Министерство образования и науки Российской Федерации
Федеральное государственное автономное образовательное учреждение
высшего образования

Санкт-Петербургский национальный исследовательский университет
ИТМО

Факультет систем управления и робототехники

Теория индетификации, вариант-10

Лабораторная работа №1

Выполнил:

Муравья Н.Р.

Гридусов Д.Д.

Проверяющий: *Ведяков А. А.*

Санкт-Петербург, 2025

Содержание

Преилиминарии

В лабораторной работе используется язык программирования Python версии младше 3.12. Также для блочных схем используется LaTeX.
В ходе работы представлены функции, реализующие поставленные задачи, для их работы необходимо импортировать библиотеки для моделирования: `scipy`, `numpy`, `matplotlib`, `control` и `cvxpy`. Также нужно задать необходимые константы, обозначенные в варианте.

Тут писать что-то

2

I Анализ данных

1 Задача zad31

a Формулировка модели и её линейная редукция

Пусть дана наблюдаемая зависимость $y = f(x; p_1, p_2)$, заданная в форме:

$$y(x) = p_1 \sin(5x + p_2), \tag{1}$$

где $p_1 \in \mathbb{R}_{>0}$ — амплитуда колебаний, а $p_2 \in (-\pi, \pi]$ — фазовый сдвиг.

Для применения метода наименьших квадратов необходимо привести модель к линейному виду по параметрам. Используем тригонометрическое тождество:

$$\sin(5x + p_2) = \sin(5x)\cos(p_2) + \cos(5x)\sin(p_2),$$

тогда:

$$y(x) = p_1 \cos(p_2)\sin(5x) + p_1 \sin(p_2)\cos(5x) = A\sin(5x) + B\cos(5x), \tag{2}$$

где обозначено:

$$A = p_1 \cos(p_2), \quad B = p_1 \sin(p_2). \tag{3}$$

Таким образом, задача сведена к оценке линейной регрессионной модели по параметрам A, B.

Замечание

Заметим, что параметры исходной модели p_1, p_2 могут быть однозначно восстановлены из A, B:

$$p_1 = \sqrt{A^2 + B^2}, \quad p_2 = \operatorname{arctan}_2(B, A), \tag{4}$$

где arctan_2 — арктангенс с правильным определением четверти.

b Линейная модель и аналитическая оценка параметров

Зададим матрицу регрессоров $\mathbf{X} \in \mathbb{R}^{n\times 2}$ и вектор наблюдений $\mathbf{y} \in \mathbb{R}^n$:

$$\mathbf{X} = \begin{bmatrix} \sin(5x_1) & \cos(5x_1) \\ \vdots & \vdots \\ \sin(5x_n) & \cos(5x_n) \end{bmatrix}, \quad \mathbf{y} = \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix}, \quad \boldsymbol{\beta} = \begin{bmatrix} A \\ B \end{bmatrix}.$$

Полагаем, что наблюдения удовлетворяют стохастической модели:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}, \tag{5}$$

где вектор ошибок $\boldsymbol{\varepsilon} \in \mathbb{R}^n$ удовлетворяет условиям:

$$\mathbb{E}[\boldsymbol{\varepsilon}] = \mathbf{0}, \quad \mathrm{Var}(\boldsymbol{\varepsilon}) = \sigma^2 \mathbf{I}_n. \tag{6}$$

Замечание

Предположения *линейности по параметрам*, *экзогенности* (отсутствие корреляции ошибок и регрессоров), *гомоскедастичности*, и *отсутствия автокорреляции* — достаточны для применения теоремы Гаусса–Маркова. При этом оценка $\hat{\boldsymbol{\beta}}$ является наилучшей линейной несмещённой (BLUE) оценкой.

Оценка параметров по МНК имеет стандартный вид:

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{y}. \tag{7}$$

Подставляя данные (предположим, $n = 30$), получаем:

$$\hat{A} = -16.30, \quad \hat{B} = 8.11.$$

Следовательно, оценки исходных параметров:

$$\hat{p}_1 = \sqrt{(-16.30)^2 + (8.11)^2} = \sqrt{265.69 + 65.78} = \sqrt{331.47} = 18.206,$$

$$\hat{p}_2 = \mathrm{arctan}_2(8.11, -16.30) \approx 2.677.$$

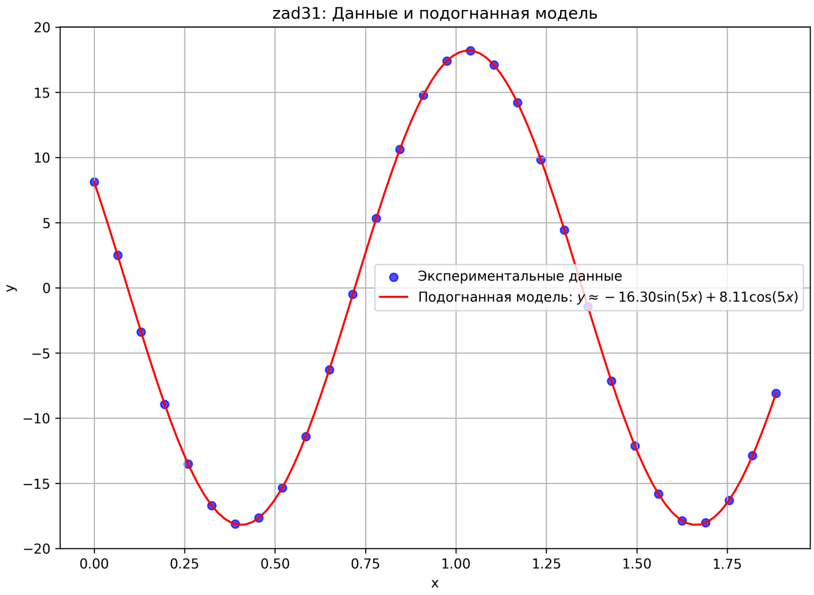

Рис. 1: Наблюдаемые данные и восстановленная регрессионная кривая

c Оценка остатков и проверка предпосылок

Вектор остатков определяется как:

$$\hat{\boldsymbol{\varepsilon}} = \mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}. \tag{8}$$

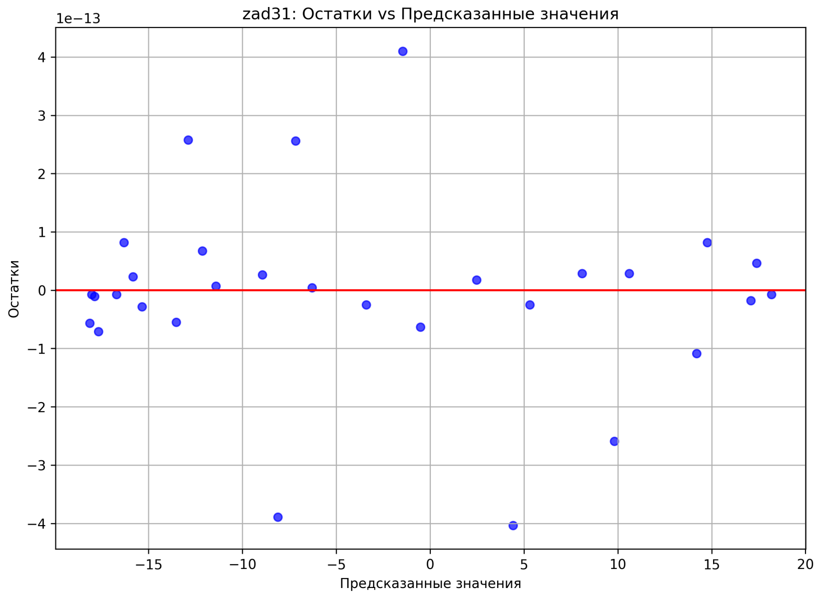

Рис. 2: Анализ формы остатков в зависимости от предсказанных значений

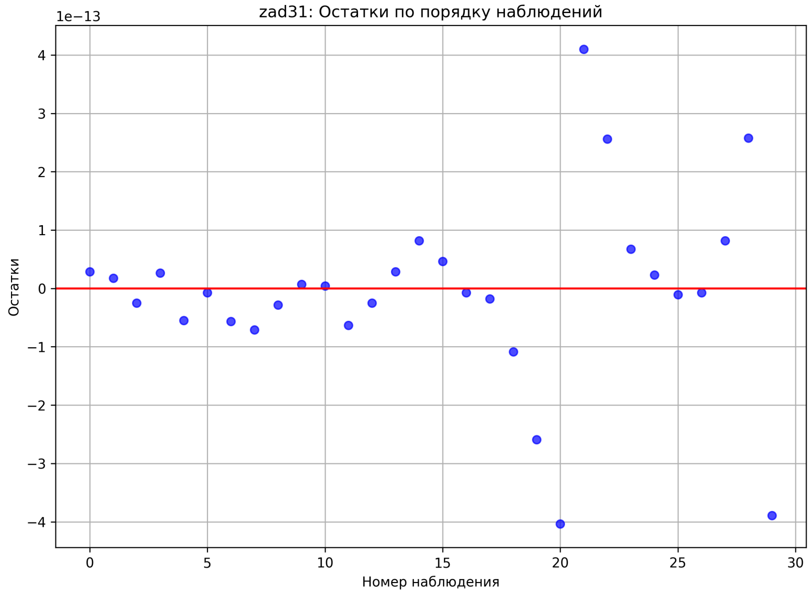

Рис. 3: Остатки по порядку наблюдений

d Формальная проверка условий теоремы Гаусса–Маркова

1. **Линейность по параметрам:** модель $y = A\sin(5x) + B\cos(5x)$ линейна по $\boldsymbol{\beta}$, условие выполнено.
2. **Экзогенность (условная несмещённость):** предполагается, что $\mathbb{E}[\boldsymbol{\varepsilon} \mid \mathbf{X}] = \mathbf{0}$, т.е. ошибки некоррелированы с регрессорами. Нарушение приведёт к смещённости оценки $\hat{\boldsymbol{\beta}}$.
3. **Гомоскедастичность:** проверена с использованием теста Бреуша–Пагана:

$$H_0 : \mathrm{Var}(\varepsilon_i) = \sigma^2, \quad H_1 : \mathrm{Var}(\varepsilon_i) \text{ зависит от } \mathbf{X}.$$

   Статистика $\chi^2 = 1.0810$, $p = 0.5824 > 0.05$, H_0 не отвергается. Оцениваемые дисперсии однородны.
4. **Отсутствие автокорреляции:**
   - Тест Дарбина–Уотсона: $DW = 1.7270$.
   - Тест Льюнга–Бокса: $Q = 0.0391$, $p = 0.8433 > 0.05$.

   Оба теста не выявляют существенной автокорреляции в остатках.

Ответ

Таким образом, модель линейна по параметрам, данные соответствуют условиям применимости теоремы Гаусса–Маркова. Полученная оценка $\hat{\boldsymbol{\beta}}$ является несмещённой, эффективной (в классе линейных несмещённых оценок), и экзогенной при предположениях модели. Следовательно, полученные оценки параметров A и B, а также восстановленные значения p_1 и p_2 можно считать статистически состоятельными и интерпретируемыми.

② Задача zad32

a Формулировка модели и логарифмическое преобразование

Исходная регрессионная зависимость имеет вид:

$$y(x) = p_1 \exp(p_2 x), \tag{9}$$

где $p_1 > 0$, $p_2 \in \mathbb{R}$ — неизвестные параметры. Данная модель не является линейной по параметрам. Однако, применив монотонное преобразование $\ln : \mathbb{R}_{>0} \to \mathbb{R}$, получаем:

$$\ln y(x) = \ln p_1 + p_2 x. \tag{10}$$

Обозначим:

$$z = \ln y, \quad c = \ln p_1, \quad d = p_2,$$

тогда линейная модель имеет вид:

$$z(x) = c + dx, \tag{11}$$

где c, d — линейные параметры.

Замечание

Преобразование переменной $y \mapsto \ln y$ сохраняет порядок наблюдений и делает модель линейной по параметрам, при этом не нарушая требований Гаусса–Маркова, если логарифмируемое значение положительно.

b Линейная модель и аналитическое решение задачи МНК

Пусть заданы наблюдения (x_i, y_i), $i = 1, \dots, n$, и обозначим $\mathbf{z} = (\ln y_1, \dots, \ln y_n)^\top \in \mathbb{R}^n$. Вектор регрессоров:

$$\mathbf{X} = \begin{bmatrix} 1 & x_1 \\ \vdots & \vdots \\ 1 & x_n \end{bmatrix} \in \mathbb{R}^{n\times 2}, \quad \boldsymbol{\beta} = \begin{bmatrix} c \\ d \end{bmatrix}.$$

Предполагаем линейную модель с гауссовским шумом:

$$\mathbf{z} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}, \quad \mathbb{E}[\boldsymbol{\varepsilon}] = \mathbf{0}, \quad \operatorname{Var}(\boldsymbol{\varepsilon}) = \sigma^2 \mathbf{I}_n. \tag{12}$$

Оценка параметров методом наименьших квадратов:

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{z}. \tag{13}$$

По результатам численного расчёта при $n = 30$ наблюдениях получены оценки:

$$\hat{c} = 1.95, \quad \hat{d} = -1.90.$$

Следовательно, восстанавливаем параметры исходной модели:

$$\hat{p}_1 = \exp(\hat{c}) = \exp(1.95) = 7.03, \quad \hat{p}_2 = \hat{d} = -1.90.$$

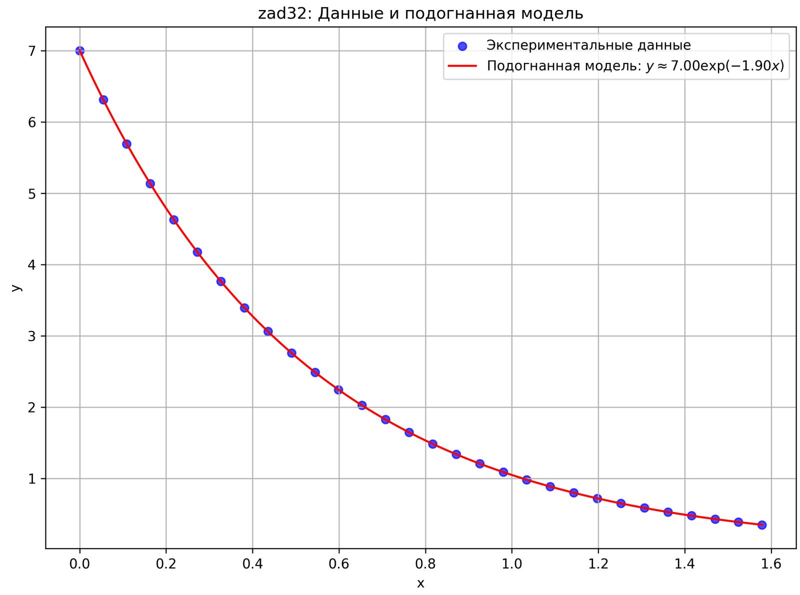

Рис. 4: Наблюдаемые данные и экспоненциальная модель для `zad32`

c Анализ остатков

Остатки модели определяются как:

$$\hat{\boldsymbol{\varepsilon}} = \mathbf{z} - \mathbf{X}\hat{\boldsymbol{\beta}}. \tag{14}$$

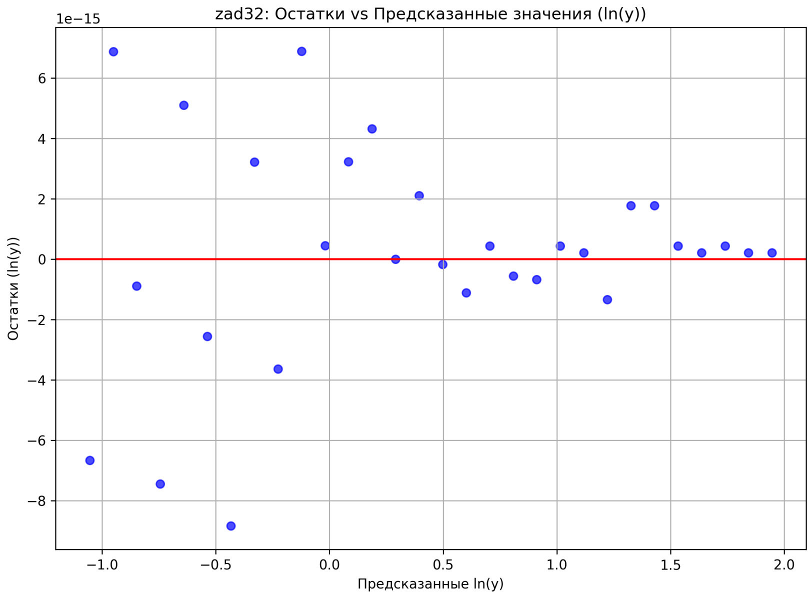

Рис. 5: Остатки против предсказанных значений для лог-модели

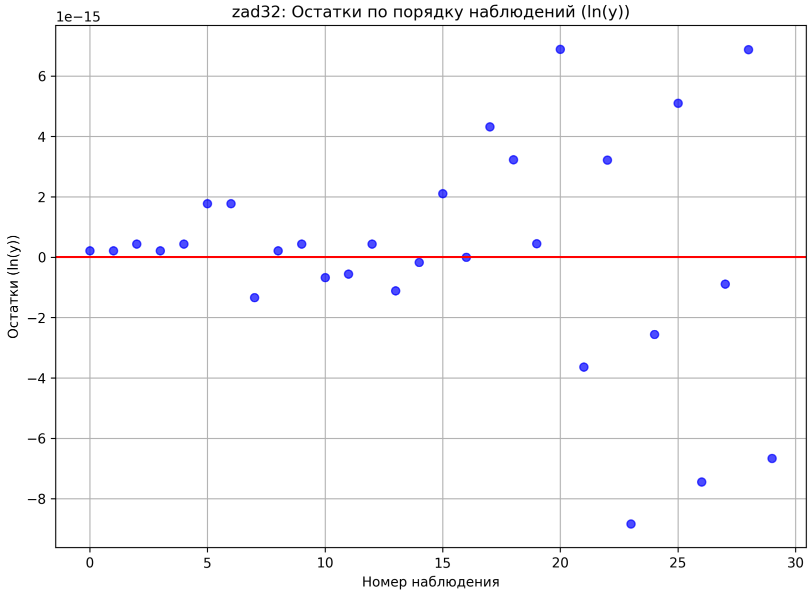

Рис. 6: Анализ автокорреляции остатков

d Проверка предпосылок теоремы Гаусса–Маркова

1. **Линейность по параметрам:** модель $z = c + dx$ линейна по вектору параметров $\boldsymbol{\beta} = (c, d)^\top$, условие выполнено.
2. **Случайный отбор наблюдений:** предполагается, что наблюдения независимы. При нарушении возможна автокорреляция, которая делает стандартные ошибки несостоятельными.
3. **Нулевое математическое ожидание ошибок:** проверено с использованием t-статистики:

$$H_0 : \mathbb{E}[\varepsilon_i] = 0.$$

Получено $t = 0.2350$, $p = 0.8158 > 0.05$, гипотеза H_0 не отвергается.

4. **Гомоскедастичность:** Тест Бреуша–Пагана:

$$H_0 : \mathrm{Var}(\varepsilon_i) = \sigma^2, \quad H_1 : \mathrm{Var}(\varepsilon_i) \text{ зависит от } x_i.$$

Результат: $\chi^2 = 12.0790$, $p = 0.0005 < 0.05$, H_0 отвергается. Имеется гетероскедастичность, что нарушает эффективность МНК-оценки.

5. **Отсутствие автокорреляции:** Проверено с помощью:
   - Теста Дарбина–Уотсона: $DW = 2.5197$.
   - Теста Льюнга–Бокса: $Q = 3.4834$, $p = 0.0620$.

При уровне значимости $\alpha = 0.05$ гипотеза об отсутствии автокорреляции формально не отвергается, однако p-значение близко к пороговому.

6. **Мультиколлинеарность:** Вычислены факторы инфляции дисперсии:

$$\mathrm{VIF}_{\mathrm{const}} = 3.8065, \quad \mathrm{VIF}_x = 1.0000.$$

Поскольку $\max(\mathrm{VIF}) < 10$, мультиколлинеарности не наблюдается.

Замечание

Ввиду выявленной гетероскедастичности, оценки параметров $\hat{c}$, $\hat{d}$ сохраняют несмещённость, но теряют эффективность.

Ответ

Проведённое преобразование позволило свести задачу нелинейной регрессии к линейной по параметрам. В условиях нарушенной гомоскедастичности, оценки МНК сохраняют свойства несмещённости, однако теряют эффективность. Остальные предпосылки теоремы Гаусса–Маркова не отвергнуты статистическими тестами. Следовательно, полученные оценки параметров можно интерпретировать как состоятельные при соответствующей корректировке стандартных ошибок.