Министерство науки и высшего образования Российской Федерации Федеральное государственное бюджетное образовательное учреждение высшего образования «Московский государственный технический университет имени Н.Э. Баумана (национальный исследовательский университет)» (МГТУ им. Н.Э. Баумана)

ФАКУЛЬТЕТ *Робототехники и комплексной автоматизации*

КАФЕДРА *Системы автоматизированного проектирования (РК-6)*

ОТЧЕТ О ВЫПОЛНЕНИИ ЛАБОРАТОРНОЙ РАБОТЫ

по дисциплине: «Вычислительная математика»

Студент	Грехова Евгения Вениаминовна
Группа	Рк6-63б
Тип задания	Лабораторная работа
Тема лабораторной работы	Нелинейная регрессия

Студент ___________ Грехова Е. В.

Преподавател ___________Соколов А.П

Оценка ________________________

Москва 2020г.

Оглавление

Задание

Дана модель экспоненциального роста:

$$\frac{d}{dt}I = (\beta - \gamma)I, \qquad (1)$$

где I- количество инфицированных людей, β - среднее число контактов, приходящихся на человека в единицу времени, и γ - это среднее число выздоровевших, приходящихся на человека в единицу времени (т.е., средняя скорость иммунизации).

Также дана SIS-модель:

$$\frac{d}{dt}I = (\beta - \gamma)I - \frac{\beta}{N}I^2, \qquad (2)$$

где N - число людей в популяции (например, число жителей страны).

Требуется:

1. Найти решение модели экспоненциального роста и SIS-модели при условии, что $I(0) = I_0$ и продемонстрировать детальный вывод этого решения. После этого приведите решение модели экспоненциального роста к форме

$$I(t) = I_0\, e^{\chi t}, \qquad (3)$$

где $\chi = (\beta - \gamma)$, и решение SIS-модели к форме

$$I(t) = \frac{I\infty}{1 + (\frac{I_\infty}{I_0} - 1)e^{-\chi t}} \qquad (4)$$

где I_∞ обозначает предельное значение $I(t)$ при $t \to \infty$, т.е. $I_\infty = \lim_{t\to\infty} I(t)$

2. Сформировать выборку данных зависимости числа информацированных от времени в данной стране. Страна выбирается исходя из первой буквы вашей фамилии.

3. Предполагая, что первые недели распространения вируса в выбранной стране рост числа зараженных описывается экспоненциальной моделью, требуется оценить значение коэффициента $\chi = (\beta - \gamma)$ с помощью нелинейной регрессии и

нормального уравнения, взяв в качестве аппроксимирующей функции решения экспоненциальной модели и начального числа инфицированных I_0=20…30 и вывести на экран в логарифмической шкале полученное решение вместе с исходными дискретными данными о числе инфицированных. Требуется подробно описать формулировку задачи регрессии в вашем случае и явно указать выражения для отдельных векторов и матрицы, входящих в нормальное уравнение.

4. Подставив найденное значение коэффициента $\chi=(\beta-\gamma)$ в решение SIS-модели, оценить значение коэффициента I_∞ с помощью нелинейной регрессии и нормального уравнения, взяв в качестве аппроксимирующей функции решение SIS-модели и вывести на экран полученное решение вместе с исходными дискретными данными о числе инфицированных. Требуется подробно описать формулировку задачи регрессии в вашем случае и явно указать выражения для отдельных векторов и матрицы, входящих в нормальное уравнение.

5. Ответить на вопросы:

(a) Какова погрешность полученной аппроксимации (решения SIS-модели) относительно нормы L_∞ ? Относительно нормы L_2 ?

(b) Какой критерий вы бы использовали для определения максимального числа дней, в течении которых использование модели экспоненциального роста оправдано? Каково такое максимальное число дней в вашем случае в соответсвии с вашим критерием?

(c) Предскажите, сколько человек в соответствии с полученной аппроксимацией (решения SIS-модели) будут и инфицированы в выбранной стране $t\to\infty$ и через сколько дней после начала эпидемии наступит ее окончание.

Цель выполнения лабораторной работы

Цель выполнения лабораторной работы – используя модель экспоненциального роста и SIS-модели проанализировать уже имеющиеся данные и предсказать дальнейшее развитие распространения вируса COVID-19 в Германии.

Выполненные задачи

1. Найдено подробное решение модели экспоненциального роста и SIS-модели.

2. Получены число инфицированных по датам.

3. Найден коэффициент χ из модели экспоненциального роста. Сформулирован критерий максимального количества дней, в течении которых использование модели экспоненциального роста оправдано.

4. Проведены исследование и анализ SIS-модели и исследованы методы улучшения SIS-модели.

5. Проведен анализ решения.

1. Решение модели экспоненциального роста и SIS-модели

1.1 Решение модели экспоненциального роста

Для решения модели экспоненциального роста $\frac{d}{dt}I = (\beta - \gamma)I$ воспользуемся методом разделения переменных:

$$\frac{d}{(\beta - \gamma)I}I = dt \qquad (5)$$

Интегрируем полученное выражение:

$$\int \frac{d}{(\beta - \gamma)I}I = \int dt \qquad (6)$$

Интегрирование левой и правой части приводит к равенству:

$$\frac{ln(I(\beta - \gamma))}{(\beta - \gamma)} = t + C1, \qquad (7)$$

$$ln(I(\beta - \gamma)) = t(\beta - \gamma) + C1 \qquad (8)$$

Выражение:

$I(\beta - \gamma) = e^{t(\beta - \gamma)}C$, где величина константы C= e^{C1} с заданным условием в начальный момент времени $I(0) = I_0$ равна : $C = (\beta - \gamma)I_0$,

принимает вид:

$$I(\beta - \gamma) = e^{t(\beta - \gamma)}(\beta - \gamma)I_0 \qquad (9)$$

$$I = I_0 e^{t(\beta - \gamma)} \qquad (10)$$

$$I = I_0 e^{\chi t}, \qquad (11)$$

где $\chi = (\beta - \gamma)$, что удовлетворяет условию поставленной задачи.

1.2 Решение SIS-модели

Для решения SIS-модели $\frac{d}{dt}I = (\beta - \gamma)I - \frac{\beta}{N}I^2$ воспользуемся методом разделения переменных :

$$\frac{d}{(\beta-\gamma)I - \frac{\beta}{N}I^2} I = dt \qquad (12)$$

Интегрируем полученное выражение:

$$\int \frac{d}{(\beta-\gamma)I - \frac{\beta}{N}I^2} I = \int dt \qquad (13)$$

Интегрирование левой и правой части приводит к равенству:

$$\frac{lnI - ln(-\beta N + \beta I + N\gamma)}{\beta - \gamma} = t + C1 \qquad (14)$$

$$ln\frac{I}{\beta I - \beta N + N\gamma} = t(\beta-\gamma) + C1 \qquad (15)$$

Выражение:

$\frac{I}{I - N\frac{(\beta-\gamma)}{\beta}} = e^{t(\beta-\gamma)}C,$ где величина константы С= e^{C1} с заданным условием в начальный момент времени $I(0) = I_0$ равна : $C = \frac{I_0}{I_0 - N\frac{(\beta-\gamma)}{\beta}}$

принимает вид:

$$\frac{I}{I - N\frac{(\beta-\gamma)}{\beta}} = e^{t(\beta-\gamma)}\frac{I_0}{I_0 - N\frac{(\beta-\gamma)}{\beta}} \qquad (16)$$

С помощью некоторых преобразований выразим I:

$$I(I_0 - \frac{(\beta-\gamma)}{\beta}N) = I_0 e^{t(\beta-\gamma)}(I - \frac{(\beta-\gamma)}{\beta}N) \qquad (17)$$

$$I(I_0 - \frac{(\beta-\gamma)}{\beta}N) = I_0 e^{t(\beta-\gamma)} I - I_0 e^{t(\beta-\gamma)} \frac{(\beta-\gamma)}{\beta}N \qquad (18)$$

$$I(I_0 - \frac{(\beta-\gamma)}{\beta}N - I_0 e^{t(\beta-\gamma)}) = -I_0 e^{t(\beta-\gamma)} \frac{(\beta-\gamma)}{\beta}N \qquad (19)$$

$$I = -\frac{I_0 e^{t(\beta-\gamma)} \frac{(\beta-\gamma)}{\beta}N}{I_0 - \frac{(\beta-\gamma)}{\beta}N - I_0 e^{t(\beta-\gamma)}} \qquad (20)$$

Разделим числитель и знаменатель дроби правой части формулы (20) на ее числитель, получим:

$$I = \frac{1}{\frac{-I_0}{I_0 e^{t(\beta-\gamma)} \frac{(\beta-\gamma)}{\beta}N} + \frac{\frac{(\beta-\gamma)}{\beta}N}{I_0 e^{t(\beta-\gamma)} \frac{(\beta-\gamma)}{\beta}N} + \frac{I_0 e^{t(\beta-\gamma)}}{I_0 e^{t(\beta-\gamma)} \frac{(\beta-\gamma)}{\beta}N}} \qquad (21)$$

Получаем зависимость I(t):

$$I(t) = \frac{1}{\frac{\beta}{(\beta-\gamma)N} + e^{-t(\beta-\gamma)}(\frac{1}{I_0} - \frac{\beta}{(\beta-\gamma)N})} \qquad (22)$$

Т.к. I_∞ обозначает предельное значение $I(t)$ при $t \to \infty$, т.е. $I_\infty = \lim_{t\to\infty} I(t)$, получаем:

$$I_\infty = \lim_{t\to\infty} I(t) = \frac{(\beta-\gamma)}{\beta}N \qquad (23)$$

Учитывая (23) в формуле (22) получаем:

$$I(t) = \frac{1}{\frac{1}{I_\infty} + e^{-t(\beta-\gamma)}(\frac{1}{I_0} - \frac{1}{I_\infty})} \qquad (24)$$

Домножив числитель и знаменатель правой части формулы (24) на I_∞ получаем:

$$I(t) = \frac{I_\infty}{1 + e^{-t(\beta-\gamma)}(\frac{I_\infty}{I_0} - 1)} \tag{25}$$

$$I(t) = \frac{I_\infty}{1 + e^{-t\chi}(\frac{I_\infty}{I_0} - 1)}, \tag{26}$$

где $\chi = (\beta - \gamma)$, что удовлетворяет условию поставленной задачи.

2. Зависимость числа зараженных от даты с начала инфицирования в Германии.

Для формирования выборки данных зависимости числа инфицированных от времени в Германии воспользуемся актуальными данными, которые представлены ниже(даты, когда число инфицированных равнялось 0 отсекаем):

Germany	2020-01-28	1
Germany	2020-01-29	4
Germany	2020-01-30	4
Germany	2020-01-31	5
Germany	2020-02-01	7
Germany	2020-02-02	8
Germany	2020-02-03	9
Germany	2020-02-04	11
Germany	2020-02-05	11
Germany	2020-02-06	11
Germany	2020-02-07	12
Germany	2020-02-08	13
Germany	2020-02-09	13
Germany	2020-02-10	13
Germany	2020-02-11	13
Germany	2020-02-12	15
Germany	2020-02-13	15
Germany	2020-02-14	15
Germany	2020-02-15	15
Germany	2020-02-16	15

Germany	2020-02-17	15
Germany	2020-02-18	15
Germany	2020-02-19	15
Germany	2020-02-20	15
Germany	2020-02-21	15
Germany	2020-02-22	15
Germany	2020-02-23	15
Germany	2020-02-24	15
Germany	2020-02-25	15
Germany	2020-02-26	17
Germany	2020-02-27	21
Germany	2020-02-28	47
Germany	2020-02-29	57
Germany	2020-03-01	111
Germany	2020-03-02	129
Germany	2020-03-03	157
Germany	2020-03-04	196
Germany	2020-03-05	262
Germany	2020-03-06	400
Germany	2020-03-07	684

Germany	2020-03-08	847
Germany	2020-03-09	902
Germany	2020-03-10	1139
Germany	2020-03-11	1296
Germany	2020-03-12	1567
Germany	2020-03-13	2369
Germany	2020-03-14	3062
Germany	2020-03-15	3795
Germany	2020-03-16	4838
Germany	2020-03-17	6012
Germany	2020-03-18	7156
Germany	2020-03-19	8198
Germany	2020-03-20	14138
Germany	2020-03-21	18187
Germany	2020-03-22	21463
Germany	2020-03-23	24774
Germany	2020-03-24	29212
Germany	2020-03-25	31554
Germany	2020-03-26	36508
Germany	2020-03-27	42288
Germany	2020-03-28	48582

Germany	2020-03-29	52547
Germany	2020-03-30	57298
Germany	2020-03-31	61913
Germany	2020-04-01	67366
Germany	2020-04-02	73522
Germany	2020-04-03	79696
Germany	2020-04-04	85778
Germany	2020-04-05	91714
Germany	2020-04-06	95391
Germany	2020-04-07	99225
Germany	2020-04-08	103228
Germany	2020-04-09	108202
Germany	2020-04-10	113525
Germany	2020-04-11	117658
Germany	2020-04-12	120479
Germany	2020-04-13	123016
Germany	2020-04-14	125098
Germany	2020-04-15	127584
Germany	2020-04-16	130450
Germany	2020-04-17	133830
Germany	2020-04-18	137439
Germany	2020-04-19	139897
Germany	2020-04-20	141672
Germany	2020-04-21	143457
Germany	2020-04-22	145694
Germany	2020-04-23	148046
Germany	2020-04-24	150383
Germany	2020-04-25	152438

Необходимые данные находятся в файле owid-covid-data.csv, откуда берем только ячейки, соответствующие выбранной стране «Germany», начиная с дня, когда в стране был обнаружен первый инфицированный, и идем последовательно для получения графика зависимости числа инфицированных «total_cases» от даты «date».

Полученный результат:

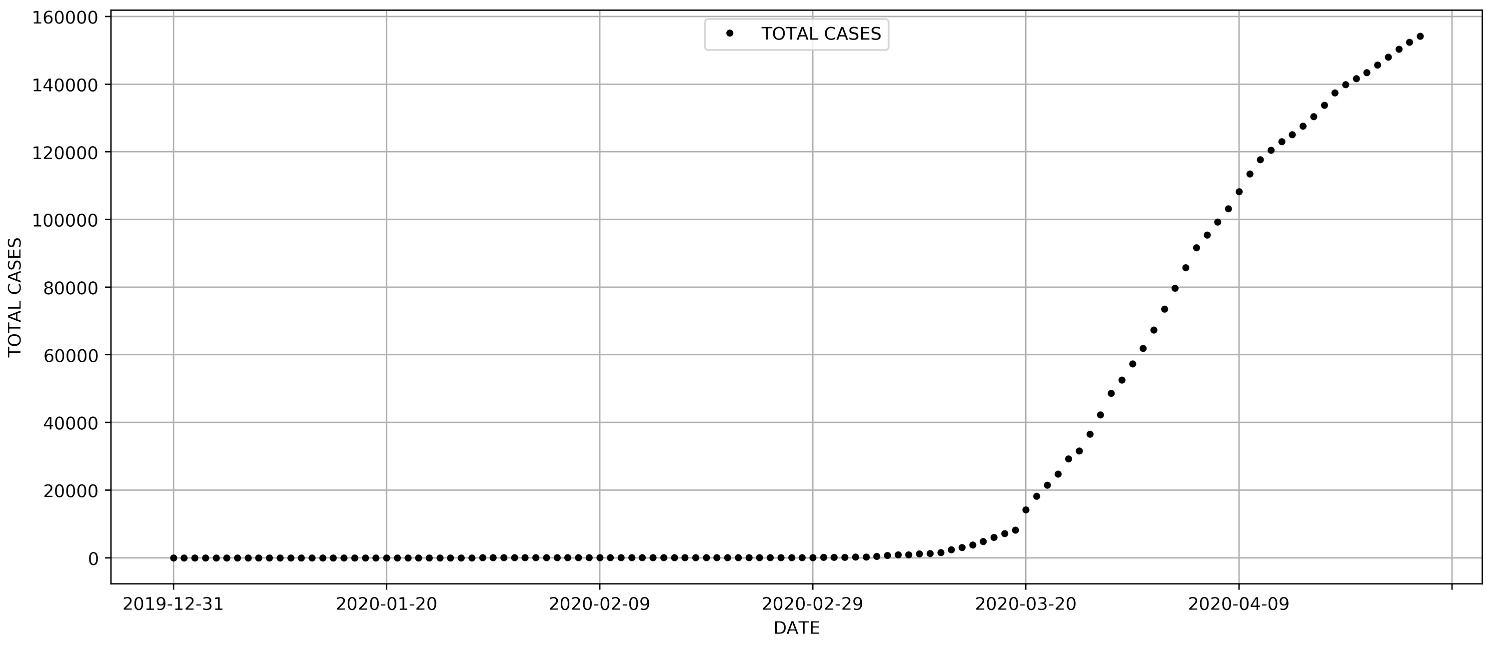

Рисунок 1. График зависимости общего числа инфицированных от времени в Германии

3. Нахождение значения коэффициента $\chi = (\beta - \gamma)$

Необходимо выбрать участок где график на рисунке 1 проявляется в экспоненциальной зависимости, исключив начальные флуктуации, где динамический процесс еще не перешел в экспоненциальный режим. В нашем случае можем заметить, что экспоненциальный рост начинается уже при достаточно большом количестве инфицированных. Первые 2/3 всего промежутка времени с момента пандемии в Гремании, мы видим незначительный рост зараженных, но в марте ситуация меняется. Проанализировав график на риунке 1 для модели экспоненциального роста возьмем начальное число инфицированных $I_0 = 684$ и количество дней, на протяжении которых график наиболее соответсвует модели экспоненциального роста, равное 25.

Чтобы коэффициент χ входил в $I(t) = I_0 e^{\chi t}$ линейно, берем натуральный логарифм от данной формулы, откуда получаем:

$$lnI = lnI_0 + \chi t = \tilde{I} \qquad (27)$$

$Y = \{y_i\}$, где i от 1 до последнего, когда еще наблюдается экспоненциальный рост модели

$$\tilde{y_i} = lnY - lnI_0 \qquad (28)$$

Метод наименьших квадратов (для минимизации суммы квадратов отклонений):

$$\sum_{i=1}^{n} (\tilde{y} - \tilde{I})^2 \longrightarrow \min \qquad (30)$$

Нормальное уравнение:

$$\chi = (X^T X)^{-1} X^T \tilde{y} \qquad (30)$$

Матрицы X и X^T:

$$X = \begin{pmatrix} t1 \\ t2 \\ \dots \\ tn \end{pmatrix} \qquad X^T = (t1, t2, \dots, tn)$$

Из нормального уравнения получим:

$$\chi = \frac{\sum_{i=1}^{n} t_i \tilde{y}}{\sum_{i=1}^{n} t_i^2} \qquad (31)$$

Откуда находим значение χ=0.21706406632246159

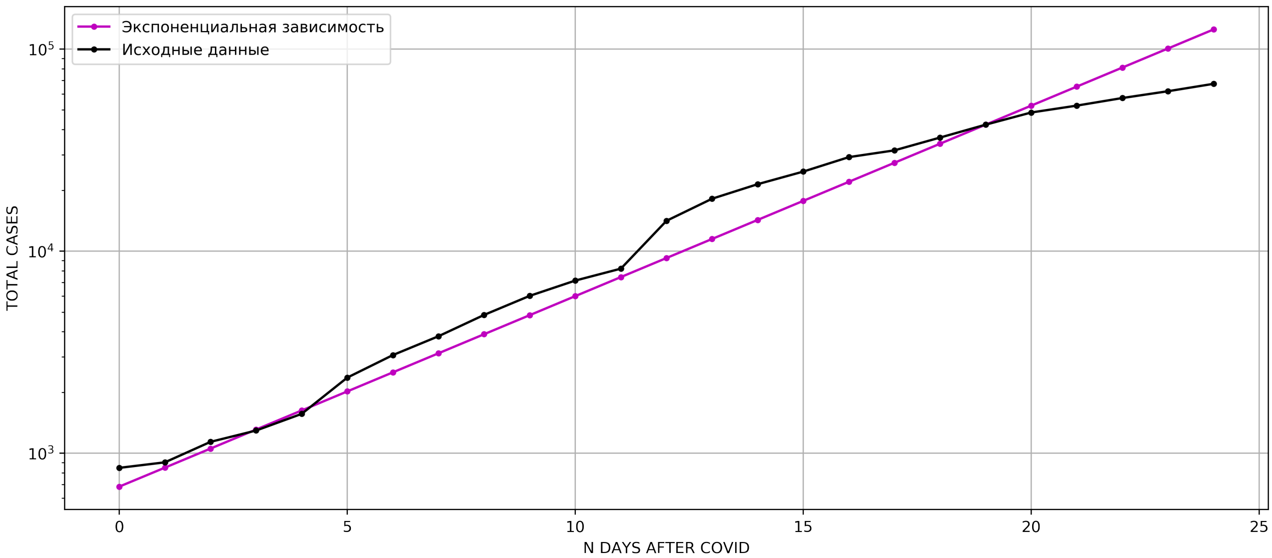

Рисунок 2. График модели экспоненциального роста

Если брать в рассмотрение меньшее количество дней, график исходных данных становится ближе к экспоненциальной модели роста (рисунок 2), но это не подходит для нашего случая, иначе появляются значительные расхождения с данными, уже имеющимися на сегодняшний день, при расчете числа заболевших при $t \to \infty$.

4. Анализ и решение SIS-модели

Используя найденное ранее значение коэффициента χ =($\beta - \gamma$) в решении SIS-модели, оценим значение коэффициента I_∞ с помощью нелинейной регрессии и нормального уравнения, взяв в качестве аппроксимирующей функции решение SIS-модели : $I(t) = \dfrac{I_\infty}{1 + e^{-t\chi}(\frac{I_\infty}{I_0} - 1)}$.

Т.к. решение SIS модели является нелинейным, то мы не можем без потери качества аппроксимации выразить коэффициент I_∞ так, что он входит линейно в формулу для решения. Поэтому в данном случае требуется требуется только сформулировать метод наименьших квадратов, а оптимальное значение I_∞ найдем с помощью минимизации суммы квадратов отклонений:

$$\sum_{i=1}^{n}(y_i - I(t_i))^2 \longrightarrow \min \qquad (32)$$

Найденные значения I_∞, I_0:

$I_\infty = 238515$
$I_0 = 769$

Используя полученные значения для I_∞, I_0 получим график, представленный на рисунке 3:

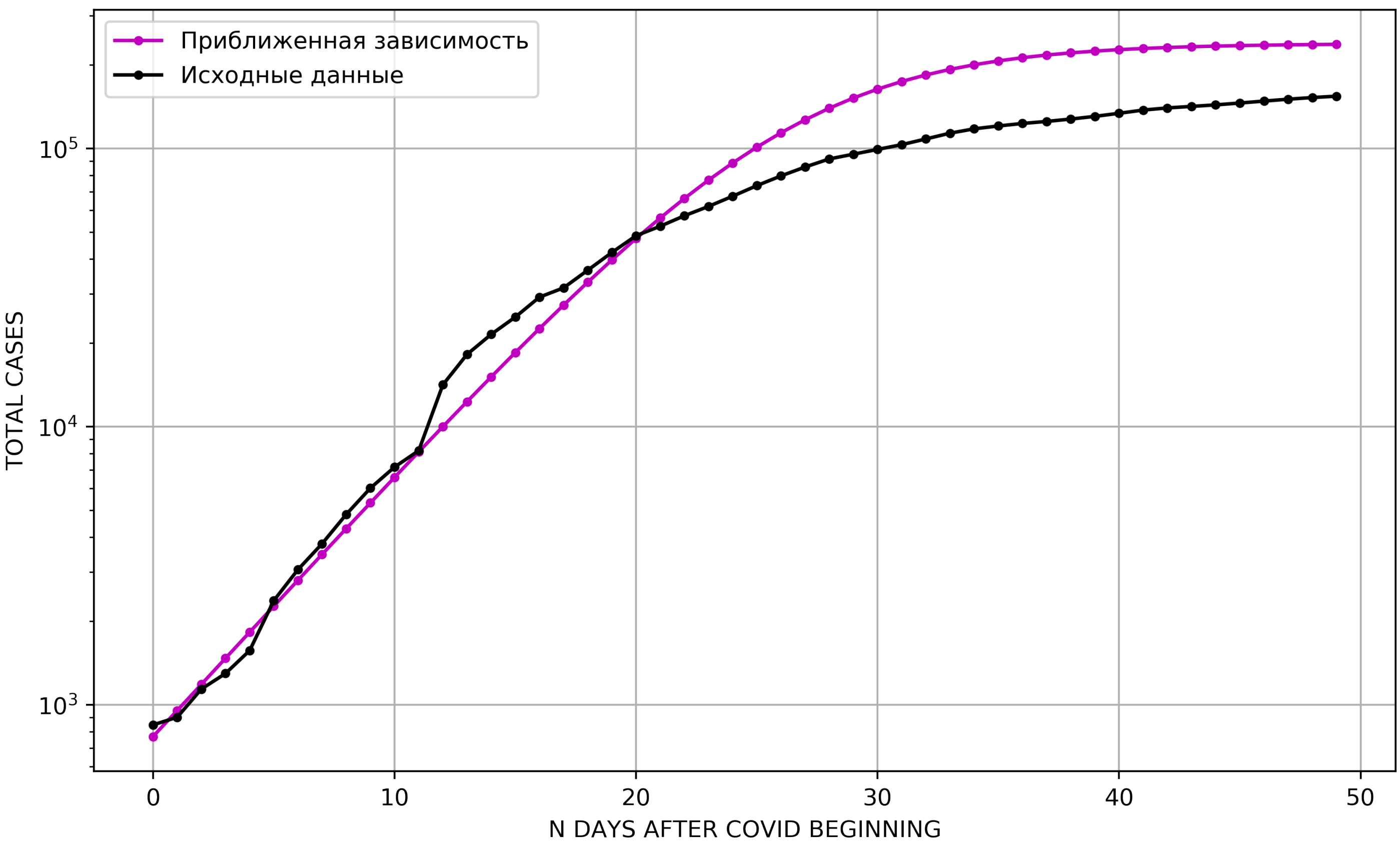

Рисунок 3. График SIS-модели

Данные остались приблизительно равными уже полученным I_∞, I_0. Значения I_∞, I_0 после оптимизации:

$I_\infty \approx 238515$
$I_0 \approx 760$

Однако на рисунке 4 можем заметить, что расхождение между приближенной зависимость и исходными данными в оптимизированном графике SIS-модели уменьшилось

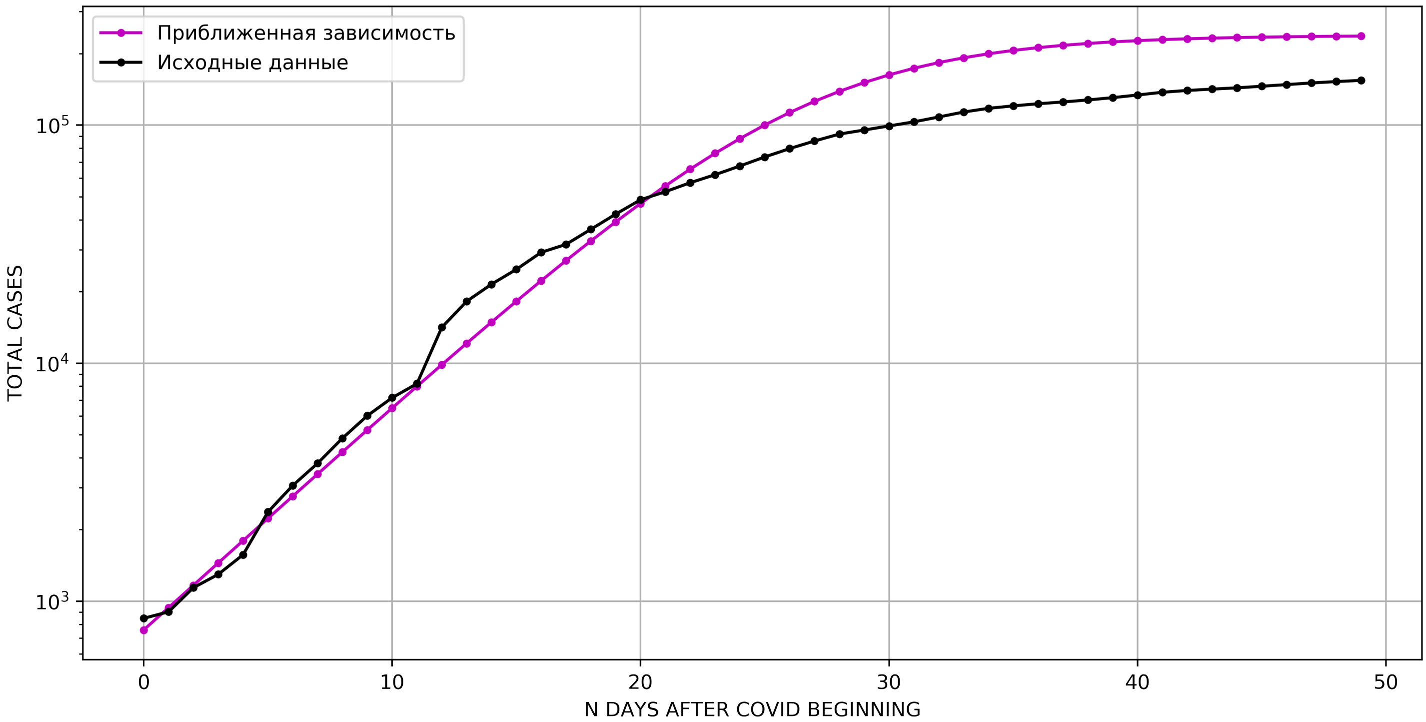

Рисунок 4. Оптимизированный график SIS-модели

5. Анализ данных

а) Погрешности аппроксимаций в пространстве L_∞. Воспользуемся формулой нормы:

$$\| f(t) \|_\infty = \max_t | f(t) |, \qquad (33)$$

Где $f(t) = I(t) - y(t)$ - разность предсказанного числа инфицированных и реально заболевших за день t.

Получили значение погрешности равное 93483.

Вычислим погрешность в пространстве L_2, для которого формула нормы имеет вид:

$$\| f(t) \|_2 = \left(\frac{1}{n} \sum_t f(t)\right)^{\frac{1}{2}} \qquad (34)$$

Где $f(t) = I(t) - y(t)$ - разность предсказанного числа инфицированных и реально заболевших за день t.

Получили значение погрешности равное 56065.5

б) Для определения максимального числа дней, в течении которых использование модели экспоненциального роста оправдано, рассматриваем отношение числа зараженных I_0 текущего дня к числу зараженных в предыдущий день. Таким образом получим получим зависимость разностей отношения числа ифицированных $e^{\chi_i} - e^{\chi_{i-1}}$

Проанализировав график рисунка 5, пришли к выводу, что для рассмотрения модели экспоненциального роста максимальное число дней, для которых использование этой модели оправдано, берем равное 25, т.к. по истечении которых видим большой скачек.

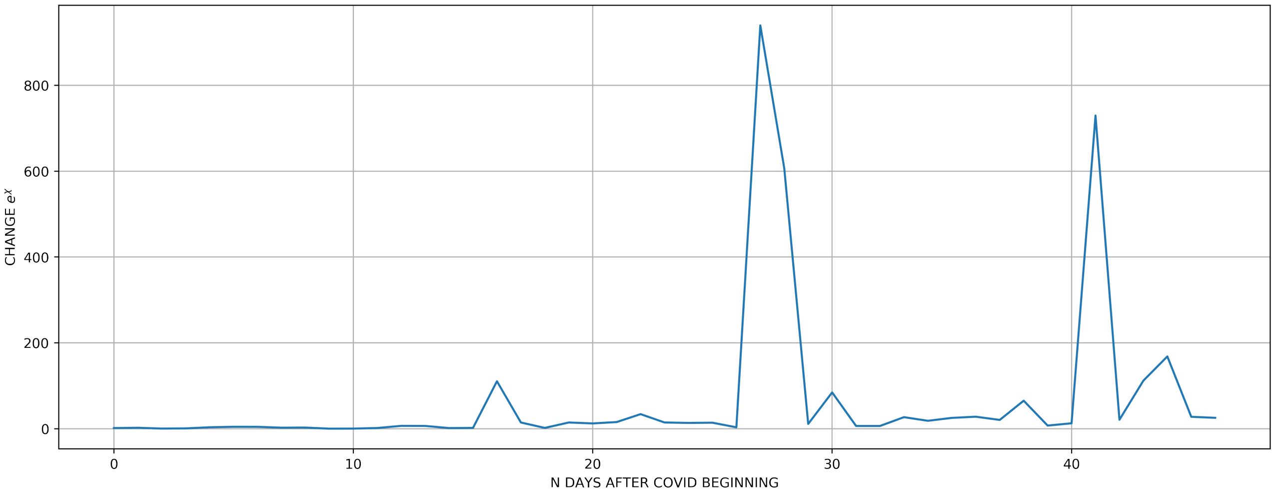

Рисунок 5. График зависимости прироста зараженных от дня

Но следует иметь в виду, что на распространение вируса влияет достаточно много факторов, начиная от введенных изоляционных ограничений и заканчивая погодными условиями (было предположение, что COVID-19 при более высокой температуре распространяется менее активно). С учетом выше написанного, можно придти к выводу, что любые предсказания касательно распространения вируса очень грубы.

в) Предполагаемое число инфицированных в Германии   при $t \to \infty$

находим при решении SIS-модели, откуда:

$$I_\infty = 238515$$

Количество дней с начала пандемии, по истечении которых наступит ее окончание, получим из уравнения SIS-модели, где при расчете количества дней будем считать, что затухание пандемии наступает с незначительным приростом числа заболевших (будем использовать прирост менее 1% от числа всего населения).

Получим:

$$t > -\frac{\ln\left(\frac{0.01}{(\frac{I_\infty}{I_0}-1)(1-1.01e^{-\chi})}\right)}{\chi} \tag{35}$$

После произведения расчета получим, что окончание пандемии в Германии наступит через 40 дней.