

HÀM NGẪU NHIÊN B-SPLINE VÀ ỨNG DỤNG VÀO DỰ BÁO

Lê Hào*

Tóm tắt

Trong bài báo này, chúng tôi đưa ra khái niệm hàm ngẫu nhiên B-spline ứng với mẫu thực nghiệm của đại lượng ngẫu nhiên hai chiều (X, Y) và chứng tỏ hàm ngẫu nhiên này phản ánh tốt mối quan hệ giữa X và Y trong trường hợp hệ số tương quan $R(X, Y)$ bé, để chứng tỏ điều này chúng tôi đã nêu và chứng minh định lý (2.4). Trong phần áp dụng, chúng tôi sử dụng định lý (2.4) để thiết lập hàm hồi quy B-spline cho việc dự báo.

Từ khóa: Hàm ngẫu nhiên B-spline, hàm hồi quy B-spline, hệ số tương quan, mẫu thực nghiệm, biến số ngẫu nhiên.

1. Giới thiệu

Việc áp dụng mô hình hồi quy phù hợp để dự báo mối tương quan của hai đại lượng ngẫu nhiên X, Y là việc rất khó khăn, thông thường khi hệ số tương quan $R(X, Y)$ quá gần 0 thì việc sử dụng một số mô hình hồi quy đã biết có thể cho ta những dự báo sai lệch lớn, ngoài ra việc cập nhật cho các mô hình hồi qui thường gặp trở ngại khi các dữ liệu được cập nhật liên tục.

Trong bài báo này, chúng tôi đề cập đến kiểu hàm $y(x) = f(x, Y)$ ($x \in D \subset \mathbb{R}$) trong đó Y là một đại lượng ngẫu nhiên, có nghĩa là ứng với mỗi giá trị $x \in D$ cụ thể thì $y(x)$ là một đại lượng ngẫu nhiên. Hàm như thế gọi là hàm ngẫu nhiên, có vai trò rất lớn trong thống kê ứng dụng.

Bài báo đề cập đến một hướng giải quyết vấn đề nói trên, thông qua việc khảo sát hàm ngẫu nhiên B-spline và áp dụng vào dự báo.

2. Các khái niệm và định lý

Trong [2] chúng tôi đã đề cập đến khái niệm và công thức của đường cong B-spline bậc 2 đi qua tất cả các điểm nút $P_k(x_k, y_k)$ ($k = 0, \dots, n$) cho trước. Đó là đường cong tham số $P(t) = (x(t), y(t))$ gồm $n-1$ cung Bezier ghép trơn, được xác lập bởi công thức:

$$\begin{aligned} P(t) &= S_0(1-t)^2 + 2T_1t(1-t) + S_1t^2 \text{ khi } 0 \leq t \leq 1 \\ P(t) &= S_1(2-t)^2 + 2T_2(t-1)(2-t) + S_2(t-1)^2 \text{ khi } 1 \leq t \leq 2 \\ P(t) &= S_2(3-t)^2 + 2T_3(t-2)(3-t) + S_3(t-2)^2 \text{ khi } 2 \leq t \leq 3 \\ &\dots\dots\dots \\ P(t) &= S_{n-3}(n-2-t)^2 + 2T_{n-2}(t-n+3)(n-2-t) + S_{n-2}(t-n+3)^2 \text{ khi } n-3 \leq t \leq n-2 \\ P(t) &= S_{n-2}(n-1-t)^2 + 2T_{n-1}(t-n+2)(n-1-t) + S_{n-1}(t-n+2)^2 \text{ khi } n-2 \leq t \leq n-1 \end{aligned} \quad (2.1)$$

Trong đó $S_0 \equiv P_0, S_{n-1} \equiv P_n$ và $S_1, S_2, \dots, S_{n-2}$ lần lượt là trung điểm của các đoạn thẳng

* ThS, Trường Đại học Phú Yên

$T_1T_2, T_2T_3, \dots, T_{n-2}T_{n-1}$.

Các điểm $T_1, T_2, \dots, T_{n-1}$ được xác định bởi hệ sau (để đảm bảo đường cong đi qua mọi điểm nút P_k):

$$\begin{array}{cccccccc}
 5T_1 & +T_2 & & & & & & = 8P_1 - 2P_0 \\
 T_1 & +6T_2 & +T_3 & & & & & = 8P_2 \\
 & T_2 & +6T_3 & +T_4 & & & & = 8P_3 \\
 \dots & \dots & \dots & \dots & \dots & \dots & \dots & \\
 & & & T_{n-3} & +6T_{n-2} & +T_{n-1} & & = 8P_{n-2} \\
 & & & T_{n-2} & +5T_{n-1} & & & = 8P_{n-1} - 2P_n
 \end{array} \quad (2.2)$$

Khi các nút $P_0, P_1, \dots, P_n$ tương ứng có hoành độ tăng dần $x_0 < x_1 < \dots < x_n$ thì hàm $x = x(t)$ tăng, khi đó công thức (2.1) của đường B-spline xác định hàm $y = L(x)$ ($x_0 \leq x \leq x_n$) có đạo hàm liên tục mọi cấp trên đoạn $[x_0, x_n]$ và $\sup_{x_0 \leq x \leq x_n} |L^{(n)}(x)| \leq M < +\infty$ ($\forall n \in \mathbb{N}$) (xem [2]).

Tiếp theo ta xét (X, Y) là đại lượng ngẫu nhiên hai chiều có hệ số tương quan $R(X, Y)$ khá bé, X là biến ngẫu nhiên nhận giá trị trong $(x_0; +\infty)$ và $(x_0, y_0), (x_1, y_1), \dots, (x_{n-1}, y_{n-1})$ là mẫu cụ thể đã biết của (X, Y) với $x_0 < x_1 < \dots < x_{n-1}$.

Giả sử cần quan sát mối quan hệ giữa X, Y khi X nhận giá trị trong khoảng (x_0, d) chứa tất các mốc x_k ($k = 0, \dots, n-1$), ta chọn thêm mốc $x_n = d$.

Định nghĩa 2.3. Với mẫu đã nói trên, hàm ngẫu nhiên B-spline là hàm $y = L(x)$ được xác định từ công thức (2.1) của đường cong B-spline bậc hai đi qua các nút:

$$P_0(x_0, y_0), P_1(x_1, y_1), \dots, P_{n-1}(x_{n-1}, y_{n-1}), P_n(x_n, Y^*)$$

Trong đó $P_n(x_n, Y^*)$ là điểm ngẫu nhiên với $Y^* = aX + b$ là đại lượng ngẫu nhiên xác định bởi công thức hồi quy của Y qua X theo nguyên tắc bình phương bé nhất, nghĩa là:

$$a = \rho \sqrt{\frac{DY}{DX}}, \quad b = EY - aEX$$

$$\text{và } \rho = R(X, Y) = \frac{\text{Cov}(X, Y)}{\sqrt{DX} \cdot \sqrt{DY}} = \frac{E(XY) - EX \cdot EY}{\sqrt{EX^2 - (EX)^2} \cdot \sqrt{EY^2 - (EY)^2}}$$

Trên phương diện lý thuyết mối quan hệ giữa X, Y được thể hiện bằng đường cong lý thuyết $y = f(x)$ đi qua các nút đã biết $(x_0, y_0), (x_1, y_1), \dots, (x_{n-1}, y_{n-1})$ và nút chưa biết (x_n, Y) , ta xem nút (x_n, Y) là cặp giá trị giả định của đại lượng ngẫu nhiên (X, Y) tại mốc $X = x_n > x_{n-1}$. Nếu cần ta xấp xỉ $f(x)$ bằng hàm B-spline đi qua vô số điểm nút của $f(x)$ nên ta xem hàm $f(x)$ có đạo hàm liên tục mọi cấp trên đoạn $[x_0, x_n]$.

Hàm $f(x)$ có đa thức nội suy với các nút $(x_0, y_0), \dots, (x_{n-1}, y_{n-1}), (x_n, Y)$ là hàm ngẫu nhiên, xác định như sau:

$$Pf(x, (x_0, x_n, \dots, x_n)) = y_0 \frac{(x-x_1)(x-x_2)\dots(x-x_n)}{(x_0-x_1)(x_0-x_2)\dots(x_0-x_n)} + y_1 \frac{(x-x_0)(x-x_2)\dots(x-x_n)}{(x_1-x_0)(x_1-x_2)\dots(x_1-x_n)} + \dots$$

$$+ y_{n-1} \frac{(x-x_0)(x-x_1)\dots(x-x_{n-2})(x-x_n)}{(x_{n-1}-x_0)(x_{n-1}-x_1)\dots(x_{n-1}-x_{n-2})(x_{n-1}-x_n)} + Y \frac{(x-x_0)(x-x_1)\dots(x-x_{n-1})}{(x_n-x_0)(x_n-x_1)\dots(x_n-x_{n-1})} \quad (x \in \mathbb{R})$$

Hàm B-spline $L(x)$ có đa thức nội suy với các nút $(x_0, y_0), \dots, (x_{n-1}, y_{n-1}), (x_n, Y^*)$ cũng là hàm ngẫu nhiên, xác định như sau:

$$PL(x, (x_0, x_n, \dots, x_n)) = y_0 \frac{(x-x_1)(x-x_2)\dots(x-x_n)}{(x_0-x_1)(x_0-x_2)\dots(x_0-x_n)} + y_1 \frac{(x-x_0)(x-x_2)\dots(x-x_n)}{(x_1-x_0)(x_1-x_2)\dots(x_1-x_n)} + \dots$$

$$+ y_{n-1} \frac{(x-x_0)(x-x_1)\dots(x-x_{n-2})(x-x_n)}{(x_{n-1}-x_0)(x_{n-1}-x_1)\dots(x_{n-1}-x_{n-2})(x_{n-1}-x_n)} + Y^* \frac{(x-x_0)(x-x_1)\dots(x-x_{n-1})}{(x_n-x_0)(x_n-x_1)\dots(x_n-x_{n-1})} \quad (x \in \mathbb{R})$$

Việc sử dụng hàm $L(x)$ để phản ánh mối tương quan X và Y có hợp lí hay không? Để trả lời cho câu hỏi đó chúng tôi tìm hiểu sự liên hệ giữa các hàm ngẫu nhiên $L(x)$, $f(x)$ và có kết quả sau:

Định lý 2.4. Với mọi $x \in [x_0, x_n]$ thì:

$$|E[L(x) - f(x)]| \leq \frac{\prod_{j=0}^n |x - x_j|}{(n+1)!} \sup_{x_0 \leq t \leq x_n} |E[L^{(n+1)}(t) - f^{(n+1)}(t)]|$$

$$\sqrt{E[L(x) - f(x)]^2} \leq \frac{\prod_{j=0}^{n-1} |x - x_j|}{\prod_{j=0}^{n-1} (x_n - x_j)} \sqrt{\min_{\alpha, \beta \in \mathbb{R}} \{E(Y - \alpha X - \beta)^2\}} + \prod_{j=0}^n |x - x_j| \sqrt{\frac{\sup_{x_0 \leq t \leq x_n} \{E[L^{(n+1)}(t) - f^{(n+1)}(t)]^2\}}{(n+1)!}}$$

Chứng minh.

$$L(x) = PL(x; (x_0, \dots, x_{n-1}, x_n)) + \prod_{j=0}^n (x - x_j) \cdot \Delta^{n+1} L(x_0, \dots, x_{n-1}, x_n, x) \quad (1)$$

$$f(x) = Pf(x; (x_0, \dots, x_{n-1}, x_n)) + \prod_{j=0}^n (x - x_j) \cdot \Delta^{n+1} f(x_0, \dots, x_{n-1}, x_n, x) \quad (2)$$

Từ (1) và (2) suy ra:

$$L(x) - f(x) = \frac{\prod_{j=0}^{n-1} (x - x_j)}{\prod_{j=0}^{n-1} (x_n - x_j)} (Y^* - Y) + \prod_{j=0}^n (x - x_j) (\Delta^{n+1} L(x_0, \dots, x_n, x) - \Delta^{n+1} f(x_0, \dots, x_n, x)) \quad (3)$$

Áp dụng định lí **Hermite – Genocchi** (xem [1]) luôn tồn tại $\xi_1, \xi_2 \in [x_0, x_n]$ sao cho:

$$E\left[\Delta^{n+1}L(x_0, \dots, x_n, x) - \Delta^{n+1}f(x_0, \dots, x_n, x)\right] = \frac{E\left[L^{(n+1)}(\xi_1) - f^{(n+1)}(\xi_1)\right]}{(n+1)!} \quad (4)$$

$$E\left[\Delta^{n+1}L(x_0, \dots, x_n, x) - \Delta^{n+1}f(x_0, \dots, x_n, x)\right]^2 \leq \frac{E\left[L^{(n+1)}(\xi_2) - f^{(n+1)}(\xi_2)\right]^2}{(n+1)!} \quad (5)$$

Từ (3) (4) và do $E(Y^* - Y) = E(aX + b - Y) = 0$ nên:

$$\left|E[L(x) - f(x)]\right| \leq \frac{\prod_{j=0}^n |x - x_j|}{(n+1)!} \sup_{x_0 \leq t \leq x_n} \left|E\left[L^{(n+1)}(t) - f^{(n+1)}(t)\right]\right|$$

Mặt khác từ (3) (5) và đề ý $E(|XY|) \leq \sqrt{EX^2 EY^2}$ (với mọi véc tơ ngẫu nhiên (X, Y)) nên:

$$\sqrt{E[L(x) - f(x)]^2} \leq \frac{\prod_{j=0}^{n-1} |x - x_j|}{\prod_{j=0}^n (x_n - x_j)} \sqrt{E(Y^* - Y)^2} + \prod_{j=0}^n |x - x_j| \sqrt{\frac{\sup_{x_0 \leq t \leq x_n} \left\{E\left[L^{(n+1)}(t) - f^{(n+1)}(t)\right]^2\right\}}{(n+1)!}}$$

Hay nói cách khác:

$$\sqrt{E[L(x) - f(x)]^2} \leq \frac{\prod_{j=0}^{n-1} |x - x_j|}{\prod_{j=0}^{n-1} (x_n - x_j)} \sqrt{\min_{\alpha, \beta \in \mathbb{R}} \left\{E(Y - \alpha X - \beta)^2\right\}} + \prod_{j=0}^n |x - x_j| \sqrt{\frac{\sup_{x_0 \leq t \leq x_n} \left\{E\left[L^{(n+1)}(t) - f^{(n+1)}(t)\right]^2\right\}}{(n+1)!}}$$

□

Nhận xét 2.5. Với $X = x \in [x_0, x_n]$ thì:

Do $x_0 < x_1 < \dots < x_{n-1} < x_n$ và $\frac{x - x_j}{x_n - x_j}$ giảm theo x_j nên với mọi $x \in [x_0, x_n]$ ta có:

$$a = \frac{x - x_0}{x_n - x_0} \geq \frac{x - x_j}{x_n - x_j} \geq \frac{x - x_{n-1}}{x_n - x_{n-1}} = b \quad (\forall j = 0, 1, \dots, n-1)$$

$$\Rightarrow A = \max(a, -b) \geq \frac{x - x_j}{x_n - x_j} \geq -\max(a, -b) = -A \text{ hay } \left| \frac{x - x_j}{x_n - x_j} \right| \leq A \quad (\forall j = 0, \dots, n-1)$$

$$\text{Suy ra: } \frac{\prod_{j=0}^{n-1} |x - x_j|}{\prod_{j=0}^{n-1} (x_n - x_j)} \leq A^n \text{ Với } A = \max\left(\frac{x - x_0}{x_n - x_0}, \frac{x_{n-1} - x}{x_n - x_{n-1}}\right)$$

Nếu ta chọn $X = x \in (2x_{n-1} - x_n, x_n)$ thì: $-1 < \frac{x - x_{n-1}}{x_n - x_{n-1}} < 1 \Rightarrow \left| \frac{x - x_{n-1}}{x_n - x_{n-1}} \right| < 1 \Rightarrow 0 \leq A < 1$

nên khi n tiến ra vô cùng ta có:

$$\frac{\prod_{j=0}^{n-1} |x - x_j|}{\prod_{j=0}^{n-1} (x_n - x_j)} \leq A^n \rightarrow 0 \quad \text{và} \quad \frac{\prod_{j=0}^n |x - x_j|}{\sqrt{(n+1)!}} \leq \frac{|x_n - x_0|^{n+1}}{\sqrt{(n+1)!}} \rightarrow 0$$

Hàm ngẫu nhiên B-spline $L(x)$ phản ánh tốt mối quan hệ giữa các đại lượng ngẫu nhiên

X, Y tại những giá trị x có $\frac{\prod_{j=0}^{n-1} |x - x_j|}{\prod_{j=0}^{n-1} (x_n - x_j)}$ và $\frac{\prod_{j=0}^n |x - x_j|}{\sqrt{(n+1)!}}$ bé.

Hàm $L(x)$ cũng phản ánh tốt mối quan hệ giữa X, Y tại những giá trị $x \in (2x_{n-1} - x_n, x_n)$ nếu chọn kích thước mẫu n đủ lớn.

3. Áp dụng

3.1. Hàm hồi quy B-spline và đường hồi quy B-spline

Trên phương diện thực nghiệm, nếu biết $(x_0, y_0), (x_1, y_1), \dots, (x_{n-1}, y_{n-1})$ là mẫu thực nghiệm của vectơ ngẫu nhiên (X, Y) với $x_0 < x_1 < \dots < x_{n-1}$, ta có thể tiến hành dự báo các giá trị của Y khi $X \in (x_0, d)$ ($x_0 < x_1 < \dots < x_{n-1} < d$) theo cách sau:

Lập thêm mốc $x_n = d$ và xác định thêm điểm nút $P_n(x_n, y^*)$ thỏa mãn:

$$y^* = ax_n + b \quad \text{và} \quad a = r \frac{\sigma_Y}{\sigma_X}, \quad b = \bar{Y} - a\bar{X}$$

với r là hệ số tương quan thực nghiệm của X, Y ứng với mẫu thực nghiệm đã cho.

Lập hàm B-spline $y = L(x)$ theo công thức (2.1) có đồ thị là đường B-spline đi qua tất cả các điểm $P_0(x_0, y_0), P_1(x_1, y_1), \dots, P_{n-1}(x_{n-1}, y_{n-1}), P_n(x_n, y^*)$.

Hàm $y = L(x)$ đó được gọi là hàm hồi quy B-spline (thực nghiệm) của Y theo X , đồ thị của nó gọi là đường hồi quy B-spline. Từ công thức của hàm hồi quy này ta có thể dự báo giá trị $Y = L(x)$ khi $X = x \in (x_0, d)$.

3.2. Bài toán

Bài toán. Quan sát mối quan hệ giữa X, Y trong đại lượng ngẫu nhiên hai chiều người ta thu được mẫu sau:

X	10	40	80	110	150	190	220	260	300	340	380
Y	200	230	250	240	180	160	150	190	200	280	300

400	420	450	480	530	570	600	640	680	720	760	800
320	290	310	270	200	180	160	200	230	250	290	270

Hãy lập các dự báo về giá trị của Y khi $X = x \in (10; 900)$

Bước 1. Lập thêm một nút dự báo tại nút $x_n = 900$ của khoảng quan sát $(10; 900)$

Từ mẫu trên ta tính được hệ số tương quan thực nghiệm $r = 0,245123532$ và các thông số khác như sau:

$$a = r \frac{\sigma_Y}{\sigma_X} = 0.053163965, \quad b = \bar{Y} - a\bar{X} = 211.5049132, \quad y^* = ax_n + b = 259.3524814$$

và lập thêm nút $P_n(900, 259.3524814)$

Bước 2. Lập công thức của đường hồi quy B-spline đi qua các nút $P_k(x_k, y_k)$ ($k = 0, \dots, n$)

Trong đó các nút $P_0, P_1, \dots, P_{n-1}$ ứng với các cặp giá trị của mẫu đã cho và nút P_n vừa lập ra trong bước 1.

Sử dụng công thức (2.2) ta dễ dàng lập được một thuật toán đơn giản cho máy tính để tìm các bộ ba điểm cơ sở cho 22 cung *Bezier* được ghép tron, cụ thể các bộ ba điểm đó là:

S0 (10.00;200.00);	T1 (36.62;237.70);	S1 (59.60;244.92);
S1 (59.60;244.92);	T2 (82.58;252.15);	S2 (95.24;250.77);
S2 (95.24;250.77);	T3 (107.90;249.39);	S3 (128.96;210.46);
S3 (128.96;210.46);	T4 (150.02;171.53);	S4 (171.01;166.48);
S4 (171.01;166.48);	T5 (192.01;161.43);	S5 (204.97;150.66);
S5 (204.97;150.66);	T6 (217.94;139.89);	S6 (239.14;169.57);
S6 (239.14;169.57);	T7 (260.34;199.26);	S7 (280.19;191.92);
S7 (280.19;191.92);	T8 (300.04;184.58);	S8 (319.73;238.91);
S8 (319.73;238.91);	T9 (339.41;293.24);	S9 (361.45;294.60);
S9 (361.45;294.60);	T10 (383.48;295.96);	S10 (391.60;313.49);
S10 (391.60;313.49);	T11 (399.72;331.01);	S11 (408.97;304.49);
S11 (408.97;304.49);	T12 (418.23;277.96);	S12 (434.58;299.59);
S12 (434.58;299.59);	T13 (450.93;321.22);	S13 (463.55;297.98);
S13 (463.55;297.98);	T14 (476.17;274.75);	S14 (504.10;232.52);
S14 (504.10;232.52);	T15 (532.03;190.30);	S15 (551.84;186.88);
S15 (551.84;186.88);	T16 (571.66;183.47);	S16 (584.83;166.18);
S16 (584.83;166.18);	T17 (598.00;148.89);	S17 (619.16;176.03);
S17 (619.16;176.03);	T18 (640.31;203.17);	S18 (660.21;217.62);
S18 (660.21;217.62);	T19 (680.11;232.06);	S19 (699.57;238.25);
S19 (699.57;238.25);	T20 (719.03;244.45);	S20 (742.36;272.85);
S20 (742.36;272.85);	T21 (765.68;301.26);	S21 (766.27;284.63);
S21 (766.27;284.63);	T22 (766.86;268.01);	S22 (900.00;259.35);

Phương trình của đường hồi quy B-spline được xác định từ công thức (2.1), trong đó mỗi thành phần của công thức có các điểm cơ sở S_i , T_{i+1} , S_{i+1} đã xác định như trên.

Bước 3. Lập giá trị dự báo của Y ứng giá trị $X = x \in (10, 900)$

Ta dễ dàng lập một thuật toán để biết chính xác giá trị $X = x$ ứng với cung *Bezier* nào

trong số 22 cung có các điểm cơ sở đã lập, sau đó tìm giá trị t tương ứng với x , rồi suy ra giá trị dự báo $y = y(t)$ của Y .

Đoạn chương trình ngắn gọn sau sẽ giúp ta làm việc này:

```
PROCEDURE TIMGIATRI(L:word; GTX:real; var GTY:real);
Var i:integer;
    Heso1,Heso2,Heso3, Delta,t1:real;
Begin
i:=0;
While (S[i].x < Gtx) and (i<L) do i:=i+1;
Heso1:=S[i].x+S[i-1].x-2*T[i].x;
Heso2:= S[i-1].x - Gtx;
Heso3:=S[i-1].x-t[i].x;

DELTA:= Sqr(Heso3) - Heso1*Heso2;
If Heso1=0 then t1:=Heso2/(2*Heso3)
    Else t1:=(Heso3+sqr(Delta))/Heso1;

GTY:= S[i-1].y*sqr(1-t1) + 2*T[i].y*t1*(1-t1) + S[i].y*Sqr(t1);
END;
```

Kết quả dự báo: Bằng một chương trình ngắn gọn trên ngôn ngữ Pascal ta dễ dàng thu được các dự báo về các giá trị của đại lượng ngẫu nhiên Y , theo các giá trị của đại lượng ngẫu nhiên X nằm trong khoảng (10; 900). Theo nhận xét 2.5 và định lý 2.4 thì các dự báo này luôn đáng tin cậy khi $X = x \in (700; 900)$ miễn sao kích thước mẫu n đủ lớn.

Dưới đây là minh họa vài dự báo từ các dữ liệu của bài toán trên:

Với giá trị $X = 830.5$, dự báo $Y = 265.44746$

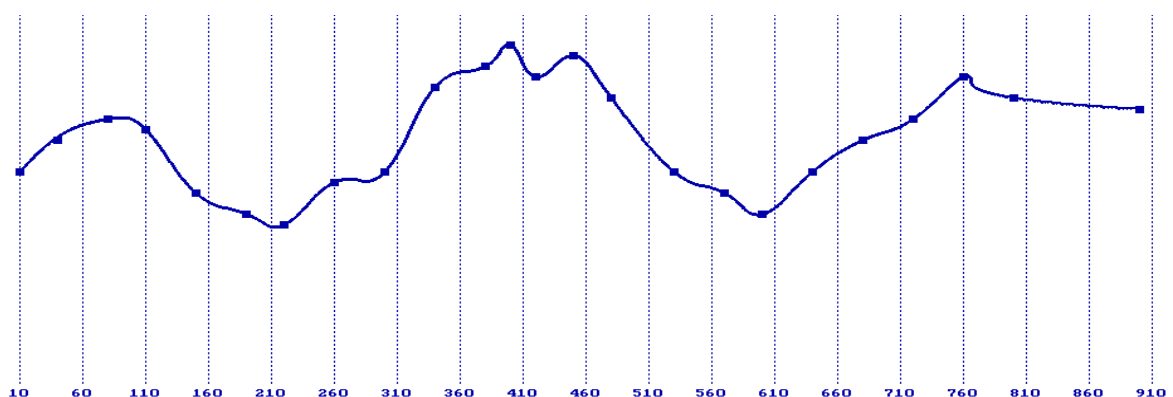
Với giá trị $X = 710.0$, dự báo $Y = 243.00347$

Với giá trị $X = 870.8$, dự báo $Y = 261.47548$

Với giá trị $X = 570.4$, dự báo $Y = 179.74631$

Với giá trị $X = 672.8$, dự báo $Y = 225.95760$

Đường hồi quy B-spline thực nghiệm cho bài toán trên như sau:



4. Kết luận

Về mặt lý thuyết: chúng tôi đã nêu và chứng minh định lý 2.4, để chứng tỏ rằng hàm ngẫu nhiên B-spline có thể phản ánh tốt mối quan hệ giữa X và Y .

Về mặt thực nghiệm: Sử dụng định lý 2.4 chúng tôi đã nêu ra khái niệm hàm hồi qui B-spline $y = L(x)$ và ứng dụng vào dự báo, cho thấy mô hình hồi qui này có thể áp dụng cho mọi kiểu dữ liệu hai chiều, đường hồi qui vạch qua mọi nút dữ liệu (x_i, y_i) và hàm hồi quy $y = L(x)$ có thể dễ dàng được cập nhật $\square$

TÀI LIỆU THAM KHẢO

- [1] W. Kahan and Richard J. Fateman (1999), *Symbolic computation of divided differences*, University of California, Berkeley.
- [2] Lê Hào (2012), *Đường cong B-spline, lý thuyết lập trình và ứng dụng*, Đề tài nghiên cứu khoa học cấp trường, Trường Đại học Phú Yên.
- [3] Lê Hào (2009), *Đường cong B-spline và ứng dụng trong đồ họa 3D*, Thông báo khoa học số 04/2009, Phú Yên.
- [4] Lê Đức Thoang (2015), *Giáo trình phương pháp tính*, Trường Đại học Phú Yên.
- [5] D.I. Kazakevits (2004), *Cơ sở lý thuyết hàm ngẫu nhiên và ứng dụng trong khí tượng thủy văn*, Nxb khoa học kỹ thuật, Hà Nội.
- [6] V.S. Pygatrep (1980), *Lý thuyết hàm ngẫu nhiên và ứng dụng*, Nxb Đại học và THCN, Hà Nội.
- [7] Đào Hữu Hồ (1998), *Xác suất - thống kê*, Chương trình giáo trình đại học, Bộ Giáo dục và Đào tạo, Hà Nội.
- [8] Đinh Văn Gắng (1999), *Lý thuyết xác suất và thống kê*, Nxb Giáo dục, Hà Nội.
- [9] Nguyễn Doãn Phước (2009), *Sử dụng phương pháp nội suy B-SPLINE để đánh giá sai số trong miền tần số của bộ biến đổi tín hiệu DAC*, Tạp chí Khoa học & Công nghệ (số 74 – 2009), Đại học Bách khoa Hà Nội.

Abstract

B-spline random functions and their forecast applications

In this paper we introduce the concept of B-spline random function with empirical samples of two-dimensional random variables (X, Y) and prove that this random function better reflects the relationship between X and Y in case the correlation coefficient $R(X, Y)$ is small, to prove this, we have introduced and proven the theorem (2.4). In the application, we have used the theorem (2.4) to establish a B-spline regression function for forecast.

Keywords: B-spline random function, B-spline regression function, correlation coefficient, empirical samples, random variable.