Санкт-Петербургский политехнический университет Петра Великого

Институт прикладной математики и механики

Кафедра «Телематика (при ЦНИИ РТК)»

Отчет по лабораторной работе

Метод максимального правдоподобия. Проверка гипотезы о распределении по критерию хи-квадрат

По дисциплине «Теория вероятностей и Математическая статистика»

Выполнил
Студент гр. 3630201/80101 ________ М. Д. Маляренко

Руководитель
к.ф.-м.н., доцент ________ А. Н. Баженов

«___»___________ 2020г.

Санкт-Петербург
2021

Содержание

Список таблиц

1 Постановка задачи

Сгенерировать выборку объёмом 100 элементов для нормального распределения $N(0,1)$. По сгенерированной выборке оценить параметры μ и σ нормального закона методом максимального правдоподобия. В качестве основной гипотезы H_0 считать, что сгенерированное распределение имеет вид $N(\hat{\mu},\hat{\sigma})$, где $\hat{\mu}$ и $\hat{\sigma}$ оценки метода максимального правдоподобия. Проверить основную гипотезу, используя критерий согласия χ^2. В качестве уровня значимости взять $\alpha = 0.05$.

Исследовать точность (чувствительность) критерия χ^2 – сгенерировать выборки равномерного распределения и распределения Лапласа из 20 элементов. Проверить их на нормальность, то есть проверить, принимает ли критерий χ^2 гипотезу, что элементы этих выборок распределены по закону $N(\hat{\mu},\hat{\sigma})$.

2 Теория

2.1 Метод максимального правдоподобия

$L(x_1, \ldots, x_n, \theta)$ — функция правдоподобия (ФП), рассматриваемая как функция неизвестного параметра θ[1][2]:

$$L(x_1, \ldots, x_n, \theta) = f(x_1, \theta) \ldots f(x_n, \theta). \tag{1}$$

Оценка максимального правдоподобия:

$$\hat{\theta}_{\text{МП}} = \arg\max L(x_1, \ldots, x_n, \theta) \tag{2}$$

Система уравнений правдоподобия (в случае дифференцируемости функции правдоподобия):

$$\frac{\partial L}{\partial \theta_k} = 0 \text{ или } \frac{\partial \ln L}{\partial \theta_k} = 0, \; k = 1, \ldots, m. \tag{3}$$

2.2 Проверка гипотезы о законе распределения. Метод χ^2

Выдвинута гипотеза H_0 о генеральном законе распределения с функцией распределения $F(x)$.

Рассматриваем случай, когда гипотетическая функция распределения $F(x)$ не содержит неизвестных параметров.

Правило проверки гипотезы о законе распределения по методу χ^2.[1]

1. Выбираем уровень значимости α.
2. Находим квантиль $\chi^2_{1-\alpha}(k-1)$ распределения χ^2 с $k-1$ степенями свободы порядка $1-\alpha$.
3. С помощью гипотетической функции распределения $F(x)$ вычисляем вероятности $p_i = P(X \in \Delta_i), i = 1, \ldots, k$.
4. Находим частоты n_i попадания элементов выборки в подмножества $\Delta_i, i = 1, \ldots, k$.
5. Вычисляем выборочное значение статистики критерия χ^2 :
$$\chi^2_B = \sum_{i=1}^{n} \frac{(n_i - np_i)^2}{np_i}.$$
6. Сравниваем χ^2_B и квантиль $\chi^2_{1-\alpha}(k-1)$.
   а) Если $\chi^2_B < \chi^2_{1-\alpha}(k-1)$, то гипотеза H_0 на данном этапе проверки принимается.
   б) Если $\chi^2_B \geq \chi^2_{1-\alpha}(k-1)$, то гипотеза H_0 отвергается, выбирается одно из альтернативных распределений, и процедура проверки повторяется.

Количество интервалов k можно определить с помощью эвристики:

$$k \approx 1.72 \cdot \sqrt[3]{n} \tag{4}$$

3 Реализация

Расчёты проводились в среде аналитических вычислений Maxima. Были написаны функции для нахождения количества промежутков разбиения с помощью эвристики 4, нахождения границ промежутков разбиения, вычисления теоретической вероятности на промежутках с помощью встроенной реализации функции нормального распределения, вычисления относительной частоты попадания элементов выборки в промежутки разбиения. Полный текст скрипта представлен в репозитории GitHub.

4 Результаты

4.1 Метод максимального правдоподобия

По методу максимального правдоподобия были получены следующие оценки параметров выборки, распределённой по закону $N(0,1)$:

$$\hat{\mu} \approx -0.035, \quad \hat{\sigma} \approx 0.992$$

4.2 Критерий согласия χ^2 для нормального распределения

- Количество промежутков $k = 8$ (4)
- Уровень значимости $\alpha = 0.05$
- Квантиль $\chi^2_{1-\alpha}(k-1) = \chi^2_{0.95}(7) = 14.0671$

В Таблице 1 представлены этапы вычисления критерия хи-квадрат для проверки гипотезы о законе распределения.

Таблица 1: Вычисление χ^2_B при проверке гипотезы H_0 о законе распределения $N(\hat{\mu}, \hat{\sigma})$ для выборки распределения $N(0,1)$

i	Δ_i	n_i	p_i	$n_i p_i$	$\frac{(n_i - np_i)^2}{np_i}$
1	$(-\infty, -1.49)$	8	0.070	7.02	0.01
2	$[-1.49, -1.05)$	9	0.084	8.35	0.04
3	$[-1.05, -0.59)$	7	0.132	13.16	2.91
4	$[-0.59, -0.15)$	20	0.169	16.94	0.57
5	$[-0.15, 0.3)$	18	0.178	17.82	0.00
6	$[0.3, 0.75)$	17	0.153	15.31	0.18
7	$[0.75, 1.2)$	13	0.108	10.75	0.45
8	$[1.20, \infty)$	8	0.106	10.62	0.64
$\sum$	–	100	1.000	100.0	$4.92 = \chi^2_B$

Табличное значение $\chi^2_{0.95}(7) = 14.0671$ больше, чем выборочное $\chi^2_B = 4.92$, это значит, что на данном этапе гипотезу H_0 можно принять.

4.3 Проверка чувствительности критерия согласия χ^2

4.3.1 Для равномерного распределения

- Количество промежутков $k = 5$
- Уровень значимости $\alpha = 0.05$
- Квантиль $\chi^2_{1-\alpha}(k-1) = \chi^2_{0.95}(4) = 9.487$

В Таблице 2 представлены этапы вычисления критерия хи-квадрат для проверки гипотезы о законе распределения.

Таблица 2: Вычисление χ^2_B при проверке гипотезы H_0 о законе распределения $N(\hat{\mu}, \hat{\sigma})$ для выборки распределения $U(-\sqrt{3}, \sqrt{3})$

i	Δ_i	n_i	p_i	$n_i p_i$	$\frac{(n_i - np_i)^2}{np_i}$
1	$(-\infty, -0.88)$	5	0.171	3.42	0.73
2	$[-0.88, -0.18)$	2	0.207	4.13	1.10
3	$[-0.18, 0.52)$	4	0.250	5.00	0.20
4	$[0.52, 1.22)$	4	0.204	4.09	0.00
5	$[1.22, \infty)$	5	0.167	3.34	0.82
$\sum$	–	20	1.000	20.0	$2.86 = \chi^2_B$

Табличное значение $\chi^2_{0.95}(7) = 9.487$ больше, чем выборочное $\chi^2_B = 2.86$, это значит, что на данном этапе гипотезу H_0 можно принять.

4.3.2 Для распределения Лапласа

- Количество промежутков $k = 5$
- Уровень значимости $\alpha = 0.05$
- Квантиль $\chi^2_{1-\alpha}(k-1) = \chi^2_{0.95}(4) = 9.487$

В Таблице 3 представлены этапы вычисления критерия хи-квадрат для проверки гипотезы о законе распределения.

Таблица 3: Вычисление χ^2_B при проверке гипотезы H_0 о законе распределения $N(\hat{\mu}, \hat{\sigma})$ для выборки распределения $L(0, 1/\sqrt{2})$

i	Δ_i	n_i	p_i	$n_i p_i$	$\frac{(n_i - np_i)^2}{np_i}$
1	$(-\infty, -0.83)$	5	0.184	3.68	0.47
2	$[-0.83, -0.31)$	3	0.182	2.98	0.00
3	$[-0.31, 0.20)$	5	0.128	3.64	0.50
4	$[0.20, 0.72)$	2	0.179	3.58	0.69
5	$[0.72, \infty)$	5	0.306	6.11	0.20
$\sum$	–	20	1.000	20.0	$1.88 = \chi^2_B$

Табличное значение $\chi^2_{0.95}(4) = 9.487$ больше, чем выборочное $\chi^2_B = 1.88$, это значит, что на данном этапе гипотезу H_0 можно принять.

Заключение

В результате выполнения лабораторной работы была изучена оценка параметров распределения по выборке распределённой согласно закону $N(0,1)$ с помощью метода максимального правдоподобия. Выяснено, что о.м.п. матожидания нормального распределения равна средневыборочному значению, а о.м.п. среднеквадратического отклонения вычисляется как стандартное выборочное отклонение.

Также были проанализированы результаты оценки гипотез о распределения выборочных элементов по критерию хи-квадрат. Для нормального распределения по закону $N(0,1)$ критерий хи-квадрат не отверг гипотезу о том, что выборка распределена по закону $N(\hat{\mu}, \hat{\sigma})$, где $\hat{\mu}$ - о.м.п. матожидания, $\hat{\sigma}$ – о.м.п. среднеквадратического отклонения.

Проверка чувствительности критерия хи-квадрат на малых выборках показала, что этот критерий может давать неверные результаты, то есть является не чувствительным на малых выборках, так как он не отверг гипотезу о нормальности распределения $U(-\sqrt{3}, \sqrt{3})$ и $L(0, 1/\sqrt{2})$

Список литературы

[1] Теоретическое приложение к лабораторным работам №5-8 по дисциплине «Математическая статистика». – СПб.: СПбПУ, 2020. – 22 с

[2] Максимов Ю.Д. Математика. Теория и практика по математической статистике. Конспект-справочник по теории вероятностей : учеб. пособие / Ю.Д. Максимов; под ред. В.И. Антонова. — СПб. : Изд-во Политехн. ун-та, 2009. — 395 с. (Математика в политехническом университете).

Приложение А. Репозиторий с исходным кодом

Исходный код скрипта для среды аналитических вычислений Maxima находится в репозитории GitHub – URL `https://github.com/malyarenko-md/TeorVer`