

САНКТ-ПЕТЕРБУРГСКИЙ ПОЛИТЕХНИЧЕСКИЙ
УНИВЕРСИТЕТ им. ПЕТРА ВЕЛИКОГО

ИНСТИТУТ ПРИКЛАДНОЙ МАТЕМАТИКИ И МЕХАНИКИ

КАФЕДРА ПРИКЛАДНОЙ МАТЕМАТИКИ

ЛАБОРАТОРНАЯ РАБОТА №5
ВЫБОРОЧНЫЕ КОЭФФИЦИЕНТЫ
КОРРЕЛЯЦИИ И ЭЛЛИПСЫ
РАССЕИВАНИЯ

Студент группы 3630102/70301

Камянский Д.В.

Преподаватель

Баженов А. Н.

САНКТ-ПЕТЕРБУРГ
2020 г.

Содержание

1 Список иллюстраций

2 Список таблиц

3 Постановка задачи

Сгенерировать двумерные выборки размерами 20, 60, 100 для двумерного нормального распределения $N(x, y, 0, 0, 1, 1, \rho)$.

Коэффициент корреляции ρ взять равным 0, 0.5, 0.9.

Каждая выборка генерируется 1000 раз и для неё вычисляются: среднее значение, среднее значение квадрата и дисперсия коэффициентов корреляции Пирсона, Спирмена и квадрантного коэффициента корреляции. Повторить все вычисления для смеси нормальных распределений:

$$f(x, y) = 0.9N(x, y, 0, 0, 1, 1, 0.9) + 0.1N(x, y, 0, 0, 10, 10, -0.9) \quad (1)$$

Изобразить сгенерированные точки на плоскости и нарисовать эллипс равновероятности.

4 Теория

1. Двумерное нормально распределение:

$$N(x, y, \bar{x}, \bar{y}, \sigma_x, \sigma_y, \rho) = \frac{1}{2\pi\sigma_x\sigma_y\sqrt{1-\rho^2}} \times \\ \times \exp\left(-\frac{1}{2(1-\rho^2)}\left[\frac{(x-\bar{x})^2}{\sigma_x^2} - 2\rho\frac{(x-\bar{x})(y-\bar{y})}{\sigma_x\sigma_y} + \frac{(y-\bar{y})^2}{\sigma_y^2}\right]\right) \quad (2)$$

2. Коэффициент корреляции Пирсона:

$$r = \frac{\frac{1}{n} \sum (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\frac{1}{n} \sum (x_i - \bar{x})^2 \frac{1}{n} \sum (y_i - \bar{y})^2}} \quad (3)$$

3. Квадрантный коэффициент корреляции:

$$r_Q = \frac{(n_1 + n_3) - (n_2 + n_4)}{n} \quad (4)$$

где n_1, n_2, n_3, n_4 – количества точек с координатами (x_i, y_i) , попавшими соответственно в I, II, III и IV квадранты декартовой системы с осями $x' = x - medx, y' = y - medy$ и с центром в точке с координатами $(medx, medy)$

4. Коэффициент корреляции Спирмана:

$$r_S = \frac{\frac{1}{n} \sum (u_i - \bar{u})(v_i - \bar{v})}{\sqrt{\frac{1}{n} \sum (u_i - \bar{u})^2 \frac{1}{n} \sum (v_i - \bar{v})^2}} \quad (5)$$

где u и v – ранги, соответствующие значениям переменной X и Y соответственно.

5 Реализация

Работы была выполнена на языке *Python3.8.2* Для генерации выборок использовался модуль *stats* библиотеки *scipy*. Для построения графиков использовалась библиотека *matplotlib*.

6 Результаты

Рис. 1: Графики двумерного нормального распределения и смеси для размера выборки $n = 20$



Рис. 2: Графики двумерного нормального распределения и смеси для размера выборки $n = 60$

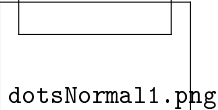


Таблица 1: Результаты для двумерного нормального распределения (??) при $p = 0.0$

Normal $n = 20, p = 0.0$			
	Pearson	Spearman	Quad
E	0.18892	0.14541	0.06000
E^2	0.05409	0.04186	0.02800
D	0.01840	0.02071	0.02440
Normal $n = 20, p = 0.5$			
	Pearson	Spearman	Quad
E	0.50363	0.52647	0.46000
E^2	-0.03769	-0.02852	-0.03800
D	0.00583	0.00535	0.01640
D	0.00411	0.00461	0.00762
Normal $n = 100, p = 0.5$			
	Pearson	Spearman	Quad
E	0.49628	0.47702	0.33200
E^2	0.25200	0.23489	0.12048
D	0.00570	0.00734	0.01026

Normal $n = 60, p = 0.0$			
	Pearson	Spearman	Quad
E	0.14612	0.05409	0.03333
E^2	0.01080	0.00965	0.00667
D	0.00865	0.00704	0.00556
Normal $n = 60, p = 0.5$			
	Pearson	Spearman	Quad
E	0.50847	0.47194	0.31333
E^2	0.26925	0.21059	0.10720
D	0.00064	0.00633	0.00404
D	0.00088	0.00073	0.00088
Normal $n = 1000, p = 0.5$			
	Pearson	Spearman	Quad
E	0.49458	0.47938	0.33320
E^2	0.24515	0.23052	0.11207
D	0.00054	0.00071	0.00105

Рис. 3: Графики двумерного нормального распределения и смеси для размера выборки $n = 100$

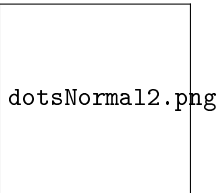


Таблица 3: Результаты для двумерного нормального распределения (??) при $p = 0.9$

Normal $n = 20, p = 0.9$			
	Pearson	Spearman	Quad
E	0.90154	0.85850	0.64000
E^2	0.81558	0.74275	0.44000
D	0.00281	0.00574	0.03040

Normal $n = 60, p = 0.9$			
	Pearson	Spearman	Quad
E	0.89761	0.88464	0.69333
E^2	0.80681	0.78457	0.48622
D	0.00112	0.00198	0.00551

Normal $n = 100, p = 0.9$			
	Pearson	Spearman	Quad
E	0.89624	0.88888	0.71600
E^2	0.80360	0.79094	0.51728
D	0.00037	0.00082	0.00462

Normal $n = 1000, p = 0.9$			
	Pearson	Spearman	Quad
E	0.89971	0.88953	0.71120
E^2	0.80951	0.79132	0.50603
D	0.00004	0.00005	0.00022

NormalMix $n = 20, p_1 = 0.9, p_2 = -0.9$			
	Pearson	Spearman	Quad
E	0.90154	0.85850	0.64000
E^2	0.12784	0.14056	0.18000
D	0.03603	0.04226	0.11240

NormalMix $n = 60, p_1 = -0.9, p_2 = -0.9$			
	Pearson	Spearman	Quad
E	0.34444	0.33502	0.24000
E^2	0.13330	0.12744	0.08711
D	0.01466	0.01521	0.02951

NormalMix $n = 100, p_1 = 0.9, p_2 = -0.9$			
	Pearson	Spearman	Quad
E	0.42503	0.39751	0.26400
E^2	0.18615	0.16254	0.07584
D	0.00550	0.00453	0.00614

NormalMix $n = 1000, p_1 = 0.9, p_2 = -0.9$			
	Pearson	Spearman	Quad
E	0.38948	0.37427	0.25080
E^2	0.15242	0.14103	0.06380
D	0.00073	0.00095	0.00090

7 Выводы

По таблицам ??, ??, ??, ??, видно, что, при увеличении объёма выборки, подсчитанные коэффициенты корреляции стремятся к теоретическим.

Ближе всех к данному коэффициенту корреляции находится коэффициент Пирсона.

По графикам видно, что при уменьшении корреляции эллипс равновероятности стремится к окружности, а при увеличении растягивается.

8 Список литературы

- [1] Модуль numpy - <https://physics.susu.ru/vorontsov/language/numpy.html>
- [2] Модуль matplotlib - <https://matplotlib.org/users/index.html>
- [3] Модуль scipy - <https://docs.scipy.org/doc/scipy/reference/>
- [4] Двумерное нормальное распределение: https://en.wikipedia.org/wiki/Multivariate_normal_distribution
- [5] Коэффициент корреляции Пирса: <http://statistica.ru/theory/koeffitsient-korrelyatsii/>
- [6] Коэффициент корреляции Спирмана: http://economic-definition.com/Exchange_Terminology/Koefficient_korrelyacii_Correlation_coefficient_eto.html

9 Приложения

Код отчёта: <https://github.com/MisterProper9000/MatStatLabs/blob/master/MatStatLab5/MatStatLab5.tex>

Код лабораторной: <https://github.com/MisterProper9000/MatStatLabs/blob/master/MatStatLab5/MatStatLab5.py>