МЕТОДИ МАТЕМАТИЧНОГО МОДЕЛЮВАННЯ

Моделювання нелінійних послідовностей з заданими параметрами регресій

В. І. АВРАМЕНКО

Дніпродзержинський державний технічний університет

Пропонується метод моделювання псевдо-випадкових нелінійних послідовностей з заданими параметрами регресії і фіксованим коефіцієнтом детермінації рівняння регресії.

Предлагается метод моделирования псевдослучайных нелинейных последовательностей с заданными параметрами регрессии и фиксированным коэффициентом детерминации уравнения регрессии.

The method of design of pseudo casual nonlinear sequences is offered with preset the parameter of regressions and fixed coefficient of determination of regression equalization.

Мета. При використанні методів статистичного моделювання, зокрема при оцінці впливу похибок вихідних даних на похибки результатів моделювання, виникає необхідність вилучення систематичних похибок від неточного завдання рівняння регресії. Для лінійних чи лінеаризованих функцій однієї змінної відомі методи генерації послідовностей з заданими параметрами рівняння регресії [1, 2]. Нижче пропонується узагальнений алгоритм генерації псевдо-випадкових послідовностей довільної довжини, обробка яких за стандартною методикою дозволяє отримати рівняння регресії з заданими значеннями параметрів регресії і коефіцієнта детермінації R^2.

1. **Постановка задачі**. Як відомо, найбільш поширеним методом визначення параметрів залежностей за результатами спостережень є метод найменших квадратів (МНК), коли розв'язується система m нормальних рівнянь виду

$$\sum_{i=1}^{n}\left(\tilde{y}_i - \varphi(\grave{A}, x_i)\right)\cdot \varphi'_{aj}(\grave{A}, x_i) = 0, \quad j = 1, 2, ..., m$$

де i=1,2,…, n – об'єм вибіркової сукупності, A=(a_1, a_2, … a_m) – вектор невідомих параметрів регресії, $\varphi(A, x_i)$ – функція регресії заданого виду. Якщо статистичні спостереження $\tilde{y}_i$ не містять випадкових збурень, тобто $\tilde{y}_i=\varphi(A, x_i)$, то параметри регресії визначаються безпомилково.

При статистичному моделюванні нелінійних залежностей з заданими параметрами регресії приймемо, що $\tilde{y}_i = \varphi(A, x_i)+z_i$, де z_i – випадкова складова, розподілена за деяким законом розподілу. Тоді нормальні рівняння приймуть вид

$$\sum_{i=1}^{n}\left((\varphi(\grave{A}, \tilde{o}_{\text{з}}) + z_{\text{з}}) - \varphi(A, x_i)\right)\cdot \varphi'_{aj}(A, x_i) = 0, \quad j = 1, 2,...,m$$

Перепишемо кожне рівняння в такому вигляді

$$\sum_{\text{з}=1}^{\text{ї}}\varphi(\grave{A}, \tilde{o}_{\text{з}})\cdot \varphi'_{aj}(A, x_i) = \sum_{i=1}^{n}\varphi(\grave{A}, \tilde{o}_{\text{з}})\cdot \varphi'_{aj}(A, x_i) + \sum_{i=1}^{n} z_i \cdot \varphi'_{aj}(A, x_i)$$

Якщо другі доданки в правих частинах рівнянь тотожньо дорівнюють нулю, то параметри регресії A(a_1, a_2, … a_m) визначаються безпомилково. Отже, умовою правильного знаходження параметрів регресії методом найменших квадратів є задоволення системи рівнянь

$$\sum_{\text{з}=1}^{\text{ї}} z_i \cdot \varphi'_{aj}(A, x_i) \equiv 0, \; j = 1, 2,..., m \qquad (1)$$

Пропонується шукати випадкові збурення z_i у такому вигляді

$$z_i = k_0 \cdot \varepsilon_i = k_0 \cdot \frac{k_1 \cdot \alpha_i + k_2 \cdot \beta_i + k_3 \cdot \gamma_i + ...}{M_c} \qquad (2)$$

де ε_i – члени стандартизованої послідовності випадкових чисел, які самі є сумою кількох стандартизованих послідовностей α, β, γ, … , взятих з деякими наперед невідомими ваговими коефіцієнтами k_1, k_2, k_3,… Величина M_c є стандартизуючим множником послідовності ε_i і обчислюється за формулою $M_c = \vec{K}\cdot\Re\cdot\vec{K}^{\grave{O}}$, де $\vec{K}$ і $\vec{K}^{\grave{O}}$ – вектори відповідно рядок і стовпець вагових коефіцієнтів k_j, $\Re$ – матриця парних коефіцієнтів послідовностей α_i, β_i, γ_i, в окремому випадку може бути діагональною.

Тоді система рівнянь (1) перетворюється на систему однорідних лінійних відносно невідомих коефіцієнтів k_j рівнянь

$$\begin{cases}\sum_{\text{з}=1}^{\text{ї}}\left(k_1\cdot\alpha_i + k_2\cdot\beta_i + k_3\cdot\gamma_i + ...\right)\cdot\varphi'_{a1}(A, x_i) = 0\\ \sum_{\text{з}=1}^{\text{ї}}\left(k_1\cdot\alpha_i + k_2\cdot\beta_i + k_3\cdot\gamma_i + ...\right)\cdot\varphi'_{a2}(A, x_i) = 0\\ \dots\dots\dots\dots\dots\dots\dots\dots\dots\dots \\ \sum_{\text{з}=1}^{\text{ї}}\left(k_1\cdot\alpha_i + k_2\cdot\beta_i + k_3\cdot\gamma_i + ...\right)\cdot\varphi'_{am}(A, x_i) = 0\end{cases} \qquad (3)$$

Як відомо, для існування нетривіального розв'язку системи однорідних лінійних рівнянь кількість невідомих повинна бути більшою за кількість рівнянь. В розглядуваному випадку кількість доданків випадкової складової ε_i повинна бути більшою за кількість m параметрів регресії. Якщо використати кількість доданків m+1 і приймаючи за замовчуванням $k_1=const$, неважко знайти потрібні значення корегуючих множників k_j, j=2, 3, …m+1. Обробка МНК масиву вихідних даних $\{\tilde{y}_i, x_i\}$, де $\tilde{y}_i=\varphi(A, x_i)+k_0\cdot\varepsilon_i$ дозволяє безпомилково в межах похибки обчислень знайти параметри регресії $a_1, a_2, \ldots a_m$.

Коефіцієнт детермінації R^2 отриманої регресії визначається величиною множника k_0. Тому що послідовність чисел ε_i після обчислення коефіцієнтів k_i стандартизується множником M_c, то дисперсія послідовності z_i дорівнює k_0^2. Якщо параметри регресії визначаються безпомилково, дисперсія регресії після використання МНК дорівнює дисперсії $D(\varphi(A, x_i))$ модельної функції на заданому інтервалі. Допускаючи незалежність збурень z_i від функції $\varphi(A, x_i)$, дисперсія аналізованої послідовності $\tilde{y}_i=\varphi(A,x_i)+k_0\cdot\varepsilon_i$ дорівнює $D(\varphi(A, x_i)) + k_0^2$. Отже, коефіцієнт детермінації

$$R^2 = \frac{D(\varphi(A, x_i))}{D(\varphi(A, x_i)) + k_0^2}.$$

Щоб отримати реалізацію $\tilde{y}_i = \varphi(A, x_i) + k_0 \cdot \varepsilon_i$ з заданим коефіцієнтом детермінації R^2, величину k_0 слід вибирати за формулою

$$k_0 = \sqrt{D(\varphi(A, x_i)) \cdot \frac{1-R^2}{R^2}} \qquad (4)$$

Розв'язок задачі спрощується, якщо члени послідовностей α_i, β_i, γ_i … будуть незалежними. Розглянемо алгоритм моделювання лінійно незалежних послідовностей.

Статистичне моделювання полягає в використанні системи $\Lambda(\lambda, \mu, \nu, \varsigma, \ldots)$ послідовностей m випадкових величин, отриманих з використанням деяких генераторів випадкових чисел. Простим лінійним перетворенням кожну з послідовностей можна звести до стандартизованого виду з числовими характеристиками $\overline{\lambda} = \overline{\mu} = \ldots = 0$, $S_\lambda = S_\mu = \ldots = 1$, де $\overline{\lambda}, \overline{\mu}$, …– середні значення, S_λ, S_μ …– величини середніх квадратичних відхилень. Як наслідок, для кожної з послідовностей слушні властивості $\sum_{i=1}^{n}\lambda_i = \sum_{i=1}^{n}\mu_i = \ldots = 0$, де n – довжина кожної реалізації і $\sum_{i=1}^{n}\lambda_i^2 = \sum_{i=1}^{n}\mu_i^2 = \ldots = n-1$.

Незалежно від джерела випадкових чисел, послідовності λ_i, μ_i … не можна вважати незалежними, особливо для коротких реалізацій. Мірою лінійної незалежності між двома послідовностями з системи є парний коефіцієнт кореляції. Для стандартизованих послідовностей слушні властивості

$$\sum_{i=1}^{n}\lambda_i \cdot \mu_i = r_{\lambda\mu}\cdot(n-1); \quad \sum_{i=1}^{n}\lambda_i \cdot \nu_i = r_{\lambda\nu}\cdot(n-1) \quad \ldots$$

Парні коефіцієнти кореляції між послідовностями системи $\Lambda(\lambda, \mu, \nu, \varsigma, \ldots)$ утворюють симетричну кореляційну матрицю $\Re$. Для лінійно незалежних послідовностей кореляційна матриця стає одиничною, отже слід відшукати таке перетворення матриці системи $\Lambda(\lambda, \mu, \nu, \varsigma, \ldots)$, яке зведе кореляційну матрицю до одиничної. Добуток матриці вихідних даних на матрицю, обернену до кореляційної, дозволяє отримати нові послідовності, які попарно незалежні, але їх кореляційна матриця не наближається до одиничної.

Для розв'язання задачі пропонується використовувати матрицю перетворення трикутного виду, елементи якої q_{ij} шукаються з системи рівнянь

$$\begin{cases} \alpha_i = \lambda_i \\ \beta_i = q_{21}\cdot\lambda_i + \mu_i \\ \gamma_i = q_{31}\cdot\lambda_i + q_{32}\cdot\mu_i + \nu_i \\ \delta_i = q_{41}\cdot\lambda_i + q_{42}\cdot\mu_i + q_{43}\cdot\nu_i + \varsigma_i \\ \ldots\ldots \end{cases}$$

де $q_{ij}, i=\overline{1,..m}, j=\overline{1,..i-1}, q_{ii}\equiv 1$ – невідомі елементи матриці перетворення Q. Виразимо їх через елементи кореляційної матриці $\Re$.

З умови $r_{\alpha\beta}=0$ отримуємо

$$\sum_{i=1}^{n}\lambda_i\cdot\mu_i = \sum_{i=1}^{n}\lambda_i\cdot(q_{21}\cdot\lambda_i+\mu_i) = q_{21}\sum_{i=1}^{n}\lambda_i^2 + \sum_{i=1}^{n}\lambda_i\cdot\mu_i =$$
$$=(n-1)\cdot(q_{21}+r_{\lambda\mu}) = 0$$

Отже $k_{21} = -r_{\lambda\mu}$.

З умов $r_{\alpha\gamma} = r_{\beta\gamma} = 0$ отримуємо систему рівнянь для знаходження невідомих q_{31} і q_{32}

$$\begin{cases} q_{31} + q_{32}\cdot r_{\lambda\mu} = -r_{\lambda\nu} \\ 0 \ + q_{32}(q_{12}\cdot r_{\lambda\mu}+1) = q_{21}\cdot r_{\lambda\mu} - r_{\mu\nu} \end{cases}$$

Для обчислення q_{41}, q_{42} і q_{43} розв'язується система рівнянь

$$\begin{cases} q_{41} + q_{42}\cdot r_{\lambda\mu} + q_{43}\cdot r_{\lambda\nu} = -r_{\lambda\varsigma} \\ 0 \ + q_{42}\cdot(q_{21}\cdot r_{\lambda\mu}+1) + q_{43}\cdot(q_{21}\cdot r_{\lambda\nu} + r_{\mu\nu}) = q_{21}\cdot r_{\lambda\varsigma} - r_{\mu\varsigma} \\ 0+0+q_{43}\cdot(q_{31}\cdot r_{\lambda\nu} + q_{32}\cdot r_{\mu\nu} + 1) = -q_{31}\cdot r_{\lambda\varsigma} - q_{32}\cdot r_{\mu\varsigma} - r_{\nu\varsigma} \end{cases}$$

і так далі. Елементи системи рівнянь можуть бути виражені через мінори різного порядку кореляційної матриці $\Re$.

Систему A випадкових послідовностей (α, β, γ, δ …) з діагональною кореляційною матрицею можна отримати як результат операції

$$A = \Lambda\cdot Q^T$$

де Q^T– транспонована матриця Q коефіцієнтів перетворення q_{ij}.

В таблиці 1 наведено приклад для однієї з реалізацій системи випадкових величин, кількість членів в кожній послідовності n=14.

Таблиця 1. Приклад генерації системи незалежних випадкових величин

Кореляційна матриця вихідної системи Λ

1	-0,148	0,191	0,330
-0,148	1	0,005	-0,149
0,191	0,005	1	-0,079
0,330	-0,149	-0,079	1

Матриця перетворення Q^T

1	0,148	-0,196	-0,343
0	1	-0,034	0,098
0	0	1	0,144
0	0	0	1

Кореляційна матриця системи A

1	0	0	0
0	1	0	0
0	0	1	0
0	0	0	1

Слід відмітити, що після перетворення випадкові величини $\alpha, \beta, \gamma, \delta$ … залишаються центрованими, але дисперсії послідовностей β_i, γ_i, δ_i … відрізняються від одиниці і дорівнюють $Q_j \cdot \Re \cdot Q_j^T$, де K_j – відповідний рядок матриці коефіцієнтів перетворення. Після врахування відповідної поправки отримана система $A(\alpha, \beta, \gamma, \delta$ …) містить стандартизовані лінійно незалежні послідовності випадкових величин.

Нижче наведені приклади реалізації задачі моделювання псевдо-випадкових послідовностей з заданими параметрами регресії для деяких класів функцій.

2. Моделювання функцій одного аргументу

2.1 Модель параболічної залежності. В якості прикладу квазілінійної функції розглянемо параболу другого степеню

$$\varphi(A,x_i)=a_0+a_1\cdot x_i+a_2\cdot x_i^2.$$

Будемо шукати регресію статистичної сукупності $\tilde{o}_{s}=\varphi(\grave{A},\tilde{o}_{s})+k_0\cdot\varepsilon_i$, де збурення має вигляд $\varepsilon_i=\dfrac{k_1\cdot\alpha_i+k_2\cdot\beta_i+k_3\cdot\gamma_i}{\grave{I}_{\tilde{n}}}$, а випадкові послідовності α_i, β_i, γ_i є стандартизовані і незалежні. Тоді система однорідних рівнянь (3) перепишеться в такому вигляді

$$\begin{cases} k_1\sum_{i=1}^{n}\alpha_s+k_2\sum_{i=1}^{n}\beta_i+k_3\sum_{i=1}^{n}\gamma_i=0 \\ k_1\sum_{i=1}^{n}\alpha_s\cdot x_i+k_2\sum_{i=1}^{n}\beta_i\cdot x_i+k_3\sum_{i=1}^{n}\gamma_i\cdot x_i=0 \\ k_1\sum_{i=1}^{n}\alpha_s\cdot x_i^2+k_2\sum_{i=1}^{n}\beta_i\cdot x_i^2+k_3\sum_{i=1}^{n}\gamma_i\cdot x_i^2=0 \end{cases}$$

Враховуючи, що послідовності α_i, β_i, γ_i є стандартизовані, перше рівняння системи завжди виконується, тому кількість складових збурення ε_i можна обмежити трьома, хоч в загальному випадку при трьох невідомих параметрах регресії a_0, a_1, a_2 кількість складових повинна бути на одиницю більше. Отже, покладаючи k_1 =1 і розв'язуючи систему рівнянь

$$\begin{cases} k_2\sum_{i=1}^{n}\beta_i\cdot x_i+k_3\sum_{i=1}^{n}\gamma_i\cdot x_i=-\sum_{s=1}^{\ddot{\imath}}\alpha_s\cdot\tilde{o}_s \\ k_2\sum_{i=1}^{n}\beta_i\cdot x_i^2+k_3\sum_{i=1}^{n}\gamma_i\cdot x_i^2=-\sum_{s=1}^{\ddot{\imath}}\alpha_s\cdot\tilde{o}_s^2 \end{cases}$$

відносно невідомих k_2 і k_3, можна отримати послідовність ε_i, використання якої дозволить отримати рівняння регресії параболи з заданими параметрами. Корегуючий множник k_0 при заданому коефіцієнті детермінації R^2 обчислюється за формулою (4).

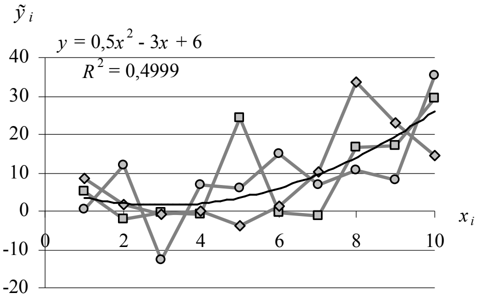

Рис. 1. Графіки теоретичної кривої і окремих реалізацій параболічної функції

Для ілюстрації на рисунку 1 наведені приклади трьох різних реалізацій, обробка кожної с яких методом найменших квадратів дозволяє отримати задані параметри регресії з коефіцієнтом детермінації R^2=0,5. В якості прикладу наведемо значення корегуючих множників для однієї з реалізацій: k_1 =1 (за замовчуванням), k_2=-0,415, k_3=-0,370, і після стандартизації послідовності ε_i k_0=8,48.

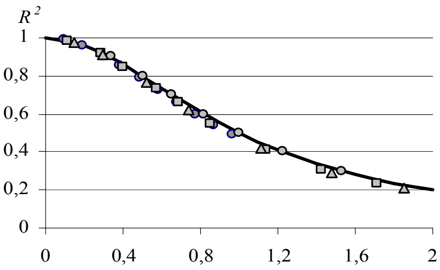

Рис. 2. Графіки залежності коефіцієнта детермінації R^2 від відношення $k_0/\sqrt{D(\varphi(A,x_i))}$

На рисунку 2 наведені графіки залежності обчислених коефіцієнтів детермінації R^2 від відношення $k_0/\sqrt{D(\varphi(A,x_i))}$, де $D(\varphi(A,x_i))$ – дисперсія теоретичної кривої на відрізку аналізу для різних реалізацій, і для порівняння теоретична крива, обчислена з використанням формули (4). Графіки свідчать, що припущення про незалежність монотонної функції регресії і елементів послідовності ε_i можна вважати слушним.

Слід відмітити, що для будь-якої функції, яка містить вільний член a_0, кількість доданків у випадковій послідовності ε_i може дорівнювати m, а не m+1, як в загальному випадку (тут, як і вище, m – кількість параметрів регресії).

2.2. Аналіз показникової функції. В якості прикладу суттєво нелінійної функції розглянемо функцію залежності пропозиції товару на ринку від ціни за одиницю продукції в такому вигляді

$$\varphi(\grave{A},\tilde{o}_s)=\grave{a}_1\cdot(\tilde{o}_s-\grave{a}_2)^{\grave{a}3}\cdot$$

В припущенні $a_1 \neq 0$, $a_3 \neq 0$ система нормальних рівнянь має вигляд

$$\begin{cases} \sum_{s=1}^{\ddot{\imath}}\left(\tilde{o}_s-\grave{a}_1\cdot(\tilde{o}_s-\grave{a}_2)^{\grave{a}3}\right)\cdot(\tilde{o}_s-\grave{a}_2)^{\grave{a}3}=0 \\ \sum_{s=1}^{\ddot{\imath}}\left(\tilde{o}_s-\grave{a}_1\cdot(\tilde{o}_s-\grave{a}_2)^{\grave{a}3}\right)\cdot(\tilde{o}_s-\grave{a}_2)^{\grave{a}3-1}=0 \\ \sum_{s=1}^{\ddot{\imath}}\left(\tilde{o}_s-\grave{a}_1\cdot(\tilde{o}_s-\grave{a}_2)^{\grave{a}3}\right)\cdot(\tilde{o}_s-\grave{a}_2)^{\grave{a}3}\cdot\ln(\tilde{o}_s-\grave{a}_2)=0 \end{cases}$$

Перепишемо систему у вигляді

$$\begin{cases} a_1\sum_{i=1}^{n}(x_i-a_2)^{2a3}=\sum_{i=1}^{n}\widetilde{y}_i\cdot(x_i-a_2)^{a3} \\ a_1\sum_{i=1}^{n}(x_i-a_2)^{2a3-1}=\sum_{i=1}^{n}\widetilde{y}_i\cdot(x_i-a_2)^{a3-1} \\ a_1\sum_{i=1}^{n}(x_i-a_2)^{2a3}\cdot\ln(x_i-a_2)=\sum_{i=1}^{n}\widetilde{y}_i\cdot(x_i-a_2)^{a3}\cdot\ln(x_i-a_2) \end{cases}$$

Для скорочення запису позначимо

$$u_i=(x_i-a_2)^{a3},\ v_i=(x_i-a_2)^{a3-1},\ w_i=(x_i-a_2)^{a3}\cdot\ln(x_i-a_2)$$

Приймемо, що $\tilde{y}_i=\varphi\,(A,\,x_i)+z_i$. У відповідності до формули (2) $z_i=k_o\cdot\varepsilon_i$, де ε_i – стандартизована лінійна комбінація чотирьох випадкових послідовностей

$$\varepsilon_i=\frac{k_1\cdot\alpha_i+k_2\cdot\beta_i+k_3\cdot\gamma_i+k_4\cdot\delta_i}{\grave{I}_{\tilde{n}}}$$

Поклавши для визначеності k_1=1, будемо розв'язувати систему лінійних рівнянь

$$\begin{cases} k_2\sum_{i=1}^{n}\beta_i\cdot u_i+k_3\sum_{i=1}^{n}\gamma_i\cdot u_i+k_4\sum_{i=1}^{n}\delta_i\cdot u_i=-\sum_{i=1}^{n}\alpha_i\cdot u_i \\ k_2\sum_{i=1}^{n}\beta_i\cdot v_i+k_3\sum_{i=1}^{n}\gamma_i\cdot v_i+k_4\sum_{i=1}^{n}\delta_i\cdot v_i=-\sum_{i=1}^{n}\alpha_i\cdot v_i \\ k_2\sum_{i=1}^{n}\beta_i\cdot w_i+k_3\sum_{i=1}^{n}\gamma_i\cdot w_i+k_4\sum_{i=1}^{n}\delta_i\cdot w_i=-\sum_{i=1}^{n}\alpha_i\cdot w_i \end{cases}$$

Після знаходження k_j і стандартизації послідовності ε_i обчислюється за формулою (4) потрібне значення множника k_0. Як результат, статистична послідовність $\tilde{y}_i=\varphi\,(A,\,x_i)+k_o\cdot\varepsilon_i$ має задані значення параметрів регресії і коефіцієнта детермінації R^2.

На рисунку 3 наведені приклади трьох послідовностей, обробка яких МНК дозволяє отримати одні і ті ж величини параметрів регресії. Для однієї з реалізацій отримані такі значення корегуючих коефіцієнтів: k_1=1,0 (за замовчуванням), k_2=-0,183, k_3=0,0482, k_4=0,0766, k_0=0,301 (після стандартизації послідовності ε_i).

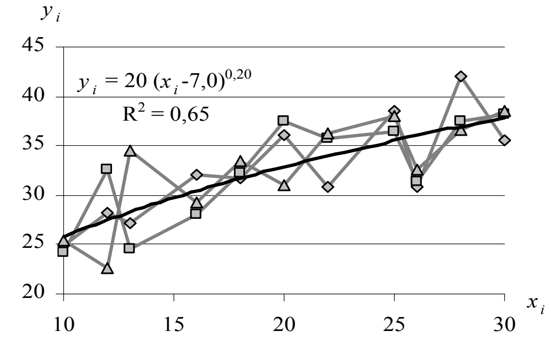

Рис. 3. Графіки теоретичної кривої і окремих реалізацій показникової функції

Слід зауважити, що описаний алгоритм генерації послідовностей з заданими параметрами регресії і коефіцієнта детермінації застосовується при використанні безпосередньо методу найменших квадратів (зокрема, при використанні функції Поиск решения чи аналогічних оптимізуючи методів) без проміжної лінеаризації статистичних даних.

3. Аналіз функцій кількох змінних

3.1 Моделювання системи з заданою кореляційною матрицею. При статистичному моделюванні функцій кількох змінних крім параметрів регресій можуть бути задані числові характеристики, в тому числі елементи кореляційної матриці системи $X(x,\ y,\ z,\ u\ \ldots)$. Для моделювання системи стандартизованих випадкових величин з заданою кореляційною матрицею використаємо систему $A(\alpha,\ \beta,\ \gamma,\ \delta\ \ldots)$ стандартизованих величин з одиничною кореляційною матрицею. Представимо кожну послідовність величин системи $X(x,\ y,\ z,\ u\ \ldots)$ у вигляді зваженої суми величин системи A

$$\begin{cases} x_i=p_{11}\cdot\alpha_i \\ y_i=p_{21}\cdot\alpha_i+p_{22}\cdot\beta_i \\ z_i=p_{31}\cdot\alpha_i+p_{32}\cdot\beta_i+p_{33}\cdot\gamma_i \\ u_i=p_{41}\cdot\alpha_i+p_{42}\cdot\beta_i+p_{43}\cdot\gamma_i+p_{44}\cdot\delta_i \\ \ldots\ldots \end{cases}$$

де p_{ij} – невідомі параметри перетворення.

Щоб після перетворення послідовності величин x_i, y_i, z_i, u_i … залишились стандартизованими, слід члени кожної послідовності поділити на значення їх середнього квадратичного відхилення. Наприклад, після перетворення дисперсія значень u_i з урахуванням незалежності стандартизованих членів послідовності $A(\alpha,\ \beta,\ \gamma,\ \delta\ \ldots)$ складе

$$D(u)=D(\alpha_i+p_{42}\cdot\beta_i+p_{43}\cdot\gamma_i+p_{44}\cdot\delta_i)=$$
$$=D(\alpha_i)+p_{42}{}^2\cdot D(\beta_i)+p_{43}{}^2\cdot D(\gamma_i)+p_{44}{}^2\cdot D(\delta_i)=$$
$$=1+p_{42}{}^2+p_{43}{}^2+p_{44}{}^2\,.$$
$$S(u)=\sqrt{1+p_{42}{}^2+p_{43}{}^2+p_{44}{}^2}$$

Обчислимо коефіцієнт парної кореляції r_{xu}

$$r_{xu}=\frac{\sum_{i=1}^{n}x_i\cdot u_i-n\cdot\bar{x}\cdot\bar{u}}{S(x)\cdot S(u)\cdot(n-1)}=$$
$$=\frac{\sum_{i=1}^{n}\alpha_i(\alpha_i+p_{42}\cdot\beta_i+p_{43}\cdot\gamma_i+p_{44}\cdot\delta_i)}{1\cdot\sqrt{1+p_{42}{}^2+p_{43}{}^2+p_{44}{}^2}\cdot(n-1)}=$$
$$=\frac{\sum_{i=1}^{n}\alpha_i{}^2+p_{42}\sum_{i=1}^{n}\alpha_i\cdot\beta_i+p_{43}\sum_{i=1}^{n}\alpha_i\cdot\gamma_i+p_{44}\sum_{i=1}^{n}\alpha_i\cdot\delta_i}{\sqrt{1+p_{42}{}^2+p_{43}{}^2+p_{44}{}^2}\cdot(n-1)}=$$
$$=\frac{1}{\sqrt{1+p_{42}{}^2+p_{43}{}^2+p_{44}{}^2}}=\frac{1}{S(u)}\,.$$

Отже, якщо кожну змінну системи $X(x,\ y,\ z\ \ldots)$ поділити на відповідне середнє квадратичне відхилення (або помножити на відповідний коефіцієнт кореляції), то отримана система випадкових величин буде стандартизованою. Таким чином, враховуючи що r_{xx}=1, будемо розв'язувати систему рівнянь

$$\begin{cases} \tilde{o}_{з}=\alpha_{з} \\ \acute{o}_{з}=r_{xy}\cdot\alpha_i+p_{22}\cdot\beta_i \\ z_i=r_{xz}\cdot\alpha_i+p_{32}\cdot\beta_i+p_{33}\cdot\gamma_i \\ u_i=r_{xu}\cdot\alpha_i+p_{42}\cdot\beta_i+p_{43}\cdot\gamma_i+p_{44}\cdot\delta_i \\ \ldots\ldots \end{cases}$$

Використовуючи задані значення кореляційної матриці системи $X(x,\ y,\ z\ \ldots)$, послідовно шукаємо невідомі параметри матриці перетворення $\overline{\overline{Ð}}$.

З рівняння

$$r_{xy}=\frac{\sum_{i=1}^{n}x_i\cdot y_i-n\cdot\bar{x}\cdot\bar{y}}{S(x)\cdot S(y)\cdot(n-1)}=\frac{\sum_{i=1}^{n}\alpha_i\cdot(r_{xy}\cdot\alpha_i+p_{22}\cdot\beta_i)}{1\cdot\sqrt{r_{xy}{}^2+p_{22}{}^2}\cdot(n-1)}=$$

$$= \frac{r_{xy}\sum_{i=1}^{n}\alpha_i^{2} + p_{22}\sum_{i=1}^{n}\alpha_i \cdot \beta_i}{\sqrt{r_{xy}^{2} + p_{22}^{2}} \cdot (n-1)} = \frac{r_{xy}}{\sqrt{r_{xy}^{2} + p_{22}^{2}}}$$

отримуємо $p_{22} = \sqrt{1 - r_{xy}^{2}}$

Аналогічно, з виразів для r_{xz} і r_{yz} можна знайти

$$p_{32} = \frac{r_{yz} - r_{xy} \cdot r_{xz}}{p_{22}}; \qquad p_{33} = \sqrt{1 - r_{xz}^{2} - p_{32}^{2}}.$$

Використовуючи вирази для парних коефіцієнтів r_{xu}, r_{yu}, r_{zu} отримуємо

$$p_{42} = \frac{r_{yu} - r_{xy} \cdot r_{xu}}{p_{32}}; \quad p_{43} = \frac{r_{zu} - r_{xz} \cdot r_{xu} - p_{32} \cdot p_{42}}{p_{42}};$$

$$p_{44} = \sqrt{1 - r_{xu}^{2} - p_{42}^{2} - p_{43}^{2}}.$$

Таким чином, параметри перетворення однозначно виражаються через елементи заданої кореляційної матриці. Для отримання матриці вихідних даних з заданою кореляційною матрицею слід обчислити добуток матриць $A \cdot \overline{\overline{D}}^{T}$, де $\overline{\overline{D}}^{T}$– транспонована матриця параметрів перетворення. В таблиці 2 наведено приклад матриці перетворення для модельного прикладу.

Таблиця 2. Матриці параметрів перетворення і кореляційна

Матриця параметрів перетворення p_{ij}			
1	0	0	0
0,20	0,98	0	0
0,30	0,45	0,84	0
0,40	0,53	0,41	0,63

Результативна кореляційна матриця			
1	0,20	0,30	0,40
0,20	1	0,50	0,60
0,30	0,50	1	0,70
0,40	0,60	0,70	1

3.2. Моделювання функції кількох змінних. Запропонована методика генерації послідовностей з заданими параметрами регресії придатна і при аналізі функцій кількох змінних. Розглянемо лінійну функцію наприклад трьох змінних $\varphi(A, x_i, y_i, z_i) = a_0 + a_1 x_i + a_2 y_i + a_3 z_i \cdot$

Враховуючи наявність вільного члена a_0, представимо збурення у вигляді суми чотирьох стандартизованих послідовностей

$$\varepsilon_i \cdot \grave{I}_{\tilde{n}} = k_0 \cdot (k_1 \cdot \alpha_i + k_2 \cdot \beta_i + k_3 \cdot \gamma_i + k_4 \cdot \delta_i)$$

Система лінійних рівнянь для знаходження множників k_j має вид (по замовчуванню k_1=1)

$$\begin{cases} k_2\sum_{i=1}^{n}\beta_i \cdot x_i + k_3\sum_{i=1}^{n}\gamma_i \cdot x_i + k_4\sum_{i=1}^{n}\delta_i \cdot x_i = -\sum_{i=1}^{n}\alpha_i \cdot x_i \\ k_2\sum_{i=1}^{n}\beta_i \cdot y_i + k_3\sum_{i=1}^{n}\gamma_i \cdot y_i + k_4\sum_{i=1}^{n}\delta_i \cdot y_i = -\sum_{i=1}^{n}\alpha_i \cdot y_i \\ k_2\sum_{i=1}^{n}\beta_i \cdot z_i + k_3\sum_{i=1}^{n}\gamma_i \cdot z_i + k_4\sum_{i=1}^{n}\delta_i \cdot z_i = -\sum_{i=1}^{n}\alpha_i \cdot z_i \end{cases}$$

Розв'язуючи систему рівнянь, можна знайти коефіцієнти k_j і використовуючи формулу (4) добитися потрібного значення коефіцієнта детермінації R^2.

В якості прикладу наведено результати моделювання лінійної функції випадкових величин. Система трьох величин має такі наперед задані характеристики: середні значення 10,00, 8,00, 6,00; середні квадратичні відхилення 1,00, 1,40, 1,80. Парні коефіцієнти кореляції r_{xy}=0,20, r_{xz}=0,30, r_{yz}=0,50. Задані параметри лінійної залежності

$\varphi(A, x_i, y_i, z_i)$=10,0+1,50$x_i$-0,50$y_i$+2,50$z_i$.

Коефіцієнт детермінації задано R^2=0,50.

В таблиці 3 наведено приклад однієї з можливих реалізацій системи аргументів $\{x_i, y_i, z_i\}$ з заданими числовими характеристиками і для неї наведено три різних реалізації функції $\varphi(A, x_i, y_i, z_i)$ з заданими параметрами регресії. Нижче для прикладу для кожної з них наведені значення корегуючих множників k_j. Тому що допоміжна послідовність ε_i стандартизована, при фіксованому значенні k_o коефіцієнт детермінації залишається незмінним, незмінні також похибки визначення параметрів регресії. Для іншого набору координат $\{x_i, y_i, z_i\}$ для досягнення такого ж значення R^2 буде інша величина k_o.

Таблиця 3. Результати моделювання системи випадкових величин

x_i	y_i	z_i		$\varphi_1(X)$	$\varphi_2(X)$	$\varphi_3(X)$
9,19	7,20	4,27		31,3	36,7	25,9
11,22	7,50	5,10		36,5	29,9	35,3
9,31	8,85	6,28		29,5	31,3	35,7
9,63	8,00	3,87		27,6	36,2	36,0
9,95	9,80	7,67		30,2	36,9	37,2
11,32	8,49	7,30		37,8	41,6	40,5
10,34	9,65	6,11		36,8	42,8	28,8
10,12	5,50	4,78		36,2	34,2	29,3
10,10	10,75	8,97		52,4	40,8	44,6
7,98	6,88	5,32		37,1	23,6	25,9
11,46	7,74	5,82		39,1	34,4	36,9
10,95	7,58	8,31		46,3	44,7	47,0
9,13	6,55	7,49		34,2	45,7	45,6
9,30	7,51	2,71		29,0	25,1	35,2
			k_0	4,85	4,85	4,85
			k_1	1	1	1
			k_2	-9,14	-0,47	0,31
			k_3	9,95	0,43	0,37
			k_4	-0,71	-0,13	1,12

Наведені результати дозволяють отримати випадкові моделі з заданими параметрами, що є важливим при дослідженні як самих моделей, так і похибок визначення параметрів регресії.

ЛІТЕРАТУРА

1. Жлуктенко В. І., Наконечний С. І., Савіна С. С. Теорія ймовірностей і математична статистика: Навч.-метод. посібник. У 2-х ч. — Ч.ІІ. Математична статистика. — К. : КНЕУ, 2001. — 336 с.
2. Толбатов Ю. А. Математична статистика та задачі оптимізації в алгоритмах і програмах — К. : Вища школа, 1994 — 356 с.

пост.08.11.12