# **Расчетное задание I-II по математической статистике**

# Введение

Расчетное задание I-II предназначено для студентов специальности 09.03.01 ИНФОРМАТИКА И ВЫЧИСЛИТЕЛЬНАЯ ТЕХНИКА (Компьютерные науки и системотехника) факультета информационных технологий НГУ.

Задание состоит из двух частей. В первой части требуется найти оценку неизвестного параметра с помощью метода максимального правдоподобия, исследовать ее свойства: несмещенность, состоятельность, построить доверительный интервал для параметра.

Во второй части требуется вычислить ряд выборочных статистик, построить графики эмпирической функции распределения, гистограммы и ядерной оценки плотности, вычислить доверительные интервалы для среднего и дисперсии.

Задания I-II рекомендуется выполнять с использованием компьютерной техники и языка Python.

Отчет по расчетному заданию должен содержать титульный лист, лист задания, текст решения задач с необходимой степенью детализации, ссылки на соответствующие теоремы, свойства, статистические таблицы, использованные при получении решения.

# **Часть I. Оценивание параметров**

Пусть имеется выборка  из распределения случайной величины с функцией распределения . В общем случае задача оценивания заключается в том, чтобы, используя статистическую информацию, доставляемую выборкой, сделать статистические выводы об истинном значении неизвестного параметра .

*Точечной* оценкой неизвестного параметра  по выборке  называется значение некоторой статистики , которое приближенно равно значению параметра :.

*Интервальной* оценкой (или *доверительным интервалом*) параметра  называют интервал , содержащий истинное значение параметра  с вероятностью .

## 1.1. Метод максимального правдоподобия

*Оценкой максимального правдоподобия* (ОМП) параметра  называется точка параметрического множества , в которой функция максимального правдоподобия  достигает наибольшего значения: .

Если для любой выборки  из выборочного пространства максимум  достигается во внутренней точке , и  дифференцируема по , то ОМП  удовлетворяет уравнению , которое называется *уравнением правдоподобия*.

***Пример 1.2***

Построить оценку максимального правдоподобия параметра  распределения Бернулли: , .

Решение:

Логарифмическая функция правдоподобия равна

=

=;

 => ,

где – среднее выборочное значение.



.

Проверим знак второй производной при :



.

Таким образом, при  функция правдоподобия достигает максимума.

***Пример 1.3***

Построить оценку максимального правдоподобия параметра  равномерного распределения на отрезке .

Решение:

Функция правдоподобия выборки равна

 =

=,

где – максимальная порядковая статистика.

При фиксированных значениях выборки (и, следовательно, при фиксированном значении ) зависимость  от  показана на рисунке 1.1. Максимум функции правдоподобия достигается в точке . Поэтому искомая оценка максимального правдоподобия есть .

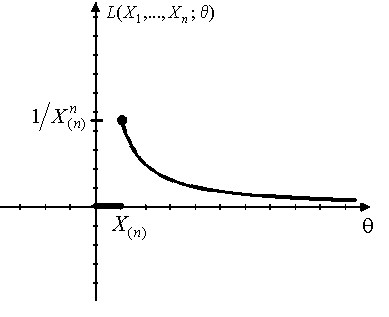


Рис. 1.1. Функция правдоподобия

## 1.2. Построение доверительного интервала с использованием центральной статистики

Статистика  называется *центральной статистикой*, если распределение  не зависит от , и при любом фиксированном  статистика  непрерывна и строго монотонна по .

С помощью центральной статистики можно построить доверительный интервал. Пусть  плотность распределения статистики .

1. Найдем такие значения , что .

2. Решим относительно , уравнения

.

3. Определяем границы доверительного интервала:

,

Для построения доверительного интервала с помощью центральной статистики основная проблема заключается в нахождении этой центральной статистики. Можно выделить класс моделей, для которых центральная статистика существует и имеет простой вид.

Пусть  – функция распределения наблюдаемой случайной величины, *монотонная* по параметру . Можно положить в качестве центральной статистики функцию , которая подчинена гамма-распределению с параметром формы .

***Пример 1.4***

Построить точный -доверительный интервал по выборке  для параметра  экспоненциального распределения , .

Решение:

Функция распределения  является монотонной (возрастающей) по параметру  (), следовательно, в качестве центральной статистики можно взять , которая подчинена гамма-распределению с функцией плотности , , ,  – объем выборки.

Тогда границы -доверительного интервала  определяются при численном решении уравнений: , , где  и  выбираются такими, что .

## 1.3. Построение асимптотического доверительного интервала

Оценки максимального правдоподобия при достаточно общих условиях являются асимптотически эффективными и асимптотически нормальными, следовательно

,

где  – функция распределения стандартного нормального закона,  – информационное количество Фишера,  – ОМП. Отсюда , тогда  – асимптотически кратчайший -доверительный интервал для .

## 1.2. Свойства оценок параметров

### 1.2.1. Несмещенность

Статистика  называется *несмещенной* оценкой параметра , если выполняется условие: .

### 1.2.2. Состоятельность

Оценка  некоторой функции  называется *состоятельной*, если , , при . То есть  , при .

Свойство состоятельности обязательно для любого правила оценивания, однако оно является асимптотическим и не связано со свойствами оценки при фиксированном объеме выборки (в отличие от свойств несмещенности и минимальной дисперсии).

**Критерий состоятельности**. Пусть  и  при . Тогда  – состоятельная оценка функции .

***Пример 1.6***

Пусть  – выборка из распределения Максвелла с функцией плотности , .

Требуется проверить оценку  на несмещенность, состоятельность.

Решение:

1. *Несмещенность.*

,

 оценка  является несмещенной оценкой параметра .

1. *Состоятельность.*

Поскольку  является несмещенной, то нам достаточно исследовать дисперсию оценки .

, ,

 по критерию состоятельности, оценка  является состоятельной.

## 1.3. Задание I

Пусть  – выборка из заданного в соответствии с вариантом закона распределения.

1.1. Найти числовые характеристики заданной модели:

* 1. математическое ожидание (1 балл);
  2. дисперсию (1 балл).

1.2. Найти точечную оценку неизвестного параметра  по методу максимального правдоподобия или методу моментов (2 балла).

1.3. Построить асимптотический доверительный интервал для  (1 балл).

# **Часть II. Выборочные характеристики, доверительные интервалы, проверка гипотез**

## 2.1. Выборочные характеристики

Пусть  — *выборка* объема , получаемая в результате  наблюдений *случайной величины*. Будем считать, что наблюдения (случайные величины)  независимы и имеют одну и ту же функцию распределения .

Реализацией выборки[[1]](#footnote-1) называется множество реализаций случайной величины .

**Определение***. Статистикой* называется измеримая функция от выборки.

Важными примерами статистик являются выборочные моменты. Для *выборочного среднего* используется обозначение

,

для *выборочной дисперсии*  и *несмещенной выборочной дисперсии*  используются обозначения

 и .

*Медиана* - корень уравнения . Если длина вариационного ряда нечетная, т. е. , то ; если , то .

*Вариационный размах*: .

## 2.2. Эмпирическая функция распределения

Введем вспомогательную случайную функцию:  – количество наблюдений в выборке случайной величины , которые меньше . Событие  означает, что в интервал  попало ровно  наблюдений, а в интервал  попало  наблюдений.

**Определение.** Функция  называется *эмпирической функцией распределения*.

## 2.3. Гистограмма

Разобьем область значений случайной величины  на  интервалов , и подсчитаем количество наблюдений, попавших в каждый интервал:  – количество наблюдений в интервале , при этом .

На графике для каждого интервала строим столбцы гистограммы шириной  и высотой .

Если область значений случайной величины бесконечная, то для построения гистограммы можно взять такие границы , чтобы в интервал  попадали все наблюдения.

Гистограмма довольно грубый способ оценивания плотности распределения, связанный с неопределенностью выбора числа интервалов , границ интервалов, а также потерей информации при группировании.

## 2.4. Ядерная оценка плотности

Пусть дана ядерная функция , удовлетворяющая условиям:

; ; ; ; .

Тогда ядерная оценка функции плотности имеет вид

,

где  – параметр размытости ядерной функции, который для сходимости оценки к функции плотности должен удовлетворять условиям , .

## 

## 2.5. Задание II

По заданному набору данных одной переменной (приложение):

2.1. Вычислить выборочные характеристики (2 балла):

* выборочное среднее,
* выборочную дисперсию,
* несмещенную выборочную дисперсию,
* размах,
* медиану.

2.2. Построить (3 балла):

* график эмпирической функции распределения,
* гистограмму,
* ядерную оценку функции плотности.

2.3. Построить 99% - доверительный интервал (в предположении, что выборка подчиняется нормальному распределению с неизвестными параметрами) (3 балла)

* для математического ожидания
* для дисперсии

**приложение**

VAR 11

1.096845e+00 -5.287447e-01 1.156123e+00 2.173618e+00 7.837808e-01 2.143643e+00

1.351229e+00 1.873702e+00 1.662140e+00 5.608621e-01 1.785185e-01 9.584880e-01

7.294377e-01 1.550325e+00 3.996763e-01 9.086647e-01 5.773610e-01 2.085788e+00

1.163254e+00 3.407647e+00 2.801604e+00 1.882082e+00 4.413660e-01 3.196276e+00

9.298778e-01 1.524579e-01 4.225722e-01 1.646672e+00 1.094645e+00 4.875063e-01

1.286636e+00 1.042466e+00 8.938204e-01 1.792667e+00 1.037169e+00 2.916249e+00

5.223442e-01 -1.654540e+00 1.953347e+00 1.449721e+00 1.681378e+00 1.515203e+00

1.736902e+00 1.315879e+00 -7.288273e-01 2.026027e+00 1.848299e+00 2.390904e+00

1.212489e+00 2.286411e+00 1.016488e+00 -7.526284e-02 1.825489e+00 -1.320122e+00

2.048697e+00 -5.064952e-01 2.337649e+00 1.179483e+00 1.298952e+00 1.379829e+00

1.855662e+00 3.125316e+00 1.787854e+00 3.305784e-01 3.513121e+00 8.786156e-01

2.065896e+00 -3.195336e-02 4.262330e-01 2.119323e+00 1.373405e+00 6.788910e-02

5.102424e-01 -2.889807e-01 -4.171255e-01 1.864159e+00 -1.418449e+00 5.905055e-01

4.030473e-01 2.394121e+00 2.019800e+00 2.175578e+00 2.052527e+00 1.173415e+00

5.688432e-01 2.613130e+00 8.452779e-01 6.408093e-01 1.354768e+00 9.001668e-01

2.008914e+00 -7.106124e-01 1.763246e+00 -8.571812e-01 1.587911e+00 2.769232e-01

-1.318476e-01 2.234600e+00 -1.772122e-01 3.018062e-01 2.253935e-01 3.818298e-01

2.133902e+00 -3.128048e-01 1.694928e-01 7.131484e-01 1.108125e+00 8.055310e-01

1.889021e+00 2.971893e+00 2.056264e+00 1.728596e+00 4.349400e-01 9.122648e-01

9.775475e-02 9.514703e-01 -1.092183e+00 1.198577e+00 8.310402e-01 4.656085e-01

1. В дальнейшем не будет делаться различий между выборкой и ее реализацией: это будет определяться контекстом. Например, если дана выборка в виде набора чисел, то это реализация выборки. [↑](#footnote-ref-1)