Санкт-Петербургский политехнический университет
Петра Великого

Институт прикладной математики и механики

Кафедра «Прикладная математика»

Отчет по лабораторной работе №5

по дисциплине
"Математическая статистика"

Выполнил студент
Группы 3630102/80101 Шао Цзяци

Проверил
доцент, к.ф.-м.н. Баженов Александр Николаевич

Санкт-Петербург
2021 г.

Содержание

Список иллюстраций

Список таблиц

1. Постановка задачи

Сгенерировать двухмерные выборки размерами 20, 60 и 100 для двухмерного 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 нее вычисляются: среднее значение, среднее значение квадрата, дисперсия коэффициентов корреляции Пирсона, Спирмена, и квадратного 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. Двухмерное нормальное распределение

Двумерная случайная величина (X, Y) называется нормальной, если её плотность вероятности определена формулой:

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

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

2.2. Корреляционный момент (ковариация) и коэффициент корреляции

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

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

Коэффициент корреляции ρ друх случайных величин X, Y:

$$\rho = \frac{K}{\sigma_x\sigma_y} \quad (3)$$

2.3. Выборочные коэффициенты корреляции

2.3.1. Выборочный коэффициент корреляции Пирсона

Выборочный коэффициент корреляции Пирсона:

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

где, $K, {s_X}^2, {s_Y}^2$ - выборочные ковориация и дисперсии случайных величин X, 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, n_4 - количество точек с координатами (x_i, y_i), попавшими соответственно в I, II, III, IV квадранты декартовой системы с осями $x' = x - medx$, $y' = y - medy$ с центром в точке с координатами $(medx, medy)$.

2.3.3. Выборочный коэффициент ранговой корреляции Спирмена

Обозначим ранги, соответствующие значениям переменной X, через u, а ранги, соответствующие значениям переменной Y, через V. *Выборочный коэффициент ранговой корреляции Спирмена*:

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

где, $\overline{u} = \overline{v} = \frac{1+2+...+n}{n} = \frac{n+1}{2}$ - 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

$$tan(2\alpha) = \frac{2\rho\sigma_x\sigma_y}{\sigma_x^2 - \sigma_y^2} \tag{8}$$

3. Реализация

Лабораторная работа выполнена с помощью встроенных средств языка программирования python в среде разработки Pycharm с дополнительными библиотеками.

- scipy
- numpy
- matplotlib
- math

Исходный код лабораторной работы размещен в Github-репозитории.
URL: `https://github.com/Shaots/shaoMathStatistic/tree/master/Lab5`

4. Результаты

4.1. Выборочные коэффициенты корреляции

$\rho = 0(3)$	r(4)	$r_S(6)$	$r_Q(5)$
E(z)	0.0016	0.0014	0.004
E(z^2)	0.0548	0.0547	0.0554
D(z)	0.0548	0.0547	0.0554
$\rho = 0.5$	r	r_S	r_Q
E(z)	0.4924	0.4657	0.3262
E(z^2)	0.2739	0.2513	0.1544
D(z)	0.0314	0.0344	0.048
$\rho = 0.9$	r	r_S	r_Q
E(z)	0.8936	0.8603	0.6918
E(z^2)	0.8012	0.7457	0.5103
D(z)	0.0027	0.0056	0.0317

Таблица 1. Двумерное нормальное распределение, n = 20

$\rho = 0(3)$	r(4)	$r_S(6)$	$r_Q(5)$
E(z)	0.0005	0.0004	-0.0029
E(z^2)	0.0172	0.0172	0.0166
D(z)	0.0172	0.0172	0.0166
$\rho = 0.5$	r	r_S	r_Q
E(z)	0.4982	0.4774	0.3353
E(z^2)	0.2578	0.2386	0.1281
D(z)	0.0096	0.0107	0.0157
$\rho = 0.9$	r	r_S	r_Q
E(z)	0.8988	0.8826	0.7077
E(z^2)	0.8085	0.7801	0.5091
D(z)	0.0007	0.0011	0.0083

Таблица 2. Двумерное нормальное распределение, n = 60

$\rho = 0(3)$	r(4)	$r_S(6)$	$r_Q(5)$
E(z)	-0.0053	-0.0068	-0.004
E(z^2)	0.0102	0.0104	0.0102
D(z)	0.0102	0.0104	0.0102
$\rho = 0.5$	r	r_S	r_Q
E(z)	0.4978	0.4784	0.3321
E(z^2)	0.2535	0.2352	0.1192
D(z)	0.0057	0.0063	0.0089
$\rho = 0.9$	r	r_S	r_Q
E(z)	0.8993	0.8867	0.7093
E(z^2)	0.8091	0.7868	0.5082
D(z)	0.0004	0.0006	0.0051

Таблица 3. Двумерное нормальное распределение, n = 100

n = 20	r	r_S	r_Q
E(z)	0.7835	0.7506	0.1142
E(z^2)	0.6231	0.5758	0.0144
D(z)	0.0092	0.0124	0.0014
n = 10	r	r_S	r_Q
E(z)	0.7907	0.7705	0.3486
E(z^2)	0.6277	0.5973	0.1258
D(z)	0.0025	0.0036	0.0043
n = 100	r	r_S	r_Q
E(z)	0.7874	0.7686	0.5721
E(z^2)	0.6216	0.5928	0.3338
D(z)	0.0016	0.0021	0.0065

Таблица 4. Смесь нормальных распределений

4.2. Эллипсы рассеивания

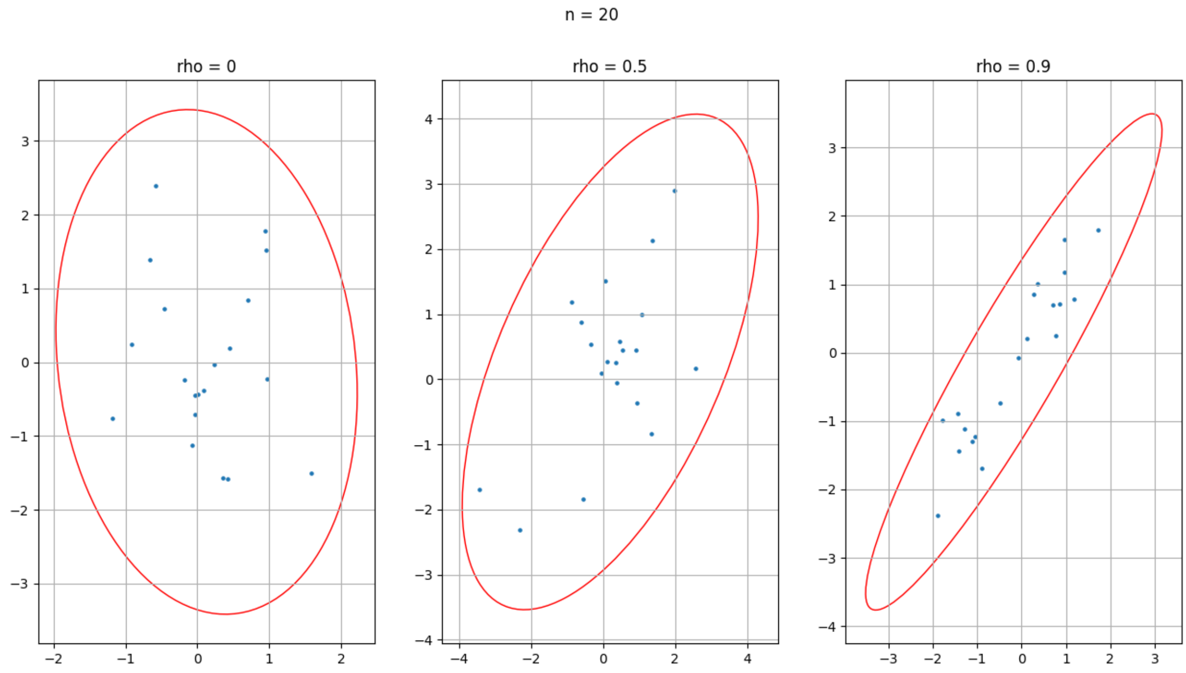

Рис. 1. Двумерное нормальное распределение, n = 20

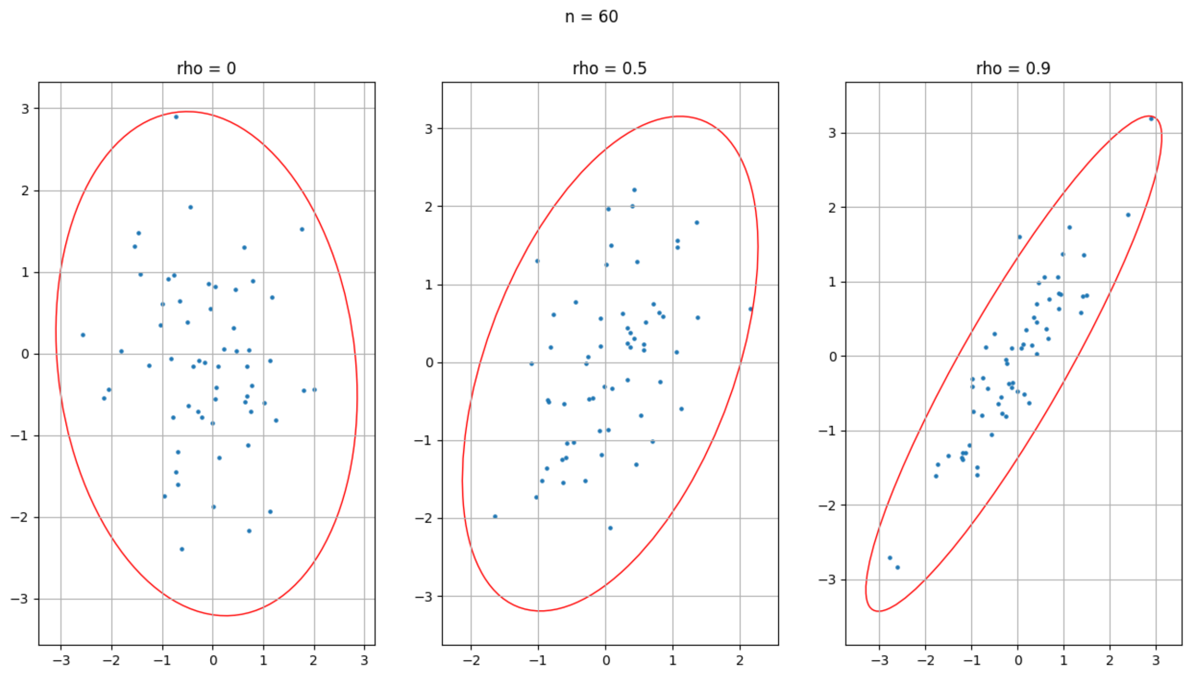

Рис. 2. Двумерное нормальное распределение, n = 60

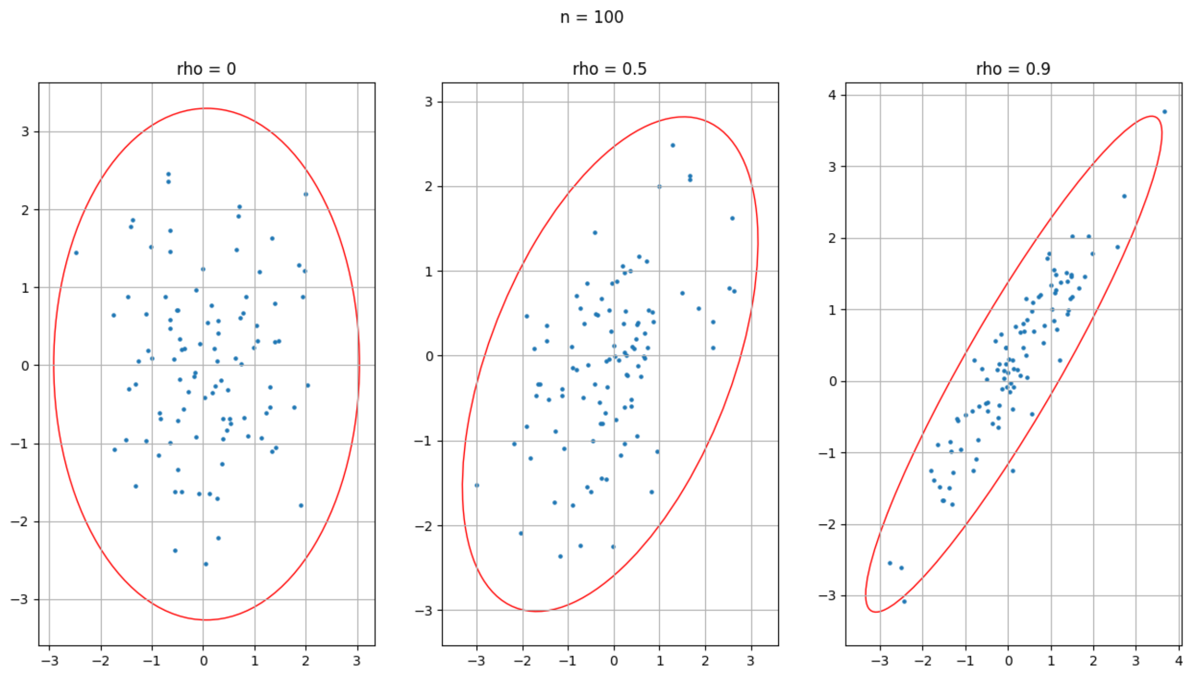

Рис. 3. Двумерное нормальное распределение, n = 100

5. Обсуждение

1) Из таблиц характеристик распределений разных размерностей и смеси распределений, вытекает что значения $E(z)$, $E(z^2)$ для величин r, r_S, r_Q в большинстве случаев подчиняются соотношению $r > r_S > r_Q$. А для дисперсии наблюдается обратное: $r < r_S < r_Q$.

2) В уравнении эллипса взял правую часть 9. На графиках эллипсов рассеивания можно наблюдать, что почти все элементы выборки попадают внутри эллипса, при этом достаточно большая часть значений концентрируется в центре эллипса, что подтверждает вывод о теоретическом значении центра эллипса, координаты которого представляются в виде среднего для исследуемых величин X, Y.