

УДК 519.2

С. В. Заболотній, д.т.н., доцент,**М. П. Рудь, к.т.н., доцент,****К. В. Іващенко, аспірант**

Черкаський державний технологічний університет

б-р Шевченка, 460, Черкаси, 18006, Україна

e-mail: s.zabolotnii@chdtu.edu.ua

**МОДЕЛЮВАННЯ ПОСЛІДОВНОГО ОЦІНЮВАННЯ ПАРАМЕТРА ЗСУВУ
АСИМЕТРИЧНО-РОЗПОДІЛЕНИХ ВИПАДКОВИХ ВЕЛИЧИН
МЕТОДОМ МАКСИМІЗАЦІЇ ПОЛІНОМА**

У статті досліджується новий алгоритм послідовного оцінювання параметра зсуву негаусових асиметрично-розподілених випадкових величин, який базується на методі максимізації полінома та моментно-кумулянтному описі. Отримано аналітичні вирази для знаходження послідовних оцінок при другому степені полінома. Шляхом статистичних випробувань методом Монте-Карло проведено порівняльний аналіз ефективності поліноміальних послідовних оцінок з відомими лінійними оцінками.

Ключові слова: послідовне оцінювання, параметр зсуву, асиметричний розподіл, стохастичні поліноми, кумулянтні коефіцієнти.

Вступ. Послідовний статистичний аналіз є спеціальним розділом математичної статистики, основною специфікою постановки задач якого є те, що рішення приймаються не на основі опрацювання вибірок фіксованого обсягу, а шляхом формування статистик із даних, що надходять у ході експерименту. Теоретичний базис підходів до вирішення задач послідовного аналізу був закладений у роботах Вальда ще у середині минулого століття [1]. Відомо, що послідовні методи надають більшу гнучкість при проведенні статистичного експерименту і, з точки зору середнього обсягу спостережень, у багатьох ситуаціях є значно ефективнішими, ніж традиційні статистичні процедури. Ще однією позитивною властивістю послідовного аналізу є можливість застосування ефективних (з обчислювальної точки зору) рекурентних алгоритмів при формуванні результуючих статистик.

У цьому дослідженні розглядається задача оцінювання параметра зсуву асиметричного розподілу, яку також можна трактувати як визначення величини постійної складової за умови наявності адитивних випадкових похибок вимірювання. Відомо [2], що існує декілька підходів до вирішення такої задачі. Найбільш оптимальним (за критерієм точності) є використання методу максимальної правдоподібності. Проте його застосування вимагає наявності апріорної інформації про вигляд закону розподілу статистичних даних та часто

призводить до достатньо складних обчислювальних алгоритмів. Дуже поширеним є робастний підхід [3], який базується на порядкових або рангових статистиках, але його послідовна реалізація також характеризується суттєвими обчислювальними затратами. Тому на практиці найчастіше використовують звичайну статистику у вигляді середнього арифметичного, для послідовного обчислення якої існує проста рекурентна модифікація [4]. Основним недоліком такого підходу є те, що дисперсія оцінки середнього арифметичного можуть суттєво перевищувати ефективні значення при відмінності розподілу статистичних даних від гаусової моделі.

Одним із альтернативних підходів до статистичного оцінювання є метод максимізації полінома (ММПл) [5]. Цей відносно новий метод базується на ймовірнісному описі у вигляді статистик вищих порядків (моментів або кумулянтів). У роботах [6–8] досліджено властивості та проаналізовано ефективність ММПл-оцінок параметра зсуву симетричних та асиметричних розподілів. Показано, що застосування ММПл при негаусовому характері статистичних даних дозволяє зменшувати дисперсію оцінок шуканого параметра на основі урахування значень статистик вищих порядків.

Метою роботи є застосування статистичного моделювання для доведення ефективності алгоритму послідовного оцінювання параметра зсуву негаусових асиметрично-розподілених

випадкових послідовностей, що базується на застосуванні ММПл при степені $r = 2$.

Математична постановка задачі послідовного оцінювання. Нехай спостерігається випадкова послідовність $x_1, x_2, \dots, x_n, \dots$, кожний елемент якої являє собою однаково розподілене та статистично незалежне вибіркове значення, що описується моделлю

$$\xi = \theta + \eta, \tag{1}$$

де $\theta = \text{const}$, η – центрована (із нульовим математичним сподіванням) асиметрично-розподілена випадкова величина, ймовірнісний характер якої описується, дисперсією (кумулянтном другого порядку) κ_2 і кумулянтними коефіцієнтами асиметрії γ_3 та ексцесу γ_4 .

Загальна задача полягає у формуванні послідовних оцінок параметра зсуву $\hat{\theta}_n$, які б враховували ймовірнісний характер стохастичної складової моделі (1).

Методи послідовного оцінювання

1. Знаходження послідовних лінійних оцінок параметра зсуву. Початково розглянемо класичне розв’язання поставленої задачі із використанням лінійної оцінки шуканого параметра (для фіксованого обсягу вибірки) у вигляді середнього арифметичного:

$$\hat{\theta} = \bar{X} = \frac{1}{n} \sum_{v=1}^n x_v. \tag{2}$$

Очевидно, що безпосереднє використання статистики (2) для отримання послідовних оцінок є затратним (з точки зору обчислювальних ресурсів), оскільки на кожному n -му кроці потребує n операцій додавання та збереження в пам’яті $n-1$ попередніх вибіркових значень x_v , $v = 1, n-1$. Тому для отримання послідовного варіанта лінійних оцінок (2) зазвичай використовують відоме [4, с. 128] рекурентне співвідношення виду

$$\hat{\theta}_n = \frac{n-1}{n} \bar{X}_{n-1} + \frac{x_n}{n} = \frac{n-1}{n} \hat{\theta}_{n-1} + \frac{x_n}{n}, \tag{3}$$

що дозволяє уточнювати оцінку шуканого параметра, здійснюючи лише одну операцію додавання і зберігаючи лише одне (попереднє) оцінне значення $\hat{\theta}_{n-1}$.

2. Знаходження послідовних ММПл-оцінок параметра зсуву. На основі [5–7] відомо, що лінійна оцінка (2) є частковим випадком ММПл-оцінок параметра зсуву при застосуванні степені полінома $r = 1$. Основною перевагою такої оцінки є простота, але спосіб

її формування не враховує ймовірнісного характеру стохастичної складової моделі (1) випадкових величин.

При застосуванні степені полінома $r = 2$ ММПл-оцінки шуканого параметра в загальному випадку можуть бути знайдені із вирішення стохастичного рівняння виду [5]

$$h_1 \left[\frac{1}{n} \sum_{v=1}^n x_v - \theta \right] + h_2 \left[\frac{1}{n} \sum_{v=1}^n x_v^2 - (\theta^2 + \kappa_2) \right] \Big|_{\theta=\hat{\theta}} = 0, \tag{4}$$

де оптимальні коефіцієнти, що знаходяться за умови забезпечення мінімуму дисперсії оцінок параметра θ , можна записати у вигляді

$$h_1 = 2\theta\gamma_3 + \kappa_2^{1/2}(2 + \gamma_4), \quad h_2 = -\gamma_3.$$

Характерною особливістю ММПл є те, що алгоритми пошуку оцінок зводяться до вирішення степеневих рівнянь відносно параметра, що оцінюється. У цьому випадку рівняння (4) зводиться до квадратичного і має аналітичний розв’язок, що дозволяє отримувати незміщену ММПл-оцінку шуканого параметра у вигляді [5]

$$\hat{\theta}^{(2)} = \frac{1}{n} \sum_{v=1}^n x_v + \kappa_2^{1/2} \frac{2 + \gamma_4}{\gamma_3} \left[\frac{n-1}{n} \kappa_2 - \frac{1}{n} \sum_{v=1}^n (x_v)^2 + \left(\frac{1}{n} \sum_{v=1}^n x_v \right)^2 \right]. \tag{5}$$

Аналіз виразу (5) свідчить, що для отримання ММПл-оцінки при степені $r = 2$, окрім лінійної статистики виду (2), додатково необхідним є обчислення квадратичної статистики $\overline{X^2} = \frac{1}{n} \sum_{v=1}^n (x_v)^2$, послідовний варіант формування якої також може бути отримано на основі рекурентного співвідношення, аналогічного до (3), тобто:

$$\overline{X^2}_n = \frac{n-1}{n} \overline{X^2}_{n-1} + \frac{(x_n)^2}{n}.$$

Таким чином, вираз для формування послідовної ММПл-оцінки (при $r = 2$) може бути записано у формі:

$$\hat{\theta}_n^{(2)} = \frac{n-1}{n} \bar{X}_{n-1} + \frac{x_n}{n} + \kappa_2^{1/2} \frac{2 + \gamma_4}{\gamma_3} \times \left[\frac{n-1}{n} \left(\kappa_2 - \overline{X^2}_{n-1} \right) + \frac{(x_n)^2}{n} + \left(\frac{n-1}{n} \bar{X}_{n-1} + \frac{x_n}{n} \right)^2 \right]. \tag{6}$$

На рис. 1 зображено модельний приклад послідовного оцінювання параметра зсуву

(при істинному значенні $\theta = 1$). Ступінь негаусового характеру моделі стохастичної складової випадкової послідовності x_n (рис. 1, а) визначається кумулянтними коефіцієнтами асиметрії $\gamma_3 = 1.4$ і ексцесу $\gamma_3 = 3$. На рис. 1, б зображено динаміку залежності послідовних оцінок параметра, що формуються на основі виразів (3) та (6), від кількості отриманих даних n .

Аналіз рис. 1 підтверджує працездатність алгоритмів послідовного оцінювання

параметра зсуву, тому що зі зростанням довжини послідовності n величини оцінних значень θ_n наближаються до істинного значення цього параметра. І хоча графіки, зображені на рис. 1, б, свідчать про потенційну перевагу (у швидкості збіжності до істинного значення) поліноміальної (при $r = 2$) ММПл-оцінки (6) порівняно з лінійною оцінкою (3), проте ці залежності є результатом однократного експерименту, що не дозволяє оцінити у цілому ступінь їх ефективності.

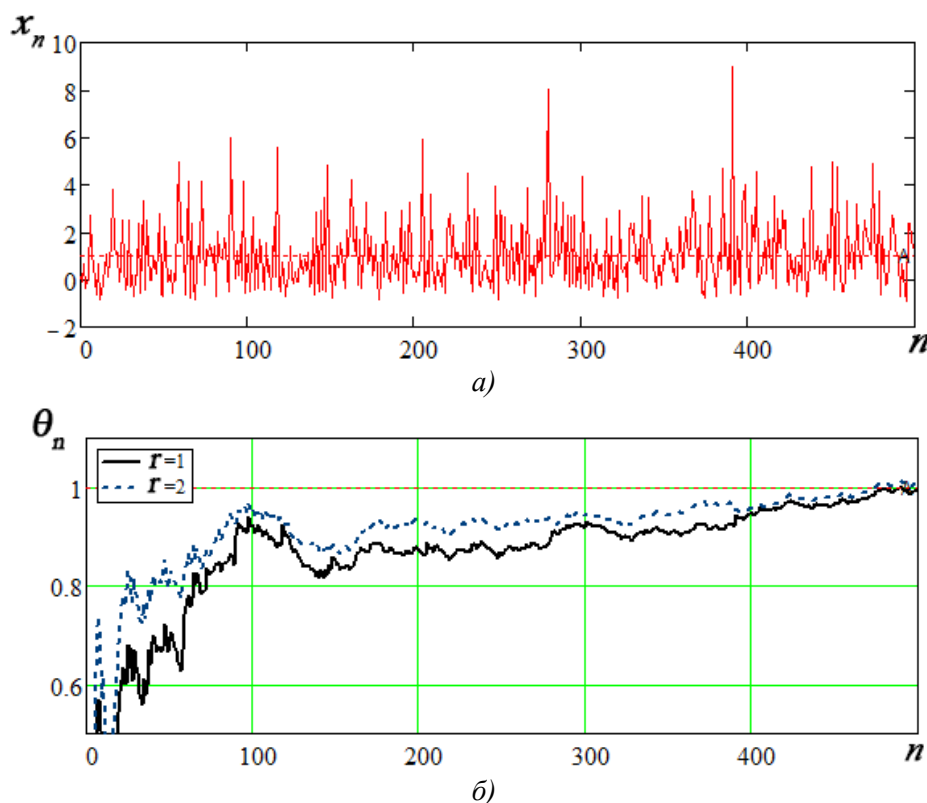


Рис. 1. Послідовне оцінювання параметра зсуву асиметричної випадкової величини: а) вхідна послідовність вибірових значень; б) значення послідовних оцінок параметра зсуву

Аналіз ефективності послідовних оцінок. Відомо, що для оцінок параметрів, які отримуються із вибірок фіксованого обсягу, основним критерієм точності є їх ступінь невизначеності, що чисельно визначається середньоквадратичним відхиленням (СКВ) від істинного значення параметра або дисперсією оцінки, яка є квадратичною функцією від СКВ. В роботі [5] показано, що співвідношення дисперсії оцінок (5) та (2) асимптотично (при $n \rightarrow \infty$) описуються виразом

$$\sigma_{\theta(r=2)}^2 = \frac{\kappa_2}{n} \left[1 - \frac{\gamma_3^2}{2 + \gamma_4} \right] = \sigma_{\theta(r=1)}^2 \left[1 - \frac{\gamma_3^2}{2 + \gamma_4} \right]. \quad (7)$$

Таким чином коефіцієнт зменшення дисперсії

$$g_2 = \sigma_{\theta(r=2)}^2 / \sigma_{\theta(r=1)}^2 = 1 - \frac{\gamma_3^2}{2 + \gamma_4} \quad (8)$$

залежить лише від ступеня негаусовості (значень кумулянтних коефіцієнтів асиметрії та ексцесу) випадкових послідовностей і, як відомо, належить діапазону $(0; 1]$.

Як вже було зазначено вище, результати однократних експериментів, приклад яких наведено на рис. 1, не дозволяють адекватно порівнювати точність різних алгоритмів послідовного оцінювання. Для порівняння можуть бути використані лише статистики, що

отримуються шляхом багаторазових експериментів з однаковими вихідними значеннями параметрів моделі (метод Монте-Карло).

На підставі отриманих результатів в програмному середовищі MATLAB розроблено набір m -функцій, що реалізують статистичне моделювання роботи алгоритмів лінійного та поліноміального послідовного оцінювання параметра зсуву негаусових випадкових величин з різними типами розподілів (експоненціальним, гамма, логнормальним, Вейбула, бігаусовим).

На рис. 2 зображено один із прикладів результатів статистичного моделювання ме-

тодом Монте-Карло. У цьому прикладі здійснюється послідовне оцінювання параметра зсуву $\theta=1$ асиметрично-розподіленої випадкової величини, ймовірнісні властивості моделі якої задаються кумулянтними коефіцієнтами асиметрії $\gamma_3=1.4$ і ексцесу $\gamma_3=3$. В такій ситуації застосування ММПл при степені полінома $r=2$ відповідно до (8) теоретично забезпечує асимптотичне (зі зростанням довжини вхідної послідовності) зменшення дисперсії оцінок (порівняно з лінійною статистикою) на коефіцієнт $g_2=0.6$.

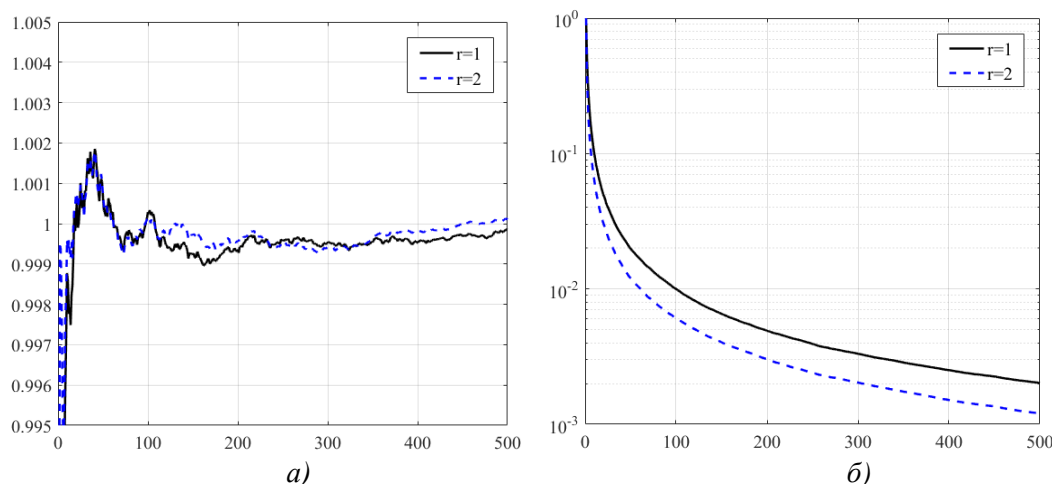


Рис. 2. Усреднені залежності статистичних характеристик послідовного оцінювання параметра зсуву асиметричної випадкової величини:

а) середні значення оцінок; б) дисперсії оцінок

Аналіз зображених на рис. 2 усереднених (за результатами 10^5 експериментів) статистичних характеристик послідовних оцінок параметра θ підтверджує теоретично передбачену ефективність застосування ММПл. Зокрема, графіки динаміки характеристик на рис. 2, б свідчать, що, наприклад, для досягнення рівня зменшення дисперсії до 10^{-2} ММПл-оцінкам необхідно у середньому на 40 % меншу кількість вимірювань ($n=60$ вибірових значень для $r=2$ порівняно з $n=100$ для лінійних оцінок при $r=1$). Ці експериментальні результати відповідають теоретично розрахованим величинам коефіцієнта зменшення дисперсії (8). Характер залежності змін середніх значень послідовних оцінок, зображений на рис. 2, а, для обох алгоритмів фактично не відрізняється і підтверджує тезу про їх асимптотичну незміщеність.

Висновки. Сукупність отриманих результатів моделювання послідовного оціню-

вання параметра зсуву випадкових величин у цілому підтверджує ефективність застосування методу максимізації полінома та моментно-кумулянтного опису для випадку, коли розподіл статистичних даних має негаусовий характер.

Застосування подібних алгоритмів забезпечить суттєве зменшення часу на прийняття рішень при діагностиці стану технічних систем і технологічних процесів.

Одним із подальших завдань для проведення досліджень у цьому напрямку є аналіз ефективності застосування чисельних ітеративних процедур, що призначені для спрощення алгоритмів формування послідовних оцінок.

Список літератури

1. Вальд А. Последовательный анализ. Москва: Физматлит, 1960. 328 с.
2. Айвазян С. А., Енюков И. С., Мешалкин Л. Д. Прикладная статистика: Основы

- моделирования и первичная обработка данных: справ. изд. Москва: Финансы и статистика, 1983. 471 с.
3. Хьюбер П. Робастность в статистике. Москва: Мир, 1984. 304 с.
 4. Макс Ж. Методы и техника обработки сигналов при физических измерениях: в 2 т. Москва: Мир, 1983. Т. 1. 312 с.
 5. Кунченко Ю. П., Лега Ю. Г. Оценка параметров случайных величин методом максимизации полинома. Киев: Наукова думка, 1991. 180 с.
 6. Warsza Z. L., Zabolotnii S. W. A polynomial estimation of measurand parameters for samples of non-Gaussian symmetrically distributed data. *Automation 2017. Advances in Intelligent Systems and Computing*, vol. 550. Cham: Springer International Publishing, 2017. P. 468–480.
 7. Warsza Z. L., Zabolotnii S. W. Estimation of measurand parameters for data from asymmetric distributions by polynomial maximization method. *Automation 2018. Advances in Intelligent Systems and Computing*, vol. 743. Cham: Springer International Publishing, 2018. P. 746–757.
 8. Warsza Z. L., Zabolotnii S. W. Polynomial estimation of the measurand parameters for samples from non-Gaussian distributions based on higher order statistics. *Advanced Mathematical and Computational Tools in Metrology and Testing XI. WORLD SCIENTIFIC*, 2018. P. 383–400.
 2. Ajvazyan, S. A. Enyukov, I. S., Meshalkin, L. D. (1983) Applied statistics: Basics of modeling and primary data processing: reference book. Moscow: Finansy i statistika, 471 p. [in Russian].
 3. Xyuber, P. (1984) Robustness in statistics. Moscow: Mir, 304 p. [in Russian].
 4. Maks, Zh. (1983) Methods and techniques for signal processing in physical measurements: in 2 vol. Moscow: Mir. T. 1, 312 p. [in Russian].
 5. Kunchenko, Yu. P., Lega, Yu. G. (1991) Estimation of random variables parameters by the method of polynomial maximization. Kiev: Naukova dumka, 180 p. [in Russian].
 6. Warsza, Z. L., Zabolotnii, S. W. (2017) A polynomial estimation of measurand parameters for samples of non-Gaussian symmetrically distributed data. *Automation 2017. Advances in Intelligent Systems and Computing*, vol. 550. Cham: Springer International Publishing, pp. 468–480.
 7. Warsza, Z. L., Zabolotnii, S. W. (2018) Estimation of measurand parameters for data from asymmetric distributions by polynomial maximization method. *Automation 2018. Advances in Intelligent Systems and Computing*, vol. 743. Cham: Springer International Publishing, pp. 746–757.
 8. Warsza, Z. L., Zabolotnii, S. W. (2018) Polynomial estimation of the measurand parameters for samples from non-Gaussian distributions based on higher order statistics. *Advanced Mathematical and Computational Tools in Metrology and Testing XI. WORLD SCIENTIFIC*, pp. 383–400.

References

1. Val'd, A. (1960) Sequential analysis. Moscow: Fizmatlit, 328 p. [in Russian].

S. V. Zabolotnii, *Dr.Tech.Sc, associate professor,*

M. P. Rud, *PhD,*

K. V. Ivashchenko, *PhD student*

Cherkasy State Technological University

Shevchenko Blvd, 460, Cherkasy, 18006, Ukraine

e-mail: s.zabolotnii@chdtu.edu.ua

MODELING OF SEQUENTIAL ESTIMATION OF BIAS PARAMETER OF ASYMMETRIC-DISTRIBUTED RANDOM VARIABLES USING POLYNOMIAL MAXIMIZATION METHOD

An original approach to finding sequential estimates of the parameter of bias of non-Gaussian asymmetric-distributed random variables is investigated in the paper. The polynomial maximization method (PlMM), which is based on the mathematical apparatus of stochastic Kunchchenko poly-

mials and a partial description of random variables by higher order statistics (moments or cumulants) is the basis of this approach. The classic approach to solving a posed problem, which is based on simple linear recurrent statistics, that does not take into account the peculiarities of probabilistic data distribution and is optimal only for Gaussian model, is analyzed. Analytical expressions for finding the estimates by polynomial maximization method at the second degree polynomial are obtained. A comparative analysis of the efficiency on the basis of the criterion of the magnitude of asymptotic dispersion of the estimates of various methods parameters is performed. It is shown that theoretical value of the coefficient of the reduction of PLMM-estimates dispersion (in comparison with linear estimates) depends on the magnitude of cumulative coefficients of asymmetry and excess of statistical data. On the basis of the received results, in MATLAB software environment a set of m-functions that realize statistical modeling by Monte-Carlo method of linear and polynomial sequential grading algorithms for the parameter of bias of non-Gaussian random variables with different types of distributions (exponential, gamma, lognormal, Weibull, double-Gaussian ones) is developed. The combination of the obtained results shows that the application of the proposed approach can provide a significant reduction in the time to make decisions when diagnosing the state of technical systems and technological processes.

Keywords: *sequential estimation, bias parameter, asymmetric distribution, stochastic polynomials, cumulative coefficients.*