

Санкт-Петербургский Политехнический Университет
им. Петра Великого

Институт прикладной математики и механики
Кафедра прикладной математики

Отчёт по лабораторной работе №4 по дисциплине “Математическая
статистика”

**Выборочные коэффициенты корреляции и эллипсы
рассеивания**

Выполнил студент:

Мишутин Д. В.

Группа:

3630102/70301

Проверил:

К.ф.-м.н., доцент

Баженов Александр Николаевич

Санкт-Петербург

2020 г.

Оглавление

| | |
|--------------------------|----|
| 1 Постановка задачи..... | 4 |
| 2 Теория..... | 4 |
| 3 Реализация..... | 4 |
| 4 Результаты..... | 4 |
| 5 Выводы..... | 15 |
| 6 Литература..... | 16 |
| 7 Приложения..... | 16 |

Список иллюстраций и таблиц

[Таблица 1 Двумерное стандартное нормальное распределение, \$n=20\$, \$r=0\$](#)

[Таблица 2 Двумерное стандартное нормальное распределение, \$n=60\$, \$r=0\$](#)

[Таблица 3 Двумерное стандартное нормальное распределение, \$n=100\$, \$r=0\$](#)

[Таблица 4 Двумерное стандартное нормальное распределение, \$n=20\$, \$r=0.5\$](#)

[Таблица 5 Двумерное стандартное нормальное распределение, \$n=60\$, \$r=0.5\$](#)

[Таблица 6 Двумерное стандартное нормальное распределение, \$n=100\$, \$r=0.5\$](#)

[Таблица 7 Двумерное стандартное нормальное распределение, \$n=20\$, \$r=0.9\$](#)

[Таблица 8 Двумерное стандартное нормальное распределение, \$n=60\$, \$r=0.9\$](#)

[Таблица 9 Двумерное стандартное нормальное распределение, \$n=100\$, \$r=0.9\$](#)

[Таблица 10 Смесь распределений, \$n=20\$](#)

[Таблица 11 Смесь распределений, \$n=60\$](#)

[Таблица 12 Смесь распределений, \$n=100\$](#)

[Рис. 1 Двумерное стандартное нормальное распределение для \$n=20\$, \$r=0\$](#)

[Рис. 2 Двумерное стандартное нормальное распределение для \$n=60\$, \$r=0\$](#)

[Рис. 3 Двумерное стандартное нормальное распределение для \$n=100\$, \$r=0\$](#)

[Рис. 4 Двумерное стандартное нормальное распределение для \$n=20\$, \$r=0.5\$](#)

[Рис. 5 Двумерное стандартное нормальное распределение для \$n=60\$, \$r=0.5\$](#)

[Рис. 6 Двумерное стандартное нормальное распределение для \$n=100\$, \$r=0.5\$](#)

[Рис. 7 Двумерное стандартное нормальное распределение для \$n=20\$, \$r=0.9\$](#)

[Рис. 8 Двумерное стандартное нормальное распределение для \$n=60\$, \$r=0.9\$](#)

[Рис. 9 Двумерное стандартное нормальное распределение для \$n=100\$, \$r=0.9\$](#)

[Рис. 10 Смесь распределений для \$n=20\$](#)

[Рис. 11 Смесь распределений для \$n=60\$](#)

[Рис. 12 Смесь распределений для \$n=100\$](#)

[Рис. 13 Эллипс рассеивания для 2-х точек при \$r=0\$](#)

[Рис. 14 Эллипс рассеивания для 2-х точек при \$r=0.5\$](#)

[Рис. 15 Эллипс рассеивания для 2-х точек при \$r=0.9\$](#)

[Рис. 16 Эллипс рассеивания для 3-х точек при \$r=0\$](#)

[Рис. 17 Эллипс рассеивания для 3-х точек при \$r=0.5\$](#)

[Рис. 18 Эллипс рассеивания для 3-х точек при \$r=0.9\$](#)

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:

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

Изобразить сгенерированные точки на плоскости и нарисовать эллипс рассеяния.

2 Теория

1. Двумерное стандартное нормальное распределение:

$$N(x, y, 0, 0, 1, 1, \rho) = \frac{1}{2\pi\sqrt{1-\rho^2}} e^{\frac{-1}{2(1-\rho^2)}(x^2 - 2\rho xy + y^2)}$$

2. Коэффициент корреляции Пирсона:

$$r_{xy} = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2 \sum_{i=1}^n (y_i - \bar{y})^2}}$$

3. Коэффициент корреляции Спирмена:

$$\rho_n = 1 - \frac{6}{n^3 - n} \sum_{i=1}^n d_i^2$$

4. Квадрантный коэффициент корреляции:

$$r_Q = \frac{1}{n} \sum_{i=1}^n \text{sign}(x_i - \text{med}_x) \text{sign}(y_i - \text{med}_y)$$

3 Реализация

Был использован язык *Python 3.8.2*: модуль *numpy* для вычисления описательных статистик, модуль *scipy* для генерации выборок на основе двумерного нормального распределения и расчёта коэффициентов корреляции, модуль *matplotlib* для построения и отображения графиков, модуль *pandas* для хранения статистических данных в таблицах и функция *display* из модуля *IPython.display* для их корректного отображения.

4 Результаты

Таблица 13 Двумерное стандартное нормальное распределение, $n=20$, $r=0$

| n=20 | Pearson | Spearman | quadrant |
|--------|---------|----------|----------|
| E(z) | 0.012 | 0.011 | 0.003 |
| E(z^2) | 0.048 | 0.05 | 0.051 |
| D(z) | 0.048 | 0.049 | 0.051 |

Таблица 14 Двумерное стандартное нормальное распределение, $n=60$, $r=0$

| n=60 | Pearson | Spearman | quadrant |
|-------------------------|----------------|-----------------|-----------------|
| E(z) | 0.004 | 0.004 | 0.003 |
| E(z²) | 0.016 | 0.017 | 0.017 |
| D(z) | 0.016 | 0.017 | 0.017 |

Таблица 15 Двумерное стандартное нормальное распределение, $n=100$, $r=0$

| n=100 | Pearson | Spearman | quadrant |
|-------------------------|----------------|-----------------|-----------------|
| E(z) | 0.001 | 0.001 | -0.001 |
| E(z²) | 0.01 | 0.01 | 0.011 |
| D(z) | 0.01 | 0.01 | 0.011 |

Таблица 16 Двумерное стандартное нормальное распределение, $n=20$, $r=0.5$

| n=20 | Pearson | Spearman | quadrant |
|-------------------------|----------------|-----------------|-----------------|
| E(z) | 0.489 | 0.46 | 0.322 |
| E(z²) | 0.271 | 0.246 | 0.148 |
| D(z) | 0.032 | 0.035 | 0.044 |

Таблица 17 Двумерное стандартное нормальное распределение, $n=60$, $r=0.5$

| n=60 | Pearson | Spearman | quadrant |
|-------------------------|----------------|-----------------|-----------------|
| E(z) | 0.496 | 0.476 | 0.331 |
| E(z²) | 0.255 | 0.237 | 0.124 |
| D(z) | 0.009 | 0.01 | 0.014 |

Таблица 18 Двумерное стандартное нормальное распределение, $n=100$, $r=0.5$

| n=100 | Pearson | Spearman | quadrant |
|-------------------------|----------------|-----------------|-----------------|
| E(z) | 0.497 | 0.477 | 0.331 |
| E(z²) | 0.253 | 0.233 | 0.118 |
| D(z) | 0.005 | 0.006 | 0.009 |

Таблица 19 Двумерное стандартное нормальное распределение, $n=20$, $r=0.9$

| n=20 | Pearson | Spearman | quadrant |
|-------------------------|----------------|-----------------|-----------------|
| E(z) | 0.896 | 0.867 | 0.696 |
| E(z²) | 0.806 | 0.756 | 0.513 |
| D(z) | 0.002 | 0.004 | 0.029 |

Таблица 20 Двумерное стандартное нормальное распределение, $n=60$, $r=0.9$

| n=60 | Pearson | Spearman | quadrant |
|-------------------------|----------------|-----------------|-----------------|
| E(z) | 0.898 | 0.883 | 0.707 |
| E(z²) | 0.808 | 0.78 | 0.508 |
| D(z) | 0.001 | 0.001 | 0.009 |

Таблица 21 Двумерное стандартное нормальное распределение, $n=100$, $r=0.9$

| n=100 | Pearson | Spearman | quadrant |
|---------------|----------------|-----------------|-----------------|
| E(z) | 0.899 | 0.886 | 0.708 |
| E(z^2) | 0.809 | 0.786 | 0.507 |
| D(z) | 0 | 0.001 | 0.005 |

Таблица 22 Смесь распределений, $n=20$

| n=20 | Pearson | Spearman | quadrant |
|---------------|----------------|-----------------|-----------------|
| E(z) | -0.08 | -0.078 | -0.05 |
| E(z^2) | 0.061 | 0.061 | 0.056 |
| D(z) | 0.054 | 0.055 | 0.054 |

Таблица 23 Смесь распределений, $n=60$

| n=60 | Pearson | Spearman | quadrant |
|---------------|----------------|-----------------|-----------------|
| E(z) | -0.092 | -0.086 | -0.06 |
| E(z^2) | 0.025 | 0.024 | 0.021 |
| D(z) | 0.016 | 0.016 | 0.017 |

Таблица 24 Смесь распределений, $n=100$

| n=100 | Pearson | Spearman | quadrant |
|---------------|----------------|-----------------|-----------------|
| E(z) | -0.097 | -0.092 | -0.063 |
| E(z^2) | 0.019 | 0.019 | 0.014 |
| D(z) | 0.01 | 0.01 | 0.01 |

Рис. 19 Двумерное стандартное нормальное распределение для $n=20$, $r=0$

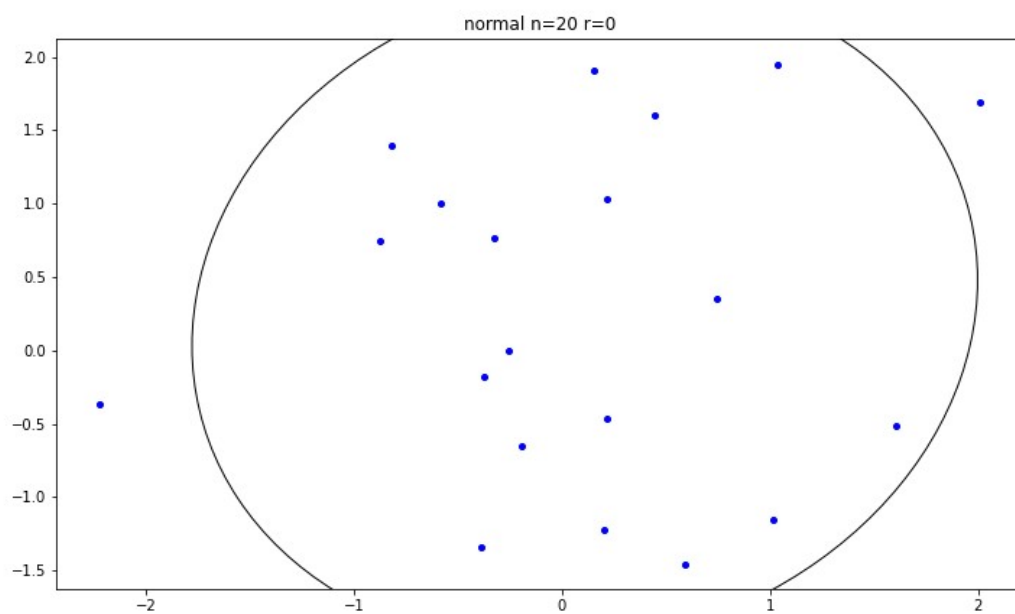


Рис. 20 Двумерное стандартное нормальное распределение для $n=60$, $r=0$

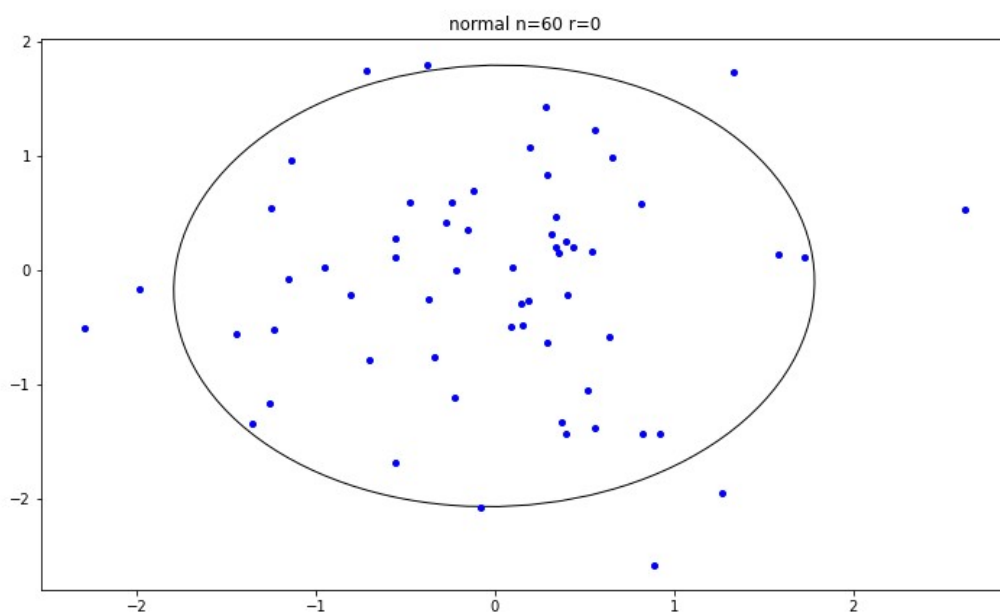


Рис. 21 Двумерное стандартное нормальное распределение для $n=100$, $r=0$

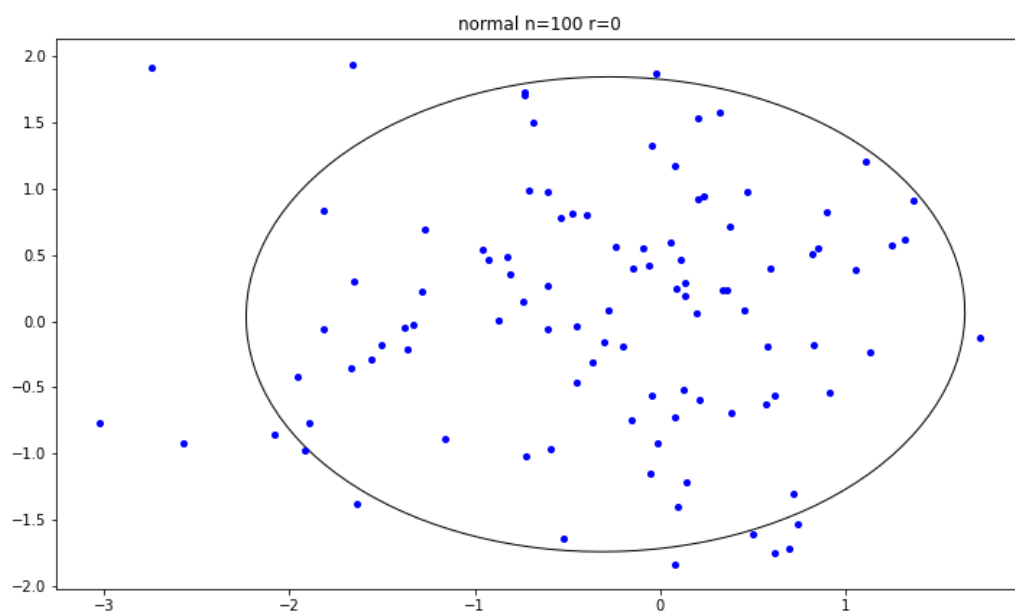


Рис. 22 Двумерное стандартное нормальное распределение для $n=20$, $r=0.5$

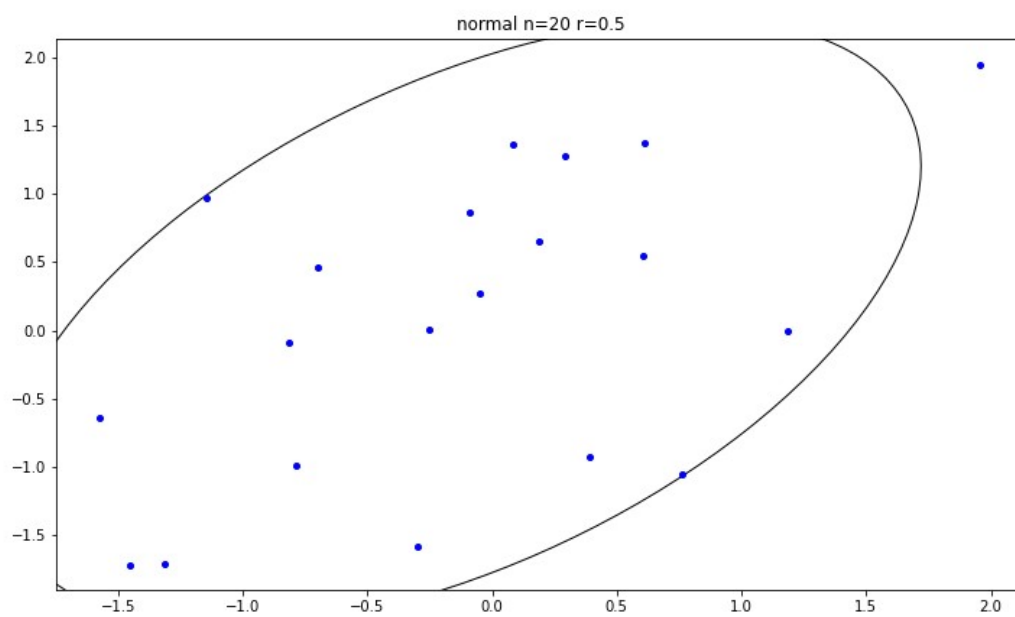


Рис. 23 Двумерное стандартное нормальное распределение для $n=60$, $r=0.5$

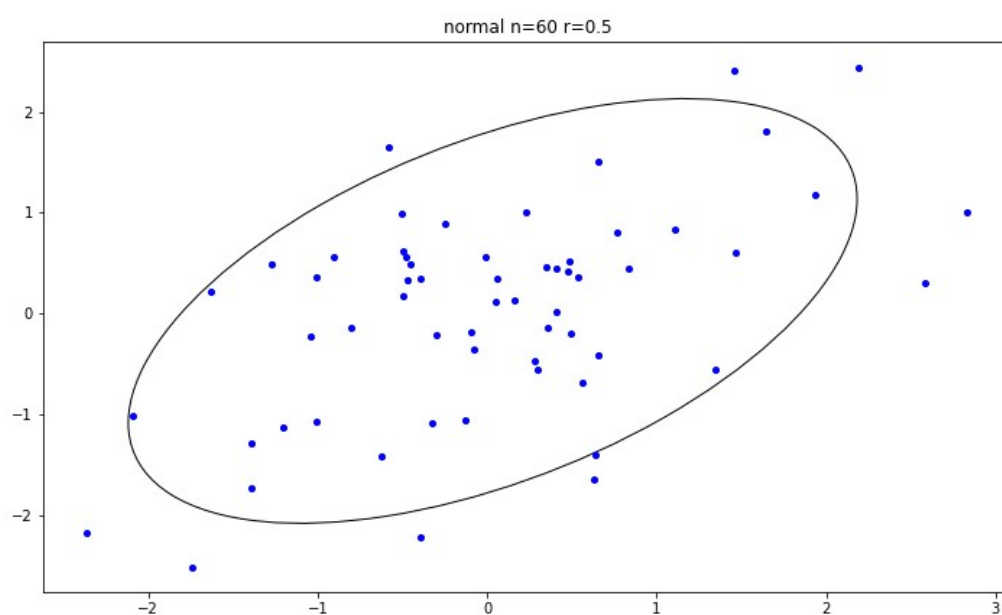


Рис. 24 Двумерное стандартное нормальное распределение для $n=100$, $r=0.5$

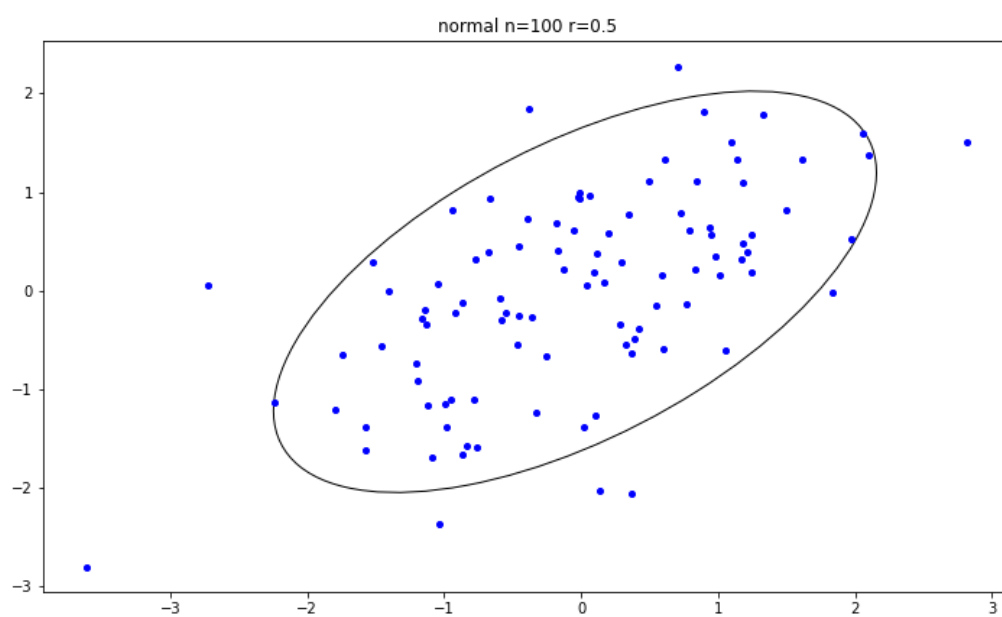


Рис. 25 Двумерное стандартное нормальное распределение для $n=20$, $r=0.9$

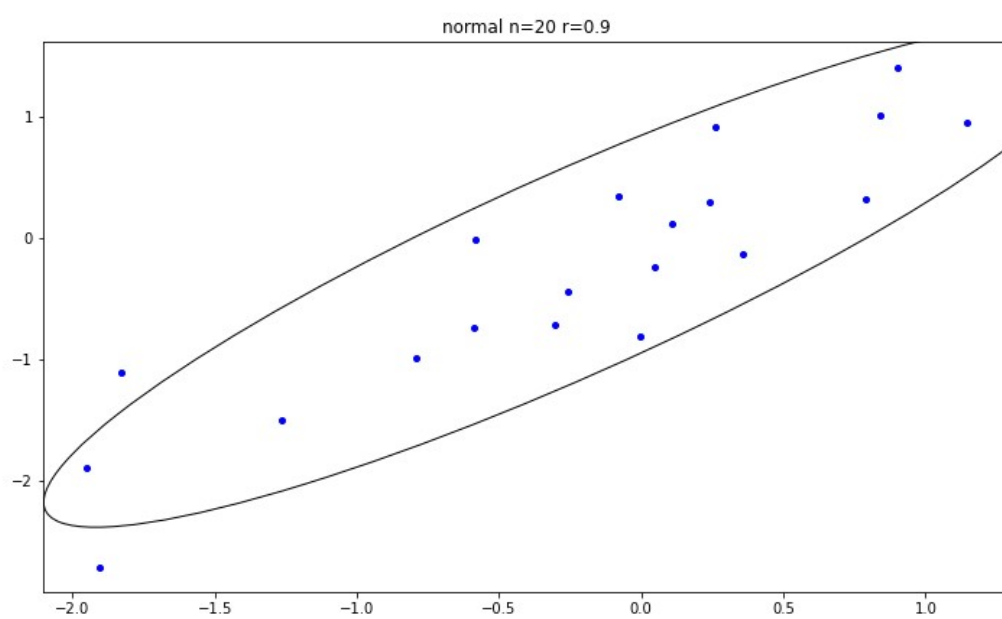


Рис. 26 Двумерное стандартное нормальное распределение для $n=60$, $r=0.9$

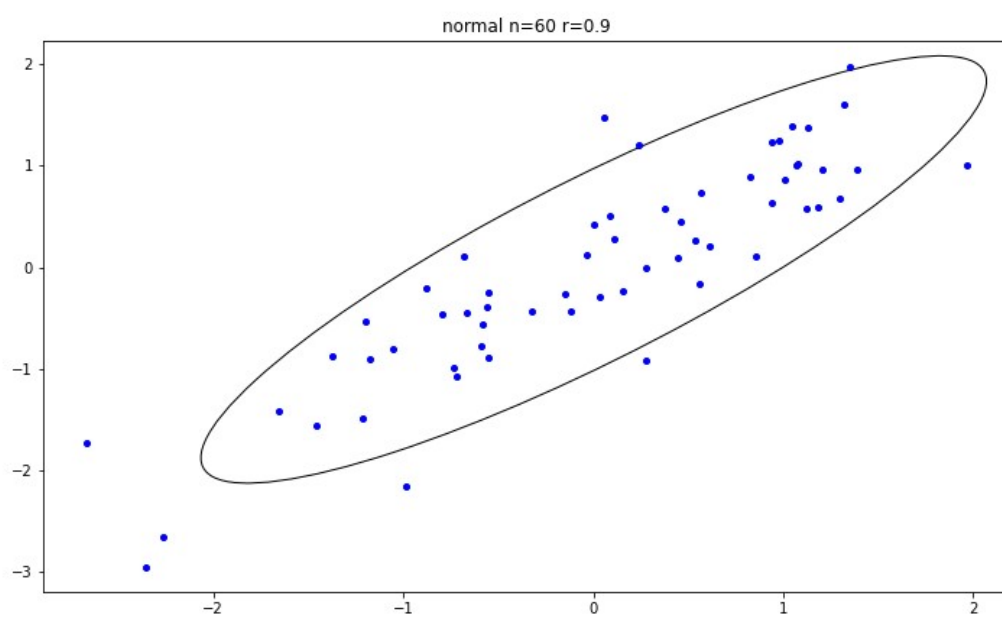


Рис. 27 Двумерное стандартное нормальное распределение для $n=100$, $r=0.9$

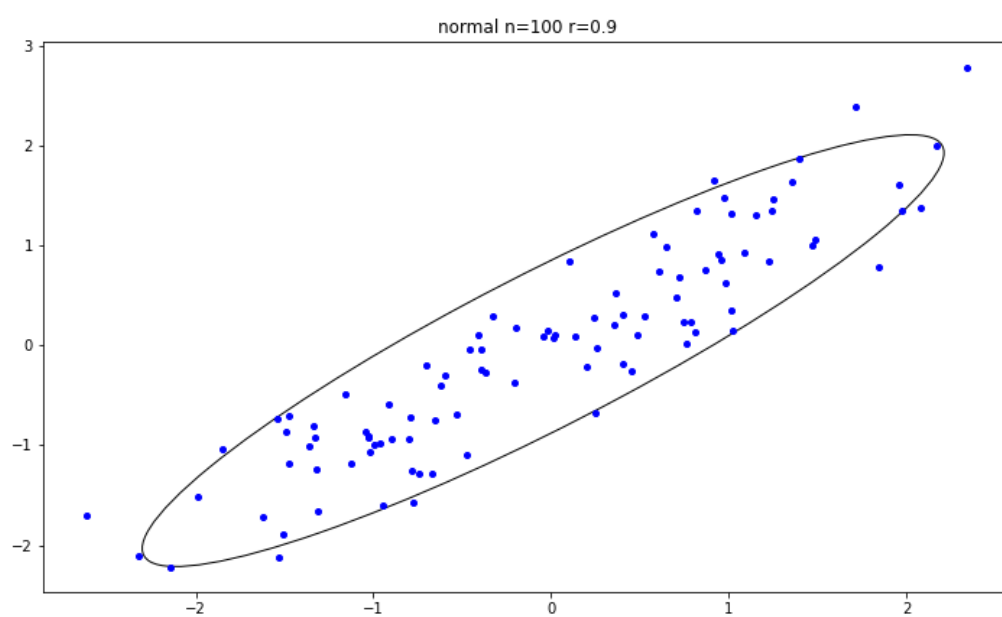


Рис. 28 Смесь распределений для $n=20$

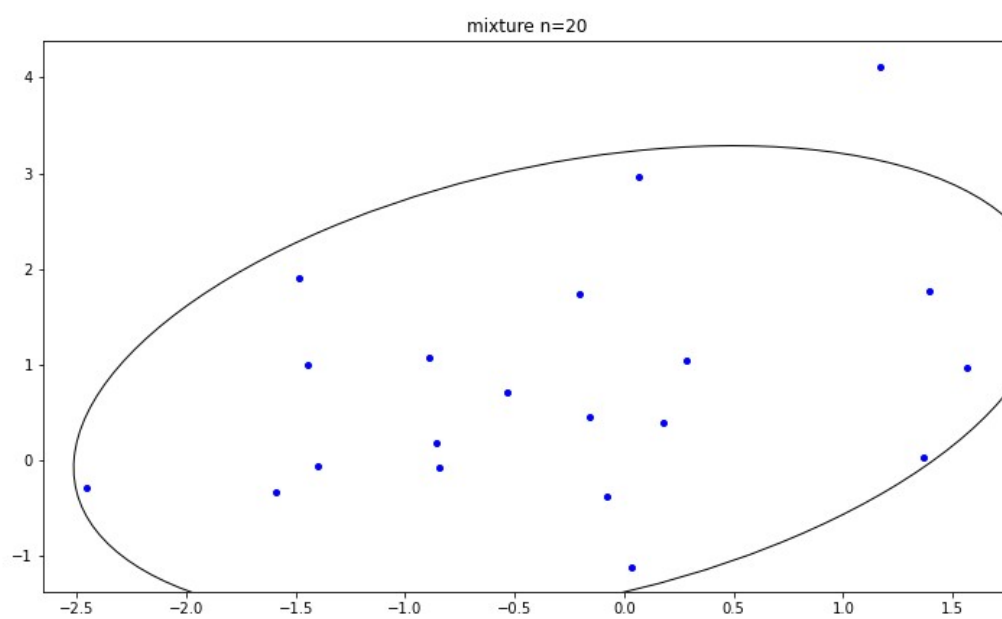


Рис. 29 Смесь распределений для $n=60$

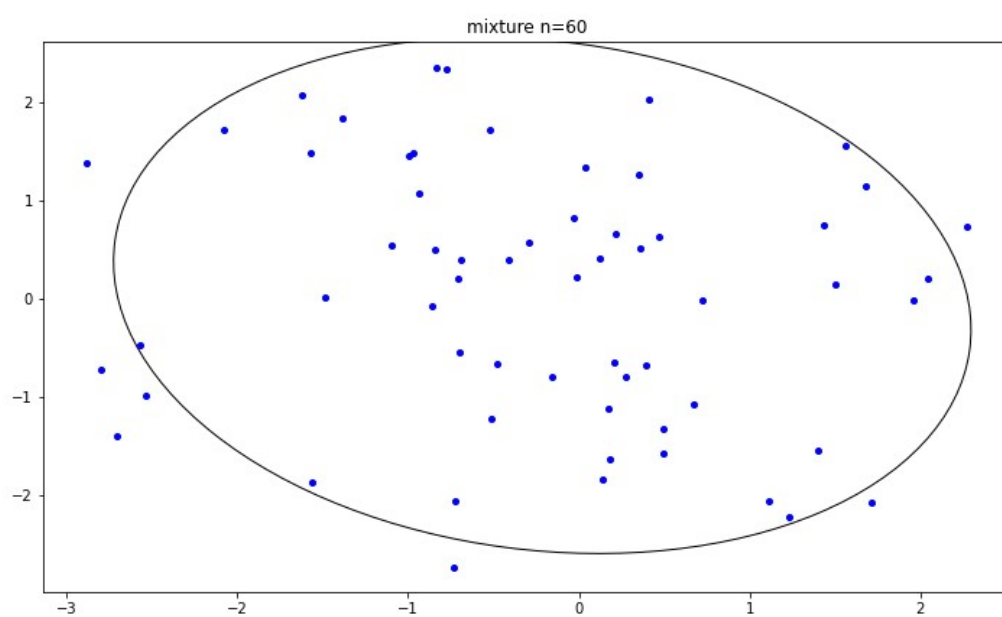


Рис. 30 Смесь распределений для $n=100$

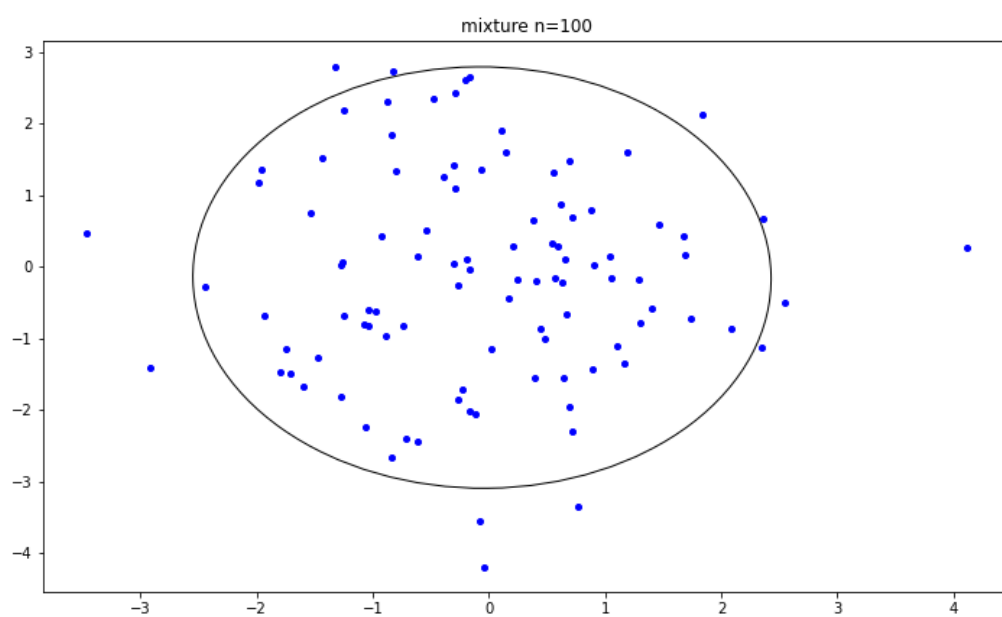


Рис. 31 Эллипс рассеивания для 2-х точек при $r=0$

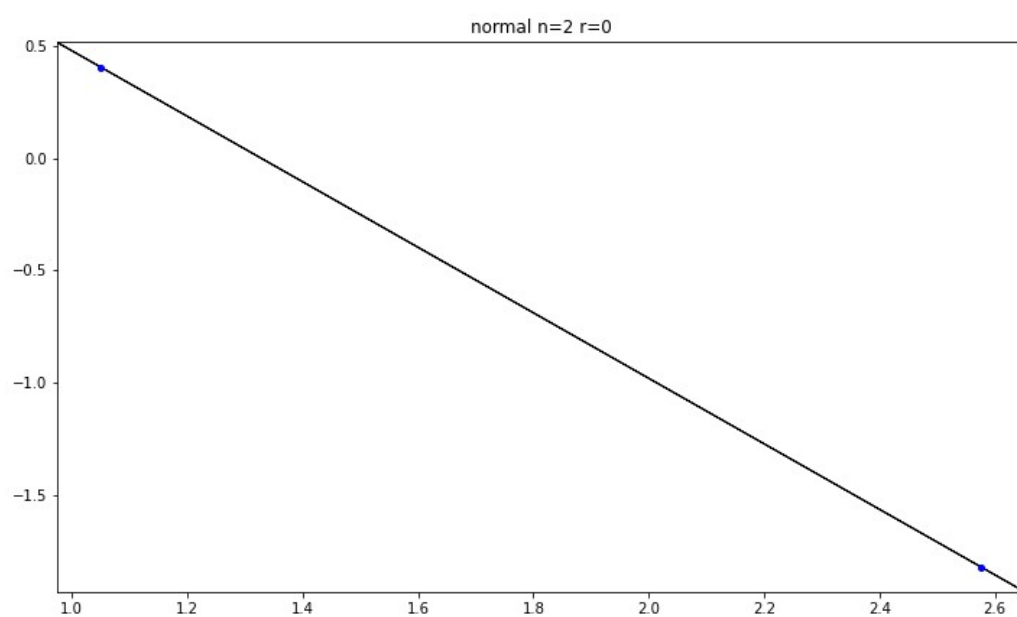


Рис. 32 Эллипс рассеивания для 2-х точек при $r=0.5$

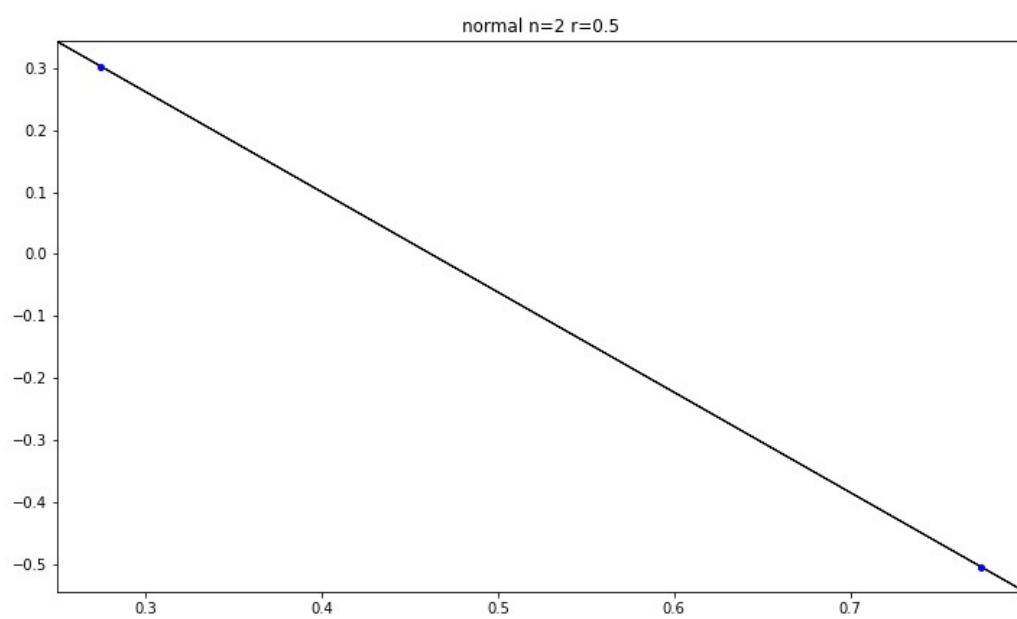


Рис. 33 Эллипс рассеивания для 2-х точек при $r=0.9$

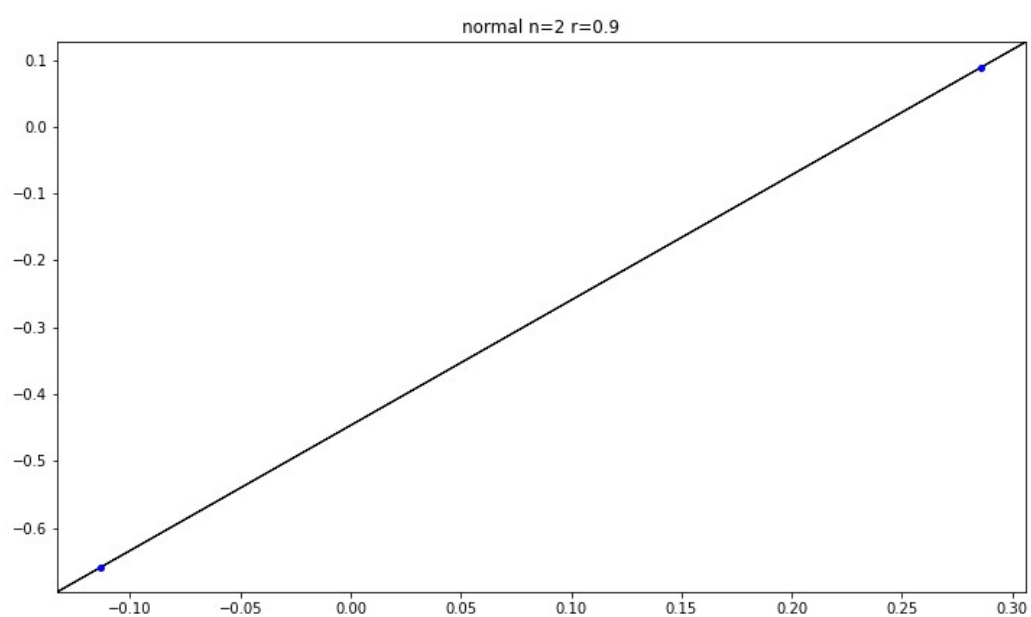


Рис. 34 Эллипс рассеивания для 3-х точек при $r=0$

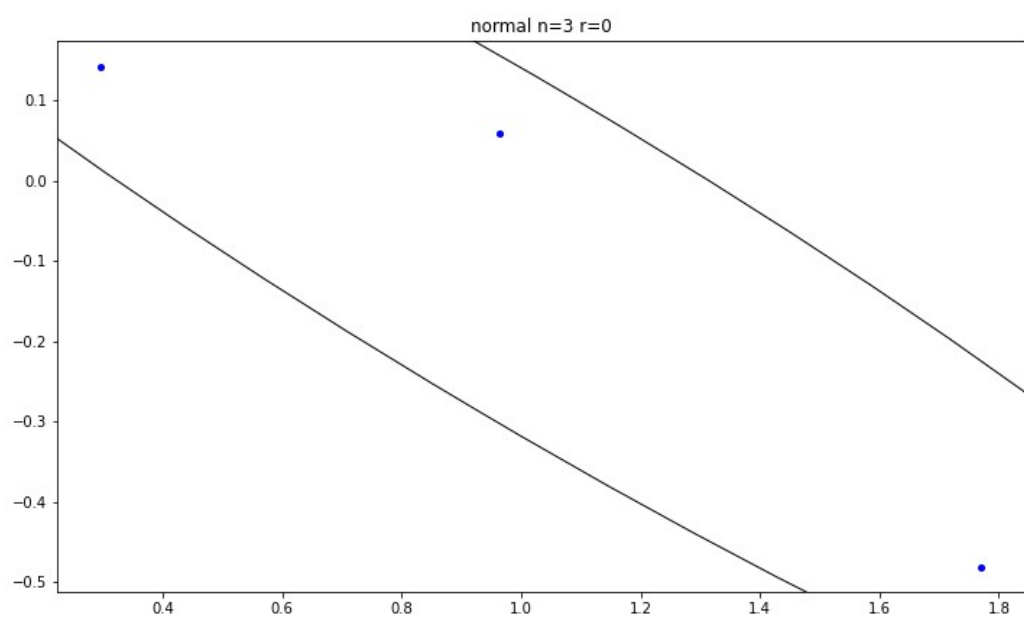


Рис. 35 Эллипс рассеивания для 3-х точек при $r=0.5$

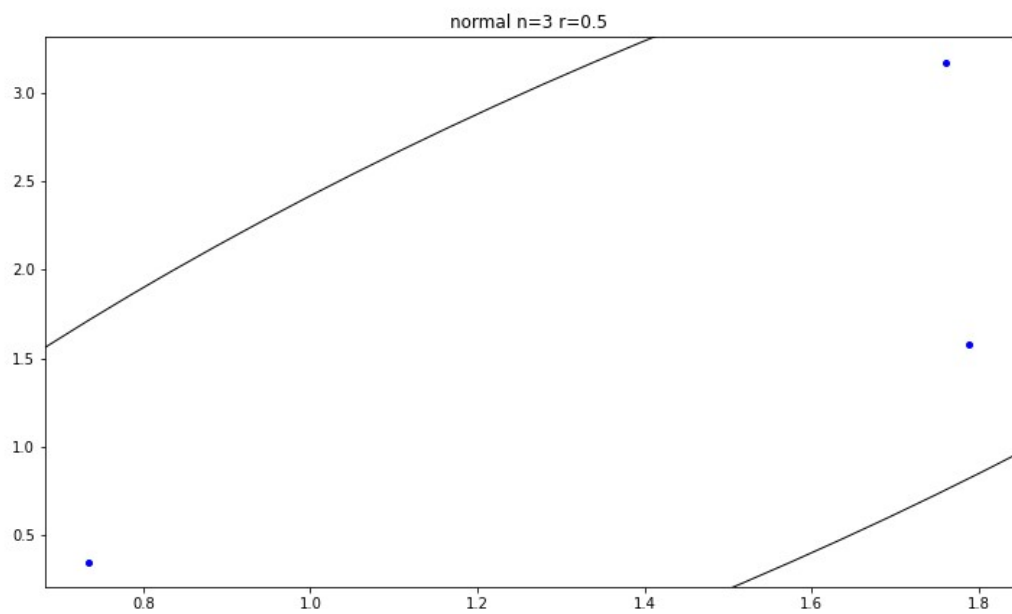
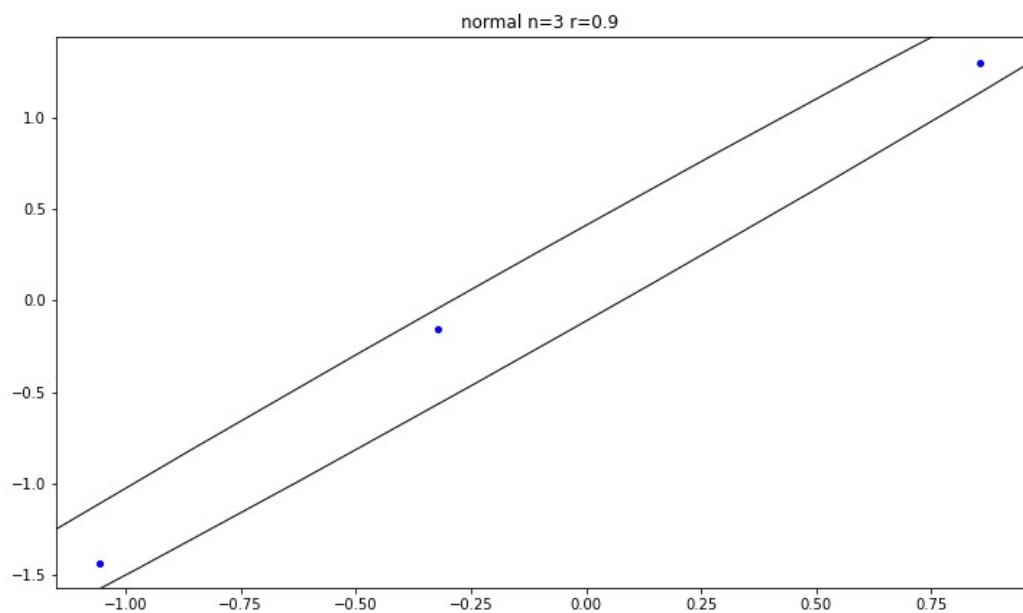


Рис. 36 Эллипс рассеивания для 3-х точек при $r=0.9$



5 Выводы

Ближе всего к теоретическому коэффициенту корреляции находится коэффициент Пирсона.

По графикам видно, что

- при увеличении объёма выборки коэффициенты корреляции стремятся к теоретическим
- при уменьшении корреляции эллипс рассеивания стремится к окружности, а при увеличении – вырождается в прямую с углом наклона в 45° против часовой стрелки
- для построения эллипса рассеивания нужно минимум 3 точки, а при 2-х точках эллипс вырождается в прямую под определённым углом

6 Литература

[Основы работы с *numpy* \(отдельная глава курса\)](#)

[Документация по *scipy*](#)

[Pandas обзор](#)

7 Приложения

[Код лабораторной](#)