ƯỚC LƯỢNG ĐỆ QUY TRONG LỚP CÁC MÔ HÌNH BIẾN DẠNG

Đinh Minh Hải

Giảng viên hướng dẫn
ThS. Nguyễn Phát Đạt

Trường Đại học Sư phạm Thành phố Hồ Chí Minh
Khoa Toán - Tin học

Ngày 2 tháng 5 năm 2024

Mục lục

Mục lục

Đặt vấn đề

Xuất phát từ một số nhu cầu:

- Nghiên cứu các hiện tượng xảy ra theo chu kỳ trong tự nhiên (nhiệt độ, lượng mưa hằng năm, ...)

Đặt vấn đề

Xuất phát từ một số nhu cầu:

- Nghiên cứu các hiện tượng xảy ra theo chu kỳ trong tự nhiên (nhiệt độ, lượng mưa hằng năm, ...)
- Đo đạc và dự báo các kết quả hiện tượng trong tự nhiên (nhiệt độ hằng ngày, lưu lượng phương tiện qua một tuyến đường, tham khảo [2], ...)

Đặt vấn đề

Xuất phát từ một số nhu cầu:

- Nghiên cứu các hiện tượng xảy ra theo chu kỳ trong tự nhiên (nhiệt độ, lượng mưa hằng năm, ...)
- Đo đạc và dự báo các kết quả hiện tượng trong tự nhiên (nhiệt độ hằng ngày, lưu lượng phương tiện qua một tuyến đường, tham khảo [2], ...)
- Xây dựng cở sở dữ liệu phục vụ quá trình ra quyết định (dữ liệu điện tâm đồ của bệnh nhân bị loạn nhịp tim, tham khảo [2], ...)

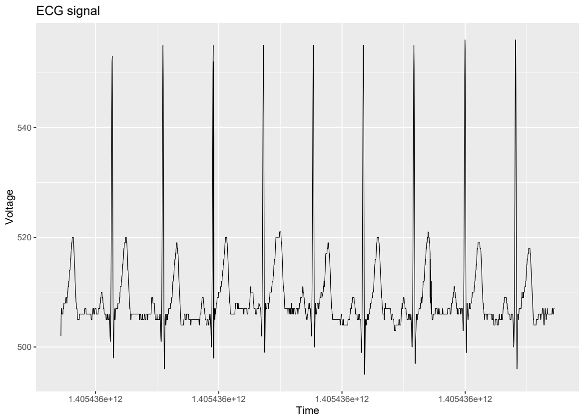
Đặt vấn đề

Xuất phát từ một số nhu cầu:

- Nghiên cứu các hiện tượng xảy ra theo chu kỳ trong tự nhiên (nhiệt độ, lượng mưa hằng năm, ...)
- Đo đạc và dự báo các kết quả hiện tượng trong tự nhiên (nhiệt độ hằng ngày, lưu lượng phương tiện qua một tuyến đường, tham khảo [2], ...)
- Xây dựng cở sở dữ liệu phục vụ quá trình ra quyết định (dữ liệu điện tâm đồ của bệnh nhân bị loạn nhịp tim, tham khảo [2], ...)

$\Rightarrow$ Có động cơ mô phỏng các dữ liệu có chu kỳ bằng phương trình toán học.

Dữ liệu tuần hoàn



Hình 1: Tín hiệu điện tim được đo từ tim một người bình thường là dữ liệu tuần hoàn

Giới thiệu mô hình biến dạng theo chu kỳ

Với mọi $n \geq 0$, các mô hình biến dạng theo chu kỳ được định nghĩa như sau[1]

$$Y_n = h(X_n) + \varepsilon_n, \tag{1}$$

[1]Bercu, B. and Fraysse, P. (2012). *A Robbins-Monro Procedure for Estimation in Semiparametric Regression Models*.

Giới thiệu mô hình biến dạng theo chu kỳ

Với mọi $n \geq 0$, các mô hình biến dạng theo chu kỳ được định nghĩa như sau[1]

$$Y_n = h(X_n) + \varepsilon_n, \tag{1}$$

trong đó

- (X_n) là dãy các quan trắc đã biết,

[1]Bercu, B. and Fraysse, P. (2012). *A Robbins-Monro Procedure for Estimation in Semiparametric Regression Models*.

Giới thiệu mô hình biến dạng theo chu kỳ

Với mọi $n \geq 0$, các mô hình biến dạng theo chu kỳ được định nghĩa như sau[1]

$$Y_n = h(X_n) + \varepsilon_n, \tag{1}$$

trong đó

- (X_n) là dãy các quan trắc đã biết,
- (Y_n) là dãy các quan trắc đã biết,

[1] Bercu, B. and Fraysse, P. (2012). *A Robbins-Monro Procedure for Estimation in Semiparametric Regression Models*.

Giới thiệu mô hình biến dạng theo chu kỳ

Với mọi $n \geq 0$, các mô hình biến dạng theo chu kỳ được định nghĩa như sau[1]

$$Y_n = h(X_n) + \varepsilon_n, \tag{1}$$

trong đó

- (X_n) là dãy các quan trắc đã biết,
- (Y_n) là dãy các quan trắc đã biết,
- (ε_n) là các sai số ngẫu nhiên chưa biết,

[1] Bercu, B. and Fraysse, P. (2012). *A Robbins-Monro Procedure for Estimation in Semiparametric Regression Models*.

Giới thiệu mô hình biến dạng theo chu kỳ

Với mọi $n \geq 0$, các mô hình biến dạng theo chu kỳ được định nghĩa như sau[1]

$$Y_n = h(X_n) + \varepsilon_n, \tag{1}$$

trong đó

- (X_n) là dãy các quan trắc đã biết,
- (Y_n) là dãy các quan trắc đã biết,
- (ε_n) là các sai số ngẫu nhiên chưa biết,
- h là hàm tuần hoàn và có dạng $h(x) = m + \sum_{k=1}^{p} a_k f(x - \theta_k)$ (với f là hàm đặc trưng chưa biết của mô hình, m là trung bình tổng quát, $\theta = (\theta_1, \ldots, \theta_p)^T$ và $a = (a_1, \ldots, a_p)^T$ lần lượt là véc tơ tham số chuyển đổi và véc tơ tham số co giãn chưa biết).

[1] Bercu, B. and Fraysse, P. (2012). *A Robbins-Monro Procedure for Estimation in Semiparametric Regression Models*.

Giới thiệu mô hình biến dạng theo chu kỳ

Với mọi $n \geq 0$, các mô hình biến dạng theo chu kỳ được định nghĩa như sau[1]

$$Y_n = h(X_n) + \varepsilon_n, \tag{1}$$

trong đó

- (X_n) là dãy các quan trắc đã biết,
- (Y_n) là dãy các quan trắc đã biết,
- (ε_n) là các sai số ngẫu nhiên chưa biết,
- h là hàm tuần hoàn và có dạng $h(x) = m + \sum_{k=1}^{p} a_k f(x - \theta_k)$ (với f là hàm đặc trưng chưa biết của mô hình, m là trung bình tổng quát, $\theta = (\theta_1, \ldots, \theta_p)^T$ và $a = (a_1, \ldots, a_p)^T$ lần lượt là véc tơ tham số chuyển đổi và véc tơ tham số co giãn chưa biết).

[1] Bercu, B. and Fraysse, P. (2012). *A Robbins-Monro Procedure for Estimation in Semiparametric Regression Models*.

Giới thiệu

Khóa luận sẽ khảo sát mô hình (1) ứng với với $n \geq 0$, $1 \leq j \leq p, 1 \leq i \leq n$ và $p \geq 2$. Khi đó, mô hình (1) trở thành

$$Y_{i,j} = a_j f\left(X_i - \theta_j\right) + v_j + \varepsilon_{i,j}, \tag{2}$$

Giới thiệu

Khóa luận sẽ khảo sát mô hình (1) ứng với với $n \geq 0$, $1 \leq j \leq p, 1 \leq i \leq n$ và $p \geq 2$. Khi đó, mô hình (1) trở thành

$$Y_{i,j} = a_j f\left(X_i - \theta_j\right) + v_j + \varepsilon_{i,j}, \tag{2}$$

trong đó

- nhiễu $(\varepsilon_{i,j})$ là dãy các biến ngẫu nhiên độc lập có trung bình 0 , phương sai $\mathbb{E}\left[\varepsilon_{i,j}^2\right] = \sigma_j^2$ và độc lập với (X_i), Khóa luận sẽ khảo sát mô hình (1) ứng với với $n \geq 0$, $1 \leq j \leq p, 1 \leq i \leq n$ và $p \geq 2$. Khi đó, mô hình (1) trở thành

$$Y_{i,j} = a_j f\left(X_i - \theta_j\right) + v_j + \varepsilon_{i,j}, \tag{2}$$

trong đó

- nhiễu $(\varepsilon_{i,j})$ là dãy các biến ngẫu nhiên độc lập có trung bình 0 , phương sai $\mathbb{E}\left[\varepsilon_{i,j}^2\right] = \sigma_j^2$ và độc lập với (X_i),
- (X_i) là dãy các biến ngẫu nhiên độc lập, cùng phân phối,

Giới thiệu

Khóa luận sẽ khảo sát mô hình (1) ứng với với $n \geq 0$, $1 \leq j \leq p, 1 \leq i \leq n$ và $p \geq 2$. Khi đó, mô hình (1) trở thành

$$Y_{i,j} = a_j f\left(X_i - \theta_j\right) + v_j + \varepsilon_{i,j}, \tag{2}$$

trong đó

- nhiễu $(\varepsilon_{i,j})$ là dãy các biến ngẫu nhiên độc lập có trung bình 0 , phương sai $\mathbb{E}\left[\varepsilon_{i,j}^2\right] = \sigma_j^2$ và độc lập với (X_i),
- (X_i) là dãy các biến ngẫu nhiên độc lập, cùng phân phối,
- f là hàm tuần hoàn chưa biết,

Giới thiệu

Khóa luận sẽ khảo sát mô hình (1) ứng với với $n \geq 0$, $1 \leq j \leq p, 1 \leq i \leq n$ và $p \geq 2$. Khi đó, mô hình (1) trở thành

$$Y_{i,j} = a_j f\left(X_i - \theta_j\right) + v_j + \varepsilon_{i,j}, \tag{2}$$

trong đó

- nhiễu $(\varepsilon_{i,j})$ là dãy các biến ngẫu nhiên độc lập có trung bình 0 , phương sai $\mathbb{E}\left[\varepsilon_{i,j}^2\right] = \sigma_j^2$ và độc lập với (X_i),
- (X_i) là dãy các biến ngẫu nhiên độc lập, cùng phân phối,
- f là hàm tuần hoàn chưa biết,
- a_j, θ_j, v_j là các số thực, với mọi $1 \leq j \leq p$.

1.2 Các giả thiết trong mô hình

Các giả thiết sau sẽ được sử dụng trong mô hình trên [2]

[2] Bercu, B. and Fraysse, P. (2012). *A Robbins-Monro Procedure for Estimation in Semiparametric Regression Models*.

1.2 Các giả thiết trong mô hình

Các giả thiết sau sẽ được sử dụng trong mô hình trên [2]

($\mathcal{H}_1$) Các quan trắc (X_i) là dãy các biến ngẫu nhiên độc lập, cùng phân phối với hàm mật độ xác suất g, dương trên giá $[-1/2; 1/2]$. Hơn nữa, g liên tục trên $\mathbb{R}$, khả vi cấp 2 và đạo hàm bị chặn.

[2] Bercu, B. and Fraysse, P. (2012). *A Robbins-Monro Procedure for Estimation in Semiparametric Regression Models*.

1.2 Các giả thiết trong mô hình

Các giả thiết sau sẽ được sử dụng trong mô hình trên [2]

($\mathcal{H}_1$) Các quan trắc (X_i) là dãy các biến ngẫu nhiên độc lập, cùng phân phối với hàm mật độ xác suất g, dương trên giá $[-1/2; 1/2]$. Hơn nữa, g liên tục trên $\mathbb{R}$, khả vi cấp 2 và đạo hàm bị chặn.
Trong suốt khóa luận, giả sử rằng ta đã biết hàm mật độ xác suất g.

[2]Bercu, B. and Fraysse, P. (2012). *A Robbins-Monro Procedure for Estimation in Semiparametric Regression Models*.

1.2 Các giả thiết trong mô hình

Các giả thiết sau sẽ được sử dụng trong mô hình trên [2]

($\mathcal{H}_1$) Các quan trắc (X_i) là dãy các biến ngẫu nhiên độc lập, cùng phân phối với hàm mật độ xác suất g, dương trên giá $[-1/2; 1/2]$. Hơn nữa, g liên tục trên $\mathbb{R}$, khả vi cấp 2 và đạo hàm bị chặn.
Trong suốt khóa luận, giả sử rằng ta đã biết hàm mật độ xác suất g.

($\mathcal{H}_2$) f là hàm đối xứng, bị chặn, tuần hoàn với chu kỳ 1 .

[2] Bercu, B. and Fraysse, P. (2012). *A Robbins-Monro Procedure for Estimation in Semiparametric Regression Models*.

1.2 Các giả thiết trong mô hình

Các giả thiết sau sẽ được sử dụng trong mô hình trên [2]

($\mathcal{H}_1$) Các quan trắc (X_i) là dãy các biến ngẫu nhiên độc lập, cùng phân phối với hàm mật độ xác suất g, dương trên giá $[-1/2; 1/2]$. Hơn nữa, g liên tục trên $\mathbb{R}$, khả vi cấp 2 và đạo hàm bị chặn.
Trong suốt khóa luận, giả sử rằng ta đã biết hàm mật độ xác suất g.

($\mathcal{H}_2$) f là hàm đối xứng, bị chặn, tuần hoàn với chu kỳ 1 .
Ta có thể tìm được hai bộ ba vector tham số (a, θ, v) và (a^*, θ^*, v^*) thỏa mãn

$$a_j f\left(x - \theta_j\right) + v_j = a_j^* f^*\left(x - \theta_j^*\right) + v_j^* . \tag{3}$$

[2] Bercu, B. and Fraysse, P. (2012). *A Robbins-Monro Procedure for Estimation in Semiparametric Regression Models.*

1.2 Các giả thiết trong mô hình

Các giả thiết sau sẽ được sử dụng trong mô hình trên [2]

($\mathcal{H}_1$) Các quan trắc (X_i) là dãy các biến ngẫu nhiên độc lập, cùng phân phối với hàm mật độ xác suất g, dương trên giá $[-1/2; 1/2]$. Hơn nữa, g liên tục trên $\mathbb{R}$, khả vi cấp 2 và đạo hàm bị chặn.
Trong suốt khóa luận, giả sử rằng ta đã biết hàm mật độ xác suất g.

($\mathcal{H}_2$) f là hàm đối xứng, bị chặn, tuần hoàn với chu kỳ 1 .
Ta có thể tìm được hai bộ ba vector tham số (a, θ, v) và (a^*, θ^*, v^*) thỏa mãn

$$a_j f\left(x - \theta_j\right) + v_j = a_j^* f^* \left(x - \theta_j^*\right) + v_j^* . \tag{3}$$

Tác giả Philippe Fraysse đề xuất thêm các giả thiết sau:

($\mathcal{H}_3$) $\int_{-1/2}^{1/2} f(x)dx = 0,$

($\mathcal{H}_4$) $a_1 = 1, \theta_1 = 0$ và $\max_{1\leq j\leq p} |\theta_j| < \frac{1}{4}$.

[2]Bercu, B. and Fraysse, P. (2012). *A Robbins-Monro Procedure for Estimation in Semiparametric Regression Models.*

Mục lục

Martingale[3]

Định nghĩa 2.1.2. Định nghĩa martingale

Giả sử rằng $X = (X_n)$ là dãy các biến ngẫu nhiên khả tích và tương thích với bộ lọc $\mathbb{F}$. Với mọi n, X được gọi là:

[3] Duflo, M. (1997). *Random Iterative Models. Applications of Mathematics(New York)*, Springer, Berlin.

Martingale[3]

Định nghĩa 2.1.2. Định nghĩa martingale

Giả sử rằng $X = (X_n)$ là dãy các biến ngẫu nhiên khả tích và tương thích với bộ lọc $\mathbb{F}$. Với mọi n, X được gọi là:

- Martingale nếu

$$\mathbb{E}\left[X_{n+1} \mid \mathcal{F}_n\right] = X_n \text{ h.c.c } . \tag{4}$$

[3] Duflo, M. (1997). *Random Iterative Models. Applications of Mathematics(New York)*, Springer, Berlin.

Martingale[3]

Định nghĩa 2.1.2. Định nghĩa martingale

Giả sử rằng $X = (X_n)$ là dãy các biến ngẫu nhiên khả tích và tương thích với bộ lọc $\mathbb{F}$. Với mọi n, X được gọi là:

- Martingale nếu
$$\mathbb{E}\left[X_{n+1} \mid \mathcal{F}_n\right] = X_n \text{ h.c.c .} \tag{4}$$

- Martingale dưới nếu
$$\mathbb{E}\left[X_{n+1} \mid \mathcal{F}_n\right] \geq X_n \text{ h.c.c .} \tag{5}$$

[3] Duflo, M. (1997). *Random Iterative Models. Applications of Mathematics(New York)*, Springer, Berlin.

Martingale[3]

Định nghĩa 2.1.2. Định nghĩa martingale

Giả sử rằng $X = (X_n)$ là dãy các biến ngẫu nhiên khả tích và tương thích với bộ lọc $\mathbb{F}$. Với mọi n, X được gọi là:

- Martingale nếu
$$\mathbb{E}\left[X_{n+1} \mid \mathcal{F}_n\right] = X_n \text{ h.c.c .} \tag{4}$$

- Martingale dưới nếu
$$\mathbb{E}\left[X_{n+1} \mid \mathcal{F}_n\right] \geq X_n \text{ h.c.c .} \tag{5}$$

- Martingale trên nếu
$$\mathbb{E}\left[X_{n+1} \mid \mathcal{F}_n\right] \leq X_n \text{ h.c.c .} \tag{6}$$

[3]Duflo, M. (1997). *Random Iterative Models. Applications of Mathematics(New York)*, Springer, Berlin.

Biến phân bình phương dự báo được[4]

Định nghĩa 2.1.3. Định nghĩa biến phân bình phương dự báo được cho martingale

Biến phân bình phương (hoặc đặc trưng bình phương) dự báo được gắn liền với martingale (X_n) với $\langle X \rangle_0 = 0$ và với mọi $n \geq 1$, được định nghĩa

[4]Duflo, M. (1997). *Random Iterative Models. Applications of Mathematics(New York)*, Springer, Berlin.

Biến phân bình phương dự báo được[4]

Định nghĩa 2.1.3. Định nghĩa biến phân bình phương dự báo được cho martingale

Biến phân bình phương (hoặc đặc trưng bình phương) dự báo được gắn liền với martingale (X_n) với $\langle X\rangle_0 = 0$ và với mọi $n \geq 1$, được định nghĩa

$$\langle X\rangle_n = \sum_{k=1}^{n} \mathbb{E}\left[(\Delta X_k)^2 \mid \mathcal{F}_{k-1}\right], \tag{7}$$

[4]Duflo, M. (1997). *Random Iterative Models. Applications of Mathematics(New York)*, Springer, Berlin.

Biến phân bình phương dự báo được[4]

Định nghĩa 2.1.3. Định nghĩa biến phân bình phương dự báo được cho martingale

Biến phân bình phương (hoặc đặc trưng bình phương) dự báo được gắn liền với martingale (X_n) với $\langle X\rangle_0 = 0$ và với mọi $n \geq 1$, được định nghĩa

$$\langle X\rangle_n = \sum_{k=1}^{n} \mathbb{E}\left[(\Delta X_k)^2 \mid \mathcal{F}_{k-1}\right], \tag{7}$$

trong đó $\Delta X_k = X_k - X_{k-1}$.
Ta ký hiệu

$$\langle X\rangle_\infty = \lim_{n\to\infty} \langle X\rangle_n. \tag{8}$$

[4]Duflo, M. (1997). *Random Iterative Models. Applications of Mathematics(New York)*, Springer, Berlin.

Định lý giới hạn trung tâm

Định lý 2.1.4. Định lý giới hạn trung tâm cho martingale

Cho (X_n) là một martingale bình phương khả tích và (a_n) là dãy các số thực dương tăng đến vô cùng. Giả sử rằng:

Định lý giới hạn trung tâm

Định lý 2.1.4. Định lý giới hạn trung tâm cho martingale

Cho (X_n) là một martingale bình phương khả tích và (a_n) là dãy các số thực dương tăng đến vô cùng. Giả sử rằng:

- Tồn tại một giới hạn tất định $l > 0$ sao cho

$$\frac{\langle X \rangle_n}{a_n} \xrightarrow{\mathcal{P}} l. \tag{10}$$

Định lý giới hạn trung tâm

Định lý 2.1.4. Định lý giới hạn trung tâm cho martingale

Cho (X_n) là một martingale bình phương khả tích và (a_n) là dãy các số thực dương tăng đến vô cùng. Giả sử rằng:

- Tồn tại một giới hạn tất định $l > 0$ sao cho

$$\frac{\langle X \rangle_n}{a_n} \xrightarrow{\mathcal{P}} l. \tag{10}$$

- Điều kiện Lindeberg

  Với mọi $\varepsilon > 0$,

$$\frac{1}{a_n} \sum_{k=1}^{n} \mathbb{E}\left[|\Delta X_k|^2 \, \mathbb{I}_{\{|\Delta X_k| \geq \varepsilon \sqrt{a_n}\}} \mid \mathcal{F}_{k-1} \right] \xrightarrow{\mathcal{P}} 0. \tag{11}$$

Định lý giới hạn trung tâm(tt)

Khi đó

$$\frac{1}{\sqrt{a_n}} X_n \xrightarrow{\mathcal{L}} \mathcal{N}(0, I). \tag{12}$$

Hơn nữa, khi $I > 0$,

$$\sqrt{a_n} \left(\frac{X_n}{\langle X \rangle_n} \right) \xrightarrow{\mathcal{L}} \mathcal{N}\left(0, I^{-1}\right). \tag{13}$$

Luật mạnh số lớn[5]

Định lý 2.1.5. Luật mạnh số lớn cho martingale

Cho (X_n) là một martingale bình phương khả tích và ký hiệu $\langle X\rangle_\infty = \lim_{n\to\infty}\langle X\rangle_n$.

- Nếu $\langle X\rangle_\infty = \infty$, ta được $\lim_{n\to\infty} \frac{X_n}{\langle X\rangle_n} = 0$ hầu chắc chắn. Hơn nữa, với mọi $\gamma > 0$, ta có

$$X_n/\langle X\rangle_n = \mathrm{o}\left(\left((\ln\langle X\rangle_n)^{1+\gamma}/\langle X\rangle_n\right)^{1/2}\right) \tag{14}$$

[5] Duflo, M. (1997). *Random Iterative Models. Applications of Mathematics(New York)*, Springer, Berlin.

Luật mạnh số lớn[5]

Định lý 2.1.5. Luật mạnh số lớn cho martingale

Cho (X_n) là một martingale bình phương khả tích và ký hiệu $\langle X\rangle_\infty = \lim_{n\to\infty}\langle X\rangle_n$.

- Nếu $\langle X\rangle_\infty = \infty$, ta được $\lim_{n\to\infty} \frac{X_n}{\langle X\rangle_n} = 0$ hầu chắc chắn. Hơn nữa, với mọi $\gamma > 0$, ta có

$$X_n/\langle X\rangle_n = \mathrm{o}\left(\left((\ln\langle X\rangle_n)^{1+\gamma}/\langle X\rangle_n\right)^{1/2}\right) \tag{14}$$

- Nếu $\langle X\rangle_\infty < \infty$ thì (X_n) hội tụ hầu chắc chắn đến biến ngẫu nhiên bình phương khả tích X_∞.

[5] Duflo, M. (1997). *Random Iterative Models. Applications of Mathematics(New York)*, Springer, Berlin.

Ước lượng Robbins-Monro[6]

Cho (γ_n) là dãy giảm các số thực dương thỏa mãn

$$\sum_{n=1}^{+\infty} \gamma_n = +\infty \text{ và } \sum_{n=1}^{+\infty} \gamma_n^2 < +\infty. \tag{18}$$

[6] Bercu, B. and Fraysse, P. (2012). *A Robbins-Monro Procedure for Estimation in Semiparametric Regression Models*.

Ước lượng Robbins-Monro[6]

Cho (γ_n) là dãy giảm các số thực dương thỏa mãn

$$\sum_{n=1}^{+\infty} \gamma_n = +\infty \text{ và } \sum_{n=1}^{+\infty} \gamma_n^2 < +\infty. \tag{18}$$

Ước lượng Robbins-Monro được xác định bởi

$$\hat{\theta}_{n+1} = \hat{\theta}_n + \gamma_{n+1} \left(T_{n+1} - f(\theta) \right), \tag{19}$$

[6] Bercu, B. and Fraysse, P. (2012). *A Robbins-Monro Procedure for Estimation in Semiparametric Regression Models*.

Ước lượng Robbins-Monro[6]

Cho (γ_n) là dãy giảm các số thực dương thỏa mãn

$$\sum_{n=1}^{+\infty} \gamma_n = +\infty \text{ và } \sum_{n=1}^{+\infty} \gamma_n^2 < +\infty. \tag{18}$$

Ước lượng Robbins-Monro được xác định bởi

$$\hat{\theta}_{n+1} = \hat{\theta}_n + \gamma_{n+1}\left(T_{n+1} - f(\theta)\right), \tag{19}$$

trong đó $f(\theta)$ đã biết và T_{n+1} là biến ngẫu nhiên thỏa mãn

$$\mathbb{E}\left[T_{n+1} \mid \mathcal{F}_n\right] = f\left(\hat{\theta}_n\right).$$

Giả sử rằng f là hàm giảm. Khi đó

$$\lim_{n\to\infty} \hat{\theta}_n = \theta \text{ h.c.c }. \tag{27}$$

[6]Bercu, B. and Fraysse, P. (2012). *A Robbins-Monro Procedure for Estimation in Semiparametric Regression Models*.

Ước lượng Robbins-Monro[6]

Cho (γ_n) là dãy giảm các số thực dương thỏa mãn

$$\sum_{n=1}^{+\infty} \gamma_n = +\infty \text{ và } \sum_{n=1}^{+\infty} \gamma_n^2 < +\infty. \tag{18}$$

Ước lượng Robbins-Monro được xác định bởi

$$\hat{\theta}_{n+1} = \hat{\theta}_n + \gamma_{n+1} \left(T_{n+1} - f(\theta) \right), \tag{19}$$

trong đó $f(\theta)$ đã biết và T_{n+1} là biến ngẫu nhiên thỏa mãn

$$\mathbb{E}\left[T_{n+1} \mid \mathcal{F}_n\right] = f\left(\hat{\theta}_n\right).$$

Giả sử rằng f là hàm giảm. Khi đó

$$\lim_{n\to\infty} \hat{\theta}_n = \theta \text{ h.c.c }. \tag{27}$$

Hơn nữa, khi $-2f'(\theta) > 1$, ta có tính tiệm cận chuẩn

$$\sqrt{n}\left(\hat{\theta}_n - \theta\right) \xrightarrow{\mathcal{L}} \mathcal{N}\left(0, \xi^2(\theta)\right). \tag{28}$$

[6]Bercu, B. and Fraysse, P. (2012). *A Robbins-Monro Procedure for Estimation in Semiparametric Regression Models*.

Hàm hạt nhân và ước lượng Nadaraya-Watson[7]

Định nghĩa 2.1.9

Cho (X_n) là dãy các biến ngẫu nhiên độc lập, cùng phân phối và hàm mật độ f chưa biết. Cho K là hàm đối xứng, dương, bị chặn và có giá compact sao cho

$$\int_{\mathbb{R}} K(x)dx = 1 \text{ và } \int_{\mathbb{R}} K^2(x)dx = v^2 \tag{29}$$

thì ta gọi K là hàm hạt nhân.

[7]Noda, K. (2007). Estimation of a Regression Function by The Parzen kernel-type Density Estimatiors, Ann Inst Statist Math, 221-234.

Hàm hạt nhân và ước lượng Nadaraya-Watson[8]

Hàm hạt nhân sử dụng chính trong khóa luận là hàm hạt nhân đều

$$K_a(x) = \frac{1}{2a} I_{\{|x| \le a\}}$$

[8]Noda, K. (2007). Estimation of a Regression Function by The Parzen kernel-type Density Estimatiors, Ann Inst Statist Math, 221-234.

Hàm hạt nhân và ước lượng Nadaraya-Watson

Ước lượng đệ quy Nadaraya-Watson của hàm f được xác định bởi

$$\widehat{f}_n(x) = \frac{\sum_{k=1}^{n} W_k(x) Y_k}{\sum_{k=1}^{n} W_k(x)} \tag{30}$$

Hàm hạt nhân và ước lượng Nadaraya-Watson

Ước lượng đệ quy Nadaraya-Watson của hàm f được xác định bởi

$$\widehat{f}_n(x) = \frac{\sum_{k=1}^{n} W_k(x) Y_k}{\sum_{k=1}^{n} W_k(x)} \tag{30}$$

trong đó

$$W_k(x) = \frac{1}{h_k} K\left(\frac{X_k - x}{h_k}\right) \tag{31}$$

Hàm hạt nhân và ước lượng Nadaraya-Watson

Ước lượng đệ quy Nadaraya-Watson của hàm f được xác định bởi

$$\widehat{f}_n(x) = \frac{\sum_{k=1}^{n} W_k(x) Y_k}{\sum_{k=1}^{n} W_k(x)} \tag{30}$$

trong đó

$$W_k(x) = \frac{1}{h_k} K\left(\frac{X_k - x}{h_k}\right) \tag{31}$$

Băng tần (h_n) là dãy các số thực dương, h_n giảm dần về 0 và nh_n tiến ra vô cùng. Với $0 < \alpha < 1$, ta thường dùng $h_n = 1/n^{\alpha}$.

Hàm hạt nhân và ước lượng Nadaraya-Watson[9]

Định lý 2.1.10. Định lý về sự hội tụ của ước lượng đệ quy

Với mọi $x \in \mathbb{R}$, nếu $\widehat{f_n}(x)$ là ước lượng đệ quy Nadaraya-Watson của hàm f thì

$$\lim_{n \to \infty} \widehat{f_n}(x) = f(x) \text{ h.c.c } . \tag{32}$$

[9]Noda, K. (2007). Estimation of a Regression Function by The Parzen kernel-type Density Estimatiors, Ann Inst Statist Math, 221-234.

Hàm hạt nhân và ước lượng Nadaraya-Watson[9]

Định lý 2.1.10. Định lý về sự hội tụ của ước lượng đệ quy

Với mọi $x \in \mathbb{R}$, nếu $\widehat{f}_n(x)$ là ước lượng đệ quy Nadaraya-Watson của hàm f thì

$$\lim_{n\to\infty} \widehat{f}_n(x) = f(x) \text{ h.c.c .} \tag{32}$$

Định lý 2.1.11. Định lý về tính tiệm cận chuẩn của ước lượng đệ quy

Giả sử rằng (ε_n) có mô-men hữu hạn, bậc lớn hơn 2. Với mọi $x \in \mathbb{R}$, nếu $\frac{1}{5} < \alpha < 1$ thì

$$\sqrt{nh_n}\left(\widehat{f}_n(x) - f(x)\right) \xrightarrow{\mathcal{L}} \mathcal{N}\left(0, \frac{\sigma^2 v^2}{(1+\alpha)g(x)}\right). \tag{33}$$

[9] Noda, K. (2007). Estimation of a Regression Function by The Parzen kernel-type Density Estimatiors, Ann Inst Statist Math, 221-234.

Mục lục

Ước lượng tham số chiều cao v

Với $1 \leq j \leq p$, một ước lượng vững $\widehat{v}_n$ của v được xác định bởi

$$\widehat{v}_{n,j} = \frac{1}{n} \sum_{i=1}^{n} \frac{Y_{i,j}}{g\left(X_i\right)}. \tag{34}$$

Ước lượng tham số chiều cao v

Định lý 3.1.1 Định lý sự hội tụ ước lượng tham số chiều cao $\widehat{v}_n$

Giả sử rằng $(\mathcal{H}_1)$ đến $(\mathcal{H}_4)$ được thỏa mãn, ta có:

- $\widehat{v}_n$ hội tụ hầu chắc chắn về v

$$\lim_{n\to\infty} \widehat{v}_n = v \text{ h.c.c } . \tag{35}$$

Ước lượng tham số chiều cao v

Định lý 3.1.1 Định lý sự hội tụ ước lượng tham số chiều cao $\widehat{v}_n$

Giả sử rằng $(\mathcal{H}_1)$ đến $(\mathcal{H}_4)$ được thỏa mãn, ta có:

- $\widehat{v}_n$ hội tụ hầu chắc chắn về v

$$\lim_{n\to\infty} \widehat{v}_n = v \text{ h.c.c }. \tag{35}$$

- Tính tiệm cận chuẩn

$$\sqrt{n}\,(\widehat{v}_n - v) \xrightarrow{\mathcal{L}} \mathcal{N}_p\left(0, \mathrm{Cov}\left(\frac{Y}{g(X)}\right)\right) \tag{36}$$

Ước lượng tham số chiều cao v

Định lý 3.1.1 Định lý sự hội tụ ước lượng tham số chiều cao $\widehat{v}_n$

Giả sử rằng $(\mathcal{H}_1)$ đến $(\mathcal{H}_4)$ được thỏa mãn, ta có:

- $\widehat{v}_n$ hội tụ hầu chắc chắn về v

$$\lim_{n\to\infty} \widehat{v}_n = v \text{ h.c.c } . \tag{35}$$

- Tính tiệm cận chuẩn

$$\sqrt{n}\left(\widehat{v}_n - v\right) \xrightarrow{\mathcal{L}} \mathcal{N}_p\left(0, \mathrm{Cov}\left(\frac{Y}{g(X)}\right)\right) \tag{36}$$

- Luật mạnh dạng toàn phương như sau

$$\lim_{n\to\infty} \frac{1}{\log(n)} \sum_{i=1}^{n} \left(\widehat{v}_i - v\right)\left(\widehat{v}_i - v\right)^T = \mathrm{Cov}\left(\frac{Y}{g(X)}\right) \text{ h.c.c } . \tag{37}$$

Ước lượng tham số chuyển đổi θ

Với mọi $t \in \mathbb{R}^p$, ta đặt hàm bổ trợ ϕ có dạng

$$\phi(t) = \mathbb{E}\left[D(X,t)\begin{pmatrix} a_1 f\left(X-\theta_1\right) \\ \vdots \\ a_p f\left(X-\theta_p\right) \end{pmatrix}\right] \tag{38}$$

$$D(X,t) = \frac{1}{g(X)} \operatorname{diag}\left(\sin\left(2\pi\left(X-t_1\right)\right), \ldots, \sin\left(2\pi\left(X-t_p\right)\right)\right) \tag{39}$$

Ước lượng tham số chuyển đổi θ

Với mọi $t \in \mathbb{R}^p$, ta đặt hàm bổ trợ ϕ có dạng

$$\phi(t) = \mathbb{E}\left[D(X,t)\begin{pmatrix} a_1 f\left(X-\theta_1\right) \\ \vdots \\ a_p f\left(X-\theta_p\right)\end{pmatrix}\right] \tag{38}$$

$$D(X,t) = \frac{1}{g(X)}\operatorname{diag}\left(\sin\left(2\pi\left(X-t_1\right)\right),\ldots,\sin\left(2\pi\left(X-t_p\right)\right)\right) \tag{39}$$

Ta định nghĩa phép chiếu của $x \in \mathbb{R}$ trên $K = [-1/4; 1/4]$ như sau

$$\pi_K(x) = \begin{cases} x \text{ nếu } |x| \leq 1/4 \\ 1/4 \text{ nêu } x \geq 1/4 \\ -1/4 \text{ nêu } x \leq -1/4 \end{cases}$$

Ước lượng tham số chuyển đổi θ

Với mọi $t \in \mathbb{R}^p$, ta đặt hàm bổ trợ ϕ có dạng

$$\phi(t) = \mathbb{E}\left[D(X,t)\begin{pmatrix} a_1 f(X-\theta_1) \\ \vdots \\ a_p f(X-\theta_p)\end{pmatrix}\right] \tag{38}$$

$$D(X,t) = \frac{1}{g(X)}\operatorname{diag}\left(\sin\left(2\pi\left(X-t_1\right)\right),\ldots,\sin\left(2\pi\left(X-t_p\right)\right)\right) \tag{39}$$

Ta định nghĩa phép chiếu của $x \in \mathbb{R}$ trên $K = [-1/4; 1/4]$ như sau

$$\pi_K(x) = \begin{cases} x \text{ nếu } |x| \leq 1/4 \\ 1/4 \text{ nêu } x \geq 1/4 \\ -1/4 \text{ nêu } x \leq -1/4 \end{cases}$$

Cho (γ_n) là dãy giảm các số thực dương thỏa mãn

$$\sum_{n=1}^{+\infty}\gamma_n = +\infty \text{ và } \sum_{n=1}^{+\infty}\gamma_n^2 < +\infty \tag{40}$$

Ước lượng tham số chuyển đổi θ

Với mọi $1 \leq j \leq p$, ta ước lượng θ_j thông qua dãy $\left(\widehat{\theta}_{n,j}\right)$ với

$$\widehat{\theta}_{n+1,j} = \pi_K \left(\widehat{\theta}_{n,j} + \gamma_{n+1} \operatorname{sign}\left(a_j f_1\right) T_{n+1,j}\right) \tag{41}$$

trong đó, giá trị ban đầu $\widehat{\theta}_0 \in K^p$ và vector ngẫu nhiên T_{n+1} được xác định bởi

$$T_{n+1} = D\left(X_{n+1}, \widehat{\theta}_n\right) Y_{n+1} \tag{42}$$

Ước lượng tham số chuyển đổi θ

Với mọi $1 \leq j \leq p$, ta ước lượng θ_j thông qua dãy $\left(\widehat{\theta}_{n,j}\right)$ với

$$\widehat{\theta}_{n+1,j} = \pi_K \left(\widehat{\theta}_{n,j} + \gamma_{n+1} \operatorname{sign}\left(a_j f_1\right) T_{n+1,j}\right) \tag{41}$$

trong đó, giá trị ban đầu $\widehat{\theta}_0 \in K^p$ và vector ngẫu nhiên T_{n+1} được xác định bởi

$$T_{n+1} = D\left(X_{n+1}, \widehat{\theta}_n\right) Y_{n+1} \tag{42}$$

và f_1 là hệ số Fourier đầu tiên của f

$$f_1 = \int_{-1/2}^{1/2} \cos(2\pi x) f(x) dx$$

Định lý 3.2.1. Định lý về sự hội tụ hầu chắc chắn của $\widehat{\theta}_n$

Giả sử rằng các giả thiết từ $(\mathcal{H}_1)$ đến $(\mathcal{H}_4)$ được thỏa mãn.

$$\lim_{n\to\infty} \widehat{\theta}_n = \theta \text{ h.c.c}$$

Hơn nữa, khi $|\theta_j| < 1/4$ thì số lần mà biến ngẫu nhiên $\widehat{\theta}_{n,j} + \gamma_{n+1} \operatorname{sign}(a_j f_1)\, T_{n+1,j}$ nằm ngoài K là hữu hạn (hầu chắc chắn).

Ước lượng tham số chuyển đổi θ

Định lý 3.2.2. Đinh lý về tính tiệm cận chuẩn của ước lượng $\widehat{\theta}_n$

Giả sử rằng các giả thiết từ $(\mathcal{H}_1)$ đến $(\mathcal{H}_4)$ được thỏa mãn. Ta cũng giả định rằng mô-men bậc lớn hơn 2 của $(\varepsilon_{i,j})$ hữu hạn và $4\pi\,|f_1|\min_{1\leq j\leq p}|a_j| > 1$. Khi đó, ta có tính tiệm cận chuẩn

$$\sqrt{n}\left(\widehat{\theta}_n - \theta\right) \xrightarrow{\mathcal{L}} \mathcal{N}_p(0, \Sigma(\theta)) \tag{43}$$

Định lý 3.2.2. Đinh lý về tính tiệm cận chuẩn của ước lượng $\widehat{\theta}_n$

Giả sử rằng các giả thiết từ $(\mathcal{H}_1)$ đến $(\mathcal{H}_4)$ được thỏa mãn. Ta cũng giả định rằng mô-men bậc lớn hơn 2 của $(\varepsilon_{i,j})$ hữu hạn và $4\pi \left|f_1\right| \min_{1\leq j\leq p} \left|a_j\right| > 1$. Khi đó, ta có tính tiệm cận chuẩn

$$\sqrt{n}\left(\widehat{\theta}_n - \theta\right) \xrightarrow{\mathcal{L}} \mathcal{N}_p(0, \Sigma(\theta)) \tag{43}$$

trong đó với mọi $t \in \mathbb{R}^p$, ta gọi hàm bổ trợ φ,

$$\varphi(t) = E\left[V(t)V(t)^T\right] \tag{44}$$

và $V(t) = \operatorname{diag}\left(\operatorname{sign}\left(a_1 f_1\right), \ldots, \operatorname{sign}\left(a_p f_1\right)\right) D(X,t) Y$.
Khi $4\pi \left|f_1\right| \min_{1\leq j\leq p} \left|a_j\right| > 1$, với mọi $1 \leq k, l \leq p$ thì

$$\Sigma(\theta)_{k,l} = \frac{\varphi(\theta)_{k,l}}{2\pi\left(\left|a_k\right| + \left|a_l\right|\right)\left|f_1\right| - 1} \tag{45}$$

Ước lượng tham số chuyển đổi θ

Định lý 3.2.3. Luật loga-lặp và luật mạnh dạng toàn phương của ước lượng $\widehat{\theta}_n$

Giả sử rằng các giả thiết từ $(\mathcal{H}_1)$ đến $(\mathcal{H}_4)$ được thỏa mãn. Hơn nữa, giả định rằng các mô-men bậc lớn hơn 2 của $(\varepsilon_{i,j})$ hữu hạn, đồng thời $4\pi\,|f_1|\min_{1\le j\le p}|a_j| > 1$. Với mọi $W \in \mathbb{R}^p$, ta có luật loga-lặp

$$\begin{aligned}
&\limsup_{n\to\infty}\left(\frac{n}{2\log(\log(n))}\right)^{1/2} W^T\left(\widehat{\theta}_n-\theta\right)\\
&= -\liminf_{n\to\infty}\left(\frac{n}{2\log(\log(n))}\right)^{1/2} W^T\left(\widehat{\theta}_n-\theta\right)\\
&= \sqrt{W^T\Sigma(\theta)W}\ \text{h.c.c}\,. \qquad (46)
\end{aligned}$$

Ước lượng tham số chuyển đổi θ

Định lý 3.2.3. Luật loga-lặp và luật mạnh dạng toàn phương của ước lượng $\widehat{\theta}_n$

Đặc biệt

$$\limsup_{n\to\infty}\left(\frac{n}{2\log(\log(n))}\right)^{1/2}\left(\widehat{\theta}_n-\theta\right)\left(\widehat{\theta}_n-\theta\right)^T=\Sigma(\theta)\ \text{h.c.c}\ . \tag{47}$$

Ước lượng tham số chuyển đổi θ

Định lý 3.2.3. Luật loga-lặp và luật mạnh dạng toàn phương của ước lượng $\widehat{\theta}_n$

Đặc biệt

$$\limsup_{n\to\infty} \left(\frac{n}{2\log(\log(n))}\right)^{1/2} \left(\widehat{\theta}_n - \theta\right)\left(\widehat{\theta}_n - \theta\right)^T = \Sigma(\theta) \text{ h.c.c .} \qquad (47)$$

Hơn nữa, ta cũng có luật mạnh dạng toàn phương

$$\lim_{n\to\infty} \frac{1}{\log(n)} \sum_{i=1}^{n} \left(\widehat{\theta}_i - \theta\right)\left(\widehat{\theta}_i - \theta\right)^T = \Sigma(\theta) \text{ h.c.c .} \qquad (48)$$

Ước lượng tham số co giãn a

Với mọi $n \geq 1$ và $1 \leq j \leq p$, gọi

$$\widehat{a}_{n,j} = \frac{1}{nf_1} \sum_{i=1}^{n} \frac{\cos\left(2\pi\left(X_i - \widehat{\theta}_{i-1,j}\right)\right)}{g\left(X_i\right)} Y_{i,j} \tag{49}$$

Ước lượng tham số co giãn a

Với mọi $n \geq 1$ và $1 \leq j \leq p$, gọi

$$\widehat{a}_{n,j} = \frac{1}{nf_1} \sum_{i=1}^{n} \frac{\cos\left(2\pi\left(X_i - \widehat{\theta}_{i-1,j}\right)\right)}{g\left(X_i\right)} Y_{i,j} \tag{49}$$

$$\tilde{a}_{n,j} = \frac{1}{n\widehat{f}_{1,n}} \sum_{i=1}^{n} \frac{\cos\left(2\pi\left(X_i - \widehat{\theta}_{i-1,j}\right)\right)}{g\left(X_i\right)} Y_{i,j} \tag{50}$$

Ước lượng tham số co giãn a

Với mọi $n \geq 1$ và $1 \leq j \leq p$, gọi

$$\widehat{a}_{n,j} = \frac{1}{nf_1} \sum_{i=1}^{n} \frac{\cos\left(2\pi\left(X_i - \widehat{\theta}_{i-1,j}\right)\right)}{g\left(X_i\right)} Y_{i,j} \tag{49}$$

$$\tilde{a}_{n,j} = \frac{1}{n\widehat{f}_{1,n}} \sum_{i=1}^{n} \frac{\cos\left(2\pi\left(X_i - \widehat{\theta}_{i-1,j}\right)\right)}{g\left(X_i\right)} Y_{i,j} \tag{50}$$

trong đó

$$\widehat{f}_{1,n} = \frac{1}{n} \sum_{i=1}^{n} \frac{\cos\left(2\pi X_i\right)}{g\left(X_i\right)} Y_{i,1}$$

Với I_p là ma trận đơn vị cấp p, e_1 là vector Euclid thứ nhất của $\mathbb{R}^p$ và M_p là ma trận vuông sao cho

$$M_p = I_p - ae_1^T \tag{51}$$

Ước lượng tham số co giãn a

Định lý 3.3.1. Định lý về sự hội tụ của ước lượng tham số co giãn $\widehat{a}_n$

Giả sử rằng các giả thiết từ $(\mathcal{H}_1)$ đến $(\mathcal{H}_4)$ được thỏa mãn. Khi đó:

1. $\widehat{a}_n$ và $\tilde{a}_n$ hội tụ hầu chắc chắn về a

$$\lim_{n\to\infty} \widehat{a}_n = a \text{ h.c.c và } \lim_{n\to\infty} \tilde{a}_n = a \text{ h.c.c} . \tag{52}$$

Ước lượng tham số co giãn a

Định lý 3.3.1. Định lý về sự hội tụ của ước lượng tham số co giãn $\widehat{a}_n$

Giả sử rằng các giả thiết từ $(\mathcal{H}_1)$ đến $(\mathcal{H}_4)$ được thỏa mãn. Khi đó:

1. $\widehat{a}_n$ và $\tilde{a}_n$ hội tụ hầu chắc chắn về a

$$\lim_{n\to\infty} \widehat{a}_n = a \text{ h.c.c và } \lim_{n\to\infty} \tilde{a}_n = a \text{ h.c.c .} \tag{52}$$

2. Tính tiệm cận chuẩn

$$\sqrt{n}\left(\widehat{a}_n - a\right) \xrightarrow{\mathcal{L}} \mathcal{N}_p(0, \Gamma(a)) \quad \text{và} \quad \sqrt{n}\left(\tilde{a}_n - a\right) \xrightarrow{\mathcal{L}} \mathcal{N}_p\left(0, M_p\Gamma(a)M_p^T\right) \tag{53}$$

trong đó $\Gamma(a)$ là ma trận hiệp phương sai

$$\Gamma(a) = \frac{1}{f_1^2}\operatorname{Cov}(C(X,\theta)Y) \tag{54}$$

Ước lượng tham số co giãn a

Ta cũng có luật mạnh dạng toàn phương

$$\lim_{n\to\infty} \frac{1}{\log(n)} \sum_{i=1}^{n} (\widehat{a}_i - a)(\widehat{a}_i - a)^T = \Gamma(a) \text{ h.c.c }, \tag{55}$$

Ta cũng có luật mạnh dạng toàn phương

$$\lim_{n\to\infty} \frac{1}{\log(n)} \sum_{i=1}^{n} (\widehat{a}_i - a)(\widehat{a}_i - a)^T = \Gamma(a) \text{ h.c.c },\tag{55}$$

$$\lim_{n\to\infty} \frac{1}{\log(n)} \sum_{i=1}^{n} (\tilde{a}_i - a)(\tilde{a} - a)^T = M_p \Gamma(a) M_p^T \text{ h.c.c }.\tag{56}$$

Ước lượng hàm hồi quy f

Ta thêm giả thiết $(\mathcal{H}_5)$: Hàm hồi quy f có tính Lipschitz và $f_1 = \int_{-1/2}^{1/2} \cos(2\pi x) f(x) dx$ đã biết trước.

Ta sử dụng ước lượng Nadaraya - Watson có trọng số $\omega_j(x)$ như sau

$$\widehat{f}_n(x) = \sum_{j=1}^{p} \omega_j(x) \widehat{f}_{n,j}(x) \tag{57}$$

Ước lượng hàm hồi quy f

Ta thêm giả thiết $(\mathcal{H}_5)$: Hàm hồi quy f có tính Lipschitz và
$f_1 = \int_{-1/2}^{1/2} \cos(2\pi x) f(x) dx$ đã biết trước.
Ta sử dụng ước lượng Nadaraya - Watson có trọng số $\omega_j(x)$ như sau

$$\widehat{f}_n(x) = \sum_{j=1}^{p} \omega_j(x) \widehat{f}_{n,j}(x) \tag{57}$$

trong đó, với mọi $1 \leq j \leq p$

$$\omega_j(x) = \omega_j(-x); \omega_j(x) \geq 0 \text{ và } \sum_{j=1}^{p} \omega_j(x) = 1 \tag{58}$$

$$\widehat{f}_{n,j}(x) = \frac{1}{\widehat{a}_{n,j}} \frac{\sum_{k=1}^{n} \left(W_{k,j}(x) + W_{k,j}(-x)\right)\left(Y_{k,j} - \widehat{v}_{k-1,j}\right)}{\sum_{k=1}^{n} \left(W_{k,j}(x) + W_{k,j}(-x)\right)} \tag{59}$$

Ước lượng hàm hồi quy f

Ta thêm giả thiết $(\mathcal{H}_5)$: Hàm hồi quy f có tính Lipschitz và $f_1 = \int_{-1/2}^{1/2} \cos(2\pi x) f(x) dx$ đã biết trước.

Ta sử dụng ước lượng Nadaraya - Watson có trọng số $\omega_j(x)$ như sau

$$\widehat{f_n}(x) = \sum_{j=1}^{p} \omega_j(x) \widehat{f}_{n,j}(x) \tag{57}$$

trong đó, với mọi $1 \leq j \leq p$

$$\omega_j(x) = \omega_j(-x); \omega_j(x) \geq 0 \text{ và } \sum_{j=1}^{p} \omega_j(x) = 1 \tag{58}$$

$$\widehat{f}_{n,j}(x) = \frac{1}{\widehat{a}_{n,j}} \frac{\sum_{k=1}^{n} \left(W_{k,j}(x) + W_{k,j}(-x)\right)\left(Y_{k,j} - \widehat{v}_{k-1,j}\right)}{\sum_{k=1}^{n} \left(W_{k,j}(x) + W_{k,j}(-x)\right)} \tag{59}$$

với

$$W_{n,j}(x) = \frac{1}{h_n} K\left(\frac{X_n - \widehat{\theta}_{n-1,j} - x}{h_n}\right) \tag{60}$$

Định lý 3.4.1

Giả sử rằng các giả thiết từ $(\mathcal{H}_1)$ đến $(\mathcal{H}_5)$ được thỏa mãn. Hơn nữa, giả định rằng các mô-men bậc lớn hơn 2 của $(\varepsilon_{i,j})$ hữu hạn. Khi đó, với mọi $x \in [-1/2; 1/2]$, ta có

$$\lim_{n\to\infty} \widehat{f}_n(x) = f(x) \text{ h.c.c .} \tag{61}$$

Ước lượng hàm hồi quy f

Định lý 3.4.2

Giả sử rằng các giả thiết từ $(\mathcal{H}_1)$ đến $(\mathcal{H}_5)$ được thỏa mãn. Ta cũng giả định rằng mô-men bậc lớn hơn 2 của $(\varepsilon_{i,j})$ hữu hạn. Lúc đó, nếu băng tần (h_n) thỏa mãn $h_n = 1/n^{\alpha}$ (với $\alpha > 1/3$) và $x \in [-1/2; 1/2]$ thì ta có tính tiệm cận chuẩn theo từng diểm như sau

- Với $x \neq 0$

$$\sqrt{nh_n}\left(\widehat{f}_n(x) - f(x)\right) \xrightarrow{\mathcal{L}} \mathcal{N}\left(0, \frac{\nu^2}{1+\alpha}\sum_{j=1}^{p}\frac{\sigma_j^2\omega_j^2(x)}{a_j^2\left(g\left(\theta_j + x\right) + g\left(\theta_j - x\right)\right)}\right) \tag{62}$$

Ước lượng hàm hồi quy f

Định lý 3.4.2

Giả sử rằng các giả thiết từ $(\mathcal{H}_1)$ đến $(\mathcal{H}_5)$ được thỏa mãn. Ta cũng giả định rằng mô-men bậc lớn hơn 2 của $(\varepsilon_{i,j})$ hữu hạn. Lúc đó, nếu băng tần (h_n) thỏa mãn $h_n = 1/n^{\alpha}$ (với $\alpha > 1/3$) và $x \in [-1/2; 1/2]$ thì ta có tính tiệm cận chuẩn theo từng diểm như sau

- Với $x \neq 0$

$$\sqrt{nh_n}\left(\widehat{f}_n(x) - f(x)\right) \xrightarrow{\mathcal{L}} \mathcal{N}\left(0, \frac{\nu^2}{1+\alpha}\sum_{j=1}^{p}\frac{\sigma_j^2\omega_j^2(x)}{a_j^2\left(g\left(\theta_j + x\right) + g\left(\theta_j - x\right)\right)}\right) \tag{62}$$

- Với $x = 0$

$$\sqrt{nh_n}\left(\widehat{f}_n(0) - f(0)\right) \xrightarrow{\mathcal{L}} \mathcal{N}\left(0, \frac{\nu^2}{1+\alpha}\sum_{j=1}^{p}\frac{\sigma_j^2\omega_j^2(0)}{a_j^2 g\left(\theta_j\right)}\right) \tag{63}$$

Mục lục

Chọn các tham số và hàm hồi quy

Mô hình có dạng

$$Y_{i,j} = a_j f\left(X_i - \theta_j\right) + v_j + \varepsilon_{i,j},$$

Mô hình có dạng

$$Y_{i,j} = a_j f\left(X_i - \theta_j\right) + v_j + \varepsilon_{i,j},$$

trong đó $1 \leq j \leq p$ và $1 \leq i \leq n$ với $p = 5$ và $n = 2000$. Ta chọn các tham số

$$v = (0, 1/3, -1, 2, -9/10)^T,$$
$$\theta = (0, 1/5, -1/20, -1/7, 1/6)^T,$$
$$a = (1, -4, 3, -5/2, 2)^T.$$

Chọn các tham số và hàm hồi quy

Mô hình có dạng

$$Y_{i,j} = a_j f\left(X_i - \theta_j\right) + v_j + \varepsilon_{i,j},$$

trong đó $1 \leq j \leq p$ và $1 \leq i \leq n$ với $p = 5$ và $n = 2000$. Ta chọn các tham số

$$v = (0, 1/3, -1, 2, -9/10)^T,$$
$$\theta = (0, 1/5, -1/20, -1/7, 1/6)^T,$$
$$a = (1, -4, 3, -5/2, 2)^T.$$

Các sai số (ε_{ij}) là dãy các biến ngẫu nhiên độc lập, cùng phân phối chuẩn $\mathcal{N}(0, 1)$.

Chọn các tham số và hàm hồi quy

Mô hình có dạng

$$Y_{i,j} = a_j f\left(X_i - \theta_j\right) + v_j + \varepsilon_{i,j},$$

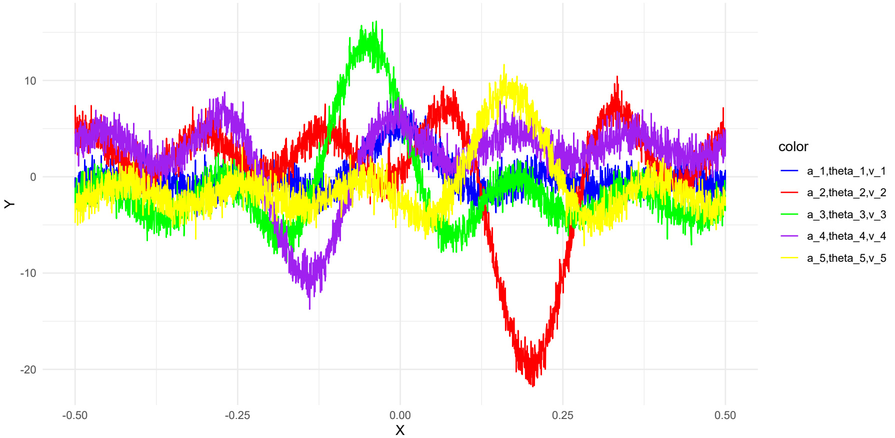
trong đó $1 \leq j \leq p$ và $1 \leq i \leq n$ với $p = 5$ và $n = 2000$. Ta chọn các tham số

$$v = (0, 1/3, -1, 2, -9/10)^T,$$
$$\theta = (0, 1/5, -1/20, -1/7, 1/6)^T,$$
$$a = (1, -4, 3, -5/2, 2)^T.$$

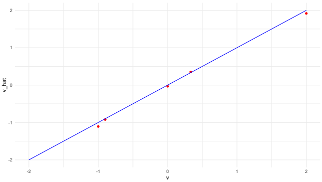
Các sai số (ε_{ij}) là dãy các biến ngẫu nhiên độc lập, cùng phân phối chuẩn $\mathcal{N}(0,1)$.
Các biến ngẫu nhiên (X_i) có phân phối đều trên $[-1/2; 1/2]$ và hàm hồi quy f xác định với mọi $x \in [-1/2; 1/2]$, được cho bởi

$$f(x) = \sum_{k=1}^{p} \cos(2k\pi x), \tag{64}$$

với $f_1 = 1/2$.

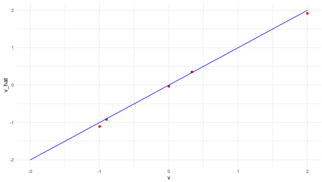


Hình 2: Dữ liệu mô phỏng

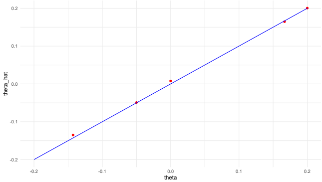


(a) Tham số chiều cao v

Ước lượng của các tham số

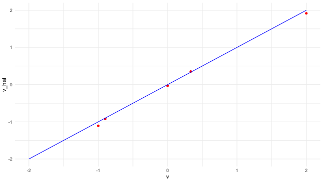


(a) Tham số chiều cao v

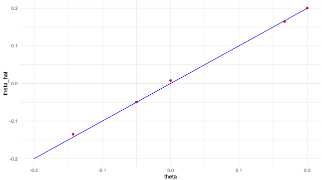


(b) Tham số chuyển θ

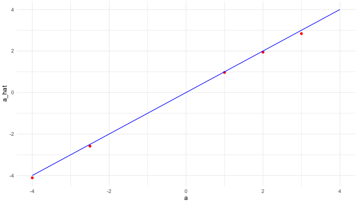
Ước lượng của các tham số



(a) Tham số chiều cao v

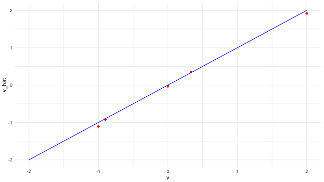


(b) Tham số chuyển θ

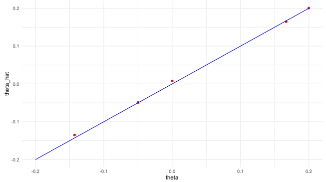


(c) Tham số co giãn a

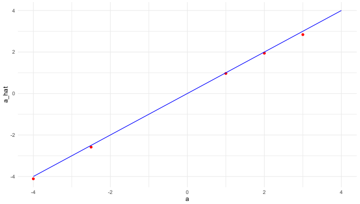
Ước lượng của các tham số



(a) Tham số chiều cao v



(b) Tham số chuyển θ

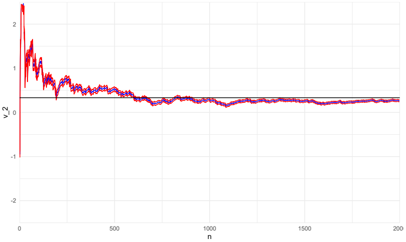


(c) Tham số co giãn a

Hình 3: Ước lượng các tham số v, θ và a (thứ tự từ trái sang, từ trên xuống)

Với $n = 2000$ và $\alpha = 5\%$, khoảng tin cậy của v_2, θ_1 lần lượt là

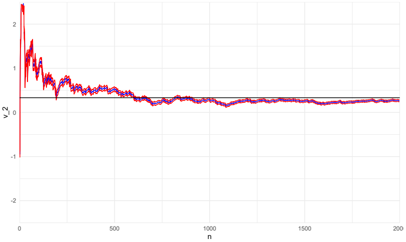
$$I_n(v_2) = [0.243044; 0.2920085], I_n(\theta_1) = [0.008646117; 0.0228825],$$



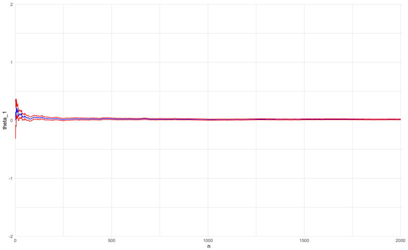
(a) Khoảng tin cậy của v_2

Với $n = 2000$ và $\alpha = 5\%$, khoảng tin cậy của v_2, θ_1 lần lượt là

$$I_n(v_2) = [0.243044; 0.2920085], I_n(\theta_1) = [0.008646117; 0.0228825],$$



(a) Khoảng tin cậy của v_2



(b) Khoảng tin cậy của θ_1

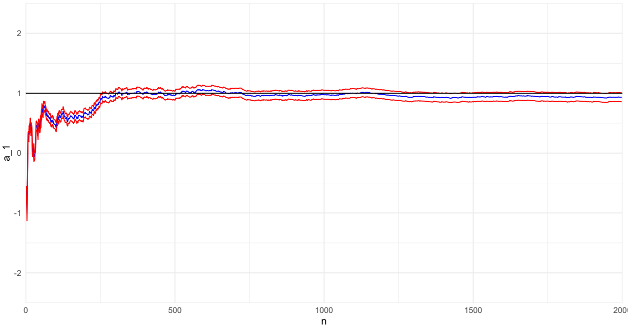
Hình 4: Khoảng tin cậy của v_2, θ_1 (thứ tự từ trái sang phải)

Khoảng tin cậy của các tham số

Với $n = 2000$ và $\alpha = 5\%$, khoảng tin cậy của a_1 là

$$I_n(a_1) = [0.8567415; 1.005915]$$

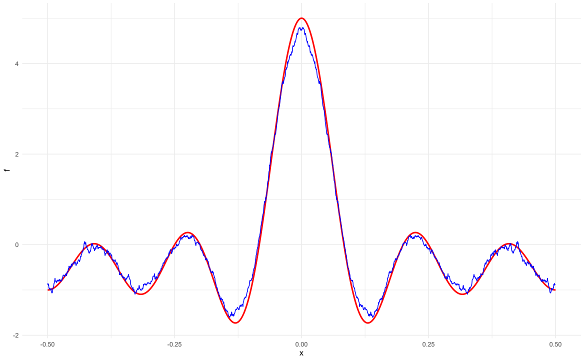
.



Hình 5: Khoảng tin cậy của a_1

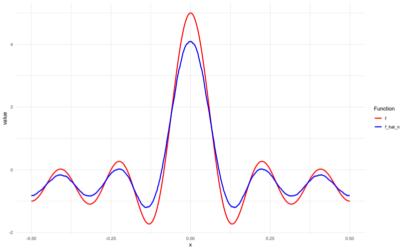
Chọn α cho băng tần (h_n) và ước lượng hàm hồi quy

Ta chọn K là hàm hạt nhân đều trên $[-1;1]$. Hơn nữa, với mọi $1 \leq j \leq p$, ta chọn $\omega_j(x) = 1/p$.



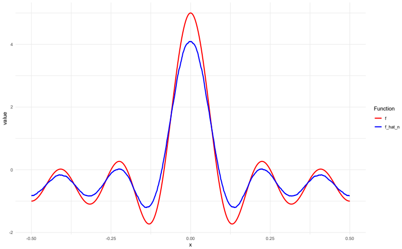
Hình 6: Hàm f được ước lượng bởi $\widehat{f_n}$ với $\alpha_1 = 9/10$.

Chọn α cho băng tần (h_n) và ước lượng hàm hồi quy

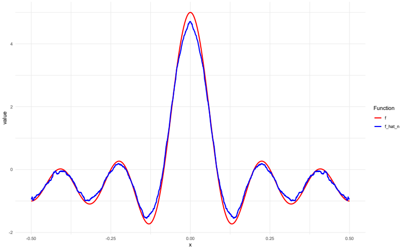


(a) Hàm f được ước lượng bởi $\widehat{f_n}$ với $\alpha_2 = 6/10$.

Chọn α cho băng tần (h_n) và ước lượng hàm hồi quy



(a) Hàm f được ước lượng bởi $\widehat{f_n}$ với $\alpha_2 = 6/10$.



(b) Hàm f được ước lượng bởi $\widehat{f_n}$ với $\alpha_3 = 8/10$.

Hình 7: Ước lượng cho f bằng $\widehat{f_n}$, trường hợp α_2, α_3 (thứ tự từ trái sang phải)

Tài liệu tham khảo I

[1] Fraysse, P. (2014). *Recursive Estimation in a Class of Models of Deformation, Journal of Statistical Planning and Inference*, 132-158.

[2] Bercu, B. and Fraysse, P. (2012). *A Robbins-Monro Procedure for Estimation in Semiparametric Regression Models*.

[3] Duflo, M. (1997). *Random Iterative Models. Applications of Mathematics(New York)*, Springer, Berlin.

[4] Bercu, B. (2014). *Asymptotic Results for Martingales with Statistical applications*.

[5] Kushner, H.J and Yin, G.G (2003). *Stochastic Approximation and Recursive Algorithms and Applications*. Applications of Mathematics(New York), Springer.

[6] Bruce M. Brown (1971). *Martingale Central Limit Theorems*.

[7] Chaabane, F. and Maaouia, F. (2000). *Théorèmes Limites Avec Poids Pour Les Martingales Vectorielles*, ESAIM PS, 4, 137–189.

Tài liệu tham khảo II

[8] Tiến, N.D. (2001). *Các mô hình xác suất và ứng dụng. Phần III, giải tích ngẫu nhiên*, NXB Đại học Quốc gia Hà Nội, 44-47.

[9] Pelletier, M. (1998). *On The Almost Sure Asymtotic Behavior of Stochastic algorithms*.

[10] Pelletier, M. (1998).*Weak convergence rates for stochastic approximation with application to multiple targets and simulated annealing*. Annals of Appli. Proba. 8, 1, 10-44.

[11] Trần Minh Phương, Nguyễn Thành Nhân (2022), *Giải tích số và ứng dụng (phần cơ bản)*, NXB Đại học Sư phạm TPHCM.

[12] Noda, K. (2007). Estimation of a Regression Function by The Parzen kernel-type Density Estimatiors, Ann Inst Statist Math, 221-234.

[13] Schuster, E.F. (1972). Joint Asymptotic Distribution of The Esimated Regression Function at a Finite Number of distinct Points, 84-88.

Vector martingale[10]

Định nghĩa 2.1.6. Khái niệm vector martingale bình phương khả tích

Trên không gian xác suất $(\Omega, \mathcal{A}, \mathbb{P})$ và $\mathbb{F} = (\mathcal{F}_n)$ là một bộ lọc. Giả sử $M = (M_n)$ là một dãy các vector ngẫu nhiên có giá trị trong $\mathbb{R}^d$ và tương thích với bộ lọc $\mathbb{F}$.

- M là một martingale bình phương khả tích nếu đối với mọi n

$$\mathbb{E}\left[\|M_n\|^2\right] < \infty \text{ và } \mathbb{E}\left[M_{n+1} - M_n \mid \mathcal{F}_n\right] = 0 \tag{15}$$

- Biến phân bình phương dự báo được của M là một dãy ngẫu nhiên $\langle M \rangle = (\langle M \rangle_n)$ các ma trận đối xứng, nửa xác định dương cỡ $d \times d$ được định nghĩa bằng cách đặt $\langle M \rangle_0 = 0$ và

$$\begin{aligned}\langle M \rangle_n - \langle M \rangle_{n-1} &= \mathbb{E}\left[(M_n - M_{n-1})(M_n - M_{n-1})^\top \mid \mathcal{F}_{n-1}\right] \\ &= \mathbb{E}\left[M_n M_n^\top \mid \mathcal{F}_{n-1}\right] - M_{n-1} M_{n-1}^\top. \end{aligned} \tag{16}$$

[10] Duflo, M. (1997). *Random Iterative Models. Applications of Mathematics(New York)*, Springer, Berlin.

Định lý giới hạn trung tâm cho vector martingale thực[11]

Định lý 2.1.7 Định lý giới hạn trung tâm cho vector martingale thực

Cho M là một vector martingale bình phương khả tích thực, đáp ứng bộ lọc $\mathbb{F} = (\mathcal{F}_n)$, có biến phân bình phương dự báo được ký hiệu bởi $\langle M \rangle$. Giả sử rằng, với một dãy xác định, thực (a_n) tăng đến $+\infty$, có thêm hai giả định sau:

- $a_n^{-1} \langle M \rangle_n \xrightarrow{\mathrm{P}} \Gamma$.
- Điều kiện Lindeberg được thỏa mãn; nói cách khác, đối với mọi $\varepsilon > 0$,

$$a_n^{-1} \sum_{k=1}^{n} \mathbb{E}\left[\|M_k - M_{k-1}\|^2 1_{(\|M_k - M_{k-1}\| \geq \varepsilon a_n^{1/2})} \mid \mathcal{F}_{k-1} \right] \xrightarrow{\mathcal{P}} 0. \tag{17}$$

Khi đó

- $a_n^{-1} M_n \xrightarrow{\text{a.s.}} 0$ và $a_n^{-1/2} M_n \xrightarrow{\mathcal{L}} \mathcal{N}(0, \Gamma)$.
- Nếu Γ là khả nghịch thì: $a_n^{1/2} \langle M \rangle_n^{-1} M_n \xrightarrow{\mathcal{L}} \mathcal{N}\left(0, \Gamma^{-1}\right)$.

[11] Duflo, M. (1997). *Random Iterative Models. Applications of Mathematics(New York)*, Springer, Berlin.