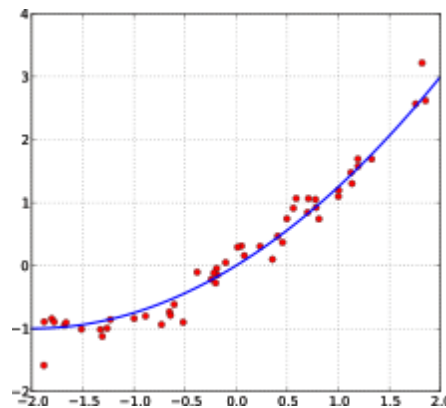


Википедия

Метод наименьших квадратов

Материал из Википедии — свободной энциклопедии

Метод наименьших квадратов (**МНК**) — математический метод, применяемый для решения различных задач, основанный на минимизации суммы квадратов отклонений некоторых функций от искомых переменных. Он может использоваться для «решения» переопределённых систем уравнений (когда количество уравнений превышает количество неизвестных), для поиска решения в случае обычных (не переопределённых) нелинейных систем уравнений, для аппроксимации точечных значений некоторой функции. МНК является одним из базовых методов регрессионного анализа для оценки неизвестных параметров регрессионных моделей по выборочным данным.



Пример кривой, проведённой через точки, имеющие нормально распределённое отклонение от истинного значения.

Содержание

История

Суть метода наименьших квадратов

Пример — система линейных уравнений

МНК в регрессионном анализе (аппроксимация данных)

МНК в случае линейной регрессии

Простейшие частные случаи

Случай полиномиальной модели

Статистические свойства МНК-оценок

Обобщенный МНК

Взвешенный МНК

См. также

Примечания

Литература

Ссылки

История

До начала XIX в. учёные не имели определённых правил для решения системы уравнений, в которой число неизвестных меньше, чем число уравнений; до этого времени употреблялись частные приёмы, зависевшие от вида уравнений и от остроумия вычислителей, и потому разные вычислители, исходя из тех же данных

наблюдений, приходили к различным выводам. Гауссу (1795) принадлежит первое применение метода, а Лежандр (1805) независимо открыл и опубликовал его под современным названием (фр. *Méthode des moindres quarrés*)^[1]. Лаплас связал метод с теорией вероятностей, а американский математик Эдвейн (1808) рассмотрел его теоретико-вероятностные приложения^[2]. Метод распространён и усовершенствован дальнейшими изысканиями Энке, Бесселя, Ганзена и других.

Работы А. А. Маркова в начале XX века позволили включить метод наименьших квадратов в теорию оценивания математической статистики, в которой он является важной и естественной частью. Усилиями Ю. Неймана, Ф.Дэвида, А. Эйткена, С. Рао было получено множество немаловажных результатов в этой области^[3].

Суть метода наименьших квадратов

Пусть \mathbf{x} — набор n неизвестных переменных (параметров), $f_i(\mathbf{x})$, $i = 1, \dots, m$, $m > n$ — совокупность функций от этого набора переменных. Задача заключается в подборе таких значений \mathbf{x} , чтобы значения этих функций были максимально близки к некоторым значениям y_i . По существу речь идет о «решении» переопределенной системы уравнений $f_i(\mathbf{x}) = y_i$, $i = 1, \dots, m$ в указанном смысле максимальной близости левой и правой частей системы. Суть МНК заключается в выборе в качестве «меры близости» суммы квадратов отклонений левых и правых частей $|f_i(\mathbf{x}) - y_i|$. Таким образом, сущность МНК может быть выражена следующим образом:

$$\sum_i e_i^2 = \sum_i (y_i - f_i(\mathbf{x}))^2 \rightarrow \min_{\mathbf{x}}.$$

В случае, если система уравнений имеет решение, то наименьшее значение суммы квадратов будет равно нулю и могут быть найдены точные решения системы уравнений аналитически или, например, различными численными методами оптимизации. Если система переопределена, то есть, говоря нестрого, количество независимых уравнений больше количества искомых переменных, то система не имеет точного решения и метод наименьших квадратов позволяет найти некоторый «оптимальный» вектор \mathbf{x} в смысле максимальной близости векторов \mathbf{y} и $\mathbf{f}(\mathbf{x})$ или максимальной близости вектора отклонений \mathbf{e} к нулю (близость понимается в смысле евклидова расстояния).

Пример — система линейных уравнений

В частности, метод наименьших квадратов может использоваться для «решения» системы линейных уравнений

$$\mathbf{Ax} = \mathbf{b},$$

где \mathbf{A} прямоугольная матрица размера $m \times n$, $m > n$ (то есть число строк матрицы \mathbf{A} больше количества искомых переменных).

Такая система уравнений в общем случае не имеет решения. Поэтому эту систему можно «решить» только в смысле выбора такого вектора \mathbf{x} , чтобы минимизировать «расстояние» между векторами \mathbf{Ax} и \mathbf{b} . Для этого можно применить критерий минимизации суммы квадратов разностей левой и правой частей уравнений системы, то есть $(\mathbf{Ax} - \mathbf{b})^T (\mathbf{Ax} - \mathbf{b}) \rightarrow \min_{\mathbf{x}}$. Нетрудно показать, что решение этой задачи минимизации приводит к решению следующей системы уравнений

$$A^T A x = A^T b \Rightarrow x = (A^T A)^{-1} A^T b.$$

Используя оператор псевдоинверсии, решение можно переписать так:

$$x = A^+ b,$$

где A^+ — псевдообратная матрица для A .

Данную задачу также можно «решить» используя так называемый *взвешенный МНК* (см. ниже), когда разные уравнения системы получают разный вес из теоретических соображений.

Строгое обоснование и установление границ содержательной применимости метода даны А. А. Марковым и А. Н. Колмогоровым.

МНК в регрессионном анализе (аппроксимация данных)

Пусть имеется n значений некоторой переменной y (это могут быть результаты наблюдений, экспериментов и т. д.) и соответствующих переменных x . Задача заключается в том, чтобы взаимосвязь между y и x аппроксимировать некоторой функцией $f(x, b)$, известной с точностью до некоторых неизвестных параметров b , то есть фактически найти наилучшие значения параметров b , максимально приближающие значения $f(x, b)$ к фактическим значениям y . Фактически это сводится к случаю «решения» переопределённой системы уравнений относительно b :

$$f(x_t, b) = y_t, t = 1, \dots, n.$$

В регрессионном анализе и в частности в эконометрике используются вероятностные модели зависимости между переменными

$$y_t = f(x_t, b) + \varepsilon_t,$$

где ε_t — так называемые *случайные ошибки* модели.

Соответственно, отклонения наблюдаемых значений y от модельных $f(x, b)$ предполагается уже в самой модели. Сущность МНК (обычного, классического) заключается в том, чтобы найти такие параметры b , при которых сумма квадратов отклонений (ошибок, для регрессионных моделей их часто называют остатками регрессии) e_t будет минимальной:

$$\hat{b}_{OLS} = \arg \min_b RSS(b),$$

где RSS — англ. *Residual Sum of Squares*^[4] определяется как:

$$RSS(b) = e^T e = \sum_{t=1}^n e_t^2 = \sum_{t=1}^n (y_t - f(x_t, b))^2.$$

В общем случае решение этой задачи может осуществляться численными методами оптимизации (минимизации). В этом случае говорят о *нелинейном МНК* (NLS или NLLS — англ. *Non-Linear Least Squares*). Во многих случаях можно получить аналитическое решение. Для решения задачи минимизации необходимо найти стационарные точки функции $RSS(b)$, продифференцировав её по неизвестным параметрам b ,

приравняв производные к нулю и решив полученную систему уравнений:

$$\sum_{t=1}^n (y_t - f(x_t, b)) \frac{\partial f(x_t, b)}{\partial b} = 0.$$

МНК в случае линейной регрессии

Пусть регрессионная зависимость является линейной:

$$y_t = \sum_{j=1}^k b_j x_{tj} + \varepsilon = \mathbf{x}_t^T \mathbf{b} + \varepsilon_t.$$

Пусть \mathbf{y} — вектор-столбец наблюдений объясняемой переменной, а \mathbf{X} — это $(n \times k)$ -матрица наблюдений факторов (строки матрицы — векторы значений факторов в данном наблюдении, по столбцам — вектор значений данного фактора во всех наблюдениях). Матричное представление линейной модели имеет вид:

$$\mathbf{y} = \mathbf{X}\mathbf{b} + \boldsymbol{\varepsilon}.$$

Тогда вектор оценок объясняемой переменной и вектор остатков регрессии будут равны

$$\hat{\mathbf{y}} = \mathbf{X}\mathbf{b}, \quad \mathbf{e} = \mathbf{y} - \hat{\mathbf{y}} = \mathbf{y} - \mathbf{X}\mathbf{b}.$$

соответственно сумма квадратов остатков регрессии будет равна

$$RSS = \mathbf{e}^T \mathbf{e} = (\mathbf{y} - \mathbf{X}\mathbf{b})^T (\mathbf{y} - \mathbf{X}\mathbf{b}).$$

Дифференцируя эту функцию по вектору параметров \mathbf{b} и приравняв производные к нулю, получим систему уравнений (в матричной форме):

$$(\mathbf{X}^T \mathbf{X})\mathbf{b} = \mathbf{X}^T \mathbf{y}.$$

В расшифрованной матричной форме эта система уравнений выглядит следующим образом:

$$\begin{pmatrix} \sum x_{t1}^2 & \sum x_{t1}x_{t2} & \sum x_{t1}x_{t3} & \dots & \sum x_{t1}x_{tk} \\ \sum x_{t2}x_{t1} & \sum x_{t2}^2 & \sum x_{t2}x_{t3} & \dots & \sum x_{t2}x_{tk} \\ \sum x_{t3}x_{t1} & \sum x_{t3}x_{t2} & \sum x_{t3}^2 & \dots & \sum x_{t3}x_{tk} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \sum x_{tk}x_{t1} & \sum x_{tk}x_{t2} & \sum x_{tk}x_{t3} & \dots & \sum x_{tk}^2 \end{pmatrix} \begin{pmatrix} b_1 \\ b_2 \\ b_3 \\ \vdots \\ b_k \end{pmatrix} = \begin{pmatrix} \sum x_{t1}y_t \\ \sum x_{t2}y_t \\ \sum x_{t3}y_t \\ \vdots \\ \sum x_{tk}y_t \end{pmatrix}, \text{ где все суммы берутся по всем}$$

допустимым значениям t .

Если в модель включена константа (как обычно), то $x_{t1} = 1$ при всех t , поэтому в левом верхнем углу матрицы системы уравнений находится количество наблюдений n , а в остальных элементах первой строки и первого столбца — просто суммы значений переменных: $\sum x_{tj}$ и первый элемент правой части системы — $\sum y_t$.

Решение этой системы уравнений и дает общую формулу МНК-оценок для линейной модели:

$$\hat{b}_{OLS} = (X^T X)^{-1} X^T y = \left(\frac{1}{n} X^T X \right)^{-1} \frac{1}{n} X^T y = V_x^{-1} C_{xy}.$$

Для аналитических целей оказывается полезным последнее представление этой формулы (в системе уравнений при делении на n , вместо сумм фигурируют средние арифметические). Если в регрессионной модели данные *центрированы*, то в этом представлении первая матрица имеет смысл выборочной ковариационной матрицы факторов, а вторая — вектор ковариаций факторов с зависимой переменной. Если кроме того данные ещё и *нормированы* на СКО (то есть в конечном итоге *стандартизированы*), то первая матрица имеет смысл выборочной корреляционной матрицы факторов, второй вектор — вектора выборочных корреляций факторов с зависимой переменной.

Немаловажное свойство МНК-оценок для моделей с константой — линия построенной регрессии проходит через центр тяжести выборочных данных, то есть выполняется равенство:

$$\bar{y} = \hat{b}_1 + \sum_{j=2}^k \hat{b}_j \bar{x}_j.$$

В частности, в крайнем случае, когда единственным регрессором является константа, получаем, что МНК-оценка единственного параметра (собственно константы) равна среднему значению объясняемой переменной. То есть среднее арифметическое, известное своими хорошими свойствами из законов больших чисел, также является МНК-оценкой — удовлетворяет критерию минимума суммы квадратов отклонений от неё.

Простейшие частные случаи

В случае парной линейной регрессии $y_t = a + bx_t + \varepsilon_t$, когда оценивается линейная зависимость одной переменной от другой, формулы расчета упрощаются (можно обойтись без матричной алгебры). Система уравнений имеет вид:

$$\begin{pmatrix} n & \sum_{t=1}^n x_t \\ \sum_{t=1}^n x_t & \sum_{t=1}^n x_t^2 \end{pmatrix} \begin{pmatrix} a \\ b \end{pmatrix} = \begin{pmatrix} \sum_{t=1}^n y_t \\ \sum_{t=1}^n x_t y_t \end{pmatrix}.$$

Отсюда несложно найти оценки коэффициентов:

$$\begin{cases} \hat{b} = \frac{n \sum_{t=1}^n x_t y_t - (\sum_{t=1}^n x_t)(\sum_{t=1}^n y_t)}{n \sum_{t=1}^n x_t^2 - (\sum_{t=1}^n x_t)^2}, \\ \hat{a} = \frac{\sum_{t=1}^n y_t - \hat{b} \sum_{t=1}^n x_t}{n}. \end{cases}$$

Несмотря на то что в общем случае модели с константой предпочтительней, в некоторых случаях из теоретических соображений известно, что константа a должна быть равна нулю. Например, в физике зависимость между напряжением и силой тока имеет вид $U = I \cdot R$; замеряя напряжение и силу тока, необходимо оценить сопротивление. В таком случае речь идёт о модели $y = bx$. В этом случае вместо системы уравнений имеем единственное уравнение

$$\left(\sum x_t^2 \right) b = \sum x_t y_t.$$

Следовательно, формула оценки единственного коэффициента имеет вид

$$\hat{b} = \frac{\sum_{t=1}^n x_t y_t}{\sum_{t=1}^n x_t^2}.$$

Случай полиномиальной модели

Если данные аппроксимируются полиномиальной функцией регрессии одной переменной $f(x) = b_0 + \sum_{i=1}^k b_i x^i$, то, воспринимая степени x^i как независимые факторы для каждого i можно оценить параметры модели исходя из общей формулы оценки параметров линейной модели. Для этого в общую формулу достаточно учесть, что при такой интерпретации $x_i x_{tj} = x_t^i x_t^j = x_t^{i+j}$ и $x_{tj} y_t = x_t^j y_t$. Следовательно, матричные уравнения в данном случае примут вид:

$$\begin{pmatrix} n & \sum_n x_t & \dots & \sum_n x_t^k \\ \sum_n x_t & \sum_n x_t^2 & \dots & \sum_n x_t^{k+1} \\ \vdots & \vdots & \ddots & \vdots \\ \sum_n x_t^k & \sum_n x_t^{k+1} & \dots & \sum_n x_t^{2k} \end{pmatrix} \begin{bmatrix} b_0 \\ b_1 \\ \vdots \\ b_k \end{bmatrix} = \begin{bmatrix} \sum_n y_t \\ \sum_n x_t y_t \\ \vdots \\ \sum_n x_t^k y_t \end{bmatrix}.$$

Статистические свойства МНК-оценок

В первую очередь, отметим, что для линейных моделей МНК-оценки являются линейными оценками, как это следует из вышеприведённой формулы. Для несмещённости МНК-оценок необходимо и достаточно выполнения важнейшего условия регрессионного анализа: условное по факторам математическое ожидание случайной ошибки должно быть равно нулю. Данное условие, в частности, выполнено, если

1. математическое ожидание случайных ошибок равно нулю, и
2. факторы и случайные ошибки — независимые случайные величины.

Первое условие можно считать выполненным всегда для моделей с константой, так как константа берёт на себя ненулевое математическое ожидание ошибок (поэтому модели с константой в общем случае предпочтительнее).

Второе условие — условие экзогенности факторов — принципиальное. Если это свойство не выполнено, то можно считать, что практически любые оценки будут крайне неудовлетворительными: они не будут даже состоятельными (то есть даже очень большой объём данных не позволяет получить качественные оценки в этом случае). В классическом случае делается более сильное предположение о детерминированности факторов, в отличие от случайной ошибки, что автоматически означает выполнение условия экзогенности. В общем случае для состоятельности оценок достаточно выполнения условия экзогенности вместе со сходимостью матрицы V_x к некоторой невырожденной матрице при увеличении объёма выборки до бесконечности.

Для того, чтобы кроме состоятельности и несмещённости, оценки (обычного) МНК были ещё и эффективными (наилучшими в классе линейных несмещённых оценок) необходимо выполнение дополнительных свойств случайной ошибки:

- Постоянная (одинаковая) дисперсия случайных ошибок во всех наблюдениях (отсутствие гетероскедастичности): $V(\epsilon_t) = \sigma^2 = \text{const.}$

- Отсутствие корреляции (автокорреляции) случайных ошибок в разных наблюдениях между собой $\text{cov}(\varepsilon_i, \varepsilon_j) = 0 \quad \forall 1 \leq i < j \leq n$.

Данные предположения можно сформулировать для ковариационной матрицы вектора случайных ошибок $V(\varepsilon) = \sigma^2 I$.

Линейная модель, удовлетворяющая таким условиям, называется *классической*. МНК-оценки для классической линейной регрессии являются несмещёнными, состоятельными и наиболее эффективными оценками в классе всех линейных несмещённых оценок (в англоязычной литературе иногда употребляют аббревиатуру *BLUE* (*Best Linear Unbiased Estimator*) — наилучшая линейная несмещённая оценка; в отечественной литературе чаще приводится теорема Гаусса — Маркова). Как нетрудно показать, ковариационная матрица вектора оценок коэффициентов будет равна:

$$V(\hat{b}_{OLS}) = \sigma^2 (X^T X)^{-1}.$$

Эффективность означает, что эта ковариационная матрица является «минимальной» (любая линейная комбинация коэффициентов, и в частности сами коэффициенты, имеют минимальную дисперсию), то есть в классе линейных несмещённых оценок оценки МНК-наилучшие. Диагональные элементы этой матрицы — дисперсии оценок коэффициентов — важные параметры качества полученных оценок. Однако рассчитать ковариационную матрицу невозможно, поскольку дисперсия случайных ошибок неизвестна. Можно доказать, что несмещённой и состоятельной (для классической линейной модели) оценкой дисперсии случайных ошибок является величина:

$$s^2 = RSS/(n - k).$$

Подставив данное значение в формулу для ковариационной матрицы и получим оценку ковариационной матрицы. Полученные оценки также являются несмещёнными и состоятельными. Важно также то, что оценка дисперсии ошибок (а значит и дисперсий коэффициентов) и оценки параметров модели являются независимыми случайными величинами, что позволяет получить тестовые статистики для проверки гипотез о коэффициентах модели.

Необходимо отметить, что если классические предположения не выполнены, МНК-оценки параметров не являются наиболее эффективными оценками (оставаясь несмещёнными и состоятельными). Однако, ещё более ухудшается оценка ковариационной матрицы — она становится смещённой и несостоятельной. Это означает, что статистические выводы о качестве построенной модели в таком случае могут быть крайне недостоверными. Одним из вариантов решения последней проблемы является применение специальных оценок ковариационной матрицы, которые являются состоятельными при нарушениях классических предположений (стандартные ошибки в форме Уайта и стандартные ошибки в форме Ньюи-Уеста). Другой подход заключается в применении так называемого обобщённого МНК.

Обобщенный МНК

Метод наименьших квадратов допускает широкое обобщение. Вместо минимизации суммы квадратов остатков можно минимизировать некоторую положительно определённую квадратичную форму от вектора остатков $e^T W e$, где W — некоторая симметрическая положительно определённая весовая матрица. Обычный МНК является частным случаем данного подхода, когда весовая матрица пропорциональна единичной матрице. Как известно, для симметрических матриц (или операторов) существует разложение $W = P^T P$. Следовательно, указанный функционал можно представить следующим образом $e^T P^T P e = (Pe)^T Pe = e_*^T e_*$, то есть этот

функционал можно представить как сумму квадратов некоторых преобразованных «остатков». Таким образом, можно выделить класс методов наименьших квадратов — LS-методы (Least Squares).

Доказано (теорема Айткена), что для обобщенной линейной регрессионной модели (в которой на ковариационную матрицу случайных ошибок не налагается никаких ограничений) наиболее эффективными (в классе линейных несмещенных оценок) являются оценки т. н. **обобщенного МНК** (**ОМНК**, **GLS** — **Generalized Least Squares**) — LS-метода с весовой матрицей, равной обратной ковариационной матрице случайных ошибок: $W = V_\epsilon^{-1}$.

Можно показать, что формула ОМНК-оценок параметров линейной модели имеет вид

$$\hat{b}_{GLS} = (X^T V^{-1} X)^{-1} X^T V^{-1} y.$$

Ковариационная матрица этих оценок соответственно будет равна

$$V(\hat{b}_{GLS}) = (X^T V^{-1} X)^{-1}.$$

Фактически сущность ОМНК заключается в определённом (линейном) преобразовании (P) исходных данных и применении обычного МНК к преобразованным данным. Цель этого преобразования — для преобразованных данных случайные ошибки уже удовлетворяют классическим предположениям.

Взвешенный МНК

В случае диагональной весовой матрицы (а значит и ковариационной матрицы случайных ошибок) имеем так называемый взвешенный МНК (WLS — Weighted Least Squares). В данном случае минимизируется взвешенная сумма квадратов остатков модели, то есть каждое наблюдение получает «вес», обратно пропорциональный дисперсии случайной ошибки в данном наблюдении: $e^T W e = \sum_{t=1}^n \frac{e_t^2}{\sigma_t^2}$. Фактически данные преобразуются взвешиванием наблюдений (делением на величину, пропорциональную предполагаемому стандартному отклонению случайных ошибок), а к взвешенным данным применяется обычный МНК.

См. также

- Обобщенный метод наименьших квадратов
- Двухшаговый метод наименьших квадратов
- Рекурсивный МНК

Примечания

- Legendre, On Least Squares. Translated from the French by Professor Henry A. Ruger and Professor Helen M. Walker, Teachers College, Columbia University, New York City. (<http://www.stat.ucla.edu/history/legendre.pdf>) Архивная копия (<http://web.archive.org/web/20110107112250/http://www.stat.ucla.edu/history/legendre.pdf>) от 7 января 2011 на Wayback Machine (англ.)
- Александрова, 2008, с. 102.
- Линник, 1962, с. 21.
- Магнус, Катышев, Пересецкий, 2007, Обозначение RSS не унифицировано. RSS может быть сокращением от regression sum of squares, а ESS — error sum of squares, то есть, RSS и ESS будут иметь обратный смысл. с. 52. Издания 2004 года..

Литература

- *Линник Ю. В.* Метод наименьших квадратов и основы математико-статистической теории обработки наблюдений. — 2-е изд. — М., 1962. (математическая теория)
- *Айвазян С. А.* Прикладная статистика. Основы эконометрики. Том 2. — М.: Юнити-Дана, 2001. — 432 с. — ISBN 5-238-00305-6.
- *Доугерти К.* Введение в эконометрику: Пер. с англ. — М.: ИНФРА-М, 1999. — 402 с. — ISBN 8-86225-458-7.
- *Кремер Н. Ш., Путко Б. А.* Эконометрика. — М.: Юнити-Дана, 2003—2004. — 311 с. — ISBN 8-86225-458-7.
- *Магнус Я. Р., Катышев П. К., Пересецкий А. А.* Эконометрика. Начальный курс. — М.: Дело, 2007. — 504 с. — ISBN 978-5-7749-0473-0.
- Эконометрика. Учебник / Под ред. Елисеевой И. И. — 2-е изд. — М.: Финансы и статистика, 2006. — 576 с. — ISBN 5-279-02786-3.
- *Александрова Н. В.* История математических терминов, понятий, обозначений: словарь-справочник. — 3-е изд.. — М.: ЛКИ, 2008. — 248 с. — ISBN 978-5-382-00839-4.
- *Витковский В. В.* Наименьшие квадраты // Энциклопедический словарь Брокгауза и Ефрона : в 86 т. (82 т. и 4 доп.). — СПб., 1890—1907.
- *Митин И. В., Русаков В. С.* Анализ и обработка экспериментальных данных- 5-е издание- 24с.

Ссылки

- Метод наименьших квадратов онлайн для зависимости $y = a + bx$ (<http://www.chem-astu.ru/science/lsq/>) с вычислением погрешностей коэффициентов и оцениванием автокорреляции.
- МНК для линейных и нелинейных зависимостей (<http://calculator-online.org/s/teoriya-veroyatnosti/method-naimenshih-kvadratov/>) (недоступная ссылка) (любой набор неизвестных коэффициентов) — на сайте «Контрольная работа РУ»

Источник — https://ru.wikipedia.org/w/index.php?title=Метод_наименьших_квадратов&oldid=103687918

Эта страница в последний раз была отредактирована 3 декабря 2019 в 19:14.

Текст доступен по лицензии Creative Commons Attribution-ShareAlike; в отдельных случаях могут действовать дополнительные условия.

Wikipedia® — зарегистрированный товарный знак некоммерческой организации Wikimedia Foundation, Inc.