

Санкт-Петербургский политехнический университет Петра Великого

Институт прикладной математики и механики

Кафедра «Телематика (при ЦНИИ РТК)»

Отчет по лабораторной работе

**Метод максимального правдоподобия
и проверка гипотез критерием хи-квадрат**

По дисциплине «Теория вероятностей и математическая статистика»

Выполнил

Студент гр.3630201/80101

В.Н. Сеннов

Руководитель

доцент к.ф.-м.н.

А.Н. Баженов

«___» _____ 202__г.

Санкт-Петербург
2020

Содержание

1	Постановка задачи	4
2	Математическое описание	5
2.1	Метод максимального правдоподобия	5
2.2	Проверка гипотезы о законе распределения методом хи-квадрат	5
3	Особенности реализации	6
4	Результаты работы программы	7
4.1	Метод максимального правдоподобия	7
4.2	Критерий хи-квадрат	7
4.3	Проверка чувствительности критерия хи-квадрат	7
	Заключение	9
	Список литературы	10
A	Репозиторий с исходным кодом	11

Список таблиц

1	Вычисление χ_B^2 при проверке гипотезы о нормальном законе распределения . .	7
2	Вычисление χ_B^2 при проверке гипотезы о нормальном законе распределения для выборки $L(0, 1)$	7
3	Вычисление χ_B^2 при проверке гипотезы о нормальном законе распределения для выборки $N(-\sqrt{3}, \sqrt{3})$	8

1 Постановка задачи

В рамках данной работы необходимо сгенерировать выборку размером 100 элементов для нормального распределения $N(0, 1)$. Для нее найти методом максимального правдоподобия оценки параметров μ и σ для функции распределения $N(\mu, \sigma)$. Проверить эту гипотезу критерием χ^2 при уровне значимости $\alpha = 0.05$. Привести таблицу вычисления χ^2 .

Также необходимо сгенерировать выборку размером 20 элементов для распределения Лапласа $L(0, 1)$ и равномерного распределения $U(-\sqrt{3}, \sqrt{3})$. Для них необходимо проделать те же действия.

2 Математическое описание

2.1 Метод максимального правдоподобия

Для выборки x_1, \dots, x_n и функции распределения f с вектором параметров θ определяют функцию правдоподобия L :

$$L(x_1, x_2, \dots, x_n, \theta) = f(x_1, \theta) \cdot f(x_2, \theta) \cdots f(x_n, \theta)$$

Оценка максимального правдоподобия $\hat{\theta}$ находится по формуле:

$$\hat{\theta} = \arg \max_{\theta} L(x_1, x_2, \dots, x_n, \theta).$$

Если f дифференцируема, то оценка максимального правдоподобия находится из системы уравнений:

$$\frac{\partial L}{\partial \theta_k} = 0 \quad \text{или} \quad \frac{\partial \ln L}{\partial \theta_k}, \quad k = 1, 2, \dots, m. \quad (1)$$

2.2 Проверка гипотезы о законе распределения методом хи-квадрат

Пусть выдвинута гипотеза H_0 о генеральном законе распределения с функцией $F(x)$, которая не содержит неизвестных параметров.

Проверить эту гипотезу можно по следующему алгоритму:

1. Выбираем уровень значимости α .
2. Разбиваем выборку на k промежутков Δ_i .
3. Находим квантиль $\chi_{1-\alpha}^2(k-1)$ по таблице [2].
4. Вычисляем $p_i = P(x \in \Delta_i)$ с помощью $F(x)$.
5. Находим частоты n_i попадания в промежуток Δ_i .
6. Вычисляем выборочное значение χ^2 :

$$\chi_B^2 = \sum_{i=1}^k \frac{(n_i - np_i)^2}{np_i} \quad (2)$$

7. Сравниваем χ_B^2 и квантиль $\chi_{1-\alpha}^2(k-1)$.

- (a) Если $\chi_B^2 < \chi_{1-\alpha}^2(k-1)$, то гипотеза H_0 принимается на данном этапе проверки.
- (b) Иначе гипотеза H_0 отвергается, выбирается другая гипотеза, для нее выполняются те же действия.

Количество интервалов k можно определить с помощью эвристики:

$$k \approx 1.72 \cdot \sqrt[3]{n} \quad (3)$$

3 Особенности реализации

Программа для выполнения лабораторной была написана на языке Python 3.8.2. Для генерации выборки был использован модуль **stats** библиотеки `scipy`.

Можно показать, что для

$$f(x, \mu, \sigma) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

решениями системы (1) являются значения $\mu = \bar{x}$, $\sigma = \sqrt{x^2 - \bar{x}^2}$. Таким образом ищались оценки для параметров μ , σ методом максимального правдоподобия.

Остальные вычисления были проведены по формулам (2) и (3).

В приложении А приведена ссылка на репозиторий с исходным кодом.

4 Результаты работы программы

4.1 Метод максимального правдоподобия

С помощью метода максимального правдоподобия были получены следующие оценки:

$$\hat{\mu} \approx -0.01617 \quad \hat{\sigma} \approx 0.9251$$

4.2 Критерий хи-квадрат

Для выборки из 100 элементов был выбран $k = 8$ по формуле (3). Уровень значимости был выбран $\alpha = 0.05$.

В таблице 1 вычислен выборочный коэффициент χ_B^2 по формуле (2).

i	Δ_i	n_i	p_i	np_i	$n_i - np_i$	$\frac{(n_i - np_i)^2}{np_i}$
1	$(-\infty; -2.03)$	1	0.0125	1.25	-0.25	0.05
2	$(-2.03; -1.45)$	2	0.0489	4.89	-2.88	1.71
3	$(-1.45; -0.87)$	20	0.1377	13.77	6.23	2.82
4	$(-0.87; -0.29)$	22	0.2425	24.25	-2.25	0.21
5	$(-0.29; 0.29)$	28	0.2675	26.75	1.25	0.06
6	$(0.29; 0.87)$	15	0.185	18.5	-3.5	0.66
7	$(0.87; 1.45)$	7	0.0801	8.01	-1.00	0.13
8	$(1.45; +\infty)$	5	0.0258	2.58	2.42	2.27
Сумма	-	100	1.0000	100.00	0.00	7.91

Таблица 1: Вычисление χ_B^2 при проверке гипотезы о нормальном законе распределения

Табличное значение $\chi_{0.95}^2(7) = 14.0671$, выборочное значение $\chi_B^2 = 7.91$, значит гипотеза может быть принята на данном этапе.

4.3 Проверка чувствительности критерия хи-квадрат

Для выборок из 20 элементов был выбран $k = 5$. Уровень значимости был выбран $\alpha = 0.05$.

В таблице 2 вычислен выборочный коэффициент χ_B^2 для выборки, соответствующей распределению Лапласа.

i	Δ_i	n_i	p_i	np_i	$n_i - np_i$	$\frac{(n_i - np_i)^2}{np_i}$
1	$(-\infty; -2.56)$	5	0.1042	2.08	2.91	4.08
2	$(-2.56; -1.65)$	0	0.1739	3.47	-3.47	3.48
3	$(-1.65; -0.74)$	4	0.2542	5.08	-1.08	0.23
4	$(-0.74; 0.17)$	5	0.2413	4.82	0.17	0.01
5	$(0.17; +\infty)$	6	0.2265	4.53	1.46	0.48
Сумма	-	20	1.0000	20.00	0.00	8.28

Таблица 2: Вычисление χ_B^2 при проверке гипотезы о нормальном законе распределения для выборки $L(0, 1)$

Табличное значение $\chi_{0.95}^2(4) = 9.4877$, выборочное значение $\chi_B^2 = 8.28$, значит гипотеза также может быть принята на данном этапе.

В таблице 3 вычислен выборочный коэффициент χ_B^2 для выборки, соответствующей равномерному распределению.

i	Δ_i	n_i	p_i	np_i	$n_i - np_i$	$\frac{(n_i - np_i)^2}{np_i}$
1	$(-\infty; -1.04)$	4	0.1586	3.17	0.82	0.22
2	$(-1.04; -0.37)$	3	0.2078	4.15	-1.15	0.32
3	$(-0.37; 0.3)$	6	0.2582	5.16	0.83	0.14
4	$(0.3; 0.97)$	2	0.211	4.22	-2.21	1.17
5	$(0.96; +\infty)$	5	0.1668	3.33	1.66	0.83
Сумма	-	20	1.0000	20.00	0.00	2.68

Таблица 3: Вычисление χ_B^2 при проверке гипотезы о нормальном законе распределения для выборки $N(-\sqrt{3}, \sqrt{3})$

Табличное значение $\chi_{0.95}^2(4) = 9.4877$, выборочное значение $\chi_B^2 = 2.68$, значит гипотеза также может быть принята на данном этапе.

Заключение

В рамках лабораторной работы была сгенерирована выборка для размером 100 элементов для нормального распределения $N(0, 1)$. Для нее найти методом максимального правдоподобия оценки параметров μ и σ для функции распределения $N(\mu, \sigma)$. Такая гипотеза была проверена критерием χ^2 . Те же действия были сделаны для выборок размером 20 элементов, соответствующих распределению Лапласа и равномерному распределению.

Было установлено, что применение метода максимального правдоподобия для нормального распределения сводится к нахождению выборочного среднего и среднеквадратичного отклонения.

Критерий χ^2 не отверг гипотезу о соответствии нормальной выборки нормальному распределению, чего и следовало ожидать.

Критерий χ^2 также не отверг гипотезу о соответствии нормальному распределению малых выборок, которые были сгенерированы в соответствии с распределением Лапласа и равномерным распределением. Это объясняется тем, что критерий χ^2 имеет асимптотический характер, а значит может давать ошибочные результаты при малых размерах выборки. Стоит также отметить, что при реализации лабораторной работы сначала генерировались выборки размером 100 элементов, и для них примерно в половине случаев критерий χ^2 отвергал гипотезу о нормальности.

Программа для лабораторной была написана языке Python 3.8.2, для генерации выборок была использована библиотека `scipy`.

Список литературы

- [1] Теоретическое приложение к лабораторным работам №5-8 по дисциплине «Математическая статистика». — Спб.: Санкт-Петербургский политехнический университет, 2020. — 12 с.
- [2] Chi-square distribution. // Wikipedia, the free encyclopedia. — https://en.wikipedia.org/wiki/Chi-square_distribution . — (дата обращения: 05.12.20)

А Репозиторий с исходным кодом

Исходный код программы для данной лабораторной размещен на сервисе GitHub.

Ссылка на репозиторий: <https://github.com/Vovan-S/TV-Lab1>.