УДК 519.862.6
DOI: https://doi.org/10.32851/2708-0366/2022.11.16

Дебела І.М.
кандидат сільськогосподарських наук, доцент,
Херсонський державний аграрно-економічний університет
ORCID: https://orcid.org/0000-0001-7990-4202

Debela Iryna
Kherson State Agrarian and Economic University

КЛАСИФІКАЦІЯ СТАНІВ СИСТЕМИ ЗА ВЕКТОРОМ ПАРАМЕТРІВ

CLASSIFICATION OF SYSTEM STATES BY VECTOR PARAMETERS

У статті описано аналітичний алгоритм класифікації станів складної системи, ідентифікованих за вектором параметрів. Алгоритм визначення стану складної системи ґрунтується на принципах дискримінантного аналізу. Множина класів системи розглядається як сукупність багатовимірних випадкових величин, визначених з точністю до значень параметрів. Основою застосування дискримінантного аналізу є припущення про нормальний розподіл багатовимірної випадкової величини, а саме вектору параметрів стану системи. Ризики класифікації оцінюються за Байєсовим вирішуючим правилом, одним із проміжних результатів якого є визначення статистичних оцінок апріорних імовірностей належності досліджуваної системи до кожного класу. Отримані оцінки використовуються в задачах оптимізації прийняття рішення в умовах потенційних економічних ризиків. Байєсовський підхід – це не новий алгоритм оптимальної класифікації, але його застосування до прикладних задач моделювання вимагає аналітичної адаптації.

Ключові слова: *дискримінантний аналіз, вектор параметрів, аналітична модель, ризик, система, класифікація станів.*

The aim of the article is to study the algorithms for classifying the states of the modeled economic system according to the vector of parameters. The economy is seen as an emergent system for which the general principles of existence and development of complex systems are valid. It is the property of emergence that can cause the structural heterogeneity of the mathematical model of the system, which in turn is the result of abrupt changes in parameters, uncertainty of the functional relationship between exogenous and endogenous model variables. The adequacy of mathematical models of such problems is determined primarily by the presence of structured input information – numerical, probabilistic, descriptive estimates of the interaction of exogenous and endogenous factors of different nature, which can be formalized mathematically as model parameters. An analytical algorithm for classifying the states of a complex system identified by a vector of parameters is described. The algorithm for determining the state of a complex system is based on the principles of discriminant analysis. The set of system classes is considered as a set of multidimensional random variables determined to the nearest parameter values. The basis for the application of discriminant analysis is the assumption of a normal distribution of a multidimensional random variable – a vector of system state parameters. In practical calculations, the classification error is interpreted as the average classification error, which can be represented as a matrix of losses (fines, risks), due to incorrect classification of the system. The average value of the classification error can be determined by the Bayesian decision rule, one of the intermediate results of which is to determine statistical estimates of a priori probabilities of belonging of the studied system to each class. The obtained estimates are used in the tasks of optimizing decision-making in conditions of potential economic risks. The Bayesian approach is not a new algorithm for optimal classification, but its application to applied modeling problems requires analytical adaptation.

Key words: *discriminant analysis, vector of parameters, analytical model, risk, system, classification of states.*

Постановка проблеми. Економіка як об'єкт моделювання може розглядатися як система, для якої справедливі загальні принципи існування та розвитку складних систем. Характеристичною особливістю економіки як складної системи є властивість емерджентності [1]. Саме ця особливість може бути причиною структурної неоднорідності математичної моделі системи, яка є наслідком різких стрибкоподібних змін параметрів, невизначеності типу функціонального зв'язку між екзогенними та ендогенними ознаками – характеристичними змінними моделі. Адекватність математичних моделей таких задач визначається насамперед наявністю структурованої вхідної інформації, а саме числових, імовірнісних, описових оцінок взаємодії екзогенних та ендогенних факторів різної природи, що можуть бути формалізовані математично як параметри моделі. Обґрунтованість, статистична значимість, можливість формалізації вхідної інформації є основними критеріями у виборі методів та алгоритмів математичного моделювання. На практиці найчастіше застосовуються теоретико-імовірнісні та статистичні методи оцінювання вхідних параметрів моделі, коли зроблені на основі вибіркових спостережень висновки переносяться на всю сукупність. Можна виділити такі основні проблеми, що виникають під час побудови економіко-математичних моделей складних економічних систем:

1) суттєва стохастичність та багатовимірність інформації; неоднорідність досліджуваних кількісних параметрів, їх залежність від якісних ознак різної природи [2, с. 968–970];

2) неможливість розділення параметрів на дві незалежні групи, такі як ендогенні та екзогенні, тобто загалом ми можемо описати лише результат їх одночасної дії без визначення параметричної структури моделі;

3) нелінійність математичної моделі відносно її параметрів [3, с. 670]; параметричний вид таких моделей невідомий, як наслідок, виникає додаткова задача пошуку алгоритму непараметричної оцінки зв'язків; коли форма математичної моделі однозначно не встановлена на основі апріорної інформації, широко використовуються лінійні регресійні моделі, оцінювання та статистичний аналіз параметрів яких не є проблемою;

4) функціональне співвідношення параметрів моделі може змінюватись протягом часового і просторового періоду моделювання; крім того, емпірична закономірність, спостережена за вибірковими характеристиками, може значно відрізнятися від вибраного функціоналу моделі, що є результатом оброблення та формалізації вхідних параметрів.

Для кожної окремої задачі моделювання складної економічної системи слід не лише попередньо оцінити параметричну структуру моделі, але й спрогнозувати наслідки перебування системи в стані з визначеними параметрами, тому дослідження методів структуризації, формалізації та оцінювання параметричних характеристик економіко-математичних моделей залишається актуальним.

Аналіз останніх досліджень і публікацій. Теорія стохастичних систем, стохастичні моделі в економіці описані в монографіях В.І. Жлуктенко, А.В. Бєгун [4], П.І. Верченко [5]. Принцип системності в аналітичних дослідженнях описаний у роботі І.В. Спільник, О.В. Ярощук [6]. Статистичний аналіз моделей змінної структури досліджується в роботах С.А. Айвазян [7], В.Е. Зотєєва [3]. Огляд методів параметричної та структурної ідентифікації математичних моделей наведений у монографії М.П. Дивак [8]. Огляд публікацій демонструє актуальність дослідження методів та алгоритмів класифікації станів модельованої економічної системи.

Формулювання цілей статті. Метою статті є дослідження алгоритмів класифікації станів модельованої економічної системи за вектором параметрів.

Виклад основного матеріалу. Основна перевага статистичних методів полягає у можливості дослідження ознак різної природи з різним механізмом впливу, оскільки ці методи оперують безрозмірними величинами – стохастичними характеристиками станів системи. Описати статистичний зв'язок означає визначити імовірнісну функцію розподілу випадкової величини на просторі елементарних подій – множини параметрів досліджуваної системи. Стан економічної системи – це сукупність параметрів стану, тобто характеристик системи на цьому етапі її функціонування. Множина S станів системи є множиною можливих n-реалізацій випадкового процесу $S=\{S_i\},(i=1\div n)$ на момент дослідження $t_0\in[0\div T]$. Моделювання стану передбачає формалізацію параметрів, першим кроком до якої є розбиття сукупності параметрів $P_j(S_i);(j=1\div m; m=l+p+k)$ на однотипні за вибраним критерієм групи [2, с. 970–972]:

– некеровані змінні $X(t_0)=\{x_j(t_0)\},(j=1\div l)$ – описові параметри, що є числовими характеристиками об'єктивних умов функціонування досліджуваної системи; компонентами $X(t_0)$ можуть бути кількісні, якісні та класифікаційні ознаки різної природи;

– умовно керовані параметри $Y(t_0)=\{y_j(t_0)\}$, $(j=1\div p)$, частину яких можна характеризувати як екзогенні чинники (характеристики стану природи); якість цієї інформації може варіюватись від повної визначеності, детермінованості до повної невизначеності;

– ендогенні параметри $C(t_0)=\{c_j(t_0)\}$, $(j=1\div k)$, характеристики внутрішньої природи досліджуваної системи.

Компонентами векторів $X(t_0)$, $Y(t_0)$, $C(t_0)$ є багатовимірні ознаки змішаної природи, такі як кількісні, якісні, описові, нормовані величини.

Розглянемо деяку модельовану стратегію поведінки досліджуваної складної системи, що може перебувати в одному з можливих станів S_j, що утворюють впорядковану множину S=$\{S_i\},(i=1\div n)$ [9]. Характеристичні параметри S_i стану системи представимо вектором випадкових величин $X=(x_i)$, $i=1\div n$. Вектор X можна інтерпретувати як формалізацію статистичного зв'язку між багатовимірними ознаками $Y=y_i$ та $C=c_i$, тобто допустимим є розбиття цього вектору на складові частини:

$$X=x_i=\begin{pmatrix} y_i \\ c_i \end{pmatrix},\ y_i=(y_{i1},y_{i2},\cdots,y_{ik}),\ C=(c_{i1},c_{i2},\cdots,c_{il}), \tag{1}$$

$$i=1\div n; k+l=n.$$

Тоді задачу прогнозування стану системи можна розділити на дві підзадачі з огляду на форму статистичного зв'язку між складовими частинами вектору X.

1) Якщо екзогенно-ендогенна структура вектору $X=(x_i)$ невідома, то постає задача прогнозування стану (визначення класу) системи [10, с. 83–85]. Задача класифікації розглядає впорядковану множину можливих станів S=$\{S_i\}$, $(i=1\div n)$ системи, кожен з яких визначається вектором параметрів $X=(x_i)$ та може належати до одного з визначених класів $K=\{\mathrm{k}_j\}$, $j=1\div N$. Теоретично множина $[X\times K]$ утворює простір $\Omega(n,q)$ випадкових подій з відомою щільністю імовірності:

$$p(x,k)=p(k)\cdot p\left(x/k\right), \tag{2}$$

де $p(k)\cdot$ – апріорні ймовірності класів; $p\left(x/k\right)$ – щільність розподілу класів (функція правдоподібності класів).

Необхідно побудувати алгоритм розподілу множини станів S=$\{S_i\}$ між класами з мінімальною похибкою класифікації. Алгоритм визначення стану складної системи побудований на застосуванні дискримінантного аналізу до множини класів $\{k_j\}$, кожен з яких характеризується множиною стохастичних величин $\xi_j=\zeta(x_i)$, що визначені однозначно або з точністю до значень параметрів x_i. В основі дискримінантного аналізу лежить припущення про те, що описи об'єктів кожного j-го класу є реалізацією багатовимірної випадкової величини ξ, розподіленої за нормальним законом з математичним сподіванням $M_j(\xi)$ та невиродженою матрицею коваріації:

$$Q_j=(cov)_{j-класу}=\frac{\sum_{i=1}^{n}\left(x_{ij}-M_j(\xi)\right)^T\cdot(x_{ij}-M_j(\xi)}{n-1}. \tag{3}$$

Тоді класифікатор може бути побудований за Байєсовим алгоритмом: за апріорними імовірностями $p(\mathrm{k}_j)$ класів, тобто станів системи, та умовною щільністю розподілу вектору параметрів $p(x,k)$. У припущенні рівно імовірних класів $p(\mathrm{k}_j)=\frac{1}{N}$, імовірність належності вектору параметрів x_i до *j*-класу буде такою:

$$p_j(x_i)==\frac{p(x_i/k_j)}{\sum_{j=1}^{N}p(x_i/k_j)}. \quad (4)$$

Очевидно, що найменша похибка класифікації відповідає найбільшому значенню $p(x_i/k_j)$. Тоді дискримінантну функцію d_j можна вибрати за таким правилом:

$$p(x_i/k_m)>p(x_i/k_n), m\neq n; m=1\div q; n=1\div q. \quad (5)$$

Для розбиття на класи довільного вектору параметрів $X=x_i$ обчислюються дискримінантні функції для усіх можливих пар $\textit{класів}(m\neq n)$:

$$d=\left[p(X/k_m)/p(X/k_n)\right], \quad (6)$$

де вектор X належить до того класу, для якого відношення умовних імовірностей буде найбільшим.

В разі рівності коваріаційних матриць двох класів min дискримінантна функція має такий вигляд:

$$d=X^T\cdot Q^{-1}\left[M_m(X)-M_n(X)\right], \quad (7)$$

де $M_m(X);M_n(X)$ – вектори математичних сподівань класів min; X^T –транспонована матриця параметрів; Q^{-1} – обернена коваріаційна матриця.

Якщо коваріаційні матриці класів різні, використовується така форма дискримінантної функції:

$$d=\left(X-M_m(X)\right)^T\cdot Q_m^{-1}[X-M_m(X)]-\left(X-M_n(X)\right)^T\cdot Q_n^{-1}\left[X-M_n(X)\right]. \quad (8)$$

В практичних обчисленнях похибка класифікації інтерпретується як середня похибка класифікації, яку можна представити у вигляді матриці втрат (штрафів, ризиків) через неправильну класифікацію стану системи. Залежно від поставленої задачі рівень втрат від помилок класифікації може бути різним. Середнє значення помилки класифікації можна визначити за Байєсовим вирішуючим правилом: відомі статистичні характеристики системи $p\left(x/k\right);p(k)\cdot$ забезпечують створення алгоритму класифікації з мінімальним середнім ризиком $\bar{R}_{min}$.

$$\bar{R}_{\min}=\underset{\bar{k}\in K}{\arg\min}\sum_{k\in K}R_{k\bar{k}}\,p\left(x/k\right)p(k), \quad (9)$$

де $R_{k\bar{k}}$ – матриця розмірністю N×N – величини ризику для кожної пари класів $(k;\bar{k})\in K$ при віднесенні стану S=$\{S_i\}$ *k*-класу до $\bar{k}$.

Найчастіше використовують платіжну матрицю стандартного виду:

$$R_{k\bar{k}}=\begin{cases}0, k=\bar{k}\\ >0, k\neq\bar{k}\end{cases}. \quad (10)$$

У разі частково або повністю невідомих параметричних характеристик *i*-го стану системи використовують непараметричні методи оцінювання класів системи, що ґрунтуються на визначенні щільності розподілу класів $p\left(x/k\right)$ щодо вектору X, але для класифікації станів використовується формула (9). Важливо провести межу між станами системи, яку називають порогом. На практиці її часто визначають інтуїтивно.

1) Якщо (1) є поділом на вектори екзогенних та ендогенних змінних, то можна побудувати математичну модель статистичного зв'язку між векторами – компонентами вектору $X=x_i$:

$$\varepsilon_i=y_i-f(c_i), \quad (11)$$

де $f(c_i)$ – невідома неперервно-диференційована векторна функція невипадкових аргументів; $\varepsilon_i=(\varepsilon_{i1},\varepsilon_{i2},\cdots,\varepsilon_{in})$ – випадковий вектор помилок з нульовим математичним сподіванням та невиродженою коваріаційною матрицею $(\sigma_{ij}) = cov(\varepsilon_i, \varepsilon_j), i; j = n$. Випадкові вектори c_i та ε_i вважають статистично незалежними з відомими щільностями розподілу $p_i(c), p_i(\varepsilon)$.

Рівняння (11) є рівнянням багатофакторної нелінійної регресії. Загалом функціональний вид регресії не відомий, і для аналізу моделі (11) використовуються непараметричні методи [10, с. 81]. Якщо структура вектору $X = x_i = \begin{pmatrix} y_i \\ c_i \end{pmatrix}$ відома, а стан системи класифікований за екзогенними ознаками c_i, постає питання оцінювання вектору ендогенних змінних $y = y_i$ за відомим (невипадковим) вектором екзогенних змінних $c = c_i$. Щільність розподілу вектору параметрів $p(X)$ можна оцінити як щільність сумісного розподілу випадкових векторів $p\left({}^{y}\!/_{c} \right)$, тобто трактувати як апостеріорну імовірність, обчислену за формулою Байєса. Тоді оцінка вектору ендогенних змінних $\bar{y}$ обчислюється як максимальна щільність вектору сумісного розподілу складових частин вектору параметрів X, тобто як мода апостеріорного розподілу ендогенних змінних $M(y)$:

$$\bar{y} = \mathrm{M}(y) = \arg\max_{\bar{y} \in y} \bar{p}\left({}^{y}\!/_{c} \right) = \arg\max_{\bar{y} \in y} \frac{\bar{p}\left({}^{y}\!/_{c} \right)}{\bar{p}(c)}, \quad (12)$$

де $\bar{p}\left({}^{y}\!/_{c} \right)$ – оцінка щільності апостеріорного розподілу випадкових векторів *y* та *c*; $\bar{p}(c)$ – оцінка щільності розподілу вектору екзогенних змінних.

Висновки. Описаний загальний алгоритм оцінювання параметрів моделі складної системи застосовується для ідентифікації параметрів моделей систем управління в умовах часткової параметричної невизначеності. Алгоритм класифікації передбачає визначення оцінок апріорних імовірностей належності досліджуваної системи до кожного класу. Отримані оцінки використовуються в алгоритмах оптимізації прийняття рішення в умовах потенційних економічних ризиків [9]. Емпірична функція розподілу апріорних імовірностей складових частин вектору параметрів визначається на основі статистичних спостережень, експертних оцінок або як евристична функція розподілу. Класифікація параметрів залежить від цілей управління та власного досвіду. Байєсове вирішуюче правило є не новим алгоритмом оптимальної класифікації, але на його основі будуються нові алгоритми класифікації та прогнозування оцінок параметрів моделі.

Список використаних джерел:

1. Геселева Н.В., Заріцька Н.М. Емерджентні властивості системи. *Бізнес-Інформ*. 2013. № 7. С. 93–97.

2. Айвазян С.А. Многомерный статистический анализ в социально-экономических исследованиях. *Экономика и математические методы*. 1977. Т. XIII. Вып. 5. С. 968–983. URL: http://www.cemi.rssi.ru/emm/files/1977-05-Aivazian_SA.pdf (дата звернення: 08.01.2022).

3. Зотеев В.Е. Численный метод нелинейного оценивания на основе разностных уравнений. *Вестник СГТУ. Серия: Физико-математические науки*. 2018. Т. 22. № 4. С. 669–701. URL: http://www.mathnet.ru/links/9bac681bdd2fcc9fd05ddf7d0aed7cd9/vsgtu1643.pdf (дата звернення: 08.01.2022).

4. Жлуктенко В.І., Бєгун А.В. Стохастичні моделі в економіці : монографія. Київ : КНЕУ, 2005. 352 с.

5. Верченко П.І. Багатокритеріальність і динаміка економічного ризику (моделі та методи) : монографія. Київ : КНЕУ, 2006. 272 с.

6. Спільник І.В., Ярощук О.В. Принцип системності в аналітичних дослідженнях. *Економічний аналіз*. 2018. Т. 28. № 2. С. 182–190.

7. Айвазян С.А., Березняцкий А.Н., Бродский Б.Е., Дарховский Б.С. Статистический анализ моделей с переменной структурой. *Прикладная эконометрика. Теория и методология*. 2015. № 39(3). С. 84–105. URL: https://cyberleninka.ru/article/n/statisticheskiy-analiz-modeley-s-peremennoy-strukturoy/viewer (дата звернення: 08.01.2022).

8. Дивак М.П., Порплиця Н.П., Дивак Т.М. Ідентифікація дискретних моделей динамічних систем з інтервальними даними : монографія. Тернопіль : ВПЦ «Економічна думка ТНЕУ», 2018. 220 с.

9. Debela I.M. Research of optimization management models in conditions of uncertainty and risks. *Global aspects of national economy development in the conditions of transformational changes*. Lviv ; Toruń : Liha-Pres, 2021. P. 115–127.

10. Милюгин В.И., Васильков М.Е. Непараметрический анализ стохастических систем с нелинейной функциональной неоднородностью. *Прикладная эконометрика. Теория и методология*. 2011. № 2(22). С. 78–92. URL: http://pe.cemi.rssi.ru/pe_2011_2_78-92.pdf (дата звернення: 13.01.2022).

11. Дебела І.М. Байєсовський метод оцінки альтернативних рішень. *Таврійський науковий вісник. Серія: Економіка*. 2021. № 8. С. 76–81.

References:

1. Geseleva N.V., Zaritskaya N.M. (2013) Emerdzhentni vlastyvosti systemy [Emergent properties of the system]. *Business information*, no. 7, pp. 93–97.

2. Aivazian S.A. (1977) Mnogomernyy statisticheskiy analiz v sotsialno-ekonomicheskikh issledovaniyakh [Multivariate statistical analysis in socio-economic research]. *Economics and mathematical methods,* vol. XIII, no. 5, pp. 968–983. Retrieved from: http://www.cemi.rssi.ru/emm/files/1977-05-Aivazian_SA.pdf (accessed 08 January 2022).

3. Zoteev V.E. (2018) Chislennyy metod nelineynogo otsenivaniya na osnove raznostnykh uravneniy [Numerical Method for Nonlinear Estimation Based on Difference Equations]. *Bulletin of SSTU. Series: Phys.-Math. Sciences*, vol. 22, no. 4, pp. 669–701. Retrieved from: http://www.mathnet.ru/links/9bac681bdd2fcc9fd05ddf7d0aed7cd9/vsgtu1643.pdf (accessed 08 January 2022).

4. Zhluktenko V.I., Begun A.V. (2005) Stokhastychni modeli v ekonomici [Stochastic models in economics]. Kyiv: KNEU. (in Ukrainian)

5. Verchenko P.I. (2006) Baghatokryterialjnistj i dynamika ekonomichnogho ryzyku (modeli ta metody) [Multicriteria and dynamics of economic risk (models and methods)]. Kyiv: KNEU. (in Ukrainian)

6. Spiljnyk I.V., Jaroshhuk O.V. (2018) Pryncyp systemnosti v analitychnykh doslidzhennjakh [The principle of systematicity in analytical research]. *Economic analysis*, vol. 28, no. 2, pp. 182–190.

7. Aivazian S.A., Bereznyatskiy A.N., Brodskiy B.E., Darkhovskiy B.S. (2015) Statisticheskiy analiz modeley s peremennoy strukturoy [Statistical analysis of models with variable structure]. *Applied Econometrics. Theory and methodology,* no. 39(3), pp. 84–105. Retrieved from: https://cyberleninka.ru/article/n/statisticheskiy-analiz-modeley-s-peremennoy-strukturoy/viewer (accessed 08 January 2022).

8. Dyvak M.P., Porplycja N.P., Dyvak T.M. (2018) Identyfikacija dyskretnykh modelej dynamichnykh system z intervaljnymy danymy [Identification of discrete models of dynamic systems with interval data]. Ternopil: Economic opinion of TNEU. (in Ukrainian)

9. Debela I.M. (2021) Doslidzhennja optymizacijnykh modelej upravlinnja v umovakh nevyznachenosti ta ryzykiv [Research of optimization management models in conditions of uncertainty and risks]. Ghlobaljni aspekty rozvytku nacionaljnoji ekonomiky v umovakh transformacijnykh zmin [Global aspets of national economy development in the conditions of transformational changes]. Lviv; Toruń: Liha-Pres, pp. 115–127.

10. Milyugin V.I., Vasilkov M.E. (2011) Neparametricheskiy analiz stokhasticheskikh sistem s nelineynoy funktsionalnoy neodnorodnostyu [Nonparametric Analysis of Stochastic Systems with Nonlinear Functional Heterogeneity]. *Applied Econometrics. Theory and methodology*, no. 2(22), pp. 78–92. Retrieved from: http://pe.cemi.rssi.ru/pe_2011_2_78-92.pdf (accessed 13 January 2022).

11. Debela I.M. (2021) Bajjesovsjkyj metod ocinky aljternatyvnykh rishenj [Bayesian method of evaluating alternative solutions]. *Taurian Scientific Bulletin. Series: Economics*, no. 8, pp. 76–81.