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

Отчет по лабораторной №1 по дисциплине «Стохастические модели и анализ данных»

Выполнил:
Чернова В.С.
Группа:
3640102/90201

Проверил:
к.ф.-м.н., доцент
Баженов А.Н.

Санкт-Петербург
2020 г

ОГЛАВЛЕНИЕ

ПОСТАНОВКА ЗАДАЧИ

Определить при каком числе измерений возможно различить выборки нормального и равномерного распределения с помощью критерия X^2.

ТЕОРЕТИЧЕСКОЕ ВВЕДЕНИЕ

1. Нормальное распределение

Нормальное распределение [1] — распределение вероятностей, которое в одномерном случае задаётся функцией плотности вероятности, совпадающей с функцией Гаусса:

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

Где μ – среднее значение, σ – среднеквадратическое отклонение (σ^2 – дисперсия).

Таким образом, одномерное нормальное распределение является двухпараметрическим семейством распределений, которое принадлежит экспоненциальному классу распределений.

2. Равномерное распределение

Равномерное распределение [2] — распределение случайной вещественной величины, принимающей значения, принадлежащие некоторому промежутку конечной длины, характеризующееся тем, что плотность вероятности на этом промежутке почти всюду постоянна.

Говорят, что случайная величина имеет непрерывное равномерное распределение на отрезке [a, b], где a, b ∈ $\mathbb{R}$, если ее плотность $f_X(x)$ имеет вид:

$$f_X(x) = \begin{cases} \frac{1}{b-a}, & x \in [a,b] \\ 0, & x \notin [a,b] \end{cases}$$

3. Критерий X^2

Критерий X^2 [3] — статистический критерий для проверки гипотезы H_0, что наблюдаемая случайная величина подчиняется некому теоретическому закону распределения.

Распределение X^2 имеет вид (рис. 1):

$$X_{k-1}^2 = \sum_{j=1}^{k} \frac{(n_j - E_j)^2}{E_j}$$

Где:

X_{k-1}^2 – распределение X^2 с k-1 степенью свободы;
n_j – количество наблюдений в j-м интервале;
E_j – ожидаемое число попаданий в j-ый интервал.

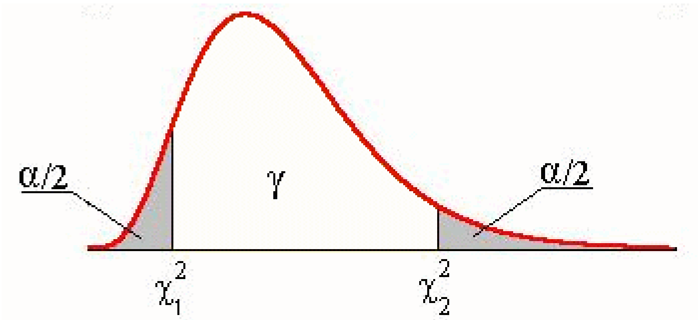


Рис. 1. Распределение X^2

В зависимости от значения критерия X^2, гипотеза H_0 может приниматься, либо отвергаться:

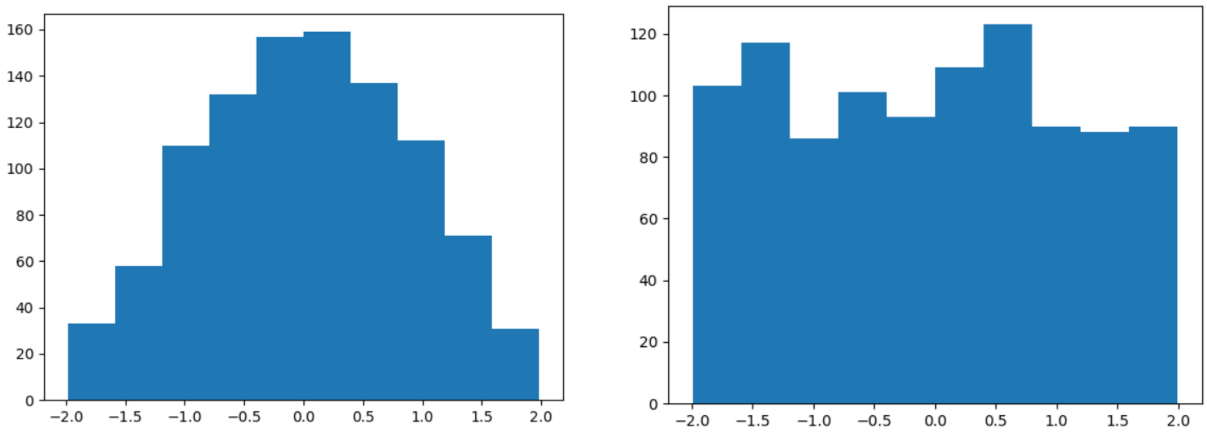
- $X_1^2 < X^2 < X_2^2$ – гипотеза H_0 выполняется;
- $X^2 < X_2^2$ – (попадает в левый "хвост" распределения). Следовательно, теоретические и практические значения очень близки;
- $X^2 > X_2^2$ – (попадает в правый "хвост" распределения) гипотеза H_0 отвергается.

X^2 смотрится по таблице квантилей [4]. Значение выбирается для заданной степени свободы и заданного уровня значимости.

ХОД РАБОТЫ

Генерировались выборки из двух распределений:

1. Выборки из нормального распределения в интервале от -2 до 2, со средним значением 0 и среднеквадратическим отклонением 1 (рис. 2а);
2. Выборки из равномерного распределения в интервале от -2 до 2 (рис. 2б).



а) б)

Рис. 2. Сгенерированная выборка из а) нормального распределения, б) равномерного распределения

Примем:

Степень свободы = 10
Уровень значимости = 0.7
Правое значение по таблице для выбранного уровня значимости – 11,7

Покажем на примере выборки из нормального распределения порядок расчета критерия Х^2. Кол-во элементов возьмем равным 80. Результаты расчетов представлены на таблицах 1-2.

Таблица 1. Порядок расчета критерия Х^2 на соответствие нормальному распределению

i	Δ_i	n_i	p_i	np_i	$n_i - np_i$	$\frac{(n_i - np_i)^2}{np_i}$
1	[-2, -1.6]	5	0.038	3.11	1.88	1.13
2	[-1.6, -1.2]	5	0.065	5.22	-0.22	0.009
3	[-0.8, -0.4]	8	0.096	7.68	0.31	0.01
4	[-0.4, 0]	13	0.13	9.94	3.05	0.94

…	…	…	…	…	…	…
10	[1.6, 2]	2	0.038	3.08	-1.08	0.38
Σ	-	80	0.93	76	5.65	5.38

Таблица 2. Порядок расчета критерия Χ^2 на соответствие равномерному распределению

i	Δ_i	n_i	p_i	np_i	$n_i - np_i$	$\frac{(n_i - np_i)^2}{np_i}$
1	[-2, -1.6]	5	0.1	8	-3	1.12
2	[-1.6, -1.2]	5	0.1	8	-3	1.12
3	[-0.8, -0.4]	8	0.1	8	0	0
4	[-0.4, 0]	13	0.1	8	5	3.12
…	…	…	…	…	…	…
10	[1.6, 2]	2	0.1	8	-6	4.5
Σ	-	80	1	80	0	14.75

Далее из нормального и равномерного распределений генерировались выборки разного размера, а затем для этих выборок считался критерий Χ^2. Таким образом для обеих выборок проверялось, подтверждается ли гипотеза принадлежности взятой выборки к нормальному и равномерному распределению. Полученные результаты представлены в таблицах 3-4.

Таблица 3. Расчет критерия Χ^2 для равномерного распределения

Размер выборки	Χ^2, сравнение с равномерным распределением	Гипотеза подтверждается?	Χ^2, сравнение с нормальным распределением	Гипотеза подтверждается?
10	6.0	Да	15.4	Нет
30	5.3	Да	11.8	Нет
50	4.8	Да	18.6	Нет
60	9.3	Да	23.7	Нет
80	4.5	Да	26.8	Нет
100	3.0	Да	32.2	Нет
150	7.7	Да	65.8	Нет

Таблица 4. Расчет критерия X^2 для нормального распределения

Размер выборки	X^2, сравнение с равномерным распределением	Гипотеза подтверждается?	X^2, сравнение с нормальным распределением	Гипотеза подтверждается?
10	8.0	Да	17.7	Нет
30	6.6	Да	15.9	Нет
50	18.8	Нет	10.3	Да
60	21.6	Нет	10.1	Да
80	15.5	Нет	5.7	Да
100	31.2	Нет	8.8	Да
150	43.6	Нет	9.5	Да

Из полученных таблиц можно сделать вывод, что при размере выборки более 60 с помощью критерия X^2 возможно достаточно точно отличить равномерное распределение от нормального и наоборот, так как верная гипотеза почти всегда подтверждается, а неверная опровергается.

ЗАКЛЮЧЕНИЕ

В процессе выполнения работы был изучен и применен на практике метод проверки соответствия какой-либо выборки заданному распределению с помощью критерия X^2. Эффективность метода была проверена на выборках различного размера из равномерного и нормального распределений. Выборки из обоих распределений брались в диапазоне от -2 до 2.

Из полученных результатов был сделан вывод, что при размере выборки более 60 с помощью критерия X^2 возможно достаточно точно отличить равномерное распределение от нормального и наоборот, так как верная гипотеза почти всегда подтверждается, а неверная опровергается.

Код написанной программы представлен по следующей ссылке: https://github.com/nika2506/stochastic_labs/tree/main/Lab_1.

СПИСОК ЛИТЕРАТУРЫ

1. Вентцель Е. С. Теория вероятностей. — 10-е изд., стер.. — М.: Academia, 2005. — 576 с. — ISBN 5-7695-2311-5
2. Непрерывное равномерное распределение [Электронный ресурс], URL: https://ru.wikipedia.org/wiki/Непрерывное_равномерное_распределение. Дата обращения: 17.10.2020
3. Прикладная статистика. Правила проверки согласия опытного распределения с теоретическим. Методические рекомендации. Часть I. Критерии типа X^2 – Новосибирск: Изд-во НГТУ, 1998. – 126 с
4. Квантили распределения хи-квадрат [Электронный ресурс], URL: https://ru.wikipedia.org/wiki/Квантили_распределения_хи-квадрат. Дата обращения: 17.10.2020