Санкт-Петербургский политехнический университет Петра Великого
Институт прикладной математики и механики
Кафедра «Прикладная математика»

КУРСОВАЯ РАБОТА ПО ДИСЦИПЛИНЕ «МАТЕМАТИЧЕСКАЯ СТАТИСТИКА»

Выполнил
студент группы 3630102/70301

Мустафаев Шамиль

Проверил
к. ф.-м. н., доцент

Баженов Александр Николаевич

Санкт-Петербург
2020

Содержание

Список таблиц

Список иллюстраций

1 Постановка задачи

Реализовать критерий Мизеса-Смирнова, проверить на выборках разной мощности и с разными функциями распределения.

2 Теория

2.1 Критерий Крамера-Мизеса-Смирнова

Данный критерий применяется для проверки гипотез вида $H_0 : F_n(x) = F(x, \theta)$ с известным набором параметров теоретического закона.
Статистика критерия Крамера-Мизеса-Смирнова(иначе - критерия ω^2) имеет вид:

$$S = n\omega^2 = \frac{1}{12n} + \sum_{i=1}^{n} (F(x_i, \Theta) - \frac{2i-1}{2n})^2, \tag{1}$$

где n - объем выборки, x_i - элемент выборки, упорядоченной по возрастанию.
При справедливости гипотезы статистика критерия должна подчиняться асимптотическому закону $a_1(S)$.[3]
Таким образом, при использовании критерия необходимо:

1. Вычислить значение статистики по формуле для S(1).
2. По таблице из [2] определить значение a функции распределения $a_1(S)$ для только что вычисленного результата.
3. Выбрать уровень значимости α.
4. Если $a \geq 1 - \alpha$, то гипотезу о согласии эмпирического и теоретического распределений отвергают, в противном случае гипотеза принята.

2.2 Распределения

- Нормальное распределение

$$N(x, 0, 1) = \frac{1}{\sqrt{2\pi}} e^{\frac{-x^2}{2}} \tag{2}$$

- Равномерное распределение

$$U(x, -\sqrt{3}, \sqrt{3}) = \begin{cases} \frac{1}{2\sqrt{3}} & |x| \leq \sqrt{3} \\ 0 & |x| > \sqrt{3} \end{cases} \tag{3}$$

3 Реализация

Лабораторная работа выполнена с помощью встроенных средств языка программирования Python в среде разработки PyCharm. Использованы библиотеки numpy для простоты использования различных статистических функций, scipy для простоты вычислений. Исходный код лабораторной работы приведён в приложении в виде ссылки на репозиторий GitHub.

4 Результаты

4.1 Результаты использования критерия ω^2

В качестве уровня значимости рекомендуется брать $\alpha = 0.1$ или 0.2.
Выберем $\alpha = 0.1$.
В качестве гипотезы рассмотрим H_0 о нормальном законе распределения $N(0,1)$.
Теперь возьмем и проверим согласованность распределений для выборок, сгенерированных по нормальному закону. Представим данные в виде таблицы:

Distribution	S = $n\omega^2$	$a_1(S)$	$1-\alpha$	Result
Normal, n = 10	0.039	0.02568	0.9	True
Normal, n = 100	0.04	0.06685	0.9	True
Normal, n = 500	0.16	0.63951	0.9	True

Таблица 1: Таблица критерия ω^2 для выборок, распределенных нормально, разных мощностей

Заметим, что мы так же могли бы взять $\alpha = 0.2$.
Теперь рассмотрим чувствительность критерия. Для этого сгенерируем несколько выборок разных мощностей, распределенных по равномерному закону $U(-\sqrt{3}, \sqrt{3})$.

Distribution	S = $n\omega^2$	$a_1(S)$	$1-\alpha$	Result
Uniform, n = 10	0.12	0.50457	0.9	Normal
Uniform, n = 100	0.29	0.85573	0.9	Normal
Uniform, n = 500	0.97	0.9971	0.9	Not normal

Таблица 2: Таблица критерия ω^2 для выборок, распределенных равномерно, разных мощностей

Для выборки равномерного распределения мощностью 100 элементов построим гистограмму:

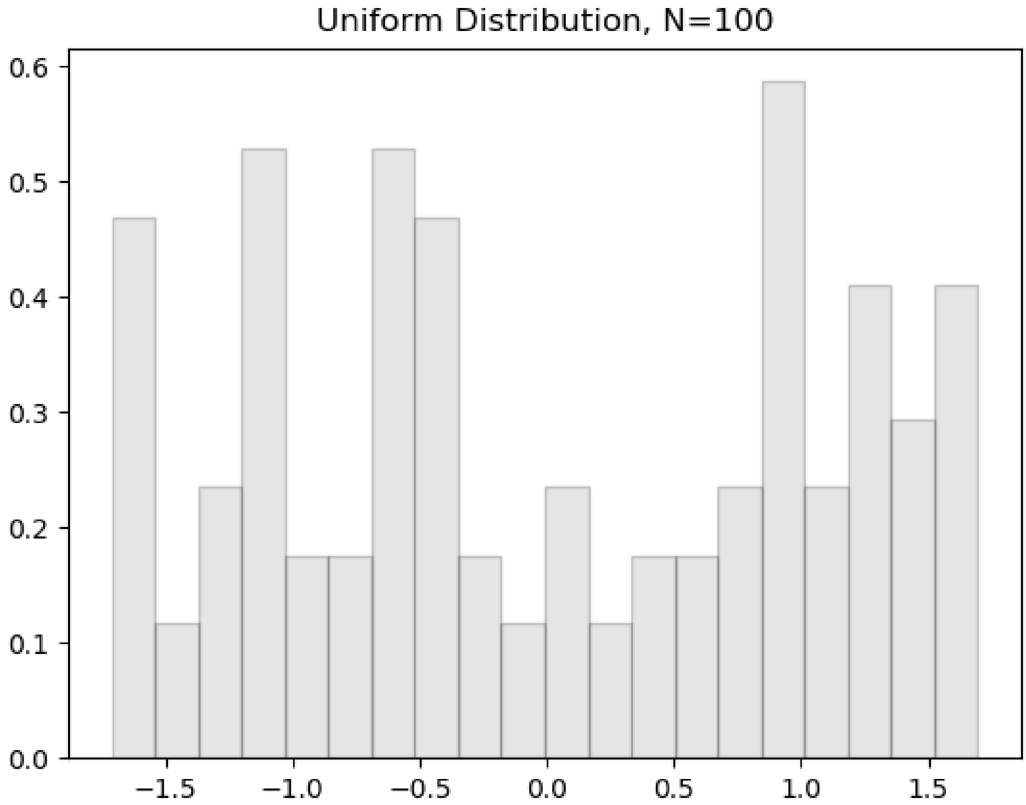

Рис. 1: Гистограмма для выборки, распределенной равномерно, мощностью 100

Проверим также выборку на нормальность с помощью критерия χ^2, описанного ранее.

i	Границы $\Delta_i(a_{i-1}, a_i]$	n_i	p_i	np_i	$n_i - np_i$	$\frac{(n_i-np_i)^2}{np_i}$
1	(-∞, -1.4]	13	0.0808	8.08	4.92	3
2	(-1.4, -0.7]	16	0.1612	16.12	-0.12	0
3	(-0.7, 0.0]	24	0.258	25.8	-1.8	0.13
4	(0.0, 0.7]	16	0.258	25.8	-9.8	3.72
5	(0.7, 1.4]	22	0.1612	16.12	5.88	2.14
6	(1.4, ∞)	9	0.0808	8.08	0.92	0.11
$\sum$	-	100	1	100	0	9.1

Таблица 3: Вычисление χ^2 в при проверке гипотезы H_0 о нормальном законе распределения $U(-\sqrt{3}, \sqrt{3})$ при n = 100

Теоретический квантиль для выбранного уровня значимости α: $\chi^2_{0.9} \approx 9.24$. Как видно из таблицы (3), $\chi^2_B = 9.1$.

$\chi^2_B < \chi^2_{0.9}$, следовательно, данный критерий также не отвергает гипотезу нормальности.

5 Обсуждение

Из таблицы (1) видно, что для нормального распределения критерий Крамера-Мизеса-Смирнова принимает верную гипотезу H_0 для различных мощностей выборки.
Также заметно, что можно использовать различные уровни значимости.
Однако, чувствительность критерия проявляется только на выборках большого размера. Для распределения $U(-\sqrt{3}, \sqrt{3})$ критерий отвергает гипотезу о нормальности распределения лишь при n = 500, но при этом при мощности выборки 100 элементов значение, получаемое из $a_1(S)$ уже близко к граничному.

6 Приложения

Код программы на GitHub, URL: `https://github.com/sh4mik/MathStat`

Список литературы

[1] СРАВНИТЕЛЬНЫЙ АНАЛИЗ МОЩНОСТИ КРИТЕРИЕВ СОГЛАСИЯ ПРИ БЛИЗКИХ КОНКУРИРУЮЩИХ ГИПОТЕЗАХ. I. ПРОВЕРКА ПРОСТЫХ ГИПОТЕЗ. Б. Ю. Лемешко, С. Б. Лемешко, С. Н. Постовалов

[2] Большев Л.Н., Смирнов Н.В. Таблицы математической статистики.

[3] Критерий Крамера-Мизеса-Смирнова