

Санкт-Петербургский политехнический университет
Петра Великого

Институт прикладной математики и механики
Кафедра «Прикладная математика»

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

Выполнил студент:

Кондратьев Д. А.

группа: 3630102/70301

Проверил:

к.ф.-м.н., доцент

Баженов Александр Николаевич

Санкт-Петербург
2020 г.

Содержание

1. Постановка задачи	2
2. Теория	2
2.1. Метод максимального правдоподобия	2
2.2. Проверка гипотезы о законе распределения генеральной совокупности. Метод хи-квадрат	2
3. Реализация	3
4. Результаты	3
4.1. Проверка гипотезы о законе распределения генеральной совокупности. Метод хи-квадрат	3
4.1.1. Стандартное нормальное распределение	3
4.1.2. Равномерное распределение	4
5. Обсуждение	5
5.0.1. Стандартное нормальное распределение	5
5.0.2. Равномерное распределение	5
6. Литература	5
7. Приложение	5

Список таблиц

1	Вычисление χ_B^2 при проверке гипотезы H_0 о нормальном законе распределения $N(x, \hat{\mu}, \hat{\sigma})$. $N(x, 0, 1)$	4
2	Вычисление χ_B^2 при проверке гипотезы H_0 о нормальном законе распределения $N(x, \hat{\mu}, \hat{\sigma})$. $U(x, -2, 2)$	4

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 .

Также проверить данную гипотезу на равномерном распределении $U(x, -2, 2)$ при размере выборки равной 20 элементам.

2. Теория

2.1. Метод максимального правдоподобия

Метод максимального правдоподобия — метод оценивания неизвестного параметра путём максимизации функции правдоподобия.

$$\hat{\theta}_{\text{МП}} = \arg \max_{\theta} L(x_1, x_2, \dots, x_n, \theta) \quad (1)$$

где L — функция правдоподобия, которая представляет собой совместную плотность вероятности независимых случайных величин x_1, x_2, \dots, x_n и является функцией неизвестного параметра θ .

$$L = f(x_1, \theta) \cdot f(x_2, \theta) \cdot \dots \cdot f(x_n, \theta) \quad (2)$$

Оценкой максимального правдоподобия будем называть такое значение $\hat{\theta}_{\text{МП}}$ из множества допустимых значений параметра θ , для которого функция правдоподобия принимает при заданных x_1, x_2, \dots, x_n максимальное значение.

Тогда при оценивании математического ожидания m и дисперсии σ^2 нормального распределения $N(m, \sigma)$ получим:

$$\ln(L) = -\frac{n}{2} \ln(2\pi) - \frac{n}{2} \ln(\sigma^2) - \frac{1}{2\sigma^2} \sum_{i=1}^n (x_i - m)^2 \quad (3)$$

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

Разобьём генеральную совокупность на k непересекающихся подмножеств $\Delta_1, \Delta_2, \dots, \Delta_k$, $\Delta_i = (a_i, a_{i+1}]$, $p_i = P(X \in \Delta_i)$, $i = 1, 2, \dots, k$ — вероятность того, что точка попала в i -ый промежуток.

Так как генеральная совокупность — это \mathbb{R} , то крайние промежутки будут бесконечными: $\Delta_1 = (-\infty, a_1]$, $\Delta_k = (a_k, \infty)$, $p_i = F(a_i) - F(a_{i-1})$.

n_i — частота попадания выборочных элементов в Δ_i , $i = 1, 2, \dots, k$.

В случае справедливости гипотезы H_0 относительно частоты $\frac{n_i}{n}$ при больших n должны быть близки к p_i , значит в качестве меры имеет смысл взять:

$$Z = \sum_{i=1}^k \frac{n}{p_i} \left(\frac{n_i}{n} - p_i \right)^2 \quad (4)$$

Тогда:

$$\chi_B^2 = \sum_{i=1}^k \frac{n}{p_i} \left(\frac{n_i}{n} - p_i \right)^2 = \sum_{i=1}^k \frac{(n_i - np_i)^2}{np_i} \quad (5)$$

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

3. Реализация

Лабораторная работа выполнена на программном языке *Python 3.8* в среде разработки *Jupyter Notebook 6.0.3*. В работе использовались следующие пакеты языка *Python*:

- *numpy* — для генерации выборки и работы с массивами;
- *tabulate* — для построения таблиц;
- *scipy.stats* — содержит необходимые распределения и критерий χ^2 .

Ссылка на исходный код лабораторной работы приведена в приложении.

4. Результаты

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

4.1.1. Стандартное нормальное распределение

Метод максимального правдоподобия:

$$\hat{\mu} \approx 0.01, \quad \hat{\sigma} \approx 1.02.$$

Критерий согласия χ^2 :
 Количество промежутков $k = 6$.
 Уровень значимости $\alpha = 0.05$.

i	Δ_i	n_i	p_i	np_i	$n_i - np_i$	$\frac{(n_i - np_i)^2}{np_i}$
1	$(-\infty, -1.01]$	16	0.1562	15.62	0.38	0.01
2	$(-1.01, -0.37]$	17	0.2004	20.04	-3.04	0.46
3	$(-0.37, 0.28]$	31	0.2517	25.17	5.83	1.35
4	$(0.28, 0.92]$	18	0.2122	21.22	-3.22	0.49
5	$(0.92, 1.56]$	10	0.1201	12.01	-2.01	0.34
6	$(1.56, \infty]$	8	0.0594	5.94	2.06	0.72
Σ	—	100	1.0000	100	0.00	$3.36 = \chi_B^2$

Таблица 1. Вычисление χ_B^2 при проверке гипотезы H_0 о нормальном законе распределения $N(x, \hat{\mu}, \hat{\sigma})$. $N(x, 0, 1)$

4.1.2. Равномерное распределение

Метод максимального правдоподобия:

$$\hat{\mu} \approx -0.07, \quad \hat{\sigma} \approx 1.2.$$

Критерий согласия χ^2 :
 Количество промежутков $k = 5$.
 Уровень значимости $\alpha = 0.05$.

i	Δ_i	n_i	p_i	np_i	$n_i - np_i$	$\frac{(n_i - np_i)^2}{np_i}$
1	$(-\infty, -1.01]$	5	0.1562	3.12	1.88	1.13
2	$(-1.01, -0.37]$	5	0.2828	5.66	-0.66	0.08
3	$(-0.37, 0.28]$	3	0.3200	6.40	-3.40	1.81
4	$(0.28, 1.56]$	4	0.1815	3.63	0.37	0.04
5	$(1.56, \infty]$	3	0.0594	1.19	1.81	2.77
Σ	—	20	1.0000	20	0.00	$5.81 = \chi_B^2$

Таблица 2. Вычисление χ_B^2 при проверке гипотезы H_0 о нормальном законе распределения $N(x, \hat{\mu}, \hat{\sigma})$. $U(x, -2, 2)$

5. Обсуждение

5.0.1. Стандартное нормальное распределение

Табличное значение квантиля: $\chi_{0.95}^2(5) = 11.07$.

Так как $\chi_B^2 < \chi_{0.95}^2(5)$, то можно заключить, что гипотеза H_0 о нормальном законе распределения $N(x, \hat{\mu}, \hat{\sigma})$ на уровне значимости $\alpha = 0.05$, согласуется с выборкой.

5.0.2. Равномерное распределение

Табличное значение квантиля: $\chi_{0.95}^2(4) = 9.49$.

Так как $\chi_B^2 < \chi_{0.95}^2(4)$, то можно заключить, что гипотезу H_0 о нормальном законе распределения $N(x, \hat{\mu}, \hat{\sigma})$ на уровне значимости $\alpha = 0.05$, нельзя опровергнуть. Такое несоответствие действительности можно объяснить довольно малым размером выборки.

6. Литература

- 1) Maximum likelihood estimation. URL: https://en.wikipedia.org/wiki/Maximum_likelihood_estimation
- 2) Chi-squared test. URL: https://en.wikipedia.org/wiki/Chi-squared_test
- 3) Таблица значений χ^2 . URL: <http://statsoft.ru/home/textbook/modules/sttable.html#chi>

7. Приложение

- 1) Код лабораторной. URL: https://github.com/DmitriiKondratev/MatStat/blob/master/Lab_7/Lab_7.ipynb
- 2) Код отчёта. URL: https://github.com/DmitriiKondratev/MatStat/blob/master/Lab_7/Lab_report_7.tex