Сплайн-регресія при обробці експериментальних даних

А.Є. ПОЛОНСЬКА, О.П.ПРИСТАВКА

Дніпропетровський національний університет

Запропоновано обчислювальну технологію апроксимації експериментальних даних усереднюючими сплайн-функціями..

Предложена вычислительная технология аппроксимации экспериментальных данных усредняющими сплайн-функциями.

The computing technology of approximation of experimental data averaging a spline-functions is offered.

Постановка проблеми. При обробці експериментальних даних виникає задача адекватного відображення фізико-хімічних процесів, що, як правило, є неоднорідними. Для обробки таких даних запропоновано використовувати сплайн-перетворення, що є найбільш прогресивними алгоритмами обробки даних. Теорія інтерполяційних та згладжуючи обчислювальних схем має на даний момент широкий розвиток. В останні роки було запропоновано локальні поліноміальні сплайни [1], що є розгалуженням усереднених сплайнів та мають стохастичну ваду.

Аналіз досліджень та постановка задачі. При обробці статистичних даних $\left\{x_i, y_i, i=\overline{1,N}\right\}$ реалізують як поліноміальні, так і квазілінійні функції регресії. Для отримання більш адекватної моделі регресії запропоновано реалізувати різного типу (інтерполяційного, згладжуючого) моделі сплайн-регресії. Як фізичного, так і математичного обгрунтування таким моделям в більшості випадків не подано. Останнє особливо відноситься до статистичних моделей сплайн-регресії, що потребують оцінки кількості вузлів, їх пошуку та точності при відтворенні регресії.

Вперше сплайн-функції в задачах статистичного аналізу представлена в роботі [2]. Застосування сплайнів в задачах екстраполяції, інтерполяції та згладжування представлено в роботах [3, 4].

Задача інтерполювання сплайнами $S(x)$ початкового масиву $\left\{x_i, y_i, i=\overline{1,N}\right\}$ формулюється як задача знаходження серед усіх $S(x)\in W_2^m[a,b]$ такої $S(x_i)=y_i$. Знаходження згладжуючого сплайну виконується із умови

$$\min_{S\in W_2^m}\left\{\sum_{i=1}^{N}\left(y_i - S\left(x_i,\vec{\theta}\right)\right)^2\right\} \tag{1}$$

Вважаючи, що мають місце нелінійні моделі регресії, що зводяться до лінійного типу, подальша задача полягає як до знаходження оцінок параметрів $\vec{\theta}$, так і у знаходженні вузлів точності наближення емпіричних даних лінійними сплайнами з реалізацією інтерполяційної процедури знаходження мінімуму залишкової дисперсії.

Виклад основного матеріалу. Обчислювальні схеми відтворення одновимірних сплайн-регресіїних моделей нелінійного типу, які наведено нижче, базуються на перетворенні останніх до лінійного типу.

Процедура відтворення нелінійних сплайн-регресійних залежностей зводиться до відтворення лінійних сплайн-регресійних моделей с подальшим визначенням коефіцієнтів нелінійної регресії.

Нехай початковий масив даних $\left\{x_i, y_i, i=\overline{1,N}\right\}$ перетворено в масив $\left\{t_i, z_i, i=\overline{1,N}\right\}$.

Лінійна сплайн-регрессія визначена у наступному вгляді:

$$S(t)=\begin{cases} b_1 t + b_0 & ,t^{(0)}\le t\le t^{(1)};\\ b_2 t + (b_1-b_2)t^{(1)} + b_0 & ,t^{(1)}\le t\le t^{(2)};\\ ------------------------ & \\ b_i t + \sum\limits_{s=1}^{i-1}(b_s - b_{s+1})t^{(s)} + b_0 & ,t^{(i-1)}\le t\le t^{(i)};\\ ------------------------ & \\ b_k t + \sum\limits_{s=1}^{k-1}(b_s - b_{s+1})t^{(s)} + b_0 & ,t^{(k-1)}\le t\le t^{(k)}; \end{cases} \tag{2}$$

Якщо накласти умову, що $S\left(t^i\right)=z^{(i)}$, то

$$b_i = \frac{z^{(i)} - z^{(i-1)}}{t^{(i)} - t^{(i-1)}} \tag{3}$$

$$S(t)=\frac{z^{(i)} - z^{(i-1)}}{t^{(i)} - t^{(i-1)}}\left(t - t^{(i-1)}\right) + z^{(i-1)},\ t^{(i-1)}\le t\le t^{(i)} \tag{4}$$

Тим самим, задача відтворення лінійної сплайн-регресії зводиться до інтерполювання лінійними сплайнами.

У загальному випадку $E\{S(t_i)\}=z_i,\quad i=\overline{1,N}$.

Якщо $N>k$ та для $t^{(i-1)}\le t_s\le t^{(i)}$ значення t_{p_i}, t_{q_i} нижньою та верхньою межами статистичних даних відрізка $\left[t^{(i-1)}, t^{(i)}\right]$, то з умови

$$\min_{z^{(i)},i=\overline{1,k}}\sum_{i=1}^{e}\sum_{e=p_i}^{q_i}\left(S(t_e) - z_e\right)^2 \tag{5}$$

З урахуванням виразу (3) отримуємо систему лінійних рівнянь:

$$
\begin{cases}
\sum\limits_{s=p_1}^{q_1}\left(\dfrac{t_s-t^{(0)}}{h_1}z^{(1)}+\dfrac{t^{(0)}-t_s}{h_1}z^{(0)}+z^{(0)}-z_s\right)\times \\
\times\left(\dfrac{t^{(0)}-t_s}{h_1}+1\right)=0; \\
\sum\limits_{s=p_1}^{q_1}\left(\dfrac{t_s-t^{(0)}}{h_1}z^{(1)}+\dfrac{t^{(0)}-t_s}{h_1}z^{(0)}+z^{(0)}-z_s\right)\times \\
\times\left(\dfrac{t_s-t^{(0)}}{h_1}+1\right)+\sum\limits_{s=p_2}^{q_2}\left(\dfrac{t_s-t^{(1)}}{h_2}z^{(2)}+\dfrac{t^{(1)}-t_s}{h_2}\times\right. \\
\left.\times z^{(1)}+z^{(1)}-z_s\right)\times\left(\dfrac{t^{(1)}-t_s}{h_2}+1\right)=0; \\
-------------------------- \\
\sum\limits_{s=p_i}^{q_i}\left(\dfrac{t_s-t^{(i-1)}}{h_i}z^{(i)}+\dfrac{t^{(i-1)}-t_s}{h_i}z^{(i-1)}+\right. \\
\left.+z^{(i-1)}-z_s\right)\left(\dfrac{t_s-t^{(i-1)}}{h_i}+1\right)+\sum\limits_{s=p_{i+1}}^{q_{i+1}}\left(\dfrac{t_s-t^{(i)}}{h_i}z^{(i+1)}+\right. \\
\left.+\dfrac{t^{(i)}-t_s}{h_i}z^{(i)}+z^{(i)}-z_s\right)\times\left(\dfrac{t^{(i)}-t_s}{h_i}+1\right)=0; \\
-------------------------- \\
\sum\limits_{s=p_k}^{q_k}\left(\dfrac{t_s-t^{(k-1)}}{h_k}z^{(k)}+\dfrac{t^{(k-1)}-t_s}{h_k}z^{(k-1)}+z^{(k-1)}-\right. \\
\left.-z_s\right)\times\left(\dfrac{t_s-t^{(k-1)}}{h_k}+1\right)=0
\end{cases}
\qquad (6)
$$

\

В матричному представленні система (6) має вигляд

$$
\begin{bmatrix}
a_{00} & a_{01}\dots a_{0k} \\
a_{10} & a_{11}\dots a_{1k} \\
-------- & \\
a_{i0} & a_{i1}\dots a_{ik} \\
-------- & \\
a_{k0} & a_{k1}\dots a_{kk}
\end{bmatrix}
\begin{bmatrix} z^{(0)} \\ z^{(1)} \\ \vdots \\ z^{(i)} \\ \vdots \\ z^{(k)} \end{bmatrix}
=
\begin{bmatrix} g_0 \\ g_1 \\ \vdots \\ g_i \\ \vdots \\ g_k \end{bmatrix}
\qquad (7)
$$

або

$$A\times Z=G, \qquad (8)$$

де

$$a_{e,e}=\sum_{s=p_e}^{q_e}\left(\frac{t_s-t^{(e-1)}}{h_e}\right)^2+\sum_{s=p_{e+1}}^{q_{e+1}}\left(\frac{t^{(e)}-t_s}{h_{e+1}}+1\right)^2,$$

$$e=\overline{1,k-1}$$

$$a_{0,0}=\sum_{s=p_1}^{q_1}\left(\frac{t^{(0)}-t_s}{h_1}\right)^2;\ a_{k,k}=\sum_{s=p_k}^{q_k}\left(\frac{t_s-t^{(k-1)}}{h_k}\right)^2;$$

$$a_{e+1,e}=\sum_{s=p_{e+1}}^{q_{e+1}}\left(\frac{t^{(e)}-t_s}{h_{e+1}}+1\right)\left(\frac{t_s-t^{(e)}}{h_{e+1}}\right),\ e=\overline{0,k-1}$$

$$a_{e-1,e}=\sum_{s=p_e}^{q_e}\left(\frac{t^{(e-1)}-t_s}{h_e}+1\right)\left(\frac{t_s-t^{(e-1)}}{h_e}\right),\ e=\overline{1,k}$$

$a_{r,e}=0$, якщо $r\ge e+2,\ e\ge r+2$;

$$g_e=\sum_{s=p_e}^{q_e}z_s\left(\frac{t_s-t^{(e-1)}}{h_e}\right)+\sum_{s=p_{e+1}}^{q_{e+1}}z_s\left(\frac{t^{(e)}-t_s}{h_{e+1}}+1\right),$$

$$e=\overline{1,k-1};$$

$$g_0=\sum_{s=p_1}^{q_1}z_s\left(\frac{t^{(0)}-t_s}{h_1}+1\right);\ g_k=\sum_{s=p_k}^{q_k}z_s\left(\frac{t_s-t^{(k-1)}}{h_k}\right);$$

$$h_e=t^{(e)}-t^{(e-1)}.$$

Окремим випадком є $h_e=h=const$.

З урахуванням випадкових складових значення z_s на проміжку $\left[t^{(e-1)},t^{(e)}\right]$ можна представити у вигляді

$$z_s=S(t_s)+\varepsilon_s=\frac{t_s-t^{(e-1)}}{h_e}z^{(e)}+\frac{t^{(e-1)}-t_s}{h_e}z^{(e-1)}+$$
$$+z^{(e-1)}+\varepsilon_e \qquad (9)$$

де $E\{\varepsilon_s\}=0,D\{\varepsilon_s\}=\sigma^2$.

З урахуванням виразу (9) умова (5) прийме наступний вигляд:

$$\min_{z^{(e)}}\sum_{i=1}^{k}\sum_{s=p_i}^{q_i}\left(z_s-\frac{t_s-t^{(e-1)}}{h_e}z^{(e)}-\frac{t^{(e-1)}-t_s}{h_e}z^{(e-1)}-\right.$$
$$\left.-z^{(e-1)}\right)^2=S^2_{çàë} \qquad (10)$$

Це еквівалентно розв'язку системи

$$A\times\hat{Z}=G, \qquad (11)$$

де $\hat{Z}^T=\left(\hat{z}^{(0)},\hat{z}^{(1)},\dots,\hat{z}^{(k)}\right)$.

Отримавши розв'язок системи (11), визначають вектор оцінок параметрів лінійної сплайн-регресійної залежності $\vec{\hat{Z}}$ та з урахуванням (3) – вектор B оцінок $\vec{\hat{B}}$. Спираючись на отримані оцінки $\vec{\hat{Z}}$, $\vec{\hat{B}}$, проводять обчислення оцінок параметрів нелінійної сплайн-регресійної залежності з урахуванням формул перерахунку.

З системи (11) випливає, що

$$E\{z_s\}=\frac{t_s-t^{(e-1)}}{h_e}z^{(e)}+\left(\frac{t^{(e-1)}-t_s}{h_e}+1\right)z^{(e-1)} \qquad (12)$$

та леми, що наведено нижче.

Лема 1.1.

$$E\{g_r\}=\sum_{s=0}^{k}a_{rs}z^{(s)}.$$

Доведення.

Дійсно,

$$E\{g_0\}=\sum_{s=p_1}^{q_1}\left(\frac{t^{(0)}-t_s}{h_1}+1\right)E\{z_s\}=$$

$$=z^{(0)}\sum_{s=p_1}^{q_1}\left(\frac{t^{(0)}-t_s}{h_1}+1\right)^2+z^{(1)}\left(\frac{t_s-t^{(0)}}{h_1}\right)\times$$

$$\times\left(\frac{t^{(0)}-t_s}{h_1}+1\right)=\sum_{s=0}^{k}a_{0s}z^{(s)};$$

$$E\{g_r\}=\sum_{s=p_r}^{q_r}\frac{t_s-t^{(r-1)}}{h_r}E\{z_s\}+$$

$$+\sum_{s=p_{r+1}}^{q_{r+1}}\left(\frac{t^{(r)}-t_s}{h_{r+1}}+1\right)E\{z_s\}=z^{(r)}\sum_{s=p_{r+1}}^{q_{r+1}}\left(\frac{t_s-t^{(r)}}{h_{r+1}}\right)^2+$$

$$+z^{(r-1)}\sum_{s=p_r}^{q_{r+1}}\left(\frac{t_s-t^{(r-1)}}{h_r}\right)\left(\frac{t^{(r-1)}-t_s}{h_r}+1\right)+z^{(r+1)}\times$$

$$\sum_{s=p_{r+1}}^{q_{r+1}}\left(\frac{t^{(r)}-t_s}{h_{r+1}}+1\right)^2,$$

або

$$E\{g_r\}=z^{(r-1)}\sum_{s=p_r}^{q_r}\left(\frac{t_s-t^{(r-1)}}{h_r}\right)\left(\frac{t^{(r-1)}-t_s}{h_r}+1\right)+$$

$$+z^{(r)}\left\{\sum_{s=p_r}^{q_r}\left(\frac{t_s-t^{(r-1)}}{h_r}\right)^2+\sum_{s=p_{r+1}}^{q_{r+1}}\left(\frac{t_s-t^{(r-1)}}{h_r}+1\right)\right\}+$$

$$+z^{(r+1)}\sum_{s=p_{r+1}}^{q_{r+1}}\left(\frac{t_s-t^{(r-1)}}{h_{r+1}}\right)\left(\frac{t^{(r-1)}-t_s}{h_{r+1}}+1\right)=\sum_{s=0}^{k}a_{rs}z^{(s)},$$

$$0<r<k \qquad (13)$$

$$E\{g_k\}=\sum_{s=p_k}^{q_k}\left(\frac{t_s-t^{(k-1)}}{h_k}\right)E\{z_s\}=z^{(k-1)}\times$$

$$\times\sum_{s=p_k}^{q_k}\left(\frac{t_s-t^{(k-1)}}{h_k}\right)\left(\frac{t^{(k-1)}-t_s}{h_k}+1\right)+z^{(k)}\times$$

$$\times\sum_{s=p_k}^{q_k}\left(\frac{t_s-t^{(k-1)}}{h_k}\right)^2=\sum_{s=0}^{k}a_{ks}z^{(s)}.$$

Тим самим лему доведено. Отже,

$$E\{g_r\}=\sum_{s=0}^{k}a_{rs}z^{(s)},\quad 0\le r\le s \qquad (14)$$

З урахуванням виразів (13), (14), має місце система

$$T\times Z=E\{G\} \qquad (15)$$

з якої випливає

$$Z=T^{-1}\times E\{G\}.$$

З іншого боку, з системи (11) з урахуванням $A=T$, отримують

$$\hat{Z}=T^{-1}\times E\{G\}.$$

Тоді математичне сподівання вектору $\hat{Z}$ має вигляд

$$E\{\hat{Z}\}=E\{T^{-1}\times G\}=T^{-1}\times E\{G\}=TT^{-1}\times Z=Z,$$

звідки випливає

$$E\{\hat{Z}^{(r)}\}=Z^{(r)},\quad r=0,1,2,\ldots,k.$$

<u>*Лема 1.2.*</u> Оцінкою коваріації вільних членів матриці G є

$$cov[g_r,g_e]=a_{re}\sigma^2.$$

Доведення.

Базуючись на визначенні коваріації, отримують

$$cov[g_r,g_e]=E\{\{g_r-E\{g_r\}\}\{g_e-E\{g_e\}\}\}=E\left\{\sum_{s=p_r}^{q_r}z_s\times\right.$$

$$\times\left(\frac{t_s-t^{(r-1)}}{h_r}\right)+\sum_{s=p_{r+1}}^{q_{r+1}}z_s\left(\frac{t^{(r)}-t_s}{h_{r+1}}+1\right)-\sum_{s=p_r}^{q_r}(z_s-\varepsilon_s)\times$$

$$\times\left(\frac{t_s-t^{(r-1)}}{h_r}\right)-\sum_{s=p_{r+1}}^{q_{r+1}}(z_s-\varepsilon_s)\left(\frac{t^{(r)}-t_s}{h_{r+1}}\right)E\left\{\sum_{s=p_e}^{q_e}z_s\times\right.$$

$$\times\left(\frac{t_s-t^{(e-1)}}{h_e}\right)+\sum_{s=p_{e+1}}^{q_{e+1}}z_s\left(\frac{t^{(r)}-t_s}{h_{r-e+1}}+1\right)-\sum_{s=p_e}^{q_e}(z_s-\varepsilon_s)\times$$

$$\left.\left.\times\left(\frac{t_s-t^{(e-1)}}{h_e}\right)-\sum_{s=p_{e+1}}^{q_{e+1}}(z_s-\varepsilon_s)\left(\frac{t^{(e)}-t_s}{h_{e+1}}+1\right)\right\}\right\}= \qquad (16)$$

$$=E\left\{\left[\sum_{s=p_r}^{q_r}\varepsilon_s\left(\frac{t_s-t^{(r-1)}}{h_r}\right)+\sum_{s=p_{r+1}}^{q_{r+1}}\varepsilon_s\left(\frac{t^{(r)}-t_s}{h_{r+1}}+1\right)\right]\times\right.$$

$$\left.\times\left[\sum_{s=p_e}^{q_e}\varepsilon_s\left(\frac{t_s-t^{(e-1)}}{h_e}\right)+\sum_{s=p_{e+1}}^{q_{e+1}}\varepsilon_s\left(\frac{t^{(e)}-t_s}{h_{e+1}}+1\right)\right]\right\}$$

З виразу (16) випливають наступні випадки:

1) якщо $r\ne e, r\ne e+1, r\ne e-1$, то

$$cov[g_r,g_e]=a_{re}\sigma^2=0;$$

2) якщо $r=e$, то

$$cov[g_e,g_e]=E\left\{\left[\sum_{s=p_e}^{q_e}\varepsilon_s\left(\frac{t_s-t^{(e-1)}}{h_e}\right)\right]^2+\right.$$

$$\left.+\left[\sum_{s=p_{e+1}}^{q_{e+1}}\varepsilon_s\left(\frac{t^{(e)}-t_s}{h_{e+1}}+1\right)\right]^2\right\}=a_{ee}\sigma^2$$

3) якщо $r=e+1$, то $cov[g_{e+1},g_e]=a_{e+1,e}\sigma^2$;

4) якщо $r=e-1$, то $cov[g_{e-1},g_e]=a_{e-1,e}\sigma^2$.

Це доводить лему.

<u>*Теорема 1.1.*</u>

Якщо S_0^2 - мінімум суми відхилень оцінки лінійного сплайну відносно емпіричних даних, N – загальна кількість варіант, k – кількість вузлів, тоді незсунена оцінка дисперсії σ^2 має вигляд

$$D=\frac{E\{S_0^2\}}{(N-k-1)}.$$

Доведення.

Базуючись на (10), вважаємо, що значення S_0^2 визначається за виразом

$$S_0^2=\sum_{i=1}^{k}\sum_{s=p_i}^{q_i}\left[z_s-\frac{t_s-t^{(i-1)}}{h_i}\hat{z}^{(i)}-\frac{t^{(i-1)}-t_s}{h_i}\hat{z}^{(i-1)}-\right.$$

$$\left.-\hat{z}^{(i-1)}\right]^2=\sum_{i=1}^{k}\sum_{s=p_i}^{q_i}z_s^2-2\sum_{r=0}^{k}g_r\hat{z}^{(r)}+\sum_{r=0}^{k}\sum_{s=0}^{k}a_{rs}z^{(r)}z^{(s)}=$$

$$=\sum_{i=1}^{k}\sum_{s=p_i}^{q_i}z_s^2-\sum_{r=0}^{k}g_r\hat{z}^{(r)}. \qquad (17)$$

Подальше доведення зводиться до знаходження математичних сподівань елементів правої частини виразу (17). При цьому реалізується оцінка

$$\left\{cov\left[\hat{z}^{(r)},\hat{z}^{(y)}\right]\right\}=\left\{\left[\left(\hat{z}^{(r)}-z^{(r)}\right)\left(\hat{z}^{(e)}-z^{(e)}\right)\right]\right\}=$$

$$=\left[\left(\hat{Z}-Z\right)\left(\hat{Z}-Z\right)^T\right]=E\left\{T^{-1}(g-E\{g\})(g-E\{g\})^T\times\right.$$

$$\left.\times T^{-1}\right\}=T^{-1}E\left\{(g-E\{g\})(g-E\{g\})^T T^{-1}\right\}=$$

$$= T^{-1}\left(T\sigma^2\right)T^{-1} = \{d_{r,e}\}\sigma^2$$

де $\{d_{r,e}\} = D = T^{-1}$.

З урахуванням лем 1.1 та 1.2 визначають шукані математичні сподівання:

$$E\left\{\sum_{i=1}^{k}\sum_{s=p_i}^{q_i} z_s^2\right\} = \sum_{i=1}^{k}\sum_{s=p_i}^{q_i}\left[D\{z_s\} + \left(E\{z_s\}\right)^2\right] = \sum_{i=1}^{k}\sum_{s=p_i}^{q_i}\left(\sigma^2 + \right.$$

$$\left. + \left(\left(\frac{t_s - t^{(i-1)}}{h_i}\right)z^{(i)} + \left(\frac{t^{(i-1)} - t_s}{h_i} + 1\right)z^{(i-1)}\right)^2\right) =$$

$$= N\sigma^2 + \sum_{i=1}^{k}\sum_{s=p_i}^{q_i}\left[z^{(i)2}\left(\frac{t_s - t^{(i-1)}}{h_i}\right)^2 + z^{(i-1)2} \times\right.$$

$$\times\left(\frac{t^{(i-1)} - t_s}{h_i} + 1\right)^2 + 2z^{(i)}z^{(i-1)}\left(\frac{t_s - t^{(i-1)}}{h_i}\right)\times$$

$$\left.\times\left(\frac{t^{(i-1)} - t_s}{h_i} + 1\right)\right] = N\sigma^2 + z^{(0)2}\sum_{s=p_1}^{q_1}\left(\frac{t^{(1)} - t_s}{h_1} + 1\right)^2 +$$

$$+ z^{(0)}z^{(1)}\sum_{s=p_1}^{q_1}\left(\frac{t_s - t^{(0)}}{h_1}\right)\left(\frac{t^{(0)} - t_s}{h_1} + 1\right) + z^{(0)}z^{(2)}\times$$

$$\times 0 + \ldots + z^{(1)}z^{(0)}\sum_{s=p_1}^{q_1}\left(\frac{t^{(0)} - t_s}{h_1} + 1\right)\left(\frac{t_s - t^{(0)}}{h_1}\right) +$$

$$+ z^{(1)2}\left[\sum_{s=p_1}^{q_1}\left(\frac{t_s - t^{(0)}}{h_1}\right)^2 + \sum_{s=p_2}^{q_2}\left(\frac{t^{(1)} - t_s}{h_2} + 1\right)^2\right] +$$

$$+ z^{(1)}z^{(2)}\sum_{s=p_2}^{q_2}\left(\frac{t_s - t^{(1)}}{h_2}\right)\left(\frac{t^{(1)} - t_s}{h_2} + 1\right) + z^{(1)}z^{(3)}\times$$

$$\times 0 + \ldots + z^{(1)}z^{(k)}\cdot 0 + \ldots = N\sigma^2 + \sum_{r=0}^{k}\sum_{e=0}^{k} a_{re}z^{(r)}z^{(e)};$$

$$E\left\{\sum_{r=0}^{k} g_r\hat{z}^{(r)}\right\} = E\left\{\hat{Z}^T G\right\} = E\left\{\sum_{r=0}^{k}\sum_{e=0}^{k}\hat{z}^{(r)}a_{re}\hat{z}^{(e)}\right\} =$$

$$= \sum_{r=0}^{k}\sum_{e=0}^{k} a_{re}E\left\{\hat{z}^{(r)}\hat{z}^{(e)}\right\} = \sum_{r=0}^{k}\sum_{e=0}^{k} a_{re}\left[\hat{z}^{(r)}\hat{z}^{(e)} + \right.$$

$$\left. + cov\left[\hat{z}^{(r)}\hat{z}^{(e)}\right]\right] = \sum_{r=0}^{k}\sum_{e=0}^{k} a_{re}z^{(r)}z^{(e)} + \sum_{r=0}^{k}\sum_{e=0}^{k} a_{re}\times$$

$$\times d_{re}\sigma^2 = \sum_{r=0}^{k}\sum_{e=0}^{k} a_{re}z^{(r)}z^{(e)} + (k+1)\sigma^2$$

Отже, має місце результат

$$E\left\{S_0^2\right\} = E\left\{\sum_{i=1}^{k}\sum_{s=p_i}^{q_i} z_s^2\right\} - E\left\{\sum_{r=0}^{k} g_r\hat{z}^{(r)}\right\} = (N - k - 1)\sigma^2.$$ Тим самим теорему доведено.

Наслідок.

При заданому математичному сподіванні точність наближенні $E\left\{S_0^2\right\}$ та дисперсії похибки заміни σ^2 емпіричних даних лінійними сплайнами кількість вузлів визначається як

$$k = \frac{(N+1)\sigma^2 - E\left\{S_0^2\right\}}{\sigma^2}$$

або

$$k = (N+1) - \frac{E\left\{S_0^2\right\}}{\sigma^2}$$

Більш складні обчислювальні схеми мають місце для поліноміальних сплайн-регресійних моделей степеню $n \geq 2$.

Висновки

1. Запропонована обчислювальна технологія апроксимації експериментальних даних усереднюючими сплайн-функціями.
2. Проведено дослідження та запропонована технологія знаходження числа вузлів при заданій похибці апроксимації.

Обчислювальну технологію реалізовано у програмному середовищі „AirNorm” для оцінки залежностей концентрацій шкідливих речовин у атмосферному повітрі від відстані.

ЛІТЕРАТУРА

1. Приставка П.О. Поліномаільні сплайни при обробці даних. Д.: Вид-во Дніпропетр. ун-ту, 2004. – 236 с
2. Rice J.R. The approximation of function. Vol.II. – Massachusets: Addison-Wessley, Reading, 1967/
3. Wegman Edward J., Wright Jan W. Splines in statistic// J. Amer. Statist. Assor., 1983, 78, №382 – pp. 351-365.
4. Приставка А.Ф. Сплайн-распределение в статистическом анализе. – Днепропетровск: Д.24, 1995 – 152с.
5. Grace Wahba. Spline models for observational data. – Philadelphia, Pennsylvania, 1990. – 169 p.

пост. 28.09.07.