Санкт–Петербургский политехнический университет
Петра Великого
Физико-механический институт
Кафедра «Прикладная математика»

Отчёт
по лабораторным работам №5-8
по дисциплине
«Математическая статистика»

Выполнил студент:
Турченко Михаил Константинович
группа: 5030102/90101

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

Санкт-Петербург
2021г.

Оглавление

1. Постановка задачи

1. Сгенерировать двумерные выборки размерами 20, 60, 100 для нормального двумерного распределения $\mathrm{N}(\mathrm{x}, \mathrm{y}, 0, 0, 1, 1, \rho)$. 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: $f(x, y)$ = $0.9\mathrm{N}(\mathrm{x}, \mathrm{y}, 0, 0, 1, 1, 0.9)$ + $0.1\mathrm{N}(\mathrm{x}, \mathrm{y}, 0, 0, 10, 10, -0.9)$. 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 α = 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. Для двух выборок размерами 20 и 100 элементов, сгенерированных согласно нормальному закону $N(x$, 0, 1), для параметров положения и масштаба построить асимптотически нормальные интервальные оценки на основе точечных оценок метода максимального правдоподобия и классические интервальные оценки на основе статистик χ^2 и Стьюдента. В качестве параметра надёжности взять γ = 0.95.

2. Теория

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

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

Компоненты X, Y двумерной нормальной случайной величины также распределены нормально с математическими ожиданиями $\overline{x}, \overline{y}$ и средними квадратическими отклонениями 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:

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

Коэффициент корреляции ρ двух случайных величин X и Y:

$$\rho = \frac{K}{\sigma_x\sigma_y}$$

2.3 Выборочные коэффициенты корреляции

2.3.1 Выборочный коэффициент корреляции Пирсона

Выборочный коэффициент корреляции Пирсона:

$$r = \frac{\frac{1}{n}\Sigma(\mathrm{x}_i - \overline{x})(\mathrm{y}_i - \overline{y})}{\sqrt{\frac{1}{n}\Sigma(\mathrm{x}_i - \overline{x})^2 * \frac{1}{n}\Sigma(\mathrm{y}_i - \overline{y})^2}} = \frac{K}{\mathrm{s}_x\mathrm{s}_y}$$

где K, s_x^2, s_y^2— выборочные ковариация и дисперсии с.в. X и Y

2.3.2 Выборочный квадрантный коэффициент корреляции

Выборочный квадрантный коэффициент корреляции

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

где n_1, n_2, n_3 и n_4 — количества точке с координатами (x_i, y_i), попавшими соответственно в I, II, III и IV квадранты декартовой системы с осями x' = x − med x, y' = y − med y и с центром в точке с координатами (med x, 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:

$$r_S = \frac{\frac{1}{n}\Sigma(\mathrm{u}_i - \overline{u})(\mathrm{v}_i - \overline{v})}{\sqrt{\frac{1}{n}\Sigma(\mathrm{u}_i - \overline{u})^2 * \frac{1}{n}\Sigma(\mathrm{v}_i - \overline{v})^2}}$$

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

2.4 Эллипсы рассеивания

Уравнение проекции эллипса рассеивания на плоскость xOy:

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

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

$$tg\ 2\alpha = \frac{2\rho\sigma_x\sigma_y}{\sigma_x^2 - \sigma_y^2}$$

2.5 Простая линейная регрессия

2.5.1 Модель простой линейной регрессии

Регрессионную модель описания данных называют простой линейной регрессией, если

$$\mathrm{y}_i = \beta_0 + \beta_1\mathrm{x}_i + \varepsilon_i$$

где x_i — заданные числа (значения фактора); y_i — наблюдаемые значения отклика; ε_i — независимые, нормально распределённые $N(0, \sigma)$ 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):

$$Q(\beta_0, \beta_1) = \sum \varepsilon_i^2 = \sum (\mathrm{y}_i - \beta_0 - \beta_1 \mathrm{x}_i)^2 \to \min_{\beta_0, \beta_1}$$

2.5.3 Расчётные формулы для МНК-оценок МНК-оценки параметров:

$$\widehat{\beta_1} = \frac{\overline{\boldsymbol{xy}} - \overline{\boldsymbol{x}} * \overline{\boldsymbol{y}}}{\overline{\boldsymbol{x}^2} - \overline{\boldsymbol{x}}^2}$$

$$\widehat{\beta_0} = \overline{\boldsymbol{y}} - \overline{\boldsymbol{x}}\widehat{\beta_1}$$

2.6 Робастные оценки коэффициентов линейной регрессии

Метод наименьших модулей:

$$\sum |\mathrm{y}_i - \beta_0 - \beta_1 \mathrm{x}_i| \to \min_{\beta_0, \beta_1}$$

$$\widehat{\beta_{1R}} = \boldsymbol{r_Q} \frac{\boldsymbol{q_y} *}{\boldsymbol{q_x} *}$$

$$\widehat{\beta_{0R}} = \boldsymbol{med\, y} - \widehat{\beta_{1R}} * \boldsymbol{med\, x}$$

$$\boldsymbol{r_Q} = \frac{\boldsymbol{1}}{\boldsymbol{n}} \sum \boldsymbol{sgn}(\mathrm{x}_i - \boldsymbol{med\, x}) * \boldsymbol{sgn}(\mathrm{y}_i - \boldsymbol{med\, y})$$

$$\boldsymbol{q_y} * = \frac{\boldsymbol{y}_{(j)} - \boldsymbol{y}_{(l)}}{\boldsymbol{k_q}(\boldsymbol{n})}; \ \boldsymbol{q_x} * = \frac{\boldsymbol{x}_{(j)} - \boldsymbol{x}_{(l)}}{\boldsymbol{k_q}(\boldsymbol{n})}$$

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

$$sgn\, z = \begin{cases} 1, z > 0 \\ 0, z = 0 \\ -1, z < 0 \end{cases}$$

Уравнение регрессии: $y = \widehat{\beta_{0R}} + \widehat{\beta_{1R}} * \boldsymbol{x}$

2.7 Метод максимального правдоподобия

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

$$L(\boldsymbol{x_1}, \dots, \boldsymbol{x_n}, \theta) = f(\boldsymbol{x_1}, \theta) f(\boldsymbol{x_2}, \theta) \dots f(\boldsymbol{x_n}, \theta)$$

Оценка максимального правдоподобия:

$$\widehat{\theta_{\text{мп}}} = arg \max_{\theta} L(x_1, \dots, x_n, \theta)$$

Система уравнений правдоподобия (в случае дифференцируемости функции правдоподобия):

$$\frac{\partial L}{\partial \theta_{\boldsymbol{k}}} = 0 \text{ или } \frac{\partial\, lnL}{\partial \theta_{\boldsymbol{k}}} = 0$$

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. Находим квантиль $\chi^2_{1-\alpha}(k-1)$ распределения хи-квадрат с k – 1 степенями свободы порядка 1 – α.

3. С помощью гипотетической функции распределения $F(x)$ вычисляем вероятности p_i = P ($X \in \Delta i$),

4. Находим частоты n_i попадания элементов выборки в подмножества Δi , i = 1, ... , k

5. Вычисляем выборочное значение статистики критерия χ^2:

$$\chi^2_\text{В} = \frac{(n_i - np_i)^2}{np_i}$$

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

а) Если $\chi^2_\text{В} < \chi^2_{1-\alpha}(k-1)$, то гипотеза H_0 на данном этапе проверки принимается.

б) Если $\chi^2_\text{В} \geq \chi^2_{1-\alpha}(k-1)$, 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 $(\text{x}_1, \text{x}_2, \dots, \text{x}_n)$ объёма n из нормальной генеральной совокупности. На её основе строим выборочное среднее $\overline{x}$ и выборочное среднее квадратическое отклонение s. Параметры m и σ нормального распределения неизвестны.

Доверительный интервал для m с доверительной вероятностью γ = 1 – α:

$$P(\overline{x} - \frac{sx}{\sqrt{n-1}} < m < \overline{x} + \frac{sx}{\sqrt{n-1}}) = 1 - \alpha$$

2.9.2 Доверительный интервал для среднего квадратического отклонения σ нормального распределения

Дана выборка $(\text{x}_1, \text{x}_2, \dots, \text{x}_n)$ объёма n из нормальной генеральной совокупности. На её основе строим выборочную дисперсию. Параметры m и σ нормального распределения неизвестны.

Задаёмся уровнем значимости α.

Доверительный интервал для σ с доверительной вероятностью γ = 1 – α:

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

2.10 Доверительные интервалы для математического ожидания m и среднего квадратического 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 распределение имеет конечные математическое ожидание и дисперсию.

$u_{1-\frac{\alpha}{2}}$ — квантиль нормального распределения $N(0, 1)$ порядка $1 - \alpha/2$.
Доверительный интервал для m с доверительной вероятностью $\gamma = 1 - \alpha$:

$$P\left(\overline{x} - \frac{su_{1-\frac{\alpha}{2}}}{\sqrt{n}} < m < \overline{x} + \frac{su_{1-\frac{\alpha}{2}}}{\sqrt{n}}\right) = 1 - \alpha$$

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

Предполагаем, что исследуемая генеральная совокупность имеет конечные первые четыре момента.

$u_{1-\frac{\alpha}{2}}$ – квантиль нормального распределения $N(0, 1)$ порядка $1 - \alpha/2$.

$e = \frac{\frac{1}{n}\Sigma(x_i - \overline{x})^4}{s^4}$ — выборочный эксцесс;

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

или

$$s(1-0.5U) \ < \ \sigma \ < \ s(1+0.5U)$$

где $U = u_{1-\frac{\alpha}{2}}\sqrt{(e+2)/n}$

Эти формулы дают доверительный интервал для σ с доверительной вероятностью $\gamma = 1 - \alpha$

3. Реализация

Лабораторная работа выполнена с помощью встроенных средств языка программирования Python в среде разработки Microsoft Visual Studio. Исходный код лабораторной работы – [MathStat/Labs 5-8 at main · TurchenkoMikhail/MathStat (github.com)]

4 Результаты

4.1 Выборочные коэффициенты корреляции

P = 0	r	r_s	r_q
E(z)	-0.009	-0.01	0.001
$E(z^2)$	0.053	0.051	0.052
D(z)	0.053	0.051	0.052
P = 0.5	r	r_s	r_q
E(z)	0.492	0.462	0.328
$E(z^2)$	0.273	0.25	0.156
D(z)	0.031	0.037	0.048
P = 0.9	r	r_s	r_q
E(z)	0.895	0.865	0.688
$E(z^2)$	0.803	0.754	0.5
D(z)	0.003	0.005	0.027

Таблица 1: Двумерное нормальное распределение, n = 20

P = 0	r	r_s	r_q
E(z)	0.01	0.009	0.008
$E(z^2)$	0.017	0.018	0.016
D(z)	0.017	0.017	0.016
P = 0.5	r	r_s	r_q
E(z)	0.493	0.472	0.33
$E(z^2)$	0.253	0.234	0.125
D(z)	0.01	0.011	0.016

P = 0.9	r	r_s	r_q
E(z)	0.9	0.884	0.707
E(z^2)	0.811	0.782	0.509
D(z)	0.001	0.001	0.008

Таблица 2: Двумерное нормальное распределение, n = 60

P = 0	r	r_s	r_q
E(z)	-0.002	-0.003	-0.002
E(z^2)	0.01	0.01	0.01
D(z)	0.01	0.01	0.01
P = 0.5	r	r_s	r_q
E(z)	0.503	0.482	0.333
E(z^2)	0.258	0.239	0.12
D(z)	0.006	0.006	0.009
P = 0.9	r	r_s	r_q
E(z)	0.9	0.887	0.708
E(z^2)	0.81	0.788	0.507
D(z)	0	0.001	0.005

Таблица 3: Двумерное нормальное распределение, n = 100

N = 20	r	r_s	r_q
E(z)	0.776	0.603	0.33
E(z^2)	0.607	0.375	0.13
D(z)	0.006	0.011	0.02
N = 60	r	r_s	r_q
E(z)	0.788	0.615	0.33
E(z^2)	0.622	0.382	0.115
D(z)	0.002	0.003	0.006

N = 100	r	r_s	r_q
E(z)	0.789	0.616	0.331
$E(z^2)$	0.623	0.381	0.113
D(z)	0.001	0.001	0.004

Таблица 4: Смесь нормальных распределений

4.2 Эллипсы рассеивания

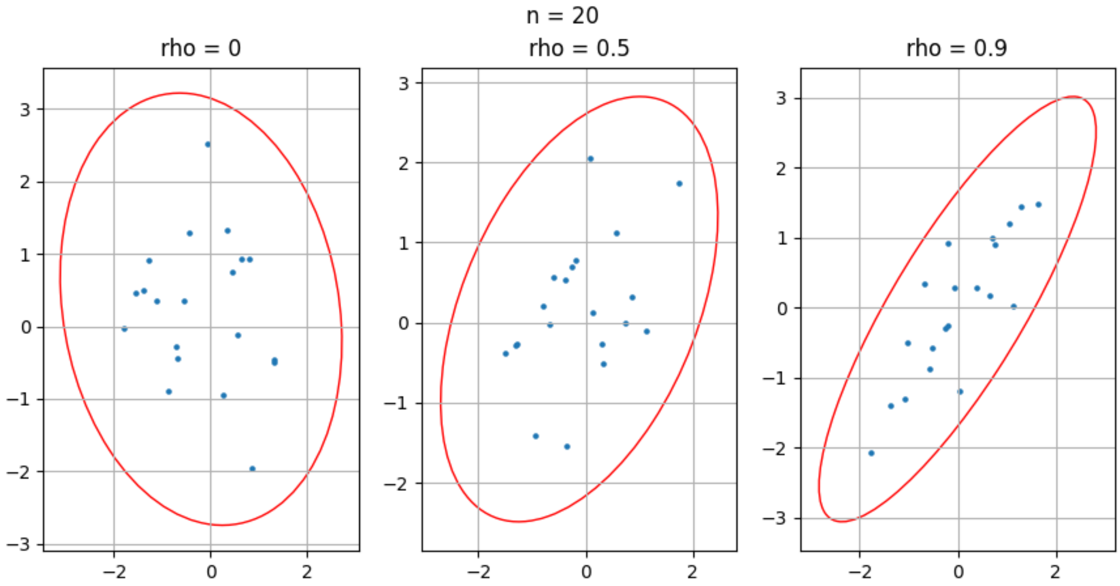

Рис.1 Эллипсы равновероятности нормального двумерного распределения, n=20

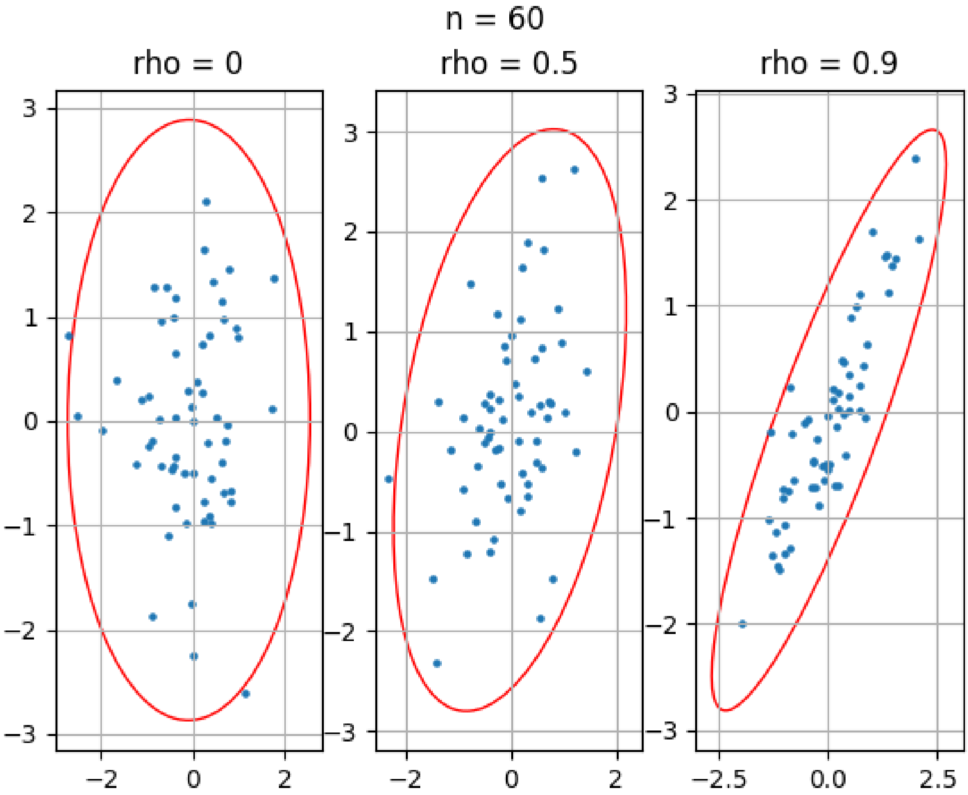

Рис.2 Эллипсы равновероятности нормального двумерного распределения, n=60

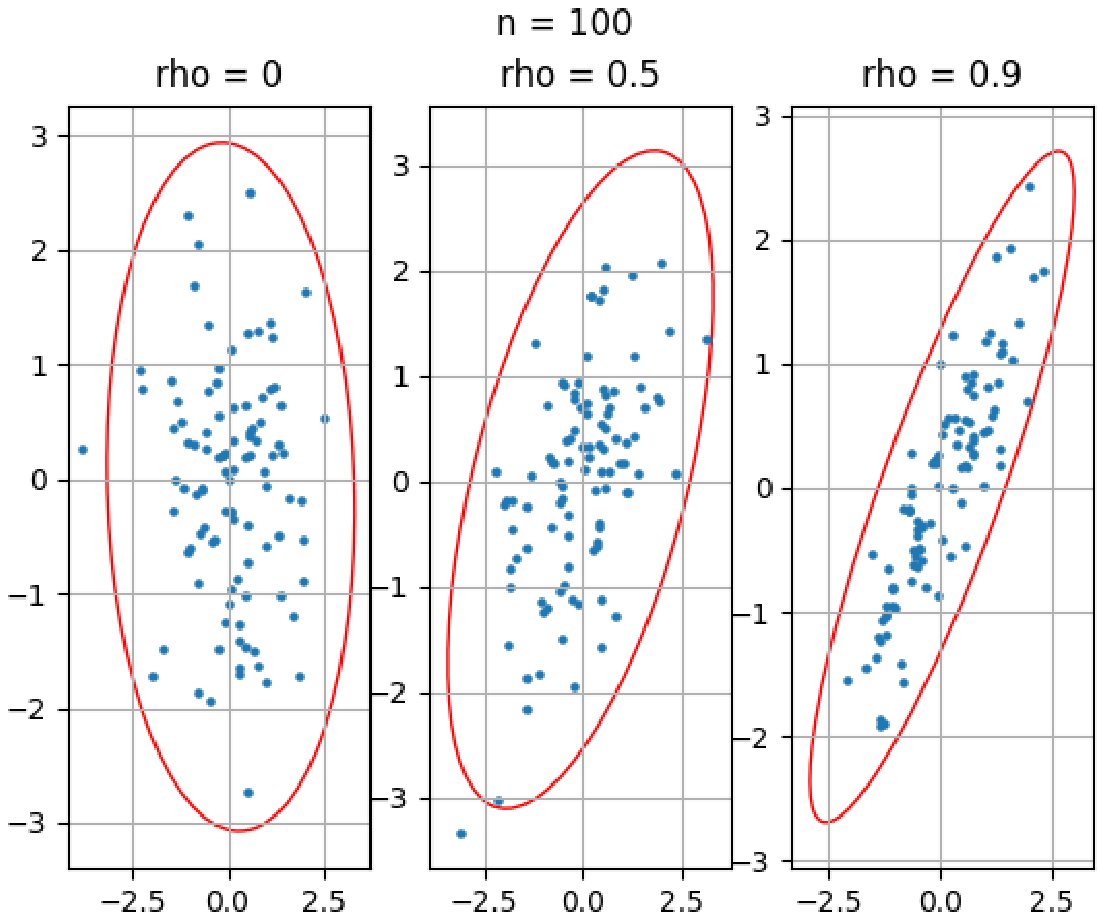

Рис.3 Эллипсы равновероятности нормального двумерного распределения, n=100

4.3 Оценки коэффициентов линейной регрессии

4.3.1 Выборка без возмущений

Критерий наименьших квадратов: $\hat{a} \approx 2.12$, $\hat{b} \approx$ 2.08·

Критерий наименьших модулей: $\hat{a} \approx 2.22$, $\hat{b} \approx 2.35$

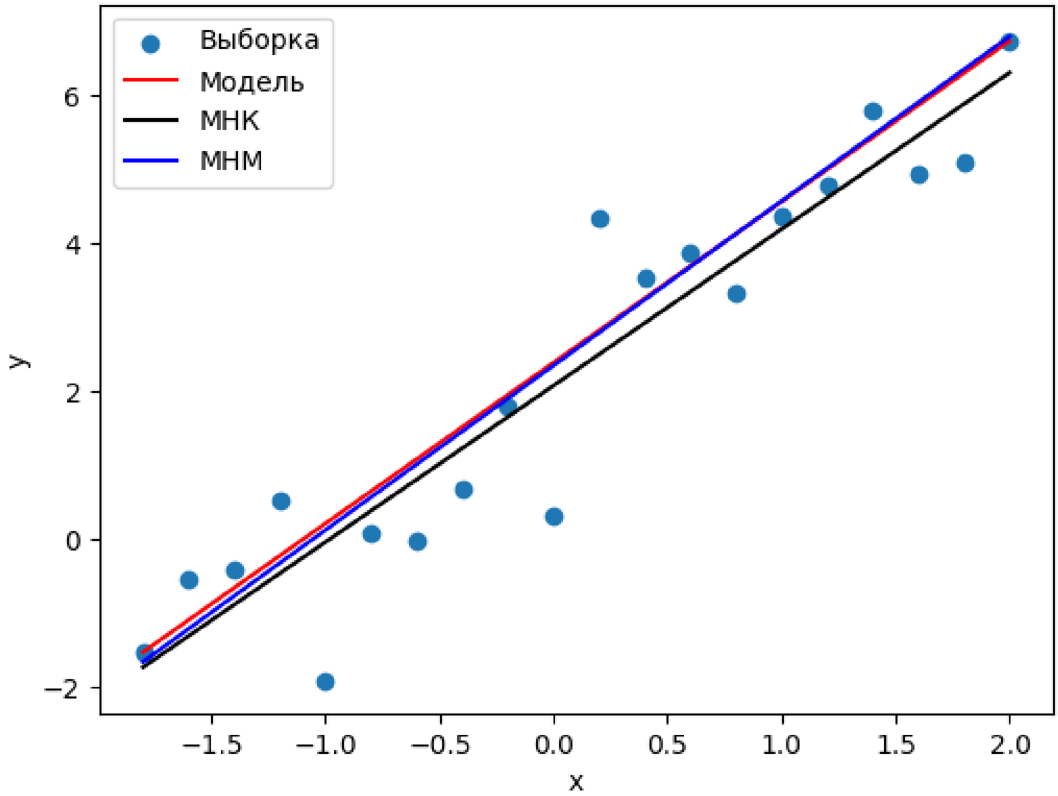

Рис.4 МНК и МНМ для выборки без возмущений

4.3.2 Выборка с возмущениями

Критерий наименьших квадратов: $\hat{a} \approx 0.59$, $\hat{b} \approx 2.47$ ·

Критерий наименьших модулей: $\hat{a} \approx 1.08$, $\hat{b} \approx 2.71$

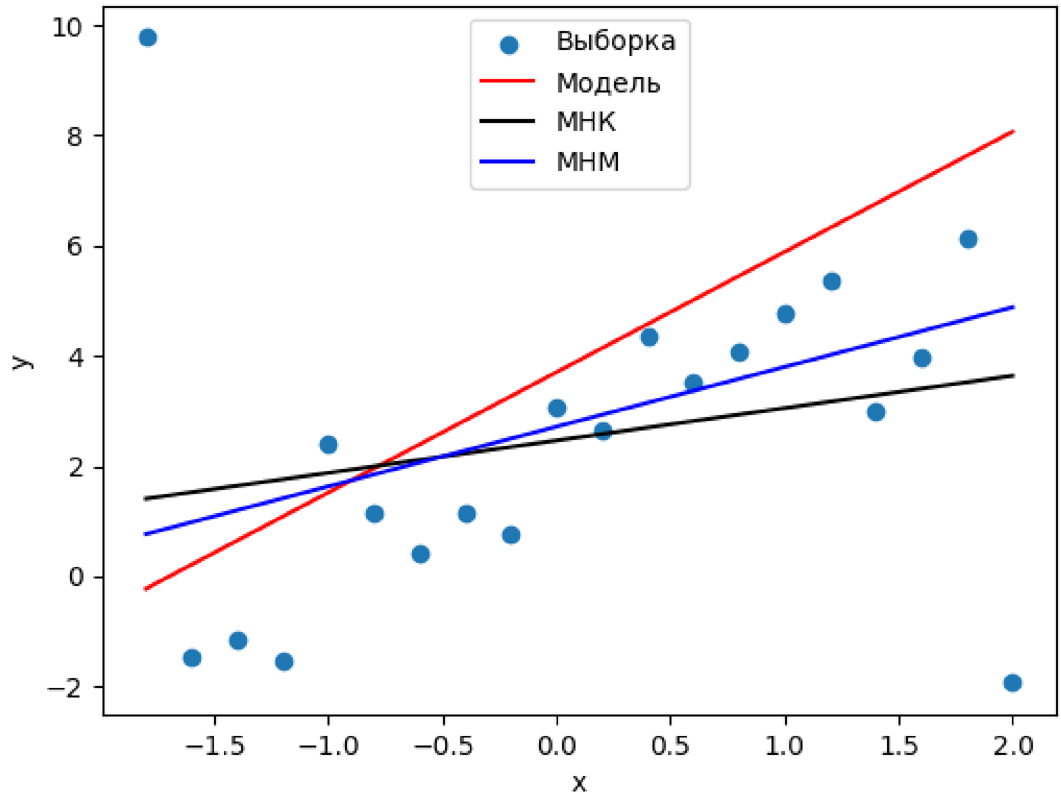

Рис.5 МНК и МНМ для выборки с возмущениями

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

Метод максимального правдоподобия: $\hat{\mu} \approx 0.08$, $\hat{\sigma} \approx 0.93$.

Критерий согласия χ^2:

Количество промежутков $k = 6$.

Уровень значимости $\alpha = 0.05$.

Тогда квантиль из таблицы $\chi^2_{1-\alpha}(k - 1) = 11.07$

i	Границы Δi	n_i	p_i	np_i	$n_i - np_i$	$\frac{(n_i - np_i)^2}{np_i}$
1	[-∞, -1.01]	8	0.1562	15.62	-7.62	3.72
2	[-1.01, -0.37]	17	0.1994	19.94	-2.94	0.43
3	[-0.73, 0.27]	28	0.2507	25.07	2.93	0.34
4	[0.27, 0.92]	22	0.2148	21.48	0.52	0.01
5	[0.92, 1.56]	18	0.1194	11.94	6.06	3.07
6	[1.56, + ∞]	7	0.0594	5.94	1.06	0.19
Σ	-	100	1	100	0	$7.77 = \chi^2_B$

Таблица 5: Вычисление χ^2_B при проверке гипотезы H_0 о нормальном законе распределения

Сравним $\chi^2_{0.95}5 = 11.07$ и найденное $\chi^2_B = 7.77$: 11.07>7.77. Следовательно, гипотезу H_0 на данном этапе проверки можно принять.

Исследование на чувствительность

1) Возьмем выборку, распределенную согласно $Laplace(x, \hat{\mu}, \frac{\hat{\sigma}}{\sqrt{2}})$

Рассмотрим гипотезу $H_0 *$, что выборка распределена согласно закону N(x, $\hat{\mu}$,$\hat{\sigma}$)

Критерий согласия χ^2:

Количество промежутков $k = 5$.

Уровень значимости $\alpha = 0.05$.

Тогда квантиль из таблицы $\chi^2_{1-\alpha}(k - 1) = 9.49$

i	Границы Δi	n_i	p_i	np_i	$n_i - np_i$	$\frac{(n_i - np_i)^2}{np_i}$
1	[-∞, -1.5]	1	0.0668	1.34	-0.34	0.08
2	[-1,5, -0.5]	2	0.2417	4.83	-2.83	1.66
3	[-0.5, 0.5]	8	0.3829	7.66	0.34	0.02
4	[0.5, 1.5]	8	0.2417	4.83	3.17	2.07
5	[1.5, + ∞]	1	0.0668	1.34	-0.34	0.08
Σ	-	20	1	20	0	3.92 = $\chi^2_В$

Таблица 6: Вычисление $\chi^2_В$ при проверке гипотезы $H_0 *$ о нормальном законе распределения

Сравним $\chi^2_{0.95}5$= 9.49 и найденное $\chi^2_В = 3.92$: 9.49 >3.92. Следовательно, гипотезу $H_0 *$ на данном этапе проверки можно принять.

2) Возьмем выборку, распределенную согласно $\mathrm{Uniform}(x, -\sqrt{3}, \sqrt{3})$

Рассмотрим гипотезу $H_0 **$, что выборка распределена согласно закону N(x, $\hat{\mu}$,$\hat{\sigma}$)

Критерий согласия χ^2:

Количество промежутков k = 5.

Уровень значимости α = 0.05.

Тогда квантиль из таблицы $\chi^2_{1-\alpha}(k - 1)$= 9.49

i	Границы Δi	n_i	p_i	np_i	$n_i - np_i$	$\frac{(n_i - np_i)^2}{np_i}$
1	[-∞, -1.5]	3	0.0668	1.34	1.66	2.07
2	[-1,5, -0.5]	4	0.2417	4.83	-0.83	0.14
3	[-0.5, 0.5]	2	0.3829	7.66	-5.66	4.18
4	[0.5, 1.5]	9	0.2417	4.83	4.17	3.59
5	[1.5, + ∞]	2	0.0668	1.34	0.66	0.33
Σ	-	20	1	20	0	10.32 = $\chi^2_В$

Таблица 7: Вычисление $\chi^2_В$ при проверке гипотезы $H_0 **$ о нормальном законе распределения

Сравним $\chi^2_{0.95}5$= 9.49 и найденное $\chi^2_В = 10.32$: 10.32 >9.49. Следовательно, гипотеза $H_0 **$ на данном этапе проверки должна быть отвергнута.

4.5 Доверительные интервалы для параметров нормального распределения

N = 20	m	σ
	-0.85 < m < 0.27	0.91 < σ < 1.76
N = 100	m	σ
	-0.29 < m < 0.23	0.99 < σ < 1.32

Таблица 8: Доверительные интервалы для параметров нормального распределения

4.6 Доверительные интервалы для параметров произвольного распределения. Асимптотический подход

N = 20	m	Σ
	-0.81 < m < 0.23	0.62 < σ < 1.74
N = 100	m	Σ
	-0.29 < m < 0.15	0.88 < σ < 1.37

Таблица 9: Доверительные интервалы для параметров произвольного распределения. Асимптотический подход

Построим графики плотности нормального распределения N(x, 0, 1) и нанесем оценки параметров распределения $[\underline{m}, \overline{m}]$ и $[\underline{m} - \underline{\sigma}, \overline{m} + \overline{\sigma}]$

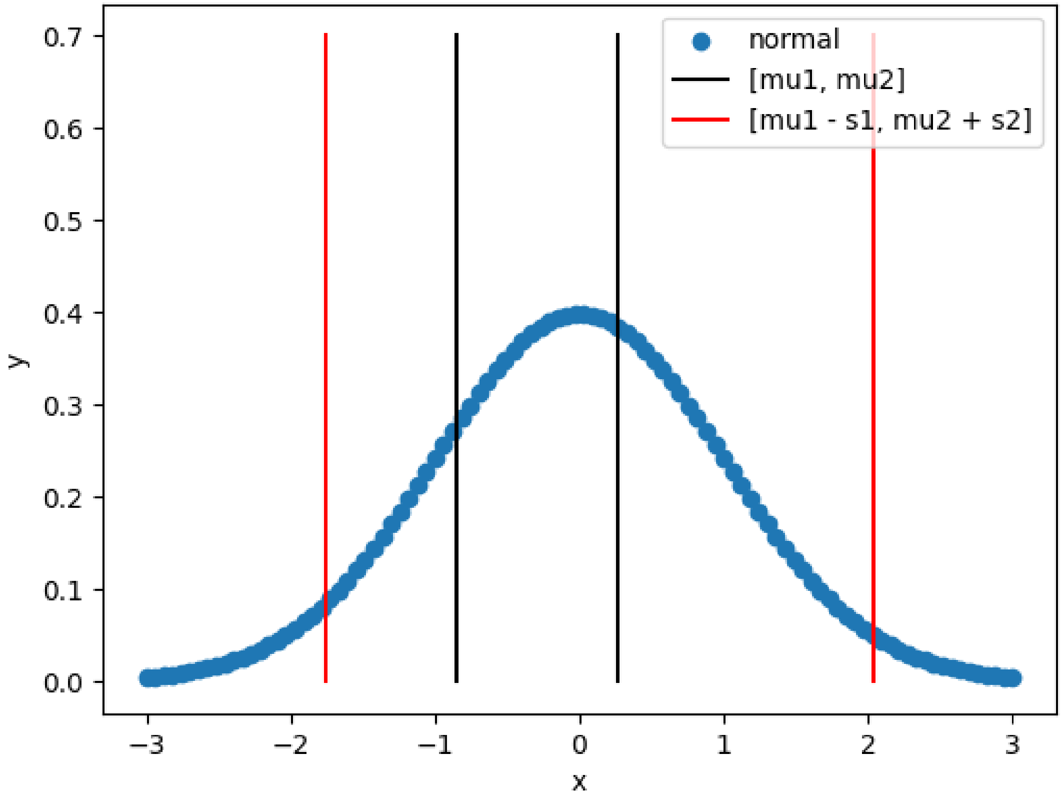

Рис.6 Доверительные интервалы для параметров нормального распределения, n = 20

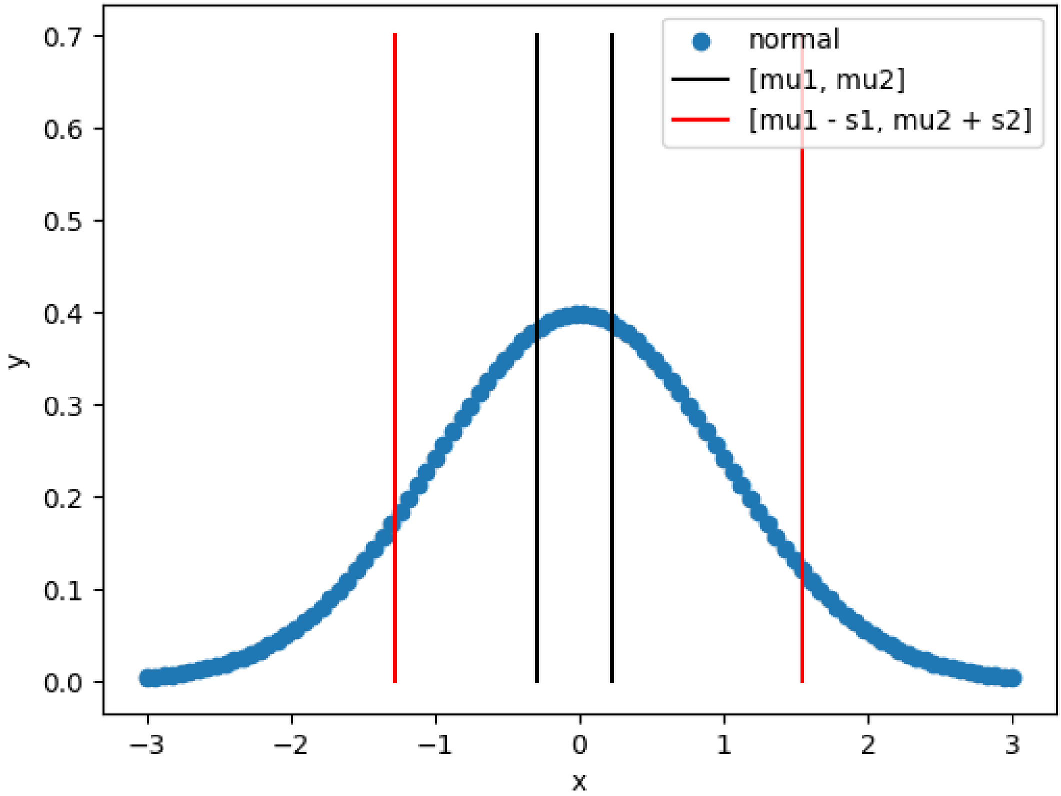

Рис.7 Доверительные интервалы для параметров нормального распределения, n = 100

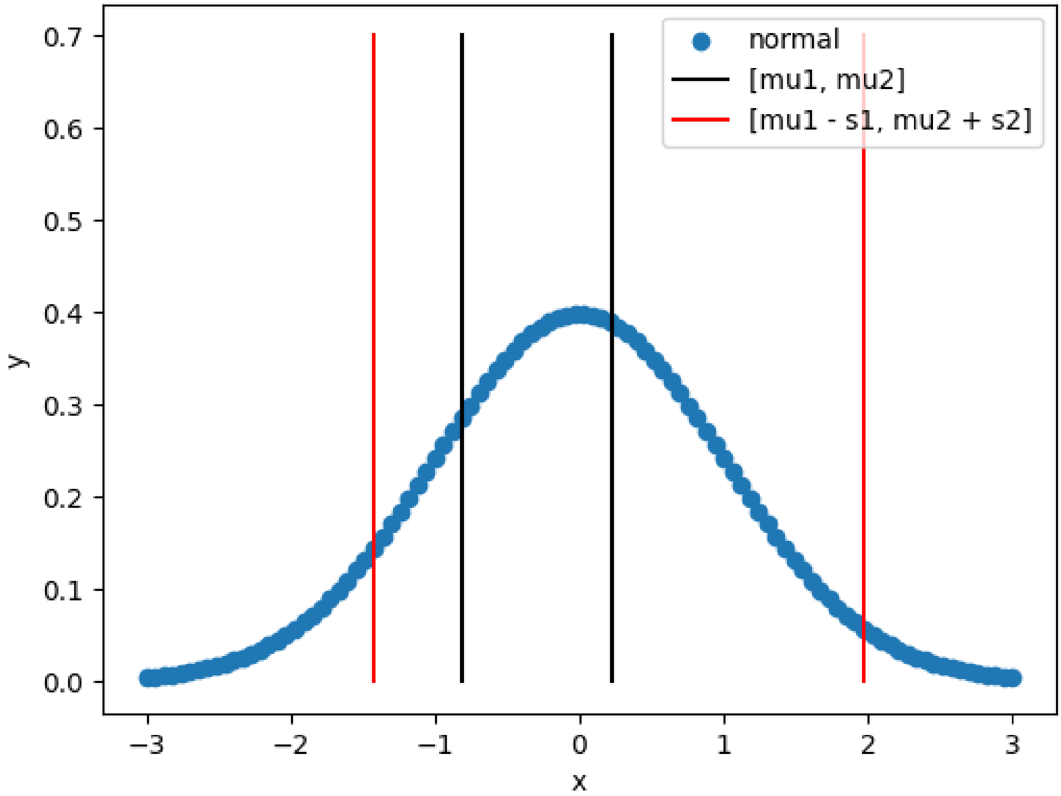

Рис.8 Доверительные интервалы для параметров произвольного распределения. Асимптотический подход, n=20

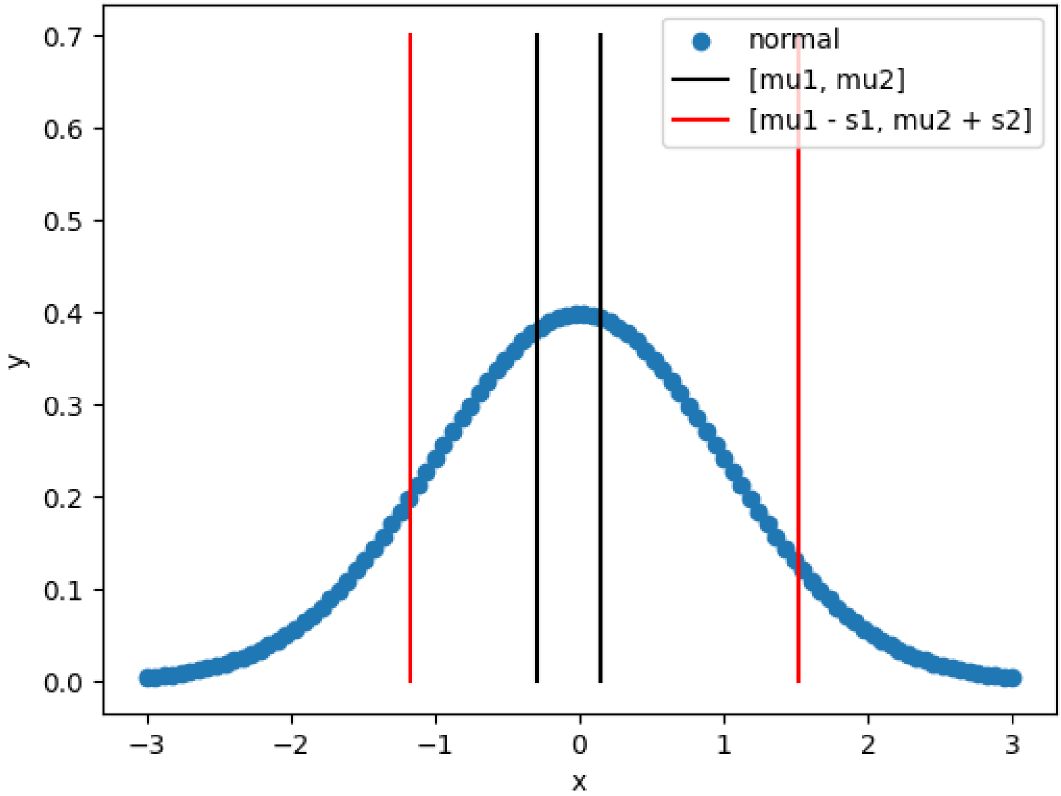

Рис.9 Доверительные интервалы для параметров произвольного распределения. Асимптотический подход, n=100

5 Обсуждение

5.1 Выборочные коэффициенты корреляции и эллипсы рассеивания

1) Сравним дисперсии выборочных коэффициентов корреляции.

Для двумерного нормального распределения дисперсии выборочных коэффициентов корреляции упорядочены следующим образом: $r < rS < rQ$.

– Для смеси нормальных распределений дисперсии выборочных коэффициентов корреляции упорядочены следующим образом: $r < rS < rQ$.

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

5.2 Оценки коэффициентов линейной регрессии

· Критерий наименьших квадратов и метод наименьших модулей практически одинаково точно оценивают коэффициенты линейной регрессии на выборке без возмущений

· Критерий наименьших модулей точнее оценивает коэффициенты линейной регрессии на выборке с возмущениями

· Критерий наименьших модулей устойчив к редким выбросам

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

· По результатам проверки на близость с помощью критерия хи-квадрат можно принять гипотезу H0 о нормальном распределении $N(x, \hat{\mu}, \hat{\sigma})$. на уровне значимости α = 0.05 для выборки, сгенерированной согласно $N(x, 0, 1)$. То есть, если взять в качестве гипотезы нормальное распределение с параметрами сдвига и масштаба равными оценкам максимального правдоподобия для μ, σ, вычисленным по выборке $\sim N(x, 0, 1)$, то критерий отразит эту согласованность. Теоретически это обосновывается состоятельностью оценки максимального правдоподобия.

· Видим так же, что критерий принял гипотезу о том, что 20-элементная выборка, сгенерированная согласно $N(x, 0, 1)$, описывается законом распределения $Laplace(x, \hat{\mu}, \frac{\hat{\sigma}}{\sqrt{2}})$ (тут опять μ, ˆ σˆ - оценки максимального правдоподобия для μ, σ, вычисленным по той же 20-элементной стандартной нормальной выборке.)

· То есть, при малых мощностях выборки критерий хи-квадрат не почувствовал разницы между нормально распределенной случайной величиной и распределенной по Лапласу. Это ожидаемый результат, ведь выборка довольно мала, законы схожи по форме и параметры масштаба и сдвига выбраны тоже так, чтобы законы максимально друг к другу приблизить.

·Тем не менее, критерий хи-квадрат смог отличить равномерно распределенную случайную величину от нормальной величины.

· По исследованию на чувствительность видим, что при небольших объемах выборки уверенности в полученных результатах нет, критерий может ошибиться. Это обусловлено тем, что теорема Пирсона говорит про асимптотическое распределение, а при малых размерах выборки результат не будет получаться достоверным. Статистика критерия χ^2 лишь асимптотически распределена по закону χ^2 $(k - 1)$, то есть значение n предполагается достаточно большим.

5.4 Доверительные интервалы для параметров распределения

· Генеральные характеристики (m = 0 и σ = 1) накрываются построенными доверительными интервалами

· Доверительные интервалы, полученные по большей выборке, являются соответственно более точными, т.е. меньшими по длине

· Доверительные интервалы для параметров нормального распределения более надёжны, так как основаны на точном, а не асимптотическом распределении