

Санкт-Петербургский политехнический университет Петра Великого

Институт прикладной математики и механики

Кафедра «Телематика (при ЦНИИ РТК)»

Отчет по лабораторным работам № 7, 8

По дисциплине «Теория вероятностей и Математическая статистика»

Выполнил

Студент гр. 3630201/80101

Печеный Н. А.

Руководитель

к.ф.-м.н., доцент

Баженов А. Н.

«___» _____ 2021г.

Санкт-Петербург
2021

Содержание

1	Постановка задачи	4
2	Теория	5
2.1	Метод максимального правдоподобия	5
2.2	Проверка гипотезы о законе распределения генеральной совокупности. Метод хи-квадрат	5
2.3	Доверительные интервалы	6
2.3.1	Оценка на основе статистики Стьюдента и хи-квадрат	6
2.3.2	Асимптотические оценки на основе центральной предельной теоремы	6
3	Реализация	7
4	Результаты	8
4.1	Проверка гипотезы о законе распределения генеральной совокупности. Метод хи-квадрат	8
4.2	Доверительные интервалы для параметров нормального распределения . . .	9
4.3	Доверительные интервалы для параметров произвольного распределения. Асимптотический подход	10
	Заключение	11
	Список Литературы	12
	Приложение А. Репозиторий с исходным кодом	13

Список таблиц

1	Вычисление χ_B^2 при проверке гипотезы H_0 о нормальном законе распределения $N(x, \hat{\mu}, \hat{\sigma})$	8
2	Вычисление χ_B^2 при проверке гипотезы о нормальном законе распределения для выборки $L(x, 0, 1)$	9
3	Вычисление χ_B^2 при проверке гипотезы о нормальном законе распределения для выборки $U(x, -\sqrt{3}, \sqrt{3})$	9
4	Интервальные оценки на основе статистик Стьюдента и хи-квадрат	10
5	Асимптотические интервальные оценки	10

1 Постановка задачи

1. Сгенерировать выборку объемом 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 (например, 20 элементов). Проверить их на нормальность.
2. Для двух выборок размерами 20 и 100 элементов, сгенерированных согласно нормальному закону $N(x, 0, 1)$, для параметров положения и масштаба построить асимптотически нормальные интервальные оценки на основе точечных оценок метода максимального правдоподобия и классические интервальные оценки на основе статистик χ^2 и Стьюдента. В качестве параметра надёжности взять $\gamma = 0.95$

2 Теория

2.1 Метод максимального правдоподобия

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

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

Оценка максимального правдоподобия:

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

Система уравнений правдоподобия (в случае дифференцируемости функции правдоподобия):

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

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 [2, с. 358] 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 $p_i = P(X \in \Delta_i), i = 1, \dots, k$.
4. Находим частоты n_i попадания элементов выборки в подмножества $\Delta_i, i = 1, \dots, k$.
5. Вычисляем выборочное значение статистики критерия χ^2 :

$$\chi_B^2 = \sum_{i=1}^n \frac{(n_i - np_i)^2}{np_i}.$$

6. Сравниваем χ_B^2 и квантиль $\chi_{1-\alpha}^2(k-1)$.
 - а) Если $\chi_B^2 < \chi_{1-\alpha}^2(k-1)$, то гипотеза H_0 на данном этапе проверки принимается.
 - б) Если $\chi_B^2 \geq \chi_{1-\alpha}^2(k-1)$, то гипотеза H_0 отвергается, выбирается одно из альтернативных распределений, и процедура проверки повторяется.

Количество интервалов k можно определить с помощью эвристики:

$$k \approx 1.72 \cdot \sqrt[3]{n} \quad (4)$$

2.3 Доверительные интервалы

Дана выборка размером $n(x_1, \dots, x_n)$ из генеральной совокупности. Для нее построим выборочное среднее \bar{x} и среднеквадратическое отклонение s .

Параметры расположения μ и масштаба σ неизвестны. Построим для них доверительный интервал с доверительной вероятностью γ .

2.3.1 Оценка на основе статистики Стьюдента и хи-квадрат

Оценка для параметра положения:

$$P\left(\bar{x} - \frac{s \cdot t_{1-\alpha/2}(n-1)}{\sqrt{n-1}} < \mu < \bar{x} + \frac{s \cdot t_{1-\alpha/2}(n-1)}{\sqrt{n-1}}\right) = \gamma, \quad (5)$$

где $1 - \alpha = \gamma$, $t_{1-\alpha/2}(n-1)$ — квантиль распределения Стьюдента с $(n-1)$ степенями свободы порядка $1 - \alpha/2$.

Оценка для параметра масштаба:

$$P\left(\frac{s\sqrt{n}}{\sqrt{\chi_{1-\alpha/2}^2(n-1)}} < \sigma < \frac{s\sqrt{n}}{\sqrt{\chi_{\alpha/2}^2(n-1)}}\right) = \gamma, \quad (6)$$

где $1 - \alpha = \gamma$, $\chi_p^2(n-1)$ — квантиль распределения хи-квадрат с $(n-1)$ степенями свободы порядка p .

Эти оценки справедливы для выборки из нормальной генеральной совокупности.

2.3.2 Асимптотические оценки на основе центральной предельной теоремы

Оценка для параметра положения:

$$P\left(\bar{x} - \frac{s \cdot u_{1-\alpha/2}}{\sqrt{n}} < \mu < \bar{x} + \frac{s \cdot u_{1-\alpha/2}}{\sqrt{n}}\right) \approx \gamma, \quad (7)$$

где $1 - \alpha = \gamma$, $u_{1-\alpha/2}$ — квантиль стандартного нормального распределения порядка $1 - \alpha/2$.

Для оценки параметра масштаба необходимо рассчитать выборочный эксцесс $e = \frac{m_4}{s^4} - 3$, где $m_4 = \frac{1}{n} \sum (x_i - \bar{x})^4$ — четвёртый выборочный центральный момент.

Парметр масштаба можно оценить так:

$$P(s(1+U)^{-0.5} < \sigma < s(1-U)^{-0.5}) \approx \gamma, \quad (8)$$

где $U = u_{1-\alpha/2} \sqrt{(e+2)/n}$, $u_{1-\alpha/2}$ — квантиль стандартного нормального распределения порядка $1 - \alpha/2$.

Эти оценки справедливы для выборки из генеральной совокупности, которая имеет конечные центральные моменты вплоть до 4 порядка и конечное матожидание.

3 Реализация

Расчёты проводились в среде аналитических вычислений *Mathematica*. Для генерации выборок и создания и отрисовки графиков были использованы библиотечные функции среды разработки. Код скрипта представлен в репозитории на GitHub, ссылка на репозиторий находится в **Приложении А**.

Можно показать, что для

$$f(x, \mu, \sigma) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

решениями системы (3) являются значения $\mu = \bar{x}$, $\sigma = \sqrt{\overline{x^2} - \bar{x}^2}$ [3]. Именно таким образом ищались оценки для параметров μ и σ методом максимального правдоподобия.

4 Результаты

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

Метод максимального правдоподобия:

$$\hat{\mu} \approx 0.03, \quad \hat{\sigma} \approx 1.11$$

Критерий согласия χ^2 :

- Количество промежутков $k = 8$ (4)
- Уровень значимости $\alpha = 0.05$
- Тогда квантиль из таблицы [2, с.358] $\chi_{1-\alpha}^2(k-1) = \chi_{0.95}^2(7) = 14.0671$

В таблице 1 приведено вычисление χ_B^2 при проверке гипотезы H_0 о нормальном законе распределения $N(x, \hat{\mu}, \hat{\sigma})$.

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

i	Δ_i	n_i	p_i	np_i	$n_i - np_i$	$\frac{(n_i - np_i)^2}{np_i}$
1	$-\infty, -1.352$	8	0.088	8.8	-0.8	0.0727
2	$-1.352, -0.905$	9	0.095	9.5	-0.5	0.0263
3	$-0.905, -0.458$	14	0.141	14.1	-0.1	0.0007
4	$-0.458, -0.011$	19	0.172	17.2	1.8	0.1884
5	$-0.011, 0.437$	17	0.173	17.3	-0.3	0.0052
6	$0.437, 0.884$	14	0.143	14.3	-0.3	0.0063
7	$0.884, 1.331$	11	0.097	9.7	1.3	0.1742
8	$1.331, +\infty$	8	0.091	9.1	-1.1	0.1329
Σ	—	100	1.000	100.0	0.0	$0.6067 = \chi_B^2$

Табличное значение $\chi_{0.95}^2(7) = 14.0671$, выборочное $\chi_B^2 = 0.6067$, это значит, что на данном этапе гипотезу H_0 можно принять.

Исследование на чувствительность.

Распределение Лапласа.

Для выборок из 20 элементов был выбран $k = 5$. Уровень значимости был выбран $\alpha = 0.05$. В таблице 2 вычислен выборочный коэффициент χ_B^2 для выборки, соответствующей распределению Лапласа $L(x, 0, 1)$

Таблица 2: Вычисление χ_B^2 при проверке гипотезы о нормальном законе распределения для выборки $L(x, 0, 1)$

i	Δ_i	n_i	p_i	np_i	$n_i - np_i$	$\frac{(n_i - np_i)^2}{np_i}$
1	$-\infty, -0.915$	1	0.2002	4.004	-3.004	2.2538
2	$-0.915, 0.054$	5	0.3259	6.518	-1.518	0.3535
3	$0.054, 1.022$	9	0.2939	5.878	3.122	1.6582
4	$1.022, 1.991$	4	0.1116	2.232	1.768	1.4004
5	$1.991, +\infty$	1	0.0684	1.368	-0.368	0.0989
Σ	—	20	1.0000	20.000	0.000	$5.7648 = \chi_B^2$

Табличное значение $\chi_{0.95}^2(4) = 9.4877$, выборочное значение $\chi_B^2 = 5.7648$, значит гипотеза также может быть принята на данном этапе.

Равномерное распределение.

В таблице 3 вычислен выборочный коэффициент χ_B^2 для выборки, соответствующей равномерному распределению.

Таблица 3: Вычисление χ_B^2 при проверке гипотезы о нормальном законе распределения для выборки $U(x, -\sqrt{3}, \sqrt{3})$

i	Δ_i	n_i	p_i	np_i	$n_i - np_i$	$\frac{(n_i - np_i)^2}{np_i}$
1	$-\infty, -1.402$	1	0.0952	1.904	-0.904	0.4292
2	$-1.402, -0.406$	4	0.2875	5.750	-1.750	0.5326
3	$-0.406, 0.589$	6	0.2875	5.750	0.250	0.0109
4	$0.589, 1.586$	8	0.2875	5.750	2.250	0.8804
5	$1.586, +\infty$	1	0.0423	0.846	0.154	0.0280
Σ	—	20	1.0000	20.000	0.000	$1.8811 = \chi_B^2$

Табличное значение $\chi_{0.95}^2(4) = 9.4877$, выборочное значение $\chi_B^2 = 1.8811$, значит гипотеза также может быть принята на данном этапе.

4.2 Доверительные интервалы для параметров нормального распределения

В таблице 4 представлены оценки на основе статистик Стьюдента и хи-квадрат.

Таблица 4: Интервальные оценки на основе статистик Стьюдента и хи-квадрат

$n = 20$	m (5)	σ (6)
	$-0.43 < m < 0.63$	$0.86 < \sigma < 1.65$
$n = 100$	m	σ
	$-0.22 < m < 0.15$	$0.82 < \sigma < 1.08$

4.3 Доверительные интервалы для параметров произвольного распределения. Асимптотический подход

В таблице 5 представлены асимптотические интервальные оценки.

Таблица 5: Асимптотические интервальные оценки

$n = 20$	m (7)	σ (8)
	$-0.38 < m < 0.58$	$0.91 < \sigma < 1.49$
$n = 100$	m	σ
	$-0.21 < m < 0.15$	$0.81 < \sigma < 1.08$

Заключение

В рамках лабораторной работы была сгенерирована выборка размером 100 элементов для нормального распределения $N(x, 0, 1)$. Для неё методом максимального правдоподобия были найдены оценки параметров μ и σ . Также для данной функции гипотеза о её нормальности была проверена критерием χ^2 . Также критерием χ^2 были проверены гипотезы о нормальности выборок из 20 элементов, соответствующих распределению Лапласа и равномерному распределению.

Было установлено, что применение метода максимального правдоподобия для нормального распределения сводится к нахождению выборочного среднего и среднеквадратического отклонения.

Критерий χ^2 не отверг гипотезу о соответствии нормальной выборки нормальному распределению, чего и следовало ожидать.

Критерий χ^2 также не отверг гипотезу о нормальности выборок, соответствующих равномерному распределению и распределению Лапласа размером 20 элементов. Это объясняется тем, что критерий χ^2 носит асимптотический характер, а значит может давать ошибки при малых размерах выборки.

Также для выборок нормального распределения размером 20 и 100 элементов были построены доверительные интервалы для параметров положения и масштаба. Интервалы были получены двумя способами: с помощью классических интервальных оценок и асимптотических оценок.

На выборках малого размера асимптотические оценки оказались точнее, что странно, но может быть списано на ошибки, возникающие при работе с выборками малого размера. При выборках большего размера разницы в доверительных интервалах почти не наблюдается, что может свидетельствовать о том, что при увеличении мощности выборки ещё сильнее можно будет наблюдать более высокую точность классических интервальных оценок.

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).
- [3] Методы нахождения оценок: метод максимального правдоподобия - <https://nsu.ru/mmф/tvims/chernova/ms/lec/node14.html> (дата обращения - 09.01.2021)

Приложение А. Репозиторий с исходным кодом

Ссылка на репозиторий GitHub с исходным кодом: <https://github.com/pchn/TeorVer>