Санкт-Петербургский политехнический университет
Петра Великого

Физико-механический институт

Кафедра "Прикладная математика"

Отчет по лабораторным работам №5-8
по дисциплине
"Математическая статистика"

Выполнил студент:

Кротиков Сергей Ильич
группа: 5030102/90101

Проверил:

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

Санкт-Петербург, 2022

СОДЕРЖАНИЕ

СПИСОК ИЛЛЮСТРАЦИЙ

СПИСОК ТАБЛИЦ

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 квадратного 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. 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. Проделать 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.

2 Теория

2.1 Двумерное нормальное распределение

Двумерная случайная величина (X, Y) называется распределённой нормально (или просто нормальной), если её плотность вероятности определена формулой:

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

Компоненты X, Y двумерной нормальной случайной величины также распределены нормально с математическими ожиданиями $\bar{x}, \bar{y}$ и СКО σ_x, σ_y соответственно. Параметр ρ называется коэффициентом корреляции.

2.2 Корреляционный момент (ковариация) и коэффициент корреляции

Корреляционный момент, иначе *ковариация*, двух случайных величин X и Y:

$$K = cov(X, Y) = M[(X-\bar{x})(Y-\bar{y})] \quad (2)$$

Коэффициент корреляции ρ двух случайных величин X и Y:

$$\rho = \frac{K}{\sigma_x\sigma_y} \quad (3)$$

2.3 Выборочные коэффициенты корреляции

Выборочный коэффициент корреляции Пирсона:

$$r = \frac{\frac{1}{n}\sum(x_i-\bar{x})(y_i-\bar{y})}{\sqrt{\frac{1}{n}\sum(x_i-\bar{x})^2\frac{1}{n}\sum(y_i-\bar{y})^2}} = \frac{K}{s_X s_Y} \quad (4)$$

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

2.3.1 Выборочный квадратный коэффициент корреляции

Выборочный квадратный коэффициент корреляции:

$$r_Q = \frac{(n_1 + n_3) - (n_2 + n_4)}{n} \tag{5}$$

где n_1, n_2, n_3 и n_4 - количества точек с координатами (x_i, y_i), попавшими соответственно в I, II, III и IV квадранты декартовой системы с осями $x' = x - medx$, $y' = y - medy$ и с центром в точке с координатами $(medx, medy)$

2.3.2 Выборочный коэффициент ранговой корреляции Спирмена

Обозначим ранги, соответствующие значениям переменной X, через u, а ранги, соответствующие переменной Y, - через v.
Выборочный коэффициент корреляции Спирмена:

$$r = \frac{\frac{1}{n}\sum(u_i - \bar{u})(v_i - \bar{v})}{\sqrt{\frac{1}{n}\sum(u_i - \bar{u})^2 \frac{1}{n}\sum(v_i - \bar{v})^2}} \tag{6}$$

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

2.4 Эллипсы рассеивания

Уравнение проекции эллипса рассеивания на плоскость xOy:

$$\frac{(x - \bar{x})^2}{\sigma_x^2} - 2\rho\frac{(x - \bar{x})(y - \bar{y})}{\sigma_x \sigma_y} + \frac{(y - \bar{y})^2}{\sigma_y^2} = const \tag{7}$$

Центр эллипса (2.4) находится в точке с координатами $(\bar{x}, \bar{y})$; оси симметрии эллипса составляют с осью Ox углы, определяемые уравнением

$$tg2\alpha = \frac{2\rho\sigma_x\sigma_y}{\sigma_x^2 - \sigma_y^2} \tag{8}$$

2.5 Простая линейная регрессия

2.5.1 Модель простой линейной регрессии

Регрессионную модель описания данных называют простой линейной регрессией, если

$$y_i = \beta_0 + \beta_1 x_i + \varepsilon_i, i = 1..n \tag{9}$$

где $x_1, ..., x_n$— заданные числа (значения фактора); $y_1, ...y_n$— наблюдаемые значения отклика; $\varepsilon_1, ..., \varepsilon_n$— 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.
В модели (2.5.1) отклик y зависит от одного фактора x, и весь разброс экспериментальных точек объясняется только погрешностями наблюдений (результатов измерений) отклика y. Погрешности результатов измерений x в этой модели полагают существенно меньшими погрешностей результатов измерений y, так что ими можно пренебречь.

2.5.2 Метод наименьших квадратов

Метод наименьших квадратов (МНК) :

$$Q(\beta_0, \beta_1) = \sum_{i=1}^{n} \varepsilon_i^2 = \sum_{i=1}^{n} (y_i - \beta_0 - \beta_1 x_i)^2 \to \min_{\beta_0, \beta_1} \tag{10}$$

2.5.3 Расчётные формулы для МНК-оценок

МНК-оценки параметров β_0, β_1:

$$\hat{\beta}_1 = \frac{\overline{xy} - \bar{x}\bar{y}}{\overline{x^2} - (\bar{x})^2} \tag{11}$$

$$\hat{\beta}_0 = \bar{y} - \bar{x}\hat{\beta}_1 \tag{12}$$

2.6 Робастные оценки коэффициентов линейной регрессии

$$\sum_{i=1}^{n} |y_i - \beta_0 - \beta_1 x_i| \to \min_{\beta_0, \beta_1} \tag{13}$$

$$\hat{\beta}_{1R} = r_Q \frac{q_y^*}{q_x^*}, \tag{14}$$

$$\hat{\beta}_{0R} = medy - \hat{\beta}_{1R} medx, \tag{15}$$

$$r_Q = \frac{1}{n} \sum_{i=1}^{n} sgn(x_i - medx) sgn(y_i - medy), \tag{16}$$

$$q_y^* = \frac{y_{(j)} - y_{(l)}}{k_q(n)}, \quad q_x^* = \frac{x_{(j)} - x_{(l)}}{k_q(n)},$$
$$\begin{cases} [\frac{n}{4}] + 1 \text{ при } \frac{n}{4} \text{ дробном}, \\ \frac{n}{4} \text{ при } \frac{n}{4} \text{ целом}. \end{cases}$$
$$j = n - l + 1$$
$$sgn(z) = \begin{cases} 1 \text{ при } z > 0 \\ 0 \text{ при } z = 0 \\ -1 \text{ при } z < 0 \end{cases} \tag{17}$$

Уравнение регрессии здесь имеет вид

$$y = \hat{\beta}_{0R} + \hat{\beta}_{1R} x \tag{18}$$

2.7 Метод максимального правдоподобия

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

$$L(x_1, \ldots, x_n, \theta) = f(x_1, \theta) f(x_2, \theta) \ldots f(x_n, \theta) \tag{19}$$

Оценка максимального правдоподобия:

$$\hat{\theta}_{\text{мп}} = \arg\max_{\theta} L(x_1, \ldots, x_n, \theta) \tag{20}$$

Система уравнений правдоподобия (в случае дифференцируемости функции правдоподобия):

$$\frac{\partial L}{\partial \theta_k} \text{ или } \frac{\partial \ln L}{\partial \theta_k}, \quad k = 1, \ldots, m \tag{21}$$

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 распределение по методу χ^2

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

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

6. Сравниваем χ^2_B и квантиль $\chi^2_{1-\alpha}(k-1)$.

   - Если $\chi^2_B < \chi^2_{1-\alpha}(\mathrm{k}-1)$, то гипотеза H_0 на данном этапе проверки принимается.
   - 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 Доверительные интервалы для параметров нормального распределения

2.9.1 Доверительный интервал для математического ожидания m нормального распределения

Дана выборка $(x_1, x_2, \ldots, x_n)$ объёма n из нормальной генеральной совокупности. На её основе строим выборочное среднее $\bar{x}$ и выборочное среднее квадратичное отклонение s. Параметры m и σ нормального распределения неизвестны. Доверительный интервал для m с доверительной вероятностью $\gamma = 1-\alpha$:

$$
\begin{gathered}
P\left(\bar{x} - \frac{sx}{\sqrt{n-1}} < m < \bar{x} + \frac{sx}{\sqrt{n-1}}\right) = 2F_T(x) - 1 = 1-\alpha, \\
P\left(\bar{x} - \frac{st_{1-\alpha/2}(n-1)}{\sqrt{n-1}} < m < \bar{x} + -\frac{st_{1-\alpha/2}(n-1)}{\sqrt{n-1}}\right) = 1-\alpha
\end{gathered} \tag{22}
$$

2.9.2 Доверительный интервал для среднего квадратичного отклонения σ нормального распределения

Дана выборка $(x_1, x_2, \ldots, x_n)$ объёма n из нормальной генеральной совокупности. На её основе строим выборочную дисперсию s^2. Параметры m и σ нормального распределения неизвестны.
Задаёмся уровнем значимости α.
Доверительный интервал для σ с доверительной вероятностью $\gamma = 1-\alpha$:

$$
P\left(\frac{s\sqrt{n}}{\sqrt{\chi^2_{1-\alpha/2}(n-1)}} < \sigma < \frac{s\sqrt{n}}{\sqrt{\chi^2_{\alpha/2}(n-1)}}\right) = 1-\alpha \tag{23}
$$

2.10 Доверительные интервалы для математического ожидания m и среднего квадратичного отклонения σ произвольного распределения при большом объёме выборки. Асимптотический подход

При большом объёме выборки для построения доверительных интервалов может быть использован асимптотический метод на основе центральной предельной теоремы.

2.10.1 Доверительный интервал для математического ожидания m произвольной генеральной совокупности при большом объёме выборки

Предполагаем, что исследуемое генеральное распределения имеет конечные математическое ожидание m и дисперсию σ^2.
$u_{1-\alpha/2}$ - квантиль нормального распределения $N(0,1)$ порядка $1-\alpha/2$.
Доверительный интервал для m с доверительной вероятностью $\gamma = 1-\alpha$:

$$P\left(\bar{x} - \frac{su_{1-\alpha/2}}{\sqrt{n}} < m < \bar{x} + - \frac{st_{1-\alpha/2}(n-1)}{\sqrt{n-1}}\right) = \gamma \tag{24}$$

2.10.2 Доверительный интервал для среднего квадратичного отклонения σ произвольной генеральной совокупности при большом объёме выборки

Предполагаем, что исследуемая генеральная совокупность имеет конечные первые четыре момента.
$u_{1-\alpha/2}$ - квантиль нормального распределения $N(0,1)$ порядка $1-\alpha/2$
$E = \frac{\mu_4}{\sigma^4} - 3$ - эксцесс генерального распределения, $e = \frac{m_4}{s^4} - 3$ - выборочный эксцесс; $m_4 = \frac{1}{n}\sum_n^{i=1}(x_i - \bar{x})^4$ - четвёртый выборочный центральный момент.

$$s(1+U)^{-1/2} < \sigma < s(1-U)^{-1/2} \tag{25}$$

или

$$s(1-0.5U)^{-1/2} < \sigma < s(1+0.5U)^{-1/2} \tag{26}$$

где $U = u_{1-\alpha/2}\sqrt{(e+2)/n}$
Формулы (2.10.2) или (2.10.2) дают доверительный интервал для σ с доверительной вероятностью $\gamma = 1 - \alpha$

3 Реализация

Лабораторная работа выполнена на языке Python версия 3.7 в среде разработки JupyterLab. Использовались дополнительные библиотеки:

1. scipy (генерация выборок)
2. statsmodels, statistics (построение эмпирических функций распределения)
3. matplotlib (визуализация)
4. numpy (вычисление ряда числовых характеристик)

В приложении находится ссылка на GitHub репозиторий с исходным кодом.

4 Результаты

4.1 Выборочные коэффициенты корреляции

$\rho = 0$	$r(2.3)$	$r_S(2.3.2)$	$r_Q(2.3.1)$
$E(z)$	0.007	-0.003	0.0
$E(z^2)$	0.024	0.023	0.04
$D(z)$	0.05	0.048	0.05
$\rho = 0.5$	r	r_S	r_Q
$E(z)$	0.513	0.471	0.4
$E(z^2)$	0.263	0.222	0.16
$D(z)$	0.031	0.035	0.048
$\rho = 0.9$	r	r_S	r_Q
$E(z)$	0.907	0.883	0.8
$E(z^2)$	0.822	0.779	0.64
$D(z)$	0.003	0.005	0.03

Таблица 1: Двумерное нормальное распределение, n = 20

$\rho = 0$	r	r_S	r_Q
$E(z)$	0.008	0.011	0.0
$E(z^2)$	0.007	0.006	0.004
$D(z)$	0.017	0.016	0.017
$\rho = 0.5$	r	r_S	r_Q
$E(z)$	0.501	0.476	0.333
$E(z^2)$	0.251	0.226	0.111
$D(z)$	0.01	0.011	0.016
$\rho = 0.9$	r	r_S	r_Q
$E(z)$	0.901	0.887	0.733
$E(z^2)$	0.811	0.786	0.538
$D(z)$	0.001	0.001	0.009

Таблица 2: Двумерное нормальное распределение, n = 60

$\rho = 0$	r	r_S	r_Q
$E(z)$	0.003	0.001	0.0
$E(z^2)$	0.005	0.004	0.006
$D(z)$	0.01	0.01	0.01
$\rho = 0.5$	r	r_S	r_Q
$E(z)$	0.501	0.483	0.32
$E(z^2)$	0.251	0.233	0.102
$D(z)$	0.006	0.007	0.009
$\rho = 0.9$	r	r_S	r_Q
$E(z)$	0.9	0.887	0.72
$E(z^2)$	0.811	0.787	0.518
$D(z)$	0.0	0.001	0.005

Таблица 3: Двумерное нормальное распределение, n = 100

$n = 20$	r	r_S	r_Q
$E(z)$	0.803	0.883	0.6
$E(z^2)$	0.645	0.779	0.36
$D(z)$	0.009	0.005	0.041
$n = 60$	r	r_S	r_Q
$E(z)$	0.792	0.887	0.6
$E(z^2)$	0.628	0.786	0.36
$D(z)$	0.003	0.001	0.011
$n = 100$	r	r_S	r_Q
$E(z)$	0.79	0.887	0.56
$E(z^2)$	0.625	0.787	0.314
$D(z)$	0.002	0.001	0.007

Таблица 4: Смесь нормальных распределений

4.2 Эллипсы рассеивания

Для уравнения эллипса выбиралась константа равная $const = 2 \cdot (2 \cdot \sigma)$

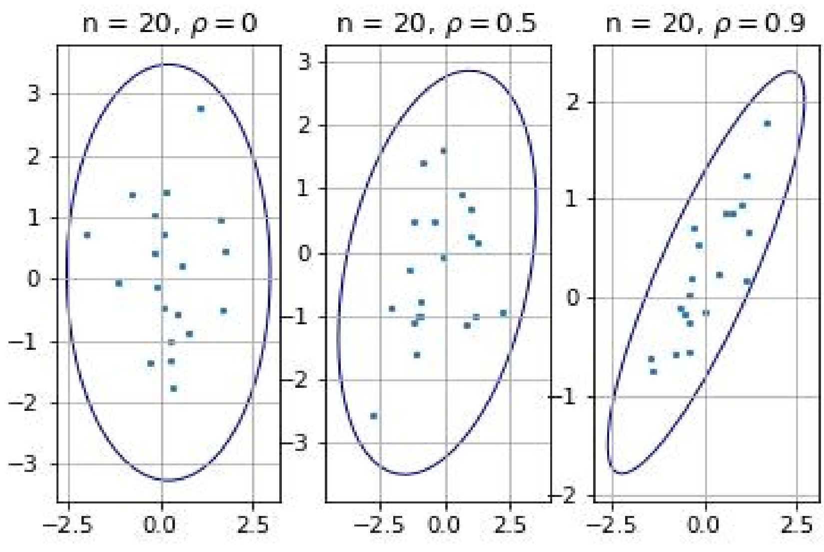

Рис. 1: Двумерное нормальное распределение, $n = 20$

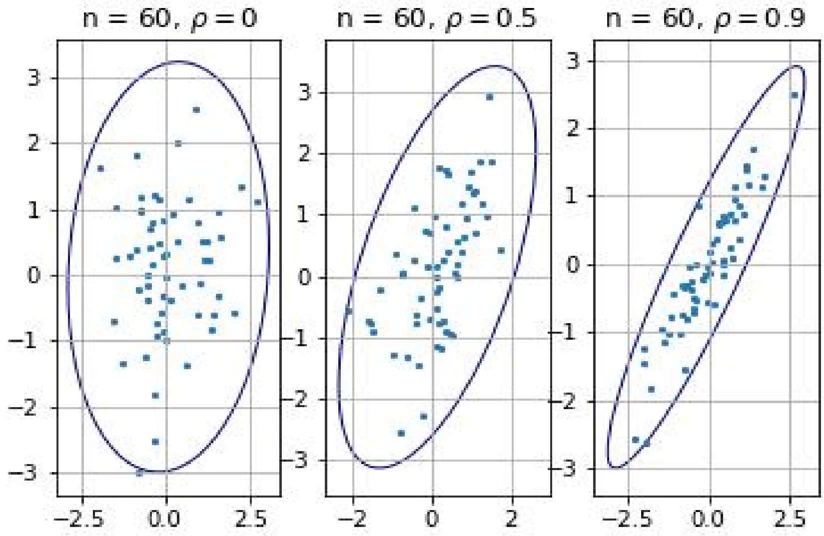

Рис. 2: Двумерное нормальное распределение, $n = 60$

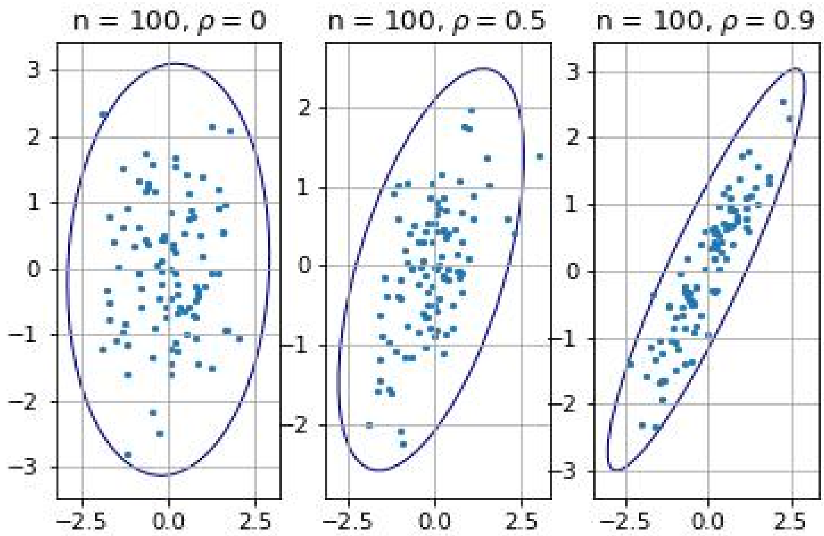

Рис. 3: Двумерное нормальное распределение, $n = 100$

4.3 Оценка коэффициентов линейной регрессии

4.3.1 Выборка без возмущений

- Критерий наименьших квадратов:

$$\hat{a} \approx 2.03, \quad \hat{b} \approx 1.73$$

- Критерий наименьших модулей:

$$\hat{a} \approx 1.89, \quad \hat{b} \approx 1.62$$

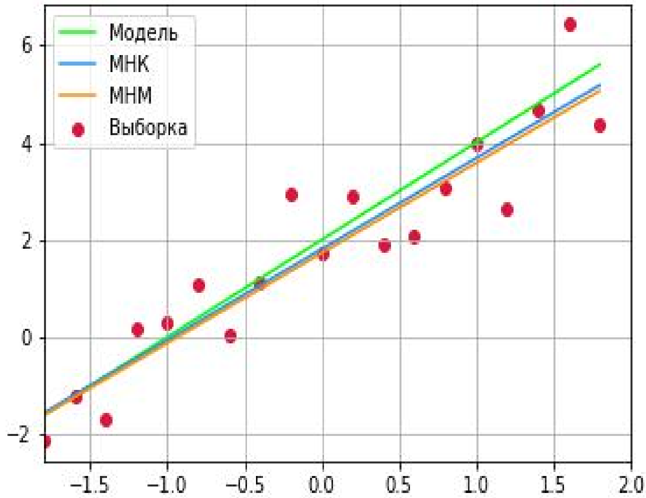

Рис. 4: Выборка без возмущений

МНК $distance \ = \ 0.49$
МНМ $distance \ = \ 2.31$

4.3.2 Выборка с возмущениями

- Критерий наименьших квадратов:

$$\hat{a} \approx 2.06, \quad \hat{b} \approx 0.44$$

- Критерий наименьших модулей:

$$\hat{a} \approx 1.68, \quad \hat{b} \approx 1.91$$

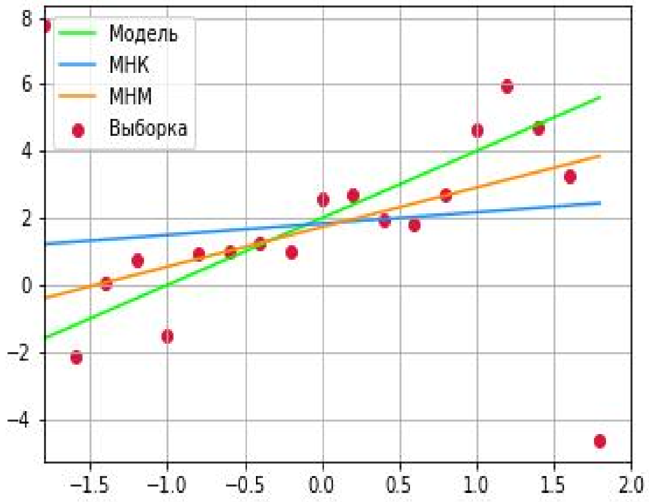

Рис. 5: Выборка без возмущений

МНК $distance = 66.24$
МНМ $distance = 2.97$

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

Метод максимального правдоподобия:

$$\hat{\mu} \approx 0.09, \quad \hat{\sigma} \approx 0.99$$

Критерий согласия χ^2:

- Количество промежутков $k = 8$
- Уровень значимости $\alpha = 0.05$
- Тогда квантиль $\chi^2_{1-\alpha}(k-1) = \chi^2_{0.95}(7)$. Из таблицы $\chi^2_{0.95}(7) \approx 14.07$.

i	$limits$	n_i	p_i	np_i	$n_i - np_i$	$\frac{(n_i-np_i)^2}{np_i}$
1	[-∞, -1.1]	15	0.1357	13.57	1.43	0.15
2	[-1.1, -0.73]	7	0.096	9.6	-2.6	0.7
3	[-0.73, -0.37]	8	0.1253	12.53	-4.53	1.64
4	[-0.37, 0.0]	13	0.1431	14.31	-1.31	0.12
5	[0.0, 0.37]	19	0.1431	14.31	4.69	1.54
6	[0.37, 0.73]	15	0.1253	12.53	2.47	0.49
7	[0.73, 1.1]	9	0.096	9.6	-0.6	0.04
8	[1.1, ∞]	14	0.1357	13.57	0.43	0.01
$\sum$	-	100	1	100	0	$4.69 = \chi_B^2$

Таблица 5: Вычисление χ_B^2 при проверке гипотезы H_0 о нормальном законе распределения $N(x, \hat{\mu}, \hat{\sigma})$

Сравнивая $\chi_B^2 = 4.69$ и $\chi_{0.95}^2(7) \approx 14.07$, видим, что $\chi_B^2 < \chi_{0.95}^2(7)$.

Исследование на чувствительность

Рассмотрим гипотезу H_0^*, что выборка распределена согласно закону $Laplace(x, \hat{\mu}, \frac{\hat{\sigma}}{\sqrt{2}})$.

Используем критерий согласия χ^2:

- Уровень значимости α= 0.05
- $n = 20$ - размер выборки
- $k := \lfloor 1 + 3.3 \lg 20 \rfloor = \lfloor 5.3 \rfloor = 5$ - количество промежутков
- Квантиль $\chi_{1-\alpha}^2(k-1) = \chi_{0.95}^2(4) \approx 9.49$

i	$limits$	n_i	p_i	np_i	$n_i - np_i$	$\frac{(n_i-np_i)^2}{np_i}$
1	[-∞, -1.1]	3	0.1357	2.71	0.29	0.03
2	[-1.1, -0.37]	4	0.2213	4.43	-1.43	0.46
3	[-0.37, 0.37]	8	0.2861	5.72	0.28	0.01
4	[0.37, 1.1]	3	0.2213	4.43	1.57	0.56
5	[1.1, ∞]	2	0.1357	2.71	-0.71	0.19
$\sum$	-	20	1	20	0	$1.25 = \chi_B^2$

Таблица 6: Вычисление χ_B^2 при проверке гипотезы H_0 о законе распределения $L(x, \hat{\mu}, \hat{\sigma})$, $n = 20$

Сравнивая $\chi_B^2 = 1.25$ и $\chi_{0.95}^2(4) \approx 9.49$, видим, что $\chi_B^2 < \chi_{0.95}^2(4)$.
Проведём аналогичный анализ для равномерного распределения

i	$limits$	n_i	p_i	np_i	$n_i - np_i$	$\frac{(n_i-np_i)^2}{np_i}$
1	[-∞, -1.1]	3	0.1357	2.71	0.29	0.03
2	[-1.1, -0.37]	7	0.2213	4.43	2.57	1.5
3	[-0.37, 0.37]	2	0.2861	5.72	-3.72	2.42
4	[0.37, 1.1]	4	0.2213	4.43	-0.43	0.04
5	[1.1, ∞]	4	0.1357	2.71	1.29	0.61
$\sum$	-	20	1	20	0	$4.6 = \chi_B^2$

Таблица 7: Вычисление χ_B^2 при проверке гипотезы H_0 о законе распределения $U(x, \hat{\mu}, \hat{\sigma})$, $n = 20$

Сравнивая $\chi_B^2 = 4.6$ и $\chi_{0.95}^2(4) \approx 9.49$, видим, что $\chi_B^2 < \chi_{0.95}^2(4)$.

4.5 Доверительные интервалы для параметров нормального распределения

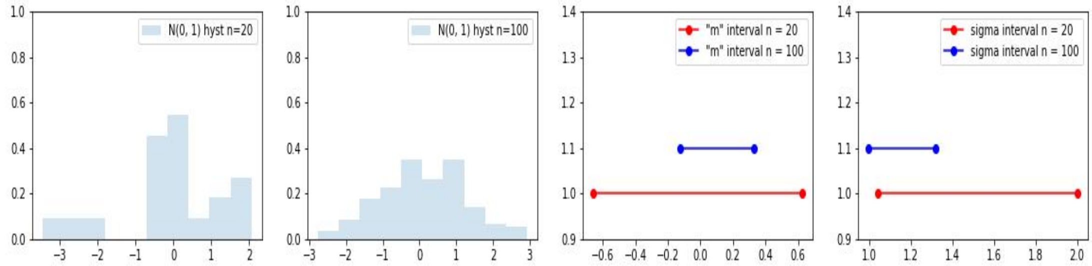

Рис. 6: Гистограммы нормальных распределений и доверительные интервалы их параметров

$n = 20$	m	σ
	$-0.66 < m < 0.62$	$1.04 < \sigma < 2$
$n = 100$	m	σ
	$-0.12 < m < 0.33$	$1.00 < \sigma < 1.36$

Таблица 8: Доверительные интервалы для параметров нормального распределения

4.6 Доверительные интервалы для параметров произвольного распределения. Асимптотический подход.

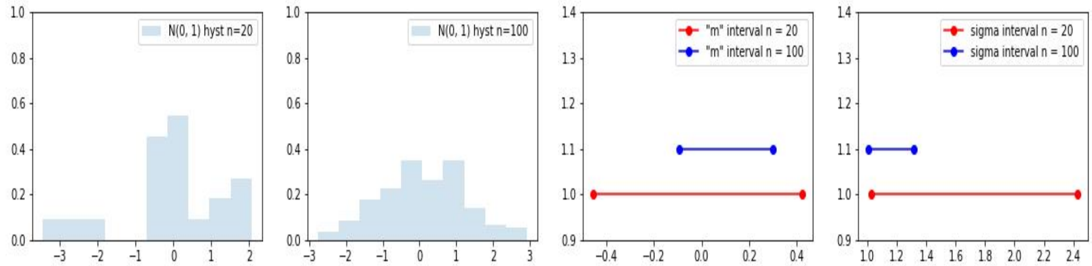

Рис. 7: Гистограммы нормальных распределений и доверительные интервалы их параметров. Асимптотический подход

$n = 20$	m	σ
	$-0.45 < m < 0.42$	$1.03 < \sigma < 2.43$
$n = 100$	m	σ
	$-0.09 < m < 0.30$	$1.01 < \sigma < 1.31$

Таблица 9: Доверительные интервалы для параметров нормального распределения

5 Обсуждение

5.1 Ядерные оценки плотности распределения

Для двумерного нормального распределения дисперсии выборочных коэффициентов корреляции упорядочены следующим образом: $r < r_S < r_Q$; для смеси распределений получили обратную картину: $r_Q < r_S < r$.
Процент попавших элементов выборки в эллипс рассеивания (95%-ная доверительная область) примерно равен его теоретическому значению (95%).

5.2 Оценки коэффициентов линейной регрессии

По полученным результатам (см. метрику удаленности модельной прямой от теоретической - *distance*) можно сказать, что используя критерий наименьших квадратов удастся точнее оценить коэффициенты линейной регрессии для выборки без возмущений. Если же редкие возмущения присутствуют, тогда лучше использовать критерий наименьших модулей.

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

Заключаем, что гипотеза H_0 о нормальном законе распределения $N(x, \hat{\mu}, \hat{\sigma})$ на уровне значимости $\alpha = 0.05$ согласуется с выборкой для нормального распределения $N(x, 0, 1)$.
Также видно, что для выборок сгенерированных по равномерному закону и закону Лапласа гипотеза H_0 оказалась принята.

5.4 Доверительные интервалы для параметров произвольного распределения

- Генеральные характеристики ($m = 0$ и $\sigma = 1$) накрываются построенными доверительными интервалами.
- Также можно сделать вывод, что для большей выборки доверительные интервалы являются соотвественно более точными, т.е. меньшими по длине.
- Кроме того, при большом объёме выборки асимптотические и классические оценки практически совпадают.

6 Приложение

Код программы GitHub URL:

https://github.com/Krotikov/matStat