Санкт-Петербургский политехнический университет
Высшая школа прикладной математики и
вычислительной физики, ИПММ

Направление подготовки
01.03.02 «Прикладная математика и информатика»

Отчет по лабораторной работе №7
по дисциплине «Математическая статистика»

Выполнил студент гр. 3630102/80201
Кирпиченко С. Р.
Руководитель
Баженов А. Н.

Санкт-Петербург
2021

Содержание

Список таблиц

1 Постановка задачи

Сгенерировать выборку объёмом 100 элементов для нормального распределения $N(x, 0, 1)$. По сгенерированной выборке оценить параметры μ и σ нормального закона методом максимального правдоподобия. В качестве основной гипотезы H_0 будем считать, что сгенерированное распределение имеет вид $N(x, \hat{\mu}, \hat{\sigma})$. Проверить основную гипотезу, используя критерий согласия χ^2. В качестве уровня значимости взять $\alpha = 0.05$. Привести таблицу вычислений χ^2.
Исследовать точность (чувствительность) критерия χ^2 - сгенерировать выборки равномерного распределения и распределения Лапласа малого объема (20 элементов). Проверить их на нормальность.

2 Теория

2.1 Метод максимального правдоподобия

$L(x_1, ..., x_n, \theta)$ – функция правдоподобия(ФП), рассматриваемая как функция неизвестного параметра θ:

$$L(x_1, ..., x_n, \theta) = f(x_1, \theta)f(x_2, \theta)...f(x_n, \theta). \tag{1}$$

Оценка максимального правдоподобия:

$$\widehat{\theta}_{\text{мп}} = \arg\max_{\theta} L(x_1, ..., x_n, \theta). \tag{2}$$

Система уравнений правдоподобия (в случае дифференцируемости функции правдоподобия):

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

2.2 Проверка гипотезы о законе распределения генеральной совокупности. Метод хи-квадрат

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. Выбираем уровень значимости α.
2. По таблице [1, с. 358] находим квантиль $\chi^2_{1-\alpha}(k-1)$ распределения хи-квадрат с $k-1$ степенями свободы порядка $1-\alpha$.
3. Вычисляем вероятности $p_i = P(X \in \Delta_i), i = 1, ..., k$, с помощью гипотетической функции распределения $F(x)$.
4. Находим частоты n_i попадания элементов выборки в подмножества $\Delta_i, i = 1, ..., k$.
5. Вычисляем выборочное значение статистики критерия χ^2:

$$\chi^2_{\text{В}} = \sum_{i=1}^{k} \frac{(n_i - np_i)^2}{np_i}.$$

6. Сравниваем $\chi^2_{\text{В}}$ и квантиль $\chi^2_{1-\alpha}(k-1)$.
   (a) Если $\chi^2_{\text{В}} < \chi^2_{1-\alpha}(k-1)$, то гипотеза H_0 на данном этапе проверки принимается.
   (b) Если $\chi^2_{\text{В}} \geq \chi^2_{1-\alpha}(k-1)$, то гипотеза H_0 отвергается, выбирается одно из альтернативных распределений, и процедура проверки повторяется.

3 Реализация

Лабораторная работа выполнена на языке Python 3.9 с использованием библиотек numpy, scipy, matplotlib, seaborn.

4 Результаты

4.1 Проверка гипотезы о законе распределения генеральной совокупности. Метод хи-квадрат

Метод максимального правдоподобия:

$$\hat{\mu} \approx 0.08 \quad \hat{\sigma} \approx 1.08$$

Критерий согласия χ^2:

1. Количество промежутков $k = 1.72\sqrt[3]{n} = 8$
2. Уровень значимости $\alpha = 0.05$
3. $\chi^2_{1-\alpha}(k-1) = \chi^2_{0.95}(7) = 14.07$

i	Δ_i	n_i	p_i	np_i	$n_i - np_i$	$\frac{(n_i - np_i)^2}{np_i}$
1	$(-\infty,\ -1.1)$	11	0.138	13.8	-2.8	0.5674
2	$(-1.1,\ -0.73)$	9	0.0886	8.86	0.14	0.0022
3	$(-0.73,\ -0.37)$	14	0.114	11.4	2.6	0.5928
4	$(-0.37,\ 0.0)$	16	0.1309	13.09	2.91	0.6478
5	$(0.0,\ 0.37)$	9	0.1341	13.41	-4.41	1.4486
6	$(0.37,\ 0.73)$	11	0.1225	12.25	-1.25	0.1282
7	$(0.73,\ 1.1)$	15	0.0999	9.99	5.01	2.5094
8	$(1.1,\ \infty)$	15	0.172	17.2	-2.2	0.2815
Σ	–	100	1	100	0	6.18=χ^2_B

Таблица 1: Проверка гипотезы H_0 на нормальной выборке

$\chi^2_B < \chi^2_{0.95}(7) \Rightarrow$ на текущем этапе гипотеза H_0 о том, что генеральная выборка имеет распределение $N(x, \hat{\mu}, \hat{\sigma})$, принимается.

4.2 Исследование на чувствительность

Рассмотрим выборку, сгенерированную по распределению Лапласа.

$$\hat{\mu} \approx -0.09 \quad \hat{\sigma} \approx 0.89$$

Критерий согласия χ^2:

1. Количество промежутков $k = 1.72\sqrt[3]{n} = 5$
2. Уровень значимости $\alpha = 0.05$
3. $\chi^2_{1-\alpha}(k-1) = \chi^2_{0.95}(4) = 9.49$

i	Δ_i	n_i	p_i	np_i	$n_i - np_i$	$\frac{(n_i - np_i)^2}{np_i}$
1	$(\infty, -1.1)$	4	0.1292	2.58	1.42	0.7753
2	$(-1.1, -0.37)$	2	0.2495	4.99	-2.99	1.7916
3	$(-0.37, 0.37)$	8	0.3171	6.34	1.66	0.4336
4	$(0.37, 1.1)$	4	0.213	4.26	-0.26	0.0159
5	$(1.1, \infty)$	2	0.0912	1.82	0.18	0.0171
Σ	–	20	1.0	20.0	0	3.03= χ^2_B

Таблица 2: Проверка гипотезы H_0 на выборке, сгенерированной по распределению Лапласа

Рассмотрим выборку, сгенерированную по равномерному распределению.

$$\hat{\mu} \approx -0.09 \quad \hat{\sigma} \approx 0.88$$

i	Δ_i	n_i	p_i	np_i	$n_i - np_i$	$\frac{(n_i - np_i)^2}{np_i}$
1	$(\infty, -1.1)$	3	0.127	2.54	0.46	0.0836
2	$(-1.1, -0.37)$	6	0.2516	5.03	0.97	0.1863
3	$(-0.37, 0.37)$	4	0.3207	6.41	-2.41	0.9087
4	$(0.37, 1.1)$	5	0.2128	4.26	0.74	0.1297
5	$(1.1, \infty)$	2	0.0879	1.76	0.24	0.0333
Σ	–	20	1.0	20.0	0	1.34= χ^2_B

Таблица 3: Проверка гипотезы H_0 на выборке, сгенерированной по равномерному распределению

Видим, что в обоих случаях гипотеза принимается.

5 Обсуждение

Оценки максимального правдоподобия показали свою состоятельность на выборке из 100 элементов, распределенной нормально. Погрешность найденных параметров наблюдается во втором знаке после запятой.
На малых выборках метод хи-квадрат при проверке на нормальность не различает выборки, распределенные равномерно и по закону Лапласа, подтверждая гипотезу в обоих случаях. Это обусловлено теоремой К. Пирсона: статистика критерия χ^2 асимптотически при $n \to \infty$ распределена по закону χ^2 с $k-1$ степенями свободы. То есть на малых выборках теория ничего не гарантирует.

Примечание

С исходным кодом работы и данного отчета можно ознакомиться в репозитории `https://github.com/Stasychbr/MatStat`

Список литературы

[1] 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).