

Санкт-Петербургский Политехнический университет Петра
Великого

Отчет по лабораторной работе №3

Анализ выбросов в распределениях

Студент:	Швачко Никита Андреевич
Преподаватель:	Баженов Александр Николаевич
Группа:	5030102/20202

Санкт-Петербург 2025

1 Формулировка задания и его формализация

Для анализа двухмерных нормальных распределений и смесей нормальных распределений необходимо:

1. Сгенерировать выборки размерами 20, 60 и 100 элементов для нормального двумерного распределения $N(x, y, 0, 0, 1, 1, \rho)$ с коэффициентами корреляции $\rho = 0.0, 0.5, 0.9$.
2. Каждая выборка генерируется 1000 раз, после чего вычисляются следующие статистики:
 - Среднее значение для коэффициентов корреляции Пирсона, Спирмена и квадратного коэффициента корреляции.
 - Среднее значение квадрата этих коэффициентов.
 - Дисперсия этих коэффициентов.
3. Повторить все вычисления для смеси нормальных распределений:

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

4. Изобразить сгенерированные точки на плоскости и нарисовать эллипс равновероятности для каждого случая.

2 Сгенерированные точки и эллипсы равновероятности

Для каждой выборки была выполнена генерация точек и построение эллипсов равновероятности. Ниже представлены графики сгенерированных точек для различных коэффициентов корреляции.

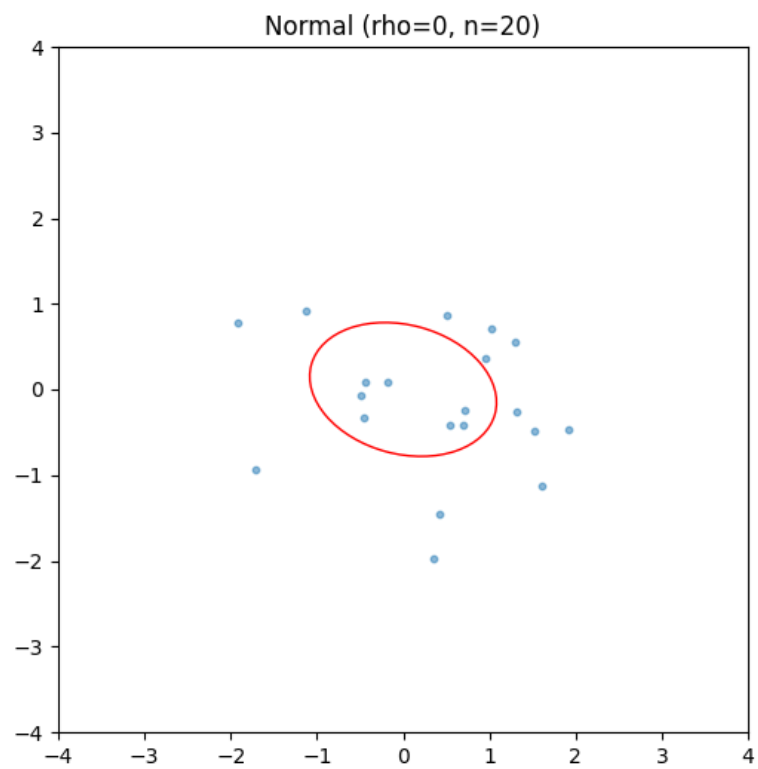


Рис. 1: Генерация точек для $\rho = 0.0$ и эллипс равновероятности для $n = 20$

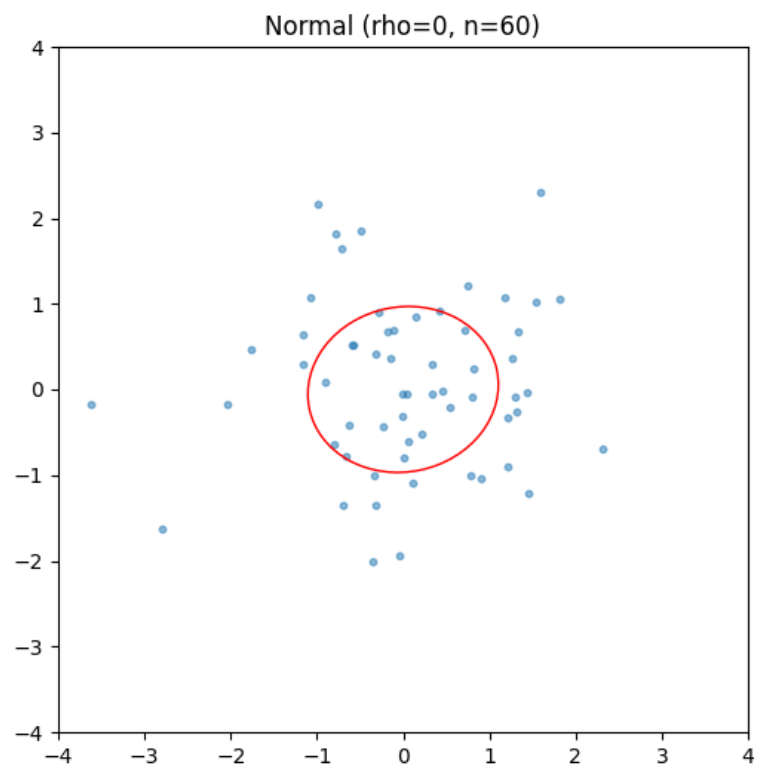


Рис. 2: Генерация точек для $\rho = 0.0$ и эллипс равновероятности для $n = 60$

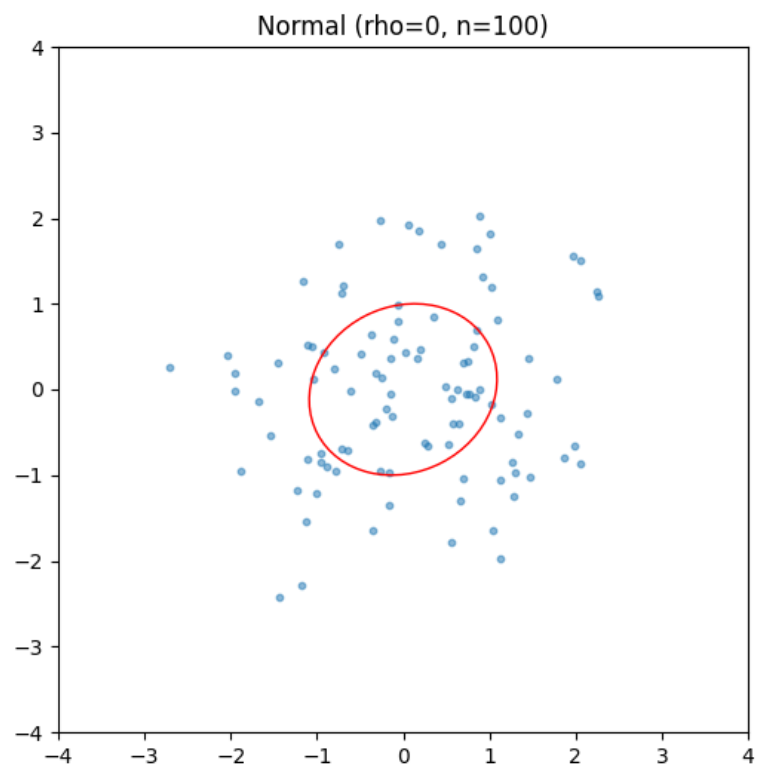


Рис. 3: Генерация точек для $\rho = 0.0$ и эллипс равновероятности для $n = 100$

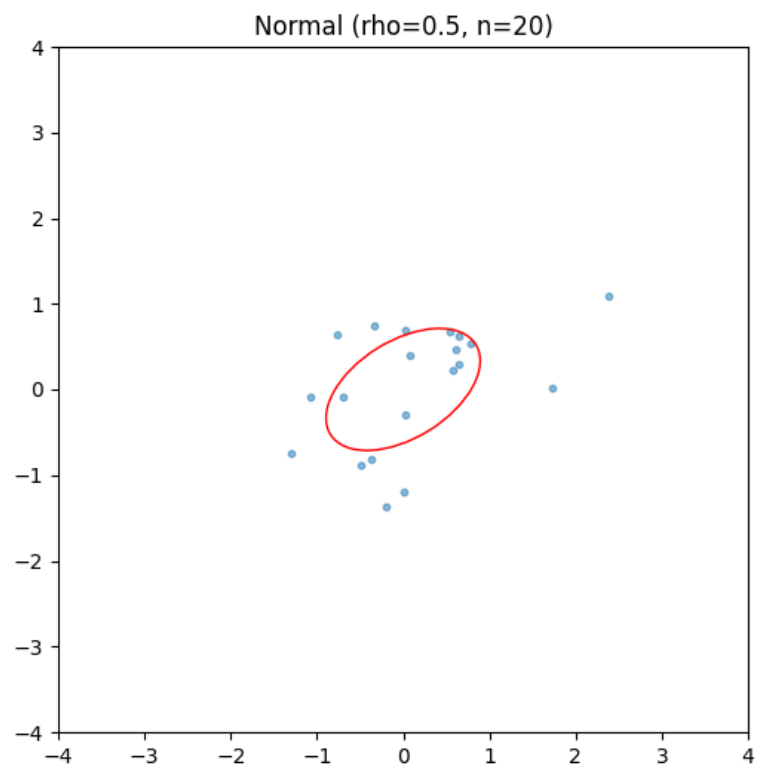


Рис. 4: Генерация точек для $\rho = 0.5$ и эллипс равновероятности для $n = 20$

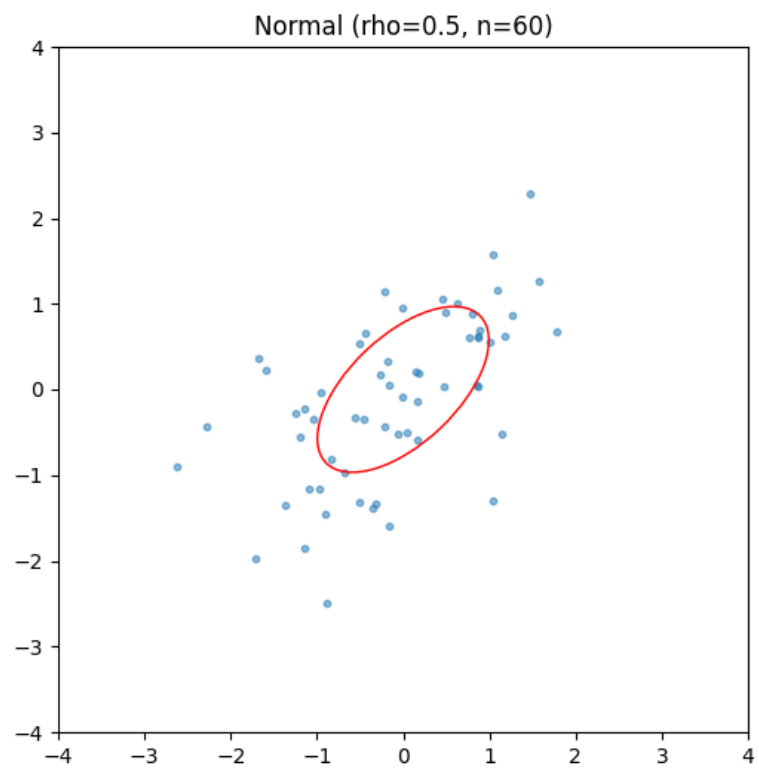


Рис. 5: Генерация точек для $\rho = 0.5$ и эллипс равновероятности для $n = 60$

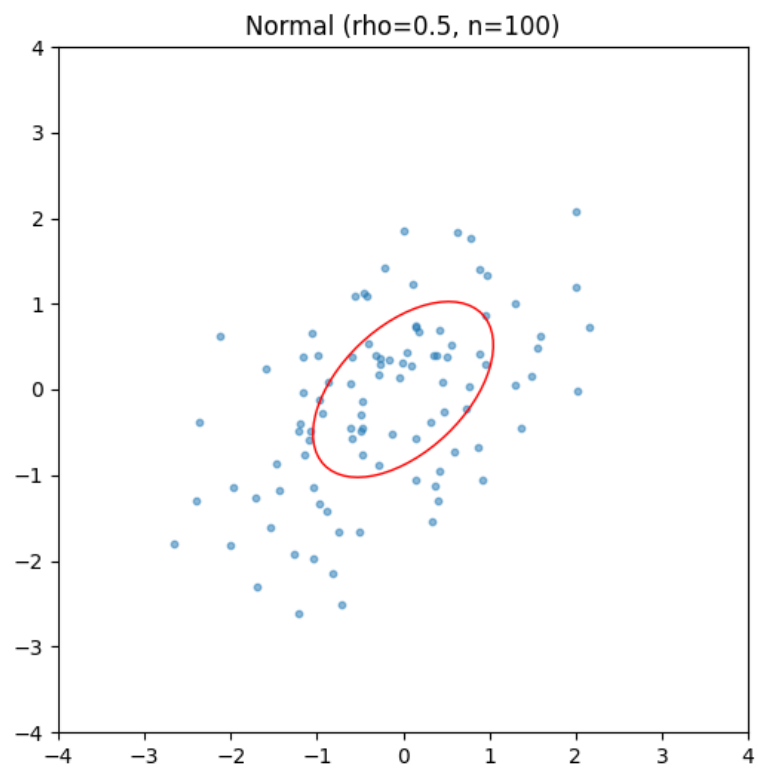


Рис. 6: Генерация точек для $\rho = 0.5$ и эллипс равновероятности для $n = 100$

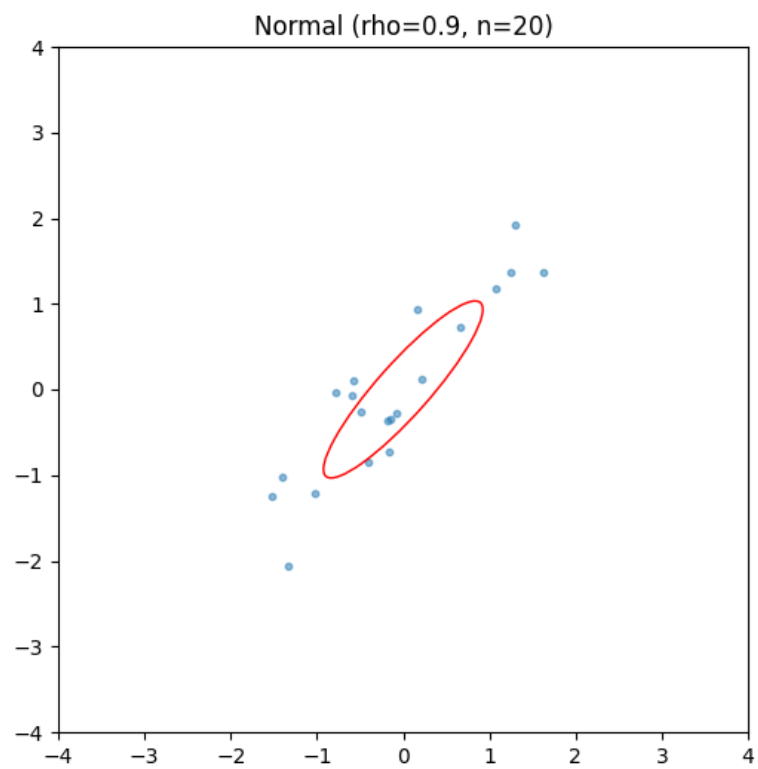


Рис. 7: Генерация точек для $\rho = 0.9$ и эллипс равновероятности для $n = 20$

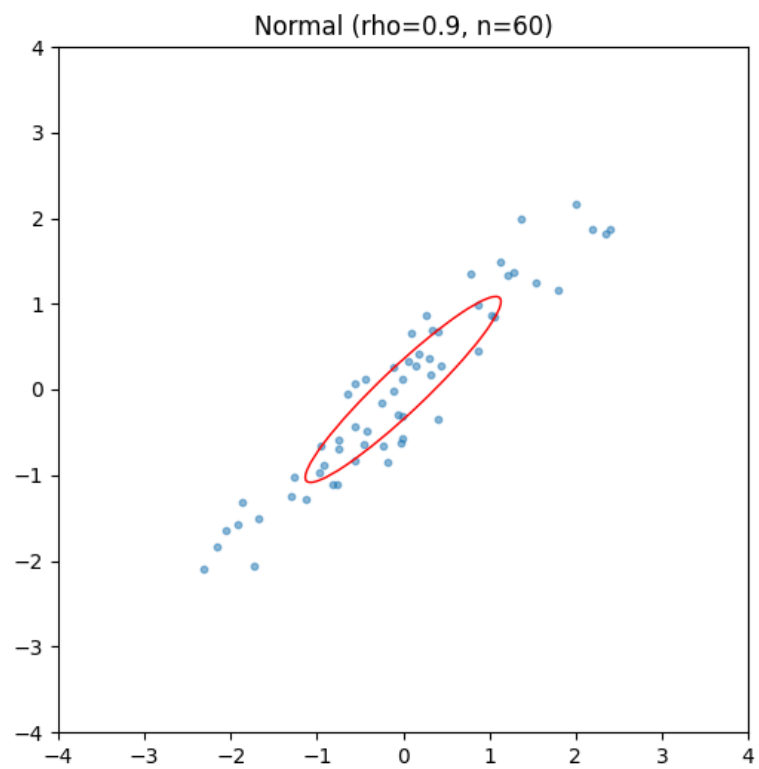


Рис. 8: Генерация точек для $\rho = 0.9$ и эллипс равновероятности для $n = 60$

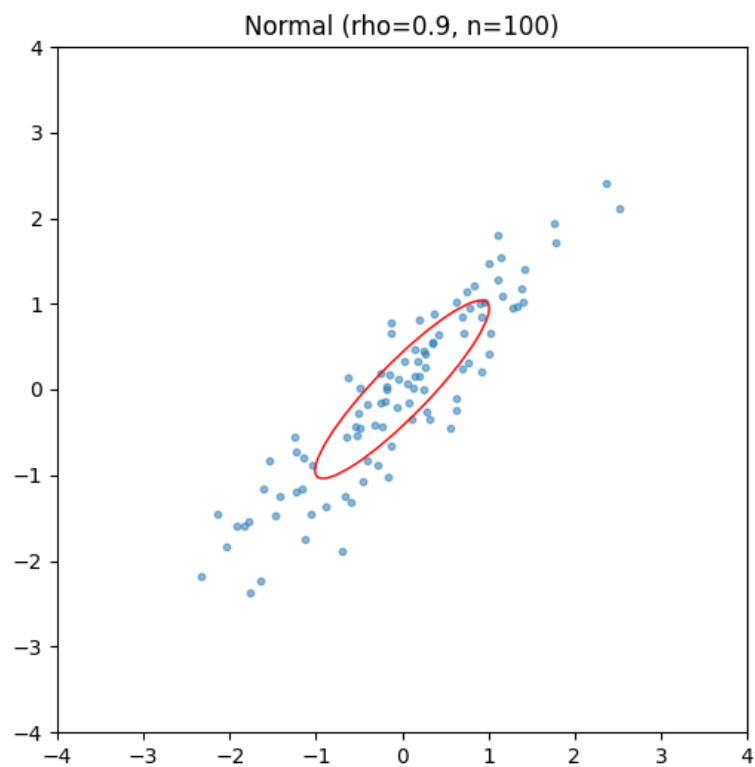


Рис. 9: Генерация точек для $\rho = 0.9$ и эллипс равновероятности для $n = 100$

3 Результаты для смеси нормальных распределений

Также были выполнены вычисления для смеси нормальных распределений. Графики точек и эллипсов равновероятности для смеси представлены ниже.

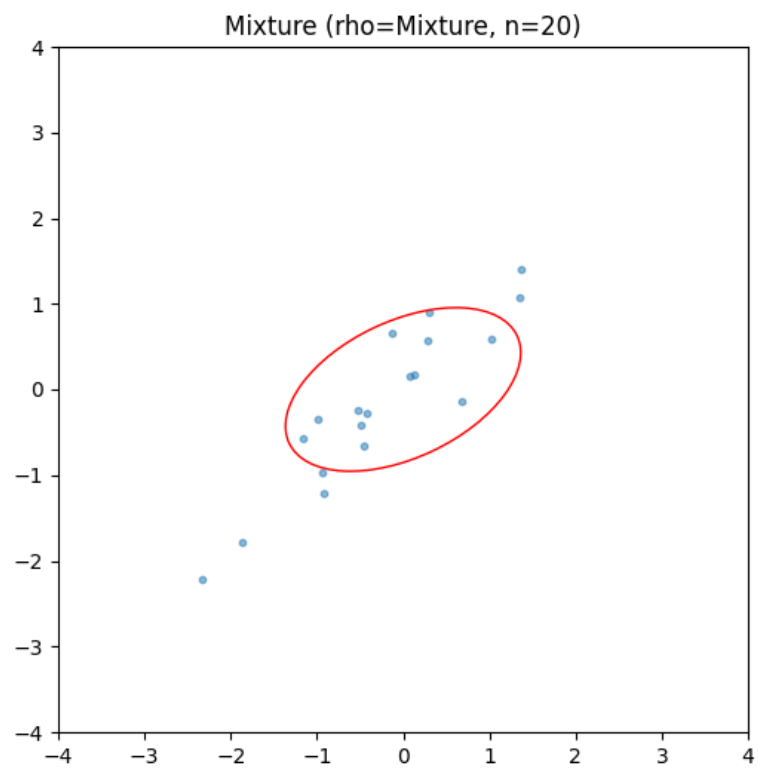


Рис. 10: Генерация точек для смеси с эллипсом равновероятности для $n = 20$

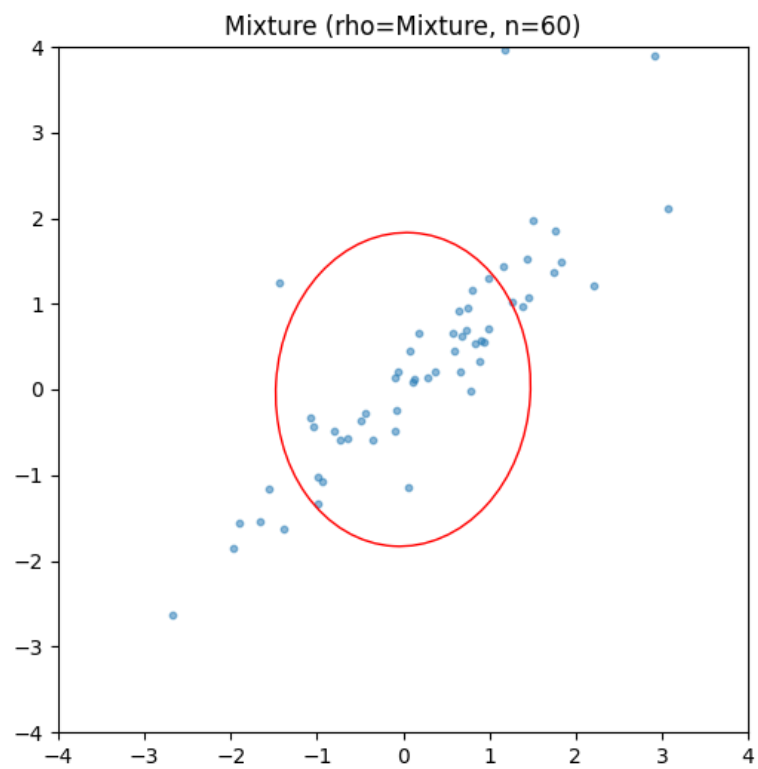


Рис. 11: Генерация точек для смеси с эллипсом равновероятности для $n = 60$

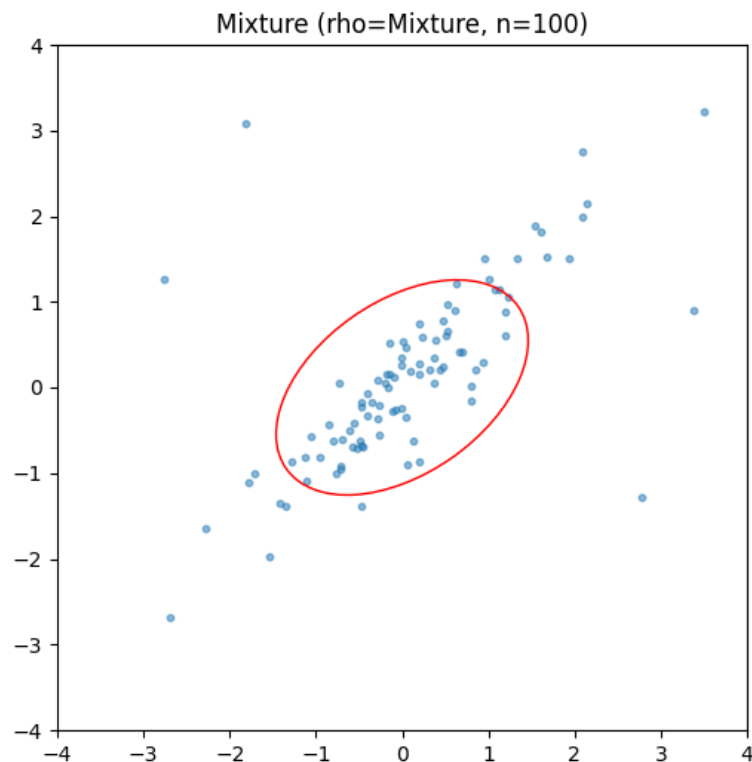


Рис. 12: Генерация точек для смеси с эллипсом равновероятности для $n = 100$

4 Результаты вычислений для коэффициентов корреляции

В результате вычислений для каждого коэффициента корреляции были получены следующие статистики (среднее значение, среднее квадрата и дисперсия) для выборок размером 20, 60 и 100.

5 Выводы

- Коэффициент корреляции Пирсона, Спирмена и квадрантный коэффициент дают схожие результаты для нормального распределения с высоким коэффициентом корреляции.
- Для смеси нормальных распределений результаты отличаются из-за смешанных компонент с различной корреляцией.
- Графики и эллипсы равновероятности демонстрируют изменение зависимости между переменными в зависимости от значения коэффициента корреляции.

Таблица 1: Средние значения и дисперсии коэффициентов корреляции для различных выборок

Распределение, коэфф корр, размер выборки	Фактический коэфф корр	Пирсона	Спирмена	Квадрантный
('Normal', 0, 20)	-0.19	-0.18	-0.23	-0.29
('Normal', 0, 60)	0.06	0.05	0.01	-0.13
('Normal', 0, 100)	0.12	0.12	0.08	0.02
('Normal', 0.5, 20)	0.47	0.47	0.43	0.59
('Normal', 0.5, 60)	0.59	0.59	0.60	0.54
('Normal', 0.5, 100)	0.50	0.51	0.51	0.32
('Normal', 0.9, 20)	0.90	0.90	0.82	0.90
('Normal', 0.9, 60)	0.95	0.95	0.94	0.84
('Normal', 0.9, 100)	0.91	0.91	0.90	0.70
('Mixture', ", 20)	0.45	0.52	0.67	0.70
('Mixture', ", 60)	0.03	0.11	0.63	0.73
('Mixture', ", 100)	0.44	0.44	0.66	0.60