Санкт-Петербургский Политехнический университет
Петра Великого
Институт прикладной математики и механики
Высшая школа прикладной математики и вычислительной физики

Отчёт
по лабораторной работе 7
по дисциплине
"математическая статистика"

Выполнил студент:

Аникин Александр Алексеевич,
группа 3630102\80201

Проверил:

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

Санкт-Петербург
2021

Содержание

Список таблиц

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. Исследовать точность (чувствительность) критерия - сгенерировать выборки равномерного распределения и распределения Лапласа малого объема, проверить их на нормальность.

2 Теория

2.1 Метод максимального правдоподобия

$L(x_1, ..., x_n, \theta)$ - функция правдоподобия, рассматриваемая как функция неизвестного параметра θ:

$$L(x_1, ..., x_n, \theta) = f(x_1, \theta)f(x-2, \theta)...f(x_n, \theta) \tag{1}$$

Оценка максимального правдоподобия:

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

Система уравнений правдоподобия (в случае дифференцируемости функции правдоподобия):

$$\frac{\partial l}{\partial \theta_k} = 0 \quad \text{или} \quad \frac{\partial \ln l}{\partial \theta_k} = 0, \quad k = 1, ..., m \tag{3}$$

2.2 Проверка гипотезы о законе распределения генеральной совокупности критерием χ^2

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 критерием χ^2:

- Выбирается уровень значимости α;
- По таблице ([1], стр. 358) выбирается квантиль $\chi^2_{1-\alpha}(k-1)$ порядка 1-α с $k-1$ степенями свободы;
- С помощью гипотетической функции распределения $F(x)$ вычисляются вероятности $p_i = P(X \in \Delta_i), \quad i = 1, ..., k$;
- Находятся частоты n_i попадания элементов выборки в поднмножества $\Delta_i, i = 1, ..$
- Вычисляется выборочное значение статистики критерия χ^2:

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

- Сравниваются χ^2_B и квантиль $\chi^2_{1-\alpha}(k-1)$:

- если $\chi^2_B < \chi^2_{1-\alpha}(k-1)$, то гипотеза H_0 на данном этапе проверки принимается;
- если $\chi^2_B \chi^2_{1-\alpha}(k-1)$, то гипотеза H_0 отвергается, выбирается одно из альтернативных распределений, и процедура проверки повторяется.

3 Реализация

Лабораторная работа выполнена на языке Python 3.8 с помощью загружаемых пакетов SciPy, NumPy. Исходный код лабораторной работы находится на GitHub репозитории.

4 Результаты

4.1 Проверка гипотезы о законе распределения генеральной совокупности критерием χ^2

4.1.1 Нормальное распределение $N(x, \hat{\mu}, \hat{\sigma})$

Основная гипотеза H_0: $F(x) = N(x, \bar{\mu}, \bar{\sigma})$.
Метод максимального правдоподобия:

$$\hat{\mu} = -0.035 \quad \hat{\sigma} = 1.041$$

Критерий χ^2 :

- Количество промежутков: $k = 7$;
- Уровень значимости: $\alpha = 0.05$;
- Квантиль распределения $\chi^2_{0.95}(6) = 12.59$;

i	границы Δ_i	n_i	p_i	np_i	$n_i - np_i$	χ^2_i
1	[-∞ , -1.0]	18	0.175	17.530	0.47	0.013
2	[-1.0 , -0.6]	11	0.114	11.374	-0.374	0.0120
3	[-0.6 , -0.2]	12	0.140	13.994	-1.994	0.284
4	[-0.2 , 0.2]	16	0.150	14.960	1.040	0.072
5	[0.2 , 0.6]	17	0.139	13.894	3.106	0.694
6	[0.6 , 1.0]	7	0.112	11.211	-4.211	1.582
7	[1.0 , ∞]	19	0.170	17.037	1.963	0.226
$\sum$	-	100	1	100	0	$\chi^2_B = 2.146$

Таблица 1: Нормальное распределение, проверка гипотезы

$\chi^2_B = 2.146 < \chi^2_{0.95}(6) = 12.59$, значит, на данном этапе проверки текущая гипотеза принимается.

4.1.2 Исследование критерия χ^2 на чувствительность

Генерируются выборки равномерного распределения и распределения Лапласа по 20 элементов и проверяется гипотеза, что полученные наборы данных являются выборками нормального распределения.

- Равномерное распределение $U(x, -1.5, 1.5)$:

  Метод максимального правдоподобия:

$$\hat{\mu} = 0.421 \quad \hat{\sigma} = 0.577$$

  Критерий χ^2 :

  - Количество промежутков: $k = 5$;
  - Уровень значимости: $\alpha = 0.05$;
  - Квантиль распределения $\chi^2_{0.95}(4) = 9.49$;

i	границы Δ_i	n_i	p_i	np_i	$n_i - np_i$	χ^2_i
1	[$-\infty$, -1.0]	1.0	0.159	3.173	-2.173	1.488
2	[-1.0 , -0.333]	3.0	0.211	4.216	-1.216	0.351
3	[-0.333 , 0.333]	6.0	0.261	5.222	0.778	0.116
4	[0.333 , 1.0]	4.0	0.211	4.216	-0.216	0.011
5	[1.0 , ∞]	6.0	0.159	3.173	2.827	2.518
$\sum$	-	20	1	20	0	$\chi^2_B = 4.484$

Таблица 2: Равномерное распределение, проверка на устойчивость

$\chi^2_B = 4.484 < \chi^2_{0.95}(6) = 9.49$, значит, на данном этапе проверки текущая гипотеза принимается.

- Распределение Лапласа $L(x, 0, 1)$:

  Метод максимального правдоподобия:

$$\hat{\mu} = 0.165 \quad \hat{\sigma} = 1.232$$

  Критерий χ^2 :

  - Количество промежутков: $k = 5$;
  - Уровень значимости: $\alpha = 0.05$;
  - Квантиль распределения $\chi^2_{0.95}(4) = 9.49$;

i	границы Δ_i	n_i	p_i	np_i	$n_i - np_i$	χ_i^2
1	[$-\infty$, -1.0]	2.0	0.159	3.173	-1.173	0.434
2	[-1.0 , -0.333]	7.0	0.211	4.216	2.784	1.839
3	[-0.333 , 0.333]	4.0	0.261	5.222	-1.222	0.286
4	[0.333 , 1.0]	3.0	0.211	4.216	-1.216	0.351
5	[1.0 , ∞]	4.0	0.159	3.173	0.827	0.215
$\sum$	-	20	1	20	0	$\chi_B^2 = 3.124$

Таблица 3: Распределение Лапласа, проверка на устойчивость

$\chi_B^2 = 3.124 < \chi^2_{0.95}(6) = 9.49$, значит, на данном этапе проверки текущая гипотеза принимается.

5 Обсуждение

Проведенное исследование показало, что метод χ^2 неэффективен для выборок малого размера - по результатам исследования на чувствительность выборки распределения Лапласа и равномерного распределения воспринимались как выборки нормального распределения, поэтому для более точной проверки гипотез о законах распределения следует проводить проверку на большем объеме данных.

Список литературы

[1] 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).