

АНОТАЦІЯ

Ткаченко О.М. Поліноміальні методи та засоби оцінювання параметрів регресії з використанням моделей негаусових помилок. – Кваліфікаційна наукова праця на правах рукопису.

Дисертація на здобуття наукового ступеня доктора філософії за спеціальністю 122 – комп'ютерні науки (12 Інформаційні технології). – Черкаський державний технологічний університет, Черкаси, 2021.

Дисертаційна робота спрямована на вирішення актуальної науково-технічної задачі створенні методів і засобів математичного і комп'ютерного моделювання процесів оцінювання параметрів регресійних залежностей шляхом адаптації ймовірнісних моделей досліджуваних даних для покращення точності отримуваних оцінок на основі урахування відмінностей статистичних властивостей регресійних помилок від гаусової ідеалізації.

У першому розділі роботи проведено огляд інформаційних джерел за тематикою дисертаційного дослідження, який свідчить, що не зважаючи на більш ніж двохсотлітню історію регресійного аналізу задача підвищення точності оцінювання параметрів на основі урахування специфіки регресійних моделей залишається актуальною. На сучасному етапі розвитку науково-технічного прогресу вирішення подібних задач неможливо уявити без використання комп'ютерних технологій та спеціалізованих програмних засобів основу яких, складають відповідні математичні моделі та обчислювальні методи.

Основна гіпотеза про потенційне підвищення ефективності опрацювання (критерієм якого є величина дисперсії оцінок інформативних параметрів) базується на тому, що використання класичного методу найменших квадратів забезпечує отримання оптимальних оцінок лише при виконанні ряду теоретичних припущень, зокрема при гаусовому (нормальному) розподілі помилок. В ситуаціях коли ефект нормалізації відсутній для підвищення ефективності (мінімізації дисперсії) оцінок регресійних параметрів застосовується адаптивний підхід, що базується на максимізації правдоподібності. Проте практичне використання параметричних методів

адаптивного оцінювання, ускладнюється суттєвими вимогами до наявності апріорної інформації про тип та значення параметрів розподілу регресійних помилок.

Відомо, що одним із компромісних підходів до вирішення задач, пов'язаних із опрацюванням негаусових даних, є напрямок, що базується на використанні нелінійних перетворень у вигляді стохастичних поліномів Кунченка, який надає додаткові важелі для підвищення точності порівняно із лінійними методами, які оптимізовані для гаусових моделей.

У другому розділі роботи запропоновано новий підхід до адаптивного знаходження оцінок параметрів на основі використання для опису випадкової складової регресійних моделей статистик вищих порядків, що надало можливість реалізаційно просто враховувати відхилення від гаусової ідеалізації в процесі синтезу результатуючих обчислювальних методів та алгоритмів.

На основі апарату стохастичних поліномів та отриманих із використанням моментно-кумулянтного опису модифікацій регресійних моделей здійснено синтез обчислювальних методів адаптивного оцінювання параметрів регресійних моделей лінійного, поліноміального і нелінійного типу. Показано, що загальна задача, алгоритмічно може бути зведена до розв'язання системи нелінійних стохастичних рівнянь із застосуванням чисельної ітераційної процедури Ньютона-Рафсона.

Проведений у третьому розділі роботи теоретичний аналіз показав, що застосування запропонованого способу поліноміального оцінювання загалом забезпечує зменшення дисперсії отримуваних оцінок, порівняно із відомими оцінками методу найменших квадратів. При цьому ступінь ефективності не залежить від типу регресійних моделей та є однаковим для всіх складових параметрів регресії. Зростання точності досягається завдяки врахуванню негаусовості регресійних помилок. Кількісно величина ефективності поліноміальних оцінок (відносно лінійних) залежить від ступеня негаусовості, яка описується значеннями кумулянтних коефіцієнтів вищих порядків.

У четвертому розділі роботи представлено опис структури розробленого програмного комплексу, набір модулів якого забезпечують як безпосереднє

вирішення задачі знаходження адаптивних оцінок параметрів регресійних залежностей так і реалізацією комп'ютерного статистичного моделювання на основі методу Монте-Карло і бутстреп-аналізу. Сукупність отриманих результатів статистичного моделювання у цілому підтверджують теоретично доведену ефективність поліноміальних оцінок відносно лінійних оцінок методу найменших квадратів. На прикладі моделі помилок із експоненціальним степеневим розподілом з також показано, що при відсутності апріорної інформації про значення параметрів регресійних помилок адаптивні поліноміальні оцінки можуть бути більш точними порівняно із класичними оцінкам (зростання точності щодо методу найменших квадратів складає до 60%, а відносно методу максимальної правдоподібності до 10%).

Наукова новизна дисертаційної роботи полягає у створенні методів математичного моделювання процесів оцінювання параметрів регресійних залежностей при негаусовому розподілі їх випадкових помилок на основі використання статистик вищих порядків, методу максимізації полінома (методу Кунченка), що дозволяє зменшити дисперсію інформативних оцінок параметрів і забезпечує високу якість статистичного опрацювання в комп'ютерних системах. Вперше розроблено методи адаптивного поліноміального оцінювання параметрів регресії, які за рахунок використання удосконалених на основі застосування для опису регресійних моделей статистик вищих порядків дозволяють підвищити точність оцінок інформативних параметрів в умовах апріорної невизначеності щодо ймовірнісного характеру негаусових регресійних помилок. Крім того, отримали подальший розвиток елементи теорії оцінювання параметрів регресії з використанням методу максимізації полінома, що дозволяє отримувати субефективні, адаптивні і компромісні з точки зору практичної реалізації рішення з оцінювання інформативних параметрів регресії за негаусового характеру випадкових помилок.

Практична цінність одержаних результатів визначається тим, що отримані методи та засоби математичного і комп'ютерного моделювання дали змогу: розробити методику отримання адаптивних оцінок інформативних параметрів регресійних залежностей при асиметричному і симетричному характері їх

випадкових помилок, яка може бути використана для побудови нових обчислювальних алгоритмів аналізу даних; синтезувати нові алгоритми адаптивного оцінювання параметрів однофакторних і багатфакторних лінійних, поліноміальних та нелінійних регресійних залежностей, які дають змогу враховувати негаусовість регресійних залишків; провести аналіз можливості та ефективності застосування чисельного методу Ньютона-Рафсона та його модифікацій для розв'язку систем нелінійних рівнянь методу максимізації полінома для знаходження оцінок інформативних параметрів лінійних та нелінійних регресійних залежностей; розробити методіку та отримати аналітичні вирази для визначення теоретичної точності поліноміальних оцінок, з використанням яких можна здійснювати порівняльний аналіз відносно оцінок класичними методами найменших квадратів та максимальної правдоподібності. На основі використання програмного пакету MATLAB і спеціалізованої для аналізу даних мови високого рівня R розроблено програмні засоби, призначені для вирішення задач статистичного моделювання знаходження оцінок параметрів регресійних залежностей.

Розроблені програмні засоби були успішно застосовані при вирішенні прикладної задачі в рамках дослідження механічних характеристик системи подачі філаменту та реологічних характеристик екструдера 3D принтера. Результати застосування запропонованих моделей та методів оцінювання дозволили забезпечити зменшення на 10-25 % величини дисперсії оцінок інформативних параметрів нелінійних регресійних моделей.

Ключові слова: регресійні моделі, негаусові помилки, моменти, кумулянти, оцінювання параметрів, стохастичні поліноми

ABSTRACT

Tkachenko O.M. Polynomial methods and tools for estimating regression parameters using Non-Gaussian error models.

Thesis for a Doctor of Philosophy degree in specialty 122 – Computer Science. – Cherkasy State Technological University. Cherkasy, 2021.

The work was solved the scientific and technical problem of development and application of mathematical and computer modeling methods for processes of estimating regression parameters under the condition of non-Gaussian character of their errors.

A new approach to the adaptive finding of parameter estimates based on the use of higher-order regression models for describing the random component is proposed, which made it possible to simply take into account deviations from Gaussian idealization in the synthesis and analysis of the effectiveness of the resulting methods.

Based on the apparatus of stochastic Kunchenko polynomials and modifications of regression models obtained using instantaneous-cumulative description, the synthesis of computational methods of adaptive parameters estimation of regression models for linear, polynomial, and nonlinear type is carried out. It is shown that the general problem can be algorithmically reduced to solving a system of nonlinear stochastic equations using a numerical Newton-Rafson iterative procedure.

The properties of polynomial estimates under the condition of the asymmetric and symmetric character of non-Gaussian errors are analyzed and their efficiency is compared with classical estimates of least squares and maximum likelihood. It is shown that the application of the proposed approach reduces the variance of polynomial estimates compared to the known least squares estimates, and the increase in accuracy is achieved by taking into account the non-Gaussian regression errors.

The developed software package, its structure, and set of modules provide both a direct solution to the problem of finding adaptive estimates of the parameters of regression dependencies and the implementation of computer statistical modeling based on the Monte Carlo method and bootstrap analysis. The set of obtained results of statistical modeling confirms the theoretically proven efficiency of polynomial

estimates. The example of the error model with exponential power distribution shows that in the absence of a priori information about the values of regression error parameters, adaptive polynomial estimates can be more accurate even compared to adaptive estimates of maximum plausibility.

Keywords: regression models, non-Gaussian errors, moments, cumulants, parameter estimation, stochastic polynomials

Список публікацій в яких опубліковані основні наукові результати дисертації:

1. Zabolotnii, S., Warsza, Z. L., & Tkachenko, O. (2018). Polynomial Estimation of Linear Regression Parameters for the Asymmetric PDF of Errors. *Automation 2018. Advances in Intelligent Systems and Computing*, 743, 758–772. doi:10.1007/978-3-319-77179-3_75 (**Scopus**)
2. Zabolotnii, S. W., Warsza, Z. L., & Tkachenko, O. (2020). Estimation of Linear Regression Parameters of Symmetric Non-Gaussian Errors by Polynomial Maximization Method. *Automation 2019. Advances in Intelligent Systems and Computing*, 920, 636–649. doi:10.1007/978-3-030-13273-6_59 (**Scopus**)
3. Zabolotnii, S., Khotunov, V., Chepynoha, A., & Tkachenko, O. (2021). Estimating parameters of linear regression with an exponential power distribution of errors by using a polynomial maximization method. *Eastern-European Journal of Enterprise Technologies*, 1(4 (109)), 64–73. <https://doi.org/10.15587/1729-4061.2021.225525> (**Scopus**)

Список публікацій які засвідчують апробацію матеріалів дисертації:

4. Заболотній С.В. Поліноміальні адаптивні процедури регресійного аналізу із використанням моделей негаусових помилок на основі статистик вищих порядків/ С.В. Заболотній, О.М. Ткаченко // Тези доповідей IV Міжнародної науково-практичної конференції «Обчислювальний інтелект (результати, проблеми, перспективи) – 2017» (ComInt – 2017): Київ, 16-18 травня 2017 р. – К: КНУ ім. Т. Шевченка, – 2017. С. 113-114
5. Заболотній С.В. Застосування методу максимізації полінома для оцінювання параметрів однофакторної лінійної регресії при негаусовому розподілі

- помилки / С.В. Заболотній, О.М. Ткаченко // Тези доповідей VI Міжнародної науково-практичної конференції «Обробка сигналів і негаусівських процесів – 2017» (ОСНП-2017): Черкаси, 24-26 травня 2017 р. – Черкаси: ЧДТУ, – 2017. С. 74-76.
6. Заболотній С.В. Аналіз ефективності поліноміальних оцінок параметрів лінійної регресії при симетричному розподілі негаусових помилок / С.В Заболотній, М.П. Рудь, О.М. Ткаченко // Сучасні прилади, матеріали і технології для неруйнівного контролю і технічної діагностики машинобудівного і нафтогазопромислового обладнання, 14-16 листопада 2017: Тези доповідей 8-ма міжнародна н/т конф. – Івано-Франківськ, 2017. – С.130-131
 7. Заболотній С.В. Особливості поліноміального оцінювання параметрів регресії при негаусовому симетричному розподілі помилок / С.В Заболотній, М.П. Рудь, О.М. Ткаченко // Автоматика та комп'ютерно-інтегровані технології у промисловості, телекомунікаціях, енергетиці та транспорті, 16-17 листопада 2017: Матеріали Всеукраїнської науково-практичної інтернет-конференції. – Кропивницький, 2017. – С.177-179
 8. Заболотній С.В., Рудь М.П., Ткаченко О.М. Застосування методу максимізації полінома для оцінювання параметрів нелінійних регресійних моделей // Праці VII Міжнародної науково-практичної конференції «Обробка сигналів і негаусівських процесів», присвяченої пам'яті професора Ю.П. Кунченка: Тези доповідей. [Електронний ресурс]. – Черкаси: ЧДТУ, 2019, С. 76-79.