Санкт-Петербургский Политехнический Университет имени Петра Великого
Институт Прикладной Математики и Механики
Кафедра "Прикладная Математика"

Отчет по лабораторным работам №5-8

по дисциплине

"Математическая Статистика"

Выполнил студент:
Тырыкин Я. А.
группа 5030102/80401
Проверил:
к.ф.-м.н., доцент
Баженов А. Н.

Санкт-Петербург
2021

Содержание

Список иллюстраций

Список таблиц

1 Постановка задачи

1.1 Лабораторная работа №5

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

Коэффициент корреляции ρ Коэффициент корреляции

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:

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

Изобразить сгенерированные точки на плоскости и нарисовать эллипс равновероятности.

2 Теория

2.1 Двумерное нормальное распределение

Двумерная случайная величина (X, Y) называется распределённой нормально (или просто нормальной), если её плотность вероятности определена формулой

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

Компоненты X, Y двумерной нормальной случайной величины также распределены нормально с математическими ожиданиями $\overline{x}, \overline{y}$ и средними квадратическими отклонениями σ_x, σ_y соответственно.

Параметр ρ называется коэффициентом корреляции.

2.2 Корреляционный момент (ковариация) и коэффициент корреляции

Корреляционный момент, иначе ковариация, двух случайных величин X и Y:

$$K = cov(X, Y) = M[(X - \overline{x})(Y - \overline{y})]. \tag{2}$$

Коэффициент корреляции ρ двух случайных величин X и Y:

$$\rho = \frac{K}{\sigma_x\sigma_y} \tag{3}$$

2.3 Выборочные коэффициенты корреляции

2.3.1 Выборочный коэффициент корреляции Пирсона

Выборочный коэффициент корреляции Пирсона:

$$r = \frac{\frac{1}{n}\sum(x_i - \overline{x})(y_i - \overline{y})}{\sqrt{\frac{1}{n^2}\sum(x_i - \overline{x})^2\sum(y_i - \overline{y})^2}} = \frac{K}{s_X s_Y} \tag{4}$$

где K, s_X^2, s_Y^2- выборочные коварияция и дисперсии случайных величин X и Y.

2.3.2 Выборочный квадрантный коэффициент корреляции

Выборочный квадрантный коэффициент корреляции

$$r_Q = \frac{(n_1 + n_3) - (n_2 + n_4)}{n} \tag{5}$$

где n_1, n_2, n_3 и 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 $(medx, medy)$.

2.3.3 Выборочный коэффициент ранговой корреляции Спирмена

Обозначим ранги, соотвествующие значениям переменной X, через u, а ранги, соотвествующие значениям переменной $Y-$ через v.

Выборочный коэффициент ранговой корреляции Спирмена:

$$r_S = \frac{\frac{1}{n}\sum(u_i - \overline{u})(v_i - \overline{v})}{\sqrt{\frac{1}{n^2}\sum(u_i - \overline{u})^2\sum(v_i - \overline{v})^2}} \tag{6}$$

где $\overline{u} = \overline{v} = \dfrac{1 + 2 + ... + n}{n} = \dfrac{n+1}{2}$ - среднее значение рангов.

2.4 Эллипсы рассеивания

Уравнение проекции эллипса рассеивания на плоскость xOy:

$$\frac{(x - \overline{x})^2}{\sigma_x^2} - 2\rho\frac{(x - \overline{x})(y - \overline{y})}{\sigma_x\sigma_y} + \frac{(y - \overline{y})^2}{\sigma_y^2} = const \tag{7}$$

Центр эллипса (7) находится в точке с координатами $(\overline{x}, \overline{y})$; оси симметрии эллипса составляют с осью Ox углы, определяемые уравнением

$$\text{tg } 2\alpha = \frac{2\rho\sigma_x\sigma_y}{\sigma_x^2 - \sigma_y^2} \tag{8}$$

3 Модульная структура программы

Лабораторная работа выполнена с применением средств языка Python версии 3.7 в среде разработки PyCharm IDE (в частности, с применением встроенных методов библиотеки SciPy и MatPlotLib). Исходной код лабораторной работы находится по ссылке в приложении к отчёту.

4 Результаты

4.1 Выборочные коэффициенты корреляции

$\rho = 0$ (3)	r (4)	r_s (6)	r_Q (5)
$E(z)$	-0.002	-0.005	0.02
$E(z^2)$	0.05	0.05	0.05
$D(z)$	0.05	0.06	0.05
$\rho = 0.5$	r	r_s	r_Q
$E(z)$	0.24	0.22	0.32
$E(z^2)$	0.16	0.15	0.17
$D(z)$	-0.08	-0.07	0.14
$\rho = 0.9$	r	r_s	r_Q
$E(z)$	0.46	0.43	0.7
$E(z^2)$	0.37	0.35	0.52
$D(z)$	-0.08	-0.08	0.03

Таблица 1: Двумерное нормальное распределение, $n = 20$

$\rho = 0$ (3)	r (4)	r_s (6)	r_Q (5)
$E(z)$	0.34	0.32	0.001
$E(z^2)$	0.28	0.27	0.02
$D(z)$	-0.05	-0.05	0.02
$\rho = 0.5$	r	r_s	r_Q
$E(z)$	0.37	0.35	0.33
$E(z^2)$	0.28	0.26	0.13
$D(z)$	-0.09	-0.09	0.02
$\rho = 0.9$	r	r_s	r_Q
$E(z)$	0.46	0.44	0.7
$E(z^2)$	0.36	0.34	0.5
$D(z)$	-0.09	-0.09	0.009

Таблица 2: Двумерное нормальное распределение, $n = 60$

$\rho = 0$ (3)	r (4)	r_s (6)	r_Q (5)
$E(z)$	0.39	0.37	0.006
$E(z^2)$	0.31	0.3	0.01
$D(z)$	-0.07	-0.07	0.01
$\rho = 0.5$	r	r_s	r_Q
$E(z)$	0.4	0.39	0.34
$E(z^2)$	0.3	0.29	0.12
$D(z)$	-0.1	-0.09	0.01
$\rho = 0.9$	r	r_s	r_Q
$E(z)$	0.46	0.44	0.7
$E(z^2)$	0.36	0.34	0.47
$D(z)$	-0.09	-0.09	0.01

Таблица 3: Двумерное нормальное распределение, $n = 100$

$\rho = 0$ (3)	r (4)	r_s (6)	r_Q (5)
$E(z)$	-0.5	0.43	0.51
$E(z^2)$	0.5	0.26	0.3
$D(z)$	0.3	0.06	0.04
$\rho = 0.5$	r	r_s	r_Q
$E(z)$	-0.65	0.47	0.56
$E(z^2)$	0.49	0.25	0.32
$D(z)$	0.05	-0.03	0.01
$\rho = 0.9$	r	r_s	r_Q
$E(z)$	-0.69	0.47	0.56
$E(z^2)$	0.51	0.24	0.323
$D(z)$	0.03	0.02	0.007

Таблица 4: Смесь нормальных распределений

4.2 Эллипсы рассеивания

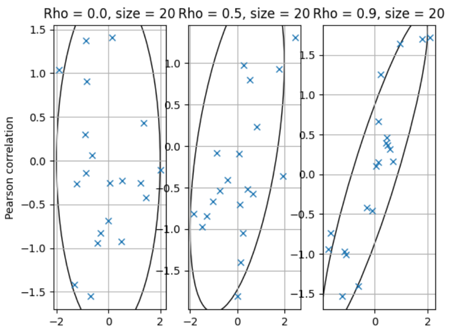

Рис. 1: Двумерное нормальное распределение, $n = 20$

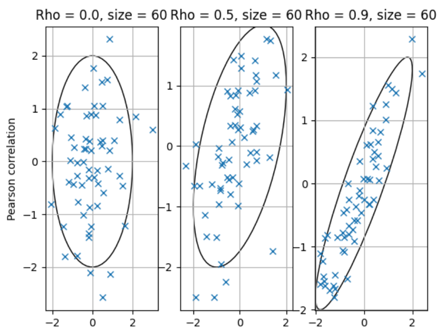

Рис. 2: Двумерное нормальное распределение, $n = 60$

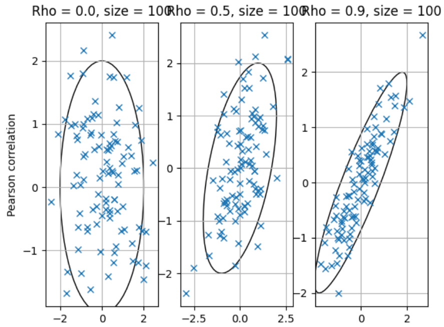

Рис. 3: Двумерное нормальное распределение, $n = 100$

5 Обсуждение

5.1 Выборочные коэффициенты корреляции и эллипсы рассеивания

Сравним дисперсии выборочных коэффициентов корреляции:

- Для двумерного нормального распределения дисперсии выборочных коэффициентов корреляции упорядочены следующим образом: $r_Q > r_s > r$.

- Для смеси нормальных распределений дисперсии выборочных коэффициентов корреляции упорядочены следующим образом: $r_Q < r_s < r$.

Процент попавших элементов выборки в эллипс рассеивания (95%-ная доверительная область) примерно равен его теоретическому значению (95%).

6 Ресурсы

Код программы, реализующей отрисовку обозначенных распределений:

https://github.com/YaroslavAggressive/Mathematical-statistics-lab-works