет. В качестве экспериментальной оценки можно n раз проверить какой результат будет у критерий в случае, когда основная гипотеза не выполняется. И считаем сколько раз он дал верный ответ. Для этого был написан скрипт, который автоматизирует данный алгоритм.

Полученный результат:

|  |  |
| --- | --- |
| Размер выборки | Мощность |
| 30 | 0,12 |
| 50 | 0,4 |
| 100 | 0.76 |

# Выводы

В ходе выполнения курсовой работы применялись фундаментальные теоретические знания на практике для анализа распределения Рэлея. Исходя из предоставленной выборки, были построены различные графические представления: эмпирическая функция распределения и гистограмма, которые позволили описать само распределение и определить его параметры. Кроме того, были получены точечные оценки при помощи метода максимального правдоподобия и метода моментов, а также интервальные оценки. Были исследованы свойства полученных точных оценок. В дополнение к этому были разработаны статистические критерии для проверки параметров данного распределения.

Список используемых источников

[1] - <https://www.sciencedirect.com/science/article/pii/S1018364720301622>

[2] - <https://www.usna.edu/NAOE/_files/documents/Courses/EN330/Rayleigh-Probability-Distribution-Applied-to-Random-Wave-Heights.pdf> (VPN)

[3] - <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5697816/>

4 - https://www.hindawi.com/journals/jps/2016/8246390/