

DOI: 10.37791/2687-0649-2024-19-2-23-41

# Использование алгоритма оптимизации DIRECT и кусочно-линейных функций для построения нелинейных регрессионных моделей по гетероскедастичным данным

*Т. Т. Каримбаев<sup>1</sup>, С. А. Медведев<sup>1</sup>, Е. Т. Захарова<sup>1</sup>*

<sup>1</sup>Агрофизический научно-исследовательский институт, Санкт-Петербург, Россия  
*glogurin@yandex.ru*

**Аннотация.** Нелинейные регрессионные модели являются важным инструментом в сельскохозяйственных исследованиях, так как множество биологических процессов теоретически и экспериментально описываются нелинейными функциями. Помимо того, что нелинейные модели позволяют точно описывать экспериментальные данные, нелинейные функции обладают свойством физической интерпретируемости параметров и являются более устойчивыми вне области определения исследуемой выборки. На данный момент существующие методы расчета коэффициентов модели: метод наименьших квадратов, взвешенный метод наименьших квадратов и обобщенный метод наименьших квадратов обладают рядом недостатков. Наиболее совершенный метод – обобщенный метод наименьших квадратов опирается на большое количество аксиом, которые часто не соблюдаются в реальных примерах, а теоретическое доказательство не является аподиктическим. В статье представляется гибкий, устойчивый и точный метод расчета коэффициентов для произвольной однофакторной регрессионной модели на основе метода оценки максимального правдоподобия. Метод теоретически обосновывается с минимальным количеством аксиом, а также приводятся примеры результатов работы программной реализации для логистической функции и функции Михаэлиса на тестовых синтетических данных и экспериментальных выборках выработки сухой массы травы в зависимости от объема азотных удобрений. Основное преимущество метода заключается в простоте теоретического доказательства и малого количества теоретических ограничений на входные параметры задачи. Также предложенный метод, в отличие от обобщенного метода наименьших квадратов, детерминированно сходится на абсолютном минимуме благодаря использованию алгоритма DIRECT, может учитывать гетероскедастичность и не требует ручной настройки параметров оптимизации для обеспечения сходимости. Также представлены соображения о возможном расширении метода для многофакторного регрессионного анализа и возможные улучшения для оценки гетероскедастичности.

**Ключевые слова:** нелинейные регрессионные модели, метод максимального правдоподобия, гетероскедастичность, C#

**Для цитирования:** Каримбаев Т.Т., Медведев С.А., Захарова Е.Т. Использование алгоритма оптимизации DIRECT и кусочно-линейных функций для построения нелинейных регрессионных моделей по гетероскедастичным данным // Прикладная информатика. 2024. Т. 19. № 2. С. 23–41. DOI: 10.37791/2687-0649-2024-19-2-23-41

# Using the DIRECT optimization algorithm and piecewise-linear functions for building nonlinear regression model

T. Karimbaev<sup>1</sup>, S. Medvedev<sup>1</sup>, E. Zakharova<sup>1</sup>

<sup>1</sup>Agrophysical Research Institute, Saint Petersburg, Russia  
glogurin@yandex.ru

**Abstract.** Nonlinear regression models are an important tool in agricultural research, as many biological processes are theoretically and experimentally described by nonlinear functions. In addition to accurately describing experimental data, nonlinear models have the property of physical interpretability of parameters and are more robust outside the domain of the studied sample. Currently, existing methods for calculating model coefficients – such as Ordinary Least Squares, Weighted Least Squares, and Generalized Least Squares – have several drawbacks. The most advanced Generalized Least Squares method relies on a large number of axioms, which are often not adhered to in real examples, and the theoretical proof is not apodictic. This article introduces a flexible, robust, and accurate method for calculating coefficients for arbitrary single-factor regression models based on the maximum likelihood estimation method. The method is theoretically justified with a minimal number of axioms, and examples of results from the software implementation are provided for the logistic function and the Michaelis function using synthetic test data and experimental samples of dry grass mass production depending on the volume of nitrogen fertilizers. The main advantage of the method lies in the simplicity of theoretical proof and the small number of theoretical constraints on the input parameters of the problem. Unlike Generalized Least Squares, the proposed method deterministically converges to the absolute minimum, thanks to the use of the DIRECT algorithm. It can account for heteroscedasticity and does not require manual tuning of optimization parameters to ensure convergence. Considerations for possible extensions of the method to multifactorial regression analysis and potential improvements for heteroscedasticity estimation are also presented.

**Keywords:** nonlinear regression models, method of maximum likelihood, heteroscedasticity, C#

**For citation:** Karimbaev T., Medvedev S., Zakharova E. Using the DIRECT optimization algorithm and piecewise-linear functions for building nonlinear regression model. *Prikladnaya informatika*=Journal of Applied Informatics, 2024, vol.19, no.2, pp.23-41 (in Russian). DOI: 10.37791/2687-0649-2024-19-2-23-41

## Введение

Многие процессы, протекающие в агроэкосистемах, можно описать только нелинейными уравнениями [1, 2]. Их коэффициенты, как правило, основываются не на фундаментальных константах, а на обобщении данных эмпирических наблюдений. Поэтому нелинейные регрессионные модели являются важным инструментом как в сель-

скохозяйственных исследованиях, так и в системах поддержки принятия решений [3, 4]. Для построения этих моделей по облаку точек необходимы соответствующие математические методы. Обычно для решения данной задачи используется метод наименьших квадратов (МНК) либо его модификация – взвешенный метод наименьших квадратов (ВМНК) или его обобщение (ОМНК) [1, 5]. МНК и ВМНК можно использовать