Санкт-Петербургский политехнический университет Петра Великого
Высшая школа интеллектуальных систем и суперкомпьютерных технологий

Расчетная работа №3б

Дисциплина: Теория вероятности и математическая статистика

Тема: Аппроксимация результатов измерений зависимых переменных

Выполнил
студент гр. 3530901/90003 __________________________ Руднев А.К.
(подпись)

Преподаватель __________________________ Никитин К.В.
(подпись)

«___» _______________ 2021 г.

Санкт-Петербург
2021

Содержание

1. Техническое задание

1.1 Исходные данные:

Исходные данные

В результате измерений при значениях независимой переменной

$$x_1, x_2, \dots x_k$$

получены следующие данные:

$$\begin{array}{l} y11, y21, \dots, yk1 \\ y12, y22, \dots, yk2 \\ \dots\dots\dots\dots\dots\dots \\ y1n, y2n, \dots, ykn \end{array}$$

Рис. 1 – Исходные данные

1.2 Задание

Задание

1. Вычислить в каждой точке x_i средние арифметические значения $\overline{yi}$, оценки дисперсий s_i^2, параметрические толерантные пределы для погрешностей, доверительные интервалы для математических ожиданий, проверить гипотезу о равенстве дисперсий в этих точках по критерию Кочрена (см. приложение 3);

2. Произвести последовательную полиномиальную аппроксимацию Прим. В качестве значений y при аппроксимации необходимо использовать средние арифметические значения $\overline{yi}$ (см. приложение 3).

2.1. Начать с нулевой степени полинома

2.2. вычислить оценки коэффициентов a полинома МНК или МНД (в зависимости от исхода проверки гипотез о равенстве дисперсий) для заданной степени полинома.

Зам. При определении a требуется обращать матрицу S_ε . Если n>k, то можно посчитать обратную матрицу S_ε^{-1} по исходной матрице S_ε , поскольку определитель S_ε не равен 0. В противном случае в качестве S_ε используется диагональная матрица с дисперсиями в каждой из k точек на диагонали.

2.3. Проверить гипотезу о степени q полинома, и если она не будет отвергнута, оценить дисперсии s_{ai}^2 и ковариационную матрицу оценок коэффициентов S_a, в противном случае увеличить степень полинома. Для проверки гипотезы используется критерий Фишера. Если число измерений n больше величины $k - q - 1$, то статистикой критерия является выражение $F = \frac{(n-k+q+1)}{(n-q-1)(n-1)} R^2$, в противном случае $F = \frac{R^2}{(k-q-1)}$.

Зам. Размерность S_a равна (q+1)*(q+1), в то время как размерность S_ε равна k*k.

2.4. Вычислить корреляционную матрицу R_a и коэффициенты корреляции $r_a(i, j)$ между оценками коэффициентов по матрице ковариации

2.5. Пусть была получена степень q полинома, прошедшая гипотезу о степени полинома. Произвести все те же действия для полинома степени, равной k-1 (вычислить коэффициенты и корреляцию между ними). Сравнить результаты для степени q и k-1 (качество аппроксимации, корреляционная матрица коэффициентов, матрица ковариации исходных данных S_ε и ее обусловленность). См указания в приложении.

3. Произвести аппроксимацию исходной зависимости другими способами. Представить полученные графики аппроксимации (полученная аппроксимирующая кривая одним цветом, точки, по которым проводилась аппроксимация маркерами одного типа и все исходные точки маркерами другого типа). Проанализировать и сравнить полученные результаты.

3.1 Произвести аппроксимацию зависимости прямой линией с помощью функций regress (использует метод R-Square), robustfit (робастная регрессия), polyfit (полиномиальная регрессия с n=1), ridge (ридж-регрессия с регуляризацией). Проанализировать полученные результаты.

3.2. Произвести полиномиальную аппроксимацию с помощью функций polyfit (polyval). Можно воспользоваться утилитой polytool, являющейся графическим интерфейсом к polyfit. Подобрать степень полинома, наилучшим способом аппроксимирующую исходную зависимость.

3.3. Произвести кусочную полиномиальную аппроксимацию с помощью функций interp1 (линейная, кубическая), pchip (полиномами Эрмита), spline (сплайны). Сравнить качество аппроксимации с предыдущими результатами.

3.4 Произвести нелинейную аппроксимацию с помощью функции nlinfit. В качестве нелинейной функции использовать произведение полинома на гармоническую функцию:

$$y(x) = (\sin\alpha x + \beta)(a_n x^n + a_{n-1} x^{n-1} + \ldots + a_1 x + a_0)$$

Рис. 2.1 – Задание (часть 1)

4. В выводах детально сравнить все использованные способы аппроксимации зависимостей, выделить преимущества и недостатки каждого из методов в смысле качества аппроксимации, трудоемкости вычислений и других факторов.

Рис. 2.2 – Задание (часть 2)

1.3 Формулы:

Формула для вычисления среднего значения:

$$\bar{y}_i = \frac{1}{n}\sum_{i=1}^{n} y_i$$

Формула для вычисления матрицы суммы:

$$\Sigma_{ii} = \frac{1}{n-1}\sum_{i=1}^{n}(y_i - \bar{y}_i)^2$$

Формула для вычисления статистики Кочрена:

$$g = \frac{s_{max}^2}{\sum_{i=1}^{n} s_i^2}$$

Формула для вычисления коэффициентов полинома:

$$\vec{a} = (X^T\Sigma^{-1}X)^{-1}X^T\Sigma^{-1}\bar{Y}$$

$$X = \begin{pmatrix} 1 & \cdots & x_1^q \\ \vdots & \ddots & \vdots \\ 1 & \cdots & x_k^q \end{pmatrix}$$

Формула для вычисления ковариационной матрицы:

$$S_{\bar{a}} = \frac{1}{n}(X^T\Sigma^{-1}X)^{-1}$$

Формула для вычисления корреляционной матрицы:

$$r_a(i,j) = \frac{k_{ij}}{\sqrt{\sigma_i^2\sigma_j^2}}$$

2. Первоначальные вычисления:

Вычислю в каждой точке x_i арифметические значения $\overline{y}_1$, оценки дисперсий s_i^2 и занесу полученные данные в таблицу 1.1

Таблица 1.1 – Вычисленные среднее значения

k	x_i	$\overline{y}_1$	s_i^2
1	-2	-9,373178	75.67656636117334
2	-1,9	-17,543774	89.53852713236
3	-1,8	-23,24373	44.08986539344444
4	-1,7	-28.007453000000005	124.28302548320113
5	-1,6	-27.890210000000003	63.869046014333335
6	-1,5	-27.75453	73.56116173788888
7	-1,4	-29.768601	105.66116462623222
8	-1,3	-18.517525	110.32705043069444
9	-1,2	-12.882235399999999	173.4619198680827
10	-1,1	-4.492787999999999	82.01255616606221
11	-1	8.974779000000002	78.77227289461001
12	-0,9	10.559793899999999	107.85857055712768
13	-0,8	19.866336	107.36334752940444
14	-0,7	16.685752	41.693446639906675
15	-0,6	21.96726	48.55289561155556
16	-0,5	24.92967	60.11915575122222
17	-0,4	17.627464999999997	99.45092714060553
18	-0,3	7.3982275	35.678336672695835
19	-0,2	10.626738999999997	41.93078178674333
20	-0,1	8.1122294	73.47849869025693
21	0	-4.6322408600000005	142.61863178645763
22	0,1	-4.763267999999999	45.47988741517333
23	0,2	1.6043634	33.16286267693782

24	0,3	-6.9361063099999996	64.3623096137037
25	0,4	4.356893	70.0198507007789
26	0,5	2.8246579	46.02998594560854
27	0,6	11.366628	93.0203040228845
28	0,7	5.1475919999999995	44.02371998146222
29	0,8	9.983110700000001	67.03111803653378
30	0,9	12.509285000000002	70.94733591242779
31	1	6.9991338999999995	83.42743248698898
32	1,1	-9.212409	38.017857175765556
33	1,2	-18.0272933	133.42987087455558
34	1,3	-34.51389	42.115206272111124
35	1,4	-41.84132	57.629727775111114
36	1,5	-51.042860000000005	55.50510209822223
37	1,6	-64.57597	74.4897733201111
38	1,7	-62.44433999999999	130.82161568266667
39	1,8	-61.18922	63.65165504177781
40	1,9	-44.32451999999999	92.73729202622224
41	2	-26.652479999999997	37.988821337333334

Параметрические толерантные пределы для погрешностей, в также доверительные интервалы приведены на рисунке 1.

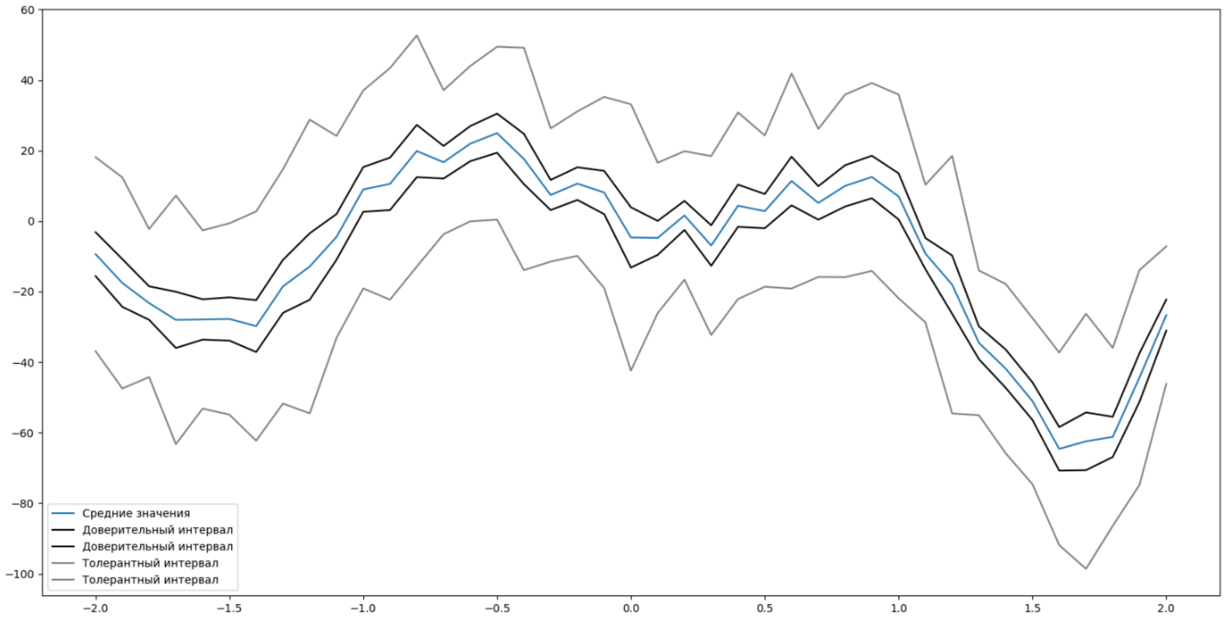

Рис. 2.1 – Доверительные интервалы и толерантных пределов

Для проверки гипотезы о равенстве дисперсий в точках по критерию Кочрена (с случае равноточности используется простой метод МНК, а в противном случае – более сложный ОМНК). Рассчитанное значение статистики критерия сравнивается с критическим значением. Если оно меньше, то делается вывод о том, что нулевая гипотеза не противоречит экспериментальным данным, и для аппроксимации используется МНК (Критическое значение критерия Кочрена взято из таблицы и равно при $\alpha = 0.95 = 0.07$):

`0.05552754704144183 < 0.07`

Рис. 2.2 – Проверка гипотеза Кочрена

Из полученных результатов можно сделать вывод, что рассчитанное значение статистики критерия меньше, чем критическое, поэтому для аппроксимации надо использовать МНК. Гипотеза о равенстве условиях в точках по Кочрену соблюдается.

3. Последовательная полиномиальная аппроксимация

Начну с нулевой степени полинома и продолжу её увеличивать на 1, так как были взяты полиномы от 1 до 39. Корреляционная диагональная матрица приведена на рисунке 3.1 и 3.2.

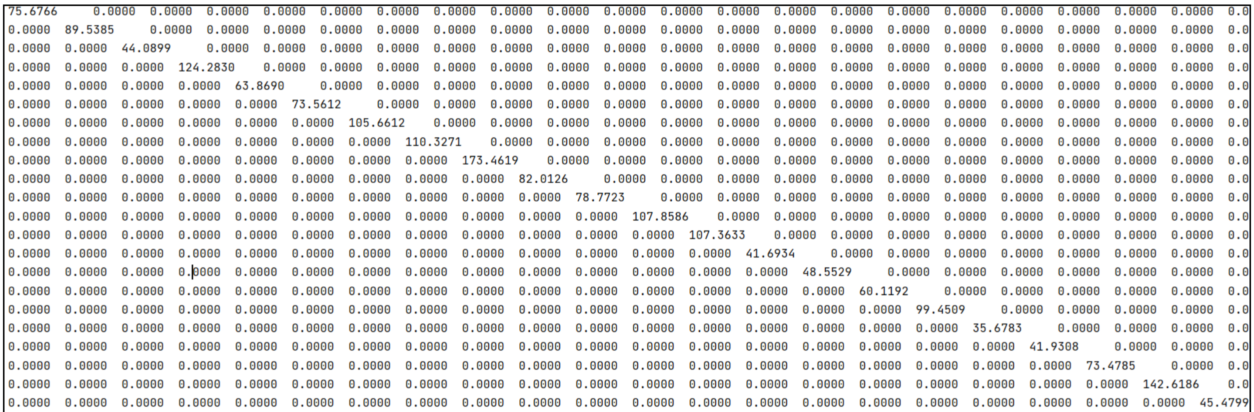

Рис. 3.1 - Корреляционная диагональная матрица (Часть 1)

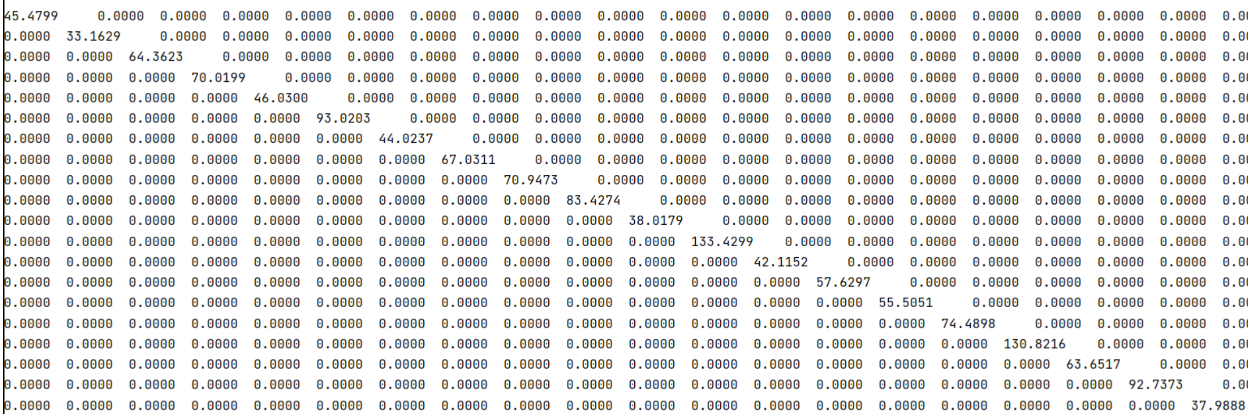

Рис. 3.2 - Корреляционная диагональная матрица (Часть 2)

Так как количество измерений в каждой точке меньше количества точек, то целесообразно использовать количество степеней свободы $k - p - 1$, $n - 1$. Критические значения и статистики Фишера представлены в таблице 1.

Таблица 1– Критические значения и статистики Фишера

q	$F_{крит}$	$F_{стат}$
7	1.7106110393385612	5.478039362613571
8	1.7116949492266318	2.4386047095531684
9	1.7128429767898328	2.121988141050622

10	1.7140609891666814	2.1420130415988377
11	1.715355592267296	2.2137360429923354
12	1.7167342508682548	2.2670721357606505

Пример полученного полинома 9 степени:

$$y(x) = 0.4775 - 27.568x - 65.4376x^2 + 58.2865x^3 - 88.4340x^4 - 42.7463x^5 - 28.9137x^6 + 11.3578x^7 - 2.7986x^8 - 0.9872x^9$$

Зависимость статистики Фишера от степени представлена на рис. 3.3

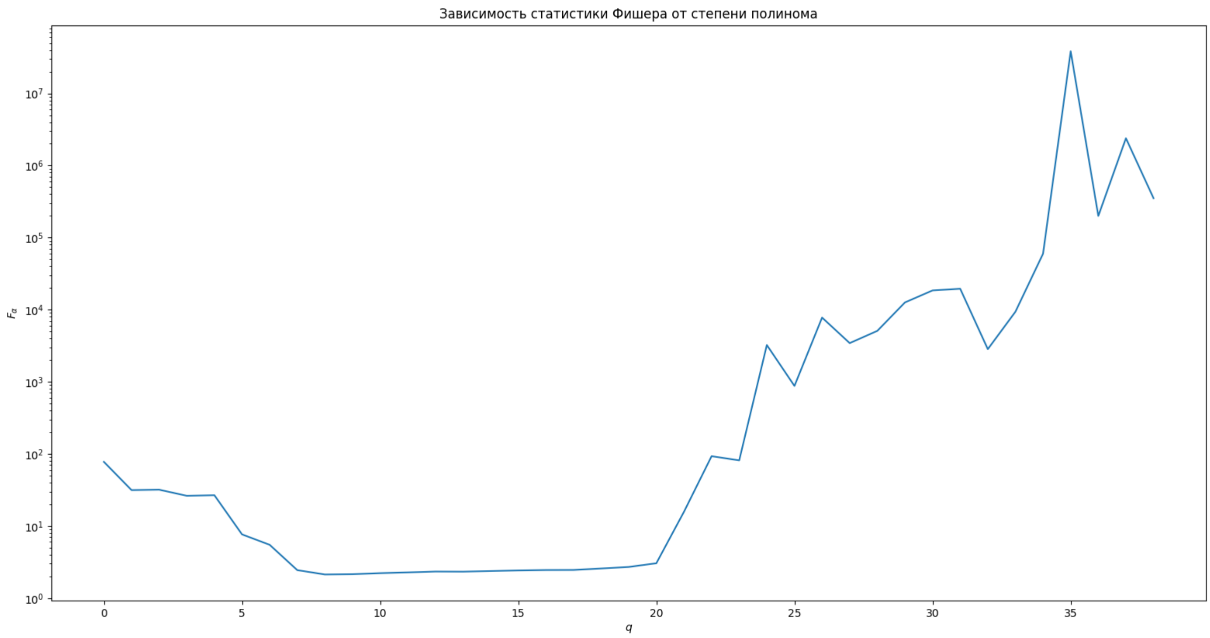

Рис. 3.3 – Зависимость статистики Фишера от степени полинома

Из полученного графика можно сделать вывод, что низкие степени полинома плохо приближают график, это и касается высоких степеней, так как они сильно зависят от погрешности исходных данных. На рисунке 3.3 можно увидеть, что наилучшее приближение дают полиномы со степенями [8, 18].

Для полинома со степенью 8 была построена ковариационная и корреляционная матрицы оценок коэффициентов. Ковариационная матрица приведена в таблице 2.

Таблица 2 – Ковариационная матрица для полинома степени 9

0.8285	0.0033	-3.1022	0.0229	2.9690	-0.0178	-1.0295	0.0028	0.1174	0.8285
0.0033	4.4665	-0.1059	-7.2967	0.3249	3.2947	-0.1628	-0.4415	0.0227	0.0033
-3.1022	-0.1059	21.0435	0.0752	-23.8280	0.0172	8.9633	-0.0052	-1.0726	-3.1022
0.0229	-7.2967	0.0752	14.6151	-0.8048	-7.2312	0.4551	1.0207	-0.0669	0.0229
2.9690	0.3249	-23.8280	-0.8048	29.5676	0.3536	-11.7408	-0.0474	1.4555	2.9690
-0.0178	3.2947	0.0172	-7.2312	0.3536	3.7631	-0.2151	-0.5486	0.0329	-0.0178
-1.0295	-0.1628	8.9633	0.4551	-11.7408	-0.2151	4.8285	0.0302	-0.6133	-1.0295
0.0028	-0.4415	-0.0052	1.0207	-0.0474	-0.5486	0.0302	0.0818	-0.0048	0.0028
0.1174	0.0227	-1.0726	-0.0669	1.4555	0.0329	-0.6133	-0.0048	0.0793	0.1174
0.8285	0.0033	-3.1022	0.0229	2.9690	-0.0178	-1.0295	0.0028	0.1174	0.8285

Далее была построена корреляционная матрица для полинома 9 степени

Таблица 3 – Корреляционная матрица для полинома степени 9

1.0000	0.0017	-0.7430	0.0066	0.5999	-0.0101	-0.5147	0.0108	0.4581	1.0000
0.0017	1.0000	-0.0109	-0.9031	0.0283	0.8036	-0.0351	-0.7304	0.0382	0.0017
-0.7430	-0.0109	1.0000	0.0043	-0.9553	0.0019	0.8892	-0.0040	-0.8303	-0.7430
0.0066	-0.9031	0.0043	1.0000	-0.0387	-0.9751	0.0542	0.9336	-0.0621	0.0066
0.5999	0.0283	-0.9553	-0.0387	1.0000	0.0335	-0.9826	-0.0305	0.9506	0.5999
-0.0101	0.8036	0.0019	-0.9751	0.0335	1.0000	-0.0505	-0.9889	0.0602	-0.0101
-0.5147	-0.0351	0.8892	0.0542	-0.9826	-0.0505	1.0000	0.0481	-0.9912	-0.5147
0.0108	-0.7304	-0.0040	0.9336	-0.0305	-0.9889	0.0481	1.0000	-0.0592	0.0108
0.4581	0.0382	-0.8303	-0.0621	0.9506	0.0602	-0.9912	-0.0592	1.0000	0.4581
1.0000	0.0017	-0.7430	0.0066	0.5999	-0.0101	-0.5147	0.0108	0.4581	1.0000

Графическое сравнение полиномов представлено на рисунке 3.4

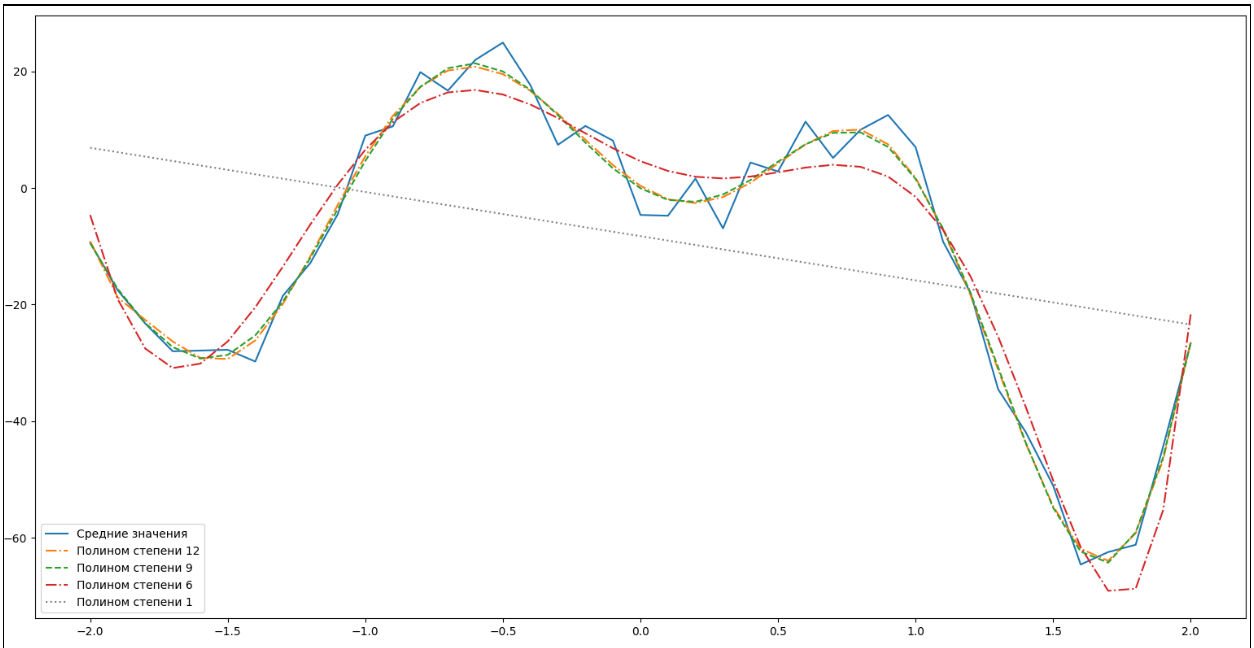

Рис. 3.4 – График полиномиальной аппроксимации разных степеней

Сравнение полученных результатов:

Из полученных результатов можно сделать вывод, что полиномиальная аппроксимация хорошо справляется с приближением, в частности для примера со степенью 9, но полиномы низкой степени (например, 1) не могут достаточно приближать функцию.

4. Аппроксимацию исходной зависимости другими способами

4.1 Аппроксимация зависимости прямой линией

Аппроксимация зависимости прямой линией будет осуществляться с помощью функции regress (используем метод R-Square), robustfit (робастная регрессия), polyfit (полиномиальная регрессия с n=1), ridge (ридж-регрессия с регуляризацией).

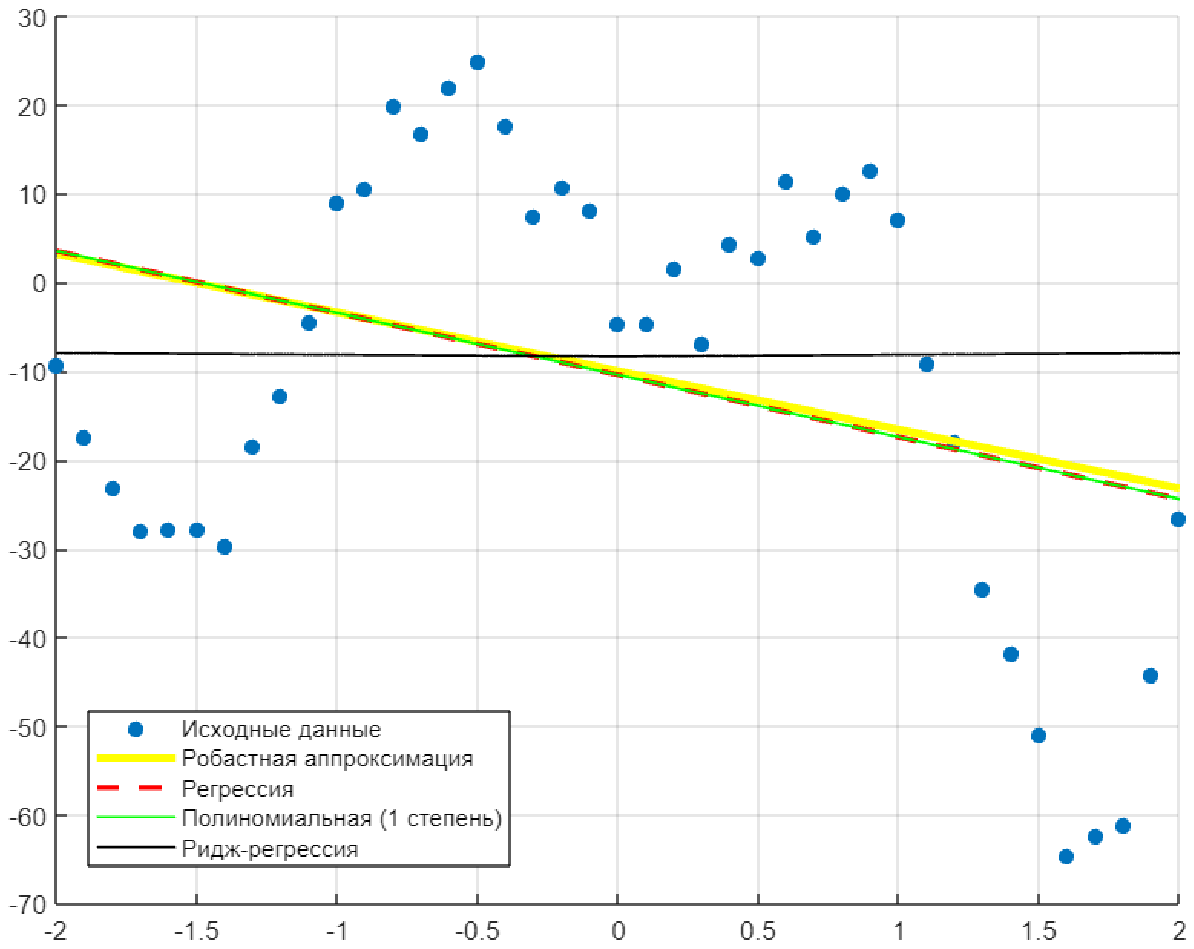

Рис. 4.1 – Аппроксимация зависимости прямой

Из полученных результатов можно сделать вывод, что линейная аппроксимация вообще не справляется с аппроксимируемой нелинейной функцией.

4.2 Полиномиальная аппроксимация

Полиномиальная аппроксимация будет осуществлена с помощью функций polyfit (polyval). Будут провидены графики для разных степеней полинома.

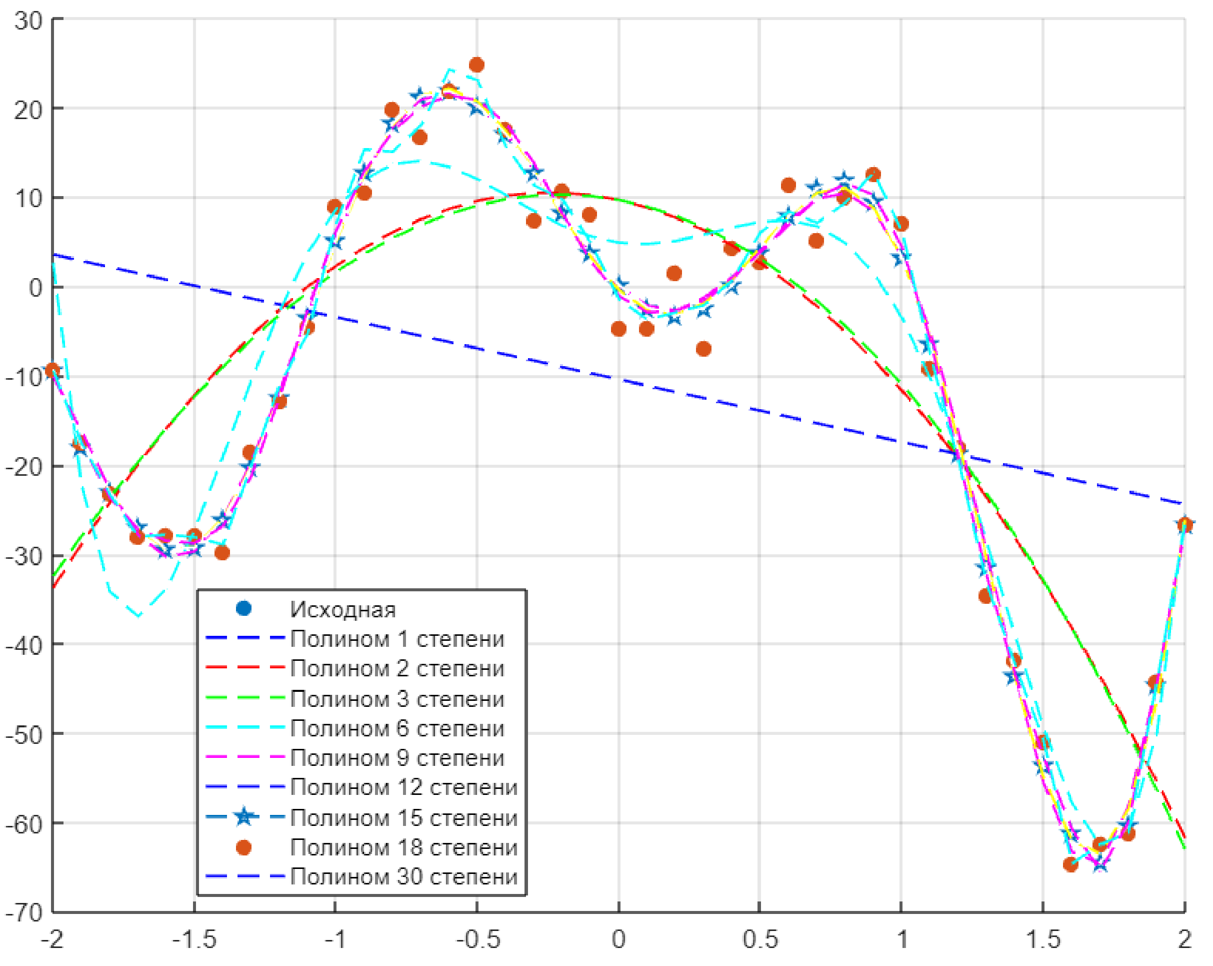

Рис. 4.2 – График полиномиальной аппроксимации

Полиномиальная аппроксимация хорошо аппроксимирует функцию, начиная с 9 степени. Результаты аппроксимация похожи на результаты полученные в пункте 3.

4.3 Кусочная полиномиальная аппроксимация

Кусочная полиномиальная аппроксимация будет осуществляться с помощью функций interp1 (линейная, кубическая), pchip (полиномами Эрмита), spline (сплайны).

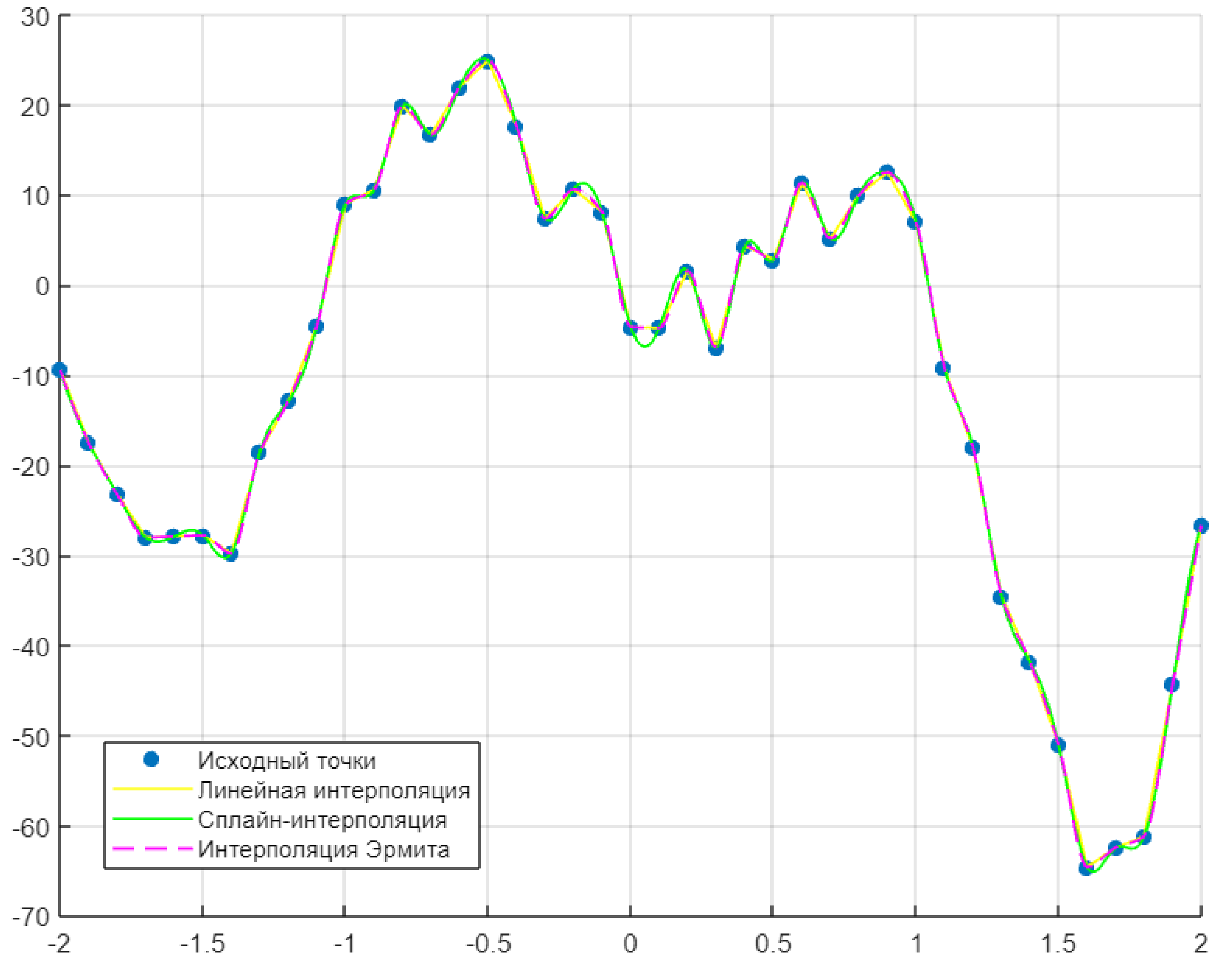

Рис. 4.3 – График кусочной полиномиальной аппроксимации

Из полученных результатов можно сказать, что кусочная полиномиальная аппроксимация очень хорошо справляется со своей задачей и почти идеально аппроксимирует функцию.

4.4 Нелинейная аппроксимация с помощью функции nlinfit

Нелинейная аппроксимация будет осуществляться с помощью функции nlinfit, а в качестве функции будет использоваться произведение полинома на гармоническую функцию.

$$y(x) = (\sin(\alpha x) + \beta)(a_n x^n + a_{n-1} x^{n-1} + \cdots + a_1 x + a_0)$$

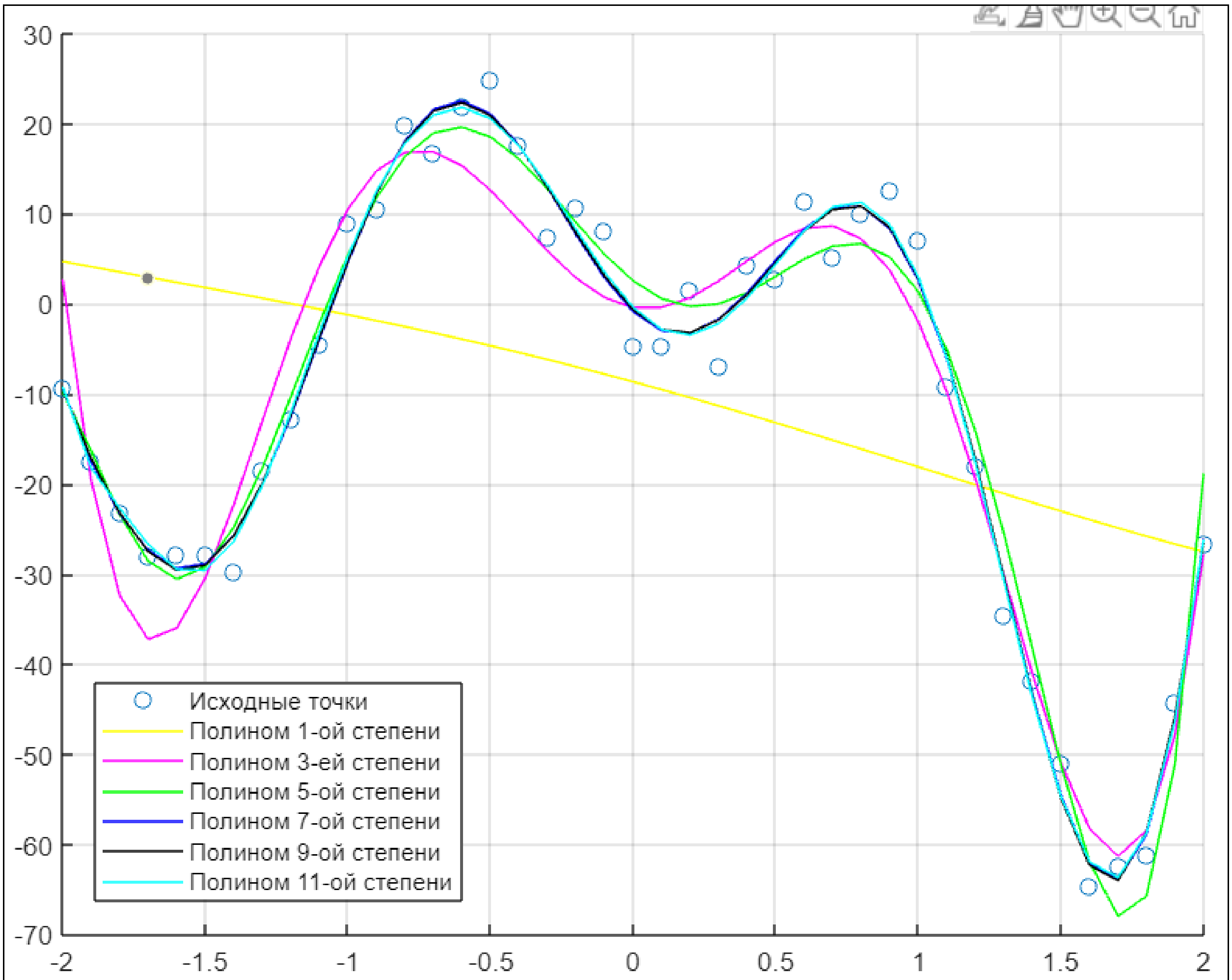

Рис. 4.4 – Нелинейная аппроксимация с помощью функции nlinfit

Результаты практически идентичные, что и у обычной полиномиальной аппроксимации.

5. Вывод

В ходе выполнения работы было проведено исследование над выборкой значений. Были вычислены средние значение и для них построены доверительные и толерантные интервалы.

Выборка из средних значений была аппроксимирована полиномами различной степени. Качество аппроксимации оценивалось с помощью критерия Фишера. Наилучшими оказался аппроксимация полиномом степени 9, так как даёт наименьший показатель статистики. Низкие степени полиномов плохо приближают функцию, а высокие очень чувствительны к погрешностям, поэтому и полином со степенью 9 приближает функцию лучше всего.

Также были применены различные виды аппроксимации с помощью MATLAB:

1) Линейная аппроксимация – никак не смогла приблизить функцию

2) Полиномиальная аппроксимация – хорошо приблизила функцию, также как и было показано в пункте 2, ведь слишком маленькая степень (например, 2 или 3) вообще не справляются с задачей, а слишком большая степень делает это достаточно плохо.

3) Кусочная полиномиальная аппроксимация – довольно хорошо приблизила функцию. Можно сказать, что это наилучший вариант среди других.

4) Нелинейная аппроксимация с помощью гармонической функции – хорошо приблизила функцию, но только не с маленькими степенями.

Таким образом, наилучшим способом аппроксимируют функцию два метода: кусочно-полиномиальная аппроксимация и полиномиальная аппроксимация, но стоит учитывать её зависимость от степени полинома