УДК 519.254:519.652

П.О. Приставка, канд. техн. наук

ОБЧИСЛЮВАЛЬНА ТЕХНОЛОГІЯ МОДЕЛЮВАННЯ ЗАЛЕЖНОСТЕЙ ЗА РЕЗУЛЬТАТАМИ СПОСТЕРЕЖЕНЬ

Інститут електроніки та систем управління НАУ, e-mail: obaybuz@inbox.ru

Подано технологію моделювання одновимірних регресій на основі лінійних комбінацій B-сплайнів. Як частковий випадок наведеної технології показано можливість відтворення функціональних залежностей, площинних кривих і замкнених контурів.

Постановка проблеми

Задача моделювання стохастичних залежностей на підставі експериментальних даних може бути вирішеною за використанням або методів інтерполяції, або методів згладжування, якщо інформативна значущість осциляцій процесу не є важливою.

Компромісним варіантом може бути застосування сплайн-методів, зокрема, поліноміальних сплайнів на основі *B*-сплайнів, близьких до інтерполяційних у середньому. Проблема виникає під час реалізації зазначених сплайнів на нерегулярних сітках вузлових точок.

Аналіз досліджень

Розв'язком задачі застосування сплайн-операторів на основі *B*-сплайнів на нерегулярних сітках вузлів є різного роду параметризація залежностей, що підлягають оцінці.

Для поліноміальних сплайнів, близьких до інтерполяційних, існує багато джерел [1; 2], де наведено відповідні обчислювальні схеми. У роботі [3] багато уваги приділено сплайн-операторам, близьким до інтерполяційних у середньому.

Виходячи з положень роботи [3], а також досліджень моделювання масивів реалізацій одновимірних випадкових величин [4], будемо шукати можливість параметризації стохастичних залежностей на підставі ймовірнісної міри.

Виклад основного матеріалу

Нехай задано ймовірнісний простір $\langle \Omega, P, A \rangle$. Введемо випадкову величину $\xi(\omega)$, що відображає Ω у R_1. При цьому будемо вважати, що $\xi(\omega)$ характеризується функцією розподілу ймовірностей

$$F(x) = P\{-\infty < \xi(\omega) < x\}.$$

Не зменшуючи загальності, припустимо, що

$$F(x) \in C^k,\ k = 2, 3, \ldots .$$

Нехай у результаті проведення експерименту одержано масив реалізацій $\xi(\omega)$:

$$X = \left\{ x_i; i = \overline{1,n} \right\}.$$

Додатково умовимося, що всі дані вихідної вибірки різняться між собою:

$$t_a \neq t_b,\ a \neq b,\ a, b = \overline{1,n}.$$

За вибіркою X побудуємо варіаційний ряд: проведемо ранжування вихідного масиву реалізацій $\xi(\omega)$, кожному варіанту поставимо у відповідність частоту, що дорівнює одиниці, випадковість $\frac{1}{n}$ та емпіричну функцію розподілу, причому

$$F_n(x_i) = \frac{i}{n},\ i = \overline{1,n}.$$

Зауважимо, що

$$P\left\{ \lim_{n\to\infty} \sup_x \left| F_n(x) - F(x) \right| = 0 \right\} = 1.$$

Для проведення моделювання сукупності реалізацій випадкової величини $\xi(\omega)$ достатньо або наперед знати аналітичний вигляд закону розподілу, або мати якісне неперервне наближення $F(x)$.

У будь-якому разі $\forall x_\alpha$, яке розподілене за законом розподілу ймовірностей $F(x)$ і відповідає ймовірності реалізації α, можна отримати із співвідношення

$$x_\alpha = F^{-1}(\alpha)$$

(відомий метод оберненої функції). Зважаючи на те, що аналітичний вигляд закону розподілу нами не ідентифікується, моделювання будемо проводити виходячи з аналітичного наближення функції $F^{-1}(\alpha)$.

За віссю абсцис відкладемо значення емпіричної функції розподілу ймовірностей вихідного варіаційного ряду, за віссю ординат – відповідні значення реалізацій. Маючи за віссю абсцис рівномірне розбиття

$$\Delta_{\frac{1}{n}} = \left\{ \frac{i}{n}; i = \overline{1,n} \right\}$$

інтервалу $[0;1]$ із кроком $\frac{1}{n}$ і припустивши, що функція

$$F^{-1}(\alpha) \in C^k,\ k = 2, 3, \ldots ,$$

будемо шукати її наближення у вигляді локальних поліноміальних сплайнів на основі B-сплайнів, близьких до інтерполяційних у середньому [3; 5]. Для остаточного визначення змодельованого масиву після побудови сплайну достатньо відібрати квантилі згідно з обсягом моделювання.

Наведемо найпростішу обчислювальну схему [4] моделювання масиву реалізацій $\xi(\omega)$ за заданим масивом

$$X=\left\{x_i; i=\overline{1,n}\right\}$$

з використанням локального сплайну

$$S_{2,0}(x,\alpha),\ \alpha\in[0;1]$$

на основі B-сплайнів другого порядку нульового ступеня уточнення

$$S_{2,0}(x,\alpha)=\sum_{i=1}^{n} x_i B_{2,\frac{1}{n}}\left(\alpha-\frac{i}{n}\right),$$

де

$$B_{2,1/n}(\alpha)=\begin{cases} 0, & |\alpha|\ge 3/2n, \\ (3+2\alpha n)^2/8, & \alpha\in[-3/2n;-1/2n], \\ 3/4-(2\alpha n)^2/4, & \alpha\in[-1/2n;1/2n], \\ (3-2\alpha n)^2/8, & \alpha\in[1/2n;3/2n], \end{cases}$$

або в розгорнутому вигляді:

$$S_{2,0}(x,\alpha)=\frac{1}{8}\big((x_{i-1}+6x_i+x_{i+1})+(-2x_{i-1}+2x_{i+1})q+$$

$$+(x_{i-1}-2x_i+x_{i+1})q^2\big), \qquad (1)$$

де

$$q=2n\left(\alpha-\frac{i}{n}\right),\ i=\overline{1,n}.$$

У випадку m-кратного збільшення кількості даних достатньо взяти m значень змінної t на кожному i-му проміжку розбиття $\Delta_{\frac{1}{n}}$:

$$q_j=-1+\frac{2}{m}\sum_{k=1}^{j}k,\quad j=\overline{1,m}.$$

Як невизначені у формулі (1) значення x_0, x_{n+1} цілком достатньо взяти величини [3]:

$$x_0=\frac{1}{3}(4x_1+x_2-2x_2);$$

$$x_{n+1}=\frac{1}{3}(4x_n+x_{n-1}-2x_{n-2}).$$

Адекватність і ефективність застосування зазначеного підходу було експериментально доведено на підставі імітаційного моделювання [4].

Теоретично обґрунтована похибка при моделюванні дорівнює похибці апроксимації, яку визначено для відповідного локального використовуваного сплайну за умови, що вибірку

$$X=\left\{x_i; i=\overline{1,n}\right\}$$

проведено якісно.

Так, у наведеному прикладі застосування сплайну $S_{2,0}(x,\alpha)$, якщо

$$x(\alpha)=F^{-1}(\alpha)\in C^2, \text{ де } \alpha\in[0;1],$$

відповідна похибка становить [3]

$$\left\|x(\alpha)-S_{2,0}(x,\alpha)\right\|_C=\frac{\left\|x''(\alpha)\right\|_C}{6\cdot n^2}+o\left(\frac{1}{n^2}\right),$$

а якщо при моделюванні застосовується локальний сплайн $S_{2,1}(x,\alpha)$ на основі B-сплайнів другого порядку першого ступеня уточнення

$$S_{2,1}(x,\alpha)=\sum_{i=1}^{n}\left(x_i-\frac{1}{6}\ \Delta^2 x_i\right)B_{2,\frac{1}{n}}\left(\alpha-\frac{i}{n}\right),$$

$$\Delta^2 x_i=x_{i-1}-2x_i+x_{i+1},$$

відповідна похибка для

$$\forall x(\alpha)\in C^3$$

становить

$$\left\|x(\alpha)-S_{2,1}(x,\alpha)\right\|_C=\frac{\left\|x^{(3)}(\alpha)\right\|_C}{12\sqrt{3}\ n^3}+o\left(\frac{1}{n^3}\right).$$

Отже, можемо вибирати конкретний тип сплайну, залежно від вимог до якості апроксимації.

При відсутності осциляції функції $x(\alpha)$ суттєво не має значення, який із локальних сплайнів застосовується. Якщо ж характер варіабельності реалізацій функції $x(\alpha)$ є суттєвим, слід використовувати типи сплайнів, що уточнюють.

Не зменшуючи загальності, у подальшому викладанні будемо розглядати сплайн $S_{2,0}(x,\alpha)$ у вигляді (1).

Проведемо моделювання реалізацій двовимірної випадкової величини

$$\vec{\zeta}=(\xi,\eta),$$

що відображає Ω у R_1. Результатом відбиття є масив

$$\left\{(x_i,y_i); i=\overline{1,n}\right\},$$

який характеризується функцією розподілу

$$G(x,y)=P\{-\infty<\xi(\omega)<x; -\infty<\eta(\omega)<y\}.$$

Будемо вважати, що аналітичний вигляд закону розподілу $\vec{\zeta}$ невідомий, і покладемо

$$G(x,y)\in C^{k_1,k_2},$$

де $k_1,k_2=2,3,\ldots$.

Можна піти шляхом, аналогічним одновимірному випадку, а саме: визначати місцезнаходження точок при заданому рівні випадковості α із співвідношення

$$(x, y)_\alpha = G^{-1}(\alpha).$$

Проте подібний підхід не є тривіальним, як на разі одновимірного випадку, навіть за наявності аналітичного наближення функції $G(x, y)$.

Розв'язок може бути знайдено, якщо припустити існування залежності однієї з одновимірних випадкових величин, що належать до вектора $\vec{\zeta}$, від іншої.

Надалі вважатимемо, що величина η залежить від ξ. Вигляд зв'язку – регресійний (хай навіть слабкий) або функціональний. Відповідно до масиву реалізацій: $\exists\ \hat{y}(x)$ або $y(x)$. Тоді ніщо не заважає здійснити наступну операцію.

Проведемо ранжування вихідного масиву за незалежною змінною x, поставивши у відповідність кожній двійці (x_i, y_i) новоутвореного ряду частоту, випадковість та значення емпіричної функції розподілу

$$F_n(x_i) = \frac{i}{n},\ i = \overline{1,n},$$

що є дискретним аналогом функції розподілу $F(x)$ випадкової величини ξ зі складу вектора $\vec{\zeta}$. Далі можемо перейти до розгляду двох послідовностей:

$$\left\{\left(x_i, \frac{i}{n}\right); i = \overline{1,n}\right\} \quad \text{та} \quad \left\{\left(y_i, \frac{i}{n}\right); i = \overline{1,n}\right\}. \qquad (2)$$

Причому поміж x_i та y_i існує однозначна відповідність – як двійка реалізації $\vec{\zeta}$. Тоді, припускаючи, що параметричне подання

$$x(\alpha), y(\alpha) \in C^k,\ k = 2,3,\ldots,$$

де

$$\alpha \in [0;1],$$

можемо за масивами (2) знайти наближення функцій $x(\alpha)$, $y(\alpha)$ у вигляді локального поліноміального сплайну на зразок зазначеної обчислювальної процедури моделювання одновимірного масиву.

Для остаточного визначення змодельованого масиву реалізацій випадкової величини $\vec{\zeta} = (\xi, \eta)$ досить відібрати значення наближень функцій $x(\alpha)$, $y(\alpha)$ при однакових значеннях аргументу.

Якщо ставиться задача визначення значення регресії $\hat{y}(x)$ для $\forall x$ з області визначення незалежної змінної, то обчислювальна процедура має такий вигляд:

– визначаємо, чому дорівнює α при заданому x;

– визначаємо, які i та q з виразу (1) відповідають даному α;

– підставляємо отримані i та q у вираз, аналогічний формулі (1), для визначення $\hat{y}(x)$:

$$S_{2,0}(y,\alpha) = \frac{1}{8}\left(\left(y_{i-1} + 6y_i + y_{i+1}\right) + \left(-2y_{i-1} + 2y_{i+1}\right)q + \right.$$
$$\left. + \left(y_{i-1} - 2y_i + y_{i+1}\right)q^2\right).$$

Отримана подібним чином регресія за рахунок високої якості апроксимації локальних сплайнів може суттєво осцилювати залежно від осциляції реалізацій y_i у послідовності (2). У цьому разі в обчислювальній процедурі відтворення непараметричної регресії $\hat{y}(x)$ слід передумовити процедуру згладжування одержаної оцінки (рис. 1).

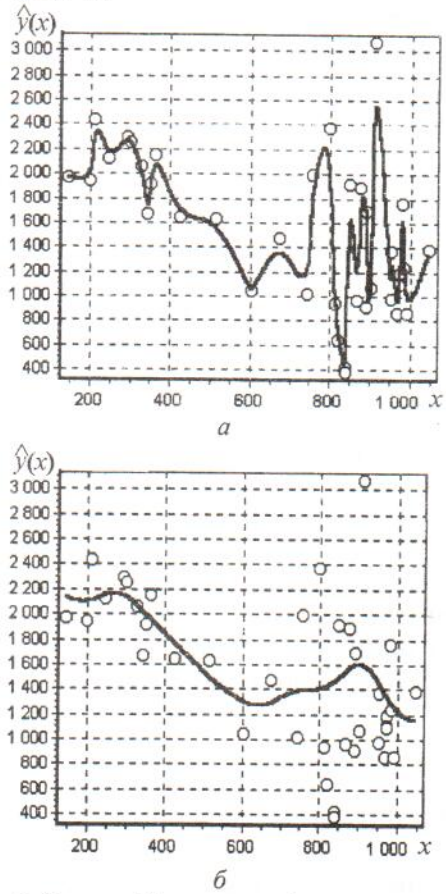
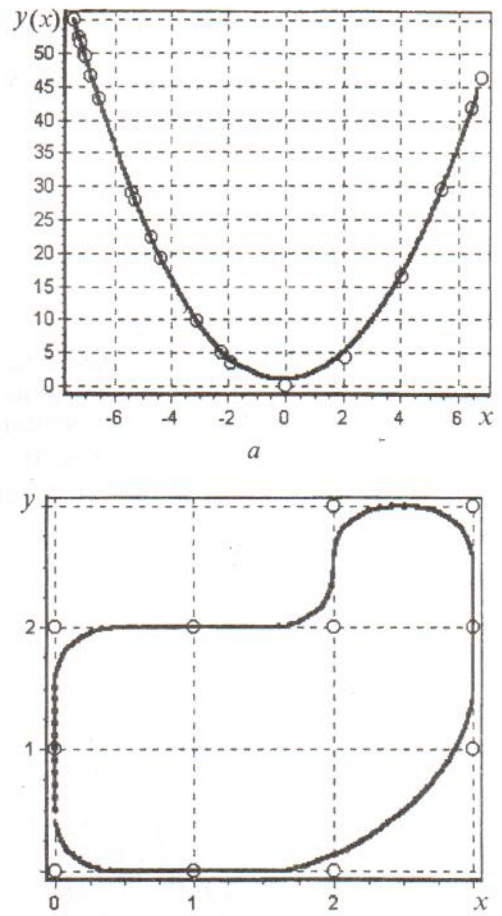


Рис. 1. Кореляційне поле та відтворена регресія: *а* – без згладжування; *б* – зі згладжуванням

Викладений підхід до відтворення регресій цілком правдивий і для відтворення функціональних залежностей $y(x)$ (рис. 2, *а*), площинних кривих, замкнених кривих (рис. 2, *б*).

Для вирішення в обчислювальному процесі задачі пошуку наближення $y(x)$ досить притримуватися викладеної технології, як і для відтворення регресій.



а

б

Рис. 2. Відтворення залежностей:
а – функціональна залежність $y = x^2$;
б – замкнений контур

Для побудови кривих у площині – підхід аналогічний, проте початковий масив

$$\left\{\left(x_i, y_i\right); i = \overline{1,n}\right\}$$

не підлягає сортуванню за жодній зі змінних. Для замкнених кривих потрібно слідкувати, щоб перша і остання точки послідовності збігалися. По суті при відтворенні функцій або кривих відбувається перехід до їх параметричного зображення. Причому значення параметра збігається або є кратним номеру відповідної точки у вихідному масиві.

Висновки

Викладена інформаційна технологія моделювання реалізацій одновимірних і технологія відтворення одновимірних регресій (функціональних залежностей, площинних кривих) за рахунок своєї простоти та ефективності застосування типів зазначених локальних сплайнів може бути рекомендована до реалізації в обчислювальному процесі обробки спостережень.

Порівняльний аналіз наведених обчислювальних схем із викладеними в роботі [1] за методами машинної графіки, більшість з яких є стандартами для реалізації в сучасному програмному забезпеченні (метод найменших квадратів, криві Без'є, раціональні *B*-сплайни (NURBS)), показує, що при практично тотожній адекватності, що надають класичні методи та запропонований підхід, реалізація, а отже, швидкодія та якість роботи останнього мають перевагу.

Список літератури

1. *Роджерс Д., Адамс Дж.* Математические основы машинной графики. – М.: Мир, 2001. – 604 с.
2. *Павлидис Т.* Алгоритмы машинной графики и обработки изображений. – М.: Радио и связь, 1986. – 400 с.
3. *Лигун А.А., Шумейко А.А.* Асимптотические методы восстановления кривых. – К.: ІМ НАН України, 1996. – 358 с.
4. *Приставка П.О., Смойловська О.О.* Обробка вибірок обмеженого обсягу з використанням поліноміальних сплайнів // Актуальні проблеми автоматизації та інформаційних технологій. – Д.: Навч. кн. – 2001. – Т.4. – С. 86–95.
5. *Бабак В.П., Білецький А.Я., Приставка О.П., Приставка П.О.* Статистична обробка даних. – К.: МІВВЦ, 2001. – 388 с.

Стаття надійшла до редакції 23.05.03.

Ф.А. Приставка

Вычислительная технология моделирования зависимостей по результатам наблюдений

Приведена технология моделирования одномерных регресий на основе линейных комбинаций В-сплайнов. Как частный случай приведенной технологии показана возможность восстановления функциональных зависимостей, плоских кривых и замкнутых контуров.

P.A. Pristavka

The calculated technology of modeling dependences of results of observers

The probabilistic modeling of nonlinear regressional, stochastic and functional dependences based on the B-splines linear combinations was considered in the article.