Алгоритмізація обробки неоднорідних даних на основі двовимірного сплайн-нормального розподілу

ПРИСТАВКА О.П.,. МАЦУГА О.М

Дніпропетровський національний університет

Розроблено обчислювальні схеми відтворення запропонованого двовимірного сплайн-нормального розподілу з однією та двома прямими склеювання, паралельними осям спостережень. Алгоритмізовано процес обробки неоднорідних даних і класифікації на основі даного розподілу.

Разработаны вычислительные схемы восстановления предложенного двумерного сплайн-нормального распределения с одной и двумя прямыми склеивания, параллельными осям наблюдений. Алгоритмизирован процесс обработки неоднородных данных и классификации на основе данного распределения.

The computing schemes of restoring of two-dimensional spline - normal distribution with one and two straight line of pasting together, parallel axes of observations are developed. The algorithm of processing of the non-uniform data and classifications is created on the basis of the given distribution.

Постановка проблеми. У задачах класифікації та діагностики для підвищення точності розпізнавання пропонується двовимірний сплайн-нормальний розподіл, в описі якого вже міститься поділяюча пряма. Актуальною для задач діагностики представляється модель двовимірного сплайн-нормального розподілу з прямими склеювання, паралельними осям спостережень, оскільки такі прямі можна трактувати як межі, що відокремлюють різні стадії протікання процесу або явища.

Застосування зазначеного розподілу в автоматизованих системах класифікації та діагностики пов'язане із задачею відтворення розподілу за вхідним масивом даних, яка досі не знайшла свого повного розв'язку. Тому в роботі поставлено за мету розробити обчислювальну схему відтворення розподілу, на її основі алгоритмізувати процес обробки неоднорідних даних і сформулювати правила класифікації.

Аналіз досліджень і публікацій. Матеріали роботи ґрунтуються на проведеному авторами дослідженні властивостей двовимірного сплайн-нормального розподілу [1]. В [1] для двовимірної випадкової величини $\vec{\zeta}=(\xi_1,\xi_2)$, на реалізації якої мають вплив різні зовнішні фактори, вводяться функції розподілу

– у вигляді двовимірного сплайн-нормального розподілу з однією прямою склеювання

$$F\left(x_1,x_2;\vec{\Theta}_1^{cn2}\right)=\begin{cases}\mathrm{N}\left(x_1,x_2;\vec{\Theta}_1^{n2}\right), & x_1(x_2)\le C,\\ \mathrm{N}\left(x_1,x_2;\vec{\Theta}_2^{n2}\right), & x_1(x_2)\ge C,\end{cases}$$

де $\vec{\Theta}_1^{cn2}=\left\{\vec{\Theta}_1^{n2},\vec{\Theta}_2^{n2},C\right\}$ – вектор параметрів розподілу; $\mathrm{N}\left(x_1,x_2;\vec{\Theta}_s^{n2}\right)$ – функція розподілу двовимірного нормального закону з параметрами $\vec{\Theta}_s^{n2}=\left\{m_{1,s},m_{2,s},\sigma_{1,s},\sigma_{2,s},r_s\right\}$, $s=1,2$; $x_1(x_2)=C$ – пряма склеювання, у точках якої функція $F\left(x_1,x_2;\vec{\Theta}_1^{cn2}\right)$ неперервна; $C=const$;

– у вигляді двовимірного сплайн-нормального розподілу з двома прямими склеювання

$$F\left(x_1,x_2;\vec{\Theta}_2^{cn2}\right)=\begin{cases}\mathrm{N}\left(x_1,x_2;\vec{\Theta}_1^{n2}\right), & x_1(x_2)\le C_1, x_2(x_1)\le C_2,\\ \mathrm{N}\left(x_1,x_2;\vec{\Theta}_2^{n2}\right), & x_1(x_2)\le C, x_2(x_1)\ge C_2,\\ \mathrm{N}\left(x_1,x_2;\vec{\Theta}_3^{n2}\right), & x_1(x_2)\ge C, x_2(x_1)\le C_2,\\ \mathrm{N}\left(x_1,x_2;\vec{\Theta}_4^{n2}\right), & x_1(x_2)\ge C, x_2(x_1)\ge C_2,\end{cases}$$

де $\vec{\Theta}_2^{cn2}=\left\{\vec{\Theta}_1^{n2},\vec{\Theta}_2^{n2},\vec{\Theta}_3^{n2},\vec{\Theta}_4^{n2},C_1,C_2\right\}$ – вектор параметрів розподілу; $\vec{\Theta}_s^{n2}=\left\{m_{1,s},m_{2,s},\sigma_{1,s},\sigma_{2,s},r_s\right\}$, $s=\overline{1,4}$; $x_1(x_2)=C_1$, $x_2(x_1)=C_2$ – прямі склеювання, у точках яких функція $F\left(x_1,x_2;\vec{\Theta}_2^{cn2}\right)$ неперервна; $C_1,C_2=const$.

Проведені авторами дослідження властивостей двовимірного сплайн-нормального розподілу з однією та двома прямими склеювання, паралельними осям спостережень, дозволили визначити умови, за яких забезпечується неперервність відповідних функцій розподілу, та досліджено функції щільностей на модальність [1].

Для двовимірного сплайн-нормального розподілу з однією прямою склеювання, з урахуванням маргінальної властивості функції розподілу, встановлено, що одновимірна випадкова величина ξ_1, як складова $\vec{\zeta}$, розподілена за сплайн-нормальним законом з одним вузлом C і з параметрами $\vec{\Theta}_1^{cn1}=\left\{m_{1,1},\sigma_{1,1},\sigma_{1,2},C\right\}$:

$$F_1\left(x_1;\vec{\Theta}_1^{cn1}\right)=\begin{cases}\Phi\left(\left(x_1-m_{1,1}\right)/\sigma_{1,1}\right), & x_1\le C,\\ \Phi\left(\left(x_1-m_{1,1}-u_C\left(\sigma_{1,1}-\sigma_{1,2}\right)\right)/\sigma_{1,2}\right), & x_1\ge C,\end{cases} \quad (1)$$

де $\Phi(\bullet)$ – функція Лапласа; $u_C=\left(C-m_{1,1}\right)/\sigma_{1,1}$;

а ξ_2 – за нормальним законом з параметрами $\vec{\Theta}_2^{n1}=\left\{m_2,\sigma_2\right\}$:

$$F_2\left(x_2;\vec{\Theta}_2^{n1}\right)=\Phi\left(\left(x_2-m_2\right)/\sigma_2\right). \quad (2)$$

З умови неперервності функції $F\left(x_1,x_2;\vec{\Theta}_1^{cn2}\right)$ у точках прямої склеювання, встановлено: $m_{1,2}=m_{1,1}+u_C\left(\sigma_{1,1}-\sigma_{1,2}\right)$, $m_{2,1}=m_{2,2}$, $\sigma_{2,1}=\sigma_{2,2}$, $r_1=r_2$. У ході дослідження функції щільності $f\left(x_1,x_2;\vec{\Theta}_1^{cn2}\right)=\partial F\left(x_1,x_2;\vec{\Theta}_1^{cn2}\right)/\partial x_1\partial x_2$ зазначеного розподілу на модальність, визначено умови, за яких вона є одно- або двомодальна [1].

Тим самим, визначено функцію двовимірного сплайн-нормального розподілу з однією прямою склеювання у вигляді

$$F\left(x_1,x_2;\vec{\Theta}_1^{cn2}\right)=\begin{cases}\mathrm{N}\left(x_1,x_2;\vec{\Theta}_1^{n2}\right), & x_1(x_2)\le C,\\ \mathrm{N}\left(x_1,x_2;\vec{\Theta}_2^{n2}\right), & x_1(x_2)\ge C,\end{cases} \quad (3)$$

де $\vec{\Theta}_1^{cn2}=\left\{\vec{\Theta}_1^{n2},\vec{\Theta}_2^{n2},C\right\}$; $\vec{\Theta}_s^{n2}=\left\{m_{1,s},m_2,\sigma_{1,s},\sigma_2,r\right\}$,

$s=1,2$; $m_{1,2}=m_{1,1}+u_C(\sigma_{1,1}-\sigma_{1,2})$; $u_C=(C-m_{1,1})/\sigma_{1,1}$.

За аналогією визначено, що для сплайн-розподілу з двома прямими склеювання і ξ_1, і ξ_2 розподілені за одновимірним сплайн-нормальним законом з одним вузлом. Як вузли виступають C_1 та C_2 відповідно. Відповідні функції розподілів мають вигляд:

$$F_1\left(x_1;\vec{\Theta}_1^{cn1}\right)=\begin{cases}\Phi\left(\left(x_1-m_{1,1}\right)/\sigma_{1,1}\right), & x_1\le C_1,\\ \Phi\left(\left(x_1-m_{1,1}-u_{C_1}\left(\sigma_{1,1}-\sigma_{1,3}\right)\right)/\sigma_{1,3}\right), & x_1\ge C_1,\end{cases} \quad (4)$$

де $\vec{\Theta}_1^{cn1}=\{m_{1,1},\sigma_{1,1},\sigma_{1,3},C_1\}$ – вектор параметрів розподілу ξ_1; $u_{C_1}=(C_1-m_{1,1})/\sigma_{1,1}$

та

$$F_2\left(x_2;\vec{\Theta}_2^{cn1}\right)=\begin{cases}\Phi\left(\left(x_2-m_{2,1}\right)/\sigma_{2,1}\right), & x_2\le C_2,\\ \Phi\left(\left(x_2-m_{2,1}-u_{C_2}\left(\sigma_{2,1}-\sigma_{2,2}\right)\right)/\sigma_{2,2}\right), & x_2\ge C_2,\end{cases} \quad (5)$$

де $\vec{\Theta}_2^{cn1}=\{m_{2,1},\sigma_{2,1},\sigma_{2,2},C_2\}$ – вектор параметрів розподілу ξ_2; $u_{C_2}=(C_2-m_{2,1})/\sigma_{2,1}$.

Встановлено, що неперервність функції $F\left(x_1,x_2;\vec{\Theta}_2^{cn2}\right)$ у точках прямих склеювання має місце за таких умов: $m_{1,3}=m_{1,1}+u_{C_1}(\sigma_{1,1}-\sigma_{1,3})$, $m_{2,2}=m_{2,1}+u_{C_2}(\sigma_{2,1}-\sigma_{2,2})$, $m_{1,1}=m_{1,2}$, $m_{1,3}=m_{1,4}$, $\sigma_{1,1}=\sigma_{1,2}$, $\sigma_{1,3}=\sigma_{1,4}$, $r_1=r_2=r_3=r_4$. У процесі дослідження функції щільності $f\left(x_1,x_2;\vec{\Theta}_2^{cn2}\right)=\partial F\left(x_1,x_2;\vec{\Theta}_2^{cn2}\right)/\partial x_1\partial x_2$ даного розподілу на модальність встановлено, що вона може бути одно-, дво-, три- або чотиримодальна, і визначено умови, за яких досягаються дані кількості мод [1].

Тим самим, одержано функцію розподілу двовимірного сплайн-нормального закону у вигляді:

$$F\left(x_1,x_2;\vec{\Theta}_2^{cn2}\right)=\begin{cases}\mathrm{N}\left(x_1,x_2;\vec{\Theta}_1^{n2}\right), & x_1(x_2)\le C_1, x_2(x_1)\le C_2,\\ \mathrm{N}\left(x_1,x_2;\vec{\Theta}_2^{n2}\right), & x_1(x_2)\le C, x_2(x_1)\ge C_2,\\ \mathrm{N}\left(x_1,x_2;\vec{\Theta}_3^{n2}\right), & x_1(x_2)\ge C, x_2(x_1)\le C_2,\\ \mathrm{N}\left(x_1,x_2;\vec{\Theta}_4^{n2}\right), & x_1(x_2)\ge C, x_2(x_1)\ge C_2,\end{cases} \quad (6)$$

де $\vec{\Theta}_2^{cn2}=\left\{\vec{\Theta}_1^{n2},\vec{\Theta}_2^{n2},\vec{\Theta}_3^{n2},\vec{\Theta}_4^{n2},C_1,C_2\right\}$;

$\vec{\Theta}_1^{n2}=\{m_{1,1},m_{2,1},\sigma_{1,1},\sigma_{2,1},r\}$; $\vec{\Theta}_2^{n2}=\{m_{1,1},m_{2,2},\sigma_{1,1},\sigma_{2,2},r\}$;

$\vec{\Theta}_3^{n2}=\{m_{1,3},m_{2,1},\sigma_{1,3},\sigma_{2,1},r\}$; $\vec{\Theta}_4^{n2}=\{m_{1,3},m_{2,2},\sigma_{1,3},\sigma_{2,2},r\}$;

$m_{1,3}=m_{1,1}+u_{C_1}(\sigma_{1,1}-\sigma_{1,3})$; $u_{C_1}=(C_1-m_{1,1})/\sigma_{1,1}$;

$m_{2,2}=m_{2,1}+u_{C_2}(\sigma_{2,1}-\sigma_{2,2})$; $u_{C_2}=(C_2-m_{2,1})/\sigma_{2,1}$.

Базуючись на одержаних результатах, у даній роботі вирішується наступна задача.

Постановка задачі. Нехай реалізацію двовимірної випадкової величини $\vec{\zeta}=(\xi_1,\xi_2)$ задано масивом спостережень $\Omega_{2,N}=\left\{x_{1,i},x_{2,i},i=\overline{1,N}\right\}$. Для складових $\vec{\zeta}$, одновимірних випадкових величин ξ_1 та ξ_2, масив $\Omega_{2,N}$ трансформується в масиви $\Omega_{1,N}^1\bigcup\Omega_{1,N}^2=\Omega_{2,N}$, $\Omega_{1,N}^1\bigcap\Omega_{1,N}^2=\varnothing$, де $\Omega_{1,N}^1=\left\{x_{1,i},i=\overline{1,N}\right\}$, $\Omega_{1,N}^2=\left\{x_{2,i},i=\overline{1,N}\right\}$.

Необхідно розробити обчислювальні схеми відтворення функції двовимірного сплайн-нормального розподілу з однією прямою склеювання (3) та двома прямими склеювання (6) за даними масиву $\Omega_{2,N}$. На їх основі сформулювати алгоритм обробки неоднорідних даних та правила класифікації.

Виклад основного матеріалу. Проведені дослідження двовимірного сплайн-нормального розподілу з однією та двома прямими склеювання дозволили запропонувати обчислювальні схеми їх відтворення, подані процедурами 1 та 2 відповідно. Суть останніх полягає у визначенні оцінок параметрів $\vec{\zeta}$ як оцінок параметрів її складових ξ_1 та ξ_2. Оцінювання коефіцієнта кореляції пропонується проводити за результатами відтворення сплайн-регресійної залежності з одним вузлом, що пов'язано із неоднорідним характером реалізацій $\vec{\zeta}$.

Процедура 1. Відтворення функції двовимірного сплайн-нормального розподілу з однією прямою склеювання (3).

1. За вибірковими даними $\Omega_{1,N}^1$ знаходження оцінки $\hat{\vec{\Theta}}_1^{cn1}=\left\{\hat{m}_{1,1},\hat{\sigma}_{1,1},\hat{\sigma}_{1,2},\hat{C}\right\}$ як оцінки вектора параметрів одновимірного сплайн-нормального розподілу з одним вузлом (1) [2].

2. За вибірковими даними $\Omega_{1,N}^2$ знаходження оцінки $\hat{\vec{\Theta}}_2^{n1}=\{\hat{m}_2,\hat{\sigma}_2\}$ як оцінки вектора параметрів одновимірного нормального розподілу (2) [4].

3. Оцінювання коефіцієнта кореляції $\hat{r}$, враховуючи, що дані масиву $\Omega_{2,N}$ (зокрема $\Omega_{1,N}^1$) неоднорідні, проводиться на основі відтворення лінійної сплайн-регресійної залежності з одним вузлом [3]

$$x_2(x_1)=\begin{cases}a_1x_1+b_1, & x_1\le\hat{C},\\ a_2\left(x_1-\hat{C}\right)+a_1\hat{C}+b_1, & x_1\ge\hat{C}.\end{cases}$$

Знаходження оцінок $\hat{a}_1,\hat{b}_1,\hat{a}_2$ сплайн-регресії здійснюється згідно [3] з розв'язку СЛАР

$$\begin{Vmatrix}D_2 & D_1 & \hat{C}E_1\\ D_1 & D_0 & E_1\\ \hat{C}E_1 & E_1 & E_2\end{Vmatrix}\cdot\begin{Vmatrix}\hat{a}_1\\ \hat{b}_1\\ \hat{a}_2\end{Vmatrix}=\begin{Vmatrix}G_1\\ G_0\\ G_2\end{Vmatrix},$$

де $D_p=\sum_{i=1}^{d}x_{1,i}^p+\hat{C}^p(N-d)$, $p=\overline{0,2}$;

$E_p=\sum_{i=d+1}^{N}\left(x_{1,i}-\hat{C}\right)^p$, $p=1,2$; $G_0=\sum_{i=1}^{N}x_{2,i}$;

$G_1=\sum_{i=1}^{d}x_{1,i}x_{2,i}+\hat{C}\sum_{i=d+1}^{N}x_{2,i}$; $G_2=\sum_{i=d+1}^{N}x_{2,i}\left(x_{1,i}-\hat{C}\right)$;

$d=\sum_{i=1}^{N}J\left(x_{1,i},x_{2,i}\right)$; $J\left(x_{1,i},x_{2,i}\right)=\begin{cases}1, & x_{1,i}\le\hat{C}\\ 0, & x_{1,i}>\hat{C}\end{cases}$

Тоді оцінка $\hat{r}$ визначається

$$\hat{r}=\frac{\hat{r}_1d+\hat{r}_2(N-d)}{N}, \quad (7)$$

де $\hat{r}_1=\frac{\hat{a}_1\hat{\sigma}_{1,1}}{\hat{\sigma}_2}$; $\hat{r}_2=\frac{\hat{a}_2\hat{\sigma}_{1,2}}{\hat{\sigma}_2}$.

4. Призначення довірчих інтервалів для параметрів $m_{1,1}$, $\sigma_{1,1}$, $\sigma_{1,2}$, m_2, σ_2 з надійною імовірністю $1-\alpha$:

$$\hat{\theta} - t_{\alpha/2,\nu}\sqrt{D(\hat{\theta})} \le \theta \le \hat{\theta} + t_{\alpha/2,\nu}\sqrt{D(\hat{\theta})}, \qquad (8)$$

де θ – значення параметра, $\hat{\theta}$ – його оцінка, $t_{\alpha/2,\nu}$ – квантиль розподілу Стьюдента з кількістю степенів вільності $\nu = N-1$.

Визначення $D(\hat{m}_{1,1})$, $D(\hat{\sigma}_{1,1})$, $D(\hat{\sigma}_{1,2})$ здійснюється [2] з дисперсійно-коваріаційної матриці $DC = S^2(T'T)^{-1}$, де S^2 – залишкова дисперсія, з умови мінімуму якої знаходиться оцінка вектора $\hat{\hat{\Theta}}_1^{cn1}$ у п.1, T – матриця системи, що визначається з умови мінімуму S^2, T' – транспонована до неї.

Згідно [4] $D(\hat{m}_2) = \hat{\sigma}_2^2/N$, $D(\hat{\sigma}_2) = \hat{\sigma}_2^2/(2N)$.

Оцінку параметра $\hat{C}$ можна розглядати як квантиль порядку $\gamma = F_1(\hat{C})$. Враховуючи, що оцінка функції щільності $f_1\left(x_1; \hat{\hat{\Theta}}_1^{cn1}\right)$ розподілу ξ_1 має розрив першого роду при $x_1 = \hat{C}$, дисперсія оцінки $\hat{C}$ може бути визначена наступним чином: [4]:

$$D(\hat{C} \pm 0) = \frac{\gamma(1-\gamma)}{N\left(f_1\left(\hat{C} \pm 0; \hat{\hat{\Theta}}_1^{cn1}\right)\right)^2}.$$

Тоді інтервальне оцінювання C проводиться з імовірністю $1-\alpha$:

$$\hat{C} - t_{\alpha/2,\nu}\sqrt{D(\hat{C}-0)} \le C \le \hat{C} + t_{\alpha/2,\nu}\sqrt{D(\hat{C}+0)}. \qquad (9)$$

З урахуванням лінійного перетворення (7) та даних роботи [3], інтервальне оцінювання r здійснюється з надійною імовірністю $1-\alpha$:

$$\frac{\hat{r}_1^{н} d + \hat{r}_2^{н}(N-d)}{N} \le r \le \frac{\hat{r}_1^{в} d + \hat{r}_2^{в}(N-d)}{N},$$

де $\hat{r}_s^{н,в} = th\left(\frac{1}{2}\ln\frac{1+\hat{r}_s}{1-\hat{r}_s} \mp \frac{u_{\alpha/2}}{\sqrt{N-2}} - \frac{\hat{r}_s}{2(N-1)}\right)$, $s = 1,2$;

$u_{\alpha/2}$ – квантиль нормального розподілу.

Процедура 2. Відтворення функції двовимірного сплайн-нормального розподілу з двома прямими склеювання (6).

1. За вибірковими даними $\Omega_{1,N}^1$ знаходження оцінки $\hat{\hat{\Theta}}_1^{cn1} = \{\hat{m}_{1,1}, \hat{\sigma}_{1,1}, \hat{\sigma}_{1,3}, \hat{C}_1\}$ як оцінки вектора параметрів одновимірного сплайн-нормального розподілу з одним вузлом (4) [2].

2. За вибірковими даними $\Omega_{1,N}^2$ знаходження оцінки $\hat{\hat{\Theta}}_2^{cn1} = \{\hat{m}_{2,1}, \hat{\sigma}_{2,1}, \hat{\sigma}_{2,2}, \hat{C}_2\}$ як оцінки вектора параметрів одновимірного нормального розподілу з одним вузлом (5) [2].

3. Оцінювання коефіцієнта кореляції $\hat{r}$, з урахуванням неоднорідності даних масивів $\Omega_{1,N}^s$, $s = 1,2$, проводиться на основі відтворення двох лінійних сплайн-регресійних залежностей з одним вузлом [3]:

$$x_2(x_1) = \begin{cases} a_1 x_1 + b_1, & x_1 \le \hat{C}_1, \\ a_2(x_1 - \hat{C}_1) + a_1\hat{C}_1 + b_1, & x_1 \ge \hat{C}_1 \end{cases}$$

та

$$x_1(x_2) = \begin{cases} a_3 x_2 + b_3, & x_2 \le \hat{C}_2, \\ a_4(x_2 - \hat{C}_2) + a_3\hat{C}_2 + b_3, & x_2 \ge \hat{C}_2. \end{cases}$$

Оцінки $\hat{a}_1, \hat{b}_1, \hat{a}_2$ знаходяться з розв'язку СЛАР [3]

$$\begin{Vmatrix} D_2 & D_1 & \hat{C}_1 E_1 \\ D_1 & D_0 & E_1 \\ \hat{C}_1 E_1 & E_1 & E_2 \end{Vmatrix} \cdot \begin{Vmatrix} \hat{a}_1 \\ \hat{b}_1 \\ \hat{a}_2 \end{Vmatrix} = \begin{Vmatrix} G_1 \\ G_0 \\ G_2 \end{Vmatrix},$$

де $D_p = \sum_{i=1}^{d'} x_{1,i}^p + \hat{C}_1^p (N - d')$, $p = \overline{0,2}$;

$E_p = \sum_{i=d'+1}^{N} (x_{1,i} - \hat{C}_1)^p$, $p = 1,2$; $G_0 = \sum_{i=1}^{N} x_{2,i}$;

$G_1 = \sum_{i=1}^{d'} x_{1,i} x_{2,i} + \hat{C}_1 \sum_{i=d'+1}^{N} x_{2,i}$; $G_2 = \sum_{i=d'+1}^{N} x_{2,i}(x_{1,i} - \hat{C}_1)$;

$d' = \sum_{i=1}^{N} J'(x_{1,i}, x_{2,i})$; $J'(x_{1,i}, x_{2,i}) = \begin{cases} 1, & x_{1,i} \le \hat{C}_1 \\ 0, & x_{1,i} > \hat{C}_1 \end{cases}$

Аналогічно оцінки $\hat{a}_3, \hat{b}_3, \hat{a}_4$ знаходяться з розв'язку СЛАР

$$\begin{Vmatrix} D_2 & D_1 & \hat{C}_2 E_1 \\ D_1 & D_0 & E_1 \\ \hat{C}_2 E_1 & E_1 & E_2 \end{Vmatrix} \cdot \begin{Vmatrix} \hat{a}_3 \\ \hat{b}_3 \\ \hat{a}_4 \end{Vmatrix} = \begin{Vmatrix} G_1 \\ G_0 \\ G_2 \end{Vmatrix},$$

де $D_p = \sum_{i=1}^{d''} x_{2,i}^p + \hat{C}_2^p (N - d'')$, $p = \overline{0,2}$;

$E_p = \sum_{i=d''+1}^{N} (x_{2,i} - \hat{C}_2)^p$, $p = 1,2$; $G_0 = \sum_{i=1}^{N} x_{1,i}$;

$G_1 = \sum_{i=1}^{d''} x_{1,i} x_{2,i} + \hat{C}_2 \sum_{i=d''+1}^{N} x_{1,i}$; $G_2 = \sum_{i=d''+1}^{N} x_{1,i}(x_{2,i} - \hat{C}_2)$;

$d'' = \sum_{i=1}^{N} J''(x_{1,i}, x_{2,i})$; $J''(x_{1,i}, x_{2,i}) = \begin{cases} 1, & x_{2,i} \le \hat{C}_2 \\ 0, & x_{2,i} > \hat{C}_2 \end{cases}$

Тоді оцінка $\hat{r}$ визначається як

$$\hat{r} = \frac{\hat{r}_1 d' + \hat{r}_2(N-d') + \hat{r}_3 d'' + \hat{r}_4(N-d'')}{2N}, \qquad (10)$$

де $\hat{r}_1 = \frac{\hat{a}_1\hat{\sigma}_{1,1}}{\hat{\sigma}_2}$; $\hat{r}_2 = \frac{\hat{a}_2\hat{\sigma}_{1,3}}{\hat{\sigma}_2}$; $\hat{r}_3 = \frac{\hat{a}_3\hat{\sigma}_{2,1}}{\hat{\sigma}_1}$; $\hat{r}_4 = \frac{\hat{a}_4\hat{\sigma}_{2,2}}{\hat{\sigma}_1}$;

$\hat{\sigma}_s$ – оцінка середньоквадратичного відхилення за масивом $\Omega_{1,N}^s$, $s = 1,2$.

4. Визначення довірчих інтервалів для параметрів розподілу (за винятком C_1, C_2 та r) здійснюється з надійною імовірністю $1-\alpha$ згідно (8) При цьому $D(\hat{m}_{1,1})$, $D(\hat{\sigma}_{1,1})$, $D(\hat{\sigma}_{1,3})$, $D(\hat{m}_{2,1})$, $D(\hat{\sigma}_{2,1})$, $D(\hat{\sigma}_{2,2})$ визначаються згідно [2] з дисперсійно-коваріаційної матриці $DC = S^2(T'T)^{-1}$, де S^2 – залишкова дисперсія, з умови мінімуму якої знаходяться оцінки векторів параметрів у пп. 1 та 2; T – матриця системи, що визначається з умови мінімуму S^2; T' – транспонована до неї. Для параметрів $\hat{C}_1$, $\hat{C}_2$, що є квантилями порядку

$\gamma_1 = F_1\left(\hat{C}_1\right)$, $\gamma_2 = F_1\left(\hat{C}_2\right)$ відповідно, довірчий інтервал визначається за аналогією з (9). Межі довірчого інтервалу для параметра $\hat{r}$ визначається з надійною імовірністю $1-\alpha$ з урахуванням лінійного перетворення (10):

$$\hat{r}^{н,в} = \frac{\hat{r}_1^{н,в} d' + \hat{r}_2^{н,в}\left(N - d'\right) + \hat{r}_3^{н,в} d'' + \hat{r}_4^{н,в}\left(N - d''\right)}{2N},$$

де $\hat{r}_s^{н,в} = th\left(\frac{1}{2}\ln\frac{1+\hat{r}_s}{1-\hat{r}_s} \mp \frac{u_{\alpha/2}}{\sqrt{N-2}} - \frac{\hat{r}_s}{2(N-1)}\right)$, $s = \overline{1,4}$.

Адекватність запропонованих обчислювальних схем підтверджена шляхом реалізації обчислювального експерименту, який полягав у такому. Проводилося моделювання масивів обсягу N, розподілених за двовимірним сплайн-нормальним розподілом з подальшим відтворенням функцій розподілу. Результати відтворення порівнювались із теоретичними функціями шляхом перевірки гіпотези про рівність знайдених оцінок параметрів та значень параметрів на основі t-тесту ($t = \left(\hat{\theta} - \theta\right)/\sigma\left(\hat{\theta}\right)$, де θ – значення параметра; $\hat{\theta}$ – його оцінка; $\sigma\left(\hat{\theta}\right)$ – середньоквадратичне оцінки). Крім того, для перевірки вірогідності відтворення було застосовано критерій згоди χ^2-Пірсона. Експерименти з моделювання проводилися 100 разів та усереднювались. Суттєвим в ході реалізації критерію χ^2-Пірсона є вибір оптимальної кількості класів по кожній змінній. Відомі співвідношення для їх визначення [4] передбачають однорідність даних, яка за початковими умовами не має місця. Використовуючи результати проведеного дослідження функції щільності двовимірного сплайн-розподілу на модальність, у роботі при застосуванні критерію кількість класів по змінній, за якою відтворюється одновимірний сплайн-нормальний розподіл, збільшувалась удвічі. Нижче (табл. 1, 2) наведено результати двох експериментів.

Таблиця 1. **Результати обчислювального експерименту для сплайн-розподілу з однією прямою склеювання**

	Значення параметра	N=50		N=100		N=300		N=1000		N=2000	
		Оцінка параметра	t-статистика ($t_{\alpha/N}$ = 2,01)	Оцінка параметра	t-статистика ($t_{\alpha/N}$ = 1,98)	Оцінка параметра	t-статистика ($t_{\alpha/N}$ = 1,97)	Оцінка параметра	t-статистика ($t_{\alpha/N}$ = 1,96)	Оцінка параметра	t-статистика ($t_{\alpha/N}$ = 1,96)
$m_{1,1}$	100	99,5	0,077	99,1	0,182	100,0	0,012	100,0	-0,025	100,0	-0,019
m_2	75	75,1	-0,053	75,0	-0,001	75,1	-0,154	75,0	0,010	75,0	-0,116
$\sigma_{1,1}$	30	32,1	-0,290	30,2	-0,042	30,7	-0,246	30,1	-0,055	30,0	-0,010
$\sigma_{1,2}$	30	9,0	0,182	9,2	0,186	9,7	0,134	9,7	0,189	9,9	0,091
σ_2	15	14,9	0,049	14,9	0,073	14,9	0,106	15,0	0,047	15,0	-0,032
r	0,6	0,58	0,152	0,59	0,176	0,60	0,062	0,59	0,289	0,60	0,128
C	90	91,5	-0,495	91,1	-0,716	90,3	-0,484	90,1	-0,843	90,0	-0,188
χ^2		0,333		0,291		0,332		0,356		0,549	

Таблиця 2. **Результати обчислювального експерименту для сплайн-розподілу з двома прямими склеювання**

	Значення параметра	N=50		N=100		N=300		N=1000		N=2000	
		Оцінка параметра	t-статистика ($t_{\alpha/N}$ = 2,01)	Оцінка параметра	t-статистика ($t_{\alpha/N}$ = 1,98)	Оцінка параметра	t-статистика ($t_{\alpha/N}$ = 1,97)	Оцінка параметра	t-статистика ($t_{\alpha/N}$ = 1,96)	Оцінка параметра	t-статистика ($t_{\alpha/N}$ = 1,96)
$m_{1,1}$	150	151,0	-0,172	149,7	0,111	149,9	0,093	150,0	0,027	150,0	0,089
$m_{2,1}$	200	200,8	-0,158	200,0	0,002	200,0	-0,019	200,0	-0,001	200,0	0,091
$\sigma_{1,1}$	25	26,8	-0,311	25,4	-0,115	25,4	-0,194	25,0	0,015	25,0	0,002
$\sigma_{1,3}$	10	9,3	0,082	8,7	0,215	9,4	0,191	9,9	0,058	9,9	0,060
$\sigma_{2,1}$	30	32,3	-0,362	30,9	-0,213	30,5	-0,201	30,1	-0,077	30,0	0,023
$\sigma_{2,2}$	15	13,3	0,207	13,4	0,216	14,3	0,226	14,9	0,057	14,9	0,073
r	0,9	0,87	0,368	0,88	0,342	0,89	0,331	0,89	0,500	0,89	0,611
C_1	160	157,8	0,125	160,4	-0,048	160,0	0,008	160,0	0,086	159,9	0,181
C_2	205	204,5	0,021	205,8	-0,070	205,2	-0,060	205,0	0,024	205,0	0,033
χ^2		0,696		0,439		0,618		0,844		0,950	

З метою автоматизації обробки неоднорідних даних запропоновано загальний алгоритм відтворення двовимірного сплайн-нормального розподілу, який передбачає:

1. Ідентифікацію одновимірних нормальних розподілів за масивами $\Omega_{1,N}^1$, $\Omega_{1,N}^2$ на основі коефіцієнтів асиметрії та ексцесу [4].
2. Якщо за масивом $\Omega_{1,N}^1$ ідентифікований нормальний розподіл, то перехід на п.3, інакше – на п.4.
3. Якщо за масивом $\Omega_{1,N}^2$ ідентифікований нормальний розподіл, то відтворення двовимірного нормального розподілу і на п.7, інакше – на п.5.
4. Якщо за масивом $\Omega_{1,N}^2$ ідентифікований нормальний розподіл, то відтворення двовимірного сплайн-нормального розподілу з однією прямою склеювання за змінною x_2 за процедурою, аналогічною процедурі 1, і перехід на п.7. Інакше перехід на п.6.
5. Відтворення двовимірного сплайн-нормального розподілу з однією прямою склеювання (3) за процедурою 1 і перехід на п.7.
6. Відтворення двовимірного сплайн-нормального розподілу з двома прямими склеювання (6) за процедурою 2.
7. Кінець роботи алгоритму.

Знайдені на основі алгоритму оцінки прямих склеювання можна розглядати як поділяючі і, як наслідок, використовувати у задачі класифікації. Для двовимірного сплайн-нормального розподілу з однією прямою склеювання має місце одна лінійна поділяюча функція $x_1(x_2)=C$. Вирішальне правило класифікації має вигляд: об'єкт, заданий точкою $M(x_1,x_2)$, належить класу

$$I(x_1,x_2)=\begin{cases}1, & x_1(x_2)\le C,\\ 2, & x_1(x_2)>C,\end{cases}$$

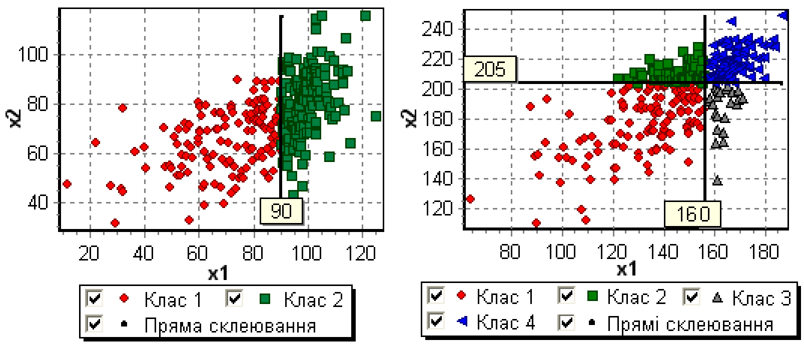
де $I(x_1,x_2)$ – індикатор класу.

Для двовимірного сплайн-нормального розподілу з двома прямими склеювання мають місце дві лінійні поділяючі функції $x_1(x_2)=C_1$, $x_2(x_1)=C_2$. І вирішальне правило має вигляд: об'єкт, заданий точкою $M(x_1,x_2)$, належить класу

$$I(x_1,x_2)=\begin{cases}1, & x_1(x_2)\le C_1, x_2(x_1)\le C_2,\\ 2, & x_1(x_2)\le C_1, x_2(x_1)>C_2,\\ 3, & x_1(x_2)>C_1, x_2(x_1)\le C_2,\\ 4, & x_1(x_2)>C_1, x_2(x_1)>C_2,\end{cases}$$

Сформульовані вирішальні правила застосовано до даних, одержаних шляхом імітаційного моделювання (рис. 1).

Запропоновані обчислювальні схеми увійшли до ядра розробленої автоматизованої системи "VerMed", призначеної для опрацювання медичних даних [5], і були апробовані на даних Кримського республіканського НДІ фізичних методів лікування та медичної кліматології ім. І.М. Сєченова.



а *б*

Рис. 1. Результати класифікації на основі двовимірного сплайн-нормального розподілу: *а* – з однією прямою склеювання; *б* – з двома прямими склеювання

Висновки

1. Розроблено процедури відтворення двовимірного сплайн-нормального розподілу з однією та двома прямими склеювання з перевіркою адекватності та вірогідності відтворення шляхом реалізації обчислювального експерименту. Сформульовано загальний алгоритм відтворення даного розподілу.
2. На основі часткових випадків двовимірного сплайн-нормального розподілу з однією та двома прямими склеювання визначено лінійні поділяючі прямі та сформульовано вирішальні правила класифікації.
3. Подальші розвідки в даному напрямку передбачають дослідження загальних випадків двовимірного сплайн-нормального розподілу з однією та двома прямими склеювання з метою формулювання вирішальних правил класифікації.

ЛІТЕРАТУРА

1. Приставка, О.П. Модель двовимірного сплайн-нормального розподілу та її властивості / О.П. Приставка, О.М. Мацуга // Актуальні проблеми автоматизації та інформаційних технологій. – 2008. – Т. 12.
2. Приставка, А.Ф. Смеси и сплайн-распределения на неоднородных нормальных пространствах / А.Ф. Приставка, О.В. Райко // Днепропетровский гос. университет. – Д., 1987. – 233 с. Деп. в ВИНИТИ 11.01.88, №33–В88.
3. Приставка, А.Ф. Вычислительные методы и программная среда корреляционного и регрессионного анализа / А.Ф. Приставка, А.И. Передерий, О.В. Райко, В.М. Остропицкий. – Д.: ДГУ, 1996. – 192 с.
4. Бабак, В.П. Статистична обробка даних / В.П. Бабак, А.Я. Білецький, О.П. Приставка, П.О. Приставка. – К.: МІВВЦ, 2001. – 388 с.