Санкт-Петербургский Политехнический Университет
им. Петра Великого

Институт прикладной математики и механики

Кафедра "Прикладная математика"

Отчёт
Лабораторная работа № 7
по дисциплине
"Математическая статистика"

Выполнил студент:
Шарапов Сергей Андреевич
группа: 3630102/70401

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

Санкт-Петербург
2020 год

Содержание

1 Список таблиц

2 Постановка задачи

Необходимо 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. В качестве ровня значимости взять $\alpha = 0,05$. Привести таблицу вычислений χ^2.

3 Теория

3.1 Метод максимального правдоподобия

Метод максимального правдоподобия – метод оценивания неизвестного параметра путём максимимзации функции правдоподобия.

$$\hat{\theta}_{\text{МП}} = argmax\mathbf{L}(x_1,x_2,\ldots,x_n,\theta) \tag{1}$$

Где $\mathbf{L}$ это функция правдоподобия, которая представляет собой совместную плотность вероятности независимых случайных величин $X_1,x_2,\ldots,x_n$ и является функцией неизвестного параметра θ

$$\mathbf{L} = f(x_1,\theta)\cdot f(x_2,\theta)\cdot\dots\cdot f(x_n,\theta) \tag{2}$$

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

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

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

Отсюда находятся выражения для оценок m и σ^2:

$$\begin{cases} m = \bar{x} \\ \sigma^2 = s^2 \end{cases} \tag{4}$$

3.2 Критерий согласия Пирсона

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

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:

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

Тогда

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

Для выполнения гипотезы H_0 должны выполняться следующие условия:

$$\chi^2_B < \chi^2_{1-\alpha}(k-1) \tag{7}$$

где $\chi^2_{1-\alpha}(k-1)$ – квантиль распределения χ^2 с $k-1$ степенями свободы порядка $1-\alpha$, где α заданный уровень значимости.

4 Реализация

Работы была выполнена на языке $Python$3.8.2 Для генерации выборок использовался модуль . Для генерации выборок и обработки функции распределения использовалась библиотека *scipy.stats*.

5 Результаты

5.1 Метод максимального правдоподобия

При подсчете оценок параметров закона нормального распределения методом максимального правдоподобия были получены следующие значения:

$$\begin{aligned} \hat{m}_{\text{МП}} &= 0.090527 \\ \hat{\sigma}^2_{\text{МП}} &= 0.963167 \end{aligned} \tag{8}$$

5.2 Критерий Пирсона

Таблица 1: Таблица вычислений χ^2

i	Δ_i	n_i	p_i	$\frac{(n_i-np_i)^2}{np_i}$
1	$(-\infty, -1.0]$	15	0.1288	0.3501
2	$(-1.0, -0.5)$	10	0.1411	1.1988
3	$(-0.5, 0.0)$	24	0.1927	1.1634
4	$(0.0, 0.5)$	19	0.2021	0.0721
5	$(0.5, 1.0)$	13	0.1629	0.6626
6	$(1.0, \infty)$	19	0.1725	0.1771
$\sum$		100	1	3.6241

$$\chi^2_B = 3.6241$$

5.3 Проверка гипотезы о нормальности для распределения Лапласа

Размер выборки $n = 25$ для распределения Лапласа

$$L\left(x, 0, \frac{1}{\sqrt{2}}\right) = \frac{1}{\sqrt{2}} e^{-\sqrt{2}|x|} \tag{9}$$

$$\begin{aligned} \hat{m}_{\text{МП}} &= 0.198045 \\ \hat{\sigma}^2_{\text{МП}} &= 0.656187 \end{aligned} \tag{10}$$

Таблица 2: Таблица вычислений χ^2

i	Δ_i	n_i	p_i	$\frac{(n_i - np_i)^2}{np_i}$
1	$(-\infty, -1.0]$	1	0.0339	0.027
2	$(-1.0, 0.0)$	12	0.3475	1.2641
3	$(0.0, 1.0)$	8	0.5078	1.7359
4	$(1.0, \infty)$	4	0.1108	0.5454

$$\chi^2_B = 3.5725$$

6 Выводы

Табличное значение квантиля $\chi^2_{1-\alpha}(k-1) = \chi^2_{0.95}(5) = 11.0705$. Полученное значение критерия согласия Пирсона для нормального распределения $\chi^2_B = 3.6241 < \chi^2_{0.95}(5)$, следовательно основная гипотеза H_0 на исходной выборке не может быть отвергнута на уровне значимости $\alpha = 0.05$.. Для распределения Лапласа полученное значение критерия Пирсона $\chi^2_B = 3.5725 < \chi^2_{0.95}(3) = 7.8147$ означает что из полученной выборки мы не можем отвергнуть гипотезу H_0 о нормальности исходного распределения. Такой результат легко объясним низким размером выборки, так как интервалы в которых мы оцениваем распределение получаются слишком большими, на которых распределение Лапласа очень схоже с нормальным.

7 Литература

Модуль numpy

Модуль matplotlib

Модуль scipy

Шевляков Г. Л. Лекции по математической статистике, 2019.

8 Приложения

Код лабораторной