

САНКТ-ПЕТЕРБУРГСКИЙ ПОЛИТЕХНИЧЕСКИЙ УНИВЕРСИТЕТ  
ПЕТРА ВЕЛИКОГО

ФИЗИКО-МЕХАНИЧЕСКИЙ ИНСТИТУТ

ВЫСШАЯ ШКОЛА ПРИКЛАДНОЙ МАТЕМАТИКИ И ВЫЧИСЛИТЕЛЬНОЙ ФИЗИКИ

**Отчет**  
**по лабораторной работе №10**  
**по дисциплине**  
**«Математическая статистика»**

Выполнил студент:  
Иванова А.С.  
группа: 5030102/00101

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

Санкт-Петербург  
2023 г.

# Содержание

<b>1</b>	<b>Постановка задачи</b>	<b>4</b>
<b>2</b>	<b>Теория</b>	<b>5</b>
2.1	Постановка задачи восстановления функциональной зависимости . . . . .	5
2.2	Варьирование неопределенности изменений . . . . .	5
2.3	Варьирование неопределенности изменений с расширением и сужением интервалов . . . . .	6
2.4	Анализ регрессионных остатков . . . . .	7
2.5	Информационное множество задачи . . . . .	7
<b>3</b>	<b>Реализация</b>	<b>8</b>
<b>4</b>	<b>Результаты</b>	<b>9</b>
4.1	Данные выборки . . . . .	9
4.2	Варьирование неопределенности изменений . . . . .	10
4.3	Варьирование неопределенности изменений с расширением и сужением интервалов . . . . .	10
4.4	Анализ регрессионных остатков . . . . .	12
4.5	Информационное множество задачи . . . . .	14
4.6	Коридор совместных зависимостей . . . . .	14
4.7	Прогноз вне области данных . . . . .	15
<b>5</b>	<b>Обсуждение</b>	<b>16</b>
5.1	Варьирование неопределенности изменений . . . . .	16
5.2	Варьирование неопределенности изменений с расширением и сужением интервалов . . . . .	16
5.3	Анализ регрессионных остатков . . . . .	16
5.4	Информационное множество задачи . . . . .	16
5.5	Коридор совместных зависимостей . . . . .	16
5.6	Прогноз вне области данных . . . . .	17

## Список иллюстраций

1	Данные выборки $\mathbf{X}_1$ . . . . .	9
2	Диаграмма рассеяния выборки $\mathbf{X}_1$ с уравновешанным интервалом неопределенности . . . . .	9
3	Диаграмма рассеяния выборки $\mathbf{X}_1$ и регрессионная прямая по модели . . . . .	10
4	Диаграмма рассеяния выборки $\mathbf{X}_1$ и регрессионная прямая по модели . . . . .	11
5	Векторы $w_0$ и $w_1$ . . . . .	11
6	Диаграмма рассеяния регрессионных остатков для модели без сужения интервалов . . . . .	12
7	Диаграмма рассеяния регрессионных остатков для модели с сужением и расширением интервалов . . . . .	12
8	Частоты элементарных подинтервалов регрессионных остатков при вычислении моды для двух моделей . . . . .	13
9	Информационное множество задачи по заданной модели – красный брус, интервальная оболочка – синий брус . . . . .	14
10	Коридор совместных зависимостей . . . . .	15
11	Коридор совместных зависимостей. Построение прогноза . . . . .	15

## Список таблиц

## 1 Постановка задачи

Имеется выборка данных с интервальной неопределенностью. Число отсчетов в выборке равно 200. Используется модель данных с уравновешенным интервалом погрешности.

$$\boldsymbol{x} = \overset{\circ}{x} + \boldsymbol{\epsilon}; \quad \boldsymbol{\epsilon} = [-\epsilon, \epsilon] \text{ для некоторого } \epsilon > 0,$$

Здесь  $\overset{\circ}{x}$  – данные некоторого прибора,  $\epsilon = 10^{-4}$  – погрешность прибора.

Нужно иллюстрировать данные выборки, построить диаграмму рассеяния, построить линейную регрессионную зависимость варьированием неопределенности изменений (без сужения и с расширением и сужением интервалов), произвести анализ регрессионных остатков, построить информационное множество по модели, проиллюстрировать коридор совместных зависимостей, построить прогноз вне области данных.

## 2 Теория

### 2.1 Постановка задачи восстановления функциональной зависимости

Пусть некоторая величина  $y$  является функцией от независимых переменных  $x_1, \dots, x_m$ :

$$y = f(\beta, x) \quad (1)$$

Где  $x = (x_1, \dots, x_m)$  является вектором независимых переменных,  $\beta = (\beta_1, \dots, \beta_p)$  – вектор параметров функции. Заметим, что переменные  $x_1, \dots, x_m$  также называются входными, а переменные  $y_1$  – выходной.

**Задача восстановления функциональной зависимости** заключается в том, чтобы, располагая набор значений  $x$  и  $y$ , найти такие  $\beta_1, \dots, \beta_p$  в выражении (1), которые соответствуют конкретной функции  $f$  из параметрического семейства. Если функция  $f$  является линейной, то можно записать

$$y = \beta_0 + \beta_1 x_1 + \dots + \beta_m x_m \quad (2)$$

В общем случае результаты измерений величин  $x_1, \dots, x_m$  и  $y$  являются интервальнозначными

$$x_1^{(k)}, \dots, x_m^{(k)}, y^{(k)}$$

Индекс  $k$  пробегает значения от 1 до  $n$ , равного полному числу измерений.

**Брусом неопределенности  $k$ -го измерения функциональной зависимости** будем называть интервальный вектор-брус, образованный интервальными результатами измерений с одинаковыми значениями индекса  $k$

$$(x_{k1}, \dots, x_{km}, y_k) \subset \mathbb{R}^{m+1} \quad (3)$$

Брус неопределенности измерения является прямым декартовым произведением интервалов неопределенности независимых переменных и зависимой переменной.

### 2.2 Варьирование неопределенности изменений

Если величину коррекции каждого интервального наблюдения  $y_i = [\overset{\circ}{y}_i - \epsilon_i, \overset{\circ}{y}_i + \epsilon_i]$  выборки  $S_n$  выражать коэффициентом его уширения  $w_i \geq 1$ , а общее изменение выборки характеризовать суммой этих коэффициентов, то минимальная коррекция выборки в виде вектора коэффициентов  $w^* = (w_1^*, \dots, w_n^*)$ , необходимая для совместности задачи построения  $y = f(x, \beta)$  может быть решена решением задачи условной оптимизации:

Найти:

$$\min_{w, \beta} \sum_{i=1}^n w_i \quad (4)$$

При ограничениях:

$$\begin{cases} \text{mid}\mathbf{x}_i - w_i\epsilon_i \leq \beta_0 + \beta_i * i \leq \text{mid}\mathbf{x}_i + w_i\epsilon_i, \\ w_i \geq 1 \end{cases} \quad (5)$$

$$i = 1, \dots, n$$

Результирующие значения коэффициентов  $w_i$ , строго превосходящие единицу, указывают на наблюдения, которые требуют уширения интервалов неопределенности для обеспечения совместности данных и модели.

## 2.3 Варьирование неопределенности изменений с расширением и сужением интервалов

Поставим задачу условной оптимизации следующим образом:

Найти:

$$\min_{w, \beta} \sum_{i=1}^n w_i \quad (6)$$

При ограничениях:

$$\begin{cases} \text{mid}\mathbf{x}_i - w_i\epsilon_i \leq \beta_0 + \beta_i * i \leq \text{mid}\mathbf{x}_i + w_i\epsilon_i, \\ w_i \geq 0 \end{cases} \quad (7)$$

$$i = 1, \dots, n$$

Отличие постановки от (4) и (5) состоит в том, что интервалы измерений могут как расширяться в случае  $w_i \geq 1$ , так и сужаться при  $0 \leq w_i < 1$

Задавшись каким-то порогом  $\alpha : 0 < \alpha \leq 1$  можно выделить области входного аргумента  $\Psi$ , в которых регрессионная зависимость хуже соответствует исходным данным. Например:

$$\Psi = \arg_i w_i \geq \alpha \quad (8)$$

Для объективного использования этого приема параметр  $\alpha$  можно брать, например, из анализа гистограммы распределения ветвора  $w$ .

Использование выделения «подозрительных» областей даёт основу для других приемов. Например, для построения кусочно-линейной регрессионной зависимости.

## 2.4 Анализ регрессионных остатков

В теоретико-вероятностной математической статистике анализ регрессионных остатков один из приемов оценки качества регрессии.

Приведем пример пояснения этого приема. «Если выбранная регрессионная модель хорошо описывает истинную зависимость, то остатки должны быть независимыми, нормально распределенными случайными величинами с нулевым средним, и в значениях должен отсутствовать тренд Анализ регрессионных остатков – это процесс проверки выполнения этих условий».

В случае интервальных выборок мы не задаемся вопросом о виде распределения остатков, а будем использовать те возможности, которые появляются при описании объектов и результатов вычислений в виде интервалов.

## 2.5 Информационное множество задачи

Информационным множеством задачи восстановления функциональной зависимости является множество всех значений параметров функции  $\beta$ , получаемое в результате обработки интервальных значений входных переменных  $x$  и выходной переменной  $y$ .

Один из главных вопросов при построении регрессии – оценивание её параметров. В зависимости от прикладных целей характер и назначение искомых оценок могут существенно различаться.

Внешняя интервальная оценка параметра определяется минимальными и максимальными значениями, которых может достигать значение параметра в информационном множестве.

В совокупности интервальные оценки параметров задают брус, описанный вокруг информационного множества и именуемый внешней интервальной оболочкой информационного множества.



### 3 Реализация

Лабораторная работа была выполнена с помощью встроенных средств языка программирования Python (библиотеки: NumPy, SciPy, Matplotlib, Pandas, Pypoman) в среде разработки Visual Studio Code. Исходный код работы приведен в приложении.

Ссылка на репозиторий с исходным кодом: <https://github.com/anivse/MathematicalStatistics>

## 4 Результаты

### 4.1 Данные выборки

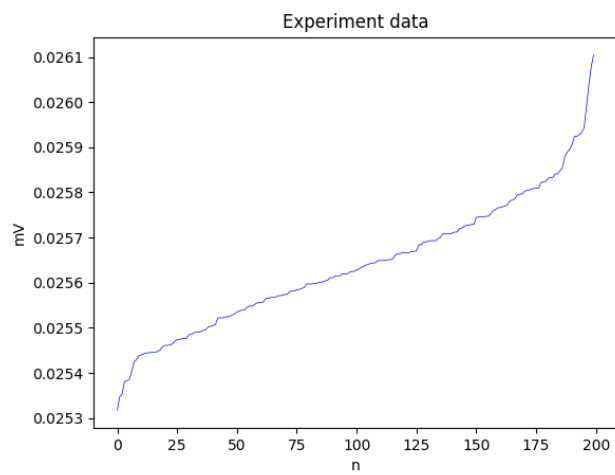


Рис. 1: Данные выборки  $X_1$

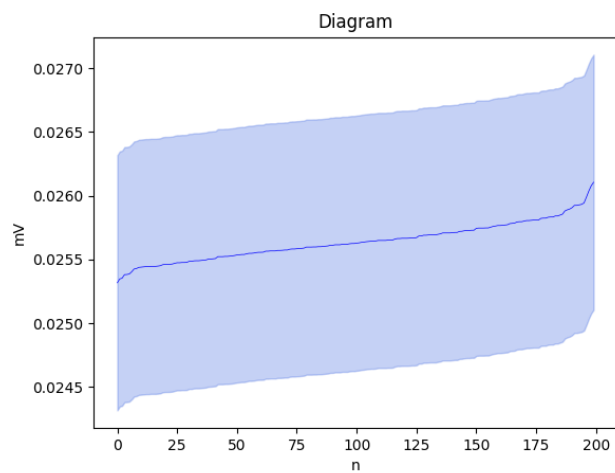


Рис. 2: Диаграмма рассеяния выборки  $X_1$  с уравновешанным интервалом неопределенности

## 4.2 Варьирование неопределенности изменений

При решении задачи линейного программирования были получены следующие результаты:

$$\beta_0 = 0.025, \beta_1 = 3.439 \cdot 10^{-6}$$

$$w_1 = (w_1^1, \dots, w_1^n), \sum_{i=1}^n w_1^i = 200.169$$

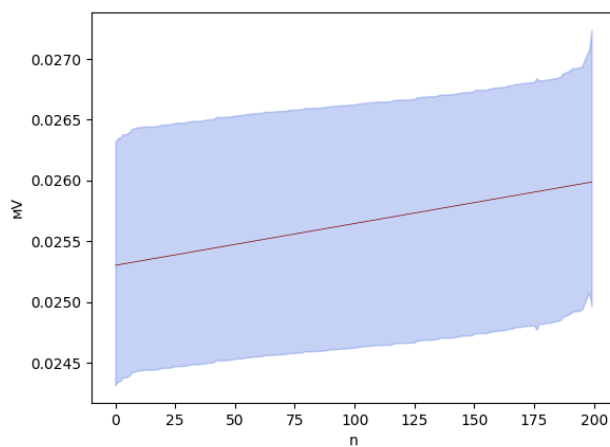


Рис. 3: Диаграмма рассеяния выборки  $X_1$  и регрессионная прямая по модели

Красным цветом обозначена регрессионная прямая модели.

## 4.3 Варьирование неопределенности изменений с расширением и сужением интервалов

При решении задачи линейного программирования были получены следующие результаты:

$$\beta_0 = 0.025, \beta_1 = 2.200 \cdot 10^{-6}$$

$$w_0 = (w_0^1, \dots, w_0^n), \sum_{i=1}^n w_0^i = 32.546$$

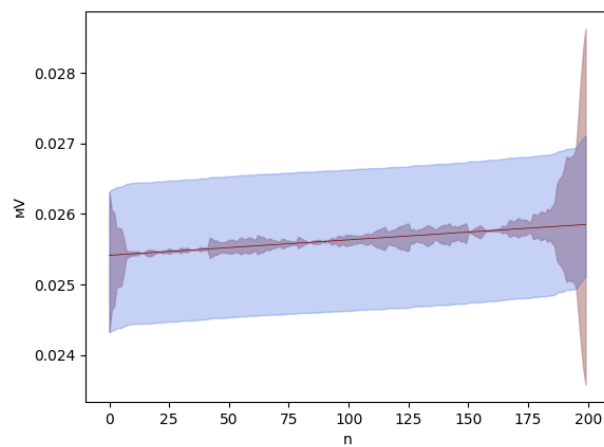


Рис. 4: Диаграмма рассеяния выборки  $\mathbf{X}_1$  и регрессионная прямая по модели

Розовым цветом обозначены скорректированные интервалы выборки  $\mathbf{X}_1$ , голубым - первоначальные интервалы, красным цветом обозначена регрессионная прямая.

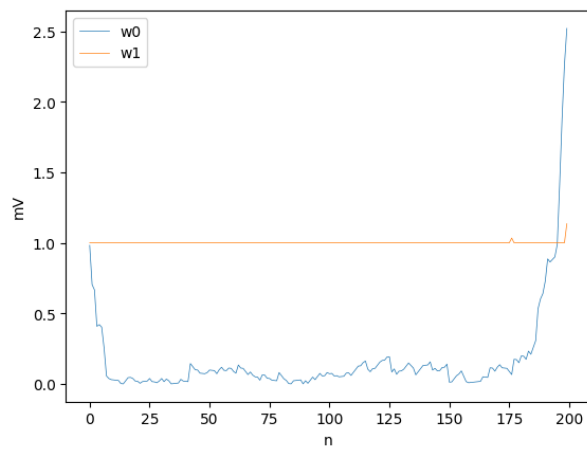


Рис. 5: Векторы  $w_0$  и  $w_1$

#### 4.4 Анализ регрессионных остатков

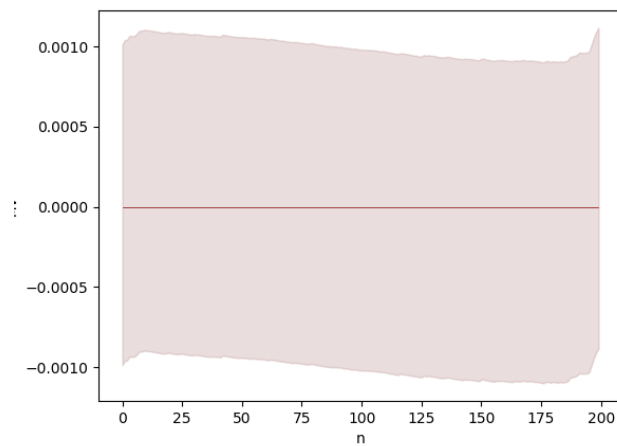


Рис. 6: Диаграмма рассеяния регрессионных остатков для модели без сужения интервалов

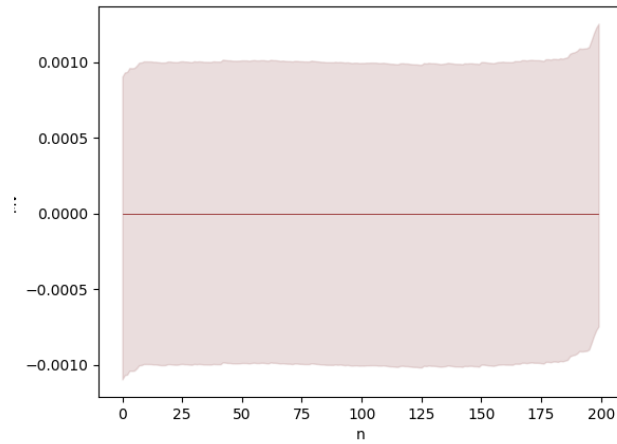


Рис. 7: Диаграмма рассеяния регрессионных остатков для модели с сужением и расширением интервалов

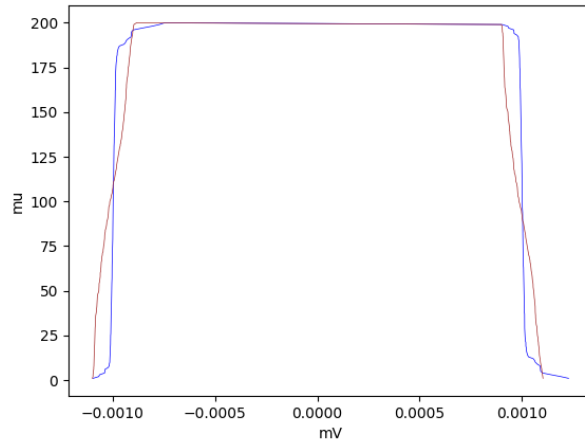


Рис. 8: Частоты элементарных подинтервалов регрессионных остатков при вычислении моды для двух моделей

**Красный график** – график частот элементарных подинтервалов регрессионных остатков при вычислении моды модели без сужения интервалов.

**Синий график** – график частот элементарных подинтервалов регрессионных остатков при вычислении моды модели с сужением и расширением интервалов.

Меры совместности регрессионных остатков:

$$\text{mode } \mathbf{X}^0 = [-0.0007, 0.0009] \quad J_i(\mathbf{X}^0) = 0.7020$$

$$\text{mode } \mathbf{X}^1 = [-0.0008, 0.0009] \quad J_i(\mathbf{X}^1) = 0.8043$$

Здесь  $\mathbf{X}^{0,1}$  – регрессионные остатки выборки  $\mathbf{X}_1$ , вычисленные с использованием разных условий оптимизации.

## 4.5 Информационное множество задачи

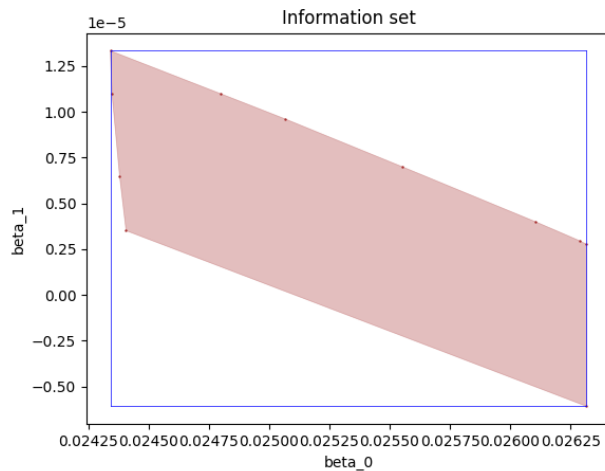


Рис. 9: Информационное множество задачи по заданной модели – красный брус, интервальная оболочка – синий брус

## 4.6 Коридор совместных зависимостей

Внешние интервальные оценки параметров модели:

$$\text{mid}\beta_0 = [2.4341e - 02, 2.6317e - 02]$$

$$\text{mid}\beta_1 = [-6.0905e - 06, 1.3344e - 05]$$

Подставляя эти значения в уравнение регрессии, получаем:

$$\mathbf{x}(k) = \text{mid}\beta_0 + \text{mid}\beta_1 \cdot k \quad (9)$$

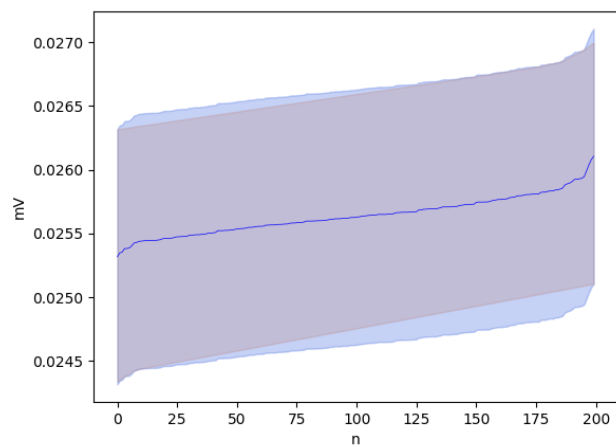


Рис. 10: Коридор совместных зависимостей

#### 4.7 Прогноз вне области данных

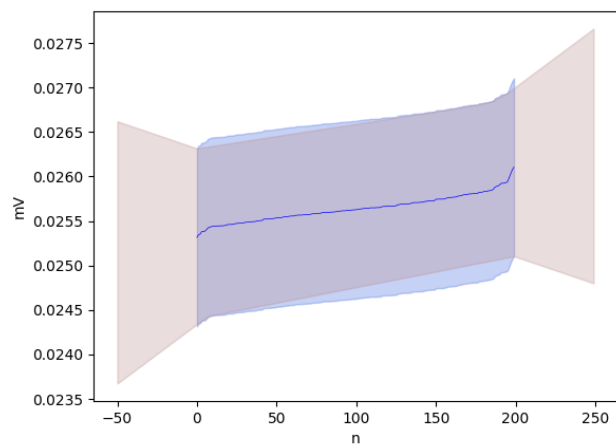


Рис. 11: Коридор совместных зависимостей. Построение прогноза



## 5 Обсуждение

### 5.1 Варьирование неопределенности изменений

Почти все компоненты вектора  $w$  оказались равны 1, то есть, расширения интервалов измерений почти не понадобилось. Недостатком полученного значения с единичными значениями  $w_1^i$  является неучет расстояний точек регрессионной зависимости до данных интервальной выборки. Данная модель "не чувствует" отклонений измерений от прямой на концах выборки — неопределенности измерений достаточно велики, чтобы покрыть этот эффект.

### 5.2 Варьирование неопределенности изменений с расширением и сужением интервалов

В результате построения модели небольшая часть интервалов (от 197 до 200) расширилась, остальные интервалы сузились. Сумма компонент вектора  $w$  уменьшилась более чем в 6 раз. Таким образом, постановка задачи с возможностью одновременного расширения и сужения радиусов неопределенности измерений позволяет более гибко подходить к задаче оптимизации.

Из графиков векторов  $w_0$  и  $w_1$  можно сделать вывод, что график вектора  $w_0$  содержит большее количество информации, чем график вектора  $w_1$ .

### 5.3 Анализ регрессионных остатков

Интервальные выборки остатков получились с весьма разными свойствами. Диаграмма рассеяния регрессионных остатков для модели с сужением и расширением интервалов выглядит более естественно, хотя из-за высокой степени совместности исходной выборки разница незначительна.

Данный вывод подтверждает и график частот элементарных подинтервалов регрессионных остатков при вычислении моды моделей. Графики очень похожи, однако график для частот элементарных подинтервалов регрессионных остатков при вычислении моды модели с расширением и сужением интервалов имеет немного более широкую внутреннюю оценку, что соответствует большей устойчивости к возмущениям данных.

Обе интервальные выборки остатков имеют довольно высокий коэффициент Жакара, что свидетельствует о высокой степени совместности выборки.

### 5.4 Информационное множество задачи

Информационное множество для представляет собой многогранный брус, а интервальные оценки, именуемые интервальной оболочкой, задают брус, описанный вокруг многогранного множества.

### 5.5 Коридор совместных зависимостей

Из внешнего вида коридора совместных зависимостей можно сделать вывод, что внутри его можно провести множество прямых и что он покрывает почти всю интервальную выборку.

## 5.6 Прогноз вне области данных

Была расширена область определения аргумента для модели и построен прогноз за пределами интервальной выборки. На основе полученных результатов можно сделать вывод, что величина неопределенности растет по мере удаления от области, в которой производились исходные измерения. Это обусловлено видом коридора зависимостей, расширяющимся за пределами области измерений.