УДК 631.524.02: 519.237.5:519.654

СТАТИСТИЧЕСКАЯ ОЦЕНКА ОШИБОК ПРИ АППРОКСИМАЦИИ ИНЖЕНЕРНЫХ ДАННЫХ НЕЛИНЕЙНЫМИ ФУНКЦИЯМИ

Анатолий Анатольевич Аникьев[1]
доктор физико-математических наук, профессор
aaanikyev@mail.ru
Эмилия Николаевна Аникьева[2]
старший преподаватель
korol_0909@mail.ru
[1]Московский государственный технический университет
им. Н.Э. Баумана
г. Москва, Россия
[2]Мичуринский государственный аграрный университет
г. Мичуринск, Россия

Аннотация. Рассмотрена процедура приближения эмпирических данных модельными законами в общем виде полиномов произвольной степени. Оценка тесноты связи между двумя наборами данных определяется величиной коэффициента корреляции и коэффициентом детерминации. Вне зависимости от порядка моделирующего полинома можно оценивать ошибку аппроксимации или по величине остатков среднеквадратичного приближения или по значениям коэффициентов корреляции или детерминации.

Ключевые слова: теснота связи, приближение, модельный закон, полином.

Задачи аппроксимации обычно возникают, когда имеется набор дискретных данных в виде таблицы как результат проведенного эксперимента, и необходимо описать полученные экспериментально зависимости некоторой аналитической функцией. В качестве аналитической функции, удовлетворяющей условию непрерывности вместе со своими производными в интервале изменения предиктора (независимой переменной), обычно выбирается полином *n*-го порядка. После выбора вида аналитической функции решается задача подбора коэффициентов полинома таким образом, чтобы полученная в результате кривая наименее уклонялась от экспериментальных точек и в то же время хорошо описывала эмпирическую зависимость. Задачу поиска такой функции можно решать методом наименьших квадратов (МНК). Название метода происходит от способа нахождения коэффициентов выбранной функции. Задается функция коэффициентов $S(a_0, a_1, a_2, ..., a_n)$, например, полинома, равная сумме квадратов отклонений выбранной функции $f(x_i)$ от экспериментальных точек [1, 2]:

$$S(a_0, a_1, a_2, ..., a_n) = \sum_{i=1}^{m} (f(x_i, a_0, a_1, a_2, ..., a_n) - y_i)^2, \quad (1)$$

где n – порядок полинома, m – количество экспериментальных точек.

Затем решается задача поиска минимума этой функции, имея в качестве параметров ее коэффициенты. Из условий экстремума функции многих переменных получается система линейных алгебраических уравнений относительно коэффициентов полинома. Если система имеет единственное решение, то коэффициенты полинома найдены, и определена аналитическая зависимость, описывающая экспериментальную зависимость с минимальным уклонением от неё. Как правило, в задачах аппроксимации $m > n$, т. е. система переопределена. Прежде всего, решение таких задач начинают с рассмотрения аппроксимации эмпирического ряда данных линейным законом, исходя из того факта, что зрение человека очень хорошо приспособлено фиксировать отклонение точек зависимости от прямой линии. Если нам задана таблица данных только двух переменных $y(x)$, и нам надо проверить, подчиняется ли исходная зависимость линейному закону, то уравнение линейной зависимости $y = kx + b$ имеет только два неизвестных коэффициента k и b. Соответственно, если таблица данных содержит 10 значений y_i и x_i, мы будем иметь 10 уравнений для определения двух неизвестных k и b. Такой подход к решению возможен, и пакет MathCad предоставляет для этого разнообразные функции. Рассмотрим возможные методы решения этой задачи на примере [2].

Пример 1. Заданы два вектора данных, предиктор x и зависимая переменная y (x). Необходимо решить задачу аппроксимации эмпирических данных линейным законом.

Способ 1. Решаем задачу поиска минимума функции (формула (8)) двух переменных k и b в предполагаемом линейном законе с помощью встроенной функции *Minerr* (k, b) в блоке {*Given* – *Minerr*}. Эмпирический ряд данных задан в таблицах предиктора x и зависимой переменной y, имеющих всего одиннадцать значений. Процедура решений показана на рис. 1, где приведен рабочий лист Mathcad.

В данном случае мы применили функцию **Minerr (x_1, x_2, …)**, которая находит решение в блоке {*Given* – *Minerr*}, используя итерационную процедуру для поиска минимума функции S (k, p) по закону наименьших квадратов. Предварительно необходимо задать начальные значения неизвестным.

Преимущество этой функции в том, что явно получается сумма отклонений и можно вывести их в служебном слове {*ERR*}. Чтобы выяснить, насколько хорошо линейный закон приближает эмпирическую зависимость, вычисляем коэффициент корреляции. В пакете MathCad эта функция имеет синтаксис *corr* (A, B**)** и возвращает коэффициент корреляции Пирсона для двух рядов данных A и B. Значение коэффициента корреляции 0.945 указывает на хорошее соответствие между двумя наборами данных - «эмпирических» значений y_i и набором значений найденной функции в точках x_i.

$x^T =$

	0	1	2	3	4	5	6	7	8	9
0	3	3.1	3.2	3.3	3.4	3.5	3.6	3.7	3.8	...

$y^T =$

	0	1	2	3	4	5	6	7	8
0	4.615	4.591	5.13	5.481	5.492	5.553	5.471	5.727	...

$$\begin{pmatrix} k \\ p \end{pmatrix} := \begin{pmatrix} 1 \\ 1 \end{pmatrix}$$

Given

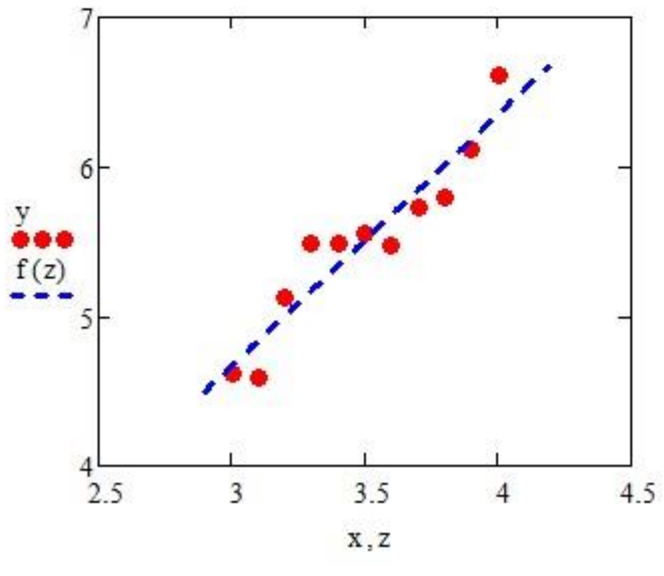
$$\sum_{i=0}^{10} \left(y_i - k \cdot x_i - p\right)^2 = 0$$

$$\begin{pmatrix} a \\ b \end{pmatrix} := \mathrm{Minerr}(k, p) \qquad \begin{pmatrix} a \\ b \end{pmatrix} = \begin{pmatrix} 1.682 \\ -0.38 \end{pmatrix}$$

$$\mathrm{ERR} = 0.372 \qquad \mathrm{line}(x, y) = \begin{pmatrix} -0.38 \\ 1.682 \end{pmatrix}$$

$$f(z) := a \cdot z + b \quad z := 2.9, 2.91 .. 4.2$$

$$i := 0 .. 10$$



$$f_i := a \cdot x_i + b$$

$$\mathrm{corr}(f, y) = 0.945$$

Рисунок 1 - Вычисление коэффициентов линейного закона по таблице эмпирических данных методом МНК: красными кружками показаны экспериментальные значения, синей пунктирной линией — найденная аналитически линейная зависимость.

Если полученная экспериментальная зависимость отличается от линейной, можно находить коэффициенты нелинейной зависимости этими же способами, поскольку в методе МНК для нахождения коэффициентов решается СЛАУ относительно коэффициентов. В следующем примере задан модельный закон дисперсии в виде вектор-столбцов по переменным $x1$ и $y1$. На рабочем листе MathCad приводится решение задачи аппроксимации с использованием функций *Minerr ( )* и *Minimize* ( ).

Пример 2. Значение предиктора задаем в интервале от 2 до 3 с шагом 0.1. Функцию моделируем квадратичным законом и добавляем случайный разброс точек с помощью псевдогенератора случайных чисел rnd (n). Обе функции одинаково хорошо справились с задачей аппроксимации, как видно из графика на рис. 2, где ромбами изображена экспериментальная функция, а синим пунктиром – полученная аналитическая функция в виде полинома второго порядка. Применение функции *Minerr (a, b, c)* аналогично рассмотренному в примере 1. Определяем начальные значения для коэффициентов полинома второго порядка. Далее в блоке {*Given – Minerr*} происходит поиск таких коэффициентов полинома второго порядка, которые обеспечивают минимум квадрата разности между рядом экспериментальных точек и искомой аналитической функцией. Затем выводятся значения этих коэффициентов и значение суммы погрешности метода. На графике строим найденную аналитическую функцию и ряд экспериментальных точек [3].

Однако, если мы выключим разброс точек, функция *Minerr* ( ) хуже справляется с задачей, и надо проводить несколько итераций, заменяя начальные значения полученными, чтобы найти приемлемое решение. В то же время функция *Minimize* ( ) отлично справилась с задачей, получив точные значения коэффициентов (рис. 2).

$x1_i := 2 + 0.1 \cdot i \qquad y1_i := 2 + 3 \cdot x1_i - (x1_i)^2 + \mathrm{rnd}(0.5)$

$x1^T =$

	0	1	2	3	4	5	6	7	8	9
0	2	2.1	2.2	2.3	2.4	2.5	2.6	2.7	2.8	...

$y1^T =$

	0	1	2	3	4	5	6	7	8
0	4.27	4.121	4.191	4	3.938	3.556	3.173	3.23	...

$$\begin{pmatrix} a1 \\ b1 \\ c1 \end{pmatrix} := \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}$$

Given

$$\sum_{i=0}^{10} \left[y1_i - a1 \cdot (x1_i)^2 - b1 \cdot x1_i - c1 \right]^2 = 0$$

$$\begin{pmatrix} m \\ n \\ p \end{pmatrix} := \mathrm{Minerr}(a1, b1, c1) \qquad \begin{pmatrix} m \\ n \\ p \end{pmatrix} = \begin{pmatrix} -1.7797676 \\ 6.7233416 \\ -2.065022 \end{pmatrix}$$

$ERR = 0.127 \qquad f1(z) := m \cdot z^2 + n \cdot z + p \qquad z := 1.9, 1.91 .. 3.1$

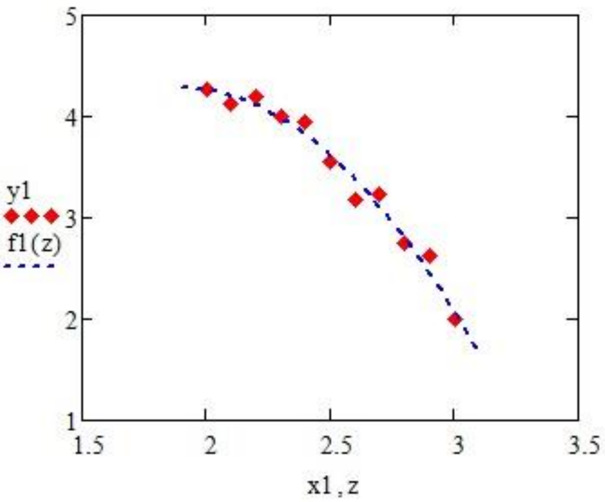


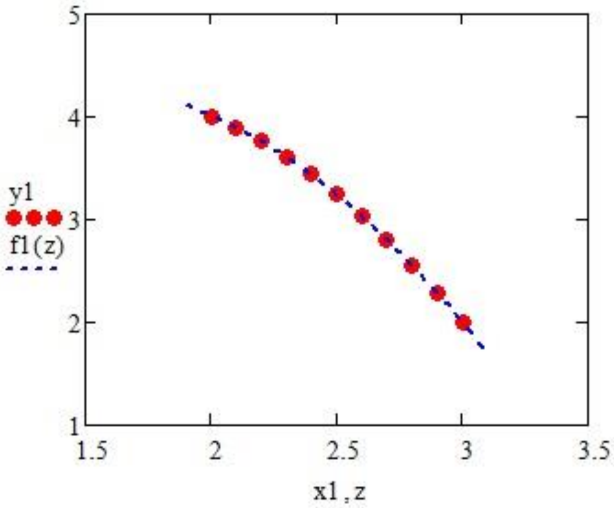
Рисунок 2 - Процедура поиска коэффициентов полинома, аппроксимирующего нелинейную зависимость: на графике ромбы — эмпирические данные, штриховая линия — приближение, полученное методом наименьших квадратов.

$$\begin{pmatrix} a1 \\ b1 \\ c1 \end{pmatrix} := \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} \qquad \text{Given}$$

$$\sum_{i=0}^{10} \left[y1_i - a1\cdot\left(x1_i\right)^2 - b1\cdot x1_i - c1 \right]^2 = 0$$

$$\begin{pmatrix} m \\ n \\ p \end{pmatrix} := \mathrm{Minerr}(a1, b1, c1) \qquad \begin{pmatrix} m \\ n \\ p \end{pmatrix} = \begin{pmatrix} -0.9336059 \\ 2.6656071 \\ 2.4133734 \end{pmatrix}$$

$$ERR = 3.958 \times 10^{-4} \qquad f1(z) := m\cdot z^2 + n\cdot z + p \qquad z := 1.9, 1.91 .. 3.1$$



$$f2(a1, b1, c1) := \sum_{i=0}^{10} \left[y1_i - a1\cdot\left(x1_i\right)^2 - b1\cdot x1_i - c1 \right]^2$$

$$\begin{pmatrix} m \\ n \\ p \end{pmatrix} := \mathrm{Minimize}(f2, m, n, p) \qquad \begin{pmatrix} m \\ n \\ p \end{pmatrix} = \begin{pmatrix} -1 \\ 3 \\ 2 \end{pmatrix}$$

Рисунок 3 - Вычисление коэффициентов полинома, аппроксимирующего нелинейную модельную зависимость без разброса точек (обозначения в формулах те же, что и на рис. 2).

Относительная погрешность аппроксимации эмпирических данных в методе МНК может быть вычислена по соотношению:

$$\delta\varepsilon = \sqrt{\frac{\sum_{i=1}^{m}(y_i - f(x_i))^2}{\sum_{i=1}^{m} y_i^2}} \cdot 100\% \,, \qquad (2)$$

а средний квадрат ошибки - по соотношению:

$$\Delta\varepsilon = \frac{1}{m} \cdot \sum_{i=1}^{m} \left[y_i - f(x_i) \right]^2 \qquad (3)$$

Качество модельного закона $f(x)$, применяемого при аппроксимации можно считать удовлетворительным, если величина (2) не превышает 10%. Кроме того, мы в каждом примере обязательно вычисляем коэффициент корреляции – величину, определяющую насколько коррелируют между собой два набора данных – эмпирическая зависимость и модельная. Степень достоверности корреляции определяется коэффициентом детерминации $D = r^2$ где r – коэффициент корреляции. При значении D > 0.7 предполагается, что теснота связи двух наборов данных достаточна, чтобы считать модельный закон дисперсии хорошо описывающим ход эмпирической зависимости. В случае любого нелинейного хода кривой, описывающей эмпирические данные можно привести исходные данные к линейному закону путем перехода к криволинейной системе координат. Затем строим коридор возможных отклонений модельного закона от экспериментального для зависимой переменной и окончательно доверительную область для всего модельного закона. Конкретный разбор построения нелинейного коридора отклонений для зависимой переменной мы приведем в следующей статье [4].

Список литературы:

1. Кобзарь А. И. Прикладная математическая статистика. М.: Физматлит. 2006.

2. Пугачев В.С. Теория вероятности и математическая статистика. М.: Физматлит. 2002.

3. Макова Н. Е. Решение уравнений средствами smath studio // Наука и Образование. 2024. Т. 7. № 1.

4. Лютый А. В., Марченко В. О., Подольская Е. Е. Метод аппроксимации данных результатов испытаний разбрасывателей минеральных удобрений // Наука и Образование. 2021. Т. 3. № 4 С. 5.

UDC 631.524.02: 519.237.5:519.654

STATISTICAL ERROR ASSESSMENT IN APPROXIMATION OF ENGINEERING DATA BY NONLINEAR FUNCTIONS

Anatoly An. Anikyev[1]
doctor of physical and mathematical sciences, professor
aaanikyev@mail.ru
Emiliya N. Anikyeva[2]
senior lecturer
korol_0909@mail.ru
[1]Bauman Moscow State Technical University
Moscow, Russia
[2]Michurinsk State Agrarian University
Michurinsk, Russia

Abstract. The procedure of approximation of empirical data by model laws in the general form of arbitrary degree polynomials is considered. The strength of the relationship between two sets of data is determined by the correlation coefficient magnitude and the coefficient of determination. Regardless of the order of the modeling polynomial, it is possible to estimate the approximation error either by the value of the residuals of the mean square approximation or by the values of the correlation or determination coefficients.

Keywords: closeness of connection, approximation, model law, polynomial.

Статья поступила в редакцию 11.11.2024; одобрена после рецензирования 20.12.2024; принята к публикации 25.12.2024.

The article was submitted 11.11.2024; approved after reviewing 20.12.2024; accepted for publication 25.12.2024.