УДК 519.233.2:519.254

О.М. Мацуга, Ю.В. Оголь

Дніпропетровський національний університет імені Олеся Гончара

ОБЧИСЛЮВАЛЬНІ СХЕМИ ВІДТВОРЕННЯ СПЛАЙН-НОРМАЛЬНОГО РОЗПОДІЛУ НА ОСНОВІ МЕТОДУ МАКСИМАЛЬНОЇ ПРАВДОПОДІБНОСТІ

Запропоновано дві обчислювальні схеми відтворення сплайн-нормального розподілу з одним та двома вузлами на основі методу максимальної правдоподібності. Перша передбачає використання чисельних методів, друга – неперервного генетичного алгоритму.

Ключові слова: *функція максимальної правдоподібності, сплайн-нормальний розподіл, вузол, метод Ньютона, генетичний алгоритм.*

Предложены две вычислительные схемы восстановления сплайн-нормального распределения с одним и двумя узлами на основе метода максимального правдоподобия. Первая предполагает использование численных методов, вторая – непрерывного генетического алгоритма.

Ключевые слова: *функция максимального правдоподобия, сплайн-нормальное распределение, узел.*

Two computing schemes for restoration of spline-normal distribution with one and two nodes are offered. The schemes are based on maximum likelihood method. First of them uses calculus of approximations. Second of them uses the real-coded genetic algorithm.

Keywords: *maximum likelihood function, spline-normal distribution, spline-Weibull distribution.*

Постановка проблеми. У практиці обробки статистичних даних багато випадків, коли імовірнісний простір неоднорідний, і відповідно моделі класичних розподілів неадекватні. У такому разі для підвищення вірогідності висновків за вхідними даними доцільно застосовувати моделі сплайн-розподілів [1]. Обчислювальні схеми відтворення таких моделей здебільшого базуються на методі найменших квадратів. Зважаючи на властивості оцінок максимальної правдоподібності, актуальна задача побудови обчислювальних схем на основі методу максимальної правдоподібності. У роботі дана задача розглядається стосовно сплайн-нормального розподілу.

Аналіз останніх досліджень і публікацій. Моделі сплайн-розподілів для аналізу неоднорідних імовірнісних структур введені та обґрунтовані О.П. Приставкою [1]. Сплайн-нормальний розподіл

АКТУАЛЬНІ ПРОБЛЕМИ АВТОМАТИЗАЦІЇ ТА ІНФОРМАЦІЙНИХ ТЕХНОЛОГІЙ. Том 14. 2010

дозволив ефективно розв'язати ряд прикладних задач теорії надійності, медичної діагностики.

Під час заходження оцінки вектора параметрів сплайн-нормального розподілу широке розповсюдження здобули ітераційна, лінійна та робасна процедури, в основу яких покладено метод найменших квадратів [2; 3]. Застосування методу максимальної правдоподібності обмежено [2] внаслідок того, що він призводить до більш складних процедур, які потребують розв'язання систем трансцендентних рівнянь або застосування нелінійної оптимізації.

Постановка задачі. Нехай результати спостережень над одновимірною випадковою величиною ξ задано у вигляді вибірки $\left\{x_i; i=\overline{1,n}\right\}$, де n – кількість спостережень; x_i – спостережуване значення в i-му експерименті.

Необхідно за вибірковими даними відтворити однин з таких сплайн-розподілів:

– сплайн-нормальний з одним вузлом з функцією щільності

$$f\left(x;\vec{\Theta}\right)=\begin{cases}\dfrac{1}{\sigma_1\sqrt{2\pi}}\exp\left(-\dfrac{1}{2}\left(\dfrac{x-m_1}{\sigma_1}\right)^2\right), -\infty < x \le x_{0,1},\\ \dfrac{1}{\sigma_2\sqrt{2\pi}}\exp\left(-\dfrac{1}{2}\left(\dfrac{x-x_{0,1}}{\sigma_2}+\dfrac{x_{0,1}-m_1}{\sigma_1}\right)^2\right), x_{0,1} < x < +\infty,\end{cases} \quad (1)$$

де $\vec{\Theta}=\left\{m_1,\sigma_1,\sigma_2,x_{0,1}\right\}$;

– сплайн-нормальний з двома вузлами з функцією щільності

$$f\left(x;\vec{\Theta}\right)=\begin{cases}\dfrac{1}{\sigma_1\sqrt{2\pi}}\exp\left(-\dfrac{1}{2}\left(\dfrac{x-m_1}{\sigma_1}\right)^2\right), -\infty < x \le x_{0,1},\\ \dfrac{1}{\sigma_2\sqrt{2\pi}}\exp\left(-\dfrac{1}{2}\left(\dfrac{x-x_{0,1}}{\sigma_2}+\dfrac{x_{0,1}-m_1}{\sigma_1}\right)^2\right), x_{0,1} < x \le x_{0,2},\\ \dfrac{1}{\sigma_3\sqrt{2\pi}}\exp\left(-\dfrac{1}{2}\left(\dfrac{x-x_{0,2}}{\sigma_3}+\dfrac{x_{0,2}-x_{0,1}}{\sigma_2}+\dfrac{x_{0,1}-m_1}{\sigma_1}\right)^2\right), x_{0,2} < x < +\infty,\end{cases} \quad (2)$$

де $\vec{\Theta}=\left\{m_1,\sigma_1,\sigma_2,\sigma_3,x_{0,1},x_{0,2}\right\}$.

Ставиться задача: розробити та реалізувати обчислювальні схеми відтворення цих сплайн-розподілів на основі методу максимальної правдоподібності.

Основний матеріал. Пропоновані обчислювальні схеми передбачають знаходження оцінок параметрів $\hat{\theta}_1, \hat{\theta}_2, \ldots, \hat{\theta}_p$ за умови, що оцінки вузлів $\hat{x}_{0,1}, \hat{x}_{0,2}, \ldots, \hat{x}_{0,k}$ відомі. Припускається, що вузли співпадають з варіантами варіаційного ряду, а їх місцезнаходження визначається шляхом перебору на основі вирішального правила. Тоді оцінки параметрів $\hat{\theta}_1, \hat{\theta}_2, \ldots, \hat{\theta}_p$ можуть бути визначені методом максимальної правдоподібності.

Коли вузли склеювання співпадають з варіантами ($\hat{x}_{0,s} = x_{n_s}$, $s = \overline{1,k}$), функція максимальної правдоподібності сплайн-розподілу має вигляд

$$L(\vec{\theta}) = \sum_{s=1}^{k+1} \sum_{i=n_{s-1}}^{n_s} \ln f_s(x_i), \tag{3}$$

де $\vec{\theta} = (\theta_1, \ldots, \theta_p)$; n_s – номер варіанти, з якою співпадає s-й вузол склеювання, $s = \overline{1,k}$; $n_0 = 1$; $n_{k+1} = n$.

Для сплайн-розподілів (1) та (2) функція (3) набуває виду:

– для сплайн-нормального розподілу з одним вузлом (1):

$$L(\vec{\theta}) = \sum_{i=1}^{n_1} \ln f_1(x_i) + \sum_{i=n_1+1}^{n} \ln f_2(x_i) = -\frac{n}{2}\ln(2\pi) - n_1 \ln\sigma_1 - \\ -\frac{1}{2\sigma_1^2}\sum_{i=1}^{n_1}(x_i - m_1)^2 - (n-n_1)\ln\sigma_2 - \frac{1}{2}\sum_{i=n_1+1}^{n}\left(\frac{x_i - x_{n_1}}{\sigma_2} + \frac{x_{n_1} - m_1}{\sigma_1}\right)^2, \tag{4}$$

де $\vec{\theta} = (m_1, \sigma_1, \sigma_2)$;

– для сплайн-нормального розподілу з двома вузлами (2):

$$L(\vec{\theta}) = \sum_{i=1}^{n_1} \ln f_1(x_i) + \sum_{i=n_1+1}^{n_2} \ln f_2(x_i) + \sum_{i=n_2+1}^{n} \ln f_3(x_i) = -\frac{n}{2}\ln(2\pi) - \\ -n_1\ln\sigma_1 - \frac{1}{2\sigma_1^2}\sum_{i=1}^{n_1}(x_i - m_1)^2 - (n_2 - n_1)\ln\sigma_2 - \frac{1}{2}\sum_{i=n_1+1}^{n_2}\left(\frac{x_i - x_{n_1}}{\sigma_2} + \frac{x_{n_1} - m_1}{\sigma_1}\right)^2 - \\ -(n-n_2)\ln\sigma_3 - \frac{1}{2}\sum_{i=n_2+1}^{n}\left(\frac{x_i - x_{n_2}}{\sigma_3} + \frac{x_{n_2} - x_{n_1}}{\sigma_2} + \frac{x_{n_1} - m_1}{\sigma_1}\right)^2, \tag{5}$$

де $\vec{\theta} = (m_1, \sigma_1, \sigma_2, \sigma_3)$.

Перша з пропонованих обчислювальних схем передбачає

знаходження оцінок $\hat{\theta}_1, \hat{\theta}_2, \ldots, \hat{\theta}_p$ за фіксованого вузла з необхідної умови максимуму функції правдоподібності, яка приводить до розв'язування системи рівнянь

$$\frac{\partial L}{\partial \theta_1} = 0, \quad \frac{\partial L}{\partial \theta_2} = 0, \quad \ldots, \quad \frac{\partial L}{\partial \theta_p} = 0,$$

яка для функцій правдоподібності (4) та (5) набуває вигляду:

– для функції (4)

$$\begin{cases} \dfrac{\partial L}{\partial m_1} = \dfrac{1}{\sigma_1^2}\sum\limits_{i=1}^{n_1}(x_i - m_1) + \dfrac{1}{\sigma_1}\sum\limits_{i=n_1+1}^{n}\left(\dfrac{x_i - x_{n_1}}{\sigma_2} + \dfrac{x_{n_1} - m_1}{\sigma_1}\right) = 0, \\ \dfrac{\partial L}{\partial \sigma_1} = -\dfrac{n_1}{\sigma_1} + \dfrac{1}{\sigma_1^3}\sum\limits_{i=1}^{n_1}(x_i - m_1)^2 + \dfrac{x_{n_1} - m_1}{\sigma_1^2}\sum\limits_{i=n_1+1}^{n}\left(\dfrac{x_i - x_{n_1}}{\sigma_2} + \dfrac{x_{n_1} - m_1}{\sigma_1}\right) = 0, \\ \dfrac{\partial L}{\partial \sigma_2} = -\dfrac{n - n_1}{\sigma_2} + \dfrac{1}{\sigma_2^2}\sum\limits_{i=n_1+1}^{n}\left(\dfrac{x_i - x_{n_1}}{\sigma_2} + \dfrac{x_{n_1} - m_1}{\sigma_1}\right)(x_i - x_{n_1}) = 0; \end{cases} \tag{6}$$

– для функції (5)

$$\begin{cases} \dfrac{\partial L}{\partial m_1} = \dfrac{1}{\sigma_1^2}\sum\limits_{i=1}^{n_1}(x_i - m_1) + \dfrac{1}{\sigma_1}\sum\limits_{i=n_1+1}^{n_2}\left(\dfrac{x_i - x_{n_1}}{\sigma_2} + \dfrac{x_{n_1} - m_1}{\sigma_1}\right) + \\ \qquad + \dfrac{1}{\sigma_1}\sum\limits_{i=n_2+1}^{n}\left(\dfrac{x_i - x_{n_2}}{\sigma_3} + \dfrac{x_{n_2} - x_{n_1}}{\sigma_2} + \dfrac{x_{n_1} - m_1}{\sigma_1}\right) = 0, \\ \dfrac{\partial L}{\partial \sigma_1} = -\dfrac{n_1}{\sigma_1} + \dfrac{1}{\sigma_1^3}\sum\limits_{i=1}^{n_1}(x_i - m_1)^2 + \dfrac{x_{n_1} - m_1}{\sigma_1^2}\sum\limits_{i=n_1+1}^{n_2}\left(\dfrac{x_i - x_{n_1}}{\sigma_2} + \dfrac{x_{n_1} - m_1}{\sigma_1}\right) + \\ \qquad + \dfrac{x_{n_1} - m_1}{\sigma_1^2}\sum\limits_{i=n_2+1}^{n}\left(\dfrac{x_i - x_{n_2}}{\sigma_3} + \dfrac{x_{n_2} - x_{n_1}}{\sigma_2} + \dfrac{x_{n_1} - m_1}{\sigma_1}\right) = 0, \\ \dfrac{\partial L}{\partial \sigma_2} = -\dfrac{n_2 - n_1}{\sigma_2} + \dfrac{1}{\sigma_2^2}\sum\limits_{i=n_1+1}^{n_2}\left(\dfrac{x_i - x_{n_1}}{\sigma_2} + \dfrac{x_{n_1} - m_1}{\sigma_1}\right)(x_i - x_{n_1}) + \\ \qquad + \dfrac{x_{n_2} - x_{n_1}}{\sigma_2^2}\sum\limits_{i=n_2+1}^{n}\left(\dfrac{x_i - x_{n_2}}{\sigma_3} + \dfrac{x_{n_2} - x_{n_1}}{\sigma_2} + \dfrac{x_{n_1} - m_1}{\sigma_1}\right) = 0, \\ \dfrac{\partial L}{\partial \sigma_3} = -\dfrac{n - n_2}{\sigma_3} + \dfrac{1}{\sigma_3^2}\sum\limits_{i=n_2+1}^{n}\left(\dfrac{x_i - x_{n_2}}{\sigma_3} + \dfrac{x_{n_2} - x_{n_1}}{\sigma_2} + \dfrac{x_{n_1} - m_1}{\sigma_1}\right)(x_i - x_{n_2}) = 0; \end{cases} \tag{7}$$

Трансцендентні системи рівнянь (6) та (7) можуть бути розв'язані будь-яким чисельним методом. У роботі пропонується використання методу випадкового пошуку з подальшим застосуванням методу

Ньютона для уточнення оцінок параметрів. Метод випадкового пошуку передбачає випадковий вибір оцінок параметрів з області їх визначення. Згідно методу Ньютона оцінки параметрів визначають ітераційно за формулою

$$\vec{\theta}'^{(l+1)} = \vec{\theta}'^{(l)} - W^{-1}\left(\vec{\theta}^{(l)}\right) f\left(\vec{\theta}^{(l)}\right), \tag{8}$$

де l – номер ітерації; знак «$'$» позначає транспонування вектора; $f\left(\vec{\theta}\right) = \left(\frac{\partial L}{\partial \theta_1}, \frac{\partial L}{\partial \theta_2}, \ldots, \frac{\partial L}{\partial \theta_p}\right)'$; $W\left(\vec{\theta}\right)$ – матриця Якобі

$$W\left(\vec{\theta}\right) = \begin{pmatrix} \frac{\partial^2 L}{\partial \theta_1^2} & \frac{\partial^2 L}{\partial \theta_1 \partial \theta_2} & \cdots & \frac{\partial^2 L}{\partial \theta_1 \partial \theta_p} \\ \frac{\partial^2 L}{\partial \theta_2 \partial \theta_1} & \frac{\partial^2 L}{\partial \theta_2^2} & \cdots & \frac{\partial^2 L}{\partial \theta_2 \partial \theta_p} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^2 L}{\partial \theta_p \partial \theta_1} & \frac{\partial^2 L}{\partial \theta_p \partial \theta_2} & \cdots & \frac{\partial^2 L}{\partial \theta_p^2} \end{pmatrix};$$

$W^{-1}\left(\vec{\theta}\right)$ – обернена до матриці Якобі.

Елементи матриці $W\left(\vec{\theta}\right)$ мають наступний вигляд для розподілів:

– сплайн-нормального з одним вузлом

$$\frac{\partial^2 L}{\partial m_1^2} = -\frac{n}{\sigma_1^2}, \quad \frac{\partial^2 L}{\partial m_1 \partial \sigma_2} = \frac{\partial^2 L}{\partial \sigma_2 \partial m_1} = -\frac{1}{\sigma_1 \sigma_2^2} \sum_{i=n_1+1}^{n} \left(x_i - x_{n_1}\right),$$

$$\frac{\partial^2 L}{\partial m_1 \partial \sigma_1} = \frac{\partial^2 L}{\partial \sigma_1 \partial m_1} = -\frac{2}{\sigma_1^3} \sum_{i=1}^{n_1} \left(x_i - m_1\right) - \frac{1}{\sigma_1^2} \sum_{i=n_1+1}^{n} \left(\frac{x_i - x_{n_1}}{\sigma_2} + 2\frac{x_{n_1} - m_1}{\sigma_1}\right),$$

$$\frac{\partial^2 L}{\partial \sigma_1^2} = \frac{n_1}{\sigma_1^2} - \frac{3}{\sigma_1^4} \sum_{i=1}^{n_1} \left(x_i - m_1\right)^2 - \frac{x_{n_1} - m_1}{\sigma_1^3} \sum_{i=n_1+1}^{n} \left(2\frac{x_i - x_{n_1}}{\sigma_2} + 3\frac{x_{n_1} - m_1}{\sigma_1}\right),$$

$$\frac{\partial^2 L}{\partial \sigma_1 \partial \sigma_2} = \frac{\partial^2 L}{\partial \sigma_2 \partial \sigma_1} = -\frac{x_{n_1} - m_1}{\sigma_1^2 \sigma_2^2} \sum_{i=n_1+1}^{n} \left(x_i - x_{n_1}\right),$$

$$\frac{\partial^2 L}{\partial \sigma_2^2} = \frac{n - n_1}{\sigma_2^2} - \frac{3}{\sigma_2^4} \sum_{i=n_1+1}^{n} \left(x_i - x_{n_1}\right)^2 - \frac{2\left(x_{n_1} - m_1\right)}{\sigma_1 \sigma_2^3} \sum_{i=n_1+1}^{n} \left(x_i - x_{n_1}\right);$$

– сплайн-нормального з двома вузлами

$$\frac{\partial^2 L}{\partial m_1^2} = -\frac{n}{\sigma_1^2}, \quad \frac{\partial^2 L}{\partial m_1 \partial \sigma_3} = \frac{\partial^2 L}{\partial \sigma_3 \partial m_1} = -\frac{1}{\sigma_1 \sigma_3^2} \sum_{i=n_2+1}^{n} \left(x_i - x_{n_2}\right),$$

$$\frac{\partial^2 L}{\partial m_1 \partial \sigma_1} = \frac{\partial^2 L}{\partial \sigma_1 \partial m_1} = -\frac{2}{\sigma_1^3} \sum_{i=1}^{n_1} \left(x_i - m_1\right) - \frac{1}{\sigma_1^2} \sum_{i=n_1+1}^{n_2} \left(\frac{x_i - x_{n_1}}{\sigma_2} + 2\frac{x_{n_1} - m_1}{\sigma_1}\right) -$$

$$-\frac{1}{\sigma_1^2} \sum_{i=n_2+1}^{n} \left(\frac{x_i - x_{n_2}}{\sigma_3} + \frac{x_{n_2} - x_{n_1}}{\sigma_2} + 2\frac{x_{n_1} - m_1}{\sigma_1}\right),$$

$$\frac{\partial^2 L}{\partial m_1 \partial \sigma_2} = \frac{\partial^2 L}{\partial \sigma_2 \partial m_1} = -\frac{1}{\sigma_1 \sigma_2^2} \left(\sum_{i=n_1+1}^{n_2} \left(x_i - x_{n_1}\right) + \left(n - n_2\right)\left(x_{n_2} - x_{n_1}\right)\right),$$

$$\frac{\partial^2 L}{\partial \sigma_1^2} = \frac{n_1}{\sigma_1^2} - \frac{3}{\sigma_1^4} \sum_{i=1}^{n_1} \left(x_i - m_1\right)^2 - \frac{x_{n_1} - m_1}{\sigma_1^3} \sum_{i=n_1+1}^{n_2} \left(2\frac{x_i - x_{n_1}}{\sigma_2} + 3\frac{x_{n_1} - m_1}{\sigma_1}\right) -$$

$$-\frac{x_{n_1} - m_1}{\sigma_1^3} \sum_{i=n_2+1}^{n} \left(2\frac{x_i - x_{n_2}}{\sigma_3} + 2\frac{x_{n_2} - x_{n_1}}{\sigma_2} + 3\frac{x_{n_1} - m_1}{\sigma_1}\right),$$

$$\frac{\partial^2 L}{\partial \sigma_1 \partial \sigma_2} = \frac{\partial^2 L}{\partial \sigma_2 \partial \sigma_1} = -\frac{x_{n_1} - m_1}{\sigma_1^2 \sigma_2^2} \left(\sum_{i=n_1+1}^{n_2} \left(x_i - x_{n_1}\right) + \left(n - n_2\right)\left(x_{n_2} - x_{n_1}\right)\right),$$

$$\frac{\partial^2 L}{\partial \sigma_1 \partial \sigma_3} = \frac{\partial^2 L}{\partial \sigma_3 \partial \sigma_1} = -\frac{x_{n_1} - m_1}{\sigma_1^2 \sigma_3^2} \sum_{i=n_2+1}^{n} \left(x_i - x_{n_2}\right),$$

$$\frac{\partial^2 L}{\partial \sigma_2^2} = \frac{n_2 - n_1}{\sigma_2^2} - \frac{1}{\sigma_2^3} \sum_{i=n_1+1}^{n_2} \left(3\frac{x_i - x_{n_1}}{\sigma_2} + 2\frac{x_{n_1} - m_1}{\sigma_1}\right)\left(x_i - x_{n_1}\right) -$$

$$-\frac{x_{n_2} - x_{n_1}}{\sigma_2^3} \sum_{i=n_2+1}^{n} \left(2\frac{x_i - x_{n_2}}{\sigma_3} + 3\frac{x_{n_2} - x_{n_1}}{\sigma_2} + 2\frac{x_{n_1} - m_1}{\sigma_1}\right),$$

$$\frac{\partial^2 L}{\partial \sigma_2 \partial \sigma_3} = \frac{\partial^2 L}{\partial \sigma_3 \partial \sigma_2} = -\frac{x_{n_2} - x_{n_1}}{\sigma_2^2 \sigma_3^2} \sum_{i=n_2+1}^{n} \left(x_i - x_{n_2}\right),$$

$$\frac{\partial^2 L}{\partial \sigma_3^2} = \frac{n - n_2}{\sigma_3^2} - \frac{1}{\sigma_3^3} \sum_{i=n_2+1}^{n} \left(3\frac{x_i - x_{n_2}}{\sigma_3} + 2\frac{x_{n_2} - x_{n_1}}{\sigma_2} + 2\frac{x_{n_1} - m_1}{\sigma_1}\right)\left(x_i - x_{n_2}\right).$$

Пропонована обчислювальна схема відтворення сплайн-розподілів подана процедурою 1.

Процедура 1.

1. Формують варіаційний ряд $\left\{x_i; i = \overline{1, n}\right\}$, де $x_i \le x_{i+1}$.

2. Вважаючи, що вузли $\hat{x}_{0,s}$, $s=\overline{1,k}$ збігаються з варіантами ряду

$$\hat{x}_{0,1}=x_{n_1}, \hat{x}_{0,2}=x_{n_2}, \ldots, \hat{x}_{0,k}=x_{n_k},$$

де $n_1=\overline{4,n-4-(k-1)}$; $n_2=\overline{n_1+1,n-4-(k-2)}$, …, $n_k=\overline{n_{k-1}+1,n-4}$,

обчислюють оцінку вектора $\hat{\vec{\theta}}^{(t)}$ для кожного варіанта t місцезнаходження вузлів шляхом розв'язання системи (6) або (7) залежно від того, який розподіл відтворюється. Для цього знаходять початкові оцінки параметрів випадковим чином з області визначення параметрів. Далі початкові оцінки уточнюють за допомогою методу Ньютона, користуючись рівністю (8) і враховуючи, що деякі оцінки параметрів мають бути додатні.

3. Визначають місцезнаходження вузлів склеювання з умови

$$\max_t L\left(\hat{\vec{\theta}}^{(t)}\right),$$

приписані їм оцінки параметрів вважають шуканими $\hat{\theta}_1=\hat{\theta}_1^{(t)},\ldots,\hat{\theta}_p=\hat{\theta}_p^{(t)}$.

4. Обчислюють дисперсій оцінок параметрів $D\{\hat{\theta}_j\}$, $j=\overline{1,p}$ з дисперсійно-коваріаційної матриці

$$DC=-I^{-1},$$

де I – інформаційна матриця, тотожна наведеній вище матриці Якобі

$$I=\begin{pmatrix} \frac{\partial^2 L}{\partial\theta_1^2} & \frac{\partial^2 L}{\partial\theta_1\partial\theta_2} & \cdots & \frac{\partial^2 L}{\partial\theta_1\partial\theta_p} \\ \frac{\partial^2 L}{\partial\theta_2\partial\theta_1} & \frac{\partial^2 L}{\partial\theta_2^2} & \cdots & \frac{\partial^2 L}{\partial\theta_2\partial\theta_p} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^2 L}{\partial\theta_p\partial\theta_1} & \frac{\partial^2 L}{\partial\theta_p\partial\theta_2} & \cdots & \frac{\partial^2 L}{\partial\theta_p^2} \end{pmatrix},$$

та здійснюють довірче оцінювання параметрів згідно співвідношення

$$\hat{\theta}_j-t_{\alpha/2,\nu}\sqrt{D\{\hat{\theta}_j\}}\le\theta_j\le\hat{\theta}_j+t_{\alpha/2,\nu}\sqrt{D\{\hat{\theta}_j\}},\ j=\overline{1,p}.$$

Друга з пропонованих обчислювальних схем передбачає знаходження оцінок $\hat{\theta}_1,\hat{\theta}_2,\ldots,\hat{\theta}_p$ за фіксованого вузла шляхом пошуку максимуму функції (3) за допомогою неперервного генетичного алгоритму [4]. Функцією пристосованості є функція (3). У роботі

застосовано варіант генетичного алгоритму, поданий процедурою 2.

Процедура 2.

1. Формують випадковим чином популяцію розв'язків $\left\{\vec{\theta}^{(1)},\ldots,\vec{\theta}^{(b)}\right\}$, яку складають вектори дійсних чисел $\vec{\theta}^{(j)}=(\theta_1,\ldots\theta_p)$.

2. Формують з поточної популяції нову, яку використовують на наступних пунктах 3–5. Для цього:

   2.1. Знаходять r розв'язків поточної популяції, яким відповідають максимальні значення функції пристосованості (3). Вони утворюють нову популяцію $\left\{\vec{\theta}'^{(1)},\vec{\theta}'^{(2)},\ldots,\vec{\theta}'^{(r)}\right\}$.

   2.2. До нової популяції також додають g розв'язків поточної, які задовольняють умову

$$\max_{j=1,m-r}\ \min_{c=1,r}\sqrt{\left(\hat{\vec{\theta}}^{(j)}-\hat{\vec{\theta}}'^{(c)}\right)^2}.$$

Тим самим одержують $\left\{\vec{\theta}'^{(1)},\ldots,\vec{\theta}'^{(r+g)}\right\}$.

3. Здійснюють схрещення розв'язків популяції $\left\{\vec{\theta}'^{(1)},\ldots,\vec{\theta}'^{(r+g)}\right\}$ по типу «кожен з кожним». Схрещення двох розв'язків $\vec{\theta}'^{(j_1)}$ та $\vec{\theta}'^{(j_2)}$ здійснюється за допомогою оператора

$$Child=0{,}5\vec{\theta}'^{(j_1)}+0{,}5\vec{\theta}'^{(j_2)}.$$

Усі елементи, що були утворені внаслідок схрещення, долучають до популяції, одержуючи популяцію $\left\{\vec{\theta}'^{(1)},\ldots,\vec{\theta}'^{((r+g)(r+g-1)/2)}\right\}$.

4. З імовірністю p_m виконують мутацію розв'язків $\vec{\theta}'^{(j)}$

$$\vec{\theta}'^{(j)}=\vec{\theta}'^{(j)}\pm C,$$

де константа $C>0$ визначається випадковим чином з області значень відповідного параметра; знак «±» обирається випадковим чином.

5. Обчислюють пристосованість популяції як суму пристосованості кожного розв'язка. Якщо пристосованість популяції не змінювалась протягом E ітерацій, закінчують роботу процедури, інакше повертаються на пункт 2, вважаючи популяцію $\left\{\vec{\theta}'^{(1)},\ldots,\vec{\theta}'^{((r+g)(r+g-1)/2)}\right\}$ поточною.

Пропонована обчислювальна схема відтворення сплайн-розподілів на основі генетичного алгоритму подана процедурою 3.

Процедура 3.

1. Формують варіаційний ряду $\{x_i ; i=\overline{1,n}\}$, де $x_i \le x_{i+1}$.

2. Вважаючи, що вузли $\hat{x}_{0,s}$, $s=\overline{1,k}$ збігаються з варіантами ряду

$$\hat{x}_{0,1} = x_{n_1}, \hat{x}_{0,2} = x_{n_2}, \ldots, \hat{x}_{0,k} = x_{n_k},$$

де $n_1 = \overline{4, n-4-(k-1)}$; $n_2 = \overline{n_1+1, n-4-(k-2)}$, …, $n_k = \overline{n_{k-1}+1, n-4}$,

обчислюють оцінки $\hat{\hat{\theta}}^{(t)}$ для кожного варіанта t місцезнаходження вузлів шляхом пошуку максимуму функцій (4) або (5) залежно від того, який розподіл відтворюється. Задачу оптимізації розв'язують генетичним алгоритмом, що поданий процедурою 2.

3. Визначають місцезнаходження вузлів склеювання з умови

$$\max_t L\left(\hat{\hat{\theta}}^{(t)}\right),$$

приписані їм оцінки параметрів вважають шуканими $\hat{\theta}_1 = \hat{\theta}_1^{(t)}, \ldots, \hat{\theta}_p = \hat{\theta}_p^{(t)}$.

4. Обчислюють дисперсій оцінок параметрів $D\{\hat{\theta}_j\}$, $j=\overline{1,p}$ та здійснюють довірче оцінювання параметрів як у процедурі 1.

Запропоновані обчислювальні схеми відтворення сплайн-нормального розподілу з одним та двома вузлами реалізовано у програмному забезпеченні Odnomer.

Адекватність обчислювальних схем підтверджено результатами обчислювальних експериментів на даних імітаційного моделювання. Нижче наведено результати експерименту, під час якого моделювалась вибірка обсягом $n=500$ зі сплайн-нормального розподілу з одним вузлом (табл. 1).

Таблиця 1

Результати обчислювального експерименту

Параметр	Значення під час моделювання	Процедура 1		Процедура 3	
		Оцінка	Дисперсія	Оцінка	Дисперсія
m_1	100	99,8	0,24	99,8	0,27
σ_1	10	9,5	0,18	9,9	0,23
σ_2	30	30,8	2,07	29,8	2,07
$x_{0,1}$	105	105,2	–	104,8	–

Нижче подано результати експерименту, в ході якого моделювалась вибірка обсягом $n=300$ зі сплайн-нормального розподілу з двома вузлами (табл. 2).

Таблиця 2

Результати обчислювального експерименту

Параметр	Значення під час моделювання	Процедура 1		Процедура 3	
		Оцінка	Дисперсія	Оцінка	Дисперсія
m_1	100	98,3	2,10	99,6	1,68
σ_1	15	16,6	3,11	15,2	2,69
σ_2	60	63,4	2,87	55,0	4,12
σ_3	25	23,0	2,64	26,2	3,48
$x_{0,1}$	90	87,9	–	91,3	–
$x_{0,2}$	135	138,8	–	138,9	–

Порівняння процедур 1 та 3 свідчить, що знайдені за їх допомогою оцінки параметрів дуже близькі та адекватні параметрам моделювання. Згідно критеріїв згоди Пірсона та Колмогорова за обома процедурами відтворення вірогідне. Перевагою процедури 1 є вища швидкість роботи.

Обчислювальні схеми також були апробовані на даних Українського державного НДІ медико-соціальних проблем інвалідності під час визначення межі «норма-патологія».

Висновки. Запропоновано та реалізовано обчислювальні схеми відтворення сплайн-нормального розподілу з одним та двома вузлами, в основу яких покладено метод максимальної правдоподібності. Результати тестування на даних імітаційного моделювання та реальних медичних даних підтвердили адекватність та вірогідність запропонованих схем.

Бібліографічні посилання

1. **Приставка О.П.** Сплайн-розподіли у статистичному аналізі / О.П. Приставка. – Д., 1995. – 152 с.
2. **Приставка А.Ф.** Смеси и сплайн-распределения на неоднородных нормальных пространствах / А.Ф. Приставка, О.В. Райко – Д., 1987. – 233 с. – Деп. в ВИНИТИ 11.01.88, №33–В88.
3. **ОСТ 54 30008-82.** Прикладная статистика. Правила определения оценок, доверительных границ параметров и теоретической функции сплайн-нормального распределения с одним узлом. – Введ. МГА 1.01.83. – 36 с. – Группа Т59.
4. **Паклин Н.** Непрерывные генетические алгоритмы – математический аппарат / Н. Паклин. – Режим доступу: http://www.basegroup.ru/library/optimization/real_coded_ga/.

Надійшла до редколегії 30.06.10