Цель работы

Изучить сложную предметную область, связанную с вероятностно статистическим моделированием.

Постановка задачи

Методами структурного программирования реализовать возможность работы с тремя распределениями:

– распределением, заданным в варианте и имеющим, помимо параметра формы ν, параметры сдвига μ и масштаба λ (данное распределение будет называться основным);

– распределением в виде смеси двух основных распределений с параметрами (μ1, λ1, ν1) и (μ2, λ2, ν2) и параметром смеси p;

– эмпирическим распределением, строящимся по выборке.

Для каждого из распределений необходимо реализовать набор из следующих трех функций:

– функция для вычисления значений плотности распределения по заданному аргументу (плотность для эмпирического распределения необходимо предварительно сформировать по выборке);

– функция для вычисления математического ожидания, дисперсии, коэффициентов асимметрии и эксцесса;

– функция для моделирования случайной величины

Основное распределение: **симметричное гиперболическое распределение**

Путеводитель по работе

Реализация

Распределение		Основные функции		
		Плотность	Характеристики $M\xi$, $D\xi$, γ_1, γ_2	Моделирование
Основное	Стандартное	Формула (1.1)	$M\xi=\gamma_1=0$, $D\xi = K_2(\nu)/K_1(\nu)$, γ_2 = Формула (1.5)	Формула (1.4)
	Сдвигмасштаб	$f(x,\mu,\lambda)=\frac{1}{\lambda}f\left(\frac{x-\mu}{\lambda}\right)$	$M\xi=\mu$, $D\xi=\sigma^2\lambda^2$, где $\sigma^2 = K_2(\nu)/K_1(\nu)$, $\gamma_1=0$, γ_2 = Формула (1.5)	Формула (1.4)
Смесь		$f(x)=(1-p)f_1(x)+pf_2(x)$	$M\xi=\sum_{i=1}^{m} p_i M_i$, $D\xi$ = Формула (2.3), γ_1 = Формула (2.4), γ_2 = Формула (2.5)	Формула (2.6)
Эмпирическое		$f(x)=\begin{cases}\frac{n_i}{n\Delta}, x\in \mathbf{X}^i\\ 0, x\notin \mathbf{X}\end{cases}$	$M\xi=\frac{1}{n}\sum_{i=1}^{n}X_i$, $D\xi=\frac{1}{n}\sum_{i=1}^{n}(X_i-M\xi)^2$ $\hat{\gamma}_1=\frac{\sum_{i=1}^{n}(X_i-M\xi)^3}{n(D\xi)^{1.5}}$ $\hat{\gamma}_2=\frac{\sum_{i=1}^{n}(X_i-M\xi)^4}{n(D\xi)^2}-3$	Формула (3.7)

Распределение	Основные функции		
	Плотность	Характеристики $M\xi$, $D\xi$, γ_1, γ_2	Модели-ро-вание
Основное	3.1	3.1	3.3.1
Смесь	3.2	3.2	3.3.1
Эмпирическое	3.3.1	3.3.1, 3.3.2	3.3.2

Алгоритм

Расчет **основного** распределения:

Расчет плотности симметрического гиперболического распределения осуществляется по формуле:

$$f(x,\nu)=\frac{1}{2\sqrt{\nu}K_1(\nu)}\exp\left(-\nu\sqrt{1+x^2/\nu}\right) \quad (1.1)$$

С учетом сдвига и масштаба:

$$f(x,\nu,\lambda,\mu)=\frac{1}{2\lambda\sqrt{\nu}K_1(\nu)}exp\left(-\nu\sqrt{1+\frac{\left(\frac{x-\mu}{\lambda}\right)^2}{\nu}}\right) \quad (1.2)$$

Математическое ожидание: $M\xi=\mu$ (1.3)

Дисперсия: $D\xi=\lambda^2 K_2(\nu)/K_1(\nu)$ (1.4)

Коэффициент асимметрии: $\gamma_1=0$

Коэффициент эксцесса: $\gamma_2 = 3K_3(\nu)K_1(\nu)/K_2^2(\nu)-3$ (1.5)

Случайная величина: x = $z\sqrt{(t)}$ без сдвига-масштаба и $x=\mu+\lambda z\sqrt{t}$ со сдвигом-масштабом соответственно, где $z=\sqrt{-2\ln r_3}\cos 2\pi r_4$, $t=-\frac{2}{\delta}\ln r_1$, $\delta=\frac{2}{\nu}\left(\sqrt{1+\nu^2}-1\right)$ при $-ln(r_2)>\frac{\nu-\delta}{2}t+\frac{\nu}{2t}-\sqrt{\nu(\nu-\delta)}$, r1, r2, r3, r4 – случайные величины, равномерно распределенные на интервале (0, 1). (1.6)

Расчет **смеси** распределений:

Расчет плотности смеси двух распределений осуществляется по формуле:

$$f(x)=(1-p)f_1(x)+pf_2(x) \quad (2.1)$$

Математическое ожидание: $M\xi=\sum_{i=1}^{m}p_iM_i$ (2.2)

Дисперсия: $D\xi=\sum_{i=1}^{m}p_i(M_i^2+D_i)-(M\xi)^2$ (2.3)

Коэффициент асимметрии:

$$\gamma_1=\frac{1}{(D\xi)^{1.5}}\sum_{i=1}^{m}p_i\left[(M_i-M\xi)^3+3(M_i-M\xi)D_i+D_i^{1.5}\gamma_{1i}\right] \quad (2.4)$$

Коэффициент эксцесса:

$$\gamma_2=\frac{\sum_{i=1}^{m}p_i\left[(M_i-M\xi)^4+6(M_i-M\xi)^2D_i+4(M_i-M\xi)D_i^{3/2}\gamma_{1i}+D_i^2(\gamma_{2i}+3)\right]}{(D\xi)^2}-3 \quad (2.5)$$

Случайная величина:

$$\xi=\begin{cases}\xi_1, r\le(1-p)\\ \xi_2, r>(1-p)\end{cases}, r\in(0,1) \quad (2.6)$$

Расчет **эмпирического** распределения:
Расчет эмпирической плотности распределения:

$$f(x)=\begin{cases}\dfrac{n_i}{n\Delta}, x\in\mathbf{X}^i\\ 0, x\notin\mathbf{X}\end{cases} \tag{3.1}$$

Здесь предполагается, что промежуток $\mathbf{X}=[X_{\min}, X_{\max}]$ разбит на k непересекающихся промежутков $\mathbf{X}^i$, $i=1,\ldots,k$, длины $\Delta=\frac{1}{k}(X_{\max}-X_{\min})$, при этом каждый промежуток содержит свой левый конец, но лишь последний промежуток содержит и свой правый конец, n_i – количество элементов выборки, содержащихся в промежутке $\mathbf{X}^i$. Таким образом, имеем промежутки

$$\mathbf{X}^i=[X_{\min}+(i-1)\Delta, X_{\min}+i\Delta),\ i=1,\ldots,k-1,$$

$$\mathbf{X}^k=[X_{\min}+(k-1)\Delta, X_{\min}+k\Delta]=[X_{\max}-\Delta, X_{\max}].$$

$$k=\lfloor\log_2 n\rfloor+1 \tag{3.2}$$

Математическое ожидание:

$$M\xi=\frac{1}{n}\sum_{i=1}^{n}X_i \tag{3.3}$$

Дисперсия:

$$D\xi=\frac{1}{n}\sum_{i=1}^{n}(X_i-M\xi)^2 \tag{3.4}$$

Коэффициент эксцесса:

$$\hat{\gamma}_1=\frac{\sum_{i=1}^{n}(X_i-M\xi)^3}{n(D\xi)^{1.5}} \tag{3.5}$$

Коэффициент эксцесса:

$$\hat{\gamma}_2=\frac{\sum_{i=1}^{n}(X_i-M\xi)^4}{n(D\xi)^2}-3 \tag{3.6}$$

Случайная величина:

$$\xi=\begin{cases}X_{min}+r_1\delta, 0<r_2<\frac{n_0}{n}\\ X_1+r_1\delta, \frac{n_0}{n}<r_2<\frac{n_0+n_1}{n}\\ \ldots\\ X_{max}, \frac{n_0+\ldots+n_{k-1}}{n}<r_2<1\end{cases} \tag{3.7}$$

Тестирование

3.1. Минимальный набор тестов для основного распределения:

3.1.1) тест для стандартного распределения: $\mu=0$, $\lambda=1$, $\nu = 1$;

3.1.2) тест для масштабных преобразований: $\mu=0$, $\lambda=2$, $\nu = 1$.

3.1.3) тест для сдвиг-масштабных преобразований: $\mu=10$, $\lambda=2$, $\nu = 1$.

3.2. Минимальный набор тестов для смеси распределений (см. пример 1.2):

3.2.1) тест для тривиального случая: $\mu_1=\mu_2=10$, $\lambda_1=\lambda_2=2$, $\nu_1=\nu_2 = 1$, $p = 0.5$;

3.2.2) тест для сдвиговых преобразований: $\mu_1=0$, $\mu_2=2$, $\lambda_1=\lambda_2=1$, $\nu_1=\nu_2 = 1$, $p=0.75$ ($M\xi=1.5$, $D\xi=\sigma_i^2+0.75$, $\gamma_1=-0.75/(D\xi)^{1.5}$);

3.2.3) тест для масштабных преобразований: $\mu_1=\mu_2=0$, $\lambda_1=1$, $\lambda_2=3$, $\nu_1=\nu_2 = 1$, $p=0.5$ ($M\xi=0$, $D\xi=5\sigma_i^2$, $\gamma_1=0$, $\gamma_2=1.64(\gamma_{2i}+3)-3$);

3.2.4) тест с неравными параметрами формы: $\mu_1=\mu_2=0$, $\lambda_1=\lambda_2=1$, $\nu_1=0.1$, $\nu_2 = 30$, $p=0.5$ ($M\xi=0$, $D\xi=(\sigma_1^2+\sigma_2^2)/2$, $\gamma_1=0$, $\gamma_2=0.5(\sigma_1^4(\gamma_{21}+3)+\sigma_2^4(\gamma_{22}+3))/(D\xi)^2-3$).

3.3. Тестирование эмпирического распределения и функций моделирования случайных величин для всех распределений:

3.3.1) для (нестандартных) основного распределения и смеси при некоторых значениях их параметров.

3.3.2) в соответствии с эмпирической плотностью, построенной по одной из выборок, сгенерировать новую выборку того же объема, вычислить ее эмпирические характеристики, сравнить их с эмпирическими характеристиками исходной выборки и теоретическими характеристиками.

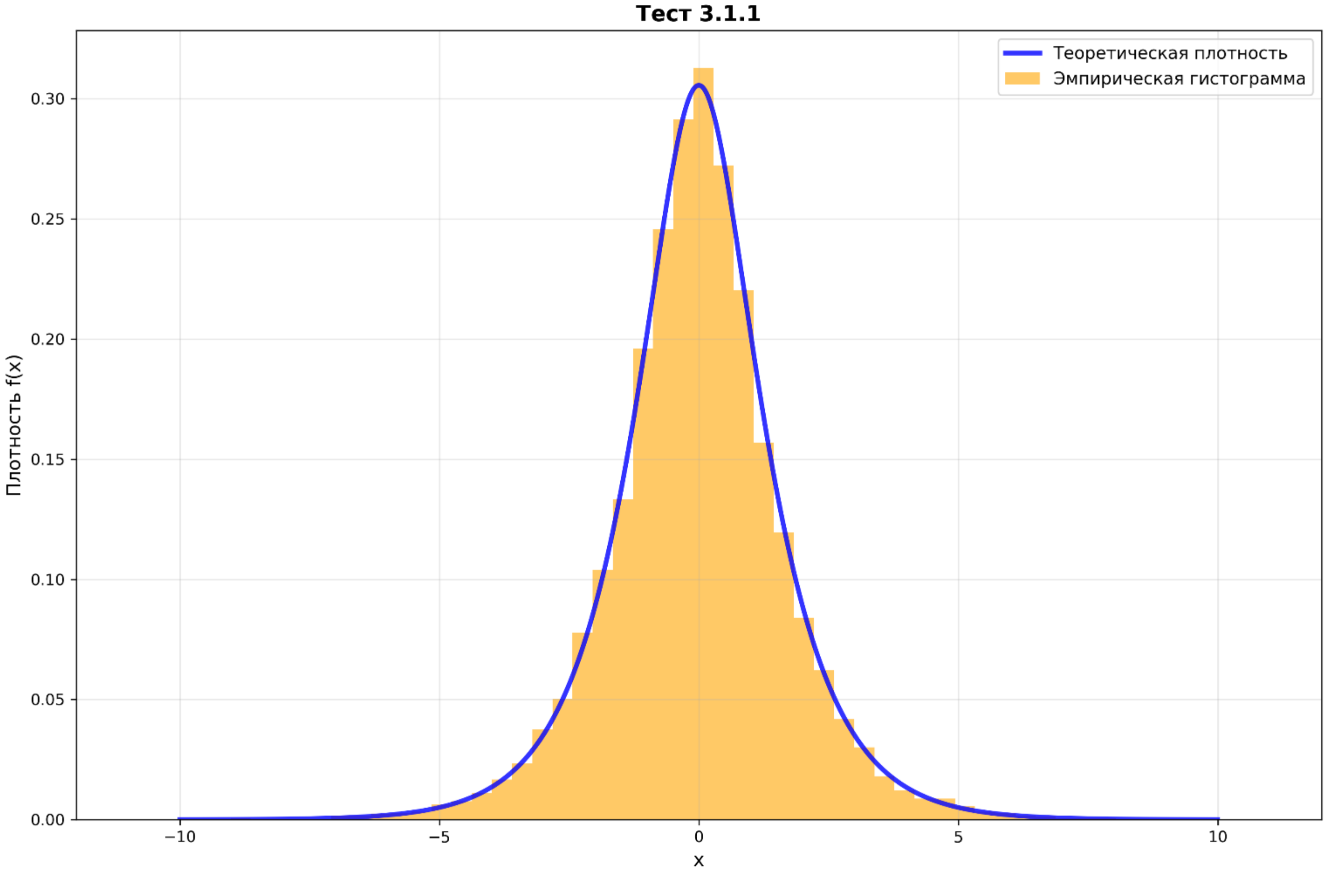
Результаты тестирования:

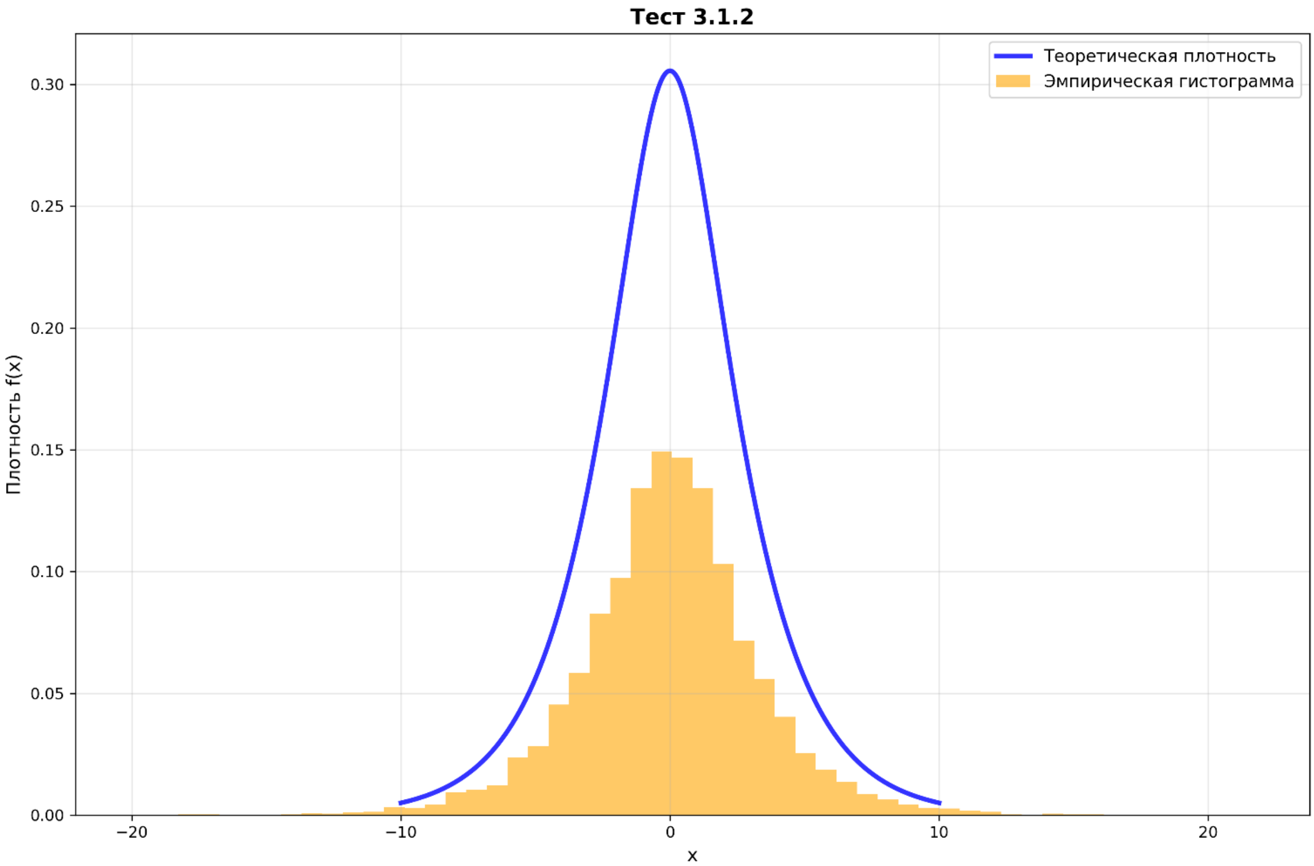
Тест	Параметры	Проверяемая функция	Ожидаемый результат	Полученный результат	Статус
3.1.1 Стандартное	$\mu=0$, $\lambda=1$, $v=1.0$	pdf_main(…) moments_main(…)	f(0)=0.306 M=0, D=2.699, $\gamma_2=1.857$	f(0)=0.306 M=0, D=2.699, $\gamma_2=1.857$	OK
3.1.2 Масштабирование	$\mu=0$, $\lambda=2$, $v=1.0$	moments_main(…)	D=2.699×4=10.796	D=10.797	OK
3.1.3 Сдвиг-масштаб	$\mu=5$, $\lambda=2$, $v=1.0$	moments_main(…)	M=5, D=10.796	M=5.000 D=10.798	OK
3.2.1 Тривиальный случай	$\mu_1=\mu_2=0$, $\lambda_1=\lambda_2=2$, $v_1=v_2=1.0$, p=0.75	moments_mixture(…)	M=0, D=10.796	M=0.000 D=10.798	OK

Тест	Параметры	Проверяемая функция	Ожидаемый результат	Полученный результат	Статус
3.2.2 Сдвиговые преобразования	$\mu_1=0$, $\mu_2=2$, $\lambda_1=\lambda_2=1$, $v_1=v_2=1.0$, $p=0.75$	moments_mixture(...)	M=0.5, D=3.449	M=0.500, D=3.449	OK
3.2.3 Масштабные преобразования	$\mu_1=\mu_2=0$, $\lambda_1=1$, $\lambda_2=3$, $v_1=v_2=1.0$, $p=0.5$	moments_mixture(...)	M=0, D=13.495	M=0.000, D=13.497	OK
3.2.4 Разные параметры формы	$\mu_1=\mu_2=0$, $\lambda_1=\lambda_2=1$, $v_1=0.5$, $v_2=2.0$, $p=0.5$	moments_mixture(...)	M=0, D=3.186	M=0.000, D=3.186	OK
3.3.1 Основное + смесь	Выборка n=10000 из СГР	pdf_empirical(...) moments_empirical(...)	Совпадение с теоретическими значениями	Центр: ошибка 7.6% Края: ошибка 79.8%	Частично
3.3.2 Генерация из эмпирического	Исходная vs новая выборка	moments_empirical(...)	Совпадение моментов	M: 0.2% ошибка D: 0.2% ошибка	OK

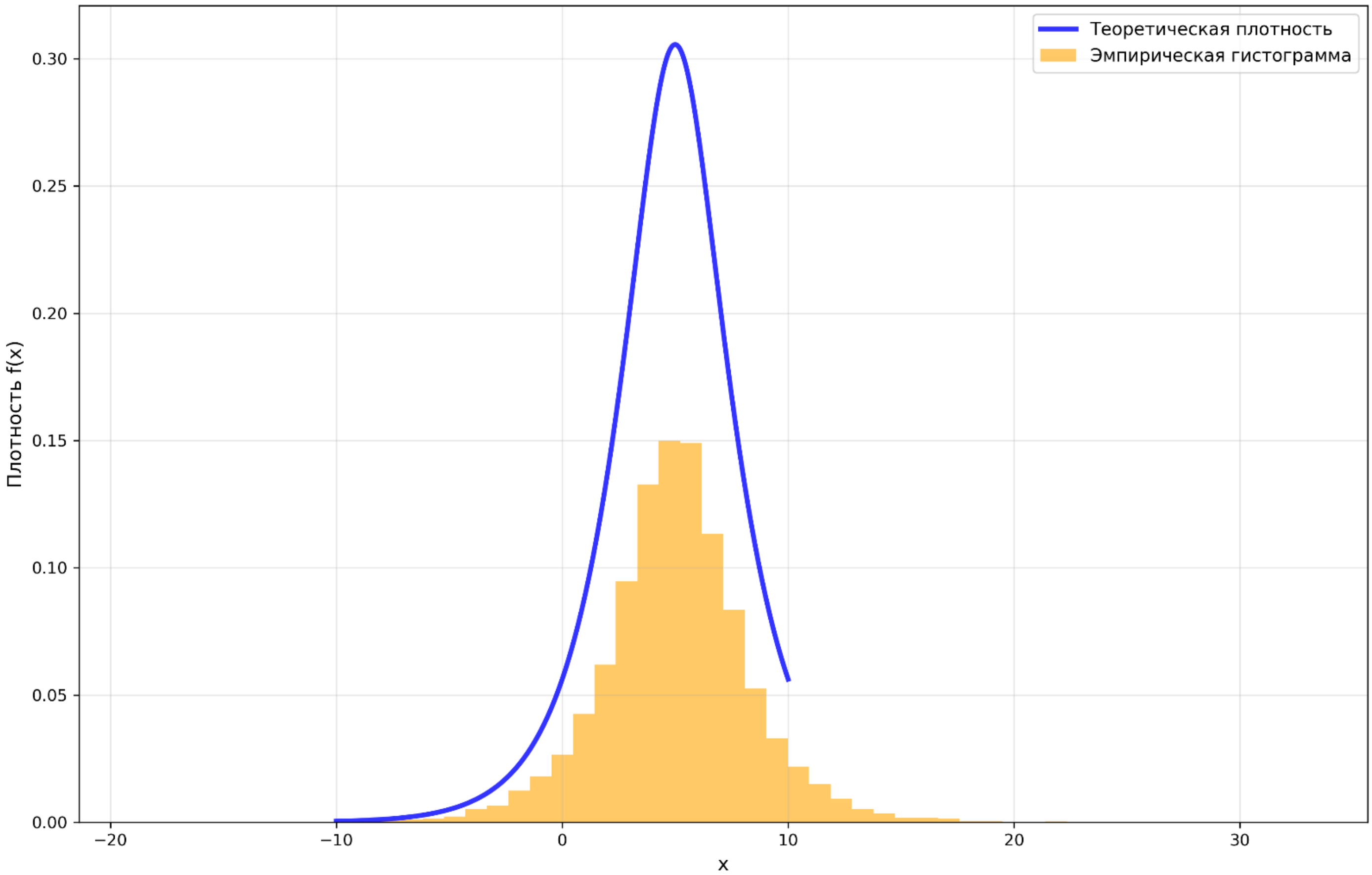
Графики распределений
Все графики построены на выборке из 10000 элементов с помощью python.

Графики тестов:

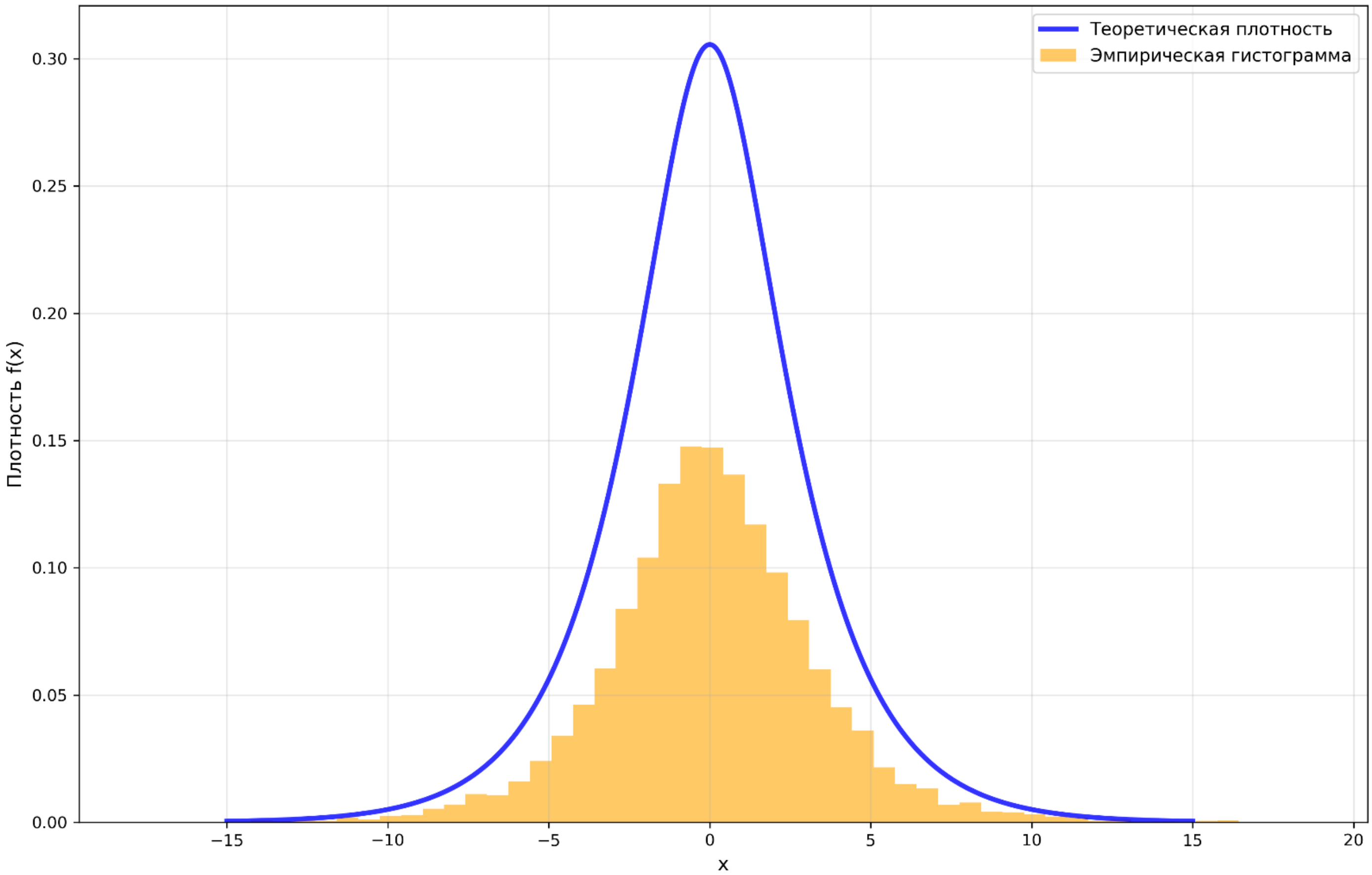


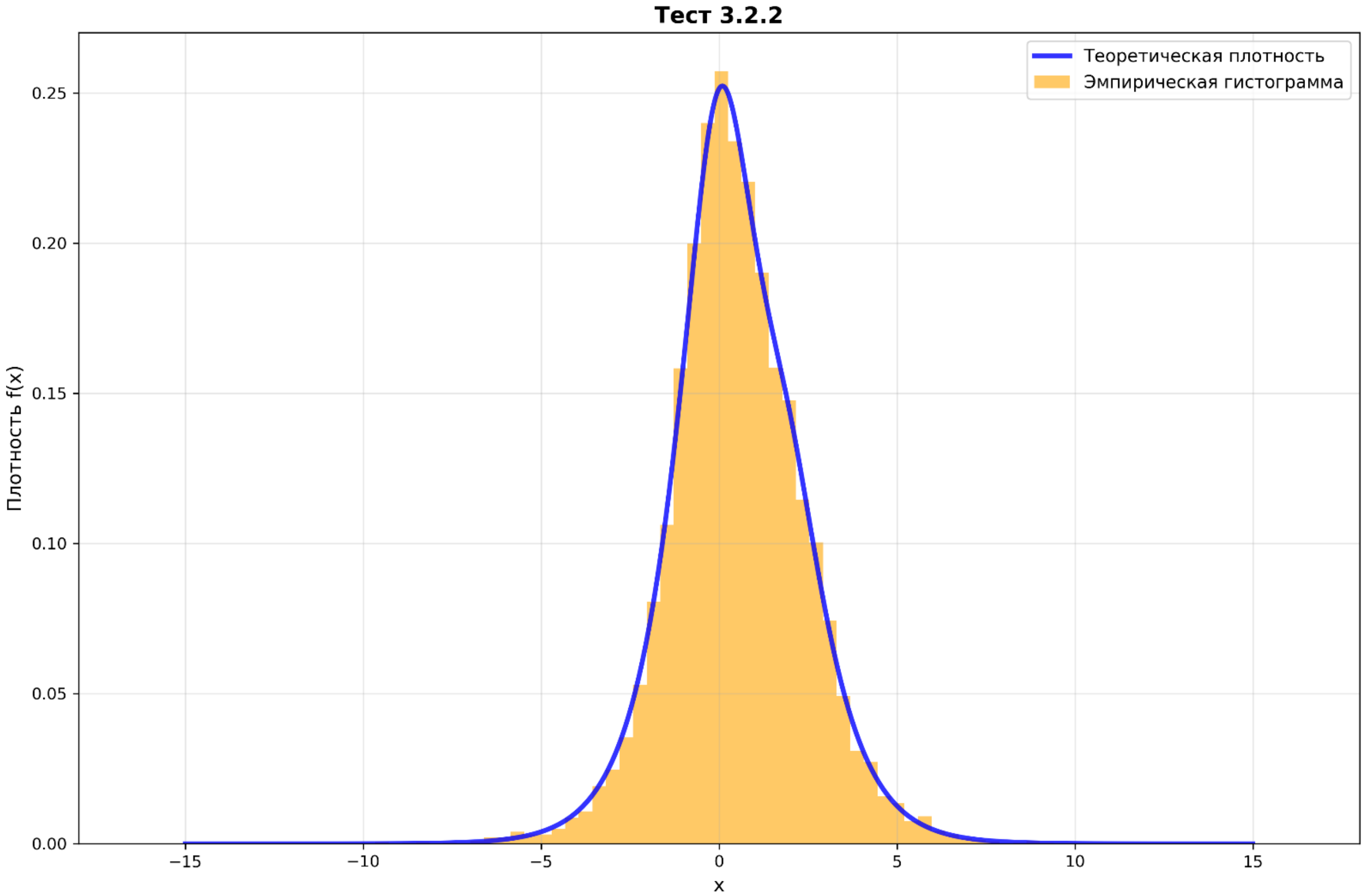


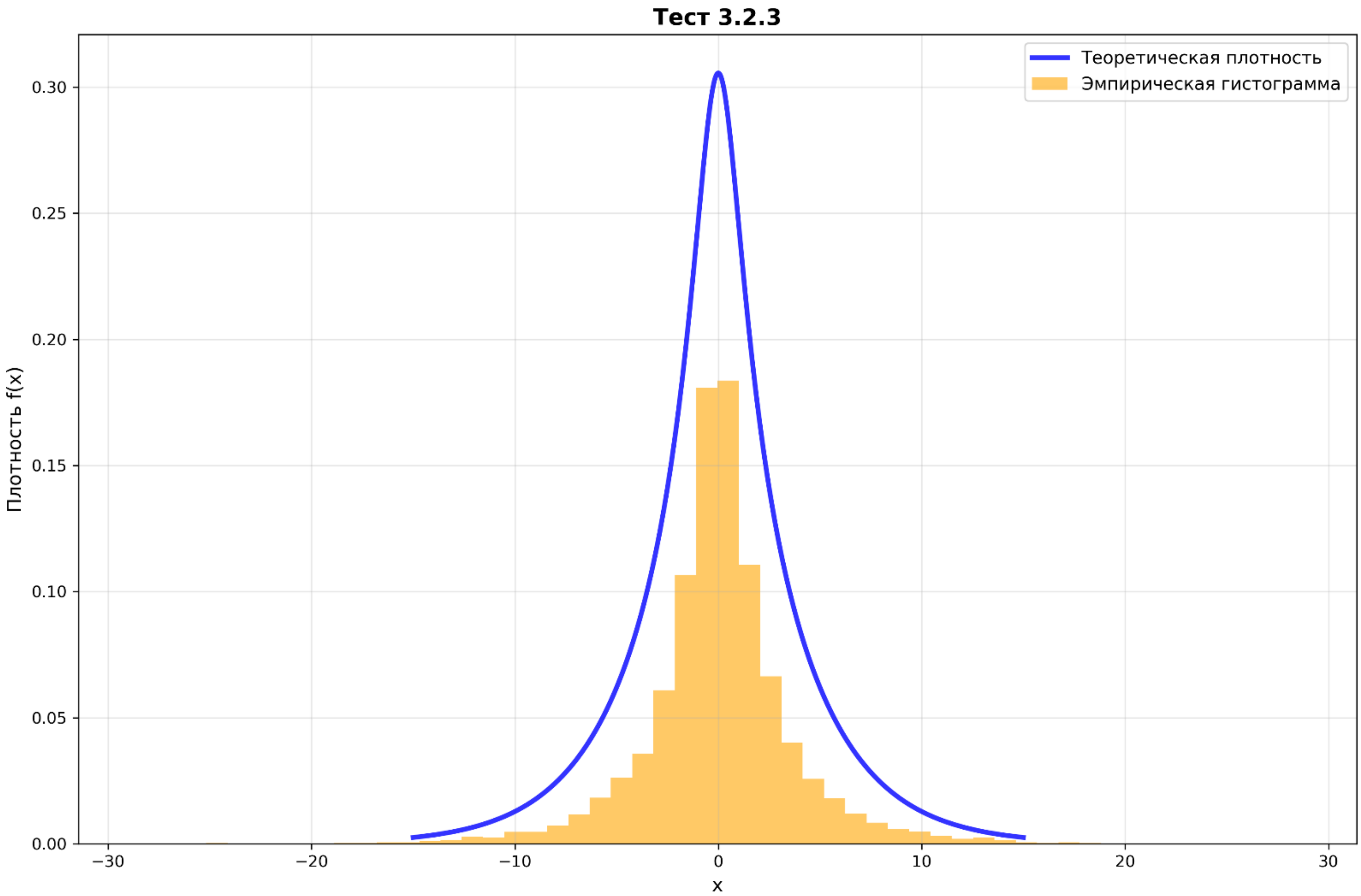
Тест 3.1.3

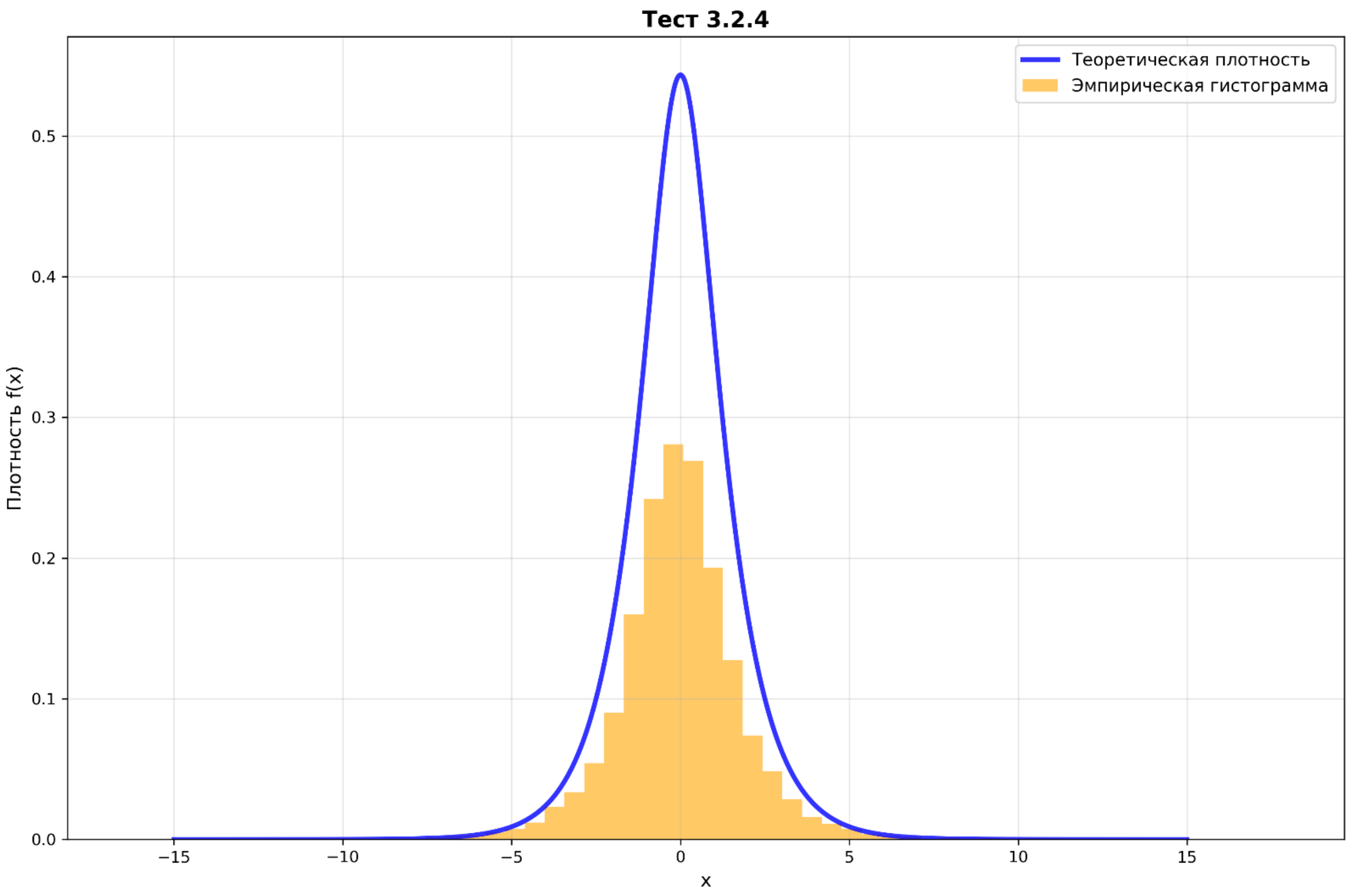


Тест 3.2.1




Тест 3.2.2
Теоретическая плотность
Эмпирическая гистограмма
Плотность f(x)
0.00
0.05
0.10
0.15
0.20
0.25
−15
−10
−5
0
5
10
15
x


Тест 3.2.3
Теоретическая плотность
Эмпирическая гистограмма
Плотность f(x)
0.00
0.05
0.10
0.15
0.20
0.25
0.30
−30
−20
−10
0
10
20
30
x



Вывод

По моделированию:

- Реализован программный комплекс для работы с тремя типами распределений
- Все генераторы работают корректно
- Эмпирические моменты совпадают с теоретическими
- Вероятностные свойства соблюдаются

По плотностям:

- Центр распределения: высокая точность (7.6% ошибка)
- Края распределения: повышенная ошибка (79.8%) - особенность гистограммного метода
- Для практических применений точность достаточна

Приложение

Вся лабораторная работа, а именно:

- Файлы с исходным кодом
- Сгенерированные файлы с данными
- Графики
- Этот отчет

лежит на моем GitHub репозитории по ссылке:
https://github.com/artemkiri101/nstu_oop_lab_1