

## 1 Estymacja gęstości prawdopodobieństwa

Omówimy teraz metody estymacji funkcji gęstości prawdopodobieństwa. Można je podzielić na dwie kategorie: parametryczne i nieparametryczne. W pierwszej znany jest funkcyjny kształt gęstości, nieznane są jedynie wartości jego parametrów. Jeśli, np. wiadomo, że rozkład jest normalny, to do ustalenia pozostaje wartość oczekiwana i wariancja. Jeśli natomiast gęstość jest całkowicie nieznaną, to mamy do czynienia z problemem nieparametrycznym. Estymować należy funkcję, a nie jedynie skończony zestaw liczb (parametrów).

Podstawą wnioskowania jest ciąg niezależnych zmiennych losowych

$$X, X_1, X_2, X_3, \dots, X_n. \quad (1.1)$$

Każda z nich ma nieznaną gęstość  $f$ .

## 2 Metody parametryczne

Metody parametryczne stosujemy jeśli znana jest postać funkcyjna gęstości, lecz nieznane są jedynie wartości jej parametrów. Istotę metod parametrycznych omówimy na przykładzie rozkładu normalnego, czyli rozkładu o gęstości

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-(x-\mu)^2/2\sigma^2}.$$

Ma on dwa parametry: wartość oczekiwaną  $\mu$  oraz wariancję  $\sigma^2$ .

Naturalnym estymatorem wartości oczekiwanej  $\mu = EX$  jest

$$\hat{\mu}_n = \frac{1}{n} \sum_{i=1}^n X_i.$$

Z praw wielkich liczb wynika jego zbieżność do  $\mu$ . Z uwagi na nią mówimy, że nasz estymator jest zgodny (estymator jest zbieżny do wartości estymowanej). Jako oszacowanie gęstości można więc przyjąć

$$\hat{f}_n(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-(x-\hat{\mu}_n)^2/2\sigma^2}.$$

Jeśli natomiast znana jest wartość oczekiwana, to wariancję  $\sigma^2 = E(X - \mu)^2$  można estymować jak poniżej:

$$\hat{\sigma}_n^2 = \frac{1}{n} \sum_{i=1}^n (X_i - \mu)^2.$$

Zgodność wynika z praw wielkich liczb. Jeśli natomiast nieznane są prawdziwe wartości obydwu parametrów, to jako estymator wariancji można przyjąć

$$\frac{1}{n} \sum_{i=1}^n (X_i - \hat{\mu}_n)^2.$$

## 3 Metody nieparametryczne

Metody nieparametryczne stosujemy, jeśli informacja wstępna (aprioryczna) o gęstości jest dalej posunięta, jeśli postać funkcyjna gęstości nie jest znana.

Idea estymacji funkcji gęstości jest przedstawiona na rys. 3.1. Wychodzi ona z tego, że

$$f(x) = \lim_{|A| \rightarrow 0} \frac{1}{|A|} \int_A f(x) dx = \lim_{|A| \rightarrow 0} \frac{P\{X \in A\}}{|A|}$$

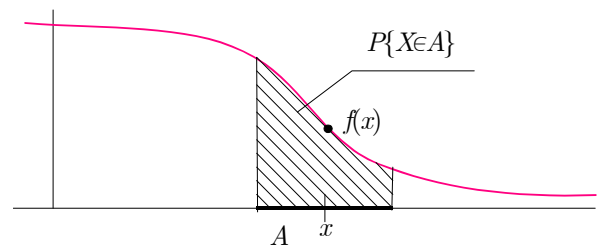
gdzie  $A$  jest odcinkiem zawierającym w swoim wnętrzu punkt  $x$ , a  $|A|$  jego długością. W procesie estymacji ułamek

$$\frac{P\{X \in A\}}{|A|}$$

szacuje się wstawiając częstość wpadania obserwacji do odcinka  $A$  w miejsce prawdopodobieństwa, co prowadzi do wyrażenia

$$\frac{\text{częstość wpadania obserwacji do odcinka } A}{\text{długość odcinka } A}. \quad (3.1)$$

Jest przy tym oczywiste, że długość odcinka  $A$  powinna maleć do zera, gdy  $n \rightarrow \infty$ .



Rys. 3.1: Idea estymacji  $f(x)$ . Pole zakreskowanej powierzchni, to  $P\{X \in A\}$ .

Przedstawimy trzy metody postępowania. W pierwszej długość odcinka jest stała, nie zmienia się wraz z  $n$ . Estymator przyjmuje na nim stałą wartość. Jego wykres nazywa się histogramem. W pozostałych długość odcinka zmienia się. W metodzie jądrowej jest ona z góry ustalona, natomiast częstość wpadania szacuje empirycznie się na podstawie obserwacji. W metodzie typu najbliższy sąsiad, z góry ustala się częstość, natomiast długość przedziału wylicza się na podstawie danych pomiarowych. Innymi słowy w metodzie jądrowej mianownik ułamka (3.1) ma charakter deterministyczny, a licznik jest losowy. W metodzie typu najbliższy sąsiad jest na odwrót, licznik jest deterministyczny, a mianownik losowy.

## 4 Histogram

Histogram jest metodą, która lokuje się pomiędzy metodami parametrycznymi i nieparametrycznymi. Jest ona niejako zapowiedzią nieparametrycznej metody jądrowej.

Załóżmy, że gęstość określona jest na odcinku  $[a, b]$ . Niech

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

gdzie  $x_{i+1} - x_i = \Delta$ . Odcinek ten został w ten sposób podzielony na  $n$  pododcinków  $A_i = [x_i, x_{i+1})$ , każdy o długości  $\Delta = (b - a)/n$ . Na każdym z nich estymator  $f_n(x)$  gęstości jest funkcją stałą równą wartości w (3.1), czyli

$$f_n(x) = \frac{\text{liczba } X_i \text{ w zbiorze } A_i}{\Delta n}, \text{ dla } x \in A_i.$$

Jest oczywiste, że

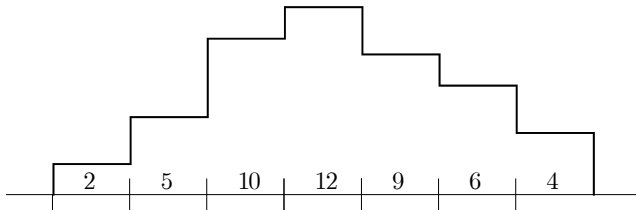
$$Ef_n(x) = \frac{1}{\Delta} P\{X \in A_i\} = \frac{1}{\Delta} \int_{A_i} f(\xi) d\xi.$$

Z prawa wielkich liczb wynika ponadto, że

$$f_n(x) \xrightarrow{n \rightarrow \infty} \frac{1}{\Delta} \int_{A_i} f(\xi) d\xi$$

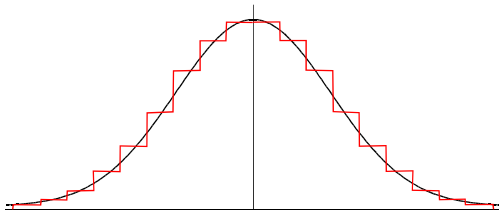
według średniej drugiego rzędu. Estymator jest niezwykle prosty, ale nie można oczekiwać zbieżności do  $f(x)$ .

Przykład histogramu pokazany jest na rys. 4.1. Estymator jest stały na kolejnych odcinkach. Na każdym jego wartości jest równa liczbie obserwacji w nim zawartych, czyli 2, 5, 10, 12, 9, 6, 4, podzielonej przez liczbę wszystkich obserwacji, czyli  $2 + 5 + 10 + 12 + 9 + 6 + 4 = 48$ .



Rys. 4.1: Przykładowy histogram. Zaznaczono liczby obserwacji w poszczególnych odcinkach.

Rys. 4.2 przedstawia natomiast gęstość rozkładu normalnego na odcinku  $[-2, 25, 2, 25]$  i granicę, do której zbiega się histogram dla  $\Delta = 4, 5/18$ , gdzie 18 jest liczbą podocinków.



Rys. 4.2: Gęstość  $N(0,1)$  i granica (funkcja schodkowa) do której zbiega się estymator histogramowy.

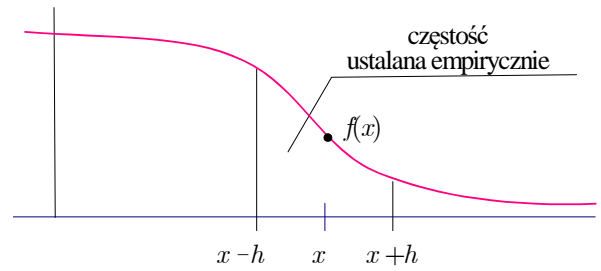
## 5 Metoda jądrowa

W metodzie jądrowej odcinek  $A$  ustala się jako  $[x-h, x+h]$ , gdzie  $h$  jest z góry przyjętą liczbą, patrz rys. 5.1, natomiast częstość wpadania obserwacji do tego zbioru ustala się empirycznie, co prowadzi do ułamka

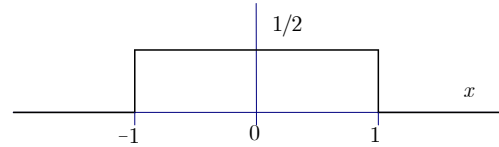
$$\begin{aligned} & \frac{\text{częstość wpadania obserwacji do odcinka } [x-h, x+h]}{2h} \\ &= \frac{\text{liczba obserwacji w odcinku } [x-h, x+h]}{2nh}. \end{aligned}$$

Aby go sprawnie wyliczyć, wprowadźmy jądro prostokątne  $K$ , patrz rys. 5.2,

$$K(v) = \begin{cases} \frac{1}{2}, & \text{dla } |v| \leq 1 \\ 0, & \text{w przeciwnym wypadku.} \end{cases} \quad (5.1)$$



Rys. 5.1: Idea estymatora jądrowego;  $h$  ustalone, częstość wyliczana empirycznie.



Rys. 5.2: Jądro prostokątne.

Zatem

$$\begin{aligned} & \text{liczba obserwacji w zbiorze } [x-h, x+h] \\ &= 2 \sum_{i=1}^n K\left(\frac{x-X_i}{h}\right), \end{aligned}$$

skąd wynika, że

$$\begin{aligned} & \frac{\text{częstość wpadania do zbioru } [x-h, x+h]}{2h} \\ &= \frac{1}{nh} \sum_{i=1}^n K\left(\frac{x-X_i}{h}\right). \end{aligned}$$

Uzależniając  $h$  od liczby obserwacji, otrzymujemy jądrowy estymator gęstości o postaci

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

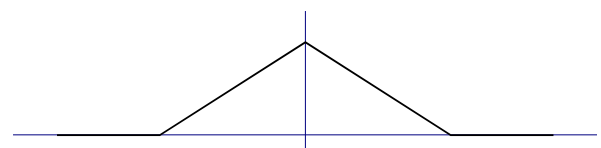
Klasę funkcji jądrowych można znacznie poszerzyć. Mogą to być funkcje ograniczone, które spełniają następujące warunki:

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

Przykłady takich funkcji, to jądro:

o trójkątne, patrz rys. 5.3,

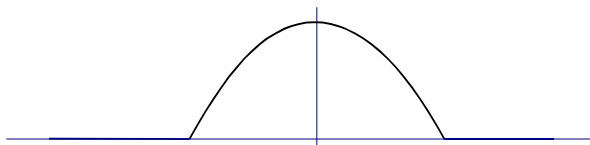
$$K(v) = \begin{cases} 1 - |v|, & \text{dla } |v| \leq 1 \\ 0, & \text{dla } |v| > 1, \end{cases}$$



Rys. 5.3: Jądro trójkątne.

o paraboliczne, rys. 5.4,

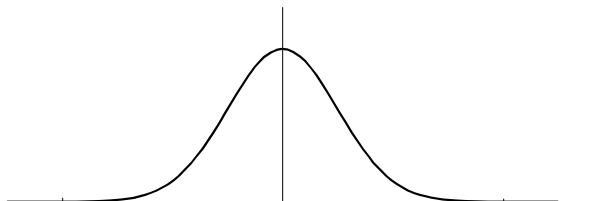
$$K(v) = \begin{cases} \frac{4}{3}(1-v^2), & \text{dla } |v| \leq 1 \\ 0, & \text{dla } |v| > 1, \end{cases}$$



Rys. 5.4: Jądro paraboliczne.

o Gaussa-Weierstrassa, rys. 5.5,

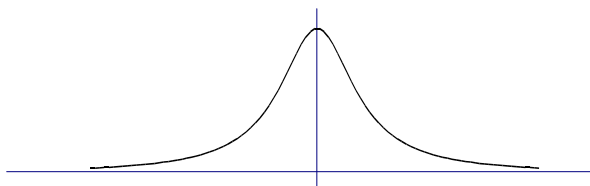
$$K(v) = \frac{1}{\sqrt{2\pi}} e^{-v^2/2}$$



Rys. 5.5: Jądro Gaussa-Weierstrassa.

o Poissona, rys. 5.6,

$$K(v) = \frac{1}{\pi} \frac{1}{1+v^2}$$



Rys. 5.6: Jądro Poissona.

o Fejéra, rys. 5.7,

$$K(v) = \frac{1}{\pi} \frac{\sin^2 v}{v^2}$$

o Lebesgue'a, rys. 5.8,

$$K(v) = \frac{1}{2} e^{-|v|}$$

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

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

Jeśli

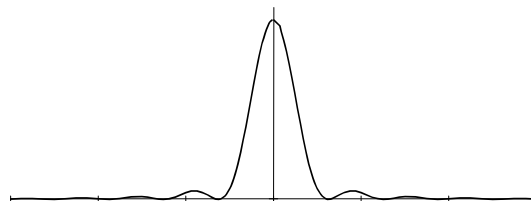
$$h(n) = cn^{-\gamma},$$

gdzie  $c > 0$ , to są one spełnione dla  $0 < \gamma < 1$ .

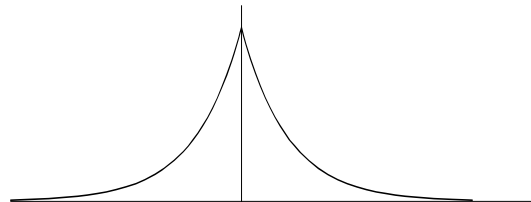
Przy funkcji jądra i ciągu liczbowym wybranymi we wskazany sposób estymator jest zgodny, tzn.

$$\lim_{n \rightarrow \infty} E(\hat{f}_n(x) - f(x))^2 = 0$$

w każdym punkcie  $x$  w którym gęstość  $f$  jest ciągła.



Rys. 5.7: Jądro Fejéra.



Rys. 5.8: Jądro Lebesgue'a.

Wykorzystując rezultaty podane w Ćwiczeniach 5.1 i 5.3, możemy napisać

$$\begin{aligned} E(\hat{f}_n(x) - f(x))^2 &= [E\hat{f}_n(x) - f(x)]^2 + \text{var}[\hat{f}_n(x)] \\ &\leq d_1^2 h^2(n) + d_2 \frac{1}{nh(n)}. \end{aligned} \quad (5.2)$$

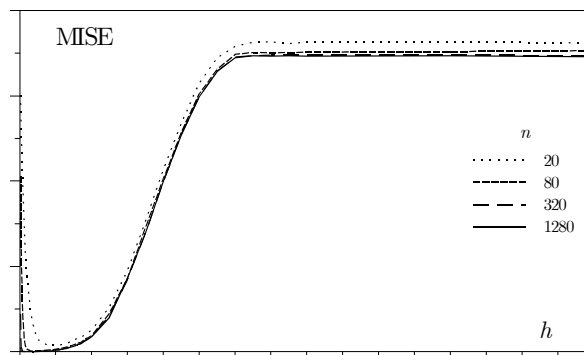
Widzimy więc, że małe  $h(n)$  zmniejsza pierwszy składnik sumy pochodzący od obciążenia  $E\hat{f}_n(x) - f(x)$  i jednocześnie niebezpiecznie zwiększa drugi pochodzący od wariancji. Właściwy dobór  $h(n)$  jest zatem kompromisem pomiędzy tymi sprzecznymi tendencjami.

Jeśli chodzi o zastosowania, to pojawiają się dwa problemy. Pierwszy to wybór funkcji jądrowej  $K$ , drugi to decyzja co do ciągu liczbowego  $h(n)$ . Należałoby je wybrać tak aby np. zminimalizować całkowity błąd średniokwadratowy

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

Analiza pokazuje, że dobór ciągu liczbowego jest znacznie istotniejszy. Z (5.2) wynika, że zdecydowanie należy przy tym unikać zbyt małego  $h(n)$ .

Komputerowe badania symulacyjne prowadzą do wyników pokazanych na rys. 5.9. Dla każdej liczby obserwacji  $n$  istnieje optymalne  $h$ . Wyniki potwierdzają wcześniejszą uwagę, że przyjęcie zbyt małej wartości jest niebezpieczne, gdyż prowadzi do gwałtownego wzrostu błędu spowodowanego wzrostem wariancji, patrz Ćwiczenie 5.3.

Rys. 5.9: Algorytm jądrowy. Zależność MISE od  $h$ .

**Ćwiczenie 5.1** Z (5.3) wynika, że dla nieujemnego jądra, obciążenie estymatora, czyli  $E\hat{f}_n(x) - f(x)$ , jest równe

$$\begin{aligned} & \frac{1}{h(n)} \int K\left(\frac{v-x}{h(n)}\right) f(v)dv - \int f(x)K(v)dv \\ &= \int K(v)f(x+h(n)v)dv - \int f(x)K(v)dv \\ &= \int K(v)[f(x+h(n)v) - f(x)]dv \end{aligned}$$

i co wartości bezwzględnej ograniczone przez

$$ch(n) \int |v|K(v)dv,$$

przy założeniu że  $|f(x+hy) - f(x)| \leq ch|y|$  dla wystarczająco małego  $h$  (Mówimy wtedy, że gęstość spełnia w punkcie  $x$  lokalny warunek Lipschitza). Dla  $d_1 = c \int |v|K(v)dv$  zatem

$$|E\hat{f}_n(x) - f(x)| \leq d_1 h(n).$$

**Ćwiczenie 5.2** Zauważmy, że przyjęte funkcje jądroowe zbiegają się do delty Diraca:

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

Zatem

$$E\hat{f}_n(x) \xrightarrow{n \rightarrow \infty} f(x) \int K(\xi)d\xi,$$

bowiem

$$E\hat{f}_n(x) = \frac{1}{h(n)} \int K\left(\frac{v-x}{h(n)}\right) f(v)dv, \quad (5.3)$$

która to wielkość zbiega się do

$$\int \delta(v-x)f(v)dv \int K(\xi)d\xi = f(x) \int K(\xi)d\xi,$$

gdy  $n \rightarrow \infty$ . Jeśli więc  $\int K(\xi)d\xi = 1$ , to

$$E\hat{f}_n(x) \xrightarrow{n \rightarrow \infty} f(x).$$

**Ćwiczenie 5.3** Ponieważ obserwacje są niezależne, zatem

$$\begin{aligned} \text{var}[\hat{f}_n(x)] &= \frac{1}{nh^2(n)} \text{var}\left[K\left(\frac{x-X}{h(n)}\right)\right] \\ &\leq \frac{1}{nh(n)} \left[\frac{1}{h(n)} EK^2\left(\frac{x-X}{h(n)}\right)\right]. \end{aligned}$$

Ze względu na to, że wyrażenie w nawiasie kwadratowym zbiega się do  $f(x) \int K^2(v)dv$ , patrz Ćwiczenie 5.2, istnieje  $d_2$ , takie że

$$\text{var}[\hat{f}_n(x)] \leq d_2 \frac{1}{nh(n)}.$$

## 6 Najbliższy sąsiad

W metodzie nazywanej "najbliższy sąsiad" (czyli NN, od *nearest neighbour*) częstość wpadania do odcinka ustala się z góry, natomiast jego długość zależy od obserwacji, patrz rys. 6.1.

Przed przystąpieniem do właściwej estymacji obserwacje (1.1) porządkuje się w nowy ciąg

$$X_{(1)}, X_{(2)}, X_{(3)}, \dots, X_{(n)}, \quad (6.1)$$

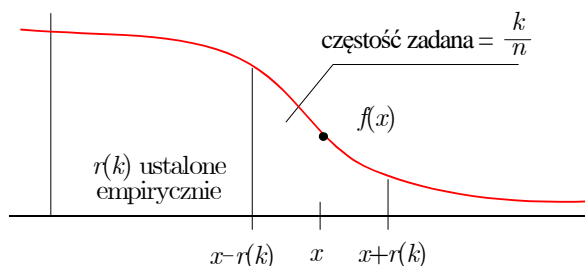
w którym

$$|X_{(1)} - x| < |X_{(2)} - x| < \dots < |X_{(n)} - x|.$$

Oznacza to, że  $X_{(1)}$  jest obserwacją położoną najbliżej punktu  $x$ ,  $X_{(2)}$  jest następną, natomiast  $X_{(n)}$  jest najbardziej oddaloną. Jako częstość wpadania do zbioru przyjmuje się ułamek  $k/n$ , gdzie  $k$  jest ustaloną z góry liczbą. Odcinkiem  $A$  jest zatem  $[x - r(k), x + r(k)]$ , gdzie  $r(k) = |X_{(k)} - x|$  jest odległością pomiędzy punktem  $x$  a  $k$ -tą najbliższą mu obserwacją. Wyrażenie w (3.1) przyjmuje zatem postać:

$$\frac{k/n}{2r(k)} = \frac{k}{2n|X_{(k)} - x|}.$$

Zadziwiające jest to, że występuje w nim bezpośrednio jedna tylko obserwacja  $X_{(k)}$ , aczkolwiek odpowiednio wyselekcjonowana.



Rys. 6.1: Idea estymatora typu najbliższy sąsiad; częstość zadana,  $r(k)$  wyliczone empirycznie.

Estymator nasz nie korzysta z wyjściowego ciągu (1.1) obserwacji uporządkowanych chronologicznie, lecz z ciągu (6.1), w którym obserwacje uporządkowane są według ich odległości od punktu  $x$ . Mówimy w związku z tym, że korzysta on ze statystyk porządkowych.

Jest oczywiste, że  $k$  powinno się zmieniać wraz z  $n$ , tak aby długość przedziału, czyli  $2a = 2|X_{(n)} - x|$ , malała do zera, co prowadzi do estymatora o postaci:

$$\bar{f}_n(x) = \frac{k(n)}{2n|X_{(k(n))} - x|}.$$

Analiza pokazuje, że warunkiem jego zbieżności do  $f(x)$  jest:

$$\lim_{n \rightarrow \infty} k(n) = \infty, \quad \lim_{n \rightarrow \infty} \frac{k(n)}{n} = 0.$$

**Uwaga 6.1** Algorytmy histogramowy, jądrowy i NN charakteryzują się pewnym podobieństwem. Zauważamy to, zapisując je w następujący sposób:

$$f_n(x) = \frac{\text{liczba } X_j \text{ w zbiorze } A_i}{n|A_i|}, \quad \text{dla } x \in A_i,$$

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

$$\hat{f}_n(x) = \frac{\text{liczba } X_j \text{ w zbiorze } A_x}{n|A_x|},$$

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

$$\tilde{f}_n(x) = \frac{\text{liczba } X_j \text{ w zbiorze } A_{k(n)}}{n|A_{k(n)}|},$$

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

## 7 Metoda rozwinięć ortogonalnych

Estymator ortogonalny skonstruowany jest według zupełnie innej zasady niż omówione powyżej. Jego istota polega na rozwinięciu funkcji gęstości w szereg ortogonalny.

Zbiór, na którym określona jest gęstość prawdopodobieństwa oznaczmy jako  $D$ . W przypadku rozkładu równomiernego jest to odcinek, dla rozkładu wykładniczego jest to półprosta  $[0, \infty)$ , dla normalnego cała prosta  $(-\infty, \infty)$ . Załóżmy dalej, że na zbiorze tym istnieje układ funkcji ortogonalnych

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

czyli takich, że

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

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

Dla gęstości  $f$  można skonstruować jej rozwinięcie ortogonalne

$$f(x) \sim \sum_{k=0}^{\infty} a_k \varphi_k(x). \quad (7.1)$$

Symbol " $\sim$ " oznacza, że

$$a_k = \int_D f(x)\varphi_k(x)dx = E\varphi_k(X). \quad (7.2)$$

Z teorii rozwinięć ortogonalnych wynika, że

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

Z uwagi na to, że współczynniki  $a_k$  rozwinięcia są wartościami oczekiwanymi, można je estymować jak poniżej:

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

W rezultacie, jako estymator gęstości przyjmuje się

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

gdzie  $N(n)$  jest pewnym ciągiem.

Zatem

$$\tilde{f}_n(x) - f(x) \sim \sum_{k=0}^{N(n)} (\tilde{a}_k - a_k) + \sum_{k=N(n)+1}^{\infty} (-a_k),$$

skąd, z uwagi na (7.1) i (7.3), wynika, że

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

Ponieważ  $E\tilde{a}_k = a_k$ , patrz (7.2), i

$$\text{var}[\tilde{a}_k] = \frac{1}{n} \text{var}[\varphi_k(X)] \leq \frac{1}{n} E\varphi_k^2(X) \leq \frac{1}{n} c,$$

gdzie  $c = \max_k(\max_x |\varphi_k(x)|)$ , zatem  $E(\tilde{a}_k - a_k)^2 \leq c/n$ . W rezultacie

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

Bez trudu wnioskujemy teraz, że jeśli

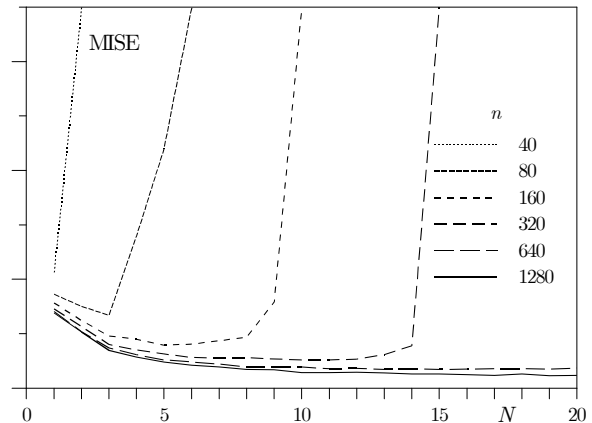
$$\lim_{n \rightarrow \infty} N(n) = \infty, \quad \lim_{n \rightarrow \infty} \frac{N(n)}{n} = 0,$$

to

$$\lim_{n \rightarrow \infty} E \int_D (\tilde{f}_n(x) - f(x))^2 dx = 0.$$

Estymator ortogonalny jest zatem zbieżny do nieznannej gęstości.

Przykładowe wyniki symulacji komputerowej przedstawiono na rys. 7.1. Widać wyraźnie, że błąd gwałtownie narasta przy zbyt dużym  $N$ . Długość rozwinięcia w szereg nie powinna być zatem duża.



Rys. 7.1: Algorytm ortogonalny. Zależność MISE od  $N$ .

Dla różnych zbiorów  $D$  należy wybrać odpowiednie układy funkcji ortogonalnych. Poniżej podano przykłady. Jeśli

- $D = [-\pi, \pi]$ , to można zastosować szereg trygonometryczny:

$$\frac{1}{\sqrt{2\pi}}, \frac{1}{\sqrt{\pi}} \sin kx, \frac{1}{\sqrt{\pi}} \cos kx, \quad k = 1, 2, \dots$$

- $D = [-\pi, \pi]$ , to można zastosować układ wielomianów Legendre'a:

$$p_k(x) = \sqrt{\frac{2k+1}{2}} P_k(x), \quad k = 0, 1, 2, \dots$$

gdzie

$$P_0(x) = 1,$$

$$P_1(x) = x,$$

$$P_2(x) = \frac{1}{2}(3x^2 - 1),$$

$$P_3(x) = \frac{1}{2}(5x^3 - 3x),$$

$$P_4(x) = \frac{1}{8}(35x^4 - 30x^2 + 3), \text{ itd.}$$

- $D = [0, \infty)$ , to można wykorzystać układ Laguerre'a:

$$l_k(x) = L_k(x)e^{-x/2},$$

gdzie

$$L_0(x) = 1,$$

$$L_1(x) = -x + 1,$$

$$L_2(x) = \frac{1}{2}x^2 - 2x + 1,$$

$$L_3(x) = -\frac{1}{6}x^3 + \frac{3}{2}x^2 - 3x + 1,$$

$$L_4(x) = \frac{1}{24}x^4 - \frac{2}{3}x^3 + 3x^2 - 4x + 1, \text{ itd.}$$

- $D = (-\infty, \infty)$ , to można stosowny jest układ Hermite'a:

$$h_k(x) = \frac{1}{\sqrt{2^k k! \sqrt{\pi}}} H_k(x) e^{-x^2/2},$$

gdzie

$$H_0(x) = 1,$$

$$H_1(x) = 2x,$$

$$H_3(x) = 8x^3 - 12x,$$

$$H_4(x) = 16x^4 - 48x^2 + 12, \text{ itd.}$$

Zwróćmy uwagę na to, że w układzie Legendre'a funkcje ortogonalne są wielomianami. W układzie Laguerre'a i Hermite'a funkcje te mają natomiast postać wielomianów  $L_k(x)$  i  $H_k(x)$  pomnożonych przez funkcje wagowe  $e^{-x/2}$  i  $e^{-x^2/2}$ .

## 8 Zakończenie

Jeśli nasza wiedza wstępna, czyli aprioryczna, o gęstości prawdopodobieństwa jest natury parametrycznej, tzn. jeśli znamy jej postać funkcyjną, problem sprowadza się do szacowania prawdziwych wartości parametrów. Szybkość z jaką maleje błąd średniokwadratowy jest rzędu  $1/n$ .

Jeśli natomiast nasza wiedza jest uboższa, jeśli postać funkcyjna nie jest znana, stosujemy metody nieparametryczne. Najprostrzy jest histogram, ale ma on nieusuwalny błąd, polegający na tym, że jest on zbieżny do funkcji, która jest jedynie przybliżeniem nieznannej gęstości. Szybkość zbieżności rzędu  $1/n$  jest jednak zachowana.

W pełni skuteczne są trzy omówione metody: jądrowa, NN i ortogonalna, aczkolwiek zbiegają się wolniej, a to ze względu na znacznie mniejszą informację aprioryczną o estymowanej gęstości.

Pierwsza jest najprostrza obliczeniowo. Jej wadą jest duża wrażliwość na dobór  $h(n)$ , bowiem, przy małej wartości tego współczynnika błąd może być bardzo duży, rzędu  $1/nh(n)$ .

Algorytm NN ma z kolei dobre własności lokalne. Obserwacje które decydują o jego wartości w punkcie  $x$  są bowiem najbliższymi jego sąsiadami, a ich odległość jest ponadto elastyczna. Jeśli bowiem wokół niego jest wiele obserwacji, sąsiedztwo to jest niewielkie, jeśli jest ich mało, sięga ono dalej. Wadą jest natomiast konieczność uszeregowania tych obserwacji w odpowiedniej kolejności i to dla każdego punktu oddzielnie.

W odróżnieniu od poprzednich, motywacja algorytmu ortogonalnego jest natury analitycznej, a nie probabilistycznej. Ma on dobre własności globalne, rozwinięciu podlega bowiem cała gęstość i żadnemu punktowi  $x$  nie poświęca się przy tym szczególnej uwagi. Zaletą jest to, że po zakończeniu obserwacji, aby obliczyć jego wartość w dowolnym punkcie wystarczy pamiętać  $N(n) + 1$  liczb, a mianowicie  $\tilde{a}_0, \tilde{a}_1, \dots, \tilde{a}_{N(n)}$ . W przypadku algorytmu jądrowego i NN niezbędne jest zachowanie w pamięci wszystkich  $n$  obserwacji  $X_1, X_2, \dots, X_n$ .