

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:
Самутичев Евгений Романович
группа: 3630102/70201

Проверил:
к.ф.-м.н., доцент
Баженов Александр Николаевич

Санкт-Петербург
2020 г.

Содержание

1	Постановка задачи	2
2	Теория	3
2.1	Точечное оценивание	3
2.1.1	Основные понятия	3
2.1.2	Метод максимального правдоподобия	3
2.2	Критерий согласия χ^2	3
3	Реализация	5
4	Результаты	6
5	Обсуждение	7
6	Приложения	8

Список иллюстраций

1 Постановка задачи

Сгенерировать выборку объёмом 100 элементов для нормального распределения $N(0, 1)$. 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 $N(\hat{\mu}, \hat{\sigma})$. Проверить основную гипотезу, используя критерий согласия χ^2 . В качестве уровня значимости взять $\alpha = 0.05$. Привести таблицу вычислений χ^2 .

2 Теория

2.1 Точечное оценивание

2.1.1 Основные понятия

Пусть имеется выборка $\{x_i\}_{i=1}^n$ из генеральной совокупности с плотностью распределения $f(x, \theta)$. Предполагается что функциональный вид зависимости задан с точностью до неизвестного параметра θ . Требуется по выборке наблюдений $\{x_i\}_{i=1}^n$ определить число $\hat{\theta}_n$ которое можно принять за значение параметра θ . Точечной оценкой неизвестного параметра θ распределения называется борелевская функция наблюдений $\hat{\theta}_n = \hat{\theta}_n(x_1, \dots, x_n)$, приближенно равная θ . Следует заметить что параметр может быть векторным, к примеру $\theta = (\mu, \sigma)$ для нормального распределения.

2.1.2 Метод максимального правдоподобия

Рассмотрим один общий метод построения точечных оценок. Для начала введем важное понятие, *функцией правдоподобия* (ФП) называется совместная плотность вероятности распределения n независимых с.в. x_1, \dots, x_n :

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

Оценкой максимального правдоподобия (о. м. п.) будем называть такое значение $\hat{\theta}_{\text{мп}}$ из множества допустимых значений параметра θ , для которого ФП принимает наибольшее значение при заданных x_1, \dots, x_n :

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

Легко обобщается на случай векторного параметра $\theta = (\theta_1, \dots, \theta_m)$:

$$\hat{\theta}_{\text{мп}} = \arg \max_{\theta_1, \dots, \theta_m} L(x_1, \dots, x_n, \theta_1, \dots, \theta_m) \quad (3)$$

Известно [1, стр. 444] что о. м. п. нормального распределения являются выборочное среднее и выборочная дисперсия:

$$\hat{\mu}_{\text{мп}} = \bar{x} \quad \hat{\sigma}_{\text{мп}} = \sqrt{s^2} \quad (4)$$

2.2 Критерий согласия χ^2

Для проверки гипотезы о законе распределения применяются различные критерии согласия. В данной работе рассматривается наиболее обоснованный и наиболее часто используемый в практике - критерий χ^2 [1, стр. 482]. И так, выдвинута гипотеза H_0 о генеральном законе распределения с функцией распределения $F(x)$. Под конкурирующей гипотезой H_1 понимается гипотеза о справедливости одного из конкурирующих распределений.

Разобьем множество значений изучаемой случайной величины X на k непересекающихся подмножеств $\Delta_1, \dots, \Delta_k$ и пусть $p_i = \mathbf{P}(X \in \Delta_k)$. Если множество значений представляет вещественную ось, то подмножества имеют вид:

$$\Delta_i = (a_{i-1}, a_i], i = 2, \dots, k-1 \quad \Delta_1 = (-\infty, a_1] \quad \Delta_k = (a_{k-1}, +\infty) \quad (5)$$

Пусть n_1, \dots, n_k - частоты попадания выборочных элементов в подмножества $\Delta_1, \dots, \Delta_k$ соответственно. В случае справедливости гипотезы относительные частоты $\frac{n_i}{n}$ должны быть

близки к p_i при $i = 1, \dots, k$. Поэтому за меру отклонения было предложено (К. Пирсоном) [1, стр. 483] выбрать значение

$$\chi_B^2 = \sum_{i=1}^k \frac{n}{p_i} \left(\frac{n_i}{n} - p_i \right)^2 = \sum_{i=1}^k \frac{(n_i - np_i)^2}{np_i} \quad (6)$$

Существует **теорема**: статистика критерия χ^2 асимптотически распределена по закону χ^2 с $k - 1$ степенями свободы. На основе этой теоремы формируется правило проверки гипотезы о законе распределения по методу χ^2 : можно принять гипотезу H_0 на уровне значимости α если $\chi_B^2 < \chi_{1-\alpha}^2$, в противном случае она отвергается.

В данной работе k и длины $\Delta_1, \dots, \Delta_k$ выбирались по правилам, которые обычно используют при построении гистограмм [2]. Правило Райса для числа интервалов:

$$k = \lceil 1.72 \sqrt[3]{n} \rceil \quad (7)$$

и правило Фридмана-Дайсона для ширины (считаем все интервалы кроме крайних одинаковой ширины)

$$a_i = \text{med } N(\hat{\mu}, \hat{\sigma}) + \left(i - \frac{k-1}{2} \right) h, \text{ где } h = 2 \frac{\text{IQR}(x_1, \dots, x_n)}{\sqrt[3]{n}}, i = 2, \dots, k-1 \quad (8)$$

, где $\text{IQR}(x_1, \dots, x_n)$ - выборочная интерквартильная широта, $\text{med } N(\hat{\mu}, \hat{\sigma})$ - медиана гипотетического распределения (т.к. предполагается что именно в окрестности медианы будет большая часть элементов выборки).

3 Реализация

Работа выполнена с использованием языка **Python** в интегрированной среде разработки **PyCharm**, были задействованы библиотеки:

- **NumPy** - векторизация вычислений, работа с массивами данных, вычисление выборочных характеристик
- **SciPy** - модуль **stats** для генерации данных, работы с распределениями, оценки методом максимального правдоподобия

Исходный код работы приведен в приложении.

4 Результаты

Оценки:

$$\hat{\mu} = 0.03 \quad \hat{\sigma} = 1.01 \quad (9)$$

Число промежутков: $k = \lceil 1.72 \cdot \sqrt[3]{100} \rceil = 8$

Таблица вычислений χ^2 :

i	Δ_i	n_i	p_i	$n_i - np_i$
1	$(-\infty, -1.79]$	2	0.0366	-1.6609
2	$(-1.79, -1.27]$	7	0.0637	0.627
3	$(-1.27, -0.75]$	12	0.121	-0.0972
4	$(-0.75, -0.23]$	22	0.1777	4.2306
5	$(-0.23, 0.29]$	22	0.202	1.8009
6	$(0.29, 0.81]$	15	0.1777	-2.7694
7	$(0.81, 1.33]$	8	0.121	-4.0972
8	$(1.33, +\infty)$	12	0.1003	1.9661

Таблица 1: Таблица вычислений χ^2

При $\alpha = 0.05$: $\chi^2_{1-\alpha}(k-1) \approx 14.0671$, а вычисленное $\chi^2_{\text{B}} = 4.1883$, видно что $\chi^2_{\text{B}} < \chi^2_{1-\alpha}(k-1)$

5 Обсуждение

Согласно результатам эксперимента, заданное по оценкам (9) распределение $N(\hat{\mu}, \hat{\sigma})$ является генеральным законом по которому построена выборка с уровнем значимости 0.05. Теоретически это обосновывается тем что оценки максимального правдоподобия состоятельны.

6 Приложения

1. Исходный код лабораторной <https://github.com/zhenyatos/statlabs/tree/master/Lab7>

Список литературы

- [1] **Вероятностные разделы математики.** Учебник для бакалавров технических направлений. // Под ред. Максимова Ю.Д. - СПб «Иван Федоров», 2001. - 592 с., илл
- [2] Wikipedia contributors. (2020, March 19). Histogram. In Wikipedia, The Free Encyclopedia. Retrieved 18:27, May 14, 2020, from <https://en.wikipedia.org/w/index.php?title=Histogram&oldid=946321806>