

Санкт-Петербургский политехнический университет
Петра Великого

Институт прикладной математики и механики
Кафедра «Прикладная математика»

ОТЧЕТ ПО ЛАБОРАТОРНОЙ РАБОТЕ №7

по дисциплине
"Математическая статистика"

Выполнила студентка
группы 3630102/80201

Деркаченко Анна Олеговна

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. Выборка нормального распределения	6
4.2. Выборка распределения Лапласа	6
4.3. Выборка равномерного распределения	7
5. Обсуждение	7

Список иллюстраций

1. Постановка задачи

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

Необходимо:

- 1) Сгенерировать выборку объемом 100 элементов
- 2) Оценить параметры μ и σ методом максимального правдоподобия, используя основную гипотезу H_0 , что сгенерированное распределение имеет вид $N(x, \hat{\mu}, \hat{\sigma})$
- 3) Проверить основную гипотезу, используя критерий согласия χ^2 и уровень значимости $\alpha = 0.05$
- 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) и проверить их на нормальность

2. Теория

2.1. Метод максимального правдоподобия

Пусть x_1, \dots, x_n — случайная выборка из генеральной совокупности с плотностью вероятности $f(x, \theta)$, $L(x_1, \dots, x_n, \theta)$ — функция правдоподобия, представляющая собой совместную плотность вероятности независимых с.в. x_1, \dots, x_n и рассматриваемая как функция неизвестного параметра θ :

$$L(x_1, \dots, x_n, \theta) = f(x_1, \theta)f(x_2, \theta)\dots f(x_n, \theta) \quad (1)$$

Оценка максимального правдоподобия - значение $\hat{\theta}$ из множества допустимых значений параметра θ , для которого функция правдоподобия принимает наибольшее значение при заданных x_1, \dots, x_n :

$$\hat{\theta} = \arg \max_{\theta} L(x_1, \dots, x_n, \theta) \quad (2)$$

Часто проще искать максимум логарифма функции правдоподобия, так как он имеет максимум в одной точке с ней:

$$\frac{\partial \ln L}{\partial \theta} = \frac{1}{L} \frac{\partial L}{\partial \theta}, L > 0 \quad (3)$$

и решать *уравнение правдоподобия*

$$\frac{\partial \ln L}{\partial \theta} = 0 \quad (4)$$

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

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:

$$\chi^2 = \sum_{i=1}^k \frac{n_i - np_i^2}{np_i} \quad (5)$$

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 $p_i = P(X \in \Delta_i), i = \overline{1, k}$
- 4) Находим частоты n_i попадания элементов выборки в подмножества $\Delta_i, i = \overline{1, k}$
- 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

Реализация лабораторной работы проводилась на языке Python в среде разработки PyCharm с использованием дополнительных библиотек:

- scipy
- numpy
- matplotlib
- math

Исходный код лабораторной работы размещен в GitHub-репозитории.

URL: <https://github.com/derkanw/Mathstat/tree/main/lab7>

4. Результаты

4.1. Выборка нормального распределения

$$\left\{ \begin{array}{l} \hat{\mu} = 0.0938 \\ \hat{\sigma} = 0.962 \\ \text{Количество промежутков } k = 8 \\ \text{Уровень значимости } \alpha = 0.05 \end{array} \right. \quad (6)$$

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

i	Границы Δ_i	n_i	p_i	np_i	$n_i - np_i$	$\frac{(n_i - np_i)^2}{np_i}$
1	$[-\infty, -1.1]$	10	0.1357	13.5666	-3.5666	0.9376
2	$[-1.1, -0.7333]$	7	0.096	9.6012	-2.6012	0.7047
3	$[-0.7333, -0.3667]$	14	0.1253	12.5256	1.4744	0.1735
4	$[-0.3667, 0.0]$	20	0.1431	14.3066	5.6934	2.2657
5	$[0.0, 0.3667]$	12	0.1431	14.3066	-2.3066	0.3719
6	$[0.3667, 0.7333]$	14	0.1253	12.5256	1.4744	0.1735
7	$[0.7333, 1.1]$	9	0.096	9.6012	-0.6012	0.0376
8	$[1.1, \infty]$	14	0.1357	13.5666	0.4334	0.0138
Σ	-	100	1	100	0	4.6785

Таблица 1. Вычисление χ^2_B при нормальном распределении

4.2. Выборка распределения Лапласа

$$\left\{ \begin{array}{l} \hat{\mu} = -0.1291 \\ \hat{\sigma} = 0.8581 \\ \text{Количество промежутков } k = 5 \\ \text{Уровень значимости } \alpha = 0.05 \end{array} \right. \quad (7)$$

Тогда квантиль $\chi^2_{1-\alpha}(k-1) = \chi^2_{0.95} \approx 9.4877$.

i	Границы Δ_i	n_i	p_i	np_i	$n_i - np_i$	$\frac{(n_i - np_i)^2}{np_i}$
1	$[-\infty, -1.1]$	2	0.1357	2.7133	-0.7133	0.1875
2	$[-1.1, -0.3667]$	5	0.2213	4.4254	0.5746	0.0746
3	$[-0.3667, 0.3667]$	9	0.2861	5.7226	3.2774	1.8769
4	$[0.3667, 1.1]$	1	0.2213	4.4254	-3.4254	2.6513
5	$[1.1, \infty]$	3	0.1357	2.7133	0.2867	0.0303
\sum	-	20	1	20	0	4.8207

Таблица 2. Вычисление χ_B^2 при нормальном распределении

4.3. Выборка равномерного распределения

$$\left\{ \begin{array}{l} \hat{\mu} = -0.2113 \\ \hat{\sigma} = 0.885 \\ \text{Количество промежутков } k = 5 \\ \text{Уровень значимости } \alpha = 0.05 \end{array} \right. \quad (8)$$

Тогда квантиль $\chi_{1-\alpha}^2(k-1) = \chi_{0.95}^2 \approx 9.4877$.

i	Границы 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	0.1357	2.7133	0.2867	0.0303
2	$[-1.1, -0.3667]$	7	0.2213	4.4254	2.5746	1.4979
3	$[-0.3667, 0.3667]$	4	0.2861	5.7226	-1.7226	0.5186
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	2.2752

Таблица 3. Вычисление χ_B^2 при нормальном распределении

5. Обсуждение

Для всех случаев справедливо неравенство $\chi_B^2 < \chi_{0.95}^2$. Для случая выборки, распределенной по закону $N(x, \hat{\mu}, \hat{\sigma})$, можно сказать, что гипотеза H_0 о нормальном распределении на уровне значимости $\alpha = 0.05$ согласуется с ней с определенной точностью. Для случаев распределения Лапласа и равномерного распределения эта оценка немного хуже приближается к параметрам нормального распределения, что объясняется малым количеством выборки. В итоге для всех случаев гипотеза H_0 принята.