Санкт-Петербургский Политехнический Университет им. Петра Великого

Институт прикладной математики и механики

Кафедра прикладной математики

Отчёт по лабораторной работе №7 по дисциплине “Математическая статистика”

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

Выполнил студент:

Мишутин Д. В.

Группа:

3630102/70301

Проверил:

К.ф.-м.н., доцент

Баженов Александр Николаевич

Санкт-Петербург

2020 г.

Оглавление

Список иллюстраций и таблиц

1 Постановка задачи

Сгенерировать выборку объёмом 100 элементов для стандартного 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.

2 Теория

Нормальное распределение:

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

2.1 Метод максимального правдоподобия (ММП)

МНМ – метод оценивания неизвестного параметра θ путём максимизации функции правдоподобия $L(X, \theta)$:

$$\hat{\theta}_{\text{ОМП}} = argmax[L(X, \theta)]$$

$$L(X, \theta) = \prod_{i=1}^{n} f(x_i, \theta)$$

Оценкой максимального правдоподобия будем называть такое значение $\hat{\theta}_{\text{ОМП}}$ из множества допустимых значений θ, для которого $L(X, \theta)$ принимает максимальное значение для заданных $x_1, \dots, x_n$.

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

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

Отсюда находятся выражения для оценок μ и σ^2:

$$\begin{cases} \mu = \overline{x} \\ \sigma^2 = s^2 \end{cases}$$

2.2 Критерий Пирсона

Разобьём генеральную совокупность на k непересекающихся подмножеств $\Delta_1, \dots, \Delta_k$, где $\Delta_i = (x_i, x_{i+1}], p_i = P(X \in \Delta_i), i = \overline{1, k}$ – вероятность того, что точка попала в i-ый промежуток.

Так как генеральная совокупность это $\mathbb{R}$, то крайние промежутки будут бесконечными: $\Delta_1 = (-\infty, x_1], \Delta_k = (x_k, \infty], p_i = F(x_i) - F(x_{i-1})$

Пусть n_i – частота попадания выборочных элементов в Δ_i.

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:

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

Тогда

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

Для выполнения гипотезы H_0 должны выполняться следующие условия:

$$\chi_B^2 < \chi_{1-\alpha}^2(k-1),$$

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

3 Реализация

Был использован язык *Python 3.8.2*: модуль *numpy* для генерации выборок на основе стандартного нормального распределения и вычисления описательных статистик, модуль *scipy.stats* для расчёта коэффициентов, модуль *pandas* для оптимального хранения статистических данных и функция *display* из модуля *IPython.display* для их корректного отображения в таблицах.

4 Результаты

4.1 Метод максимального правдоподобия (ММП)

При подсчёте оценок параметров закона нормального распределения с помощью МНМ были получены следующие результаты:

$$\hat{\mu}_{\text{ОМП}} = 0.0598$$

$$\widehat{\sigma^2}_{\text{ОМП}} = 1.0079$$

4.2 Критерий Пирсона

Таблица 3 Вычисления χ^2

i	Δ_i	n	p_i	χ_B^2

1	-1	15.0	0.1465	0.0083
2	-0.5	14.0	0.1428	0.0055
3	0	16.0	0.1870	0.3909
4	0.5	27.0	0.1925	3.1189
5	1	10.0	0.1557	1.9923
6	inf	18.0	0.1755	0.0118

$$\chi_B^2 = 5.5277$$

5.3 Проверка гипотезы о нормальности для равномерного распределения

Размер выборки $n = 20$, заданный отрезок $[-2, 2]$.

$$U(x, -2,2) = \begin{cases} \frac{1}{4}, x \in [-2, 2] \\ 0, \text{иначе} \end{cases}$$

$$\hat{\mu}_{\text{ОМП}} = -0.0309$$

$$\widehat{\sigma^2}_{\text{ОМП}} = 0.8724$$

Таблица 4 Вычисления χ^2

i	Δ_i	n	p_i	χ_B^2
1	-2	0.0	0.0120	0.2399
2	1	18.0	0.8693	0.0216
3	4	2.0	0.1187	0.0587
4	inf	0.0	0.0000	0.0000

$$\chi_B^2 = 0.3203$$

5 Выводы

Табличное значение квантиля $\chi^2_{0.95}(5) = 11.0705$. Полученное значение критерия согласия Пирсона для нормального распределения $\chi_B^2 = 5.5277 < 11.0705$, следовательно основная гипотеза H_0 не может быть опровергнута на уровне значимости $\alpha = 0.05$.

Для равномерного распределения полученное значение критерия Пирсона $\chi_B^2 = 0.3203 < \chi^2_{0.95}(3) = 7.81473$ означает, что из полученной выборки мы не можем опровергнуть гипотезу H_0 о нормальности данного распределения.

6 Литература

Основы работы с *numpy* (отдельная глава курса)

Документация по *scipy*

Pandas обзор

Таблица значений χ^2

7 Приложения

[Код лабораторной]