УДК 519.87

Об оценке параметров трехмерной равномерно-регрессионной модели

И.В. Пономарев

Алтайский государственный университет (Барнаул, Россия)

On Estimation of Parameters of a Three-Dimensional Uniform Regression Model

I.V. Ponomarev

Altai State University (Barnaul, Russia)

В процессе моделирования зависимости между различными показателями анализа исследователь сталкивается с вычислительными трудностями при оценке параметров модели. Обычно получение пригодной для изучения модели получают последовательным уточнением входящих в ее состав признаков, а следовательно, многократным повторением вычислительного алгоритма. При этом вычислительная сложность этих алгоритмов начинает играть значительную роль в моделировании. Для снижения числа итераций применяется определенный набор показателей, отвечающих за качество построенной модели и способных «просигнализировать» о необходимости корректировки модели. В регрессионном моделировании такими показателями являются величина функционала качества и параметры модели. Они способны дать ответ на вопрос о целесообразности построения той или иной модели и являются индикаторами качества полученной функциональной зависимости.

В данной работе изучаются методы и алгоритмы построения и оценки основных показателей L_∞-регрессии — показатель качества и параметры модели. В первой части исследования описываются наиболее эффективные вычислительные процедуры определения параметров в случае трехмерной равномерно-регрессионной модели, указывается сложность этих алгоритмов и дается геометрическая интерпретация. Во второй части приводится ряд теорем об оценках значений параметров трехмерной L_∞-регрессии, приводится формула для получения показателя однородности выборки.

Ключевые слова: равномерно-регрессионная модель, функционал качества, выпуклая оболочка, вычислительная сложность.

Typically, a researcher faces computational difficulties in assessing the parameters of a certain model when modeling the relationship between different indicators of analysis. A suitable model is generally obtained by a sequential refinement of the features included in its composition and, therefore, by performing multiple repetitions of computational algorithms. At the same time, the computational complexity of these algorithms begins to play a significant role in modeling. A certain set of indicators is used to reduce the number of iterations. These indicators are responsible for the quality of the constructed model and capable of "signaling" about the need to adjust the model. The model parameters and the quality functional value are such indicators in regression modeling. They are able to answer the question of the appropriateness of building a particular model and are indicators of the quality of the resulting functional dependence.

In this paper, we study methods and algorithms for constructing and evaluating the main indicators of L_∞-regression — a quality indicator and model parameters. The first part of the paper describes the most efficient computational procedures for determining parameters in the case of a three-dimensional uniform regression model, indicates the complexity of these algorithms, and gives a geometric interpretation. In the second part of the paper, a series of theorems on estimating the values of the parameters of three-dimensional L_∞-regression is presented, and a formula for calculation of an indicator of sample uniformity is provided.

Key words: uniform regression model, quality functional, convex hull, computational complexity.

DOI 10.14258/izvasu(2020)4-17

Введение. Рассмотрим конечное подмножество Ω трехмерного арифметического евклидова пространства R^3:

$$\Omega = \{(x_i, y_i, z_i) : i = 1, 2, \ldots, N\},$$

которое можно представить как результат N экспериментов.

Поставим задачу исследовать зависимость между координатами точек Ω на основе Чебышевской нормы равномерного отклонения. Для этого, следуя работе [1], определим алгоритм построения линейной зависимости (L_∞ регрессии). Очевидно, что в данном случае определены три задачи, соответствующие L_∞-регрессиям z на (x, y); x на (y, z); y на (x, z):

$$\begin{aligned} z_i &= k_{zx}x_i + k_{zy}y_i + b_z, \\ x_i &= k_{xy}y_i + k_{xz}z_i + b_x, \\ y_i &= k_{yx}x_i + k_{yz}z_i + b_y. \end{aligned}$$

Для каждой такой регрессии определен соответствующий функционал качества:

$$\begin{aligned} \alpha_\infty^z &= \min_{k_{zx},k_{zy},b_z} \max_{i=1,\ldots,n} \left| k_{zx}x_i + k_{zy}y_i + b_z - z_i \right|, \\ \alpha_\infty^x &= \min_{k_{xy},k_{xz},b_x} \max_{i=1,\ldots,n} \left| k_{xy}y_i + k_{xz}z_i + b_x - x_i \right|, \\ \alpha_\infty^y &= \min_{k_{yx},k_{yz},b_y} \max_{i=1,\ldots,n} \left| k_{yx}x_i + k_{yz}z_i + b_y - y_i \right|. \end{aligned}$$

1. Алгоритм построения модели. Рассмотрим для определенности задачу построения L_∞-регрессии z на (x, y). Для нахождения ее параметров возможно использовать ряд методов.

I) Задачу нахождения α_∞^z можно свести к следующей задаче линейного программирования:

$$\begin{cases} \min\limits_{u,v,k_{zx},k_{zy}} (u - v), \\ u \geq k_{zx}x_i + k_{zy}y_i - z_i, \quad i = 1, \ldots, N, \\ v \leq k_{zx}x_j + k_{zy}y_j - z_j, \quad j = 1, \ldots, N. \end{cases} \tag{1}$$

где u — верхняя огибающая, v — нижняя огибающая. Решение задачи (1) осуществляется симплекс-методом и имеет сложность $O(N^2)$ [2]. Данный метод является универсальным и может применяться к решению задачи о равномерно-регрессионной модели любой размерности.

Заметим, что геометрически точки зрения величина α_∞^z равна минимуму ширины слоя, ограниченного двумя параллельными плоскостями, содержащими Ω, ширина берется вдоль оси z. Таким образом, эта величина равна минимальной ширине выпуклой оболочки Ω в направлении базисного вектора k [3, 4].

II) В вычислительной геометрии разработаны различные методы построения выпуклой оболочки множества. Однако не все из них могут быть обобщены на многомерный случай. Наиболее предпочтительным для нашей задачи оказался алгоритм Quickhull, позволяющий проводить вычисления практически при любой размерности множества Ω и имеющий небольшую чувствительность к погрешностям вычислений. Сложность данного алгоритма $O(N \log N)$ [5]. Данный алгоритм реализован в математических пакетах GNU Octave и MatLab.

III) Для сокращения размерности выборки воспользуемся частными регрессионными моделями, т.е. для L_∞-регрессии z на (x, y) определены две частные равномерно-регрессионные модели

$$\begin{aligned} z_i &= k_x x_i + b_x, \\ z_i &= k_y y_i + b_y, \end{aligned}$$

где $i = 1, \ldots, N$ и точки $(x_i, y_i, z_i) \in \Omega$. Каждая из этих частных моделей определяет направляющий вектор для соответствующей L_∞-регрессии. В совокупности эти два вектора задают плоскость определяющую общую равномерно-нечеткую регрессию $z_i = k_{zx}x_i + k_{zy}y_i + b_z$. Для построения частных моделей можно воспользоваться различными методами построения выпуклой оболочки двумерного множества [6].

2. Оценка параметров трехмерной L_∞ регрессии. Построение равномерно-регрессионной модели в трехмерном случае сопряжено с необходимостью построения выпуклого множества, что бывает затруднительно. Следуя работе [7], докажем ряд теорем об ограничениях на параметры модели.

Теорема 1. Справедливы неравенства

$$\left|k_{xz}^0\right| \leq \frac{\alpha_\infty^x}{\alpha_\infty^z}, \quad \left|k_{zx}^0\right| \leq \frac{\alpha_\infty^z}{\alpha_\infty^x}, \quad \left|k_{xy}^0\right| \leq \frac{\alpha_\infty^x}{\alpha_\infty^y},$$
$$\left|k_{yz}^0\right| \leq \frac{\alpha_\infty^y}{\alpha_\infty^z}, \quad \left|k_{yx}^0\right| \leq \frac{\alpha_\infty^y}{\alpha_\infty^x}, \quad \left|k_{zy}^0\right| \leq \frac{\alpha_\infty^z}{\alpha_\infty^y},$$

где k_{zx}^0, k_{zy}^0, k_{xz}^0, k_{xy}^0, k_{yx}^0, k_{yz}^0 — угловые коэффициенты соответствующих плоскостей регрессии.

Доказательство. Пусть слой, содержащий множество Ω, минимальный в направлении оси OZ — задан системой уравнений:

$$\begin{cases} z = k_{zx}^0 x + k_{zy}^0 y + b_z^0 + \alpha_\infty^z, \\ z = k_{zx}^0 x + k_{zy}^0 y + b_z^0 \end{cases} \tag{2}$$

минимальный в направлении оси OX — задан системой уравнений:

$$\begin{cases} x = k_{xy}^0 y + k_{xz}^0 z + b_x^0 + \alpha_\infty^x, \\ x = k_{xy}^0 y + k_{xz}^0 z + b_x^0 \end{cases} \tag{3}$$

минимальный в направлении оси OY — задан системой уравнений:

$$\begin{cases} y = k_{yx}^0 x + k_{yz}^0 z + b_y^0 + \alpha_\infty^y, \\ y = k_{yx}^0 x + k_{yz}^0 z + b_y^0 \end{cases} . \tag{4}$$

Отрезки, высекаемые на осях OZ, OX, OY данными слоями, будут соответственно:

$$\begin{pmatrix} \alpha_\infty^z & \frac{\alpha_\infty^z}{\left|k_{zx}^0\right|} & \frac{\alpha_\infty^z}{\left|k_{zy}^0\right|} \\ \frac{\alpha_\infty^x}{\left|k_{xz}^0\right|} & \alpha_\infty^x & \frac{\alpha_\infty^x}{\left|k_{xy}^0\right|} \\ \frac{\alpha_\infty^y}{\left|k_{yz}^0\right|} & \frac{\alpha_\infty^y}{\left|k_{yx}^0\right|} & \alpha_\infty^y \end{pmatrix}$$

В силу минимальности слоя (2) в направлении оси OZ имеем:

$$\alpha_\infty^z \leq \frac{\alpha_\infty^x}{|k_{xz}^0|},$$
$$\alpha_\infty^z \leq \frac{\alpha_\infty^y}{|k_{yz}^0|}.$$

В силу минимальности слоя (3) в направлении оси OX имеем:

$$\alpha_\infty^x \leq \frac{\alpha_\infty^z}{|k_{zx}^0|},$$
$$\alpha_\infty^x \leq \frac{\alpha_\infty^y}{|k_{yx}^0|}.$$

В силу минимальности слоя (4) в направлении оси OY имеем:

$$\alpha_\infty^y \leq \frac{\alpha_\infty^z}{|k_{zy}^0|},$$
$$\alpha_\infty^y \leq \frac{\alpha_\infty^x}{|k_{xy}^0|}.$$

Из полученных неравенств следует утверждение теоремы.

Полученные неравенства показывают взаимосвязь между коэффициентами регрессии и значениями функционала качества. Они могут быть использованы при проведении анализа однородности выборки [8, 9].

Теорема 2. Справедливо равенство

$$V = \frac{\alpha_\infty^x \alpha_\infty^y \alpha_\infty^z}{\gamma}, \qquad (5)$$

где V — объем параллелепипеда, полученного пересечением минимальных слоев; $\gamma = 1 - k_{\mathrm{xy}}k_{\mathrm{yx}} - k_{\mathrm{xy}}k_{\mathrm{yz}}k_{\mathrm{zx}} - k_{\mathrm{xz}}k_{\mathrm{yx}}k_{\mathrm{zy}} - k_{\mathrm{xz}}k_{\mathrm{zx}} - k_{\mathrm{yz}}k_{\mathrm{zy}}$.

Доказательство. Ребро искомого параллелепипеда, параллельное плоскостям (3) и (4)

$$m_{12} = \frac{\alpha_\infty^z}{\gamma}\Big(k_{\mathrm{xz}} + k_{\mathrm{xy}}k_{\mathrm{yz}}, k_{\mathrm{xz}}k_{\mathrm{yx}} + k_{\mathrm{yz}}, 1 - k_{\mathrm{xy}}k_{\mathrm{yx}}\Big),$$

ребро параллелепипеда, параллельное плоскостям (2) и (4)

$$m_{13} = \frac{\alpha_\infty^x}{\gamma}\Big(1 - k_{\mathrm{yz}}k_{\mathrm{zy}}, k_{\mathrm{yx}} + k_{\mathrm{yz}}k_{\mathrm{zx}}, k_{\mathrm{zx}} + k_{\mathrm{yx}}k_{\mathrm{zy}}\Big),$$

ребро параллелепипеда, параллельное плоскостям (2) и (3)

$$m_{14} = \frac{\alpha_\infty^y}{\gamma}\Big(k_{\mathrm{xy}} + k_{\mathrm{xz}}k_{\mathrm{zy}}, 1 - k_{\mathrm{xz}}k_{\mathrm{zx}}, k_{\mathrm{xy}}k_{\mathrm{zx}} + k_{\mathrm{zy}}\Big).$$

Вычисляя объем параллелепипеда, натянутого на векторы m_{12}, m_{13}, m_{14}, получим искомое утверждение.

В данном случае объем получаемого параллелепипеда V можно интерпретировать как максимальный разброс выборки, при котором показатели регрессий не будут изменяться.

Заметим, что величину γ можно связать с таким фундаментальным понятием, как перманент матрицы [10].

Определение 1. Перманентом квадратной $n \times n$ матрицы $A = ||a_{ij}||$ называют сумму

$$Per(A) = \sum_\pi a_{1\pi_1} a_{2\pi_2} \dots a_{n\pi_n}$$

по всем перестановкам $\pi \in S_n$.

Легко показать, что

$$\gamma = 2 - Per\begin{pmatrix} 1 & k_{xy} & k_{xz} \\ k_{yx} & 1 & k_{yz} \\ k_{zx} & k_{zy} & 1 \end{pmatrix}$$

Теорема 3. Справедливо неравенство:

$$V \geq \frac{\alpha_\infty^x \alpha_\infty^y \alpha_\infty^z}{6}.$$

Доказательство. Каждое слагаемое в $\gamma = 1 - k_{\mathrm{xy}}k_{\mathrm{yx}} - k_{\mathrm{xy}}k_{\mathrm{yz}}k_{\mathrm{zx}} - k_{\mathrm{xz}}k_{\mathrm{yx}}k_{\mathrm{zy}} - k_{\mathrm{xz}}k_{\mathrm{zx}} - k_{\mathrm{yz}}k_{\mathrm{zy}}$, в силу теоремы 1, оценивается по модулю сверху единицей. Следовательно, знаменатель в равенстве (5) не превышает 6.

Данная теорема дает нижнюю оценку множества, содержащего Ω, при котором основные показатели L_∞-регрессии остаются неизменными. Это оценка позволяет значительно уменьшить объем первоначальной выборки, не теряя качество изучаемой модели.

Заключение и выводы. В данной статье была рассмотрена задача о построении и оценки параметров равномерно-регрессионной модели в трехмерном пространстве. Данная модель предназначается для анализа статистических данных, не привязана к закону распределения этих данных. Исследование данной модели осуществлено методами выпуклого анализа с использованием геометрической интерпретации модели. Приведенные алгоритмы и методы оценки качества модели имеют допустимую точность и могут быть реализованы для выборок большого объема.

Доказанные равенства позволяют уменьшить вычислительную сложность алгоритма, что положительно скажется на скорости работы основанных на них программ. Еще одним направлением в использовании полученных оценок является сокращение объема выборки с сохранением первоначальной регрессионной зависимости.

Библиографический список

1. Ponomarev I.V., Slavsky V.V. Uniformly fuzzy model of linear regression // Journal of Mathematical Sciences. 2012. Vol. 186, Issue 3. DOI: https://doi.org/10.1007/s10958-012-1002-1.

2. Карманов В.Г. Математическое программирование. 3-е изд. М., 1986.

3. Сантало Луи А. Интегральная геометрия и геометрические вероятности. пер. с англ. / под ред. Р.В. Амбарцумяна. М., 1983.

4. Половинкин Е.С., Балашов М.В. Элементы выпуклого и сильно выпуклого анализа. М., 2004.

5. Barber C.B., Dobkin D.P., Huhdanpa H.T. The Quickhull Algorithm for Convex Hulls // ACM Transactions on Mathematical Software. 1996. Vol. 22, № 4.

6. David M. Mount. Computational Geometry. University of Maryland, 2002.

7. Махаева Т.П., Пономарев И.В. Об оценке степени однородности выборки в равномерно-регрессионной модели // Известия Алт. гос. ун-та. 2020. № 1(111). DOI: https://doi.org/10.14258/izvasu(2020)1-19.

8. Брюс П., Брюс Э. Практическая статистика для специалистов Data Science : пер. с англ. СПб., 2018.

9. Кобзарь А.И. Прикладная математическая статистика. М., 2006.

10. Минк Х. Перманенты. М., 1982.