Санкт-Петербургский политехнический университет
Петра Великого

Институт прикладной математики и механики
Кафедра «Прикладная математика»

Отчёт
по лабораторной работе №5
по дисциплине
«Математическая статистика»

Выполнил студент:

Кондратьев Д. А.
группа: 3630102/70301

Проверил:

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

Санкт-Петербург
2020 г.

Содержание

Список таблиц

Список иллюстраций

1. Постановка задачи

Сгенерировать двумерные выборки размерами 20, 60, 100 для нормального двумерного 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:

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

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:

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

Компоненты X, Y двумерной нормальной случайной величины также распределены нормально с математическими ожиданиями $\overline{x}, \overline{y}$ и средними квадратическими отклонениями σ_x, σ_y соответственно [1, с. 133-134].

Параметр ρ называется коэффициентом корреляции.

2.2. Корреляционный момент(ковариация) и коэффициент корреляции

Корреляционным моментом, иначе *ковариацией*, двух случайных величин X и Y называется математическое ожидание произведения отклонений этих случайных величин от их математических ожиданий [1, с. 141].

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

Коэффициентом корреляции ρ двух случайных величин X и Y называется отношение их корреляционного момента к произведению их средних квадратических отклонений:

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

Коэффициент корреляции — это нормированная числовая характеристика, являющаяся мерой близости зависимости между случайными величинами к линейной [1, с. 150].

2.3. Выборочные коэффициенты корреляции

2.3.1. Выборочный коэффициент корреляции Пирсона

Пусть по выборке значений $\{x_i, y_i\}_1^n$ двумерной с.в. (X, Y) требуется оценить коэффициент корреляции $\rho = \frac{cov(X,Y)}{\sqrt{DXDY}}$. Естественной оценкой для ρ служит его статистический аналог в виде выборочного коэффициента корреляции, предложенного К.Пирсоном, —

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

где K, s_X^2, s_Y^2 — выборочные ковариации и дисперсии с.в. X и Y [1, с. 535].

2.3.2. Выборочный квадрантный коэффициент корреляции

Кроме выборочного коэффициента корреляции Пирсона, существуют и другие оценки степени взаимосвязи между случайными величинами. К ним относится *выборочный квадрантный коэффициент корреляции*:

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

где $n_1, n_2, n_3 n_4$ — количества точке с координатами (x_i, y_i), попавшими соответственно в I, II, III и IV квадранты декартовой системы с осями $x' = x - med\, x, y' = y - med\, y$ и с центром в точке с координатами $(med\, x, med\, y)$ [1, с. 539].

2.3.3. Выборочный коэффициент ранговой корреляции Спирмена

На практике нередко требуется оценить степень взаимодействия между качественными признаками изучаемого объекта. Качественным называется признак, который нельзя измерить точно, но который позволяет сравнивать изучаемые объекты между собой и располагать их в порядке убывания или возрастания их качества. Для этого объекты выстраиваются в определённом порядке в соответствии с рассматриваемым признаком. Процесс упорядочения называется *ранжированием*, и каждому члену упорядоченной последовательности объектов присваивается ранг, или порядковый номер. Например, объекту с наименьшим значением признака присваивается ранг 1, следующему за ним объекту — ранг 2, и т.д. Таким образом, происходит сравнение каждого объекта со всеми объектами изучаемой выборки.

Если объект обладает не одним, а двумя качественными признаками — переменными X и Y , то для исследования их взаимосвязи используют выборочный коэффициент корреляции между двумя последовательностями рангов этих признаков.

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

Выборочный коэффициент ранговой корреляции Спирмена определяется как выборочный коэффициент корреляции Пирсона между рангами u, v переменных X, Y :

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

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

3. Реализация

Лабораторная работа выполнена на программном языке *Python 3.8* в среде разработки *Jupyter Notebook 6.0.3*. В работе использовались следующие пакеты языка *Python*:

- *numpy* — для генерации выборки и работы с массивами;
- *matplotlib.pyplot* и *seaborn* — для построения эллипсов рассеивания;
- *scipy.stats* — содержит все необходимые распределения.

Ссылка на исходный код лабораторной работы приведена в приложении.

4. Результаты

4.1. Выборочные коэффициенты корреляции

$\rho = 0.0$ (3)	r (4)	r_S (6)	r_Q (5)
$E(z)$	0.01	0.00	0.00
$E(z^2)$	0.05	0.05	0.05
$D(z)$	0.0531	0.0520	0.0515

$\rho = 0.5$	r	r_S	r_Q
$E(z)$	0.49	0.46	0.32
$E(z^2)$	0.27	0.25	0.15
$D(z)$	0.0307	0.0348	0.0480

$\rho = 0.9$	r	r_S	r_Q
$E(z)$	0.896	0.866	0.70
$E(z^2)$	0.805	0.755	0.51
$D(z)$	0.0022	0.0048	0.0291

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

$\rho = 0.0$	r	r_S	r_Q
$E(z)$	0.00	0.00	-0.00
$E(z^2)$	0.02	0.02	0.02
$D(z)$	0.0172	0.0175	0.0171

$\rho = 0.5$	r	r_S	r_Q
$E(z)$	0.494	0.47	0.32
$E(z^2)$	0.254	0.23	0.12
$D(z)$	0.0097	0.0109	0.0147

$\rho = 0.9$	r	r_S	r_Q
$E(z)$	0.8988	0.883	0.706
$E(z^2)$	0.8086	0.781	0.507
$D(z)$	0.0007	0.0013	0.0089

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

$\rho = 0.0$	r	r_S	r_Q
$E(z)$	-0.001	0.000	-0.00
$E(z^2)$	0.010	0.010	0.01
$D(z)$	0.0098	0.0098	0.0105

$\rho = 0.5$	r	r_S	r_Q
$E(z)$	0.496	0.476	0.333
$E(z^2)$	0.251	0.233	0.120
$D(z)$	0.0056	0.0064	0.0089

$\rho = 0.9$	r	r_S	r_Q
$E(z)$	0.8995	0.8868	0.712
$E(z^2)$	0.8094	0.7871	0.511
$D(z)$	0.0004	0.0006	0.0051

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

$n = 20$	r	r_S	r_Q
$E(z)$	0.783	0.75	0.56
$E(z^2)$	0.622	0.57	0.35
$D(z)$	0.0085	0.0128	0.0396

$n = 60$	r	r_S	r_Q
$E(z)$	0.791	0.768	0.57
$E(z^2)$	0.628	0.594	0.34
$D(z)$	0.0024	0.0033	0.0111

$n = 100$	r	r_S	r_Q
$E(z)$	0.789	0.771	0.575
$E(z^2)$	0.624	0.596	0.337
$D(z)$	0.0015	0.0019	0.0063

Таблица 4. Смесь нормальных распределений

4.2. Эллипсы рассеивания

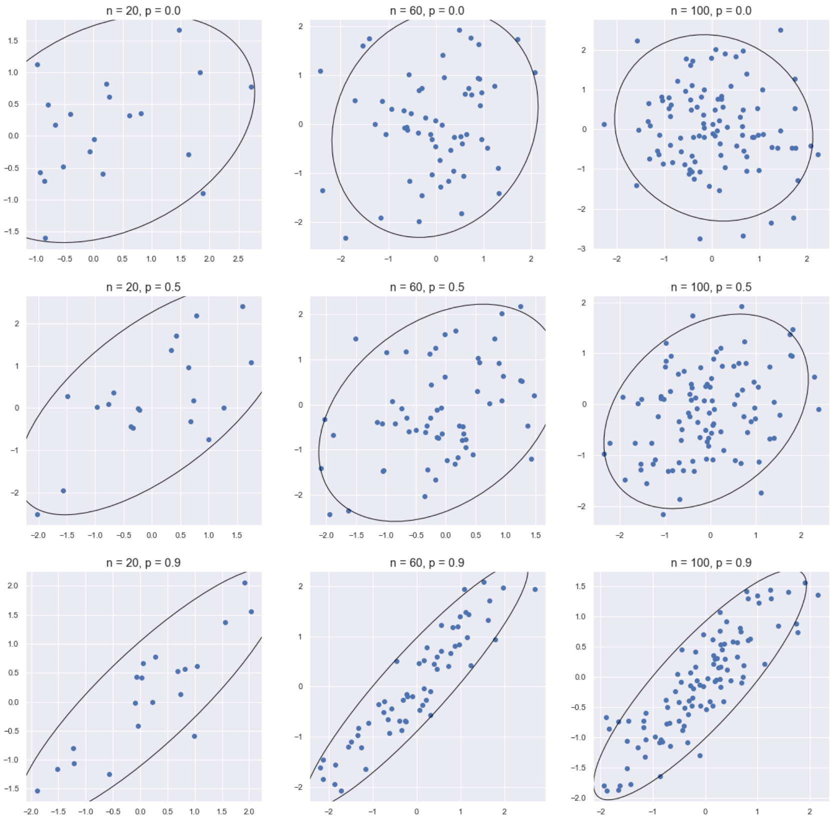

Рис. 1. Двумерное нормальное распределение

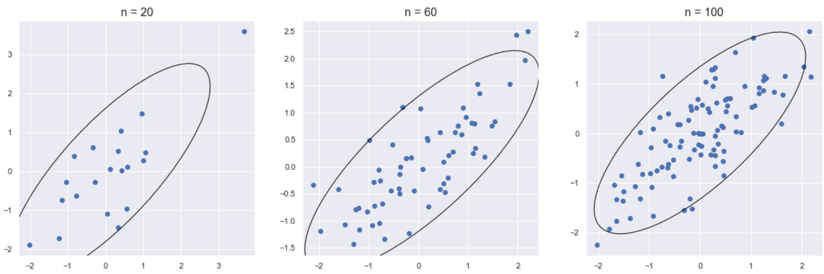

Рис. 2. Смесь нормальных распределений

5. Обсуждение

Исходя из полученных результатов можно сделать следующие выводы:

- Верны следующие соотношения для дисперсий выборочных коэффициентов корреляции:
  - для двумерного нормального распределения: $r < r_S < r_Q$;
  - для смеси нормальных распределений: $r < r_S < r_Q$.
- Процент попавших элементов выборки в эллипс рассеивания (95%-ная доверительная область) примерно равен его теоретическому значению (95%).
- При уменьшении корреляции эллипс равновероятности стремится к окружности, а при увеличении — растягивается.

6. Литература

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

2) Correlation and dependence. URL: `https://en.wikipedia.org/wiki/Correlation_and_dependence`

7. Приложение

1) Код лабораторной. URL: `https://github.com/DmitriiKondratev/MatStat/blob/master/Lab_5/Lab_5.ipynb`

2) Код отчёта. URL: `https://github.com/DmitriiKondratev/MatStat/blob/master/Lab_5/Lab_report_5.tex`