УДК 519.254:519.652
ББК В172.8

П.О. Приставка, канд. техн. наук, доц.

ОБЧИСЛЮВАЛЬНА ТЕХНОЛОГІЯ ОБРОБКИ МАСИВІВ РЕАЛІЗАЦІЙ ДВОВИМІРНОЇ ВИПАДКОВОЇ ВЕЛИЧИНИ

Подано технологію обробки масивів реалізацій двовимірної випадкової величини: підготовка даних до опрацювання, непараметрична оцінка функції щільності та розподілу ймовірностей.

Нехай маємо двовимірну векторну випадкову величину

$$\vec{\xi} = (\xi_1(\omega), \xi_2(\omega)),$$

яка визначена на σ -алгебрі F і яка набуває дійсні значення з

$$\mathbf{R}_2 = \{-\infty < t < \infty\} \bigcap \{-\infty < q < \infty\},$$

такі, що

$$\{\omega : -\infty < \xi_1(\omega) < t; \infty < \xi_2(\omega) < q\} \in \mathrm{F},\ P\{\omega : -\infty < \xi_1(\omega) < \infty; -\infty < \xi_2(\omega) < \infty\} = 1,$$

тобто будемо вважати, що існує відбиття $\Omega \xrightarrow{\vec{\xi}(\omega)} \mathbf{R}_2$ [1; 2].

Формально двовимірну випадкову величину можна розглядати як сукупність одновимірних. Тим самим властивості одновимірних випадкових величин ототожнюються з властивостями двовимірної (у тому числі і класифікація на дискретні та неперервні дійсні випадкові величини). Враховуючи, що ймовірнісний простір $\langle \Omega, \mathrm{F}, P \rangle$ відбивається через $\vec{\xi}(\omega)$ у простір дійсних чисел, заданих за Борелєм $\langle \Omega, \mathrm{F}, P \rangle \xrightarrow{\vec{\xi}(\omega)} \langle \mathbf{R}_2, B, P \rangle$, вводять функцію розподілу ймовірностей двовимірної випадкової величини у вигляді

$$F(t,q) = P\{\omega : -\infty < \xi_1(\omega) < t; -\infty < \xi_2(\omega) < q\}.$$

Геометрично $F(t,q)$ оцінює ймовірність потрапляння випадкової точки в кут, заданий у двовимірному евклідовому просторі зі сторонами, ортогональними відповідним вісям координат. Властивості функції $F(t,q)$ або еквівалентні властивостям одновимірних функцій розподілу, або мають певні особливості. Так, імовірність того, що реалізація випадкової величини $\vec{\xi} = (\xi_1, \xi_2)$ потрапляє у прямокутну область

$$I_2 = E_2(b_1, b_2) - E_2(a_1, b_2) - E_2(b_1, a_2) + E_2(a_1, a_2),$$

буде дорівнювати

$$P\{a_1 \le \xi_1 < b_1, a_2 \le \xi_2 < b_2\} = F(b_1, b_2) - F(a_1, b_2) - F(b_1, a_2) + F(a_1, a_2).$$

Якщо існує

$$\lim_{\substack{t \to t^0 + 0, \\ q \to q^0 + 0}} F(t,q) = P\{\xi_1 = t^0, \xi_2 = q^0\} + F(t^0, q^0),$$

то функція $F(t,q)$ неперервна праворуч. При цьому вводиться функція щільності розподілу ймовірностей

$$f(t,q) = \frac{\partial^2 F(t,q)}{\partial t \partial q},$$

отже $F(t,q) = \int\limits_{-\infty}^{t} \int\limits_{-\infty}^{q} f(\tau, \rho) d\tau d\rho$.

Якщо

$$F(t,q) = F(t)\ F(q),$$

то випадкові величини ξ_1, ξ_2 – незалежні.

Нехай маємо масив реалізацій

$$\{t_k, q_k;\ k = \overline{1,n}\} \tag{1}$$

деякої неперервної двовимірної випадкової величини, тобто маємо деякий об'єкт дослідження, що характеризується ознаками T, Q. У результаті проведення експерименту одержано вибірку реалізацій випадкової величини $\vec{\xi} = (\xi_1, \xi_2)$, яка характеризує ймовірнісно-статистичні властивості об'єкта. Необхідно оцінити кількісні показники вибірки та функції щільності і розподілу ймовірностей $\vec{\xi} = (\xi_1, \xi_2)$.

Слід звернути увагу, що реалізації двовимірної випадкової величини є точками на площині. Візуальне зображення масиву (1) називають кореляційним полем. Косинус кута, у який видно кореляційне поле або, косинус кута між векторами спостережень $\vec{T} = \{t_k;\ k = \overline{1,n}\}$ і $\vec{Q} = \{q_k;\ k = \overline{1,n}\}$ дорівнює величині, що визначає лінійний стохастичний зв'язок поміж одновимірними випадковими величинами ξ_1 і ξ_2 та має назву парний коефіцієнт кореляції r [3]:

$$r = \frac{M\{(\xi_1 - M\{\xi_1\})(\xi_2 - M\{\xi_2\})\}}{\sqrt{D\{\xi_1\}D\{\xi_2\}}} = \frac{\operatorname{cov}\{\xi_1, \xi_2\}}{\sigma\{\xi_1\}\sigma\{\xi_2\}}.$$

Серед важливих властивостей коефіцієнта кореляції: $|r| \le 1$; якщо $r = 0$, то ξ_1, ξ_2 – незалежні випадкові величини; якщо $|r| = 1$, то поміж ξ_1 та ξ_2 відбувається функціональний зв'язок. Статистичне визначення значення r випливає з виразу

$$\hat{r} = \left(\overline{tq} - \overline{t}\ \ \overline{q}\right) / \left(\sigma_t \sigma_q\right), \tag{2}$$

де $\overline{t} = \frac{1}{n}\sum_{k=1}^{n} t_k$;

$$\overline{q} = \frac{1}{n}\sum_{k=1}^{n} q_k;$$

$$\overline{tq} = \frac{1}{n}\sum_{k=1}^{n} t_k q_k;$$

$$\sigma_k^2 = \frac{1}{n-1}\sum_{k=1}^{n}\left(t_k - \overline{t}\right)^2;$$

$$\sigma_q^2 = \frac{1}{n-1}\sum_{k=1}^{n}\left(q_k - \overline{q}\right)^2. \tag{3}$$

Оцінки, подані виразами (2), (3), прийнято вважати головними статистичними характеристиками вибірки за результатами спостережень реалізацій двовимірної випадкової величини. Іноді за такі характеристики об'єкта спостережень за ознаками T, Q цілком достатньо вважати двовимірне середньостатистичне

$$\overline{M} = \{M\{\xi_1\}, M\{\xi_2\}\} = \{\overline{t}, \overline{q}\}$$

(центроїд вибірки) та дисперсійно-коваріаційну матрицю A:

$$A = \begin{Vmatrix} D\{\xi_1\} & \operatorname{cov}\{\xi_1, \xi_2\} \\ \operatorname{cov}\{\xi_2, \xi_1\} & D\{\xi_2\} \end{Vmatrix} = \begin{Vmatrix} \sigma_t^2 & \sigma_t \sigma_q r \\ \sigma_q \sigma_t r & \sigma_q^2 \end{Vmatrix}.$$

Якщо ознаки T, Q вимірюються в різних одиницях (наприклад, грамах та метрах), рекомендується після оцінки статистичних показників стандартизувати значення кожної ознаки через перехід до безвимірних величин:

$$t_k^* = (t_k - \bar{t})/\sigma_t ;$$

$$q_k^* = (q_k - \bar{q})/\sigma_q , \quad k = \overline{1,n}.$$

У цьому разі полегшується задача інтерпретації лінійної залежності ознак (звісно, якщо така залежність існує). У разі, коли обробці підлягає стандартизований масив

$$\{t_k^*, q_k^*;\ k = \overline{1,n}\}, \tag{4}$$

неважко впевнитися, що

$$\xi_1^* = (\xi_1 - M\{\xi_1\})\big/\sqrt{D\{\xi_1\}} ;$$

$$\xi_2^* = (\xi_2 - M\{\xi_2\})\big/\sqrt{D\{\xi_2\}} ;$$

$$r = \operatorname{cov}\{\xi_1^*, \xi_2^*\} ;$$

$$\bar{M}^* = \{M\{\xi_1^*\}, M\{\xi_2^*\}\} = \{0,0\},$$

а матриця A перетворюється в кореляційну матрицю R :

$$R = A^* = \begin{Vmatrix} 1 & r \\ r & 1 \end{Vmatrix}.$$

Наведені статистичні характеристики дозволяють проведення ймовірнісної оцінки випадкової величини. Так, якщо $\bar{\xi}$ має двовимірний нормальний розподіл, то функція щільності, одержана за результатами обробки масивів (1), (4), відповідно буде визначатися:

$$f(t,q) = \frac{1}{2\pi\sigma_t\sigma_q\sqrt{1-r^2}} \exp\left(-\frac{1}{2(1-r^2)}\left(\left(\frac{t-\bar{t}}{\sigma_t}\right)^2 - 2r\,\frac{t-\bar{t}}{\sigma_t}\cdot\frac{q-\bar{q}}{\sigma_q} + \left(\frac{q-\bar{q}}{\sigma_q}\right)^2\right)\right),$$

$$f(t^*, q^*) = \frac{1}{2\pi\sqrt{1-r^2}} \exp\left(-\frac{1}{2(1-r^2)}\left(\left(t^*\right)^2 - 2r\cdot t^*\cdot q^* + \left(q^*\right)^2\right)\right).$$

Якщо, при цьому, випадкові величини ξ_1, ξ_2 незалежні, то їх сумісна функція щільності для зазначених масивів відповідно дорівнює

$$f(t,q) = \frac{1}{2\pi\sigma_t\sigma_q} \exp\left(-\frac{1}{2}\left(\left(\frac{t-\bar{t}}{\sigma_t}\right)^2 + \left(\frac{q-\bar{q}}{\sigma_q}\right)^2\right)\right);$$

$$f(t^*, q^*) = \frac{1}{2\pi} \exp\left(-\frac{1}{2}\left(\left(t^*\right)^2 + \left(q^*\right)^2\right)\right).$$

Наведений приклад імовірнісної оцінки наочно доводить, що процедура подібної оцінки набагато спрощується у випадку незалежності відповідних одновимірних випадкових величин. Існує можливість, у разі, коли така залежність існує, виконати лінійне перетворення величин ξ_1, ξ_2, яке призводить до одержання незалежних випадкових величин ξ_1', ξ_2'. Тобто точки двовимірного простору, визначені масивом (1), необхідно подати у термінах двох головних ортогональних вісей (компонент) [3]. У загальному випадку вісі задаються лініями, для яких сума квадратів відстаней до всіляких точок масиву (1) мінімальна. Отже, є можливість говорити про дисперсії проекцій точок на напрямок компоненти (вісі). Залишається виділити лише ті компоненти, на долю яких припадає суттєвий відсоток загальної дисперсії вибірки. Розв'язання цієї задачі носить назву методу головних компонент. Таким чином, головні компоненти – це такі комбінації початкових ознак:

$$t_k' = \alpha_{T'T} t_k + \alpha_{T'Q} q_k ;$$

$$q_k' = \alpha_{Q'T} t_k + \alpha_{Q'Q} q_k ,$$

де $\alpha_{TT}^2+\alpha_{TQ}^2=1$; $\alpha_{Q'T}^2+\alpha_{Q'Q}^2=1$; α_{TT}, $\alpha_{T'Q}$, $\alpha_{Q'T}$, $\alpha_{Q'Q}$ – відповідно косинуси кутів між вісями T' і T, T' і Q, Q' і T, Q' і Q.

Вісь T' визначається умовою максимуму дисперсії вибірки (1) вздовж неї, вісь Q' –умовою ортогональності до T'. Щодо α_{TT}, $\alpha_{T'Q}$, $\alpha_{Q'T}$, $\alpha_{Q'Q}$,то: метод одержання напрямків головних осей T', Q' ґрунтується на знаходженні власних чисел і власних векторів дисперсійно-коваріаційної матриці A (кореляційної матриці R). Власний вектор з індексом T' являє собою набір коефіцієнтів $A_{T'}=\{\alpha_{TT},\alpha_{T'Q}\}$, а відповідне йому власне число $\lambda_{T'}=1+|r|$ дорівнює дисперсії компоненти T'. Так як при використанні матриці R сума власних чисел дорівнює кількості ознак, то знайшовши відношення $\frac{\lambda_{T'}}{2}$, одержують частку дисперсії вибірки, що відповідає T'-му напрямку. Те ж саме стосується і Q'-го напрямку (тут маємо $\lambda_{Q'}=1-|r|$). Якщо позначити дисперсії головних компонент $\sigma_{t'}'^2$, $\sigma_{q'}'^2$, виявляється, що

$$\sigma_{t'}^2+\sigma_{q'}^2=\sigma_t^2+\sigma_q^2.$$

Отже, початково закладена у даних сумарна варіабельність при переході до нових змінних не змінюється, а лише перерозподіляється. Крім того, нові змінні (компоненти) на відміну від початкових ознак набули таку цінну властивість, як відсутність кореляції поміж себе.

Зворотне перетворення для повернення до вихідних ознак має вигляд

$$t_k=\alpha_{TT'}t_k'+\alpha_{QT'}q_k';$$

$$q_k=\alpha_{TQ'}t_k'+\alpha_{QQ'}q_k',$$

де $\alpha_{TT'}$, $\alpha_{QT'}$, $\alpha_{TQ'}$, $\alpha_{QQ'}$ – косинуси кутів між відповідними вісями.

Можна навести і більш простішу, з точки зору обчислювальної складності, процедуру приведення залежних випадкових величин ξ_1, ξ_2 до незалежних ξ_1', ξ_2' [2]. Суть процедури полягає у використанні відомих формул повороту ортогональної системи на деякий кут φ:

$$\xi_1'=\xi_1\cos\varphi+\xi_2\sin\varphi;$$

$$\xi_2'=-\xi_1\sin\varphi+\xi_2\cos\varphi.$$

Величина кута φ, при повороті на який приходимо до незалежності ξ_1' і ξ_2', визначається із співвідношення

$$tg2\varphi=2r\sigma_t\sigma_q\Big/\left(\sigma_t^2-\sigma_q^2\right).$$

Таким чином, для подальшої обробки можемо використовувати масив $\{t_k',q_k';\ k=\overline{1,n}\}$ реалізацій двовимірної випадкової величини $\vec{\xi}'=(\xi_1',\xi_2')$, одержаний із вихідного (або стандартизованого вихідного) масиву реалізацій $\vec{\xi}=(\xi_1,\xi_2)$. Не зменшуючи загальності, у подальшому викладенні застосовуватимемо позначення такого масиву у вигляді (1).

Отже, нехай маємо вибірку (1) (стандартизовану чи ні) реалізацій двовимірної випадкової величини $\vec{\xi}=(\xi_1,\xi_2)$ у припущенні про незалежність одновимірних ξ_1, ξ_2. При цьому $t_{\min}$, $t_{\max}$, $q_{\min}$, $q_{\max}$ – мінімальні та максимальні значення відповідно масивів $\{t_k;\ k=1,n\}$ і $\{q_k;\ k=1,n\}$. Як і в одновимірному випадку, введемо до розгляду варіаційний ряд двовимірної випадкової величини

$$\vec{\xi}=(\xi_1,\xi_2):\{(t_l,q_l),n_l,p_l;l=\overline{1,s}\},$$

де s – кількість варіант варіаційного ряду; n_l – частота варіанти (t_l,q_l), $l=\overline{1,s}$; $p_s=n_s/n$ – випадковість варіанти.

Зафіксуємо два розбиття Δ_{h_t}, Δ_{h_q} осей T, Q точками

$t_i = t_{\min} + (i+1/2)h_t$, $i = \overline{0, m_t - 1}$;

$q_j = q_{\min} + (j+1/2)h_t$, $j = \overline{0, m_q - 1}$,

де $h_t = ((t_{\max} - t_{\min})/m_t)$;

$h_q = ((q_{\max} - q_{\min})/m_q)$,

відповідно до яких задається Δ_{h_t, h_q} розбиття площини спостереження на класи (відповідна m_t, m_q – кількість класів). Клас охарактеризуємо варіантою, що знаходиться у середині класу: (i,j)-му класу поставимо у відповідність точку (t_i, q_j). Як і в одновимірному випадку, для двовимірного варіаційного ряду, розбитого на класи, вводяться поняття частоти і випадковості.

Частота n_{ij} – кількість варіант вихідного ряду, що потрапили до (i,j)-го класу. При знаходженні частот класів необхідно для кожної l-ї варіанти ($l = \overline{1,s}$) виконати такі дії: знайти індекси $i = [t_l/h_t]$, $j = [q_l/h_q]$ класу, до якого слід її віднести (тут $[\cdot]$ - ціла частина), відповідно збільшуючи на одиницю величину n_{ij} при віднесенні точки до певного класу.

Випадковість p_{ij} – відношення частоти варіанти до загальної кількості спостережень випадкової величини у класі:

$$p_{ij} = n_{ij}/n;$$

$$\sum_{i=0}^{m_t-1}\sum_{j=0}^{m_q-1} p_{ij} = 1.$$

Графік із стовпців, що задається сукупністю

$$\{(t_i, q_j), p_{ij}; i = \overline{0, m_t}, j = \overline{0, m_q - 1}\}$$

називають гістограмою випадковостей випадкової величини $\vec{\xi} = (\xi_1, \xi_2)$. Величина p_{ij} є інтегральною характеристикою величини $\vec{\xi}$, а саме – ймовірністю реалізації $\vec{\xi}$ в межах (i,j)-го класу або ймовірністю події

$$\{\omega : t_i - \tfrac{h_t}{2} \le \xi_1(\omega) < t_i + \tfrac{h_t}{2}; q_i - \tfrac{h_q}{2} \le \xi_2(\omega) < q_i + \tfrac{h_q}{2}\}.$$

Неважко показати зв'язок випадковості і функції щільності розподілу ймовірностей двовимірної випадкової величини. Якщо через $\overline{f}_{ij}$ позначити усереднене значення функції $f(t,q)$ на (i,j)-му елементі розбиття Δ_{h_t, h_q}, де

$$\overline{f}_{ij} = \frac{1}{h_t h_q}\int_{i-h_t/2}^{i+h_t/2}\int_{j-h_q/2}^{j+h_q/2} f(\tau,\rho)\,d\tau\,d\rho,$$

то буде правильним

$$\overline{f}_{ij} = p_{ij}\ \ h_t\ \ h_q.$$

Тим самим приходимо до того, що масив випадковостей

$$\{(t_i, q_j), p_{ij}; i = \overline{0, m_t}, j = \overline{0, m_q - 1}\}$$

при значній кількості n спостережень у вихідному масиві (1) є з точністю до константи (у даному разі $h_t\ h_q$) нічим іншим, як масивом усереднених значень функції щільності $f(t,q)$ на розбитті Δ_{h_t, h_q}. Позначивши через $p(t,q)$ так звану функцію випадковостей, будемо вважати, що якщо буде знайдено наближення $p(t,q)$, то воно з точністю до константи буде й оцінкою функції $f(t,q)$. Згідно з роботою [4], пропонується шукати наближення функції $p(t,q)$ у вигляді локального поліноміального сплайну на основі B-сплайнів другого порядку, близького до інтерполяційного у середньому:

$$S_{2,0}(p,t,q)=\sum_{i=0}^{m_t-1}\sum_{j=0}^{m_q-1}B_{2,h_t}(t-ih_t)B_{2,h_q}(q-jh_q)\,p_{ij},$$

де з урахуванням відповідного аргументу [5]:

$$B_{2,h_w}(w)=\begin{cases}0, & |w|\ge 3h_w/2,\\ (3+2w/h_w)^2/8, & w\in[-3h_w/2;-h_w/2],\\ 3/4-(2w/h_w)^2/4, & w\in[-h_w/2;h_w/2],\\ (3-2w/h_w)^2/8, & w\in[h_w/2;3h_w/2]\end{cases}\qquad w=t,q.$$

Можливість застосування сплайну $S_{2,0}(p,t,q)$ під час обробки статистичних даних було обґрунтовано як за результатами імітаційного моделювання [6], так і теоретично. Зазначений сплайн за рахунок свого аналітичного зображення, дозволяє одержати будь-яку кількість даних для подальшого аналізу та має високі апроксимаційні властивості [4]. Так, для $\forall p(t,q)\in C^{2,2}$ з урахуванням того, що $\|p(t,q)\|=1$, відбувається

$$\|p(t,q)-S_{2,0}(p,t,q)\|\le\frac{h_t^2}{6}\|p''_{t^2}(t,q)\|+\frac{h_q^2}{6}\|p''_{q^2}(t,q)\|+\frac{h_t^2h_q^2}{36}\|p^{(4)}_{t^2q^2}(t,q)\|+\varepsilon+O(h^4),$$

де $h=\max\{h_t,h_q\}$;

$\varepsilon=\max\limits_{i,j}\{\varepsilon_{i,j}\}$;

$\varepsilon_{ij}=p_{ij}-\overline{p}_{ij}$;

$$\overline{p}_{ij}=\frac{1}{h_th_q}\int_{i-h_t/2}^{i+h_t/2}\int_{j-h_q/2}^{j+h_q/2}p(\tau,\rho)\,d\tau\,d\rho.$$

Ще однією інтегральною характеристикою варіаційного ряду, розбитого на класи, є емпірична функція розподілу ймовірностей двовимірної випадкової величини, яка визначається із співвідношень:

$$F_{ij}=P\{\omega:t_{\min}\le\xi_1(\omega)<t_i+\tfrac{h_t}{2};q_{\min}\le\xi_2(\omega)<q_j+\tfrac{h_q}{2}\}=\sum_{ii=0}^{i}\sum_{jj=0}^{j}p_{ij},\ i=\overline{0,m_t},j=\overline{0,m_q-1}.$$

Знову ж таки, є зрозумілим, що при представницькому обсязі n вихідного масиву (1), масив $\{(t_i,q_j),F_{ij};i=\overline{0,m_t},j=\overline{0,m_q-1}\}$ презентує значення функції $F(t,q)$ у центральних точках розбиття Δ_{h_t,h_q}. У зв'язку з тим, що емпірична функція розподілу ймовірностей є неперервною ліворуч, але не є неперервною праворуч, як точки, що визначають розбиття Δ_{h_t,h_q}, краще узяти такі, що одержані із співвідношень:

$t_i=t_{\min}+(i+1)h_t,\ i=\overline{0,m_t-1}$;

$q_j=q_{\min}+(j+1)h_t,\ j=\overline{0,m_q-1}$.

Тоді маємо можливість шукати наближення функції розподілу ймовірностей $F(t,q)$ двовимірної випадкової величини ξ за масивом $\{(t_i,q_j),F_{ij};i=0,m_t,j=0,m_q-1\}$ у вигляді локального поліноміального сплайну на основі B-сплайнів третього порядку, близького до інтерполяційного у середньому:

$$S_{3,0}(F,t,q)=\sum_{i=0}^{m_t-1}\sum_{j=0}^{m_q-1}B_{3,h_t}(t-ih_t)B_{3,h_q}(q-jh_q)F_{ij},$$

де з урахуванням відповідного аргументу [5]:

$$B_{3,h_w}(w)=\frac{1}{48}\begin{cases} 0, & |w|\geq 2h_w, \\ (4+2w/h_w)^3, & w\in[-2h_w,-h_w], \\ -3(2w/h_w)^3-12(2w/h_w)^2+32, & w\in[-h_w,0], \\ 3(2w/h_w)^3-12(2w/h_w)^2+32, & w\in[0,h_w], \\ (4-2w/h_w)^3, & w\in[h_w,2h_w], \end{cases} \qquad w=t,q.$$

Як і $S_{2,0}(p,t,q)$, сплайну $S_{3,0}(F,t,q)$ притаманні високі апроксимаційні властивості. У разі коли $F(t,q)\in C^{3,3}$ з урахуванням того, що $\|F(t,q)\|=1$, відбувається [7]:

$$\|p(t,q)-S_{3,0}(p,t)\|\leq\frac{5h_t^2}{24}\|p''_{t^2}(t,q)\|+\frac{5h_q^2}{24}\|p''_{q^2}(t,q)\|+\frac{25h_t^2h_q^2}{576}\|p^{(4)}_{t^2q^2}(t,q)\|+\tilde{\varepsilon}+O(h^4),$$

де $h=\max\{h_t,h_q\}$;

$\tilde{\varepsilon}=\max\limits_{i,j}\{\tilde{\varepsilon}_{i,j}\}$;

$\tilde{\varepsilon}_{ij}=F_{ij}-\overline{F}_{ij}$;

$\overline{F}_{ij}=\dfrac{1}{h_th_q}\int\limits_i^{i+h_t}\int\limits_j^{j+h_q}F(\tau,\rho)\,d\tau\,d\rho$.

Всі зазначені заходи з урахуванням підготовки обчислювального процесу дозволяють ефективно проводити обробку масивів реалізації двовимірних неперервних випадкових величин. Отже, пропонується така послідовність дій при опрацюванні масивів на кшталт формули (1): проведення первинного статистичного аналізу (обчислення статистичних характеристик, стандартизація даних при необхідності, виявлення наявності залежності ознак, відтворення нормального закону розподілу); при наявній залежності ознак застосування перетворення (5) для переходу до нової системи координат; побудова варіаційного ряду, розбитого на класи, аналіз випадковостей, емпіричної функції розподілу; непараметричне відтворення функції щільності та розподілу за використанням сплайнів $S_{2,0}(p,t,q)$ та $S_{3,0}(F,t,q)$; застосування до результатів обробки перетворення (6) для повернення до вихідної системи координат та дестандартизація, якщо стандартизація проводилася.

Список літератури

1. *Уилкс С.* Математическая статистика. – М.: Наука, 1967. – 632 с.

2. *Коваленко И.Н., Гнеденко Б.В.* Теория вероятностей: Учеб. – К.: Вища шк., 1990. – 328 с.

3. *Бабак В.П., Білецький А.Я., Приставка О.П., Приставка П.О.* Статистична обробка даних. – К.: МІВВЦ, 2001. – 388 с.

4. *Приставка П.О.* Оцінка функції щільності розподілу двох змінних поліноміальним сплайном на основі В-сплайнів // Актуальні проблеми автоматизації та інформаційних технологій.– Д.: Навч. кн. – 2001.– Т.5.– С. 3–12.

5. *Лигун А.А., Шумейко А.А.* Асимптотические методы восстановления кривых. – К.: ІМ НАН України, 1996. - 358 с.

6. *Білецький А.Я., Приставка П.О., Фоменко Г.В.* Оцінка вірогідності непараметричного відтворення функції щільності розподілу ймовірностей двох змінних // Вісн. НАУ. – 2001. –№4(11). – С. 121–126.

7. *Білецький А.Я., Приставка П.О.* Наближення функцій двох змінних поліноміальним сплайном на основі В-сплайнів третього порядку / Вісн. НАУ. – 2002. – №1(12). – С. 64–69.

Стаття надійшла до редакції 29.10.02.