Санкт-Петербургский политехнический университет

Институт прикладной математики и механики

Кафедра прикладной математики и информатики

**Отчет 8**

**«Математическая статистика»**

Группа 3630102/70301

Выполнил студент Ли Жуйци

Преподаватель Баженов А.Н.

Санкт-Петербург

2020

Оглавление

[Постановка задачи 2](#_Toc38231920)

[теория ..](#теория)...........................................................................................................................................................4

# [Результат.....](#результат)...................................................................... .........................................................6

[обсуждение](#обсуждение) ....................................................................................................................................8

[Список литературы ........................................................................................................................................9](#список)

# 

# Постановка задачи

# Для двух выборок размерами 20 и 100 элементов, сгенерированных согласно нормальному закону N(x;0;1) , для параметров положения и масштаба построить асимптотически нормальные интервальные оценки на основе точечных оценок метода максимального правдоподобия и классические интервальные оценки на основе статистик 2 χ и Стьюдента. В качестве параметра надёжности взять γ = 0,95.

### 2 [Теория](#теория)

### 2.1 Метод максимального правдоподобия

Пусть есть выборка X1, ... , Xn {\displaystyle X\_{1},\ldots ,X\_{n}} из распределения{\displaystyle \mathbb {P} \_{\theta }} P0 , где {\displaystyle \theta \in \Theta } — неизвестные параметры. Пусть {\displaystyle L(\mathbf {x} \mid \theta )\colon \Theta \to \mathbb {R} }L(x | ) : —> R функция правдоподобия, где x Точечная оценка

 {\displaystyle {\hat {\theta }}\_{\mathrm {M\Pi } }={\hat {\theta }}\_{\mathrm {M\Pi } }(X\_{1},\ldots ,X\_{n})=\mathop {\rm {argmax}} \limits \_{\theta \in \Theta }L(X\_{1},\ldots ,X\_{n}\mid \theta )}

называется оце́нкой максима́льного правдоподо́бия параметра  {\displaystyle \theta }. Таким образом оценка максимального правдоподобия — это такая оценка, которая максимизирует функцию правдоподобия при фиксированной реализации выборки.

Часто вместо функции логарифимическую L правдоподобия {\displaystyle L}L используют функциюправдоподобия l = lnL  {\displaystyle l=\ln L}Так как функция x->lnx,x>0 монотонно возрастает{\displaystyle x\to \ln x,\;x>0}  на всей области определения,максимум  любой функции {\displaystyle L(\theta )}L()является максимумом функции {\displaystyle \ln L(\theta )}lnL(), и наоборот. Таким образом,

{\displaystyle {\hat {\theta }}\_{\mathrm {M\Pi } }=\mathop {\rm {argmax}} \limits \_{\theta \in \Theta }l(X\_{1},\ldots ,X\_{n}\mid \theta )} 

Если функция правдоподобия дифференцируема, то необходимое условие экстремума — равенство нулю её градиента

{\displaystyle g(\theta )={\frac {\partial l(\mathbf {x} ,\theta \_{0})}{\partial \theta }}=0}

Достаточное условие экстремума может быть сформулировано как отрицательная определённость гессиана — матрицы вторых производных:



{\displaystyle H={\frac {\partial ^{2}l(\mathbf {x} ,\theta \_{0})}{\partial \theta \partial \theta ^{T}}}}

Важное значение для оценки свойств оценок метода максимального правдоподобия играет так называемая информационная матрица , равная по определению:



{\displaystyle I(\theta )=E[g(\theta )g(\theta )^{T}]}

В оптимальной точке информационная матрица совпадает с математическим ожиданием гессиана, взятым со знаком минус:

{\displaystyle I=-E(H\_{0})}  *I=-E(H0)*

i

**2.2 Доверительные интервалы для параметров нормального распределения**

**2.2.1 Доверительный интервал для математического ожидания m нормального распределения**

Дана выборка (𝑥1, 𝑥2, ... , 𝑥𝑛) объёма 𝑛 из нормальной генеральной совокупности. На её основе строим выборочное среднее 𝑥 и выборочное среднее квадратическое отклонение 𝑠. Параметры 𝑚 и 𝜎 нормального распределения неизвестны. Доказано, что случайная величина

T= \*

называемая статистикой Стьюдента, распределена по закону Стьюдента с 𝑛 − 1 степенями свободы. Пусть 𝑓𝑇 (𝑥) — плотность вероятности этого распределения. Тогда

# P(-x< \*<x)=P(-x<\*)=

# Здесь — функция распределения Стьюдента с 𝑛 − 1 степенями свободы.

# Полагаем 2−1 = 1−𝛼, где 𝛼 — выбранный уровень значимости. Тогда = 1 − 𝛼/2. Пусть (𝑛 − 1) — квантиль распределения Стьюдента с 𝑛 − 1 степенями свободы и порядка 1 − 𝛼/2. Из предыдущих равенств мы получаем

# P( -) = 2−1 = 1−𝛼

# P( -) = 2−1 = 1−𝛼

# 2.2.2 Доверительный интервал для среднего квадратического отклонения 𝜎 нормального распределения

# Дана выборка (𝑥1, 𝑥2, ... , 𝑥𝑛) объёма 𝑛 из нормальной генеральной совокупности. На её основе строим выборочную дисперсию . Параметры 𝑚 и

# 𝜎 нормального распределения неизвестны. Доказано, что случайная величина 𝑛/ распределена по закону 𝜒 2 с 𝑛 − 1 степенями свободы. Задаёмся уровнем значимости 𝛼 и, например, с помощью встроенных средств языка программирования R (функция qchisq) находим квантили ) и ). Это значит, что

# P (𝜒 2(n-1) <)) =

# P (𝜒 2(n-1) <)) =

# Тогда

# P () <𝜒 2(n-1) <))= P (𝜒 2(n-1) <))=1- - =1-

# Отсюда

# 

# Окончательно

# 

# 3 Результат

# 3.1 Доверительные интервалы для параметров нормального распределения

# Нормальный интервал n = 20

|  |  |  |  |
| --- | --- | --- | --- |
|  | Interval a | Interval b | Interval a-m |
| Ожидание n=20 | 0.675 | 0.207 | -0.882 |
| Дисперсия n=20 | 0.754 | 1.260 | -0.506 |
| Ожидание n=100 | -0.218 | 0.184 | -0.402 |
| Дисперсия n=100 | 0.871 | 1.183 | -0.402 |

# -0.6750904325107864 < m < 0.2075706344734736 -0.88266106698426

# Дисперсия

# 0.7541050613650225< m < 1.259864476256713 -0.5057594148916905

# Нормальный интервал n = 100

# Мат.ожидание

# -0.21837799429143964< d < 0.18421708622238173 -0.40259508051382137

# Дисперсия

# 0.870925857726643 < d < 1.1831306755071396 -0.3122048177804966

# 2

|  |  |  |  |
| --- | --- | --- | --- |
|  | Interval a | Interval b | Interval a-m |
| Ожидание Сьютенд n=20 | -0.633 | 0.165 | -0.798 |
| Дисперсия n=20 | 0.765 | 1.470 | -0.705 |
| Ожидание Сьютенд n=100 | -0.221 | 0.187 | -0.408 |
| Дисперсия n=100 | 0.901 | 1.192 | -0.291 |

# Классический интервал n = 20

# ожидание - Сьютенд

# -0.6332133817223725< m < 0.1656935836850598 -0.7989069654074323

# Дисперсия - Хи-квадрат

# 0.7652472069682972 < d <1.470485851122993 -0.7052386441546958

# Классический интервал n = 100

# ожидание - Сьютенд

# -0.2218900235449898<m< 0.18772911547593188 -0.4096191390209217

# Дисперси - Хи-квадрат

# 0.9017378086432567 <d<1.1922610043282371 -0.2905231956849804

# обсуждение

# Доверительные интервалы для параметров распределения

# ∙ Генеральные характеристики (𝑚 = 0 и 𝜎 = 1) накрываются построенными доверительными интервалами

# ∙ Доверительные интервалы, полученные по большей выборке, являются соответственно более точными, т.е. меньшими по длине

# ∙ Доверительные интервалы для параметров нормального распределения более надёжны, так как основаны на точном, а не асимптотическом распределении

# Список литературы

### 1: Доверительный интервал

# <https://ru.wikipedia.org/wiki/%D0%94%D0%BE%D0%B2%D0%B5%D1%80%D0%B8%D1%82%D0%B5%D0%BB%D1%8C%D0%BD%D1%8B%D0%B9_%D0%B8%D0%BD%D1%82%D0%B5%D1%80%D0%B2%D0%B0%D0%BB>

<http://www.machinelearning.ru/wiki/index.php?title=%D0%94%D0%BE%D0%B2%D0%B5%D1%80%D0%B8%D1%82%D0%B5%D0%BB%D1%8C%D0%BD%D1%8B%D0%B9_%D0%B8%D0%BD%D1%82%D0%B5%D1%80%D0%B2%D0%B0%D0%BB>