

САНКТ-ПЕТЕРБУРГСКИЙ ПОЛИТЕХНИЧЕСКИЙ УНИВЕРСИТЕТ

Институт прикладной математики и механики

Высшая школа прикладной математики и вычислительной физики

Отчет
по лабораторным работам №5-6
по дисциплине
«Математическая статистика»

Выполнил студент:
Колосков Александр
группа: 3630102/80301

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

Санкт-Петербург
2021 г.

Содержание

	Страница
1 Постановка задачи	4
2 Теория	4
2.1 Двумерное нормальное распределение	4
2.2 Корреляционный момент и коэффициент корреляции	4
2.3 Выборочные коэффициенты корреляции	5
2.3.1 Выборочный коэффициент корреляции Пирсона	5
2.3.2 Выборочный квадрантный коэффициент корреляции	5
2.3.3 Выборочный коэффициент ранговой корреляции Спирмена	5
2.4 Эллипсы рассеивания	5
2.5 Простая линейная регрессия	5
2.5.1 Модель простой линейной регрессии	5
2.5.2 Метод наименьших квадратов	6
2.5.3 Расчётные формулы для МНК-оценок	6
2.6 Робастные оценки коэффициентов линейной регрессии	6
2.7 Метод максимального правдоподобия	6
2.8 Проверка гипотезы о законе распределения генеральной совокупности. Метод хи-квадрат	7
2.9 Критерий Фишера	7
3 Реализация	7
4 Результаты	8
4.1 Выборочные коэффициенты корреляции	8
4.2 Эллипсы рассеивания	9
4.3 Оценки коэффициентов линейной регрессии	11
4.3.1 Выборка без возмущений	11
4.3.2 Выборка с возмущениями	11
4.4 Проверка гипотезы о законе распределения генеральной совокупности. Метод хи-квадрат	11
4.5 Исследование сигнала с использованием критерия Фишера	12
5 Обсуждение	14
5.1 Выборочные коэффициенты корреляции и эллипсы рассеивания	14
5.2 Оценки коэффициентов линейной регрессии	14
5.3 Проверка гипотезы о нормальном распределении выборки методом хи-квадрат	14
5.4 Исследование сигнала с использованием критерия Фишера	14

Список иллюстраций

	Страница
1 Эллипсы рассеивания. $\rho = 0(4)$	9
2 Эллипсы рассеивания. $\rho = 0.5(4)$	10
3 Эллипсы рассеивания. $\rho = 0.9(4)$	10
4 Выборка без возмущений	11
5 Выборка с возмущениями	11
6 Входной сигнал	12
7 Гистограмма сигнала	13
8 Входной сигнал без возмущение	13
9 Разделенные области сигнала	13

Список таблиц

	Страница
1 Выборочные коэффициенты корреляции двумерного нормального распределения. $\rho = 0(4)$	8
2 Выборочные коэффициенты корреляции двумерного нормального распределения. $\rho = 0.5(4)$	8
3 Выборочные коэффициенты корреляции двумерного нормального распределения. $\rho = 0.9(4)$	8
4 Выборочные коэффициенты корреляции смешанного распределения. (1)	9
5 Проверка гипотезы о законе $N(\hat{\mu}, \hat{\sigma})$ для выборки на 100 элементов с распределением $N(0, 1)$	12
6 Проверка гипотезы о законе $N(\hat{\mu}, \hat{\sigma})$ для выборки на 20 элементов с распределением $U([\hat{\mu} - 3\hat{\sigma}, \hat{\mu} + 3\hat{\sigma}])$	12
7 Проверка гипотезы о законе $N(\hat{\mu}, \hat{\sigma})$ для выборки на 20 элементов с распределением $L(\hat{\mu}, \frac{\hat{\sigma}}{\sqrt{2}})$	12
8 Характеристики выделенных областей	14

1 Постановка задачи

1. 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:

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

Изобразить сгенерированные точки на плоскости и нарисовать эллипс равновероятности.

2. Найти оценки коэффициентов линейной регрессии $y_i = a + bx_i + e_i$, используя 20 точек на отрезке $[-1.8; 2]$ с равномерным шагом равным 0.2. Ошибку e_i считать нормально распределённой с параметрами (0,1). В качестве эталонной зависимости взять $y_i = 2 + 2x_i + e_i$. При построении оценок коэффициентов использовать два критерия: критерий наименьших квадратов и критерий наименьших модулей. Прodelать то же самое для выборки, у которой в значения y_1 и y_{20} вносятся возмущения 10 и -10.
3. Сгенерировать выборку объёмом 100 элементов для нормального распределения $N(x, 0, 1)$. По сгенерированной выборке оценить параметры μ и σ нормального закона методом максимального правдоподобия. В качестве основной гипотезы H_0 будем считать, что сгенерированное распределение имеет вид $N(x, \hat{\mu}, \hat{\sigma})$. Проверить основную гипотезу, используя критерий согласия χ^2 . В качестве уровня значимости взять $\alpha = 0.05$. Привести таблицу вычислений χ^2 . Исследовать точность (чувствительность) критерия χ^2 - сгенерировать выборки равномерного распределения и распределения Лапласа малого объема (например, 20 элементов). Проверить их на нормальность.
4. Провести дисперсионный анализ с применением критерия Фишера по данным регистраторов для сигнала длиной 1024 - устранить явные выбросы использованием медианного фильтра; разведочным анализом с применением гистограммы определить области однородности и переходные области для сглаженного сигнала; с помощью критерия Фишера определить тип областей.

2 Теория

2.1 Двумерное нормальное распределение

Двумерная случайная величина (X, Y) называется распределенной нормально, если её плотность вероятности определяется формулой

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

где $\bar{x}, \bar{y}, \sigma_x, \sigma_y$ - математические ожидания и средние квадратические отклонения компонент X, Y соответственно, а ρ_{XY} - коэффициент корреляции.

2.2 Корреляционный момент и коэффициент корреляции

Корреляционный момент (ковариация) двух случайных величин X, Y :

$$K_{XY} = \text{cov}(X, Y) = \mathbf{M}[(X - \bar{x})(Y - \bar{y})]. \quad (3)$$

Коэффициент корреляции ρ_{XY} случайных величин X, Y :

$$\rho_{XY} = \frac{K_{XY}}{\sigma_x\sigma_y}. \quad (4)$$

Ковариационной матрицей случайного вектора (X, Y) называется симметричная матрица вида

$$K = \begin{pmatrix} D_X & K_{XY} \\ K_{YX} & D_Y \end{pmatrix}. \quad (5)$$

Кореляционной матрицей случайного вектора (X, Y) называется нормированная ковариационная матрица вида

$$R = \begin{pmatrix} 1 & \rho_{XY} \\ \rho_{YX} & 1 \end{pmatrix}. \quad (6)$$

2.3 Выборочные коэффициенты корреляции

2.3.1 Выборочный коэффициент корреляции Пирсона

Выборочный коэффициент корреляции Пирсона:

$$r = \frac{\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2 \frac{1}{n} \sum_{i=1}^n (y_i - \bar{y})^2}} = \frac{K_{XY}}{s_X s_Y}, \quad (7)$$

где K , s_X^2 , s_Y^2 – выборочные ковариация и дисперсии случайных величин X, Y .

2.3.2 Выборочный квадрантный коэффициент корреляции

$$r_Q = \frac{(n_1 + n_3) - (n_2 + n_4)}{n}, \quad (8)$$

где n_1, n_2, n_3, n_4 – количества точек с координатами (x_i, y_i) , попавшими соответственно в I, II, III и IV квадранты декартовой системы с осями $x' = x - \text{med } x$, $y' = y - \text{med } y$ и с центром в точке с координатами $(\text{med } x, \text{med } y)$.

2.3.3 Выборочный коэффициент ранговой корреляции Спирмена

Обозначим ранги, соответствующие значениям переменной X , через u , а ранги, соответствующие значениям переменной Y , – через v .

Выборочный коэффициент ранговой корреляции Спирмена:

$$r_S = \frac{\frac{1}{n} \sum_{i=1}^n (u_i - \bar{u})(v_i - \bar{v})}{\sqrt{\frac{1}{n} \sum_{i=1}^n (u_i - \bar{u})^2 \frac{1}{n} \sum_{i=1}^n (v_i - \bar{v})^2}}, \quad (9)$$

где $\bar{u} = \bar{v} = \frac{1+2+\dots+n}{n} = \frac{n+1}{2}$ – среднее значение рангов.

2.4 Эллипсы рассеивания

Уравнение проекции эллипса рассеивания на плоскость xOy :

$$\frac{(x - \bar{x})^2}{\sigma_x^2} - 2\rho_{XY} \frac{(x - \bar{x})(y - \bar{y})}{\sigma_x \sigma_y} + \frac{(y - \bar{y})^2}{\sigma_y^2} = C, \quad C - \text{const.} \quad (10)$$

Центр эллипса (10) находится в точке с координатами (\bar{x}, \bar{y}) , оси симметрии эллипса составляют с осью Ox углы, определяемые уравнением

$$\tan 2\alpha = \frac{2\rho_{XY}\sigma_x\sigma_y}{\sigma_x^2 - \sigma_y^2}. \quad (11)$$

2.5 Простая линейная регрессия

2.5.1 Модель простой линейной регрессии

Регрессионную модель описания данных называют *простой линейной регрессией*, если

$$y_i = \beta_0 + \beta_1 x_i + \varepsilon_i, \quad i = 1, \dots, n, \quad (12)$$

где x_1, \dots, x_n – заданные числа (значения фактора); y_1, \dots, y_n – наблюдаемые значения отклика; $\varepsilon_1, \dots, \varepsilon_n$ – независимые, нормально распределенные $N(0, \sigma)$ с нулевым математическим ожиданием и одинаковой (неизвестной) дисперсией случайные величины (ненаблюдаемые); β_0, β_1 – неизвестные параметры, подлежащие оцениванию.

2.5.2 Метод наименьших квадратов

Метод наименьших квадратов (МНК):

$$Q(\beta_0, \beta_1) = \sum_{i=1}^n \varepsilon_i^2 = \sum_{i=1}^n (y_i - \beta_0 - \beta_1 x_i)^2 \rightarrow \min_{\beta_0, \beta_1}. \quad (13)$$

2.5.3 Расчётные формулы для МНК-оценок

МНК-оценки параметров β_0 и β_1 :

$$\hat{\beta}_1 = \frac{\overline{xy} - \bar{x} \cdot \bar{y}}{\overline{x^2} - (\bar{x})^2}, \quad (14)$$

$$\hat{\beta}_0 = \bar{y} - \bar{x} \hat{\beta}_1. \quad (15)$$

2.6 Робастные оценки коэффициентов линейной регрессии

Метод наименьших модулей:

$$\sum_{i=1}^n |y_i - \beta_0 - \beta_1 x_i| \rightarrow \min_{\beta_0, \beta_1}. \quad (16)$$

$$\hat{\beta}_{1R} = r_Q \frac{q_y^*}{q_x^*}, \quad (17)$$

$$\hat{\beta}_{0R} = \text{med } y - \hat{\beta}_{1R} \text{ med } x, \quad (18)$$

$$r_Q = \frac{1}{n} \sum_{i=1}^n \text{sign}(x_i - \text{med } x) \text{sign}(y_i - \text{med } y), \quad (19)$$

$$q_y^* = \frac{y_{(j)} - y_{(l)}}{k_q(n)}, \quad q_x^* = \frac{x_{(j)} - x_{(l)}}{k_q(n)} \quad (20)$$

$$l = \begin{cases} [n/4] + 1 & \text{при } n/4 \text{ дробном,} \\ n/4 & \text{при } n/4 \text{ целом.} \end{cases}$$

$$j = n - l + 1.$$

$$\text{sign } z = \begin{cases} 1 & \text{при } z > 0, \\ 0 & \text{при } z = 0, \\ -1 & \text{при } z < 0. \end{cases}$$

Уравнение регрессии здесь имеет вид

$$y = \hat{\beta}_{0R} + \hat{\beta}_{1R} \cdot x. \quad (21)$$

$$k_q(20) = 1.491.$$

2.7 Метод максимального правдоподобия

$L(x_1, \dots, x_n, \theta)$ — функция правдоподобия (ФП), рассматриваемая как функция неизвестного параметра θ :

$$L(x_1, \dots, x_n, \theta) = f(x_1, \theta) f(x_2, \theta) \dots f(x_n, \theta). \quad (22)$$

Оценка максимального правдоподобия:

$$\hat{\theta}_{\text{МП}} = \arg \max_{\theta} L(x_1, \dots, x_n, \theta). \quad (23)$$

Система уравнений правдоподобия (в случае дифференцируемости функции правдоподобия):

$$\frac{\partial L}{\partial \theta_k} = 0 \quad \text{или} \quad \frac{\partial \ln L}{\partial \theta_k} = 0, \quad k = 1, \dots, m. \quad (24)$$

2.8 Проверка гипотезы о законе распределения генеральной совокупности. Метод хи-квадрат

Выдвинута гипотеза H_0 о генеральном законе распределения с функцией распределения $F(x)$.

Рассматриваем случай, когда гипотетическая функция распределения $F(x)$ не содержит неизвестных параметров.

Правило проверки гипотезы о законе распределения по методу χ^2

1. Выбираем уровень значимости α .
2. По таблице [1, с. 358] находим квантиль $\chi^2_{1-\alpha}(k-1)$ распределения хи-квадрат с $k-1$ степенями свободы порядка $1-\alpha$.
3. Вычисляем вероятности $p_i = P(X \in \Delta_i), i = 1, \dots, k$, с помощью гипотетической функции распределения $F(x)$.
4. Находим частоты n_i попадания элементов выборки в подмножества $\Delta_i, i = 1, \dots, k$.
5. Вычисляем выборочное значение статистики критерия χ^2 :

$$\chi^2_B = \sum_{i=1}^k \frac{(n_i - np_i)^2}{np_i}. \quad (25)$$

6. Сравниваем χ^2_B и квантиль $\chi^2_{1-\alpha}(k-1)$.
 - (a) Если $\chi^2_B < \chi^2_{1-\alpha}(k-1)$, то гипотеза H_0 на данном этапе проверки принимается.
 - (b) Если $\chi^2_B \geq \chi^2_{1-\alpha}(k-1)$, то гипотеза H_0 отвергается, выбирается одно из альтернативных распределений, и процедура проверки повторяется.

2.9 Критерий Фишера

Внутригрупповая дисперсия:

$$s^2_{In} = \frac{1}{k} \sum_{i=1}^k s^2_i, \quad s_i = \frac{\sum_{j=1}^{n_i} (x_{ij} - \bar{x}_i)^2}{n_i - 1}, \quad (26)$$

где k - количество разбиений выборки, n_i - количество элементов в i -ом разбиении.

Межгрупповая дисперсия:

$$s^2_{Inter} = \frac{k}{k-1} \sum_{i=1}^k (\bar{x}_i - \bar{x})^2 \quad (27)$$

Значение критерия Фишера:

$$F = \frac{s^2_{Inter}}{s^2_{In}} \quad (28)$$

3 Реализация

Лабораторная работа выполнена на языке Python в среде PyCharm с использованием библиотек numpy, scipy.stats, matplotlib.pyplot, statsmodels

4 Результаты

4.1 Выборочные коэффициенты корреляции

$n=20$	$r(7)$	$r_S(9)$	$r_Q(8)$
$E(z)$	-0.0077	-0.0037	0.0067
$E(z^2)$	0.0498	0.0498	0.0497
$D(z)$	0.0497	0.0497	0.0496
$n=60$	r	r_S	r_Q
$E(z)$	-0.0057	-0.0066	-0.0168
$E(z^2)$	0.0535	0.0529	0.0507
$D(z)$	0.0535	0.0528	0.0504
$n=100$	r	r_S	r_Q
$E(z)$	-0.0048	-0.0069	-0.0088
$E(z^2)$	0.0517	0.051	0.056
$D(z)$	0.0517	0.051	0.056

Таблица 1: Выборочные коэффициенты корреляции двумерного нормального распределения. $\rho = 0(4)$

$n=20$	$r(7)$	$r_S(9)$	$r_Q(8)$
$E(z)$	0.4922	0.4701	0.3296
$E(z^2)$	0.252	0.2318	0.1233
$D(z)$	0.0098	0.0108	0.0147
$n=60$	r	r_S	r_Q
$E(z)$	0.4986	0.4768	0.334
$E(z^2)$	0.2574	0.2379	0.1269
$D(z)$	0.0088	0.0106	0.0153
$n=100$	r	r_S	r_Q
$E(z)$	0.4988	0.4744	0.329
$E(z^2)$	0.2584	0.2359	0.1237
$D(z)$	0.0097	0.0109	0.0155

Таблица 2: Выборочные коэффициенты корреляции двумерного нормального распределения. $\rho = 0.5(4)$

$n=20$	$r(7)$	$r_S(9)$	$r_Q(8)$
$E(z)$	0.8997	0.8869	0.7142
$E(z^2)$	0.81	0.7872	0.5151
$D(z)$	0.0004	0.0007	0.005
$n=60$	r	r_S	r_Q
$E(z)$	0.8977	0.8843	0.7133
$E(z^2)$	0.8062	0.7826	0.5138
$D(z)$	0.0004	0.0007	0.0051
$n=100$	r	r_S	r_Q
$E(z)$	0.8985	0.8854	0.7118
$E(z^2)$	0.8076	0.7846	0.512
$D(z)$	0.0004	0.0006	0.0053

Таблица 3: Выборочные коэффициенты корреляции двумерного нормального распределения. $\rho = 0.9(4)$

$n=20$	r	r_S	r_Q
$E(z)$	-0.3039	0.4771	0.1509
$E(z^2)$	0.5442	0.3042	0.2747
$D(z)$	0.4518	0.0765	0.2519
$n=60$	r	r_S	r_Q
$E(z)$	-0.6378	0.4742	0.3501
$E(z^2)$	0.487	0.2513	0.2017
$D(z)$	0.0802	0.0264	0.0791
$n=100$	r	r_S	r_Q
$E(z)$	-0.6888	0.4756	0.3935
$E(z^2)$	0.5051	0.2422	0.2105
$D(z)$	0.0307	0.016	0.0556

Таблица 4: Выборочные коэффициенты корреляции смешанного распределения. (1)

4.2 Эллипсы рассеивания

Для всех экспериментов – Const = 3

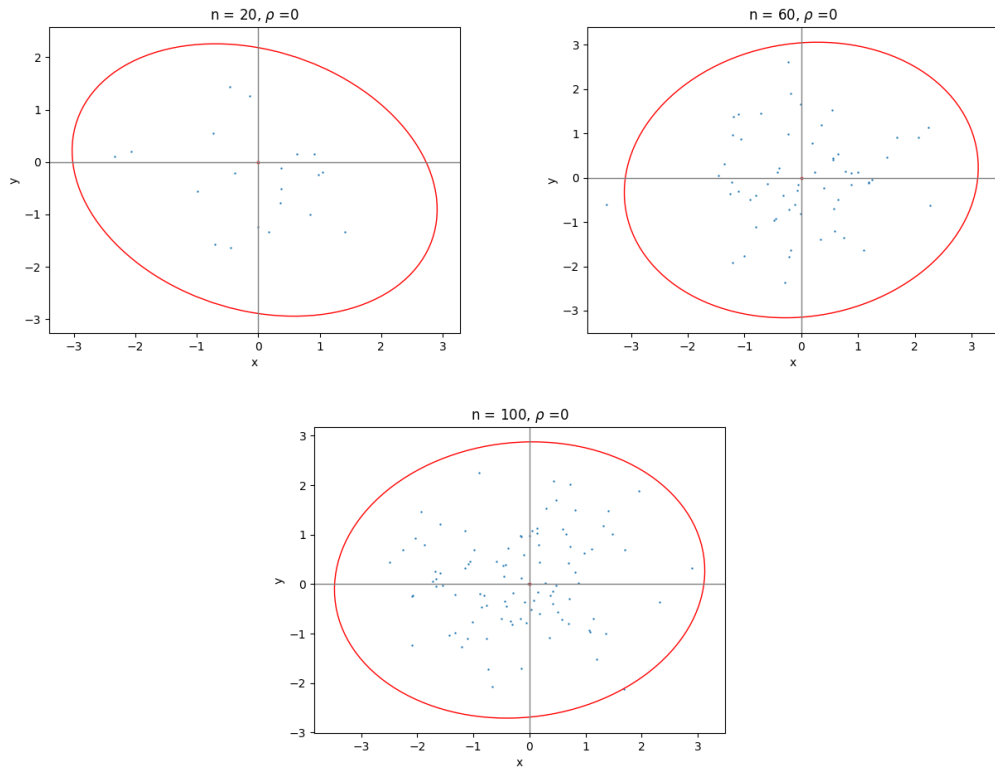


Рис. 1: Эллипсы рассеивания. $\rho = 0(4)$

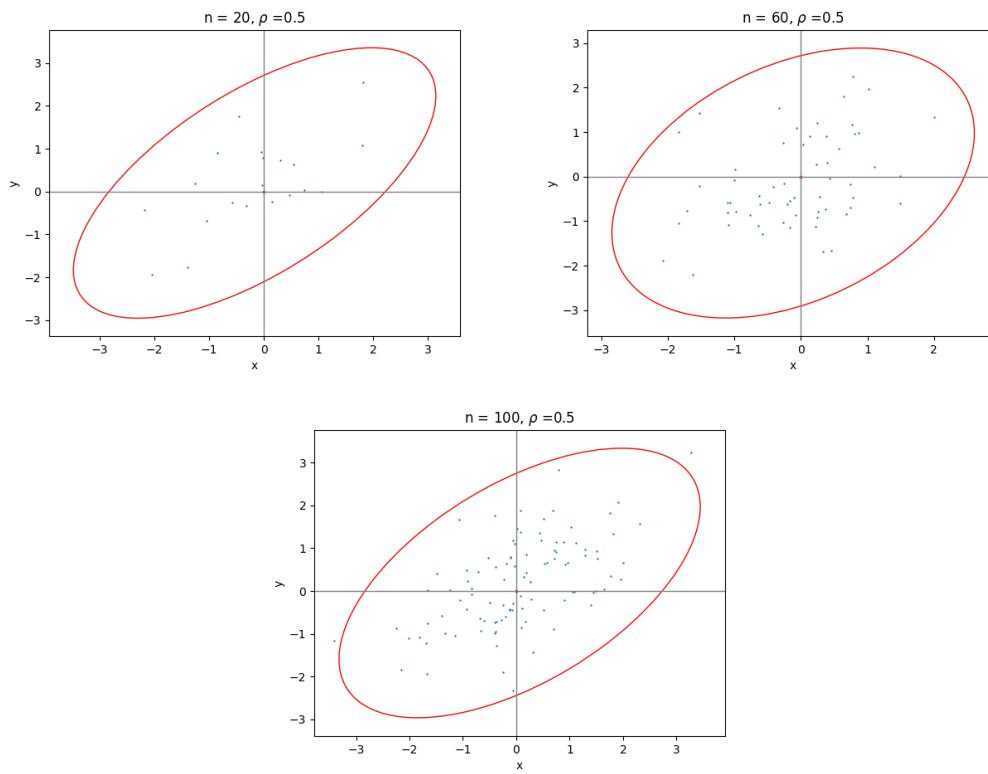


Рис. 2: Эллипсы рассеивания. $\rho = 0.5(4)$

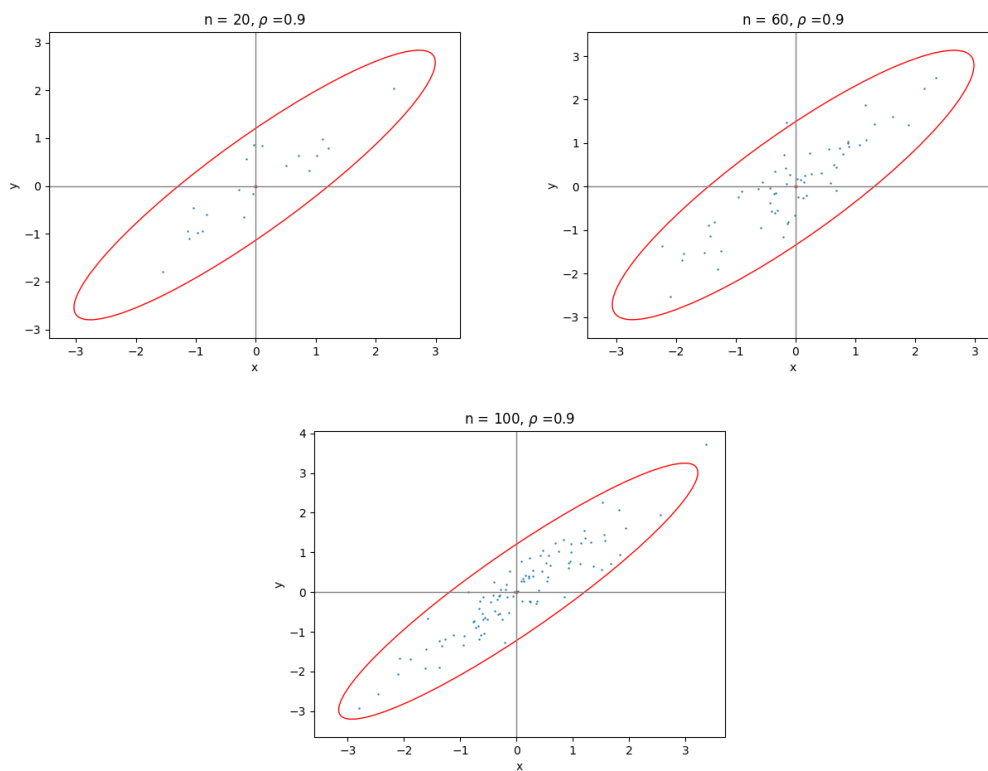


Рис. 3: Эллипсы рассеивания. $\rho = 0.9(4)$

4.3 Оценки коэффициентов линейной регрессии

4.3.1 Выборка без возмущений

- Критерий наименьших квадратов:

$$\hat{a} = 1.8986 \quad \hat{b} = 1.9533$$

- Критерий наименьших модулей:

$$\hat{a} = 1.9332 \quad \hat{b} = 1.6795$$

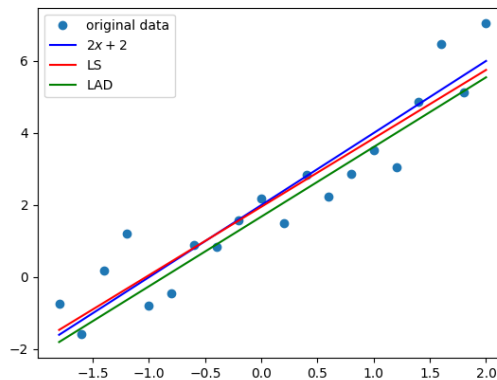


Рис. 4: Выборка без возмущений

4.3.2 Выборка с возмущениями

- Критерий наименьших квадратов:

$$\hat{a} = 0.7764 \quad \hat{b} = 1.9022$$

- Критерий наименьших модулей:

$$\hat{a} = 2.6911 \quad \hat{b} = 1.8674$$

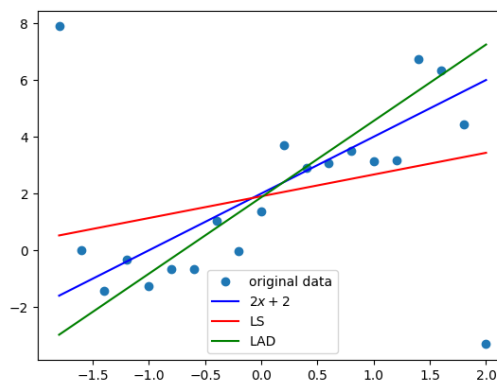


Рис. 5: Выборка с возмущениями

4.4 Проверка гипотезы о законе распределения генеральной совокупности. Метод хи-квадрат

Метод максимального правдоподобия:

$$\hat{\mu} = -0.039 \quad \hat{\sigma} = 1.098$$

Критерий согласия χ^2 :

- Количество промежутков $k = 6$.
- Уровень значимости $\alpha = 0.05$
- Квантиль $\chi^2_{1-\alpha}(k-1) = \chi^2_{0.95}(5) = 11.07$

i	Δ_i	n_i	p_i	np_i	$n_i - np_i$	$\frac{(n_i - np_i)^2}{np_i}$
1	$[-\infty, -3.337]$	0	0.0013	0.135	-0.135	-1.0
2	$[-3.337, -1.688]$	7	0.0655	6.5457	0.4543	0.0694
3	$[-1.688, -0.038]$	45	0.4332	43.3193	1.6807	0.0388
4	$[-0.038, 1.610]$	41	0.4332	43.3193	-2.3193	-0.0535
5	$[1.610, 3.259]$	7	0.0655	6.5457	0.4543	0.0694
6	$[3.259, \infty]$	0	0.013	0.135	-0.135	-1.0
Σ	—	100	1.000	100.0	0.00	$\chi^2_B = 0.387$

Таблица 5: Проверка гипотезы о законе $N(\hat{\mu}, \hat{\sigma})$ для выборки на 100 элементов с распределением $N(0, 1)$

i	Δ_i	n_i	p_i	np_i	$n_i - np_i$	$\frac{(n_i - np_i)^2}{np_i}$
1	$[-\infty, -3.337]$	0	0.0013	0.027	-0.027	-1.0
2	$[-3.337, -1.688]$	6	0.0655	1.3091	4.6909	3.5831
3	$[-1.688, -0.038]$	5	0.4332	8.6639	-3.6639	-0.4229
4	$[-0.038, 1.610]$	5	0.4332	8.6639	-3.6639	-0.4229
5	$[1.610, 3.259]$	4	0.0655	1.3091	2.6909	2.0554
6	$[3.259, \infty]$	0	0.013	0.027	-0.027	-1.0
Σ	—	20	1.000	20.0	0.00	$\chi^2_B = 25.465$

Таблица 6: Проверка гипотезы о законе $N(\hat{\mu}, \hat{\sigma})$ для выборки на 20 элементов с распределением $U([\hat{\mu} - 3\hat{\sigma}, \hat{\mu} + 3\hat{\sigma}])$

i	Δ_i	n_i	p_i	np_i	$n_i - np_i$	$\frac{(n_i - np_i)^2}{np_i}$
1	$[-\infty, -3.337]$	0	0.0013	0.027	-0.027	-1.0
2	$[-3.337, -1.688]$	0	0.0655	1.3091	-1.3091	-1.0
3	$[-1.688, -0.038]$	8	0.4332	8.6693	-0.6639	-0.0766
4	$[-0.038, 1.610]$	11	0.4332	8.6639	2.3361	0.2696
5	$[1.610, 3.259]$	1	0.0655	1.3091	-0.3091	-0.2361
6	$[3.259, \infty]$	0	0.013	0.027	-0.027	-1.0
Σ	—	20	1.000	20.0	0.00	$\chi^2_B = 2.0898$

Таблица 7: Проверка гипотезы о законе $N(\hat{\mu}, \hat{\sigma})$ для выборки на 20 элементов с распределением $L(\hat{\mu}, \frac{\hat{\sigma}}{\sqrt{2}})$

4.5 Исследование сигнала с использованием критерия Фишера

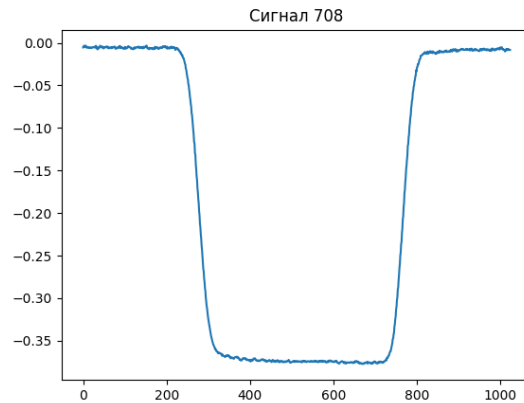


Рис. 6: Входной сигнал

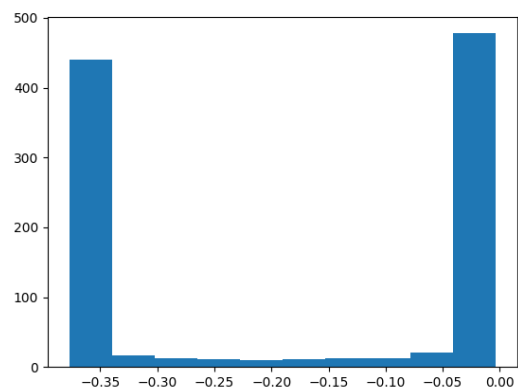


Рис. 7: Гистограмма сигнала

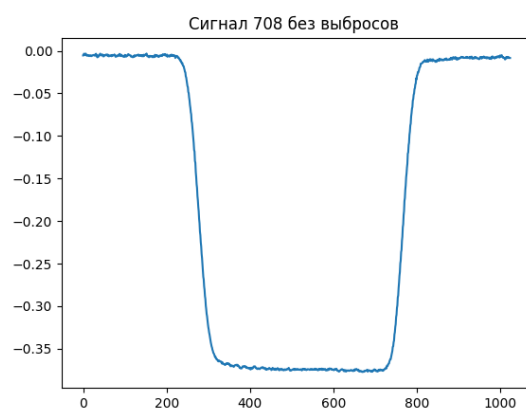


Рис. 8: Входной сигнал без возмущение

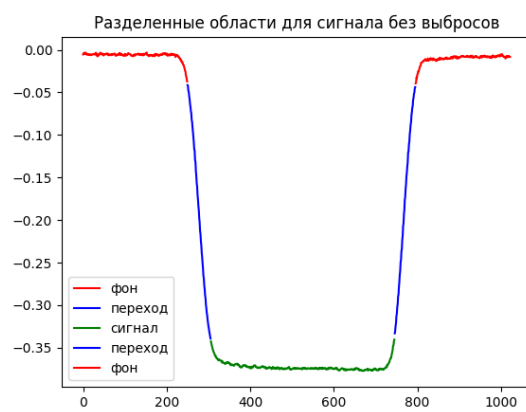


Рис. 9: Разделенные области сигнала

Промежуток	Тип	Количество разбиений	Критерий Фишера
1	Фон	5	0.186
2	Переход	4	20.5
3	Сигнал	4	0.047
4	Переход	5	20.33
5	Фон	4	0.195

Таблица 8: Характеристики выделенных областей

5 Обсуждение

5.1 Выборочные коэффициенты корреляции и эллипсы рассеивания

Для двумерного нормального распределения дисперсии выборочных коэффициентов корреляции упорядочены следующим образом: $D(r) < D(r_S) < D(r_Q)$. Аналогичные неравенства наблюдаются для величины $|\rho - r_{ind}|$, из чего можно сделать вывод, что оценка Пирсона коэффициента корреляции (7) является оптимальной в анализе двумерного нормального распределения.

Для смеси нормальных распределений наименьшая выборочная дисперсия наблюдается у коэффициента корреляции Спирмена. Кроме того коэффициент Спирмена наиболее устойчив к изменению размеров анализируемой выборки, из чего можно сделать вывод, что оценка Спирмена коэффициента корреляции (9) является оптимальной в анализе смеси нормальных распределений.

Процент попавших элементов выборки в эллипс рассеивания (95%-ная доверительная область) примерно равен его теоретическому значению.

5.2 Оценки коэффициентов линейной регрессии

Для сравнительно небольшой выборки ($n=20$) без возмущений критерий наименьших квадратов и критерий наименьших модулей дают сравнимые результаты (с небольшим выигрышем МНК в оценке коэффициента сдвига и небольшим выигрышем МНМ в оценке коэффициента наклона)

Метод наименьших модулей оказывается более устойчивым в оценке коэффициента наклона для выборки с возмущениями на краях.

5.3 Проверка гипотезы о нормальном распределении выборки методом хи-квадрат

По результатам проверки на близость с помощью критерия хи-квадрат можно принять гипотезу о нормальном распределении $N(\hat{\mu}, \hat{\sigma})$ на уровне значимости $\alpha = 0.05$ для выборки, сгенерированной согласно $N(0, 1)$.

Для небольшой выборки критерий хи-квадрат не почувствовал разницы между распределением $N(\hat{\mu}, \hat{\sigma})$ и распределением $L(\hat{\mu}, \frac{\hat{\sigma}}{\sqrt{2}})$, что можно объяснить малым размером выборки и схожестью законов при заданных параметрах.

Гипотезу о распределении выборки $U([\hat{\mu} - 3\hat{\sigma}, \hat{\mu} + 3\hat{\sigma}])$ по закону $N(\hat{\mu}, \hat{\sigma})$ метод отвергнул.

5.4 Исследование сигнала с использованием критерия Фишера

Промежутки 1, 3 и 5 (фон и сигнал) - области однородности, так как значение критерия Фишера близко к нулю. Промежутки 2 и 5 (переходы) - области неоднородности, так как значение критерия Фишера для них много больше единицы.

Репозиторий

<https://github.com/KoloskovAleksandr/MathStatLabs2021>

Список литературы

- [1] Максимов Ю.Д. Математика. Теория и практика по математической статистике. Конспект-справочник по теории вероятностей : учеб. пособие / Ю.Д. Максимов; под ред. В.И. Антонова. — СПб. : Изд-во Политехн. ун-та, 2009. — 395 с. (Математика в политехническом университете).