Санкт-Петербургский Политехнический университет

Петра Великого

Институт прикладной математики и механики

Высшая школа прикладной математики и вычислительной физики

Отчёт

по лабораторной работе 5

по дисциплине

"математическая статистика"

Выполнил студент:

Аникин Александр Алексеевич,

группа 3630102\80201

Проверил:

к.ф.-м.н., доцент

Баженов Александр Николаевич

Санкт-Петербург

2021

Содержание

Список иллюстраций

Список таблиц

1 Постановка задачи

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

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

Изобразить сгенерированные точки на плоскости и нарисовать эллипс рассеивания.

2 Теория

2.1 Двумерное нормальное распределение

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

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

Компоненты X, Y двумерной нормальной случайной величины также распределены нормально с математическими ожиданиями $\bar{x}, \bar{y}$ и средними квадратическими отклонениями σ_x, σ_y соответственно ([1], с. 133-134). Параметр ρ называется коэффициентом корреляции:

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

где K - корреляционный момент (иначе ковариация) случайных величин X и Y:

$$K = cov(X, Y) = M[(X - \bar{x})(Y - \bar{y})] \tag{4}$$

2.2 Выборочные коэффициенты корреляции

2.2.1 Коэффициент корреляции Пирсона

Выборочный коэффициент корреляции Пирсона:

$$r_P = \frac{\sum_{i=0}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=0}^{n}(x_i - \bar{x})^2 \sum_{i=0}^{n}(y_i - \bar{y})^2}} = \frac{K}{S_X S_Y}, \tag{5}$$

где S_X и S_Y - дисперсии случайных величин X и Y

2.2.2 Коэффициент корреляции Спирмена

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

$$r_S = \frac{\sum_{i=0}^{n}(u_i - \bar{u})(v_i - \bar{v})}{\sqrt{\sum_{i=0}^{n}(u_i - \bar{u})^2 \sum_{i=0}^{n}(v_i - \bar{v})^2}}, \tag{6}$$

где $\bar{u} = \bar{v} = \frac{1+2+...+n}{n} = \frac{n+1}{2}$ - среднее значение рангов ([1], с. 540-541).

2.2.3 Квадрантный коэффициент корреляции

Выборочный квадрантный коэффициент корреляции:

$$r_Q = \frac{(n_1 + n_3) - (n_2 + n_4)}{n}, \tag{7}$$

где $n1, n2, n3, n4$ — количества точек с координатами (x_i, y_i), попавшими соответственно в I, II, III, IV квадранты декартовой системы с осями $x' = x - med(x)$, $y' = y - med(y)$ и с центром в точке $(med(x), med(y))$ ([1], с. 539).

2.3 Эллипс рассеивания

Уравнение проекции эллипса рассеивания на плоскость xOy:

$$\frac{(x - \bar{x}^2)}{\sigma_x} - 2\rho\frac{(x - \bar{x})(y - \bar{y})}{\sigma_x\sigma_y} + \frac{(y - \bar{y})^2}{\sigma_y^2} = c, \qquad c = const \tag{8}$$

Принцип выбора константы c:

Выразим из уравнения y:

$$y_{1,2} = \sigma_y\left(\rho\frac{x - \bar{x}}{\sigma_x} \pm \sqrt{\frac{(x - \bar{x})^2}{\sigma_x^2}(\rho^2 - 1) + c}\right) + \bar{y} \tag{9}$$

Заметим, что

$$\frac{(x - \bar{x})^2}{\sigma_x^2}(\rho^2 - 1) + c \geq 0 \tag{10}$$

поэтому

$$c = \max(-\frac{(x_i - \bar{x})^2}{\sigma_x^2}(\rho^2 - 1)), \qquad i = 1, ..., n \tag{11}$$

3 Реализация

Лабораторная работа выполнена на языке Python 3.8 с помощью загружаемых пакетов SciPy, MatPlotLib, NumPy. Исходный код лабораторной работы находится на GitHub репозитории.

4 Результаты

4.1 Выборочные коэффициенты корреляции

$\rho = 0$	r_P[5]	r_S[6]	r_Q[7]
$E(z)$	0.003	0.005	0.002
$E(z^2)$	0.053	0.055	0.050
$D(z)$	0.053	0.055	0.050
$\rho = 0.5$	r_P	r_S	r_Q
$E(z)$	0.485	0.457	0.319
$E(z^2)$	0.267	0.246	0.145
$D(z)$	0.031	0.0376	0.043
$\rho = 0.9$	r_P	r_S	r_Q
$E(z)$	0.895	0.865	0.717
$E(z^2)$	0.803	0.752	0.538
$D(z)$	0.002	0.005	0.025

Таблица 1: Нормальное двумерное распределение [2], 20 элементов

$\rho = 0$	r_P	r_S	r_Q
$E(z)$	-0.002	-0.001	0.001
$E(z^2)$	0.016	0.017	0.017
$D(z)$	0.016	0.017	0.017
$\rho = 0.5$	r_P	r_S	r_Q
$E(z)$	0.500	0.478	0.339
$E(z^2)$	0.259	0.239	0.129
$D(z)$	0.009	0.011	0.014
$\rho = 0.9$	r_P	r_S	r_Q
$E(z)$	0.900	0.884	0.713
$E(z^2)$	0.811	0.783	0.517
$D(z)$	0.001	0.001	0.008

Таблица 2: Нормальное двумерное распределение, 60 элементов

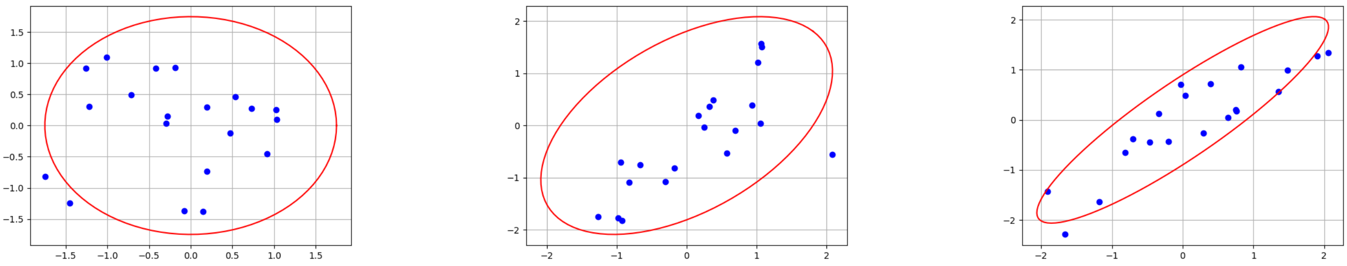
$\rho = 0$	r_P	r_S	r_Q
$E(z)$	0.001	-0.001	-0.002
$E(z^2)$	0.010	0.010	0.009
$D(z)$	0.010	0.010	0.009
$\rho = 0.5$	r_P	r_S	r_Q
$E(z)$	0.500	0.481	0.336
$E(z^2)$	0.255	0.237	0.122
$D(z)$	0.006	0.006	0.009
$\rho = 0.9$	r_P	r_S	r_Q
$E(z)$	0.899	0.885	0.709
$E(z^2)$	0.808	0.785	0.508
$D(z)$	0.0002	0.001	0.005

Таблица 3: Нормальное двумерное распределение, 100 элементов

$n = 20$	r_P	r_S	r_Q
$E(z)$	-0.099	-0.091	-0.056
$E(z^2)$	0.057	0.055	0.050
$D(z)$	0.048	0.047	0.047
$n = 60$	r_P	r_S	r_Q
$E(z)$	-0.101	-0.096	-0.066
$E(z^2)$	0.027	0.026	0.021
$D(z)$	0.017	0.016	0.017
$n = 100$	r_P	r_S	r_Q
$E(z)$	-0.099	-0.094	-0.065
$E(z^2)$	0.019	0.019	0.014
$D(z)$	0.010	0.010	0.010

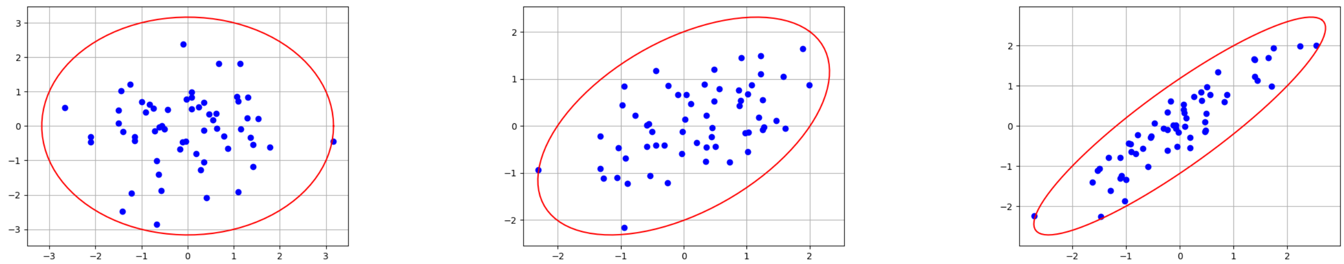
Таблица 4: Смешанное нормальное двумерное распределение[1]

4.2 Эллипсы рассеивания



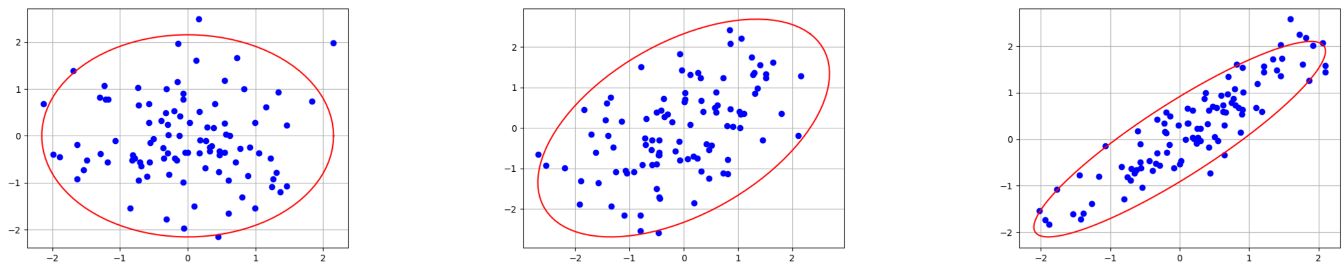
$\rho = 0$ $\rho = 0.5$ $\rho = 0.9$

Рис. 1: Эллипсы рассеивания [8] для выборки из 20 элементов



$\rho = 0$ $\rho = 0.5$ $\rho = 0.9$

Рис. 2: Эллипсы рассеивания для выборки из 60 элементов



$\rho = 0$ $\rho = 0.5$ $\rho = 0.9$

Рис. 3: Эллипсы рассеивания для выборки из 100 элементов

Список литературы

[1] Вероятностные разделы математики. Учебник для бакалавров технических направлений.//Под ред. Максимова Ю.Д. — Спб.: «Иван Федоров», 2001. — 592 с., илл.