

Санкт-Петербургский политехнический университет
Петра Великого

Институт прикладной математики и механики
Кафедра «Прикладная математика»

ОТЧЕТ ПО ЛАБОРАТОРНОЙ РАБОТЕ №7

по дисциплине
"Математическая статистика"

Выполнил студент
Группы 3630102/80101

Шао Цзяци

Проверил
доцент, к.ф.-м.н.

Баженов Александр Николаевич

Санкт-Петербург
2021 г.

Содержание

1. Постановка задачи	4
2. Теория	4
2.1. Метод максимального правдоподобия	4
2.2. Проверка гипотезы о законе распределения генеральной совокупности. Метод хи-квадрат	4
3. Реализация	5
4. Результаты	6
4.1. Нормальное распределение	6
4.2. Распределение Лапласа	6
4.3. Равномерное распределение	7
5. Обсуждение	7

Список таблиц

1	Вычисление χ_B^2 при нормальном распределении	6
2	Вычисление χ_B^2 при распределении Лапласа	7
3	Вычисление χ_B^2 при равномерном распределении	7

1. Постановка задачи

Дано нормальное распределение $N(x, 0, 1)$

- 1) Сгенерировать выборку объемом 100 элементов для нормального распределения $N(x, 0, 1)$.
- 2) 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 $N(x, \hat{\mu}, \hat{\sigma})$
- 3) Проверить основную гипотезу, используя критерий согласия χ^2 . В качестве уровня значимости взять $\alpha = 0.05$.
- 4) Привести таблицу вычислений χ^2 .
- 5) Исследовать точность (чувствительность) критерия χ^2 - сгенерировать выборки равномерного распределения и распределения Лапласа малого объема (например, 20 элементов). Проверить их на нормальность.

2. Теория

2.1. Метод максимального правдоподобия

Метод максимального правдоподобия, англ - *maximum likelihood estimation* - это метод оценивания неизвестного параметра путём максимизации функции правдоподобия.

Пусть есть выборка X_1, \dots, X_n из распределения с плотностью $f(x, \theta)$, θ - неизвестный параметр. Пусть $L(x|\theta)$ - функция правдоподобия. Оценка

$$\theta_{MLE} = \hat{\theta}_{MLE}(X_1, \dots, X_n) = \arg \max_{\theta} L(x_1, \dots, x_n, \theta) \quad (1)$$

называется оценкой максимального правдоподобия параметра θ . Таким образом оценка максимального правдоподобия — это такая оценка, которая максимизирует функцию правдоподобия при фиксированной реализации выборки.

Часто вместо функции правдоподобия L используют логарифмическую функцию правдоподобия $l = \ln L$.

$$\frac{\partial \ln L}{\partial \theta} = \frac{1}{L} \frac{\partial L}{\partial \theta}, L > 0 \quad (2)$$

и решать уравнение правдоподобия

$$\frac{\partial \ln L}{\partial \theta} = 0 \quad (3)$$

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

Пусть выдвинута гипотеза H_0 о виде закона распределения $F(x)$. Необходимо оценить его параметры и проверить закон в целом.

Для проверки гипотезы о законе распределения чаще всего применяется критерий согласия χ^2 . Пусть гипотетическая функция распределения $F(x)$ не содержит неизвестных параметров.

Разобьём генеральную совокупность, т.е. множество значений изучаемой случайной величины X на k непересекающихся равных подмножеств $\Delta_1, \Delta_2, \dots, \Delta_k$, где k выбирается согласованным с n и берется аналогичному при построении гистограмм $k \approx 1.72\sqrt[3]{n}$ или по формуле Старджесса $k \approx 1 + 3.3 \lg n$.

Пусть $p_i = P(X \in \Delta_i), i = 1, 2, \dots, k$. Если генеральная совокупность - вся вещественная ось, $p_i = F(a_i) - F(a_{i-1}), i = 1, 2, \dots, k$. При этом $\sum_{i=1}^k p_i = 1$ и $p_i > 0, i = 1, 2, \dots, k$.

Пусть, далее, n_1, n_2, \dots, n_k — частоты попадания выборочных элементов в подмножества $\Delta_1, \Delta_2, \dots, \Delta_k$ соответственно.

Если гипотеза H_0 справедлива, то относительные частоты $\frac{n_i}{n} \rightarrow p_i, i = \overline{1, k}$. Следовательно, мера отклонения выборочного распределения от гипотетического с использованием коэффициентов Пирсона:

$$\chi^2 = \sum_{i=1}^k \frac{n_i - np_i}{np_i}^2 \quad (4)$$

Теорема К.Пирсона: статистика критерия χ^2 асимптотически распределена по закону χ^2 с $k - 1$ степенями свободы.

Правило проверки гипотезы о законе распределения по методу χ^2

- 1) Выбираем уровень значимости α
- 2) Находим квантиль $\chi_{1-\alpha}^2(k - 1)$ распределения хи-квадрат с $k - 1$ степенями свободы порядка $1 - \alpha$
- 3) С помощью гипотетической функции распределения $F(x)$ вычисляем вероятности $p_i = P(X \in \Delta_i), i = 1, 2, \dots, k$
- 4) Находим частоты n_i попадания элементов выборки в подмножества $\Delta_i, i = 1, 2, \dots, k$
- 5) Вычисляем выборочное значение статистики критерия χ^2
- 6) Сравниваем χ_B^2 и квантиль $\chi_{1-\alpha}^2(k - 1)$:
 - если $\chi_B^2 < \chi_{1-\alpha}^2(k - 1)$, то гипотеза H_0 на данном этапе проверки принимается
 - иначе гипотеза H_0 отвергается, выбирается одно из альтернативных распределений, и процедура проверки повторяется

Замечание: при ситуации $\chi_B^2 \approx \chi_{1-\alpha}^2(k - 1)$ стоит увеличить объем выборки (например, в 2 раза), чтобы требуемое неравенство было более четким.

Замечание: Изучено, что если для каких-либо подмножеств $\Delta_i, i = \overline{1, k}$ условие $np_i \geq 5$ не выполняется, то следует объединить соседние подмножества (промежутки). Это условие выдвигается требованием близости величин $\frac{(n_i - np_i)}{\sqrt{np_i}}$. Тогда случайная величина будет распределена по закону, близкому к хи-квадрат. Такая близость обеспечивается достаточной численностью элементов в подмножествах Δ_i .

3. Реализация

Лабораторная работа выполнена с помощью встроенных средств языка программирования python в среде разработки Rucharm с дополнительными библиотеками.

- scipy
- numpy
- matplotlib
- math

Исходный код лабораторной работы размещен в Github-репозитории.
URL: <https://github.com/Shaots/shaoMathStatistic/tree/master/Lab7>

4. Результаты

4.1. Нормальное распределение

$$\begin{cases} \mu = -0.0574 \\ \sigma = 0.9653 \\ \text{Количество промежутков } k = 8 \\ \text{Уровень значимости } \alpha = 0.05 \end{cases} \quad (5)$$

Квантиль $\chi^2_{1-\alpha}(k-1) = \chi^2_{0.95} = 14.0671$.

i	$\Delta_i = [a_{i-1}, a_i)$	n_i	p_i	np_i	$n_i - np_i$	$\frac{(n_i - np_i)^2}{np_i}$
1	$[-\infty, -1.1]$	13	0.1357	13.5666	-0.5666	0.0237
2	$[-1.1, -0.7333]$	10	0.096	9.6012	0.3988	0.0166
3	$[-0.7333, -0.3667]$	13	0.1253	12.5256	0.4744	0.018
4	$[-0.3667, 0.0]$	16	0.1431	14.3066	1.6934	0.2004
5	$[0.0, 0.3667]$	12	0.1431	14.3066	-2.3066	0.3719
6	$[0.3667, 0.7333]$	15	0.1253	12.5256	2.4744	0.4888
7	$[0.7333, 1.1]$	11	0.096	9.6012	1.3988	0.2038
8	$[1.1, \infty]$	10	0.1357	13.5666	-3.5666	0.9376
sum	—	100	1	100	0	2.2608

Таблица 1. Вычисление χ^2_B при нормальном распределении

4.2. Распределение Лапласа

$$\begin{cases} \mu = 0.1597 \\ \sigma = 1.2274 \\ \text{Количество промежутков } k = 5 \\ \text{Уровень значимости } \alpha = 0.05 \end{cases} \quad (6)$$

Квантиль $\chi^2_{1-\alpha}(k-1) = \chi^2_{0.95} = 9.48773$.

i	$\Delta_i = [a_{i-1}, a_i)$	n_i	p_i	np_i	$n_i - np_i$	$\frac{(n_i - np_i)^2}{np_i}$
1	$[-\infty, -1.1]$	1	0.1357	2.7133	-1.7133	1.0819
2	$[-1.1, -0.3667]$	2	0.2213	4.4254	-2.4254	1.3292
3	$[-0.3667, 0.3667]$	11	0.2861	5.7226	5.2774	4.8667
4	$[0.3667, 1.1]$	4	0.2213	4.4254	-0.4254	0.0409
5	$[1.1, \infty]$	2	0.1357	2.7133	-0.7133	0.1875
sum	—	20	1	20	0	7.5062

Таблица 2. Вычисление χ_B^2 при распределении Лапласа

4.3. Равномерное распределение

$$\begin{cases} \mu = -0.161 \\ \sigma = 0.8001 \\ \text{Количество промежутков } k = 5 \\ \text{Уровень значимости } \alpha = 0.05 \end{cases} \quad (7)$$

Квантиль $\chi_{1-\alpha}^2(k-1) = \chi_{0.95}^2 = 9.48773$.

i	$\Delta_i = [a_{i-1}, a_i)$	n_i	p_i	np_i	$n_i - np_i$	$\frac{(n_i - np_i)^2}{np_i}$
1	$[-\infty, -1.1]$	1	0.1357	2.7133	-1.7133	1.0819
2	$[-1.1, -0.3667]$	10	0.2213	4.4254	5.2774	7.0224
3	$[-0.3667, 0.3667]$	3	0.2861	5.7226	-2.7226	1.2953
4	$[0.3667, 1.1]$	5	0.2213	4.4254	0.5746	0.0746
5	$[1.1, \infty]$	1	0.1357	2.7133	-1.7133	1.0819
sum	—	20	1	20	0	10.5561

Таблица 3. Вычисление χ_B^2 при равномерном распределении

5. Обсуждение

Заключаем, что гипотеза H_0 о нормальном законе распределения $N(x, \hat{\mu}, \hat{\sigma})$ на уровне значимости $\alpha = 0.05$ согласуется с выборкой для нормального распределения $N(x, 0, 1)$.

Также видно, что для выборок сгенерированных по равномерному закону и закону Лапласа гипотеза H_0 оказалась принята.