Санкт-Петербургский Политехнический
Университет
Петра Великого

Институт прикладной математики и механики

Кафедра "Прикладная математика"

Отчёт
Лабораторная работа №5
по дисциплине
"Математическая статистика"

Выполнил студент:
Салихов С.Р.
группа: 3630102/70401

Проверил:
к.ф-м.н., доцент
Баженов Александр Николавич

Санкт-Петербург
2020 г.

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, ρ).
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

$$N(x,y,\overline{x},\overline{y},\delta_x,\delta_y,\rho) = \frac{1}{2\pi\delta_x\delta_y\sqrt{1-\rho^2}} * exp\Big\{\ -\tfrac{1}{2(1-\rho^2)}\big[\tfrac{(x-\overline{x})^2}{\delta_x^2} - 2\rho\tfrac{(x-\overline{x})(y-\overline{y})}{\delta_x\delta_y} + \tfrac{(y-\overline{y})^2}{\delta_y^2}\big]\ \Big\}$$

Компоненты X, Y двумерной нормальной случайной величины также распределены нормально с математическими ожиданиями x, y и средними квадратическими отклонениями δ_x, δ_y соответственно.
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.

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

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

$$\rho = \frac{K}{\delta_x\delta_y}$$

Коэффициент корреляции — это нормированная числовая характеристика, являющаяся мерой близости зависимости между случайными величинами к линейной.

2.3 Выборочные коэффициенты корреляции

2.3.1 Выборочный коэффициент корреляции Пирсона

Пусть по выборке значений

$\{x_i, y_i\}_1^n$ 1 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, —

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

где K, s_X^2, s_Y^2 - выборочные ковариации и дисперсии с.в. X и Y.

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

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

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 $x' = x - medx, y' = y - medy$ и с центром в точке с координатами (med x, med y).

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

$$r = \frac{\frac{1}{n}\sum(u_i - \overline{u})(\upsilon_i - \overline{\upsilon})}{\sqrt{\frac{1}{n}\sum(u_i - \overline{u})^2 \frac{1}{n}\sum(\upsilon_i - \overline{\upsilon})^2)}}$$

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

3 Реализация

Для генерации выборки был использован *Python* 3.7 и модуль numpy. Для отрисовки графиков использовался модуль matplotlib. scipy.stats для обработки функций распределений.

4 Результаты

4.1 Выборочные коэффициенты корреляции

Таблица 1: Результаты для двумерного нормального распределения при n = 20

$\rho = 0$	r	r_S	r_Q
E(z)	-0.005	-0.004	-0.002
E(z^2)	0.057	0.057	0.056
D(z)	0.057	0.057	0.056
$\rho = 0.5$	r	r_S	r_Q
E(z)	0.481	0.455	0.315
E(z^2)	0.263	0.242	0.148
D(z)	0.031	0.034	0.048
$\rho = 0.9$	r	r_S	r_Q
E(z)	0.897	0.866	0.699
E(z^2)	0.808	0.754	0.516
D(z)	0.002	0.004	0.027

Таблица 2: Результаты для двумерного нормального распределения при n = 60

$\rho = 0$	r	r_S	r_Q
E(z)	-0.002	-0.002	-0.001
E(z^2)	0.016	0.017	0.017
D(z)	0.016	0.017	0.017
$\rho = 0.5$	r	r_S	r_Q
E(z)	0.499	0.477	0.328
E(z^2)	0.258	0.238	0.122
D(z)	0.009	0.010	0.014
$\rho = 0.9$	r	r_S	r_Q
E(z)	0.898	0.881	0.704
E(z^2)	0.807	0.779	0.504
D(z)	0.0007	0.0011	0.008

Таблица 3: Результаты для двумерного нормального распределения при n = 100

$\rho = 0$	r	r_S	r_Q
E(z)	-0.004	-0.004	-0.002
E(z^2)	0.009	0.009	0.1
D(z)	0.009	0.009	0.010
$\rho = 0.5$	r	r_S	r_Q
E(z)	0.496	0.477	0.330
E(z^2)	0.252	0.234	0.117
D(z)	0.006	0.006	0.009
$\rho = 0.9$	r	r_S	r_Q
E(z)	0.899	0.887	0.710
E(z^2)	0.809	0.787	0.510
D(z)	0.0004	0.0006	0.005

Таблица 4: Смесь нормальных распределений

n = 20	r	r_S	r_Q
E(z)	0.384	0.364	0.251
E(z^2)	0.187	0.174	0.113
D(z)	0.039	0.042	0.051
n = 60	r	r_S	r_Q
E(z)	0.397	0.378	0.260
E(z^2)	0.170	0.156	0.084
D(z)	0.012	0.013	0.016
n = 100	r	r_S	r_Q
E(z)	0.397	0.380	0.259
E(z^2)	0.165	0.152	0.077
D(z)	0.007	0.008	0.009

4.2 Эллипсы рассеивания

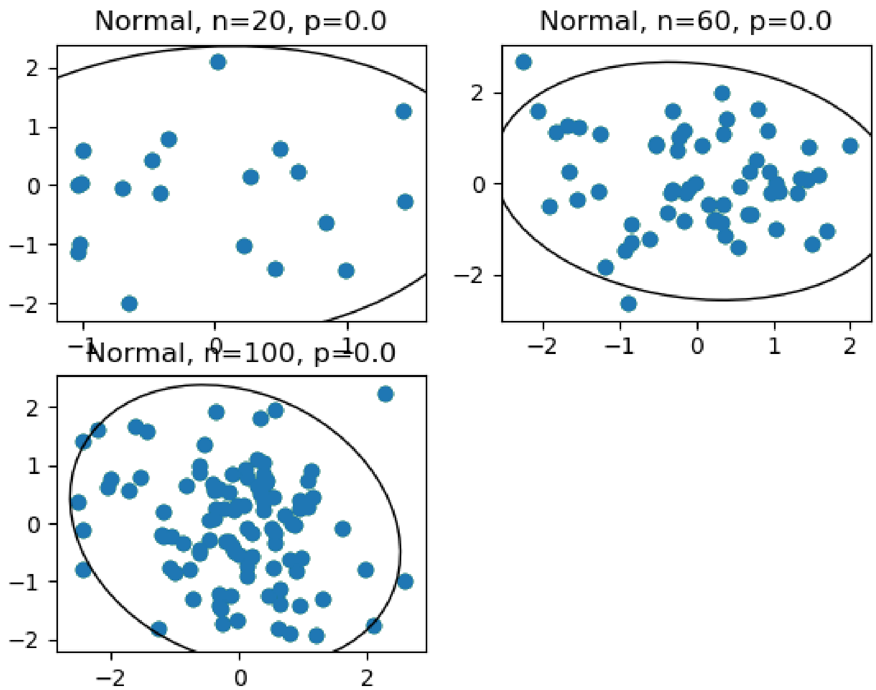

Рис. 1: Двумерное нормальное распределение, при $\rho = 0$

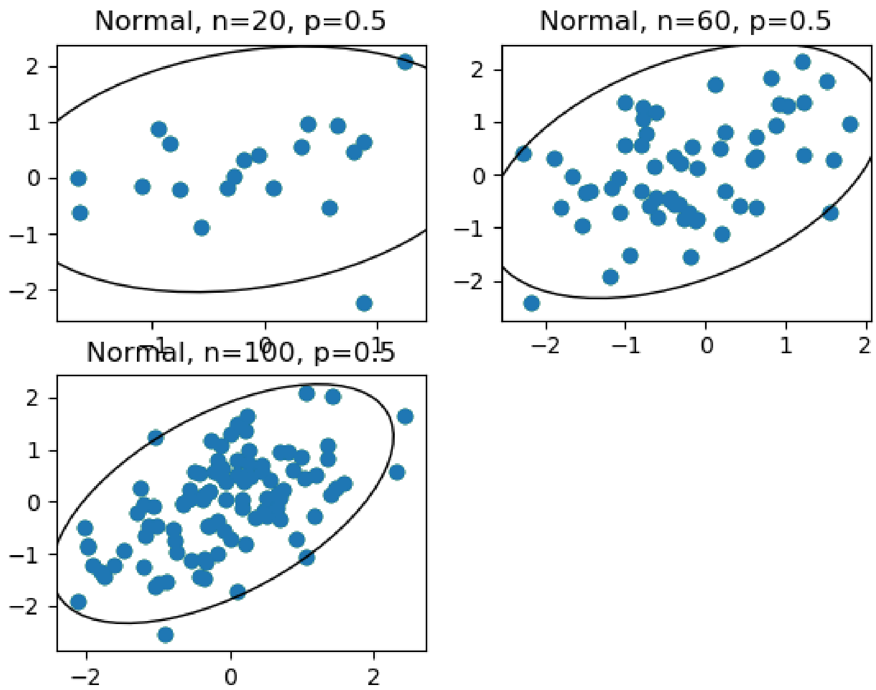

Рис. 2: Двумерное нормальное распределение, при $\rho = 0.5$

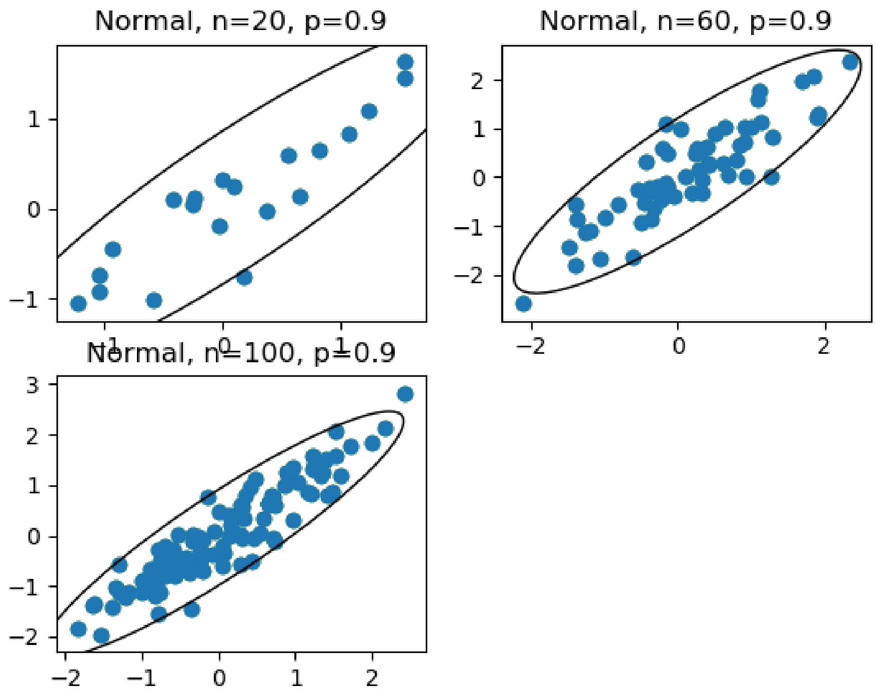

Рис. 3: Двумерное нормальное распределение, при $\rho = 0.9$

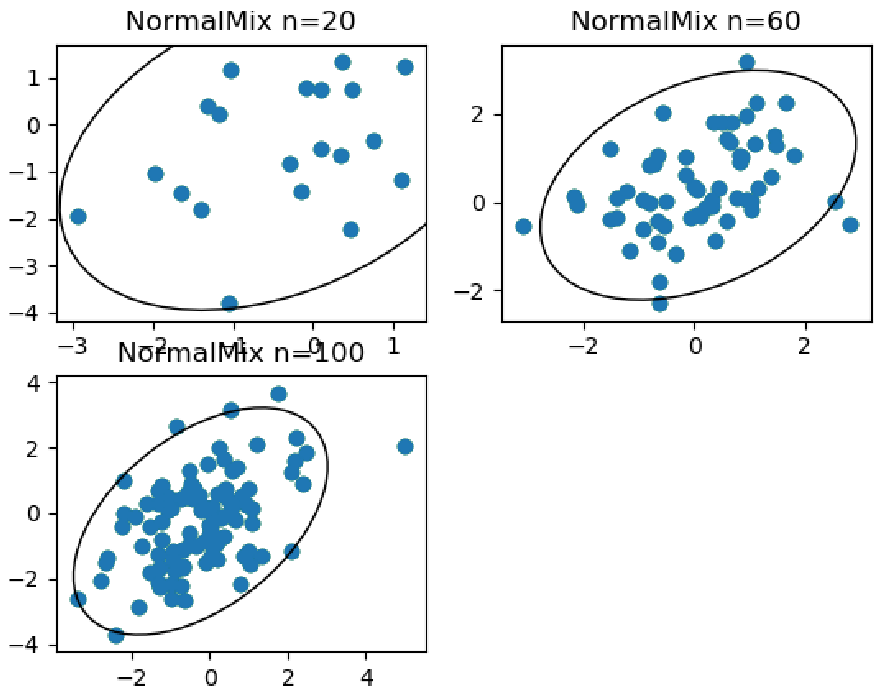

Рис. 4: Смесь нормальных распределений

5 Обсуждение

5.1 Выборочные коэффициенты корреляции и эллипсы рассеивания

1)Сравним дисперсии выборочных коэффициентов корреляции.
– Для двумерного нормального распределения дисперсии выборочных коэффициентов корреляции упорядочены следующим образом: r $< r_S < r_Q$.
– Для смеси нормальных распределений дисперсии выборочных коэффициентов корреляции упорядочены следующим образом: $r_Q < r_S <$ r.

2)Процент попавших элементов выборки в эллипс рассеивания (95%-я доверительная область) примерно равен его теоретическому значению (95%). При малом количестве событий $<\approx 12$ Процент попавших элементов в эллипс рассеивания примерно равен 100%.

5.2 Сравнение с теоретическими данными

По таблицам видно, что при увеличении объёма выборки, подсчитанные коэффициенты стремятся к теоретическим.
По графикам видно, что при уменьшении корреляции эллипс равновероятности стремится к окружности, а при увеличении растягивается.

6 Литература

Модуль numpy

Модуль matplotlib

Модуль scipy

Многомерное нормальное распределение
Корреляция

7 Приложения

Код лаборатрной
Код отчёта