

1 Wstęp

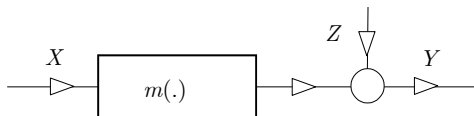
Wszystkie rozważania, także oznaczenia, pozostają w ścisłym związku z 2. ESTYMACJA GĘSTOŚCI.

2 Funkcja regresji

Omówimy teraz problem estymacji nieliniowej charakterystyki w systemie jak na rys. 2.1. Sygnał wejściowy X i zakłócenie pomiarowe Z na wyjściu mają charakter losowy i są wzajemnie niezależne. Ponadto $EZ = 0$. Sygnał wyjściowy, to $Y = m(U) + Z$. Ponieważ

$$E\{Y|X = x\} = m(x),$$

zatem nieznana charakterystyka jest funkcją regresji.



Rys. 2.1: System statyczny o charakterystyce $m(x)$.

Naszym celem jest wykrycie nieliniowej charakterystyki, czyli estymacja funkcji regresji $m(x)$ na podstawie niezależnych par pomiarów

$$(X_1, Y_1), (X_2, Y_2), (X_3, Y_3), \dots, (X_n, Y_n), \quad (2.1)$$

czyli obserwacji pary (X, Y) . Przedstawimy kilka algorytmów realizujących ten cel.

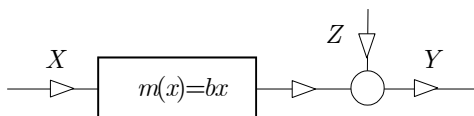
Łączną gęstość pary (X, Y) oznaczamy jako $f(x, y)$, gęstością zmiennej X jest $f(x)$, natomiast $f(y|x)$ jest odpowiednią gęstością warunkową.

3 Problem parametryczny

Jeśli znana jest postać funkcyjna charakterystyki $m(\cdot)$ i należy ustalić wartości jej parametrów, to problem jest parametryczny. Jest tak, gdy np. charakterystyka jest wielomianem znanego stopnia i zadanie polega na estymacji jego współczynników. Dla dalszych rozważań przyjmujemy, że $m(x) = bx$, tzn. że

$$Y = bX + Z,$$

gdzie b jest nieznanym współczynnikiem wzmocnienia (parametrem), rys. 3.1.



Rys. 3.1: System liniowy.

3.1 Metoda momentów

Metoda momentów bazuje na tym, że b można wyrazić przy pomocy momentów, ponieważ

$$b = \frac{EY}{EX}.$$

Naturalnymi ich estymatorami są odpowiednio

$$\frac{1}{n} \sum_{i=1}^n Y_i \text{ oraz } \frac{1}{n} \sum_{i=1}^n X_i,$$

co prowadzi do estymatora

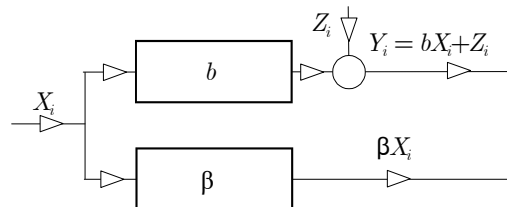
$$\tilde{b}_n = \frac{\sum_{i=1}^n Y_i}{\sum_{i=1}^n X_i}.$$

Jego zbieżność do b z prawdopodobieństwem 1 jest oczywista.

3.2 Metoda najmniejszych kwadratów

Jako model systemu rzeczywistego przyjmujemy wzmacniacz o nieustalonym na razie wzmocnieniu β . Jeśli zatem na jego wejście zostały podany sygnał X_i , to na wyjściu pojawiłaby się wartość βX_i , różna rzecz jasna od Y_i , rys. 3.2, a to z dwóch powodów. Po pierwsze trudno liczyć na to aby β było równe b , po drugie z uwagi na zakłócenie Z_i . Niedopasowanie modelu do systemu, ocenimy jako $(Y_i - \beta X_i)^2$. Mając na względzie cały ciąg obserwacji, weźmiemy pod uwagę wskaźnik globalny będący sumą lokalnych $(Y_i - \beta X_i)^2$ odnoszących się do poszczególnych par (X_i, Y_i) :

$$Q(\beta) = \sum_{i=1}^n (Y_i - \beta X_i)^2.$$



Rys. 3.2: Ilustracja metody najmniejszych kwadratów.

Jako estymator nieznanego b przyjmujemy β , które minimalizuje ten wskaźnik. Przyrównując zatem pochodną $dQ(\beta)/d\beta$ do zera, otrzymujemy równanie

$$\sum_{i=1}^n (Y_i X_i + \beta X_i) X_i = 0$$

i jego rozwiązane

$$\hat{b}_n = \frac{\sum_{i=1}^n Y_i X_i}{\sum_{i=1}^n X_i^2}$$

przyjmujemy jako estymator nieznanego b .

Z prawa wielkich liczb wynika zbieżność

$$\frac{1}{n} \sum_{i=1}^n Y_i X_i \xrightarrow{n \rightarrow \infty} EYX \text{ oraz } \frac{1}{n} \sum_{i=1}^n X_i^2 \xrightarrow{n \rightarrow \infty} EX^2$$

z prawdopodobieństwem 1. W rezultacie

$$\hat{b}_n \xrightarrow{n \rightarrow \infty} b$$

z prawdopodobieństwem 1.

Przedstawiony sposób postępowania nazywa się metodą najmniejszych kwadratów kojarzoną powszechnie z nazwiskiem Gauss.

4 Algorytm histogramowy

Histogram gęstości prowadzi w naturalny sposób do estymatora regresji, który teraz omówimy. Można uważać, że jest on pośrednikiem pomiędzy estymatorami parametrycznymi i nieparametrycznymi.

W rozważaniach będziemy operować wskaźnikiem zbioru A :

$$I_{\{A\}}(x) = \begin{cases} 1, & \text{dla } x \in A \\ 0, & \text{dla } x \notin A. \end{cases}$$

Założmy, że gęstość $f(x)$ określona jest na odcinku $[a, b]$, na którym rozmieszczamy punkty $x_0, x_1, x_2, \dots, x_n$ w ten sposób, że

$$a = x_0 < x_1 < x_2 < \dots < x_{n-1} < x_n = b,$$

i $x_{i+1} - x_i = \Delta$. Został on w ten sposób podzielony na n pododcinków $A_i = [x_i, x_{i+1}]$, każdy o długości $\Delta = (b - a)/n$.

Estymator regresji definiujemy jak poniżej:

$$m_n(x) = \frac{\frac{1}{n} \sum_{j=1}^n Y_j I_{\{A_i\}}(X_j)}{\frac{1}{n} \sum_{j=1}^n I_{\{A_i\}}(X_j)}, \quad \text{dla } x \in A_i. \quad (4.1)$$

Jego licznik jest średnią obserwacji Y_j , których X_j wpadły do odcinka A_i , mianownik natomiast jest częstością wpadania X_j do tego odcinka.

Oznaczając licznik estymatora jako $\xi_n(x)$, otrzymujemy

$$\begin{aligned} E\xi_n(x) &= E\{Y I_{\{A_i\}}(X)\} \\ &= \int_{-\infty}^{\infty} \int_{x_i}^{x_{i+1}} y f(x, y) dy dx \\ &= \int_{x_i}^{x_{i+1}} \left[\int_{-\infty}^{\infty} y f(y|x) dy \right] f(x) dx \\ &= \int_{x_i}^{x_{i+1}} m(x) f(x) dx = \int_{A_i} m(x) f(x) dx, \end{aligned}$$

skąd, odwołując się do prawa wielkich liczb, wnioskujemy, że

$$E\xi_n(x) \xrightarrow{n \rightarrow \infty} \int_{A_i} m(v) f(v) dv$$

według prawdopodobieństwa. Oznaczając następnie mianownik przez $\eta_n(x)$ i postępując podobnie wnioskujemy, że

$$E\eta_n(x) \xrightarrow{n \rightarrow \infty} \int_{A_i} f(v) dv.$$

W rezultacie, możemy napisać, że

$$m_n(x) \xrightarrow{n \rightarrow \infty} \frac{\int_{A_i} m(v) f(v) dv}{\int_{A_i} f(v) dv}$$

według prawdopodobieństwa. Funkcja graniczna ma kształt schodkowy, na całym odcinku A_i przyjmuje taką samą wartość.

Ponieważ długość odcinka A_i jest stała, estymator jest obciążony nieusuwalnym błędem i nie zbiega się do $m(x)$. Jeśli chcemy zatem osiągnąć tę zbieżność musimy uzależnić długość tego odcinka od liczby pomiarów n i spowodować, aby malała ona do zera. Tak właśnie postępuje estymator jądrowy omówiony w następnym paragrafie.

5 Algorytm jądrowy

Teraz przyjmujemy, że nasza informacja wstępna o systemie jest uboższa, że nie znamy postaci funkcyjnej charakterystyki m . Zadanie jej wyznaczenia ma zatem charakter nieparametryczny. Algorytmy przedstawione w kolejnych paragrafach w sposób oczywisty nawiązują do poznanych wcześniej estymatorów gęstości.

Jako pierwszy omówimy algorytm jądrowy jak poniżej:

$$\hat{m}_n(x) = \frac{\frac{1}{nh(n)} \sum_{i=1}^n Y_i K\left(\frac{x - X_i}{h(n)}\right)}{\frac{1}{nh(n)} \sum_{i=1}^n K\left(\frac{x - X_i}{h(n)}\right)}. \quad (5.1)$$

Jądrem K może być dowolna ograniczona funkcja spełniająca warunek:

$$\int K(v) dv = 1, \quad \lim_{|v| \rightarrow \infty} |v| K(v) = 0. \quad (5.2)$$

Przykładami takich funkcji mogą być jądra omówione wcześniej przy estymacji gęstości.

Ciąg nieujemnych liczb $\{h(n)\}$ spełnia natomiast warunki

$$\lim_{n \rightarrow \infty} h(n) = 0, \quad \lim_{n \rightarrow \infty} nh(n) = \infty. \quad (5.3)$$

Intuicyjne uzasadnienie algorytmu (5.1) jest dość oczywiste. Oznaczając licznik w (5.1) przez $\xi_n(x)$, zauważamy bowiem, że

$$\begin{aligned} E\xi_n(x) &= \frac{1}{h(n)} E\left\{m(X) K\left(\frac{x - X}{h(n)}\right)\right\} \\ &= \frac{1}{h(n)} \int m(v) K\left(\frac{x - v}{h(n)}\right) f(v) dv. \end{aligned}$$

Odwołując się do Uwagi 5.1 stwierdzamy, że powyższa wartość dąży do $m(x) f(x) \int K(v) dv$, gdy $h(n) \rightarrow 0$. Ponadto

$$\begin{aligned} nh(n) \text{ var}[\xi_n(x)] &= \frac{1}{h(n)} \text{ var}\left[Y K\left(\frac{x - X}{h(n)}\right)\right] \\ &\leq \frac{1}{h(n)} E\left\{Y^2 K^2\left(\frac{x - X}{h(n)}\right)\right\} \\ &= \frac{1}{h(n)} \int [m^2(v) + \sigma_Z^2] K^2\left(\frac{x - v}{h(n)}\right) f(v) dv, \end{aligned}$$

która to wielkość, na mocy tej samej Uwagi i (5.3), zbiega się do $[m^2(v) + \sigma_Z^2] \int K^2(v) dv$. Wykazaliśmy więc, że

$$\xi_n(x) \xrightarrow{n} m(x) f(x)$$

według średniej drugiego rzędu.

Ponieważ, z tych samych powodów $\eta_n(x) \xrightarrow{n} f(x)$ w tym samym sensie, dochodzimy do ostatecznego wniosku, że

$$\hat{m}_n(x) \xrightarrow{n} m(x),$$

według prawdopodobieństwa. Estymator jest zatem zgodny, skutecznie wykrywa nieznaną charakterystykę.

W szczególności, jako jądro można przyjąć funkcję prostokątną

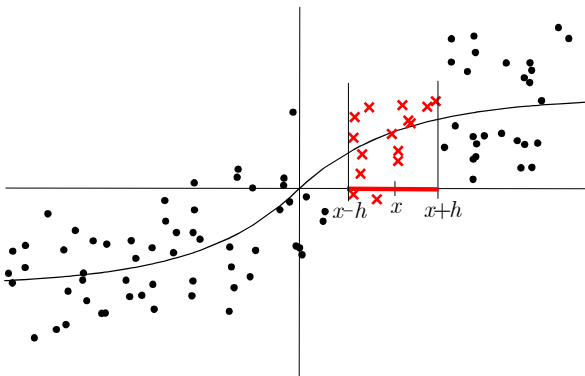
$$K(v) = \begin{cases} \frac{1}{2}, & \text{dla } |v| \leq 1 \\ 0, & \text{w przeciwnym wypadku,} \end{cases}$$

dla której

$$\hat{m}_n(x) = \frac{\frac{1}{n} \sum_{j=1}^n Y_j I_{\{(x-h(n), x+h(n))\}}(X_j)}{\frac{1}{n} \sum_{j=1}^n I_{\{(x-h(n), x+h(n))\}}(X_j)},$$

gdzie licznik jest średnią tych Y_j , których X_j leżą w odcinku $(x - h(n), x + h(n))$, natomiast mianownik częstością wpadania X_i do tego odcinka.

Zasadę działania algorytmu ilustruje rys. 5.1, na którym $m(x) = (1 - e^{-|x|}) \text{sign}(x)$. Pokazano na nim także obserwacje (Y_i, X_i) . Wartość estymatora w punkcie x wyznacza się na podstawie zaznaczonych symbolem \times .



Rys. 5.1: Estymator jądrowy, jądro prostokątne, przedział $[x - h, x + h]$ ustalony z góry. Linia ciągła oznacza charakterystykę $m(x) = (1 - e^{-|x|}) \text{sign}(x)$.

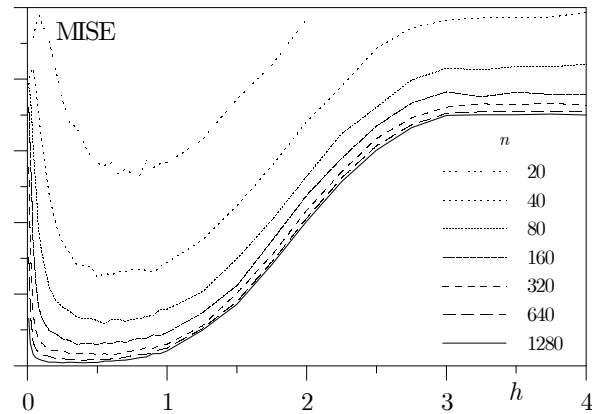
Jako globalny błąd algorytmu przyjęto

$$\text{MISE} = E \int (\hat{m}_n(x) - m(x))^2 dx.$$

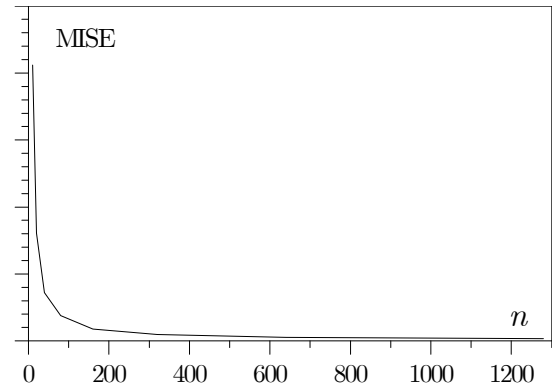
Wyniki komputerowego eksperymentu przedstawione na rys. 5.2 pokazują, że dla każdego n istnieje optymalne h . Wykres przedstawiający MISE dla optymalnie wybranego h przedstawia rys. 5.3. Błąd maleje szybko przy wzroście n , aczkolwiek wolniej niż $1/n$, co jest własnością obserwowaną także przy estymacji gęstości prawdopodobieństwa.

Uwaga 5.1 Dla funkcji jądrowych wybranych zgodnie z założeniem,

$$\frac{1}{h} \int K\left(\frac{x-v}{h}\right) g(v) dv \xrightarrow{h \rightarrow 0} g(x) \int K(\xi) d\xi.$$



Rys. 5.2: Estymator jądrowy. MISE dla różnych h .



Rys. 5.3: MISE dla optymalnego h .

Aby to uzasadnić, zauważmy, że

$$\frac{1}{h} K\left(\frac{v}{h}\right) \xrightarrow{h \rightarrow 0} \delta(v) \int K(\xi) d\xi,$$

gdzie $\delta(v)$ jest impulsem Diraca, skąd wynika, że, dla ustalonego x ,

$$\frac{1}{h} K\left(\frac{v-x}{h}\right) \xrightarrow{h \rightarrow 0} \delta(v-x) \int K(\xi) d\xi.$$

Zatem

$$\frac{1}{h} \int K\left(\frac{x-v}{h}\right) g(v) dv \xrightarrow{h \rightarrow 0} \int \delta(v-x) g(v) dv \int K(\xi) d\xi,$$

która to wielkość jest równa $g(x) \int K(\xi) d\xi$, co kończy uzasadnienie.

6 Najbliższy sąsiad

Algorytm, podobnie jak wszystkie typu NN, nie korzysta bezpośrednio z ciągu obserwacji (2.1), lecz z ciągu uporządkowanego

$$(X_{(1)}, Y_{[1]}), (X_{(2)}, Y_{[1]}), \dots, (X_{(n)}, Y_{[1]}). \quad (6.1)$$

Porządek jest taki, że

$$|X_{(1)} - x| < |X_{(2)} - x| < \dots < |X_{(n)} - x|, \quad (6.2)$$

co oznacza, że najbliższym punktu x jest $X_{(1)}$, następnie $X_{(2)}$ itd. Pozycja, na której znajdzie się para (Y_i, X_i) w ciągu (6.1) zależy od odległości X_i od x . Jeśli odległość ta jest

j -ta co do wielkości, to pozycją tą jest j . Estymator jest następującej postaci:

$$\bar{m}_n(u) = \frac{1}{k(n)} \sum_{i=1}^{k(n)} Y_{[i]},$$

gdzie podobnie jak przy estymacji gęstości,

$$\lim_{n \rightarrow \infty} k_n = \infty, \quad \lim_{n \rightarrow \infty} \frac{k_n}{n} = 0.$$

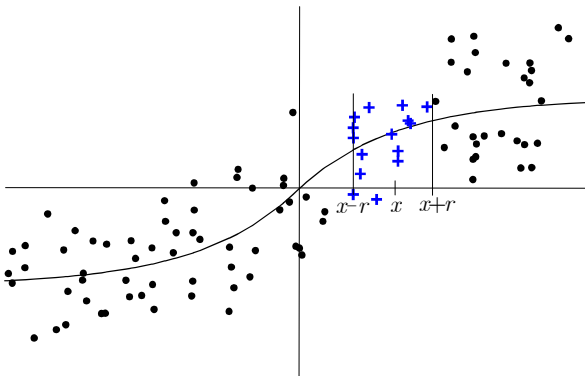
Jest on zgodny, tzn.

$$\bar{m}_n(u) \xrightarrow{n \rightarrow \infty} m(x)$$

według prawdopodobieństwa.

Intuicyjne uzasadnienie jest oczywiste. Jest on bowiem uśrednioną wartością $k(n)$ obserwacji spośród Y_1, Y_2, \dots, Y_n , dla których odpowiadające im X_1, X_2, \dots, X_n leżą najbliżej punktu x .

W sytuacji jak w § 5 (jądro prostokątne) można zastosować algorytm NN. Dla $k(n) = 15$ jego wartość w punkcie x jest wyliczana na podstawie 15 obserwacji Y_i zaznaczonych symbolem $+$. Odpowiadające im X_i znajdują się wśród piętnastu najbliższych punktowi x . Leżą one na odcinku $[x - r, x + r]$, gdzie $r = |x - X_{(15)}|$, jedna z tych obserwacji na jego brzegu, lewym. Warto estymatora jest średnią z tych Y_i , czyli średnią z $Y_{[1]}, Y_{[2]}, \dots, Y_{[15]}$.



Rys. 6.1: Estymator NN, $k = 15$. Przedział $[x - r, x + r]$, gdzie $r = |x - X_{(15)}|$, wyznaczony na podstawie obserwacji X_i .

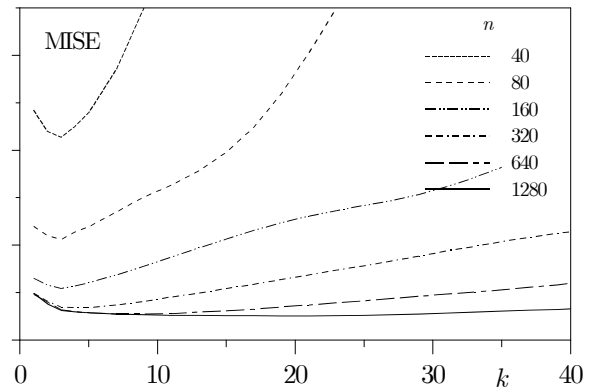
Zależność błędu

$$\text{MISE} = E \int (\bar{m}_n(u) - m(u))^2 du$$

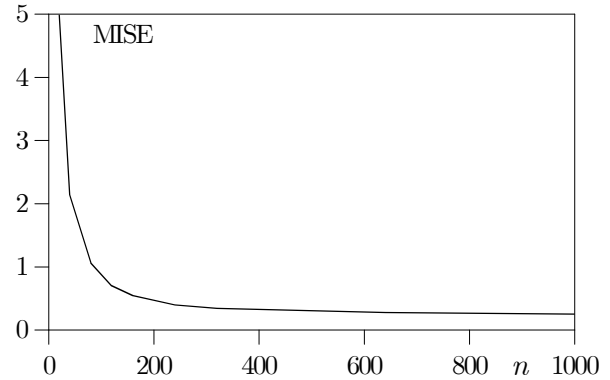
od k oraz n otrzymana na podstawie komputerowych badań eksperymentalnych pokazana jest na rys. 6.2. Konieczność precyzyjnego doboru k zdecydowanie maleje gdy zwiększa się liczba obserwacji. Błąd dla optymalnie dobranego k przedstawia rys. 6.3.

Uwaga 6.1 Podobnie jak estymatory gęstości, przedstawione algorytmy regresyjne tzn. histogramowy, jądrowy i NN charakteryzują się pewnym podobieństwem. Spozstrzegamy to zapisując je jak poniżej:

$$m_n(x) = \frac{\sum_{j=1}^n Y_j(Y_j \text{ takie, że } X_j \in A)}{\text{liczba } X_j \text{ takich, że } X_j \in A_i}, \quad \text{dla } x \in A_i,$$



Rys. 6.2: Algorytm NN. Zależność MISE od k .



Rys. 6.3: Estymator NN. MISE dla optymalnie dobranego k .

gdzie $A_i = [x_i, x_{i+1}]$,

$$\hat{m}_n(x) = \frac{\sum_{j=1}^n Y_j(\text{takie, że } X_j \in A_x)}{\text{liczba } X_j \text{ takich, że } X_j \in A_x},$$

(dla jądra prostokątnego), gdzie $A_x = [x - h(n), x + h(n)]$,

$$\bar{m}_n(x) = \frac{\sum_{j=1}^n Y_i(\text{takie, że } X_j \in A_{k(n)})}{\text{liczba } X_j \text{ takich, że } X_j \in A_{k(n)}},$$

gdzie $A_{k(n)} = [x - r(k(n)), x + r(k(n))]$.

7 Metoda rozwinięć ortogonalnych

Zbiór, na którym określona jest nieliniowa charakterystyka oznaczmy jako D . Może to być odcinek, półprosta lub cała prosta. Na zbiorze tym istnieje układ funkcji ortogonalnych

$$\varphi_0(x), \varphi_1(x), \varphi_2(x), \dots,$$

czyli takich, że

$$\int_D \varphi_i(x) \varphi_j(x) dx = 0, \quad \text{dla } i \neq j,$$

$$\int_D \varphi_i^2(x) dx = 1,$$

wspólnie ograniczonych, co oznacza, że istnieje c takie, że $\max_x |\varphi_k(x)| < c$ dla wszystkich k .

Zacniemy od tego, że

$$m(x) = \frac{g(x)}{f(x)}, \quad (7.1)$$

gdzie $g(x) = m(x)f(x)$. Przy założeniu, że $\int_D g^2(x)dx < \infty$, funkcję g można rozwinąć w szereg ortogonalny

$$g(x) \sim \sum_{k=0}^{\infty} a_k \varphi_k(x), \quad (7.2)$$

gdzie

$$\begin{aligned} a_k &= \int_D g(x) \varphi_k(x) dx = \int_D m(x) \varphi_k(x) f(x) dx \\ &= E\{m(X) \varphi_k(X)\} = E\{Y \varphi_k(X)\}, \end{aligned}$$

ponieważ $Y = m(U) + Z$. Ponadto

$$\int_D g^2(x) dx = \sum_{k=0}^{\infty} a_k^2. \quad (7.3)$$

Jako oszacowanie współczynnika a_k można przyjąć:

$$\tilde{a}_k = \frac{1}{n} \sum_{i=1}^n Y_i \varphi_k(X_i),$$

co prowadzi do następującego estymatora funkcji $g(x)$:

$$\tilde{g}_n(x) = \sum_{k=0}^{N(n)} \tilde{a}_k \varphi_k(x).$$

Ponieważ

$$\begin{aligned} \tilde{g}_n(x) - g(x) &\sim \sum_{k=0}^{N(n)} (\tilde{a}_k - a_k) \varphi_k(x) \\ &+ \sum_{k=N(n)+1}^{\infty} (-a_k), \end{aligned}$$

zatem, z uwagi na (7.2) i (7.3),

$$\begin{aligned} E \int_D (\tilde{g}_n(x) - g(x))^2 dx &= \sum_{k=0}^{N(n)} E(\tilde{a}_k - a_k)^2 \\ &+ \sum_{k=N(n)+1}^{\infty} a_k^2. \end{aligned}$$

Jest teraz oczywiste, że $E\tilde{a}_k = a_k$ oraz

$$\begin{aligned} \text{var}[\tilde{a}_k] &= \text{var} \left[\frac{1}{n} \sum_{i=1}^n Y_i \varphi_k(X_i) \right] = \frac{1}{n} \text{var}[Y \varphi_k(x)] \\ &= \frac{1}{n} E(Y \varphi_k(x))^2 \leq \frac{1}{n} c^2 EY^2, \end{aligned}$$

skąd wynika, że $E(\tilde{a}_k - a_k)^2 \leq \frac{1}{n} c^2 EY^2$. Wykazaliśmy zatem, że

$$\begin{aligned} E \int_D (\tilde{g}_n(x) - g(x))^2 dx &\leq \frac{N(n)}{n} c^2 EY^2 \\ &+ \sum_{k=N(n)+1}^{\infty} a_k^2. \end{aligned}$$

Jeśli zatem

$$N(n) \xrightarrow{n \rightarrow \infty} \infty, \quad \frac{N(n)}{n} \xrightarrow{n \rightarrow \infty} 0,$$

$$E \int_D (\tilde{g}_n(x) - g(x))^2 dx \xrightarrow{n \rightarrow \infty} 0$$

Postępując podobnie wobec mianownika w (7.1) i mając na uwadze, że $f(x) \sim \sum_{k=0}^{\infty} b_k \varphi_k(x)$, gdzie $b_k = \int_D f(x) \varphi_k(x) dx = E\varphi_k(X)$, jako estymator gęstości $f(x)$ przyjmujemy $\tilde{f}_n(x) = \sum_{k=0}^{N(n)} \tilde{b}_k \varphi_k(x)$ i argumentując jak powyżej dochodzimy do wniosku, że $E \int_D (\tilde{f}_n(x) - f(x))^2 dx \xrightarrow{n \rightarrow \infty} 0$. Wykazane własności pozwalają żywić nadzieję, że także $E(\tilde{g}_n(x) - g(x))^2 dx \xrightarrow{n \rightarrow \infty} 0$ i $E(\tilde{f}_n(x) - f(x))^2 dx \xrightarrow{n \rightarrow \infty} 0$ w ustalonym punkcie x . Jeśli tak, to przyjmując

$$\tilde{m}_n(x) = \frac{\sum_{k=0}^{N(n)} \tilde{a}_k \varphi_k(x)}{\sum_{k=0}^{N(n)} \tilde{b}_k \varphi_k(x)}$$

jako estymator nieznannej charakterystyki, możemy napisać:

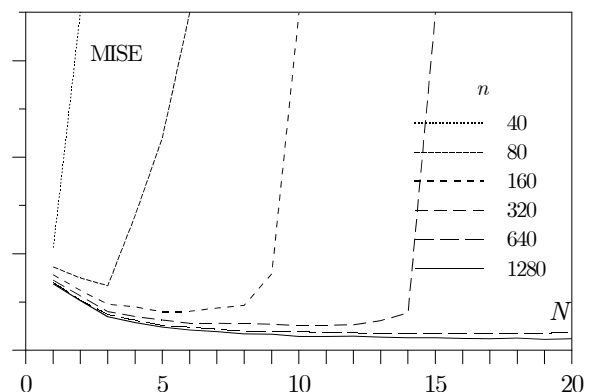
$$\tilde{m}_n(x) \xrightarrow{n \rightarrow \infty} m(x)$$

według prawdopodobieństwa.

W sytuacji jak w § 5 (jądro prostokątne) z ograniczeniem $x \in [-\pi, \pi]$, zastosowano układ trygonometryczny. Rezultaty symulacji komputerowej pokazano na rys. 7.1, na którym

$$\text{MISE} = E \int_{-\pi}^{\pi} (\tilde{m}_n(x) - m(x))^2 dx.$$

Optymalna wartość N narasta wraz z liczbą obserwacji. MISE odpowiadające optymalnemu N pokazano na rys. 7.2.



Rys. 7.1: Estymator ortogonalny. MISE w zależności od N .

8 Systemy dynamiczne

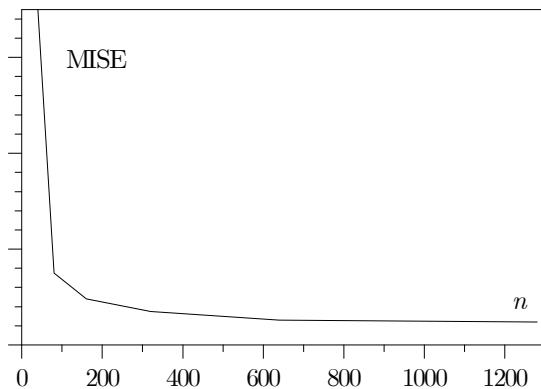
8.1 System Hammersteina

System Hammersteina składa się z dwóch podsystemów. Pierwszy jest nieliniowy i ma charakterystykę $m(u)$, Drugi jest liniowy, dynamiczny i opisywany jest równaniem

$$W_{n+1} = aW_n,$$

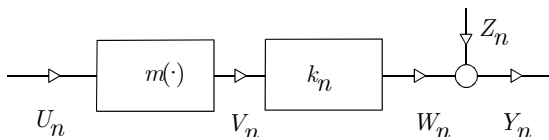
co oznacza, że jego odpowiedzią impulsową jest $k_0 = 0$, $k(n) = a^n$, $n = 1, 2, 3, \dots$. Zatem

$$Y_n = \sum_{i=0}^{\infty} k_i m(U_{n-i}) + Z_i = k_1 m(U_{n-1}) + \xi_n + Z_n, \quad (8.1)$$



Rys. 7.2: Estymator ortogonalny. MISE odpowiadające optymalnemu N .

gdzie $\xi_n = \sum_{i=1}^{\infty} k_i m(U_{n-i})$ nie zależy od U_{n-1} , natomiast Z_n jest zakłóceniem pomiarowym o średniej zero niezależnym od wejścia. Wszystkie sygnały mają charakter stacjonarny i ponadto $Em(U) = 0$.



Rys. 8.1: System Hammersteina.

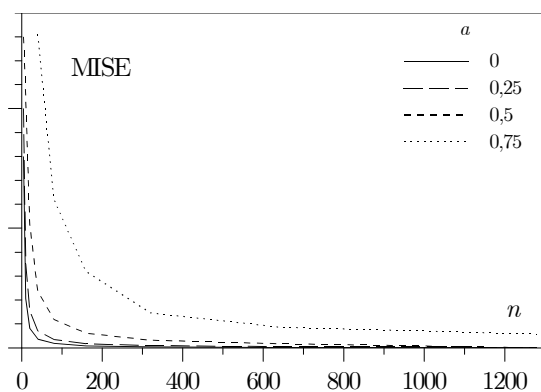
Z założeń i (8.1) wynika, że

$$E\{Y_n | U_{n-1} = u\} = \gamma m(u),$$

gdzie $\gamma = k_1$ jest nieznaną liczbą. Ponieważ charakterystyka m jest funkcją regresji, do jej estymacji zastosujemy omówiony wcześniej algorytm jądrowy, co prowadzi do estymatora:

$$\tilde{m}_n(u) = \frac{\sum_{i=1}^n Y_{i+1} K\left(\frac{u - U_i}{h(n)}\right)}{\sum_{i=1}^n K\left(\frac{u - U_i}{h(n)}\right)}.$$

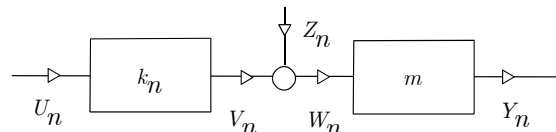
Wyniki symulacji komputerowej pokazane są na rys. 8.2. Widać wyraźnie, że wzrost a , czyli zwiększanie się roli podsystemu dynamicznego, prowadzi do narastania błędu.



Rys. 8.2: MISE dla systemu Hammersteina.

8.2 System Wienera

W systemie Wienera podsystemy połączone są w odwrotnej kolejności, nieliniowy znajduje się za dynamicznym. Problem wykrywania nieliniowości jest znacznie trudniejszy.



Rys. 8.3: System Wienera.

Jeśli jednak sygnał wejściowy U_n ma rozkład normalny, to sygnał W_n także. W sytuacji takiej

$$E\{U_n | Y_{n+1} = y\} = cm^{-1}(y),$$

gdzie c jest pewną nieznaną liczbą. Spostrzeżenie to można wykorzystać do konstrukcji estymatora odwrotności $cm^{-1}(y)$ nieznannej charakterystyki o postaci:

$$\frac{\sum_{i=1}^n U_i K\left(\frac{u - Y_{i+1}}{h(n)}\right)}{\sum_{i=1}^n K\left(\frac{u - Y_{i+1}}{h(n)}\right)}.$$

9 Zakończenie

W mocy pozostają wszystkie uwagi poczynione w Zakończeniu tekstu dotyczącego estymacji gęstości prawdopodobieństwa.

Tutaj można dodać, że metoda najmniejszych kwadratów zapewnia większą skuteczność od metody momentów, a to ze względu na zawarty w niej proceder optymalizacji.

Zwróćmy ponadto uwagę, że prezentowane algorytmy można stosować do wykrywania nieliniowości zarówno w systemach statycznych jak również dynamicznych.