УДК 681. 513

НЕЛІНІЙНА НОРМАЛІЗАЦІЯ СТАТИСТИЧНИХ ПОКАЗНИКІВ ДЛЯ ЗАДАЧІ ПОБУДОВИ ІНТЕГРАЛЬНИХ ІНДЕКСІВ

Р.В. Волощук, В.С. Степашко

Міжнародний науково-навчальний центр інформаційних технологій та систем НАН та МОН України, 03680, Київ, просп. Академіка Глушкова, 40

volrom@bigmir.net; *stepashko@irtc.org.ua*

В статті запропоновано новий метод нормалізації статистичних показників, що характеризують складні соціально-економічні процеси, для задач інтегрального оцінювання стану економічної безпеки.

Ключові слова: *складний процес, статистичний показник, нормалізація, інтегральний індекс, економічна безпека*.

В статье предложен новий метод нормализации статистических показателей, характеризующих сложные социально-экономические процессы, для задач интегрального оценивания состояния экономической безопасности.

Ключевые слова: *сложный процесс, статистический показатель, нормализация, интегральный индекс, экономическая безопасность*.

A new method for normalization of statistical indicators characterizing complex social and economic processes is proposed in the paper for tasks of integral evaluation of the economic safety status.

Keywords: *complex process, statistical indicator, normalization, integral index, economic safety*.

Вступ. Потреба в конструюванні узагальненого (інтегрального, агрегованого) показника, або інтегрального індексу, для групи взаємозв'язаних первинних показників, які в сукупності характеризують деяку систему (процес або явище), часто постає в задачах комплексного аналізу стану складних систем – наприклад, у соціології, а також у задачах оцінювання економічної або екологічної безпеки тощо.

Значення первинних показників отримуються різними методами збирання статистичних даних. Але для аналізу та загальної оцінки складної системи як єдиного цілого важливо мати один агрегований показник – інтегральний індекс, який безпосередньо не вимірюється, але певним чином інтегрує первині показники і обчислюється на основі їхніх статистичних значень [1]. При цьому статистичні показники мають бути певним чином нормалізовані – найчастіше приведенням до шкали [0; 1].

Виконаний аналіз наявних способів нормування [2] показав, що в задачі кількісного оцінювання стану економічної безпеки [3] найбільш адекватним є метод нелінійної (точніше, кусково-лінійної) нормалізації [4]. Проте він має певні недоліки, які обговорюються в даній статті. Для подолання цих недоліків пропонується новий метод нелінійної нормалізації.

Постановка задачі. Оцінюється загальний стан заданої складної системи (багатовимірного процесу). Нехай $x_i, i=\overline{1,m}$, – первинні взаємозв'язані показники функціонування цієї системи, які в сукупності характеризують її стан. Для однозначної інтерпретації, оцінювання та порівняння між собою кожен з показників x_i має бути нормованим, зазвичай приведеним до інтервалу $0 \le \tilde{x}_i \le 1$, де $\tilde{x}_i$ – нормоване значення показника, причому $\tilde{x}_i = 1$ має відповідати найкращим (оптимальним) значенням цього показника, а $\tilde{x}_i = 0$ – найгіршим (неприпустимим) його значенням. Це значення $\tilde{x}_i$ будемо називати індикатором первинного показника x_i. Після нормалізації значень кожного з заданої системи первинних показників інтегральний індекс цієї системи обчислюється як сума цих нормалізованих величин з певними ваговими коефіцієнтами (однаковими або ні).

Якщо $\tilde{x}_i$ – значення i-го нормалізованого показника, то інтегральний показник (індекс) може мати вигляд лінійної згортки m показників:

$$I = \sum_{i=1}^{m} k_i \tilde{x}_i , \tag{1}$$

де k_i – ваговий коефіцієнт i-го показника.

Інтегральний індекс як синтетичний показник має певні переваги: комплексність отриманої інформації, швидкість, простота використання; може служити інструментом обліку, аналізу та планування; індикатор стану і критерій порівняльного оцінювання стану складної системи; показник ефективності прийнятих рішень та повноти їх реалізації тощо [5].

Важливим етапом при розрахунку інтегрального індексу є визначення вагових коефіцієнтів. Найбільш поширеними серед підходів до визначення вагових коефіцієнтів є *експертні* методи встановлення ваг показників. Їх перевага – простота застосування, їх недолік – можлива недостовірність даних, певна суб'єктивність, трудомісткість збору та обробки результатів, проте експертні методи мають значну суб'єктивну складову.

Серед наявних *розрахункових* підходів до визначення вагових коефіцієнтів можна виділити метод парної кореляції та метод, що базується на використанні показника варіації [5]. Проте найбільш формалізованим і ефективним, як показано в [6], слід вважати підхід до визначення вагових коефіцієнтів на основі методу головних компонент (МГК).

Кусково-лінійна нормалізація статистичних показників. Наступним етапом при побудові інтегрального індексу є нормування показників. При цьому слід враховувати, що будь-який статистичний показник може бути як *стимулятором* – коли його зростання сприяє покращенню соціально-економічного стану держави, галузі або регіону, так і *дестимулятором* – коли його зростання призводить до погіршення стану. Більше того, виявляється, що значна частина статистичних показників можуть бути в одному діапазоні своїх значень стимуляторами, а в іншому – дестимуляторами.

Серед різних підходів до нормування [2] вирізняється нормалізація на основі врахування вектора характеристичних значень показників [4]. Цей підхід є ефективним варіантом нелінійної нормалізації (на відміну від стандартного нормування) статистичних значень первинних показників з урахуванням їхніх так званих *характеристичних значень*, які вказують діапазони оптимальних, припустимих і критичних величин показників.

Для кожного первинного показника-стимулятора x_j задано межі так званих *оптимальних* $x_{\text{опт}}$ (вище яких значення показника вважаються найкращими), *порогових* $x_{\text{пор}}$ (вище яких значення показника вважаються припустимими) та *граничних* $x_{\text{гр}}$ (нижче яких значення показника вважаються припустимими) характеристичних значень, причому $x_{\text{гр}} < x_{\text{пор}} < x_{\text{опт}}$. Відповідні значення задаються також для показників-дестимуляторів, для яких $x_{\text{опт}} < x_{\text{пор}} < x_{\text{гр}}$.

Якщо ж певний статистичний показник може бути в одному діапазоні своїх значень стимулятором, а в іншому – дестимулятором, тоді задається 6 характеристичних значень: двосторонні *нижні* та *верхні* оптимальні $x^{\text{н}}_{\text{опт}}, x^{\text{в}}_{\text{опт}}$ (що відповідають найкращим умовам для економічної системи), порогові $x^{\text{н}}_{\text{пор}}, x^{\text{в}}_{\text{пор}}$ (які бажано не перетинати) та граничні $x^{\text{н}}_{\text{гр}}, x^{\text{в}}_{\text{гр}}$ (перетинати які *неприпустимо* або фізично неможливо) значення. Відповідне розміщення характеристичних величин на осі значень показника показано на рис. 1.

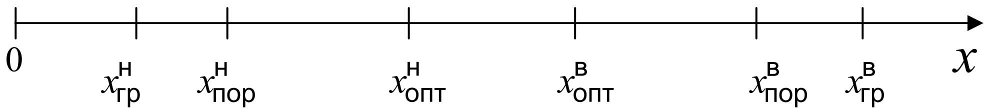

Рис 1. Характеристичні величини

У разі первинного показника-стимулятора x_j формула нормалізації в діапазоні [$x_{\text{гр}}$, $x_{\text{опт}}$] має вигляд [1]:

$$\tilde{x}_i = \min\left[0.5 \max\left(0, (x_i - x^{\text{н}}_{\text{гр}}) \big/ (x^{\text{н}}_{\text{пор}} - x^{\text{н}}_{\text{гр}})\right)\right]. \tag{2}$$

У разі первинного показника-дестимулятора x_i відповідна формула нормалізації в діапазоні [$x_{\text{опт}}$, $x_{\text{гр}}$] є такою:

$$\tilde{x}_j = \min\left[0.5 \max\left(0, (x^{\text{в}}_{\text{гр}} - x_j) \big/ (x^{\text{в}}_{\text{гр}} - x^{\text{в}}_{\text{пор}})\right)\right] \tag{3}$$

Зрозуміло, що у разі двосторонніх порогових і граничних обмежень на первинний показник x_j для нормалізації застосовуються формули (2) та (3). Відповідний нормувальний графік з урахуванням усіх характеристичних величин показано на рис. 1. При цьому для прикладу припускається, що пороговому значенню має відповідати нормалізована величина на рівні 0.5.

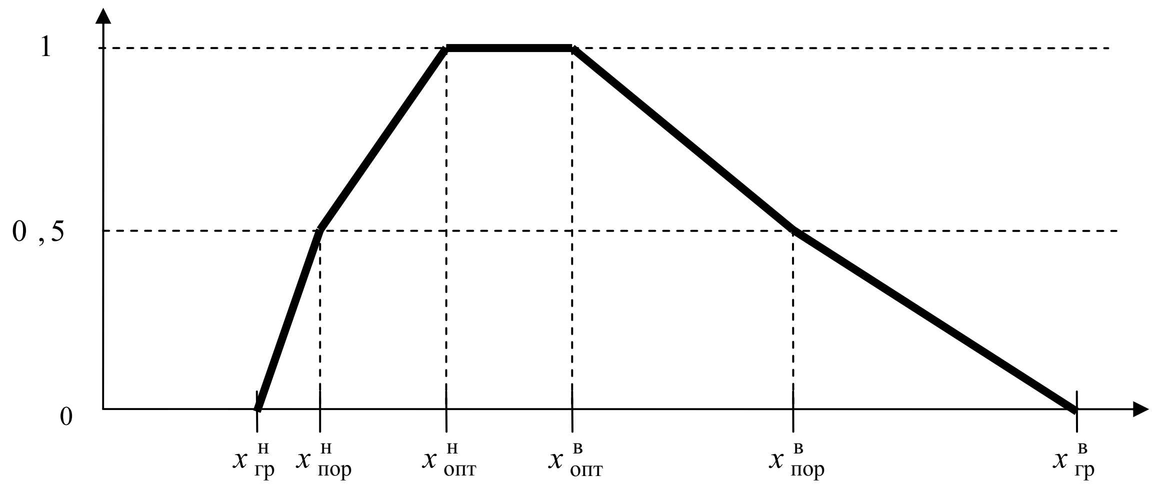

Рис.2 Кусково-лінійна нормувальна функція для статистичних показників

Проблемна особливість кусково-лінійної нормалізації. Очевидним недоліком такого підходу є передусім те, що в характеристичних точках (оптимальних, порогових та граничних) нормувальна функція негладка, вона має розриви похідної. Це означає, що в цих точках приріст величини «штрафу» на значення показника міняється стрибком.

З цього постають очевидні вимоги до потрібних властивостей функції гладкої нелінійної нормалізації:

1. При незначному відхиленні від оптимальної характеристичної точки нормалізоване значення має мало відрізнятись від 1. Це означає, що перша похідна від потрібної нормувальної функції в цій точці має дорівнювати 0.

2. В пороговій характеристичній точці функція має бути гладкою і мати перегин. Це означає, що друга похідна від потрібної функції в цій точці має дорівнювати 0.

3. При незначному відхиленні від граничної характеристичної точки нормалізоване значення має мало відрізнятись від 0. Це означає, що перша похідна від потрібної нормувальної функції в цій точці має дорівнювати 0.

Розглянемо задачу побудови функції нелінійної нормалізації, яка б відповідала поданим вище вимогам. Спершу для визначеності проаналізуємо задачу нормалізації для показника-*стимулятора*, для якого задані такі нижні характеристичні значення: граничне $x^{н}_{гр}$, порогове $x^{н}_{пор}$ та оптимальне $x^{н}_{опт}$, причому $0 \le x^{н}_{гр} < x^{н}_{пор} < x^{н}_{опт}$.

Пропоноване вирішення проблеми. Для подолання проблеми негладкості вказаної вище кусково-лінійної нормувальної функції пропонується проаналізувати неперервну функцію, що має такий вигляд:

$$y = ax^{b}e^{-cx}, \quad a, c > 0, b > 1 . \tag{4}$$

Її графік подано на рис. 4.

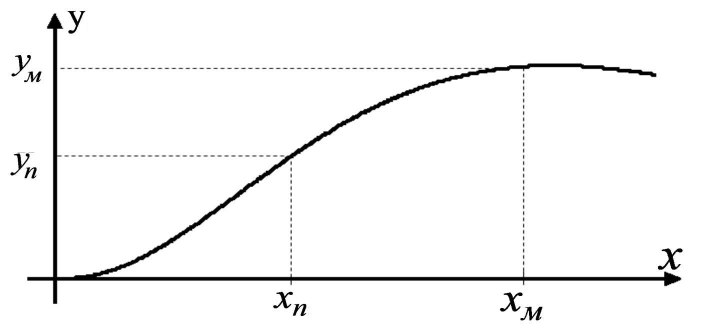

Рис.3. Графік аналітичної функції (4)

Функцію (4), яка, як видно з рис. 3, у проміжку від 0 до $x_м$ загалом відповідає сформульованим вище вимогам, можна легко пристосувати до задачі нормалізації, перенісши початок координат з точки 0 в точку $x^н_{гр}$ (зсув аргумента). При цьому отримується така нелінійна нормувальна функція:

$$y(x) = a(x - x^н_{гр})^b e^{-c(x - x^н_{гр})}, \quad a,\ c > 0,\ b > 1. \tag{5}$$

. Тепер задача полягає в тому, щоб виразити коефіцієнти *a, b, c* через задані «нижні» характеристичні значення.

Визначення параметрів функції (5). Для спрощення аналізу зробимо в (5) заміну змінної: $z = x - x^н_{гр}$, тоді (5) набуває вигляду (4):

$$y(z) = az^b e^{-cz},\ a,\ c > 0,\ b > 1. \tag{6}$$

Очевидно, що тут $y(0) = 0$. Визначимо для функції (7) спочатку точку максимуму $z_м$, а далі точку перегину $z_п$. З умови для першої похідної отримуємо рівняння $bz^{-1} - c = 0$, звідки координата *максимуму*:

$$z_м = \frac{b}{c}\ . \tag{7}$$

Беремо другу похідну і прирівнюємо її до 0:

$$bz^{-1}\left[(b-1)z^{-1} - c\right] - c\left(bz^{-1} - c\right) = 0, \tag{8}$$

звідки отримуємо квадратне рівняння для точки перегину:

$$c^2 z^2 - 2bcz + b(b-1) = 0. \tag{9}$$

Враховуючи те, що нас цікавить інтервал від нуля до точки максимуму, з рівняння (10) маємо, що *перегин* функції (7) знаходиться в точці

$$z_п = \frac{b - \sqrt{b}}{c}. \tag{10}$$

Тепер установимо, як параметри *a, b, c* функцій (5) або (7) виражаються через характеристичні величини статистичного показника-стимулятора $x^н_{гр}$, $x^н_{пор}$ та $x^н_{опт}$. Враховуючи заміну (6), введемо позначення для еквівалентних «характеристичних величин» змінної *z*:

– оптимальне значення: $z_с = x^н_{опт} - x^н_{гр}$,

– порогове значення: $z_{\text{р}} = x^{\text{н}}_{\text{пор}} - x^{\text{н}}_{\text{гр}}$,

– граничне значення: $z_{\text{g}} = x^{\text{н}}_{\text{гр}} - x^{\text{н}}_{\text{гр}} = 0$.

Очевидно, що максимум (7) відповідає оптимальному значенню $z_{\text{м}} = z_{\text{с}}$, а точка перегину – пороговому значенню $z_{\text{п}} = z_{\text{р}}$. Тоді (7) і (10) складають систему двох рівнянь для визначення двох невідомих *b, c* через еквівалентні оптимальне і порогове значення:

$$\begin{cases} \dfrac{b}{c} = z_c \\ \dfrac{b - \sqrt{b}}{c} = z_p \end{cases} \tag{11}$$

Поділивши друге рівняння на перше і розв’язавши отримане нелінійне рівняння, маємо значення параметра *b*:

$$b = \frac{(z_{\text{с}})^2}{(z_{\text{с}} - z_{\text{р}})^2}. \tag{12}$$

Тоді з першого рівняння системи маємо значення параметра *c*:

$$c = \frac{z_{\text{с}}}{(z_{\text{с}} - z_{\text{р}})^2}. \tag{13}$$

Щоб (6) була нормувальною функцією, потрібно, щоб у точці максимуму z_c (або вона ж $z_м$) досліджувана крива дорівнювала 1. Тобто для визначення нормувального коефіцієнта *a* треба в формулу (6) підставити z_c замість z і прирівняти *y* до 1, тоді з отриманого рівняння маємо:

$$a = \frac{1}{(z_c)^b e^{-c(z_c)}}. \tag{14}$$

Замість *b* і *c* можна підставити вирази (12, 13), але очевидно, що *a* простіше обчислюється через уже відомі *b* і *c* (оскільки $z_м = z_c = \dfrac{b}{c}$):

$$a = \frac{1}{\left(\dfrac{b}{c}\right)^b e^{-b}} \tag{15}$$

Отже, тепер визначено всі три параметри *a, b, c,* і функція (6) відносно змінної z має всі потрібні нам властивості. Якщо тепер у формули (12, 13) підставити вирази для z_c, z_p, z_g, отримаємо, що параметри *b, c* і *a* виражаються через усі три «нижні» характеристичні величини:

$$b = \frac{(z_{\text{м}})^2}{(z_{\text{м}} - z_{\text{п}})^2} = \frac{(x^{\text{н}}_{\text{опт}} - x^{\text{н}}_{\text{гр}})^2}{(x^{\text{н}}_{\text{опт}} - x^{\text{н}}_{\text{пор}})^2}, \tag{16}$$

$$c = \frac{z_{\text{м}}}{(z_{\text{м}} - z_{\text{п}})^2} = \frac{x^{\text{н}}_{\text{опт}} - x^{\text{н}}_{\text{гр}}}{(x^{\text{н}}_{\text{опт}} - x^{\text{н}}_{\text{пор}})^2}, \tag{17}$$

$$a = \frac{1}{(z_c)^b e^{-cz_c}} = \frac{1}{\left(\frac{b}{c}\right)^b e^{-b}} = \frac{1}{\left(x^{\text{н}}_{\text{опт}} - x^{\text{н}}_{\text{гр}}\right)^b e^{-b}}. \qquad (18)$$

Отже, у випадку стимулятора нормувальна функція має вигляд:

$$\widetilde{x}_s(x) = a(x - x^{\text{н}}_{\text{гр}})^b e^{-c(x - x^{\text{н}}_{\text{гр}})}, \qquad (19)$$

де a, b, і c визначаються з (16) – (18), причому видно, що a, $c > 0$, $b > 1$.

Задача нормалізації для показника-дестимулятора. Для такого показника нормувальна функція має бути дзеркальним відображенням до функції стимулятора, причому починатись вона має з точки $x^{\text{в}}_{\text{опт}}$ (рис. 2):

$$\widetilde{x}_d = 1 - a(x - x^{\text{в}}_{\text{опт}})^b e^{-c(x - x^{\text{в}}_{\text{опт}})}. \qquad (20)$$

Тут параметри a, b і c, отримані за такою ж послідовністю дій, виражаються через усі три відповідно «верхні» характеристичні величини:

$$b = \frac{(x^{\text{в}}_{\text{гр}} - x^{\text{в}}_{\text{опт}})^2}{(x^{\text{в}}_{\text{гр}} - x^{\text{в}}_{\text{пр}})^2}, \qquad (21)$$

$$c = \frac{x^{\text{в}}_{\text{гр}} - x^{\text{в}}_{\text{опт}}}{(x^{\text{в}}_{\text{гр}} - x^{\text{в}}_{\text{пр}})^2}, \qquad (22)$$

$$a = \frac{1}{\left(x^{\text{в}}_{\text{гр}} - x^{\text{в}}_{\text{опт}}\right)^b e^{-b}}. \qquad (23)$$

Формування загальної нормувальної функції. Тоді загальна нормувальна функція для стимулятора-дестимулятора матиме вигляд:

$$\widetilde{x} = \begin{cases} \widetilde{x}_s, \ x \in [x^{\text{н}}_{\text{гр}}, x^{\text{н}}_{\text{опт}}) \\ 1, \ \ x \in [x^{\text{н}}_{\text{опт}}, x^{\text{в}}_{\text{опт}}] \\ \widetilde{x}_d, \ x \in (x^{\text{в}}_{\text{опт}}, x^{\text{в}}_{\text{гр}}] \end{cases} \qquad (24)$$

Відповідний графік подано на рис. 4.

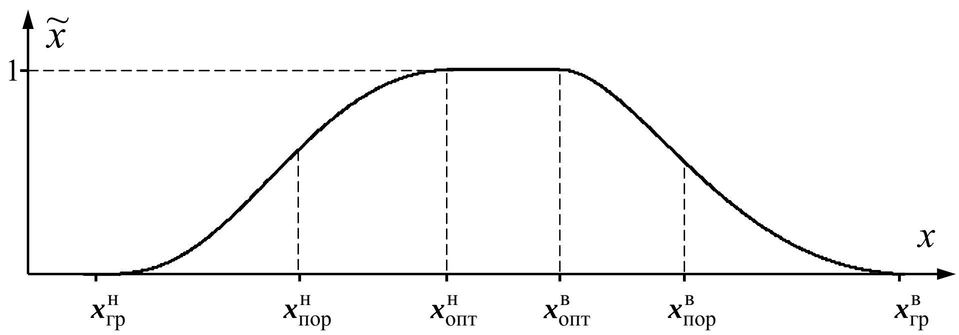

Рис. 4 Функція нелінійної нормалізації

Висновки. В роботі розглянуто важливий аспект задачі кількісного інтегрального оцінювання стану складних систем на прикладі економічної безпеки держави, пов'язаний з нормуванням статистичних показників.

Визначено основний недолік кусково-лінійної нормалізації показників, застосовуваної у методиці розрахунку інтегрального індексу економічної безпеки, пов'язаний з тим, що в характеристичних точках (оптимальних, порогових та граничних) вона негладка, тобто має розриви першої похідної.

Запропоновано нову гладку функцію нелінійної нормалізації, яка має нульову першу похідну в оптимальній і граничній характеристичних точках, а також перегин у пороговій характеристичній точці. Аналітично отримано формули, що виражають коефіцієнти введеної нормувальної функції через задані характеристичні значення статистичних показників як для випадків, коли конкретний показник є стимулятором або дестимулятором, так і для комбінованого варіанту стимулятора-дестимулятора.

Література

1. Степашко В.С., Мельник І.М., Волощук Р.В. Моделі синтезу інтегральної оцінки стану складної системи взаємозв'язаних первинних показників // Моделювання та керування станом еколого-економічних систем регіону. Зб. наук. праць. – Вип. 3. – К.: Міжнар. науково-навчальний центр інформаційних технологій та систем НАН та МОН України, 2006. – С. 57-62.

2. Волощук Р.В. Підходи до нормування економічних показників. Індуктивне моделювання складних систем: Зб. наук. пр. — К.: МННЦ ІТС НАН та МОН України, 2009. — Вип. 1. — С. 17-25.

3. Степашко В.С., Мельник И.М., Кваша Т.К. Определение интегральных индексов для оценивания экономической безопасности государства // Экономическая безопасность государства и информационно-технологические аспекты ее обеспечения / Под общей ред. Г.К.Вороновского, И.В.Недина. – К.: Знания Украины, 2006. – 664 с. / – С. 472-482.

4. Степашко В.С. Про задачу нормалізації економічних показників // "Економіко-математичне моделювання соціально-економічних систем". Збірник наукових праць. – Вип. 9. – Київ: Міжнар. науково-навч. центр інформаційних технологій і систем НАН та МОН України, 2005. – С. 32-36.

5. Райская Н.Н., Сергиенко Я.В., Френкель А.А. Использование интегральных индексов в анализе циклических изменений российской экономики // Под ред.: Е. Г. Ясин, XI международная научная конференция по проблемам развития экономики и общества: В 3 кн. – М.: Издательский дом НИУ ВШЭ, 2011. – Кн. 1., С. 120-130.

6. Волощук Р.В. Порівняльний аналіз підходів до визначення вагових коефіцієнтів інтегральних індексів стану складних систем // Індуктивне моделювання складних систем. Зб. наук. праць. — Вип. 5. — К.: МННЦ ІТС НАН та МОН України, 2013.— С. 151-165.