**УДК 519.233.2:519.254**

О.М. Мацуга, Г.С. Шубіна

*Дніпропетровський національний університет ім. Олеся Гончара*

# ОБЧИСЛЮВАЛЬНІ СХЕМИ ІДЕНТИФІКАЦІЇ СПЛАЙН-РОЗПОДІЛІВ ЗА ЙМОВІРНІСНИМ ПАПЕРОМ

**Удосконалено запропоновані обчислювальні схеми автоматизованої ідентифікації сплайн-розподілів за ймовірнісним папером. Здійснено їх тестування на даних імітаційного моделювання з нормального, логарифмічно нормального, експоненціального та Вейбулла сплайн-розподілів.**

**Ключові слова:** *ідентифікація, ймовірнісний папір, обчислювальна схема, сплайн-розподіл, вузол склеювання.*

**Усовершенствованы предложенные вычислительные схемы автоматизированной идентификации сплайн-распределений по вероятностной бумаге. Проведено их тестирование на данных имитационного моделирования из нормального, логарифмически нормального, экспоненциального и Вейбулла сплайн-распределений.**

**Ключевые слова:** *идентификация, вероятностная бумага, вычислительная схема, сплайн-распределение, узел склеивания.*

**The computing schemes of spline-distributions automated identification using probability paper are improved. They are tested on modeling data from normal, logarithmic normal, exponential and Weibull spline-distributions.**

**Key words:** *identification, probability paper, computing scheme, spline-distribution, node.*

**Постановка проблеми.** В інформаційному забезпеченні обробки статистичних даних актуальна проблема ідентифікації моделі розподілу за вибірковими даними. Один з найбільш потужних засобів ідентифікації є аналіз імовірнісного паперу, який робиться візуально. У роботі [1] авторами запропоновані обчислювальні схеми автоматизованої ідентифікації сплайн-розподілів за ймовірнісним папером. Їх ідея полягає в автоматизованому пошуку вузлів склеювання сплайн-розподілу на ймовірнісному папері. Проте результати тестування виявили, що запропоновані обчислювальні схеми інколи призводять до ідентифікації сплайн-розподілу із завеликою кількістю вузлів. Дана робота націлена на подолання цього недоліку. Крім того, запропоновані в [1] обчислювальні схеми відпрацьовані лише на даних з нормальних сплайн-розподілів. Актуальною також представляється задача їх апробації на даних зі сплайн-розподілів і з інших родин.

**Аналіз останніх досліджень і публікацій.** Модель сплайн-розподілу введена та обґрунтована професором О.П. Приставкою [2]. Вона може бути візуально ідентифікована за ймовірнісним папером, на якому «чистому розподілу» відповідає пряма, а сплайн-розподілу – ламана лінія [2]. Вузли, в яких ламана змінює кут нахилу, відповідають вузлам склеювання сплайн-розподілу.

Нехай задано вибірку $\left\{x_i; i=\overline{1,N}\right\}$, за якою побудовано варіаційний ряд $\left\{x_i, n_i, p_i; i=\overline{1,r}\right\}$, де $x_1 < x_2 < \ldots < x_r$; $r$ – кількість варіант; $x_i$ – значення $i$-ї варіанти; $n_i$ – частота $i$-ї варіанти; $\sum\limits_{i=1}^{r} n_i = N$; $p_i = \dfrac{n_i}{N}$ – відносна частота $i$-ї варіанти; $\sum\limits_{i=1}^{r} p_i = 1$. У кожній варіанті розраховано значення емпіричної функції розподілу $F_N\left(x_i\right) = \sum\limits_{j=1}^{i} p_j$, $i=\overline{1,r}$.

Для ідентифікації сплайн-розподілу з родини розподілів $F\left(x\right)$ за вибіркою $\left\{x_i; i=\overline{1,N}\right\}$ будується ймовірнісний папір. З цією метою здійснюється перетворення функції $F\left(x\right)$ для надання їй лінійного вигляду. Справедливі такі перетворення для родин розподілів:

– нормального $\varphi\left(F\left(x\right)\right) = \Phi^{-1}\left(F\left(x\right)\right)$, $\varphi\left(x\right) = x$, де $\Phi^{-1}\left(\bullet\right)$ – обернена до функції Лапласа;
– логарифмічно нормального: $\varphi\left(F\left(x\right)\right) = \Phi^{-1}\left(F\left(x\right)\right)$, $\varphi\left(x\right) = \ln x$;
– експоненціального $\varphi\left(F\left(x\right)\right) = \ln\dfrac{1}{1-F\left(x\right)}$, $\varphi\left(x\right) = x$;
– Вейбулла $\varphi\left(F\left(x\right)\right) = \ln\ln\dfrac{1}{1-F\left(x\right)}$, $\varphi\left(x\right) = \ln x$.

На ймовірнісному папері відображається лінеаризована емпірична функція розподілу $F_N(x)$ у вигляді масиву $\left\{ y_i = \varphi(x_i), z_i = \phi\left(F_N(x_i)\right); i = \overline{1,r} \right\}$.

У роботі [1] авторами запропоновано три обчислювальні схеми автоматизованого пошуку вузлів склеювання сплайн-розподілу на ймовірнісному папері.

**Обчислювальна схема 1.** *Пошук вузлів склеювання сплайн-розподілу на основі вимірювання кутів нахилу регресій.*

Фіксується варіанта $x_t$ варіаційного ряду і формуються масиви

$$\left\{ y_i, z_i; i = \overline{t-l,t} \right\}, \tag{1}$$

$$\left\{ y_i, z_i; i = \overline{t,t+l} \right\}, \tag{2}$$

за якими відтворюються відповідні лінійні регресії:

$$z = k_1 y + b_1, \tag{3}$$

$$z = k_2 y + b_2. \tag{4}$$

Параметри $k_1$ та $k_2$ цих моделей являють собою тангенси кутів нахилу регресій до вісі абсцис, тому в якості різниці між кутами нахилу регресій можна ввести показник

$$K = \left| arctg\left|k_2\right| - arctg\left|k_1\right| \right|. \tag{5}$$

Ті варіанти, в яких значення показника $K$ максимальні, доцільно вважати вузлами склеювання сплайн-розподілу.

1. Задати довжину «плеча» $l$, яка визначає довжини масивів (1) і (2).
2. Для кожної варіанти $x_t$, $t = \overline{l,r-l}$:

   2.1. Знайти оцінки параметрів $\hat{k}_1$ та $\hat{k}_2$ лінійних регресії (3) та (4) на основі масивів (1) та (2) відповідно, скориставшись методом найменших квадратів [3].

   2.2. Обчислити чергове значення показника $K$ за формулою (5), і тим самим, сформувати масив $\left\{ (x_t, K_t); t = \overline{l,r-l} \right\}$.

3. Здійснити згладжування (наприклад, медіанне) даних масиву $\left\{ (x_t, K_t); t = \overline{l,r-l} \right\}$ з метою вилучення шумів (опціонально).
4. Знайти вузли склеювання сплайн-розподілу шляхом визначення варіант $x_t$, $t = \overline{l,r-l}$, яким відповідають максимальні елементи масиву $\left\{ K_t; t = \overline{l,r-l} \right\}$. Максимальні елементи шукаються лише серед

тих, що перевищують заданий поріг. В якості порогу може бути використаний певний відсоток від максимального значення показника *K*, наприклад, значення $0,9K_{\max}$, де $K_{\max}$ – максимальне значення у масиві $\left\{K_t; t=\overline{l,r-l}\right\}$.

**Обчислювальна схема 2.** *Пошук вузлів склеювання сплайн-розподілу на основі порівняння двох регресійних прямих.*

Відмінність від попереднього випадку полягає в тому, що визначення наявності відмінності кутів нахилу регресій (3) та (4) зводиться до перевірки статистичної гіпотези $\mathrm{H}_0: k_1 = k_2$ за альтернативи $\mathrm{H}_1: k_1 \neq k_2$ на основі статистики [3]

$$K = \left| \frac{\hat{k}_1 - \hat{k}_2}{\sqrt{\dfrac{S^2_{\text{Зал},1}}{(l+1)S_1^2} + \dfrac{S^2_{\text{Зал},2}}{(l+1)S_2^2}}} \right|, \tag{6}$$

де $S^2_{\text{Зал},1}$ та $S_1^2$ обчислюються на основі масиву (1) як:

$$S^2_{\text{Зал},1} = \frac{1}{l+1}\sum_{i=t-l}^{t}\left(z_i - \hat{k}_1 y_i - \hat{b}_1\right)^2;$$

$$S_1^2 = \frac{1}{l+1}\sum_{i=t-l}^{t}\left(y_i - \bar{y}_1\right)^2; \qquad \bar{y}_1 = \frac{1}{l+1}\sum_{i=t-l}^{t} y_i;$$

$S^2_{\text{Зал},2}$ та $S_2^2$ розраховуються за масивом (2) як:

$$S^2_{\text{Зал},2} = \frac{1}{l+1}\sum_{i=t}^{t+l}\left(z_i - \hat{k}_2 y_i - \hat{b}_2\right)^2;$$

$$S_2^2 = \frac{1}{l+1}\sum_{i=t}^{t+l}\left(y_i - \bar{y}_2\right)^2; \qquad \bar{y}_2 = \frac{1}{l+1}\sum_{i=t}^{t+l} y_i.$$

Отже, обчислювальна схема 2 відрізняється від обчислювальної схеми 1 лише тим, що у пункті 2.2 чергове значення показника *K* розраховується за формулою (6).

**Обчислювальна схема 3.** *Пошук вузлів склеювання сплайн-розподілу на основі порівняння двох регресійних прямих, оцінених шляхом відтворення лінійної сплайн-регресії.*

У попередніх випадках регресії (3) та (4) на ймовірнісному папері будуються шляхом середньоквадратичного наближення до точок масивів (1) та (2), тому вони не обов'язково перетинаються. У даному випадку регресії будуються перетинними. Для реалізації цієї ідеї фіксується варіанта $x_t$ варіаційного ряду і розглядається масив

$$\left\{ y_i, z_i; i = \overline{t-l, t+l} \right\}, \tag{7}$$

за яким відтворюється лінійна сплайн-регресія вигляду

$$z = \begin{cases} k_1 y + b_1, & y \le y_t, \\ k_2 \left( y - y_t \right) + k_1 y_t + b_1, & y \ge y_t. \end{cases} \tag{8}$$

Параметри $k_1$ та $k_2$, як і у першій схемі, являють собою тангенси кутів нахилу регресій до вісі абсцис.

1. Задати довжину «плеча» $l$, яка визначає довжину масиву (7).

2. Для кожної варіанти $x_t$, $t = \overline{l, r-l}$:

   2.1. Обчислити оцінки параметрів $\hat{k}_1$ та $\hat{k}_2$ сплайн-регресії за масивом вигляду (7) [4].

   2.2. Обчислити чергове значення показника *K* за формулою (5), і тим самим, сформувати масив $\left\{ \left( x_t, K_t \right); t = \overline{l, r-l} \right\}$.

3. Провести згладжування (наприклад, медіанне) даних масиву $\left\{ \left( x_t, K_t \right); t = \overline{l, r-l} \right\}$ з метою вилучення шумів (опціонально).

4. Знайти вузли склеювання сплайн-розподілу шляхом визначення варіант $x_t$, $t = \overline{l, r-l}$, яким відповідають максимальні елементи масиву $\left\{ K_t; t = \overline{l, r-l} \right\}$. Максимальні елементи шукаються лише серед тих, що перевищують заданий поріг. Поріг може бути обраний як в обчислювальній схемі 1.

Як зазначалось вище, усі три обчислювальні схеми можуть знаходити надмірну кількість вузлів склеювання. Це пов'язано із тим, що на графіку залежності показника *K* від вузла склеювання може мати місце декілька близько розташованих локальних максимумів (рис. 1).

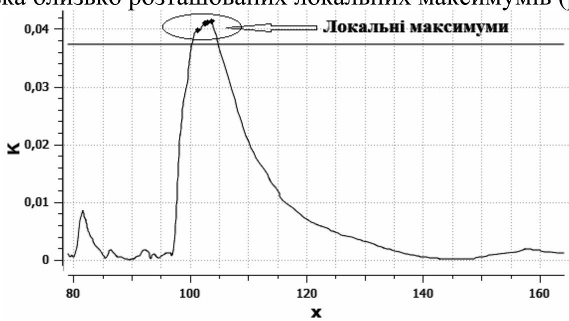


**Рис. 1. Графік залежності показника *K* від вузла склеювання**

Слід зазначити, що наявність пункту 3 (проведення згладжування) у кожній обчислювальній схемі націлена на подолання даної проблеми, проте не вирішує її повністю. Тому була поставлена наступна задача.

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

Ставиться задача ідентифікувати за цією вибіркою сплайн-розподіл шляхом автоматизованого визначення кількості його вузлів.

Для вирішення цієї задачі необхідно вдосконалити обчислювальні схеми 1–3 пошуку вузлів склеювання сплайн-розподілу на ймовірнісному папері, подолавши проблему знаходження декількох близько розташованих вузлів, та провести їх тестування на даних з нормального, логарифмічно нормального, експоненціального та Вейбулла сплайн-розподілів.

**Основний матеріал.** Пропонується модифікувати обчислювальні схеми 1–3, додавши до них ще один пункт, суть якого в наступному.

Нехай на основі пунктів 1–4 обчислювальних схем 1–3 знайдено оцінки вузлів склеювання $\left\{\hat{x}_{\text{в},i}; i=\overline{1,s}\right\}$. Для вилучення «зайвих» оцінок вузлів склеювання пропонуються два методи.

*1. Метод інтервального оцінювання квантилів*

Знайдені оцінки вузлів склеювання являють собою оцінки квантилів розподілу. Для кожного вузла на основі його оцінки будується довірчий інтервал як довірчий інтервал на квантиль. Якщо довірчі інтервали двох вузлів перетинаються, то залишається той вузол, для якого відповідне значення показника $K$ більше, інший – видаляється.

Інтервальне оцінювання здійснюється згідно з нерівністю [3]

$$\hat{x}_{\text{в},i} - t_{1-\alpha/2,v}\sqrt{D\left(\hat{x}_{\text{в},i}\right)} \le x_{\text{в},i} \le \hat{x}_{\text{в},i} + t_{1-\alpha/2,v}\sqrt{D\left(\hat{x}_{\text{в},i}\right)},$$

де $D\left(\hat{x}_{\text{в},i}\right)$ – дисперсія оцінки вузла

$$D\left(\hat{x}_{\text{в},i}\right) = \frac{\hat{F}\left(\hat{x}_{\text{в},i}\right)\left(1-\hat{F}\left(\hat{x}_{\text{в},i}\right)\right)}{N\hat{f}^2\left(\hat{x}_{\text{в},i}\right)};$$

$\hat{F}\left(\hat{x}_{\text{в},i}\right)$ та $\hat{f}\left(\hat{x}_{\text{в},i}\right)$ – значення оцінок функцій розподілу та щільності розподілу ймовірностей відповідно; $\alpha$ – імовірність «промаху» значення оцінки вузла повз довірчий інтервал (як правило, $\alpha = 0{,}05$).

Оцінювання функцій розподілу $\hat{F}(x)$ та щільності розподілу ймовірностей $\hat{f}(x)$ пропонується проводити за допомогою локальних поліноміальних сплайнів на основі *B*-сплайнів [5]. Їх перевага у високій швидкості роботи під час реалізації на ЕОМ та малій похибці апроксимації [5].

*2. Метод сусідів*

Суть методу полягає в тому, що попарно перевіряються знайдені оцінки вузли склеювання. Якщо для оцінок вузлів $\hat{x}_{\text{в},k}$ та $\hat{x}_{\text{в},l}$ різниця індексів $|k-l|$ менша за деякий заданий параметр $dist$, то такі оцінки вважаються близькими. Залишається той вузол, для якого відповідне значення показника *K* більше, інший – видаляється.

Запропоновані в роботі [1] обчислювальні схеми пошуку вузлів склеювання сплайн-розподілу та вищеописані модифікації цих схем було реалізовано у вигляді програмного забезпечення «SplineDistribution». Адекватність обчислювальних схем підтверджено результатами обчислювальних експериментів на даних імітаційного моделювання з нормального, логарифмічно нормального, експоненціального та Вейбулла сплайн-розподілів.

Нижче наводяться результати експерименту, під час якого моделювалась вибірка обсягом $n = 500$ зі сплайн-логнормального розподілу з одним вузлом з параметрами $x_0 = 0{,}8$, $m = 0{,}1$, $\sigma_1 = 0{,}5$, $\sigma_2 = 0{,}09$. На ймовірнісному папері логарифмічно нормального розподілу (рис. 2) чітко видно дві прямі з різними кутами нахилу, що відповідає сплайн-розподілу з одним вузлом.

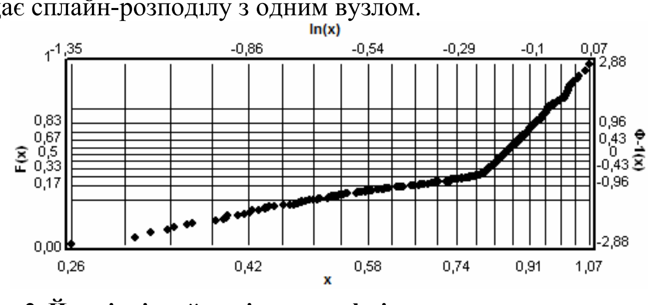


**Рис. 2. Ймовірнісний папір логарифмічно нормального розподілу в експерименті 1**

У таблиці 1 наведено результати ідентифікації за обчислювальними схемами на основі вимірювання кутів (ВК), порівняння двох регресійних прямих (ПРП), відтворення лінійної сплайн-регресії (ВСР) та модифікованими схемами (м. – застосоване медіанне згладжування, ІОК – інтервальне оцінювання квантилів, МС – метод сусідів).

*Таблиця 1*

**Ідентифіковані вузли сплайн-розподілу в експерименті 1**

| **Обчислювальна схема** | **Оцінки вузлів** | **Параметри обчислювальної схеми** |
|---|---|---|
| ВК | 0,8001 | Довжина «плеча» – 40, поріг – 90 % |
| ВК м. | 0,8013 | – |
| ВК ІОК | 0,8001 | – |
| ВК МС | 0,8001 | – |
| ВК м. ІОК | 0,8013 | – |
| ВК м. МС | 0,8013 | dist – 7 |
| ПРП | 0,8001 | Довжина «плеча» – 40, поріг – 90 % |
| ПРП м. | 0,8027 | – |
| ПРП ІОК | 0,8001 | – |
| ПРП МС | 0,8001 | – |
| ПРП м. ІОК | 0,8027 | – |
| ПРП м. МС | 0,8027 | dist – 7 |
| ВСР | 0,79; 0,8021; 0,8027 | Довжина «плеча» – 40, поріг – 90 % |
| ВСР м. | 0,8007 | – |
| ВСР ІОК | 0,79; 0,8021; 0,8027 | – |
| ВСР МС | 0,79 | dist – 7 |
| ВСР м. ІОК | 0,8007 | – |
| ВСР м. МС | 0,8007 | – |

Обчислювальні схеми на основі вимірювання кутів та порівняння двох регресійних прямих дозволили ідентифікувати один вузол склеювання, близький до заданого під час моделювання. При цьому не було потреби у застосуванні медіанного згладжування та інших модифікацій схем.

За обчислювальною схемою на основі відтворення лінійної сплайн-регресії ідентифіковано три близько розташовані вузли склеювання. Вирішити проблему надмірної кількості вузлів дозволило як застосування медіанного згладжування, так і методу сусідів. У кожному з випадків ідентифікувався один вузол склеювання.

При цьому всі ідентифіковані вузли склеювання за значенням дуже близькі до параметра моделювання. Найближчі значення оцінки вузла склеювання до параметра моделювання дала обчислювальна схема 2, що реалізує метод на основі порівняння двох регресійних прямих.

У цілому, результати чисельних експериментів засвідчили, що для якісної ідентифікації сплайн-розподілів з одним вузлом у більшості випадків достатньо базових обчислювальних схем 1 чи 2 або застосування медіанного згладжування. Для сплайн-розподілів з двома вузлами обов'язкове застосування методів інтервального оцінювання квантилів або сусідів.

Розглянемо результати експерименту, під час якого моделювалась вибірка обсягом $n = 300$ зі сплайн-нормального розподілу з двома вузлами з параметрами $x_0 = 95$, $x_1 = 120$, $m = 100$, $\sigma_1 = 10$, $\sigma_2 = 30$, $\sigma_3 = 10$. На ймовірнісному папері нормального розподілу (рис. 3) чітко виділяються три прямі з різними кутами нахилу, що відповідає випадку сплайн-розподілу з двома вузлами.

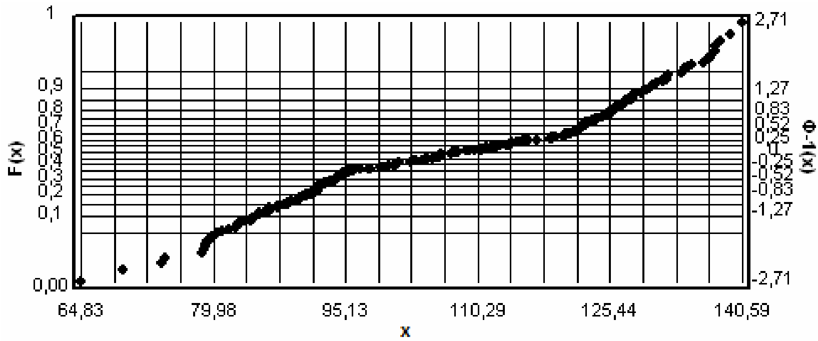


**Рис. 3. Ймовірнісний папір нормального розподілу в експерименті 2**

Результати ідентифікації вузлів сплайн-розподілу за різними методами для цього експерименту наведено у таблиці 2.

Усі три базові обчислювальні схеми ідентифікували надмірну кількість вузлів склеювання.

Застосування медіанного згладжування не виявилось ефективним для обчислювальних схем на основі вимірювання кутів та порівняння двох регресійних прямих. За його допомогою кількість ідентифікованих вузлів зменшилась, але не до потрібної кількості. Застосування методів інтервального оцінювання квантилів та сусідів як у сукупності з медіанним згладжуванням так і без нього дозволило одержати очікуваний результат – ідентифікувати два вузли склеювання, які мають місце в дійсності.

*Таблиця 2*

**Ідентифіковані вузли сплайн-розподілу в експерименті 2**

| **Обчислювальна схема** | **Оцінки вузлів** | **Параметри обчислювальної схеми** |
|---|---|---|
| ВК | 94,9; 95,0; 121,3; 122,0; 122,6; 123,4 | Довжина «плеча» – 50, поріг – 90 % |
| ВК м. | 94,9; 121,5; 122,5 | – |
| ВК ІОК | 95,0; 123,4 | – |
| ВК МС | 95,0; 123,4 | dist – 19 |
| ВК м. ІОК | 94,9; 122,5 | – |
| ВК м. МС | 94,9; 122,5 | dist – 13 |
| ПРП | 94,6; 95,0; 120,3; 121,3 | Довжина «плеча» – 50, поріг – 90 % |
| ПРП м. | 94,9; 120,9 | – |
| ПРП ІОК | 94,9; 121,3 | – |
| ПРП МС | 94,9; 121,3 | dist – 6 |
| ПРП м. ІОК | 94,9; 120,9 | – |
| ПРП м. МС | 94,9; 120,9 | – |
| ВСР | 94,3; 94,8; 121,5; 121,9; 122,0; 122,6; 123,4 | Довжина «плеча» – 50, поріг – 90 % |
| ВСР м. | 94,7; 122,3 | – |
| ВСР ІОК | 94,8; 123,4 | – |
| ВСР МС | 94,8; 123,4 | dist – 9 |
| ВСР м. ІОК | 94,7; 122,3 | – |
| ВСР м. МС | 94,7; 122,3 | – |

Для обчислювальної схеми 3 на основі відтворення лінійної сплайн-регресії застосування лише медіанного згладжування дозволило ідентифікувати два вузли склеювання.

Найближчі значення оцінок вузлів склеювання до параметрів моделювання дала обчислювальна схема 2.

З метою перевірки якості ідентифікації на основі запропонованих обчислювальних схем та їх модифікацій був проведений такий експеримент. Моделювалось по 100 вибірок обсягом 300 елементів з кожного сплайн-розподілу і за кожною вибіркою проводилась ідентифікація вузлів склеювання. Успіхом вважалась подія, коли ідентифікована така кількість вузлів, яка була задана під час моделювання. Результати експерименту (відсоток успіхів) наведено у таблиці 3. Під час експерименту використано такі параметри обчислювальних схем: довжина «плеча» – 40, поріг – 8 5 %, dist – 10.

*Таблиця 3*

**Відсоток випадків, коли ідентифікована правильна кількість вузлів сплайн-розподілів**

| Обчислювальна схема | Сплайн-розподіл з одним вузлом | Сплайн-розподіл з двома вузлами |
|---|---|---|
| ВК | 77,5 | 21,3 |
| ВК м. | 100,0 | 72,5 |
| ВК ІОК | 96,3 | 70,0 |
| ВК МС | 97,8 | 71,8 |
| ВК м. ІОК | 100,0 | 90,0 |
| ВК м. МС | 100,0 | 91,2 |
| ПРП | 91,3 | 25,0 |
| ПРП м. | 100,0 | 68,8 |
| ПРП ІОК | 98,8 | 75,0 |
| ПРП МС | 100,0 | 75,0 |
| ПРП м. ІОК | 100,0 | 86,3 |
| ПРП м. МС | 100,0 | 87,1 |
| ВСР | 33,7 | 6,3 |
| ВСР м. | 100,0 | 71,3 |
| ВСР ІОК | 65,0 | 61,3 |
| ВСР МС | 71,2 | 65,5 |
| ВСР м. ІОК | 100,0 | 87,5 |
| ВСР м. МС | 100,0 | 90,0 |

Дані таблиці 3 свідчать, що для сплайн-розподілів з одним вузлом базова обчислювальна схема 2 без жодних модифікацій дозволяє у більш ніж 90 % випадках ідентифікувати правильну кількість вузлів склеювання. Усі три обчислювальні схеми у сукупності з медіанним згладжуванням завжди забезпечують ідентифікацію правильної кількості вузлів склеювання. Отже, для сплайн-розподілів з одним вузлом немає потреби у застосуванні методів інтервального оцінювання квантилів та сусідів.

Для сплайн-розподілів з двома вузлами базові обчислювальні схеми є неефективні. Застосування медіанного згладжування частково покращує ситуацію. Усі обчислювальні схеми з медіанним згладжуванням дозволяють ідентифікувати правильну кількість вузлів склеювання приблизно у 70 % випадках, але це не можна вважати прийнятним результатом. Застосування обчислювальних схем без медіанного згладжування, але з методом інтервального оцінювання квантилів або сусідів також не забезпечує достатньо якісних результатів. Ефективним є застосування обчислювальних схем одночасно з медіанним згладжуванням та методом інтервального оцінювання квантилів або сусідів. У такому разі правильна кількість вузлів склеювання ідентифікується приблизно у 90 %, а у випадку застосування обчислювальної схеми 1 з медіанним згладжуванням та методом сусідів – у більш ніж 90 % випадках.

**Висновки.** Було вдосконалено обчислювальні схеми ідентифікації сплайн-розподілів на основі ймовірнісного паперу. Модифікація полягала у застосуванні до ідентифікованих вузлів склеювання методу інтервального оцінювання квантилів та методу сусідів для вилучення надмірних вузлів склеювання. Останній метод виявився дуже ефективним та зручним у реалізації та використанні.

Модифіковані обчислювальні схеми було додано до ядра програмного забезпечення «SplineDistribution».

Здійснено тестування модифікованих обчислювальних схем на даних імітаційного моделювання сплайн-розподілів з класів нормального, логарифмічно нормального, експоненціального та Вейбулла. Результати тестування засвідчили, що модифіковані обчислювальні схеми дозволяють у більш ніж у 90 % випадків ідентифікувати сплайн-розподіл із правильною кількістю вузлів.

У подальшому доцільним представляється дослідження впливу параметрів обчислювальних схем на результати ідентифікації.

### Бібліографічні посилання


1. **Мацуга О.М.** Обчислювальна схема ідентифікації розподілів з класу нормального / О.М. Мацуга, Г.С. Шубіна // Актуальні проблеми автоматизації та інформаційних технологій. – 2011. – Т.15. – С. 161– 172.
2. **Приставка О.П.** Сплайн-розподіли у статистичному аналізі / О.П. Приставка. – Д., 1995. – 152 с.
3. **Бабак В.П.** Статистична обробка даних / В.П. Бабак, А.Я. Білецький, О.П. Приставка, П.О. Приставка. – К., 2001. – 388 с.
4. **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., 1996. – 192 с.
5. **Приставка П.О.** Поліноміальні сплайни при обробці даних / П.О. Приставка. – Д., 2004. – 236 с.


*Надійшла до редколегії 07.05.2012*