Санкт-Петербургский Политехнический Университет Петра Великого

Физико-механический институт

Отчет по лабораторной работе №3 по математической статистике

Студент: Клыков Александр Юрьевич

Группа: 5030102/20101

Преподаватель: Баженов Александр Николаевич

Санкт-Петербург
2025 год

Задание

1 Коэффициент корреляции

1.1 Теоретическая справка

Двумерная случайная величина $\langle X, Y\rangle$ называется ***распределенной нормально***, если ее плотность вероятности определяется по формуле:

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

Компоненты X, Y двумерной СВ также распределены нормально с матожиданиями $\overline{x}$, $\overline{y}$ и СКО σ_x, σ_y соответственно. Параметр ρ называется *коэффициентом корреляции*

Ковариацией двух СВ X***,*** Y называется величина

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

.

Коэффициентом корреляции двух СВ X***,*** Y называется величина

$$(X, Y) = \frac{(X, Y)}{\sigma_x\sigma_y}$$

.

Выборочным коэффициентом корреляции Пирсона называется величина

$$(X, Y) = \frac{(X, Y)}{s_X s_Y},$$

где s_X^2, s_Y^2 - выборочные дисперсии X, Y.

Выборочным квадрантным коэффициентом корреляции называется величина

$$(x, y) = \frac{(n_1 + n_3) - (n_2 + n_4)}{n},$$

где n_1, n_2, n_3, n_4 - количества точек с координатами $(x_i, y_i)^T$ попавшими соответственно в I, II, III, IV квадранты ДСК $Ox'y'$, где $x' = x - x$, $y' = y - y$.

Обозначим ранги, соответствующие значениям переменной X через u, а ранги, соответствующие значениям переменной Y - v, тогда ***выборочным коэффициентом ранговой корреляции Спирмена*** называется величина

$$(x, y) = \frac{\frac{1}{n}\sum\limits_{i=1}^{n}(u_i - \overline{u})(v - \overline{v})}{\sqrt{\frac{1}{n^2}\sum\limits_{i=1}^{n}(u_i - \overline{u})^2 \sum\limits_{j=1}^{n}(v_i - \overline{v})^2}},$$

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

Уравнение проекции эллипса рассеивания на плоскость Oxy имеет вид

$$\frac{(x - \overline{x})^2}{\sigma_x^2} - 2(x, y)\frac{(x - \overline{x})(y - \overline{y})}{\sigma_x \sigma_y} + \frac{(y - \overline{y})^2}{\sigma_y^2} =,$$

центр эллипса находится в точке с координатами $(\overline{x}, \overline{y})^T$, оси симметрии составляют с осью Ox углы α, определяемые равенством

$$\alpha = \frac{1}{2}\arctan\frac{2(x, y)\sigma_x\sigma_y}{\sigma_x^2 - \sigma_y^2} + \frac{\pi k}{2}, k \in Z$$

1.2 Постановка задачи

Сгенерировать двумерные выборки размерами 20, 60, 100 для нормального двумерного распределения $(x, y, 0, 0, 1, 1,)$. Коэффициент корреляции взять равным 0, 0.5, 0.9. Каждая выборка генерируется 1000 раз и для нее вычисляются среднее значение, среднее значение квадрата и дисперсия коэффициентов корреляции Пирсона, Спирмена и ККК. Повторить все вычисления для смеси нормальных распределений:

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

изобразить сгенерированные точки на плоскости и нарисовать эллипс равновероятности.

1.3 Результаты

Размер выборки	Метрика	$\rho = 0$	$\rho = 0.5$	$\rho = 0.9$
3*20	Среднее значение	-0.002	0.489	0.894
	Средний квадрат	0.056	0.269	0.802
	Дисперсия	0.056	0.030	0.002
3*60	Среднее значение	-0.004	0.495	0.898
	Средний квадрат	0.016	0.254	0.808
	Дисперсия	0.016	0.009	0.001
3*100	Среднее значение	0.006	0.499	0.898
	Средний квадрат	0.010	0.255	0.807
	Дисперсия	0.010	0.006	0.000

Таблица 1: Коэффициент Пирсона

Размер выборки	Метрика	$\rho = 0$	$\rho = 0.5$	$\rho = 0.9$
3*20	Среднее значение	0.496	0.669	0.856
	Средний квадрат	0.259	0.459	0.738
	Дисперсия	0.013	0.011	0.006
3*60	Среднее значение	0.499	0.667	0.857
	Средний квадрат	0.253	0.448	0.736
	Дисперсия	0.004	0.003	0.002
3*100	Среднее значение	0.501	0.668	0.857
	Средний квадрат	0.254	0.448	0.735
	Дисперсия	0.002	0.002	0.001

Таблица 2: Квадрантный коэффициент корреляции

Размер выборки	Метрика	$\rho = 0$	$\rho = 0.5$	$\rho = 0.9$
3*20	Среднее значение	-0.002	0.458	0.866
	Средний квадрат	0.055	0.244	0.755
	Дисперсия	0.055	0.034	0.005
3*60	Среднее значение	-0.004	0.476	0.884
	Средний квадрат	0.016	0.237	0.782
	Дисперсия	0.016	0.010	0.001
3*100	Среднее значение	0.006	0.479	0.886
	Средний квадрат	0.010	0.237	0.786
	Дисперсия	0.010	0.007	0.001

Таблица 3: Ранговый коэффициент Спирмена

Размер выборки	Пирсон	Квадрант	Спирмен
20	0.870	0.842	0.841
60	0.877	0.840	0.859
100	0.878	0.842	0.865

Таблица 4: Коэффициенты корреляции для смеси нормальных распределений

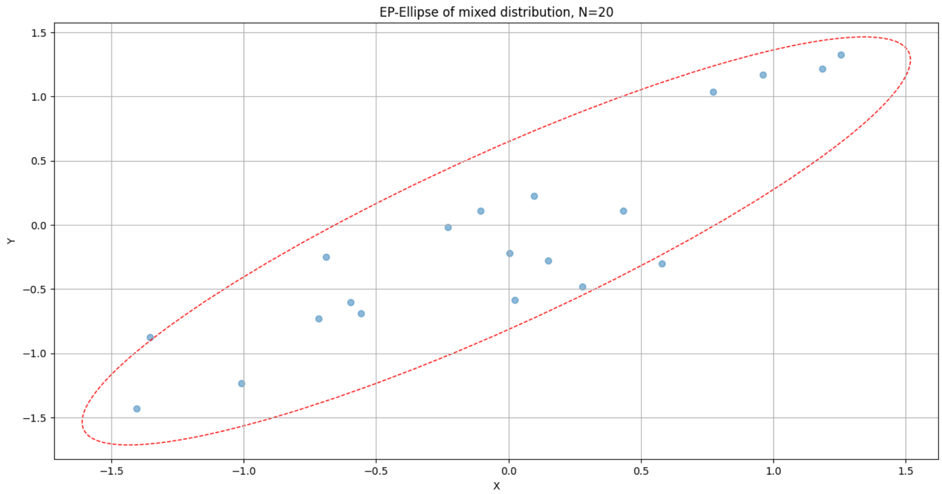

Рис. 1: Эллипс равновероятности смеси распределений, $N = 20$

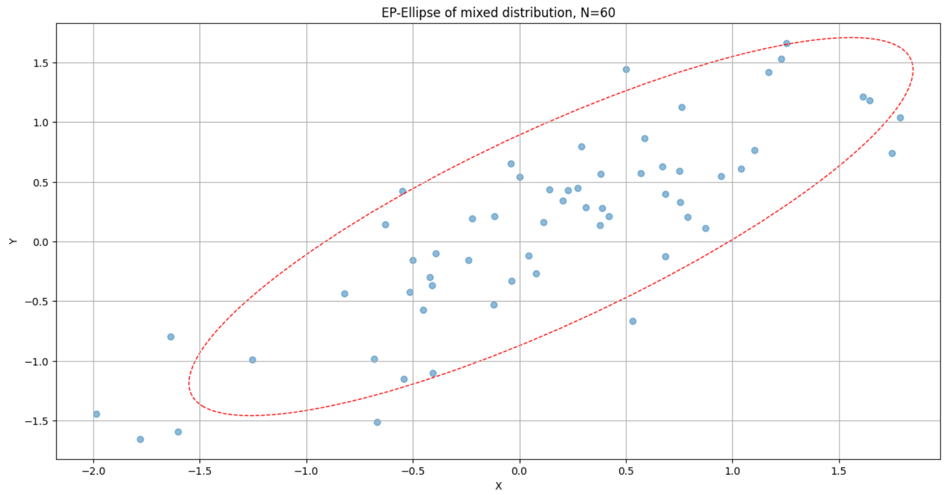

Рис. 2: Эллипс равновероятности смеси распределений, $N = 60$

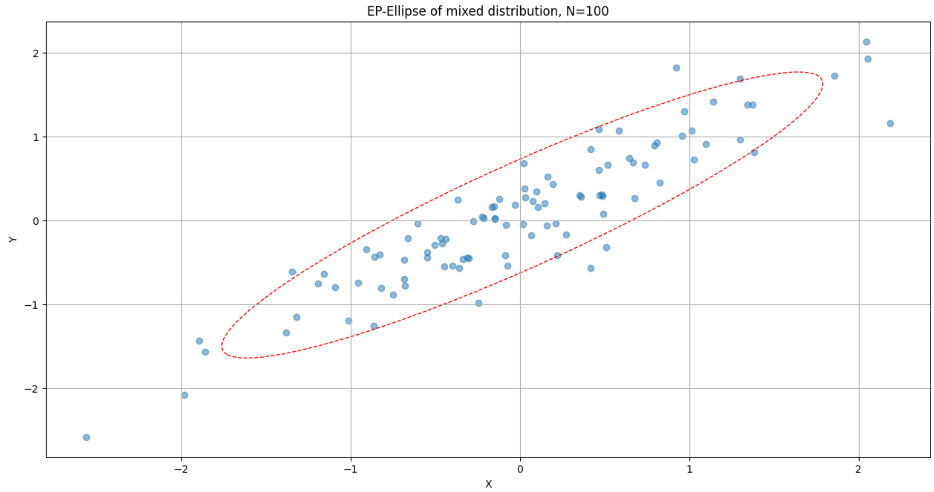

Рис. 3: Эллипс равновероятности смеси распределений, $N = 100$

Выводы

В ходе выполнения лабораторной работы были исследованы зависимости между переменными в двумерных выборках, сгенерированных из нормального распределения с различными коэффициентами корреляции, а также из смеси нормальных распределений.

Для каждого случая были рассчитаны коэффициенты корреляции Пирсона, Спирмена и квадрантный коэффициент корреляции. Полученные значения показали:

- Все три коэффициента хорошо отражают силу линейной зависимости между переменными: при увеличении ρ их значения приближаются к 1.

- При $\rho = 0$ все коэффициенты в среднем близки к нулю, что соответствует отсутствию корреляции.

- Квадрантный и Спирменовский коэффициенты оказываются менее чувствительными к выбросам по сравнению с Пирсоновским.

- При увеличении объёма выборки (от $N = 20$ до $N = 100$) дисперсия коэффициентов уменьшается, а оценки становятся более точными и стабильными.

- Для смеси распределений коэффициенты корреляции сохраняют высокие значения, особенно у Спирмена, что демонстрирует его устойчивость к аномальным значениям.

Также на графиках были изображены эллипсы равновероятности, показывающие характер распределения точек. При $\rho = 0.9$ эллипсы вытянуты вдоль прямой корреляции, тогда как при $\rho = 0$ они приближаются к окружности.

Таким образом, работа наглядно демонстрирует поведение различных мер корреляции при разных типах и параметрах распределений.