

JACEK KORONACKI
Warszawa
W. WERTZ
Wien

Rekurencyjna estymacja gęstości prawdopodobieństwa

(Praca wpłynęła do Redakcji 7.04.1983)



1. WSTĘP

1.1. Intencją tej pracy jest przedstawienie zwięzłego przeglądu teorii rekurencyjnej estymacji gęstości rozkładu prawdopodobieństwa. Rozważane tu ujęcie estymacji gęstości prawdopodobieństwa ma charakter ujęcia nieparametrycznego. Niniejszy przegląd nie rości sobie pretensji do kompletności, chodzi w nim raczej o danie intuicyjnego wglądu we własności rekurencyjnych metod estymacji i tym sposobem, zorientowany jest raczej na podkreślenie praktycznego znaczenia tych metod.

Metody rekurencyjne stanowią naturalny pierwowzór metod ogólniejszych - sekwencyjnych. W porównaniu z sumaryczną ilością literatury poświęconej estymacji gęstości prawdopodobieństwa, jej część poświęcona teorii sekwencyjnej jest znikoma i wiele problemów tej teorii pozostaje dotąd otwartych.

Autorzy tego przeglądu mają nadzieję uzupełnić go w bliskiej przyszłości częścią drugą, dotyczącą własności metod sekwencyjnych (patrz Carroll [6], Deheuvels [13], Farrell [21], Srivastava [48] i ostatnio Stute [52] oraz Wertz [66]).

Wyczerpujący przegląd teorii i metod nieparametrycznej estymacji gęstości prawdopodobieństwa, szczegółowo omawiający wyniki opublikowane do roku 1977 i obejmujący metody tu nierozważane (bo nierekurencyjne), podał Wertz [64]. Bibliografia przedmiotu, do roku 1977, zebrana została przez Werta i B.Schneider [44].

Ponieważ artykuł ten adresowany jest do czytelnika polskiego, warto przynajmniej wymienić nazwiska tych specjalistów krajowych, z których prac tu wprawdzie nie korzystamy, ale którzy od lat zajmują się problematyką nieparametrycznej estymacji gęstości (oraz pokrewną, np. estymacją funkcji regresji i zastosowaniami tych klas estymatorów). I tak, badaniem (m.in.) podstawowych problemów zbieżności zajmują się W.Greblicki i jego współpracownicy, zwłaszcza A.Krzyżak i M.Pawlak. Porównaniami różnych typów estymatorów, zwłaszcza opartymi na wynikach numerycznych, zajmuje się grupa M.Krzyński, przede wszystkim P.Stolarski.

Niniejszy artykuł jest w istocie zmodyfikowaną wersją pracy Werta [66].

W rozdziale drugim przedyskutowane są podstawowe własności asymptotyczne - zgodność, asymptotyczna nieobciążoność i asymptotyczna wariancja - najbardziej typowych estymatorów rekurencyjnych. Nacisk jest tu położony na uzyskanie wspomnianych rezultatów przy możliwie najslabszych założeniach,

zwłaszcza o szacowanej gęstości. Ze względu na możliwość późniejszych porównań estymatorów rekurencyjnych z nierekurencyjnymi, część rozdziału poświęcona jest tym ostatnim. Pierwsze paragrafy rozdziału 2 mają charakter wprowadzający - pojęcie estymatora rekurencyjnego pojawia się w paragrafie 2.4.

Rozdział trzeci poświęcony jest analizie prędkości zbieżności średniokwadratowej różnych typów estymatorów. W rozdziale czwartym rozważone są prawa iterowanego logarytmu, w rozdziale piątym omówiony jest problem asymptotycznej normalności. Wszystkie te rozważania prowadzi się przy mocniejszych założeniach od tych, jakie wystarczało przyjąć w rozdziale 2; w rozdziałach tych systematycznie stosujemy dwa różne zbiory założeń.

W celu uproszczenia zapisu ograniczymy się głównie do rozważania przypadku obserwacji jednowymiarowych, chociaż większość wyników można by uogólnić na przypadek wielowymiarowy. Ażeby wyniki uczynić bardziej przejrzystymi, oddzielnie będziemy rozważać wielkości wariancji i obciążenia estymatorów (p.n.p. paragrafy 3.4, 3.5 i 3.6).

W rozdziałach 2 - 5 nacisk położony jest na lokalne własności estymatorów (przez własności lokalne rozumiemy własności estymatorów w ustalonym punkcie jego określoności). W rozdziale szóstym zamieszczone są pewne informacje o badaniu globalnych własności estymatorów gęstości, mianowicie o badaniu ich zbieżności jednostajnej i (przede wszystkim) o badaniu asymptotycznych własności średniego scałkowanego błędu kwadratowego.

Chcąc uczynić nasz przegląd bardziej przejrzystym, zdecydowaliśmy się na stosunkowo "drobny" podział kolejnych roz-

działów. Twierdzenia, lematy, wnioski i uwagi nie są w rezultacie oddzielnie numerowane; odwołując się do nich opatrujemy je numerami paragrafów, w których się znajdują.

1.2. W dalszym ciągu będziemy stosować następujące oznaczenia:

$R(N)$	zbiór liczb rzeczywistych (naturalnych),
$R^r, r \in N,$	r -wymiarowa przestrzeń euklidesowa (R^1 będziemy też oznaczać symbolem R),
1	funkcja charakterystyczna zbioru,
$\ \cdot\ _\infty$	istotny kres górny funkcji,
f	gęstość rozkładu prawdopodobieństwa,
$f' (f'')$	pochodna (druga pochodna) funkcji f ,
$E_f(V_f)$	wartość oczekiwana (wariancja) obliczana przy gęstości f ,
$\xrightarrow{P}$	zbieżność według prawdopodobieństwa,
$N(\xi, \delta^2)$	rozkład normalny o parametrach ξ, δ^2 ,
Φ	dystybuanta rozkładu $N(0, 1)$,
$\lambda^n, P^n, \dots$	miary produktowe.

1.3. Niech będzie dana przestrzeń statystyczna $(X, \mathcal{X}, \mathcal{P})$, przy czym X jest przestrzenią $R = R^1$ lub $R^r (1 < r < \infty)$ i $\mathcal{X}$ jest odpowiednim σ -ciałem borelowskim, i niech dla każdej miary probabilistycznej $P \in \mathcal{P}$ istnieje gęstość f względem miary Lebesgue'a λ na $(X, \mathcal{X})$. Niech ξ oznacza zmienną losową określoną na pewnej przestrzeni probabilistycznej $(\Omega, \mathcal{A}, P)$ o wartościach w X i rozkładzie $P \in \mathcal{P}$; tzn. $P[\omega \in \Omega: \xi(\omega) \in A] = P(A) = \int_A f d\lambda$ dla każdego zbioru $A \in \mathcal{X}$. Niech dalej $\xi_1, \dots,$

$\xi_n, \dots$ będą niezależnymi zmiennymi losowymi o tym samym rozkładzie co zmienna losowa ξ . Przyjmijmy jeszcze następujące oznaczenia: $\xi^n := (\xi_1, \dots, \xi_n)$, x_n oznacza realizację (wartość) zmiennej losowej ξ_n , $x^n := (x_1, \dots, x_n)$ - realizację wektora ξ^n . Niech wreszcie $\mathcal{F}$ oznacza rodzinę gęstości odpowiadających rodzinie rozkładów $\mathcal{P}$. O rodzinach tych zakładamy, że nie dają się sparametryzować parametrem skończenie wymiarowym ($\mathcal{F}$ może być na przykład rodziną wszystkich gęstości ciągłych lub rodziną gęstości dwukrotnie różniczkowalnych i ograniczonych).

Nasze zadanie polega na szacowaniu nieznannej gęstości $f \in \mathcal{F}$ na podstawie zaobserwowanej wartości próby losowej, x^n . Przy ustalonym rozmiarze n próby losowej, estymatorem gęstości f nazywamy każdą mierzalną funkcję $\hat{f}_n: X^n \times X \rightarrow \mathbb{R}$. Oczywiście, na estymator $\hat{f}_n$ nakładamy zwykle jakieś dodatkowe ograniczenia, na przykład aby należał do pewnej przestrzeni funkcyjnej.

Problem estymacji, w szczególności więc problem "jakości" estymatora, sformułujemy w języku teorii decyzji, opierając się na kwadratowej funkcji straty. Za funkcję straty przyjmujemy mianowicie funkcję $L(\hat{f}_n, f) = \int_X [\hat{f}_n(\xi^n, x) - f(x)]^2 d\mu(x)$, gdzie μ jest stosownie wybraną miarą na $(X, \mathcal{X})$. Ryzyko przyjmie zatem postać $R(\hat{f}_n, f) = E_f \int_{X^n} [\hat{f}_n(\xi^n, x) - f(x)]^2 d\mu(x)$. W szczególności, ryzyko estymatora $\hat{f}_n$ gęstości f , gdy miara μ skoncentrowana jest w punkcie x (i ma masę równą 1), przyjmuje postać błędu średniokwadratowego (BSK) estymatora $\hat{f}_n$ w tym punkcie

$$(1.1) \quad \text{BSK}(x) := \text{BSK} = E_f [\hat{f}_n(\xi^n; x) - f(x)]^2.$$

BSK daje lokalną miarę jakości estymatora. Odpowiednio, przy $\mu = \lambda$, ryzyko równa się średniemu scałkowanemu błędowi kwadratowemu (SSBK) i stanowi globalną miarę jakości estymatora:

$$(1.2) \quad \begin{aligned} \text{SSBK} &= E_f \int_X [\hat{f}_n(\xi^n; x) - f(x)]^2 d\lambda(x) = \\ &= \int_{X^n} \int_X [\hat{f}_n(x^n; x) - f(x)]^2 d\lambda(x) dP^n(x^n). \end{aligned}$$

Łatwo zauważyć, że przy stosownych założeniach dodatkowych mamy

$$(1.3) \quad \begin{aligned} R(\hat{f}_n, f) &= E_f \int [\hat{f}_n(\xi^n; x) - f(x)]^2 d\mu(x) = \\ &= \int E_f [\hat{f}_n(\xi^n; x) - E_f \hat{f}_n(\xi^n; x)]^2 d\mu(x) + \\ &+ \int [E_f \hat{f}_n(\xi^n; x) - f(x)]^2 d\mu(x). \end{aligned}$$

W szczególności, w przypadku BSK (1.1), równość (1.3) zachodzi bez żadnych założeń dodatkowych. Równość (1.3) pokazuje, że - przy ewentualnych warunkach dodatkowych - ryzyko dla kwadratowej funkcji straty daje się rozbić na dwie całki względem miary μ : funkcja podcałkowa pierwszej całki równa jest w punkcie wariancji $V_f \hat{f}_n(\xi^n; x)$ estymatora $\hat{f}_n$, natomiast funkcja podcałkowa drugiej całki równa jest w punkcie x kwadratowi obciążenia, $[B(\hat{f}_n(\xi^n; x), f(x))]^2$, estymatora $\hat{f}_n$.

Kończąc ten punkt wypada podkreślić, że problem estymacji może być sformułowany w sposób znacznie ogólniejszy niż to zostało uczynione wyżej; p. [29], [30]. Odnotujmy też na marginesie, że niektóre z rozważanych niżej zagadnień można uogólnić na przypadek zmiennych losowych $\xi_1, \dots, \xi_n, \dots$ o tej samej (nieznanej) gęstości, ale nie niezależnych; p. [64] oraz paragraf 6.2.

1.4. Zanim przejdziemy do omówienia estymatorów rekurencyjnych, przytoczymy jeszcze dwa ważne fakty ogólne, dotyczące klasy wszystkich estymatorów gęstości (opartych na próbie n -elementowej ξ^n).

Mówimy, że estymator $\hat{f}_n$ jest nieobciążony w klasie $\mathcal{F}$ gęstości f określonych na prostej R , jeżeli dla każdego $x \in R$ i każdej gęstości $f \in \mathcal{F}$

$$E_{f, \xi^n}(\hat{f}_n; x) = f(x).$$

Bickel i Lehmann [3] wykazali, że jeżeli $\mathcal{F}$ jest wypukłą rodziną gęstości (tzn. jeżeli f i g należą do $\mathcal{F}$, to dla każdego $a \in [0, 1]$ gęstość $af + (1 - a)g \in \mathcal{F}$) zawierającą gęstości podwójnie wykładnicze,

$$\frac{c}{2} \exp(-c|x|), \quad c > 0,$$

to w klasie tej nie istnieje nieobciążony estymator gęstości. Stwierdzenie to wyklucza zatem istnienie estymatorów nieobciążonych w dostatecznie bogatych klasach $\mathcal{F}$. Zarazem nie

wyklucza oczywiście istnienia estymatorów asymptotycznie nieobciążonych, tzn. takich, że

$$\lim_{n \rightarrow \infty} E_{\hat{f}_n}(\xi_n^x) = f(x) \quad \text{dla każdego } x \in R.$$

(Twierdzenie Bickela-Lehmana można wzmocnić, jeśli ograniczyć się do estymatorów jądrowych, rozważanych w następnych rozdziałach; p.[70].)

Po drugie, okazuje się, że dla dostatecznie bogatych rodzin gęstości $\mathcal{F}$ nie istnieją estymatory jednostajnie (ze względu na tę rodzinę) zgodne. Z twierdzenia podanego przez Farrella [21] wynika mianowicie, że jeżeli $\mathcal{F}$ jest rodziną gęstości f na prostej R , różniczkowalnych w sposób ciągły i takich, że $\sup_{x \in R} f(x) \leq \alpha$ dla pewnej ustalonej liczby $\alpha \geq 3$, to dla każdego estymatora f_n spełniona jest nierówność

$$\sup_{f \in \mathcal{F}} E_f[\hat{f}_n(\xi_n^0) - f(0)]^2 \geq 1/16.$$

Twierdzenie Farrella nie wyklucza istnienia estymatorów zgodnych. Podaje natomiast ograniczenie dolne na maksymalne (w klasie $\mathcal{F}$ możliwych gęstości) ryzyko estymatorów opartych na próbie losowej o ustalonej liczności.

2. ESTYMATORY REKURENCYJNE. WŁASNOŚCI PODSTAWOWE

2.1. Jedną z najpopularniejszych metod estymacji krzywych gęstości jest konstrukcja tzw. estymatorów jądrowych. Estymatory jądrowe wprowadził Rosenblatt [37]. Idea konstrukcji takich estymatorów podjęta została przez Parzena [33] i odtąd

poświęcono jej bardzo wiele prac (p. [64], [44]).

Podstawą tej idei jest oparcie konstrukcji estymatora nieznanej gęstości f (określonej na prostej) na ilorazie różnicowym dystrybuanty empirycznej:

$$\hat{f}_n(x^n; x) = [F_n(x^n; x + b_n) - F_n(x^n; x - b_n)] / 2b_n,$$

gdzie F_n oznacza dystrybuantę empiryczną nieznanego rozkładu,

$$F_n(x^n; x) = \frac{1}{n} \sum_{i=1}^n 1_{(-\infty, x]}(x_i),$$

a b_n jest współczynnikiem dodatnim. Estymator $\hat{f}_n$ możemy zapisać w postaci równoważnej

$$\begin{aligned} \hat{f}_n(x^n; x) &= \frac{1}{2b_n} \left[\sum_{i=1}^n 1_{(-\infty, x+b_n]}(x_i) - \sum_{i=1}^n 1_{(-\infty, x-b_n]}(x_i) \right] = \\ &= \frac{1}{2b_n} \sum_{i=1}^n 1_{(x-b_n, x+b_n]}(x_i) = \frac{1}{n} \sum_{i=1}^n \frac{1}{b_n} K\left(\frac{x-x_i}{b_n}\right), \end{aligned}$$

gdzie $K(t) = \frac{1}{2} 1_{[-1, 1]}(t)$.

Naturalne jest zatem określenie estymatora ogólniejszej postaci

$$(2.1) \quad \hat{f}_n(x^n; x) = \frac{1}{n} \sum_{i=1}^n \frac{1}{b_n} K\left(\frac{x-x_i}{b_n}\right),$$

gdzie $K: \mathbb{R} \rightarrow \mathbb{R}$ jest funkcją mierzalną i taką, że

$$(2.2) \quad \int K(t) dt = 1 \quad \text{oraz} \quad \int tK(t) dt = 0,$$

ciąg (b_n) spełnia zaś warunki

$$(2.3) \quad b_n > 0, \quad \lim_{n \rightarrow \infty} b_n = 0, \quad \lim_{n \rightarrow \infty} n b_n = \infty.$$

(Sens dwóch pierwszych warunków (2.3) jest oczywisty, natomiast sens wymagania, aby ciąg (b_n) nie dążył do zera zbyt szybko, wyjaśni się niżej.) Estymatory opisanego typu noszą nazwę estymatorów jądrowych. Estymator (2.1) będziemy odtąd nazywać estymatorem Rosenblatt-Parzena (ERP), funkcję K będziemy nazywać jądrem, współczynnik b_n - szerokością pasma. Będziemy też odtąd zawsze zakładać, że jądro K jest nieujemne. Zauważmy, że przy przyjętych założeniach funkcja $x \mapsto f_n(x_1, \dots, x_n; x)$ jest sama, dla każdego zbioru obserwacji x^n , gęstością prawdopodobieństwa. Zauważmy też, że własności tej funkcji dotyczące jej regularności (np. różniczkowalność, całkowalność w kwadracie), są niejako powtórzeniem własności jądra K .

2.2. Estymatory (2.1) można łatwo uogólnić na przypadek zmiennych losowych r -wymiarowych. Można na przykład przyjąć (p.[5])

$$(2.4) \quad \hat{f}_n(x_1, \dots, x_n; x) = \frac{1}{n} \sum_{i=1}^n \frac{1}{b_n^r} K\left(\frac{x - x_i}{b_n}\right),$$

gdzie $K : \mathbb{R}^r \rightarrow \mathbb{R}$ i warunki (2.3) zastąpione są następującymi:

$$(2.5) \quad b_n > 0, \quad \lim_{n \rightarrow \infty} b_n = 0, \quad \lim_{n \rightarrow \infty} n b_n^{(r)} = \infty.$$

Ogólniejszą definicję zaproponowała Văduva [54]. Niech mianowicie $x_i = (x_i^{(1)}, \dots, x_i^{(r)})$, $x = (x^{(1)}, \dots, x^{(r)})$, $b_n = (b_n^{(1)}, \dots, b_n^{(r)})$ oraz

$$(2.6) \quad b_n^{(j)} > 0, \quad \lim_{n \rightarrow \infty} \|b_n\| = 0, \quad \lim_{n \rightarrow \infty} n b_n^{(1)} \dots b_n^{(r)} = \infty.$$

Estymator $\hat{f}_n$ określamy wówczas następująco:

$$(2.7) \quad \hat{f}_n(x_1, \dots, x_n; x) = \frac{1}{n} \sum_{i=1}^n (b_n^{(1)} \dots b_n^{(r)})^{-1} \times \\ \times K\left(\frac{x^{(1)} - x_i^{(1)}}{b_n^{(1)}}, \dots, \frac{x^{(r)} - x_i^{(r)}}{b_n^{(r)}}\right).$$

Estymatory jądrowe gęstości określonych na ogólniejszych przestrzeniach prób (grupach topologicznych, przestrzeniach jednorodnych itd.) opisane są np. w [9], [63] i [65].

2.3. Inny rodzaj uogólnień dotyczy samej postaci estymatora, bez względu na to, czy określony jest na prostej, czy na przestrzeni R^r . Niech (K_n) będzie ciągiem funkcji mierzalnych (określonych na R lub R^r). Przy różnych założeniach Watson i Leadbetter [60], [61], Văduva [56] i inni autorzy badali estymator o postaci

$$(2.8) \quad \hat{f}_n(x_1, \dots, x_n; x) = \frac{1}{n} \sum_{i=1}^n K_n(x - x_i).$$

Sens czynionych założeń sprowadzał się do tego, by zapewnić ciągłowi (K_n) własność bycia "jednostką aproksymacyjną", tzn. by

$$(2.9) \quad \lim_{n \rightarrow \infty} \int K_n(x-y) \varphi(y) dy = \varphi(x)$$

dla każdej dostatecznie gładkiej funkcji φ . (Watson i Leadbetter [59], Hašimov [28], Mirzahmedov i Hašimov [32] badali estymator (2.8) metodami analizy fourierowskiej.) Whittle [67], Földes i Révész [23], Földes [22], Walter i Blum [58] oraz Susarla i Walter [53] zaproponowali jeszcze ogólniejszą klasę estymatorów:

$$(2.10) \quad \hat{f}_n(x_1, \dots, x_n; x) = \frac{1}{n} \sum_{i=1}^n K_n(x, x_i),$$

gdzie (K_n) jest ciągiem funkcji mierzalnych (określonych na R^2 lub R^{2r}); zasadniczą własnością funkcji K_n powinno być spełnianie związku analogicznego do (2.9),

$$(2.9') \quad \lim_{n \rightarrow \infty} \int K_n(x, y) \varphi(y) dy = \varphi(x)$$

dla każdej dostatecznie gładkiej funkcji φ . (Ze względu na własności (2.9) i (2.9') niektórzy autorzy nazywają estymatory typu (2.8) i (2.10) estymatorami z ciągami deltowymi; ang. delta - sequence estimators.) Okazuje się, że wiele rodzajów estymatorów gęstości - np. estymatory (2.1), (2.4) i (2.7), estymatory oparte na rozwinięciach ortogonalnych, histogramy,

niektóre estymatory oparte na interpolowaniu funkcji - to szczególne przypadki estymatora typu (2.10).

Dla celów teorii estymacji rekurencyjnej przydatna jest następująca ogólna definicja estymatora gęstości, zaproponowana przez Deheuvelsa [15]:

$$(2.11) \quad f_n(x_1, \dots, x_n; x) = \sum_{k=1}^n K_{n,k}(x, x_k),$$

gdzie mierzalne funkcje $K_{n,k}$ spełniają warunki analogiczne do warunku (2.9').

2.4. Ciąg $(\hat{f}_n)$ estymatorów o postaci (2.11) nazywamy rekurencyjnym (lub krócej, ale niezbyt poprawnie, $\hat{f}_n$ nazywamy estymatorem rekurencyjnym), jeżeli dla każdego $n \in \mathbb{N}$, każdego x oraz każdego ciągu x^∞ obserwacji

$$\hat{f}_{n+1}(x_1, \dots, x_{n+1}; x) = R_n(\hat{f}_n(x_1, \dots, x_n; x), x_{n+1}, x),$$

gdzie R_n są funkcjami mierzalnymi, określonymi na przestrzeni o stosownym wymiarze.

Podana ogólna definicja estymatorów rekurencyjnych należy do Deheuvelsa [16]. Deheuvels udowodnił także następujące twierdzenie

TWIERDZENIE [16]. Niech $\hat{f}_n$ będzie estymatorem rekurencyjnym o postaci (2.11), funkcje $u \mapsto R_n(u, x_{n+1}, x)$ i $y \mapsto K_{n,k}(x, y)$ niech będą różniczkowalne, przy

czym druga z nich nie jest stałą. Wówczas estymator $\hat{f}_n$ można przedstawić w postaci

$$\hat{f}_n(x_1, \dots, x_n; x) = c_n(x) \sum_{k=1}^n K_k^*(x, x_k),$$

gdzie $c_n(x)$ są pewnymi stałymi (zależnymi od x) i K_k^* są jądrami mierzalnymi.

2.5. W dalszym ciągu pracy będziemy rozważać w głównej mierze następujące typy estymatorów jądrowych (przy założeniu $X = R$):

(i) ERP (2.1);

(ii) estymator Deheuvelsa, w skrócie ED, o postaci

$$(2.12) \quad \hat{f}_n(x_1, \dots, x_n; x) = \left[\sum_{i=1}^n b_i H(b_i) \right]^{-1} \sum_{k=1}^n H(b_k) K\left(\frac{x-x_k}{b_k}\right),$$

gdzie H jest pewną ustaloną funkcją dodatnią, K jest jądrem, $a(b_n)$ jest ciągiem współczynników dodatnich;

(iii) estymator zaproponowany przez Wolvertona i Wagnera [68] oraz niezależnie przez Yamato [69], w skrócie EWWY,

$$(2.13) \quad \hat{f}_n(x_1, \dots, x_n; x) = \frac{1}{n} \sum_{k=1}^n \frac{1}{b_k} K\left(\frac{x-x_k}{b_k}\right);$$

- (iv) inny estymator Deheuvelsa, przedstawiony w [13], w skrócie ED^* ,

$$(2.14) \quad \hat{f}_n(x_1, \dots, x_n; x) = \left[\sum_{i=1}^n b_i \right]^{-1} \sum_{k=1}^n K\left(\frac{x-x_k}{b_k}\right).$$

Estymator (i) nie jest oczywiście estymatorem rekurencyjnym i rozważany będzie ze względu na możliwość jego porównań z estymatorami rekurencyjnymi. Estymatory (iii) i (iv) są szczególnymi przypadkami estymatora (ii), powstającymi przez podstawienie odpowiednio $H(b) = 1/b$ i $H \equiv 1$. Jak to zobaczymy dalej, pewne własności każdego z estymatorów (iii) i (iv) są asymptotycznie optymalne. Z kolei estymator (ii) jest szczególnym przypadkiem poniższego

$$(2.15) \quad \hat{f}_n(x_1, \dots, x_n; x) = \left[\sum_{i=1}^n b_i H(b_i) \right]^{-1} \sum_{k=1}^n H(b_k) \tilde{K}\left(x, \frac{x-x_k}{b_k}\right),$$

będącego szczególnym przypadkiem estymatora (2.11). Estymator typu (2.15) badany był w [13] przy założeniu, że funkcja $y \mapsto \tilde{K}(x, y)$ jest dla każdego x gęstością prawdopodobieństwa.

2.6. W następnych paragrafach tego rozdziału wykazemy zgodność (w tym lub innym sensie) i asymptotyczną nieobciążoność oraz obliczymy wartości asymptotycznych wariancji wymienionych wyżej estymatorów jądrowych. Wszystkie te rozważania będziemy prowadzić przy możliwie łagodnych założeniach o nieznannej gęstości f (i jądrze K). Podstawowe znaczenie zgod-

ności estymatorów jest oczywiste. Asymptotyczna nieobciążoność jest również istotna, zwłaszcza w świetle twierdzenia Bickela-Lehmanna (p.punkt 1.4). Obliczenie asymptotycznych wariancji da nam pierwszą możliwość porównania różnych estymatorów.

W rozdziale 3 rozważymy, już przy mocniejszych założeniach prędkość malenia błędu średniokwadratowego, BSK. Zgodnie z równością (1.3) prędkość ta wyznaczona jest przez prędkość malenia wariancji oraz obciążenia estymatora w punkcie x . Okazuje się, że, przy ustalonym jądrze K , wybór dużej szerokości pasma daje małą wariancję, ale jednocześnie powiększa obciążenie. Na przykład, w przypadku ERP (2.1) z jądrem spełniającym warunek (2.2), w szczególności zatem w przypadku ilorazu różnicowego dystrybuanty empirycznej, otrzymuje się

$$V_{\hat{f}_n}(\xi^n; x) = O\left(\frac{1}{nb_n}\right) \quad \text{oraz} \quad B(\hat{f}_n(\xi^n; x), f(x)) = o(b_n)$$

jeżeli gęstość $f : \mathbb{R} \rightarrow \mathbb{R}$ jest różniczkowalna i $\|f'\|_\infty < \infty$.

Intuicyjnie, można oczekiwać, że mała wariancja oznacza uzyskanie oszacowań nieznanej gęstości nadmiernie wygładzonych, duża zaś - nadmiernie oscylacyjnych. Pewne ściśle i bardzo proste potwierdzenie tej intuicji daje odpowiednie prawo iterowanego logarytmu; p. § 4.5.

2.7. W dalszym ciągu tego rozdziału bardzo istotną rolę będzie pełnić podany niżej lemat, który poprzedzimy jeszcze stosownymi definicjami.

DEFINICJA. Punkt $x \in R^r$ jest punktem Lebesgue'a funkcji

$$\bar{f} : R^r \longrightarrow R, \text{ jeżeli}$$

$$\lambda(S) \xrightarrow[S \in \delta]{} 0 \quad \frac{1}{\lambda(S)} \int_S |\bar{f}(x+y) - \bar{f}(x)| dy = 0,$$

gdzie δ jest rodziną kul otwartych w R^r o środkach w zerze.

Na mocy twierdzenia o wartości średniej, jeżeli $\bar{f}$ jest ciągła w punkcie x , to punkt ten jest punktem Lebesgue'a. Dobrze znany jest także fakt ogólniejszy, a mianowicie, że jeżeli $\bar{f}$ jest funkcją całkowalną, to prawie każdy punkt przestrzeni R^r jest jej punktem Lebesgue'a (p.np. [49]).

DEFINICJA. Niech będą dane dwie funkcje całkowalne

$\bar{f}, \bar{K} : R^r \longrightarrow R$. Będziemy mówili, że w punkcie $x \in R^r$ zachodzi warunek $[D_x]$, jeżeli spełniony jest jeden z poniższych warunków:

- (1) funkcja $\bar{f}$ jest prawie wszędzie ograniczona, x jest punktem Lebesgue'a funkcji $\bar{f}$;
- (2) funkcja $\bar{K}$ ma zwarty nośnik, x jest punktem Lebesgue'a funkcji $\bar{f}$;
- (3) $\lim_{\|t\| \rightarrow \infty} \|t\|^r \bar{K}(t) = 0$, funkcja $\bar{f}$ jest ciągła w punkcie x ;
- (4) $\int_0^\infty u^{r-1} L(u) du < \infty$, x jest punktem Lebesgue'a funkcji $\bar{f}$,
przy czym $L(u) := \sup_{\|x\| \geq u} |\bar{K}(x)|$;

$$(5) \quad \int_{\mathbb{R}^r} L^*(y) dy < \infty, \quad x \text{ jest punktem Lebesgue'a funkcji } \bar{f},$$

$$\text{przy czym } L^*(y) := \sup_{\|x\| \geq \|y\|} |\bar{K}(x)|.$$

Odnajdujemy, że związek $\lim_{\|t\| \rightarrow \infty} \|t\|^r \bar{K}(t) = 0$ występujący w warunku (3) równoważny jest następującemu

$$\lim_{u \rightarrow \infty} u^r L(u) = 0,$$

który z kolei jest nieco słabszy od żądania $\int_0^\infty u^{r-1} L(u) du < \infty$, występującego w warunku (4).

LEMAT ([33], [17], [18]). Niech będzie spełniony warunek $[D_x]$

$\|\bar{K}\|_\infty < \infty$, niech (b_n) będzie ciągiem liczb dodatnich

$\lim_{n \rightarrow \infty} b_n = 0$. Wówczas

$$\lim_{n \rightarrow \infty} \frac{1}{b_n^r} \int_{\mathbb{R}^r} \bar{K}\left(\frac{x-\xi}{b_n}\right) \bar{f}(\xi) d\xi = \bar{f}(x) \int_{\mathbb{R}^r} \bar{K}(y) dy.$$

D o w ó d. Udowodnimy prawdziwość lematu tylko przy warunkach (1) i (2). Przy założeniu warunku (3) dowód lematu podał Parzen [33]. Przy założeniu warunku (5) lemat został udowodniony przez Steina [49] (p. rozdz. III. 2, odpowiednia część twierdzenia 2). Prawdziwość lematu przy założeniu (4) jest wnioskiem z twierdzenia Steina (por. [18]).

Zauważmy, że

$$(2.16) \quad \frac{1}{b_n^r} \int_{\mathbb{R}^r} \bar{K}\left(\frac{x-\xi}{b_n}\right) \bar{f}(\xi) d\xi - \bar{f}(x) \int_{\mathbb{R}^r} \bar{K}(y) dy =$$

$$= \frac{1}{b_n^r} \int_{R^r} \bar{K}\left(\frac{y}{b_n}\right) [\bar{f}(x-y) - \bar{f}(x)] dy.$$

Dalej, oznaczając $S(x, \delta) := \{y: \|y-x\| < \delta\}$, łatwo zauważyć, że dla każdego ustalonego $\delta > 0$

$$\begin{aligned} (2.17) \quad & \int_{\|y\| < \delta b_n} |\bar{f}(x-y) - \bar{f}(x)| b_n^{-r} |\bar{K}\left(\frac{y}{b_n}\right)| dy = \\ & = \int_{S(x, \delta b_n)} |\bar{f}(z) - \bar{f}(x)| b_n^{-r} |\bar{K}\left(\frac{x-z}{b_n}\right)| dz. \end{aligned}$$

Ale

$$\lambda(S(x, \delta b_n)) = \lambda(S(0, \delta b_n)) = b_n^r \lambda(S(0, \delta))$$

i zapisując całkę (2.17) w postaci równoważnej, otrzymujemy

$$\begin{aligned} & \int_{S(x, \delta b_n)} |\bar{f}(z) - \bar{f}(x)| \frac{\lambda(S(0, \delta))}{\lambda(S(x, \delta b_n))} |\bar{K}\left(\frac{x-z}{b_n}\right)| dz \leq \\ & \leq \|\bar{K}\|_{\infty} \cdot \lambda(S(0, \delta)) \int_{S(x, \delta b_n)} \frac{|\bar{f}(z) - \bar{f}(x)|}{\lambda(S(x, \delta b_n))} dz \rightarrow 0, \end{aligned}$$

gdy $n \rightarrow \infty$, jeśli tylko x jest punktem Lebesgue'a. Wobec równości (2.16) pozostaje wykazać, że całka

$$I_n := \int_{\|y\| \geq \delta b_n} |\bar{f}(x-y) - \bar{f}(x)| b_n^{-r} |\bar{K}\left(\frac{y}{b_n}\right)| dy$$

może być uczyniona dowolnie małą przez wybór liczby δ dostatecznie dużej.

(1) $\|\bar{f}\|_{\infty} < \infty$. Wówczas

$$I_n \leq (\|\bar{f}\|_{\infty} + |\bar{f}(x)|) \int_{\|y\| \geq \delta} |\bar{K}(y)| dy$$

i całkę I_n można uczynić dowolnie małą.

(2) Jeżeli funkcja $\bar{K}$ ma zwarty nośnik, np. $K(t) = 0$ dla $\|t\| \geq \delta_0$, to wystarczy wziąć $\delta \geq \delta_0$, aby otrzymać $I_n = 0$. $\square$

Korzystając niżej z podanego lematu, będziemy zawsze przyjmować, że $\bar{K}$ jest albo jądrem K , albo kwadratem jądra, $\bar{K} = K^2$, oraz że $\bar{f} = f$. Z punktu widzenia naszych zastosowań za najciekawsze należy zatem uznać warunki (2), (4) i (5), w nich bowiem nie zakłada się nic o nieznanej gęstości f .

U w a g a. Od tej chwili, dla większej przejrzystości, będziemy rozważać jedynie przypadek jednowymiarowy, $r = 1$.

Podamy teraz prosty wniosek z podanego lematu.

WNIOSEK. Niech ξ będzie zmienną losową o gęstości f , K - gęstością istotnie ograniczoną i niech dla funkcji $\bar{f} = f$ i $\bar{K} = K$ zachodzi warunek $[D_x]$. Niech b_n będzie ciągiem liczb dodatnich, $\lim_{n \rightarrow \infty} b_n = 0$. Wówczas

$$\lim_{n \rightarrow \infty} E_f \frac{1}{b_n} K\left(\frac{x - \xi}{b_n}\right) = f(x)$$

oraz

$$\lim_{n \rightarrow \infty} b_n V_f \left[\frac{1}{b_n} K \left(\frac{x-f}{b_n} \right) \right] = f(x) \int K^2(y) dy.$$

2.8. Zastosujmy teraz wniosek 2.7 do ERP, EWWY, ED oraz ED*.

TWIERDZENIE ([66]). Niech K będzie gęstością, $\|K\|_\infty < \infty$, niech warunek $[D_x]$ zachodzi dla funkcji f i K oraz niech ciąg (b_n) spełnia związki (2.3). Wówczas

(i) Dla ERP, danego wzorem (2.1), mamy

$$(2.18) \quad \lim_{n \rightarrow \infty} E_{f_n} \hat{f}_n (\xi_1, \dots, \xi_n; x) = f(x)$$

oraz

$$(2.19) \quad \lim_{n \rightarrow \infty} nb_n \cdot V_{f_n} \hat{f}_n (\xi_1, \dots, \xi_n; x) = f(x) \int K^2(y) dy.$$

(ii) Dla ED, danego wzorem (2.12), prawdziwy jest związek (2.18) oraz jeżeli ponadto granica

$$(2.20) \quad \lim_{n \rightarrow \infty} \frac{nb_n}{\left[\sum_{i=1}^n b_i H(b_i) \right]^2} \sum_{k=1}^n b_k H^2(b_k) = \alpha, \quad 0 < \alpha < \infty,$$

istnieje, to

$$(2.21) \quad \lim_{n \rightarrow \infty} nb_n \cdot V_{f_n} \hat{f}_n (\xi_1, \dots, \xi_n; x) = \alpha f(x) \int K^2(y) dy.$$

(iii) Dla EWWY, danego wzorem (2.13), prawdziwy jest związek (2.18) oraz jeśli ponadto granica

$$(2.22) \quad \lim_{n \rightarrow \infty} (b_n/n) \sum_{k=1}^n b_k^{-1} = \alpha, \quad 0 < \alpha < \infty,$$

istnieje, to prawdziwy jest także związek (2.21).

(iv) Dla ED*, danego wzorem (2.14), prawdziwy jest związek (2.18) oraz jeśli ponadto granica

$$(2.23) \quad \lim_{n \rightarrow \infty} n b_n \left(\sum_{i=1}^n b_i \right)^{-1} = \alpha, \quad 0 < \alpha < \infty,$$

istnieje, to prawdziwy jest także związek (2.21).

(Oczywiście związki (2.22) i (2.23) są szczególnymi przypadkami związku (2.20)).

S z k i c d o w o d u. W przypadku ERP dowód jest trywialny - wystarczy zauważyć, że wówczas

$$E_f \hat{f}_n(\xi^n; x) = E_f \frac{1}{b_n} K\left(\frac{x - \xi_1}{b_n}\right)$$

oraz

$$V_f \hat{f}_n(\xi^n; x) = \frac{1}{n} V_f \left[\frac{1}{b_n} K\left(\frac{x - \xi_1}{b_n}\right) \right].$$

W pozostałych przypadkach dowód sprowadza się za każdym razem do skorzystania z lematu Toeplitza (p.np. [31]). W przypadku ERP i EWWY, w celu wykazania związku (2.18) wystarczy zamiast

warunku $nb_n \rightarrow \infty$ założyć $\lim_{n \rightarrow \infty} b_n = 0$. (Dowodząc prawdziwości związku (2.18) w przypadku ED i ED*, warunek $nb_n \rightarrow \infty$ można by zastąpić innymi, założenie jednak tego właśnie warunku jest najbardziej naturalne.) $\square$

Przy znacznie ostrzejszych założeniach teza (i) udowodniona została przez Parzena [33], teza (iii) - przez Yamato [69], tezy (ii) i (iv) natomiast - przez Deheuvelsa [15].

U w a g a. Często przyjmuje się szerokość pasma $b_n = \gamma \cdot n^{-\beta}$, $\gamma > 0$, $0 < \beta < 1$; przy takiej szerokości pasma $\alpha = 1/(1+\beta)$ w przypadku (iii) oraz $\alpha = 1 - \beta$ w przypadku (iv). Zauważmy też, że na mocy nierówności Schwartza

$$\left(\sum_{i=1}^n b_i\right)^{-1} \leq \frac{\sum_{i=1}^n b_i H^2(b_i)}{\left[\sum_{i=1}^n b_i H(b_i)\right]^2}$$

i stąd funkcja $H \equiv 1$ daje najmniejszą wariancję asymptotyczną w klasie wszystkich estymatorów Deheuvelsa (2.12). W rezultacie wariancje badanych estymatorów spełniają asymptotycznie następujące nierówności:

$$V(iv) \leq V(iii) \leq V(i) \quad \text{oraz} \quad V(iv) \leq V(ii).$$

Banon [1] skonstruował (stosunkowo skomplikowany) estymator rekurencyjny o wariancji asymptotycznej mniejszej od wariancji ED* (2.14). (Jak to już zaznaczaliśmy, dalej zajmować się będziemy badaniem BSK, uwzględniającego także wielkość obciążenia.)

2.9. Przejdziemy teraz do problemu punktowej zgodności estymatorów jądrowych. Ponieważ w rozdziale 4 podamy dokładną prędkość zbieżności (twierdzenie 4.1), formułując poniższe twierdzenie o zgodności, ograniczymy się dla uproszczenia do przypadku EWWY.

TWIERDZENIE ([17]). Niech K będzie gęstością, $\|K\|_\infty < \infty$, niech warunek $[D_x]$ zachodzi dla funkcji f i K oraz niech ciąg (b_n) spełnia związki (2.3). Niech $\hat{f}_n$ oznacza EWWY. Wówczas

$$E_f[\hat{f}_n(\xi_1, \dots, \xi_n; x) - f(x)]^2 \longrightarrow 0$$

i stąd

$$\hat{f}_n(\xi_1, \dots, \xi_n; x) \xrightarrow{P} f(x).$$

Jeżeli ponadto

$$(2.24) \quad \lim_{n \rightarrow \infty} \frac{nb_n}{\log \log n} = \infty,$$

to

$$P\left[\lim_{n \rightarrow \infty} \hat{f}_n(\xi_1, \dots, \xi_n; x) = f(x)\right] = 1.$$

U w a g a (p. [17]). Dla średniokwadratowej zgodności EWWY wystarcza zamiast warunku $nb_n \longrightarrow \infty$ przyjąć warunek

słabszy $n^{-2} \sum_{i=1}^n b_i^{-1} \rightarrow 0$ (por. podany niżej dowód twierdzenia);

w przypadku mocnej zgodności EWWY warunek (2.24) można zastąpić innym, mianowicie

$$(2.25) \quad \sum_{n=1}^{\infty} n^{-2} b_n^{-1} < \infty$$

(żaden z warunków (2.24) i (2.25) nie implikuje drugiego).

D o w ó d t w i e r d z e n i a. Oczywiście

$$\begin{aligned} E_f [\hat{f}_n(\xi^n; x) - f(x)]^2 &= \\ &= V_f \hat{f}_n(\xi^n; x) + [E_f \hat{f}_n(\xi^n; x) - f(x)]^2. \end{aligned}$$

Drugi ze składników prawej strony powyższej równości dąży do zera na mocy twierdzenia 2.8 (warunek $nb_n \rightarrow \infty$ można przy tym zastąpić słabszym, $b_n \rightarrow 0$; p. szkic dowodu tego twierdzenia). Dalej, stosując nierówność Czebyszewa i oznaczając $n_i = \frac{1}{b_i} K\left(\frac{x - \xi_i}{b_i}\right)$, otrzymujemy

$$(2.26) \quad P[|\hat{f}_n(\xi^n, x) - E_f \hat{f}_n(\xi^n, x)| \geq \varepsilon] \leq$$

$$\leq \frac{1}{\varepsilon^2} E_f [\hat{f}_n(\xi^n, x) - E_f \hat{f}_n(\xi^n, x)]^2 =$$

$$= \frac{1}{\varepsilon^2 n^2} E_f \sum_{i=1}^n (\eta_i - E \eta_i)^2 =$$

$$= \frac{1}{\varepsilon^2 n^2} \sum_{i=1}^n (E_f \eta_i^2 - [E_f \eta_i]^2) \leq \frac{1}{\varepsilon^2 n^2} \sum_{i=1}^n E_f \eta_i^2.$$

Ale

$$\begin{aligned} E_f \eta_i^2 &= E_f \left(\frac{1}{b_i} K \left(\frac{x - \xi_i}{b_i} \right) \right)^2 \leq \frac{1}{b_i} \|K\|_{\infty} E_f \frac{1}{b_i} K \left(\frac{x - \xi_i}{b_i} \right) = \\ &= \frac{1}{b_i} \|K\|_{\infty} E_f \eta_i. \end{aligned}$$

Ponieważ $E_f \eta_i \rightarrow f(x)$, więc istnieje taka stała skończona c , że dla każdego i

$$|E_f \eta_i| \leq c.$$

Zatem, wracając do (2.26),

$$\frac{1}{\varepsilon^2 n^2} \sum_{i=1}^n E_f \eta_i^2 \leq \frac{1}{\varepsilon^2 n^2} c \|K\|_{\infty} \sum_{i=1}^n b_i^{-1}$$

i na mocy założenia (p. uwaga wyżej) prawa strona tej nierówności dąży do zera, co dowodzi zgodności średniokwadratowej i słabej.

Dowód mocnej zgodności przy warunku dodatkowym (2.24) jest nieco złożony i opiera się na nierówności Bennetta [2] (oraz oczywiście na lemacie 2.7). Dowód ten bardzo się upraszcza, gdy założyć warunek dodatkowy (2.25). Zauważmy mianowicie, że ponieważ $E_{\hat{f}_n}(\xi^n; x) \rightarrow f(x)$, w celu wykazania

punktowej mocnej zgodności EWWY, wystarczy wykazać, iż

$$\hat{f}_n(\xi^n; x) - E_f \hat{f}_n(\xi^n, x) = \frac{1}{n} \sum_{i=1}^n (\eta_i - E_f \eta_i) \longrightarrow 0 \quad \text{p.n.}$$

Ale, przy założeniu (2.25) fakt ten wynika natychmiast z mocnego prawa wielkich liczb Kołmogorowa (p.n.p. [8], str. 121, [31], str 253). $\square$

2.10. Przy identycznych założeniach o funkcjach f i K oraz przy odpowiednich założeniach o ciągu (b_n) można wykazać słabą i mocną zgodność punktową ERP (2.1) i $ED^*(2.14)$ (p., odpowiednio, [18] i [17]). Słabą zgodność ERP uzyskuje się przy założeniu (2.3), mocną - przy założeniu dodatkowym

$$(2.27) \quad \sum_{n=1}^{\infty} \exp(-\alpha n b_n) < \infty \quad \text{dla każdego} \quad \alpha > 0.$$

Mocna zgodność punktowa ED^* zachodzi przy warunku $\sum_{n=1}^{\infty} b_n = \infty$.

U w a g a . Przy założeniu dodatkowym

$$(2.28) \quad \lim_{n \rightarrow \infty} \frac{n b_n}{\log n} = \infty,$$

mocniejszym (w ogólności) od założenia (2.27), oraz pod warunkiem $f(x) > 0$, mocna zgodność punktowa ERP jest prostym wnioskiem z lematu Teichera, stanowiącego pewną wersję mocnego prawa wielkich liczb dla tablic zmiennych losowych wierszowo niezależnych ([54], twierdzenie 1; w mianowniku wzoru (1)

tego twierdzenia pominięty jest czynnik $\log n$).

(Warunki (2.27) i (2.28) są oczywiście równoważne, jeżeli szerokość pasma $b_n = \gamma n^{-\beta}$, $\gamma > 0$, $0 < \beta < 1$, $n = 1, 2, \dots$)

2.11. Omówione wyżej twierdzenia o zgodności estymatorów jądrowych mają charakter lokalny - dotyczą zgodności punktowej. Podamy teraz twierdzenie o zgodności w normie przestrzeni funkcji całkowalnych, L_1 (inne istniejące rezultaty dotyczące "globalnej zgodności" estymatorów jądrowych wymienimy w rozdziale 6). Poniższe twierdzenie jest natychmiastowym wnioskiem z lematu Glicka [24], będącego uogólnieniem twierdzenia Scheffégo o zbieżności [40] i orzekającego, że jeżeli estymator $\hat{f}_n(x^n, \cdot)$ jest gęstością i jest funkcją mierzalną swoich argumentów $x_1, \dots, x_n, x$, to z jego zgodności punktowej (słabej lub mocnej) dla prawie wszystkich x wynika zgodność (odpowiednio słaba lub mocna) w normie przestrzeni L_1 .

TWIERDZENIE ([17], [18]). Niech K będzie gęstością,

$\|K\|_\infty < \infty$, niech warunek $[D_x]$ zachodzi dla funkcji f i K w prawie wszystkich punktach $x \in R^{(1)}$ oraz niech ciąg (b_n) będzie dodatni i $b_n \rightarrow 0$. Wówczas z własności

$$\hat{f}_n(\xi^n; x) \xrightarrow{P} f(x)$$

wynika własność

⁽¹⁾ Przypomnijmy, że wobec całkowalności f prawie wszystkie punkty $x \in R$ są punktami Lebesgue'a tej funkcji.

$$\int |\hat{f}_n(\xi^n; x) - f(x)| dx \xrightarrow{P} 0$$

oraz z własności

$$P\left[\lim_{n \rightarrow \infty} \hat{f}_n(\xi^n; x) = f(x)\right] = 1$$

wynika

$$P\left[\lim_{n \rightarrow \infty} \int |\hat{f}_n(\xi^n, x) - f(x)| dx = 0\right] = 1,$$

gdzie $\hat{f}_n$ oznacza dowolny z estymatorów (2.1), (2.12), (2.13), (2.14).

U w a g a . Zastosowanie do globalnej oceny estymatora gęstości funkcji straty określonej metryką przestrzeni L_1 jest pomysłem bardzo naturalnym: jeżeli mianowicie P i Q są miarami probabilistycznymi (na prostej), funkcje f i g są zaś ich gęstościami (względem miary Lebesgue'a), to

$$\sup_A |P(A) - Q(A)| = \frac{1}{2} \int |f(x) - g(x)| dx,$$

gdzie kres górny brany jest po wszystkich zbiorach borelowskich A . Mocną zgodność w metryce L_1 możemy zatem zinterpretować w następujący sposób: jeżeli P_f oznacza prawdziwy rozkład prawdopodobieństwa zmiennych losowych $\xi_1, \xi_2, \dots$, natomiast $\hat{f}_n(\xi^n; \cdot)$ oznacza estymator rozkładu P_f ,

$$P_{\hat{f}_n(\xi^n; \cdot)}^{(A)} := \int_A \hat{f}_n(\xi^n; x) dx,$$

to

$$P \left[\lim_{n \rightarrow \infty} \sup_A |P_{\hat{f}_n(\xi^n; \cdot)}^{(A)} - P_f^{(A)}| = 0 \right] = 1.$$

2.12. Estymator rekurencyjny oparty na rozwinięciach ortogonalnych. Estymacja gęstości oparta na rozwinięciach ortogonalnych funkcji została zaproponowana przez Čencova [71]. Rutkowski [39] zaproponował pewną rekurencyjną wersję tego typu estymatorów. Niech $f \in L_2$ i (φ_j) będzie bazą ortonormalną w L_2 , skąd

$$f(x) = \sum_{j=0}^{\infty} a_j \varphi_j(x), \quad \text{przy czym } a_j = E_f \varphi_j(\xi).$$

Estymator rekurencyjny zaproponowany przez Rutkowskiego ma postać

$$\hat{f}_n(x_1, \dots, x_n; x) = \frac{1}{n} \sum_{i=1}^n \sum_{j=0}^{k(i)} \varphi_j(x_i) \varphi_j(x),$$

gdzie $(k(i))$ jest ciągiem liczb naturalnych, $\lim_{i \rightarrow \infty} k(i) = \infty$. Jeżeli $|\varphi_j(x)| \leq c_j$ dla wszystkich punktów x i każdego indeksu j oraz

$$\sum_{n=1}^{\infty} n^{-2} \left[\sum_{j=0}^{k(n)} c_j^2 \right]^2 < \infty,$$

to ciąg powyższych estymatorów jest mocno zgodny.

3. BŁĄD ŚREDNIOKWADRATOWY. ANALIZA ASYMPTOTYCZNA

3.1. Bardziej szczegółowa analiza asymptotycznych własności estymatorów rozważanych w poprzednim rozdziale wymaga poczynienia mocniejszych założeń o gęstości f . Podobnie jak w poprzednim rozdziale, dla porównania z estymatorami rekurencyjnymi, niżej analizujemy również ERP.

W dalszym ciągu tego rozdziału będziemy zakładać jeden z następujących zbiorów warunków:

(3.1) Gęstość f jest różniczkowalna, pochodna f' jest istotnie ograniczona, funkcje K oraz $z \mapsto zK(z)$ są całkowalne, $\int zK(z)dz = 0$.

(3.2) Gęstość f jest dwukrotnie różniczkowalna, druga pochodna f'' jest istotnie ograniczona, funkcje K oraz $z \mapsto z^2K(z)$ są całkowalne, funkcja K jest parzysta.

U w a g a. Jak należy oczekiwać, zamieszczone niżej twierdzenia pozostają prawdziwe przy zastąpieniu w (3.2) założenia o parzystości jądra K wymaganiem, aby zbiór $\{x \in \mathbb{R}: K(x) \neq K(-x)\}$ był miary Lebesgue'a zero.

3.2. Poniższy lemat będzie pełnił dalej podobną rolę do tej, jaką w rozdziale 2 pełnił lemat 2.7.

LEMAT ([66]). (1) Przyjmijmy założenia (3.1) oraz $b_k > 0$,

$\lim_{k \rightarrow \infty} b_k = 0$. Wówczas

$$(3.3) \quad E_f \left[\frac{1}{b_k} K \left(\frac{x-\xi}{b_k} \right) \right] - f(x) \int K(z) dz = o(b_k),$$

gdy $k \rightarrow \infty$, oraz, jeżeli ponadto $\|K\|_\infty < \infty$,

$$(3.4) \quad b_k V_f \left[\frac{1}{b_k} K \left(\frac{x-\xi}{b_k} \right) \right] - f(x) \int K^2(z) dz + \\ + b_k f^2(x) \left[\int K(z) dz \right]^2 = o(b_k).$$

(Jeżeli pominąć warunek $\int zK(z) dz = 0$, prawa strona oszacowania (3.3) przyjmuje postać

$$b_k \|f'\|_\infty \int |zK(z)| dz = o(b_k),$$

prawa strona oszacowania (3.4) natomiast

$$b_k \|f'\|_\infty \int |zK^2(z)| dz + o(b_k^2).$$

(2) Przyjmijmy założenia (3.2) oraz $b_k > 0$,

$\lim_{k \rightarrow \infty} b_k = 0$. Wówczas

$$(3.5) \quad E_f \left[\frac{1}{b_k} K \left(\frac{x-\xi}{b_k} \right) \right] - f(x) \int K(z) dz = \\ = \frac{b_k^2}{2} f''(x) \int z^2 K(z) dz + o(b_k^2)$$

oraz, jeżeli ponadto $\|K\|_{\infty} < \infty$,

$$\begin{aligned} (3.6) \quad b_k V_f \left[\frac{1}{b_k} K \left(\frac{x-z}{b_k} \right) \right] &= f(x) \int K^2(z) dz + b_k f^2(x) \left[\int K(z) dz \right]^2 = \\ &= \frac{b_k^2}{2} f''(x) \int z^2 K^2(z) dz + o(b_k^2). \end{aligned}$$

D o w ó d. (1) Wykażemy prawdziwość oszacowania (3.3).

Oczywiście

$$\int \frac{1}{b_k} K \left(\frac{x-z}{b_k} \right) f(z) dz = \int K(y) f(x - b_k y) dy.$$

Nas interesuje zatem wielkość

$$(3.7) \quad \left| \int K(y) f(x - b_k y) dy - \int K(y) f(x) dy \right|.$$

Wobec różniczkowalności funkcji f ,

$$f(x+t) - f(x) = t f'(x) + t \varepsilon_x(t),$$

gdzie $\lim_{t \rightarrow 0} \varepsilon_x(t) = 0$ i ε_x jest pewną funkcją, której postać zależy od x . Wielkość (3.7) jest zatem równa następującej w poniższych obliczeniach korzystamy z założenia (3.1)):

$$\begin{aligned} (3.8) \quad & \left| \int K(y) [-b_k y f'(x) - b_k y \varepsilon_x(-b_k y)] dy \right| = \\ & = \left| b_k \int K(y) y \varepsilon_x(-b_k y) dy \right| \leq b_k \int |K(y) y| \cdot |\varepsilon_x(-b_k y)| dy. \end{aligned}$$

Pozostaje wykazać, że prawa strona nierówności (3.8) jest rzędu $o(b_k)$. Zauważmy w tym celu, że

$$\begin{aligned} |g_x(t)| &= \left| \frac{f(x+t) - f(x)}{t} - f'(x) \right| = \\ &= |f'(x + v_{x,t} \cdot t) - f'(x)| \leq 2 \|f'\|_\infty < \infty, \end{aligned}$$

skąd funkcja podcałkowa

$$|yK(y)| \cdot |g_x(-b_k y)|$$

jest ograniczona przez funkcję całkowalną

$$|yK(y)| \cdot 2 \|f'\|_\infty$$

i możemy przejść do granicy pod znakiem całki (twierdzenie Lebesgue'a o zbieżności majoryzowanej):

$$\begin{aligned} \lim_{n \rightarrow \infty} \int |yK(y)| \cdot |g_x(-b_k y)| dy &= \\ = \int \lim_{n \rightarrow \infty} [|yK(y)| \cdot |g_x(-b_k y)|] dy &= 0. \end{aligned}$$

W ten sposób wzór (3.3) został udowodniony. Dowód prawdziwości oszacowania (3.4) przebiega zupełnie analogicznie i przeto go pomijamy.

(2). Podobnie jak to uczyniliśmy wyżej, udowodnimy tylko pierwszą tezę części (2) lematu, tzn. wzór (3.5). Wiadomo, że

$$f''(x) = \lim_{h \downarrow 0} \frac{f(x+h) + f(x-h) - 2f(x)}{h^2}.$$

Stąd

$$(3.9) \quad f(x+h) + f(x-h) = 2f(x) + h^2 f''(x) + h^2 g_x(h),$$

gdzie $\lim_{h \rightarrow 0} g_x(h) = 0$. Przy tym, ponieważ

$$f(x \pm h) = f(x) \pm hf'(x) + \frac{h^2}{2} f''(x \pm v_x(\pm h)),$$

otrzymujemy, że $|g_x(h)| \leq 2 \|f''\|_\infty$.

Korzystając z założenia parzystości jądra K oraz całkując przez podstawienie, otrzymujemy

$$\begin{aligned} (3.10) \quad & \frac{1}{b_k} \int K\left(\frac{x-z}{b_k}\right) f(z) dz = \\ & = \frac{1}{2} \left[\int \frac{1}{b_k} K\left(\frac{x-z}{b_k}\right) f(z) dz + \int \frac{1}{b_k} K\left(\frac{z-x}{b_k}\right) f(z) dz \right] = \\ & = \frac{1}{2} \int K(y) [f(x - b_k y) + f(x + b_k y)] dy. \end{aligned}$$

Wobec własności (3.9) prawa strona równości (3.10) równa się

$$f(x) \int K(y) dy + \frac{f''(x)}{2} b_k^2 \int y^2 K(y) dy + \frac{1}{2} \int K(y) b_k^2 y^2 g_x(b_k y) dy$$

i, jak poprzednio, dla zakończenia dowodu wzoru (3.5) wystarczy do ostatniej całki powyższego wyrażenia zastosować twier-

dzenie Lebesgue'a. $\square$

U w a g a. Powyższy lemat pozwoli nam oszacować prędkość zbieżności wartości oczekiwanej oraz wariancji estymatorów f_n do ich wartości asymptotycznych. Warto tu odnotować, że rząd prędkości zbieżności wartości oczekiwanej dowolnego ustalonego estymatora jądrowego $\hat{f}_n(\xi^n, x)$ nie polepszy się, jeżeli gęstość f będzie różniczkowalna więcej razy niż dwukrotnie, ale jądro K będzie nieujemne; fakt ten wynika natychmiast z oszacowania (3.5). (Przyspieszenie rzędu prędkości zbieżności można wówczas uzyskać bądź konstruując jądra przyjmujące wartości obydwu znaków, bądź odpowiednio randomizując szerokość pasma b_n ; problemami tymi nie będziemy się jednak zajmować.)

3.3. W dalszym ciągu pracy potrzebny nam będzie następujący lemat:

LEMAT ([66]). Niech jądro $K > 0$, $\|K\|_\infty < \infty$, punkty $x, y \in \mathbb{R}$, $x \neq y$, $b_k \rightarrow 0$ oraz niech będzie spełniony jeden z następujących warunków:

- (1) gęstość f jest istotnie ograniczona;
- (2) $\lim_{|z| \rightarrow \infty} zK(z) = 0$;
- (3) gęstość $f \in L_2$, punkty x, y są punktami Lebesgue'a gęstości f

$$\int_0^\infty \sup_{0 < |u| \leq z} K(u) dz < \infty.$$

Wówczas

$$\lim_{k \rightarrow \infty} E_f \left[\frac{1}{b_k} K\left(\frac{x-\xi}{b_k}\right) K\left(\frac{y-\xi}{b_k}\right) \right] = 0.$$

Dowód pomijamy (p.[66]). Jego idea pokrywa się z ideą dowodu lematu 2.7.

3.4. U w a g a. Od tej chwili będziemy zawsze zakładać, że ciąg (b_n) spełnia warunki (2.3), jądro K natomiast - warunki (2.2) oraz $\|K\|_\infty < \infty$.

TWIERDZENIE (Asymptotyczna nieobciążoność). (1) Przyjmujemy założenia (3.1). Wówczas

$$E_f \hat{f}_n(\xi^n; x) - f(x) = o(b_n) \quad \text{dla ERP;}$$

$$= o\left(\frac{\sum_{i=1}^n b_i^2 H(b_i)}{\sum_{i=1}^n b_i H(b_i)}\right), \quad \text{jeżeli ponadto } \sum_{k=1}^{\infty} b_k^2 H(b_k) = \infty,$$

dla ED;

$$= o\left(\sum_{i=1}^n b_i / n\right), \quad \text{jeżeli ponadto } \sum_{k=1}^{\infty} b_k = \infty, \quad \text{dla EWWY.}$$

(Jeżeli pominąć warunek $\int zK(z)dz = 0$, małe wyższego rzędu ("o") należy zastąpić małymi tego samego rzędu ("O"); nakładanie warunków dodatkowych na ciąg (b_n) nie jest wówczas potrzebne.)

(2) Przyjmujemy założenia (3.2). Wówczas

$$E_f \hat{f}_n(\xi^n; x) - f(x) =$$

$$= \frac{f''(x)}{2} \cdot \int z^2 K(z) dz \cdot b_n^2 + o(b_n^2) \quad \text{dla ERP};$$

$$= \frac{f''(x)}{2} \cdot \int z^2 K(z) dz \cdot \frac{\sum_{i=1}^n b_i^3 H(b_i)}{\sum_{i=1}^n b_i H(b_i)} \cdot [1 + o(1)],$$

$$\text{jeżeli ponadto } \sum_{k=1}^{\infty} b_k^3 H(b_k) = \infty, \quad \text{dla ED};$$

$$= \frac{f''(x)}{2} \cdot \int z^2 K(z) dz \cdot \frac{\sum_{i=1}^n b_i^2}{n} \cdot [1 + o(1)], \quad \text{jeżeli po-}$$

nadto

$$\sum_{k=1}^{\infty} b_k^2 = \infty, \quad \text{dla EWWY}.$$

D o w ó d jest elementarny. Przykładowo, ażeby uczynić oczywistym sens warunków dodatkowych nakładanych na ciąg (b_n) , rozważmy krótko problem obciążenia EWWY przy założeniu (3.1). Na mocy oszacowania (3.3)

$$E_f \hat{f}_n(\xi^n; x) = \frac{1}{n} \sum_{i=1}^n E_f \left[\frac{1}{b_i} K\left(\frac{x - \xi_i}{b_i}\right) \right] = f(x) + \frac{1}{n} \sum_{i=1}^n \psi_i,$$

gdzie ciąg (ψ_i) spełnia warunek $\psi_i = o(b_i)$, $i \rightarrow \infty$.

Założywszy, że $\sum_{k=1}^{\infty} b_k = \infty$, zastosowanie np. twierdzenia Stolza daje natychmiast

$$\lim_{n \rightarrow \infty} \frac{\sum_{i=1}^n y_i}{\sum_{i=1}^n b_i} = 0,$$

co dowodzi podanej tezy. Oczywiście, jeżeli $\sum_{k=1}^{\infty} b_k < \infty$, badane obciążenie jest rzędu $o(n^{-1})$. $\square$

Dla uproszczenia zapisu w występujących dalej wyrażeniach będziemy używać symbolu " $\sim$ ", oznaczającego "równość asymptotyczną", pomijając małe wyższego rzędu - zamiast pisać $A_n = c \cdot a_n + o(a_n)$, będziemy pisać $A_n \sim c \cdot a_n$.

3.5.

TWIERDZENIE (Asymptotyczna wariancja). (1) Przyjmijmy założenia (3.1). Wówczas

$$\begin{aligned} V_{f\hat{f}_n}(\xi^n; x) &\sim \frac{1}{nb_n} \cdot f(x) \cdot \int K^2(z) dz - \frac{1}{n} f^2(x) \quad \text{dla ERP;} \\ &\sim \frac{\sum_{i=1}^n b_i H^2(b_i)}{\left[\sum_{i=1}^n b_i H(b_i) \right]^2} : f(x) \cdot \int K^2(z) dz - \frac{\sum_{i=1}^n b_i^2 H^2(b_i)}{\left[\sum_{i=1}^n b_i H(b_i) \right]^2} \cdot f^2(x), \end{aligned}$$

jeżeli ponadto $\sum_{k=1}^{\infty} b_k^2 H^2(b_k) = \infty$, dla ED;

$$\sim \frac{1}{n^2} \cdot \sum_{i=1}^n \frac{1}{b_i} \cdot f(x) \cdot \int K^2(z) dz - \frac{1}{n} \cdot f^2(x) \quad \text{dla EWWY.}$$

(2) Przyjmijmy założenia (3.2). Wówczas

$$V_{f\hat{f}_n}(\xi^n; x) \sim$$

$$\sim \frac{1}{nb_n} \cdot f(x) \cdot \int K^2(z) dz - \frac{1}{n} \cdot f^2(x) +$$

$$+ \frac{b_n}{2n} \cdot f''(x) \cdot \int z^2 K^2(z) dz \quad \text{dla ERP;}$$

$$\sim \left[\sum_{i=1}^n b_i H(b_i) \right]^{-2} \cdot \left[\sum_{k=1}^n b_k H^2(b_k) \cdot f(x) \cdot \int K^2(z) dz - \right.$$

$$\left. - \sum_{k=1}^n b_k^2 H^2(b_k) \cdot f^2(x) + \right.$$

$$\left. + \sum_{k=1}^n b_k^3 H^2(b_k) \cdot \frac{f''(x)}{2} \cdot \int z^2 K^2(z) dz \right], \text{ jeżeli ponadto}$$

$$\sum_{k=1}^{\infty} b_k^3 H^2(b_k) = \infty, \quad \text{dla ED;}$$

$$\sim \frac{1}{n^2} \sum_{i=1}^n \frac{1}{b_i} \cdot f(x) \cdot \int K^2(z) dz - \frac{1}{n} \cdot f^2(x) +$$

$$+ \frac{1}{n^2} \sum_{k=1}^n b_k \cdot \frac{f''(x)}{2} \cdot \int z^2 K^2(z) dz,$$

$$\text{jeżeli ponadto } \sum_{k=1}^{\infty} b_k = \infty, \quad \text{dla EWWY.}$$

Dowód jest oczywisty, przeto go pomijamy.

3.6. Połączenie twierdzeń 3.4 i 3.5 daje twierdzenie o asymptotycznej postaci BSK estymatorów ERP, ED i EWWY. Odpowiednie twierdzenie sformułujemy tylko dla przypadku zachodzenia założeń (3.2).

Oznaczmy

$$c_1 := \left[\frac{f''(x)}{2} \int z^2 K(z) dz \right]^2,$$

$$c_2 := f(x) \int K^2(z) dz.$$

TWIERDZENIE (BSK). Przyjmijmy założenia (3.2). Wówczas

$$BSK = E_f [\hat{f}_n(\xi^n; x) - f(x)]^2$$

$$\sim b_n^4 c_1 + (nb_n)^{-1} c_2 - \frac{1}{n} \cdot f^2(x) \quad \text{dla ERP;}$$

$$BSK \sim \left[\frac{\sum_{k=1}^n b_k^3 H(b_k)}{\sum_{k=1}^n b_k H(b_k)} \right]^2 \cdot c_1 +$$

$$+ \left[\sum_{i=1}^n b_i H(b_i) \right]^{-2} \cdot \left[\sum_{k=1}^n b_k H^2(b_k) \cdot c_2 - f^2(x) \cdot \sum_{k=1}^n b_k^2 H^2(b_k) \right],$$

jeżeli ponadto

$$\sum_{k=1}^{\infty} b_k^3 H^2(b_k) = \quad , \quad \sum_{k=1}^{\infty} b_k^3 H(b_k) = \infty , \quad \text{dla ED;}$$

$$BSK \sim \left(\sum_{k=1}^n b_k^2 / n \right)^2 \cdot c_1 + \frac{1}{n^2} \sum_{k=1}^n \frac{1}{b_k} \cdot c_2 - \frac{1}{n} f^2(x) ,$$

$$\text{jeżeli ponadto } \sum_{k=1}^{\infty} b_k^2 = \infty , \text{ dla EWWY.}$$

U w a g a. Przy innych założeniach rozwinięcia powyższego typu podane zostały przez Deheuvelsa [15].

3.7. Rozważymy obecnie problem asymptotycznie optymalnego doboru szerokości pasma b_n , funkcji H oraz jądra K estymatorów ERP oraz ED; rozważenie estymatorów typu ED oznacza oczywiście uwzględnienie ED^* i EWWY. Ograniczymy się do przypadku szerokości pasma o postaci $b_n = \gamma \cdot n^{-\beta}$, $\gamma > 0$, $0 > \beta > 1$, oraz funkcji H o postaci $H(b_n) = b_n^\rho = \gamma^\rho n^{-\beta\rho}$, $\rho \in \mathbb{R}$.

(1) Estymator Rosenblatt-Parzena (ERP). Mamy $BSK = c_1 \gamma^4 n^{-4\beta} + c_2 \gamma^{-1} n^{-(1-\beta)} + o(n^{-4\beta}) + o(n^{-(1-\beta)})$.

Optymalne współczynniki β i γ wynoszą odpowiednio (przy pominięciu małych wyższego rzędu)

$$\beta = 1/5, \quad \gamma = (c_2/4c_1)^{1/5}.$$

Wówczas

$$\begin{aligned} BSK &\sim c_1^{1/5} c_2^{4/5} \cdot 5 \cdot 2^{-8/5} n^{-4/5} = \\ &= 1,65938 c_1^{1/5} \cdot c_2^{4/5} \cdot n^{-4/5} \end{aligned}$$

(p.[38]; we wzorach Rosenblatt-Parzena (21) i (22) powinny być pominięte czynniki, odpowiednio, 4 i $2^{3/5}$). Pozostaje teraz przez odpowiedni dobór jądra K zminimalizować iloczyn $c_1^{1/5} \cdot c_2^{4/5}$. Ażeby problem uczynić dobrze określonym, weźmiemy pod uwagę

jedynie takie jądra, dla których $\int z^2 K(z) dz = 1$. Wówczas funkcjonował $K \mapsto \int K^2(z) dz$ osiąga minimum dla funkcji

$$K_0(z) = \begin{cases} \frac{3}{4\sqrt{5}} \left(1 - \frac{z^2}{5}\right), & |z| \leq \sqrt{5}, \\ 0, & z < -\sqrt{5} \text{ lub } z > \sqrt{5} \end{cases}$$

(p. [20], [38]). Problem minimalizacji BSK w przypadku ERP jest rozwiązany.

(2) Estymator Deheuvelsa (ED). W tym przypadku

$$\begin{aligned} \text{BSK} \sim c_1 \gamma^4 \left[\frac{1-\beta(1+\varphi)}{1-\beta(3+\varphi)} \right]^2 \cdot n^{-4\beta} + \\ + c_2 \gamma^{-1} \left[\frac{1-\beta(1+\varphi)}{1-\beta(1+2\varphi)} \right]^2 \cdot n^{-1+\beta}. \end{aligned}$$

Optymalne współczynniki β i γ wynoszą odpowiednio

$$\beta = 1/5, \quad \gamma = \left(\frac{c_2}{c_1} \cdot \frac{2-\varphi}{40} \right)^{1/5}.$$

Wówczas $\text{BSK} \sim 5 \cdot 40^{-4/5} \frac{(4-\varphi)^2}{(2-\varphi)^{6/5}} c_1^{1/5} \cdot c_2^{4/5} \cdot n^{-4/5}$. Optymal-

na wartość φ wynosi $\varphi = -1$; jak z tego wynika, EWWY jest najlepszym estymatorem w klasie estymatorów ED ze względu na BSK (por. [16]). Jądro optymalne ma znowu postać funkcji K_0 .

Reasumując, optymalne wartości BSK estymatorów jądrowych wynoszą $\Delta \cdot c_1^{1/5} \cdot c_2^{4/5} \cdot n^{-4/5}$, gdzie

$$\Delta = \begin{cases} 1,66 & \text{dla ERP,} \\ 1,75 & \text{dla EWWY,} \\ 1,82 & \text{dla ED}^* \text{ (estymatora Deheuvelsa z } H \equiv 1). \end{cases}$$

Nie powinno nas dziwić, że najlepszy wynik daje ERP. Z drugiej strony, wyższość EWWY nad ED^* nie przeczy uwadze 2.8 - pomimo że wariancja ED^* jest asymptotycznie mniejsza od wariancji EWWY, asymptotyczna wartość BSK tego drugiego estymatora jest mniejsza.

U w a g a. Deheuvels [16] rozważał inne postaci szerokości pasma niż $b_n = \gamma n^{-\beta}$ i podał interesujące uwagi o ich wpływie na wartość BSK. My zauważymy tu tylko, że przyjmując $b_n = \gamma n^{-\beta} / \log n$ otrzymuje się znowu optymalną wartość $\beta = 1/5$, ale tym razem (asymptotyczna) wartość BSK wyznaczona jest jedynie przez wariancję estymatora. I, niejako odwrotnie, przyjmując $b_n = \gamma n^{-\beta} \cdot \log n$, optymalna wartość $\beta = 1/5$, ale wartość BSK zdeterminowana jest przez obciążenie estymatora.

4. PRAWO ITEROWANEGO LOGARYTMU

4.1. Podamy teraz prawo iterowanego logarytmu dla ED (i stąd, w szczególności, dla EWWY; prawo iterowanego logarytmu dla ERP podał Hall [27] -p. paragraf 4.4). Jego dowodu nie będziemy podawać, otrzymuje się go bowiem drogą elementarnych (choć dość znużających) rachunków z prawa iterowanego logarytmu Kołmogorowa (dla zmiennych losowych [niejednakowo!] ograniczonych; p.np. [8], str. 343, [34], str. 358).

TWIERDZENIE ([66]). Niech $f(x) > 0$. Niech będzie spełniony warunek (2.20), granica

$$(4.1) \quad \lim_{n \rightarrow \infty} \frac{\log \log \left\{ (nb_n)^{-1} \left[\sum_{k=1}^n b_k H(b_k) \right]^2 \right\}}{\log \log n} = \delta$$

istnieje, $0 < \delta < \infty$, oraz

$$(4.2) \quad \lim_{n \rightarrow \infty} nb_n \left[\sum_{k=1}^n b_k H(b_k) \right]^{-2} \cdot H^2(b_n) \cdot \log \log \frac{\left[\sum_{k=1}^n b_k H(b_k) \right]^2}{nb_n} = 0.$$

Wówczas dla ED mamy

$$P \left[\limsup_{n \rightarrow \infty} \left(\frac{nb_n}{\log \log n} \right)^{1/2} \cdot \frac{\hat{f}_n(\xi^n; x) - E_f \hat{f}_n(\xi^n; x)}{(2\delta f(x) \cdot \int K^2(z) dz)^{1/2}} = 1 \right] = 1.$$

Jeżeli ponadto:

(i) spełnione są założenia (3.1) oraz ciąg

$$(4.3) \quad \left(\frac{nb_n}{\log \log n} \right)^{1/2} \cdot \frac{\sum_{k=1}^n b_k^2 H(b_k)}{\sum_{k=1}^n b_k H(b_k)}$$

jest ograniczony i $\sum_{k=1}^{\infty} b_k^2 H(b_k) = \infty$; lub

(ii) spełnione są założenia (3.2) oraz

$$(4.4) \quad \lim_{n \rightarrow \infty} \left(\frac{nb_n}{\log \log n} \right)^{1/2} \cdot \frac{\sum_{k=1}^n b_k^3 H(b_k)}{\sum_{k=1}^n b_k H(b_k)} = 0;$$

to dla ED mamy

$$(4.5) \quad P \left[\limsup_{n \rightarrow \infty} \left(\frac{nb_n}{\log \log n} \right)^{1/2} \cdot \frac{\hat{f}_n(\xi_n^x; x) - f(x)}{(2\alpha \delta f(x) \int K^2(z) dz)^{1/2}} = 1 \right] = 1$$

Dla EWWY poszczególne założenia przyjmują następującą postać: warunek (2.20) przyjmuje postać (2.22), dalej zaś

$$(4.1') \quad \text{warunek (4.1):} \quad \lim_{n \rightarrow \infty} \frac{\log \log (n/b_n)}{\log \log n} = \delta \quad (0 < \delta < \infty),$$

$$(4.2') \quad \text{warunek (4.2):} \quad \lim_{n \rightarrow \infty} (nb_n)^{-1} \log \log (n/b_n) = 0;$$

$$(4.3') \quad \text{warunek (4.3):} \quad \text{ciąg} \left(\frac{b_n}{n \log \log n} \right)^{1/2} \cdot \sum_{k=1}^n b_k$$

jest ograniczony i $\sum_{k=1}^{\infty} b_k = \infty$;

$$(4.4') \quad \text{warunek (4.4):} \quad \lim_{n \rightarrow \infty} \left(\frac{b_n}{n \log \log n} \right) \cdot \sum_{k=1}^n b_k^2 = 0.$$

Decydując się w przypadku ED na wybór takich szerokości pasma b_n i funkcji H jak w paragrafie 3.7 otrzymuje się

z ogólnych warunków (2.20) oraz (4.1) - (4.2) następujące:

$$\beta(1+2\rho) < 1, \quad \beta(1+\rho) < 1, \quad \beta(1-\rho) < 1$$

$$\text{oraz } \beta \geq 1/3$$

w miejsce warunku (4.3) i $\beta \geq 1/5$ w miejsce warunku (4.4).
W przypadku założenia warunków (3.1) i (4.3), optymalna wartość β wynosi $\beta = 1/3$, w przypadku zaś założenia warunków (3.2) i (4.4) optymalna wartość $\beta = 1/5$. We wszystkich tych przypadkach, tzn. gdy b_n i H są postaci takiej jak w paragrafie 3.7,

$$\sigma = 1 \quad \text{i} \quad \alpha = \frac{[1 - \beta(1+\rho)]^2}{1 - \beta(1+2\rho)}.$$

Stąd, związek (4.5) przyjmuje wówczas postać

$$(4.6) \quad P \left[\limsup_{n \rightarrow \infty} \left(\frac{f_n^{1-\beta}}{\log \log n} \right)^{1/2} \cdot \frac{\hat{f}_n(\xi_n^n; x) - f(x)}{\left\{ 2 \cdot \frac{[1 - \beta(1+\rho)]^2}{1 - \beta(1+2\rho)} \cdot f(x) \int K^2(z) dz \right\}^{1/2}} = 1 \right] = 1.$$

(Przy nieco innych założeniach podobne twierdzenie udowodnione jest w [15]).

Nижеj podamy pewne uwagi związane z powyższym twierdzeniem, mające charakter bądź uwag technicznych i informacyjnych, bądź dyskusji praw iterowanego logarytmu dla estymatorów gę-

tości (p. zwłaszcza paragraf 4.5).

4.2. Warunek $f(x) > 0$ ma zasadnicze znaczenie przy dokładnym (takim, jakie daje prawo iterowanego logarytmu) szacowaniu prędkości zbieżności. Łatwo zauważyć, że w przypadku gdy gęstość f znika w otoczeniu punktu x , uzyskiwana w tym punkcie asymptotyczna prędkość zbieżności jest dowolnie duża. Zarazem, twierdzenie 3.4 pozwala podać dokładną prędkość zbieżności, jeśli tylko $f''(x) \neq 0$.

4.3. dla EWWY podobny rezultat podany jest w pracy [62], przy pewnych dodatkowych warunkach technicznych nałożonych na ciąg (b_n) . Faktycznie jednak, metoda dowodu przyjęta w [62] - korzystająca z twierdzenia o mocnym przybliżeniu dla pewnego estymatora pomocniczego - daje wynik słabszy, bo tylko nierówność zamiast równości (4.6). Co więcej, szacując obciążenie $E_f \hat{f}_n - f$, autorzy omawianej pracy przyjmują pewne założenie dodatkowe o funkcji charakterystycznej gęstości f ; w istocie, przyjęte założenie implikuje istnienie i ograniczoność drugiej pochodnej f'' . (Odnotujmy tu jeszcze, że Wegman i Davies podają równość 4.5 dla granicy zwykłej, a nie górnej, co nie zawsze jest prawdą.)

4.4. Hall [27] podał prawo iterowanego logarytmu, z $\alpha = \tilde{\sigma} = 1$. dla ERP. Jego dowód nie jest elementarny, korzysta bowiem z mocnego przybliżenia ciągu procesów $n^{1/2}(F_n(\cdot) - F(\cdot))$, gdzie F_n jest dystrybuantą empiryczną

n niezależnych zmiennych losowych o dystrybuancie F , ciągiem mostów Browna. Autorom niniejszej pracy nie udało się udowodnić takiego prawa dla ERP w sposób elementarny.

4.5. Twierdzenie 4.1 opisuje dokładną prędkość zbieżności prawie na pewno. Przy założeniu (3.1) lub (3.2) optymalne wartości szerokości pasma b_n wynoszą odpowiednio $\gamma n^{-1/3}$ lub $\gamma n^{-1/5}$. Równość (4.6) nie zależy od przyjętej wartości współczynnika γ i stąd prędkość zbieżności jest tym większa, im większa jest wartość tego współczynnika. Dzieje się tak dlatego, że czynnik $(\log \log n)^{-1/2}$ czyni obciążenie estymatora asymptotycznie pomijalnym i, zgodnie z twierdzeniem 3.5, duże szerokości pasma b_n dają dużą prędkość zbieżności różnicy $\hat{f}_n - E_f \hat{f}_n$. Zauważmy, że optymalna wartość parametru β jest taka sama, jak otrzymana dla BSK (por. par. 3.7); uzasadnione byłoby zatem wybranie np. takiej wartości współczynnika γ , która minimalizuje asymptotyczną wartość BSK.

W przypadku gdy spełnione jest założenie (3.2), łatwo podać taki ciąg (b_n) , by asymptotyczne obciążenie estymatora było wprawdzie skończone, ale nie zerowe. Okazuje się, że sytuacja taka ma zaskakująco ciekawą interpretację. Rozważmy mianowicie EWWY (dla ERP analogiczny wynik uzyskuje się natychmiast). Niech

$$b_n = \gamma n^{-1/5} \log \log^{1/5} n.$$

Proste rachunki wykazują, że wówczas warunki (2.22), (4.1')

i (4.2') nadal są spełnione, dwa pierwsze z tymi samymi wartościami α i $\tilde{\sigma}$,

$$\alpha = \frac{1}{1+\beta} = \frac{5}{6}, \quad \tilde{\sigma} = 1,$$

warunek (4.4') przyjmuje natomiast postać

$$\lim_{n \rightarrow \infty} \left(\frac{b_n}{n \log \log n} \right)^{1/2} \cdot \sum_{k=1}^n b_k^2 = \frac{5}{3} \gamma^{5/2}.$$

Ten ostatni warunek odpowiedzialny jest za asymptotyczną wartość obciążenia i ostatecznie prawo iterowanego logarytmu przyjmuje postać (por. twierdzenia 3.4 i 4.1)

$$(4.7) \quad P \left[\limsup_{n \rightarrow \infty} \frac{\gamma^{1/2} n^{2/5}}{\log \log^{2/5} n} \cdot \frac{\hat{f}_n(\xi^n; x) - f(x)}{(5/3 \cdot f(x) \int K^2(z) dz)^{1/2}} = 1 + \frac{\gamma^{5/2} f''(x) \int z^2 K(z) dz}{-2(5/3 \cdot f(x) \int K^2(z) dz)^{1/2}} \right] = 1.$$

Założmy teraz, że gęstość f jest lokalnie ściśle wypukła w punkcie x (tzn. ściśle wypukła w pewnym otoczeniu x). Wówczas $f''(x) > 0$ i wartość górnej granicy występującej w związku (4.7) rośnie wraz ze wzrostem współczynnika szerokości pasma, γ . Odpowiednio, jeżeli gęstość f jest lokalnie ściśle wklęsła w punkcie x , wartość tej górnej granicy maleje wraz ze wzrostem współczynnika γ (począwszy od pewnej wartości współczynnika γ rozważana granica jest ujemna!). Oznacza to, iż duże wartości szerokości pasma prowadzą (asympt-

totocznie) do otrzymania oszacowań nieznanej gęstości nadmier-
nie wygładzonych - w sensie granicy górnej (4.7) oszacowania
 $\hat{f}_n(\xi^n; x)$ przekraczają wartość $f(x)$, gdy f jest wypukła, i
są mniejsze od $f(x)$, gdy gęstość f jest wklęsła. W ten spo-
sób prawo iterowanego logarytmu (4.6) daje ściśle (asympto-
tyczne) uzasadnienie intuicji podanej w paragrafie 2.6.

4.6. Kończąc ten rozdział, wspomnimy jeszcze tylko, że dla
estymatorów nierekurencyjnych znane są dokładne oszacowania
prędkości zbieżności zmiennych losowych

$\sup_{x \in J} (|\hat{f}_n(\xi^n; x) - f(x)| / \sqrt{f(x)})$, gdzie J jest przedziałem zawar-
tym (tzw. "prawa logarytmu"; p.np. [10], twierdzenie 6.2.5,
[50], [51], [13]).

5. ASYMPTOTYCZNA NORMALNOŚĆ

5.1.

TWIERDZENIE ([66]). Niech $f(x) > 0$. Wówczas dla ERP mamy

$$(5.1) \quad \lim_{n \rightarrow \infty} \sup_{y \in \mathbb{R}} |P \left\{ \frac{\sqrt{nb_n} [\hat{f}_n(\xi^n; x) - E_f \hat{f}_n(\xi^n; x)]}{f(x) \cdot K^2 \int dz} \leq y \right\} - \Phi(y)| = 0.$$

Jeżeli granica

$$(5.2) \quad \lim_{n \rightarrow \infty} (nb_n)^{1+\frac{\delta}{2}} \frac{\sum_{k=1}^n b_k H^{2+\delta}(b_k)}{\left[\sum_{k=1}^n b_k H(b_k) \right]^{2+\delta}} = 0 \quad \text{dla pewnego } \delta > 0$$

oraz spełniony jest związek (2.20), to teza (5.1) zachodzi
dla ED.

Jeżeli dla ERP spełnione jest dodatkowo założenie (3.1) oraz ciąg (nb_n^3) jest ograniczony lub spełnione jest dodatkowo założenie (3.2) oraz $\lim_{n \rightarrow \infty} nb_n^5 = 0$, to

$$\lim_{n \rightarrow \infty} \sup_{y \in \mathbb{R}} |P\{\sqrt{nb_n} \frac{\hat{f}_n(\xi^n; x) - f(x)}{[f(x) \cdot \int K^2(z) dz]^{1/2}} \leq y\} - \Phi(y)| = 0.$$

Jeżeli dla ED spełnione jest dodatkowo założenie (3.1) oraz ciąg

$$(5.3) \quad \sqrt{nb_n} \frac{\sum_{k=1}^n b_k^2 H(b_k)}{\sum_{k=1}^n b_k H(b_k)} \quad \text{jest ograniczony i} \quad \sum_{k=1}^{\infty} b_k^2 H(b_k) = \infty$$

lub spełnione jest dodatkowo założenie (3.2) oraz

$$(5.4) \quad \lim_{n \rightarrow \infty} \sqrt{nb_n} \frac{\sum_{k=1}^n b_k^3 H(b_k)}{\sum_{k=1}^n b_k H(b_k)} = 0,$$

to

$$\lim_{n \rightarrow \infty} \sup_{y \in \mathbb{R}} |P\{\sqrt{nb_n} \frac{\hat{f}_n(\xi^n; x) - f(x)}{[\alpha \cdot f(x) \cdot \int K^2(z) dz]^{1/2}} \leq y\} - \Phi(y)| = 0.$$

Dla EWWY warunek (5.2) równoważny jest następującemu

$$\lim_{n \rightarrow \infty} (b_n/n)^{1+\delta} \cdot \sum_{k=1}^n b_k^{-(1+\delta)} = 0 \quad \text{dla pewnego } \delta > 0,$$

warunek (5.3) równoważny jest warunkowi: ciąg

$$(b_n/n)^{1/2} \cdot \sum_{k=1}^n b_k \quad \text{jest ograniczony i} \quad \sum_{k=1}^{\infty} b_k = \infty,$$

wreszcie (5.4) równoważny jest warunkowi

$$(b_n/n)^{1/2} \cdot \sum_{k=1}^n b_k \longrightarrow 0.$$

Zasadniczo, dowód twierdzenia 5.1 sprowadza się do wykorzystania twierdzenia Berry'ego-Esseena.

5.2.

TWIERDZENIE ([66]). Jeżeli spełnione są założenia lematu

3.3, $x_1, \dots, x_k$ są różnymi punktami na prostej, przy czym $f(x_j) > 0$, $j = 1, \dots, k$, to dla ERP mamy

$$(\eta_1, \dots, \eta_k) \xrightarrow{D} N(0, I),$$

gdzie $N(0, I)$ oznacza zmienną losową o k -wymiarowym rozkładzie normalnym z zerowym wektorem wartości oczekiwanych i jednostkową macierzą kowariancji oraz

$$(5.5) \quad \eta_j := \frac{\sqrt{nb_n} [\hat{f}_n(\xi^n; x_j) - E_f \hat{f}_n(\xi^n; x_j)]}{(f(x_j) \int K^2(z) dz)^{1/2}}, \quad j = 1, \dots, k.$$

Jeżeli ponadto spełnione są warunki (2.20) oraz

$$\sum_{k=1}^{\infty} b_k H^2(b_k) = \infty, \quad \text{to dla ED mamy}$$

$$(\eta_1/\alpha, \dots, \eta_k/\alpha) \xrightarrow{D} N(0, I),$$

gdzie η_j dane są przez (5.5).

Dowód polega na zastosowaniu twierdzenia Crámera-Wolda i lematu 3.3.

Przy stosownych założeniach powyższe twierdzenie można by sformułować dla różnic $\hat{f}_n(\xi^n; x_j) - f(x_j)$.

6. INFORMACJA O INNYCH WYNIKACH

6.1. Rekurencyjny charakter EWWY sugeruje zastosowanie do analizy jego własności asymptotycznych twierdzeń o metodach aproksymacji stochastycznej działających w przestrzeniach Hilberta (p. np. [36], [41] - [43], [57]). Wykorzystując własności procesu Robbinsa-Monro, Schmetterer udowodnił następujące twierdzenie

TWIERDZENIE ([42], [43]). Niech f i K będą gęstościami całkowalnymi w kwadracie, ciąg $b_n \downarrow 0$. Niech $\|\cdot\|$ oznacza normę w przestrzeni L_2 . Wówczas

$$(6.1) \quad P_f \left[\lim_{n \rightarrow \infty} \|\hat{f}_n(\xi^n; \cdot) - f(\cdot)\| = 0 \right] = 1$$

oraz dla każdego $\varepsilon > 0$ istnieje taka stała $C(\varepsilon)$, że dla dostatecznie dużych n

$$(6.2) \quad P_f \left[\|\hat{f}_n(\xi^n; \cdot) - f(\cdot)\| \geq \varepsilon \right] \leq e^{-nC(\varepsilon)}.$$

(Nieco słabsza wersja tego twierdzenia - z prawą stroną nierówności (6.2) o postaci $e^{-nb_n C(\varepsilon)}$ - udowodniona została przez Révész [36]; wcześniej, ale przy znacznie mocniejszych założeniach, nierówność ta wyprowadzona została przez Rejtő i Révész [35].)

6.2. Stosując podobne podejście, L.Györfi udowodnił zbieżność (6.1) EWWY w problemie z zależnymi obserwacjami.

TWIERDZENIE ([25]). Niech $\dots, \xi_{-n}, \dots, \xi_{-1}, \xi_0, \xi_1, \dots, \xi_n, \dots$ będzie ciągiem zmiennych losowych o tym samym rozkładzie danym gęstością f . Niech ciąg (ξ_n) będzie ściśle stacjonarny i ergodyczny. Niech rozkład warunkowy Q_n , gdzie

$$Q_n(A) := P[\xi_n \in A | \xi_{n-1}, \xi_{n-2}, \dots, \xi_1, \xi_0, \xi_{-1}, \dots],$$

ma dla każdego n gęstość f_n oraz $E \|f_1\| < \infty$ ($\|\cdot\|$ oznacza normę w przestrzeni L_2). Niech K będzie gęstością całkowalną w kwadracie, $b_n \rightarrow 0$, $\sum_{n=1}^{\infty} (nb_n)^{-1} < \infty$ oraz

$$(6.3) \quad \sup_n \frac{1}{n} \cdot \sum_{k=1}^{n-1} k \cdot \int \left| \frac{b_{k+1}}{b_k} K\left(\frac{b_{k+1}}{b_k} x\right) - K(x) \right| dx < \infty.$$

Wówczas dla EWWY spełniony jest związek (6.1).

(Sens warunku (6.3) przedyskutowany jest w [25]. Podane twierdzenie uogólnia wcześniejszy wynik Györfiego, podany w [26].)

6.3. W rozdziale 2 omówiliśmy zgodność punktową estymatorów jądrowych. Jak to zobaczyliśmy, zgodność ta implikuje zgodność estymatorów jądrowych w metryce przestrzeni L_1 (p. paragraf 2.11). Kończąc nasz przegląd, wymienimy jeszcze prace dotyczące zgodności jednostajnej EWWY i ED: [11], [13], [15]. Analogiczne rezultaty dla ERP udowodnione zostały np. w pracach [45], [47] i [19]. Prac tych nie będziemy tu już omawiać, wspomnimy jedynie, że badanie zgodności jednostajnej wymaga już czynienia pewnych założeń o nieznannej gęstości, w szczególności wymaga jej jednostajnej ciągłości.

PRACE CYTOWANE

- [1] G.B a n o n, Sur un estimateur non paramétrique de la densité de probabilité, Rev.Statist.Appl. 24 (1976), 61-73.
- [2] G.B e n n e t t, Probability inequalities for the sum of independent random variables, J.Amer.Statist.Assoc. 57 (1962), 33-45.
- [3] P.B i c k e l, E.L e h m a n n, Unbiased estimation in convex families, Ann.Math.Statist. 40 (1969), 1523-1535.
- [4] P.J.B i c k e l, M.R o s e n b l a t t, On some global measures of the deviations of density function estimates, Ann.Statist. 1(1973), 1071-1095. Corrections: Ann. Statist. 3(1975), 1370.

-
- [5] Th.C a c o u l l o s, Estimation of a multivariate density, Ann.Inst.Statist.Math. 18 (1966), 179-189.
- [6] R.J.C a r r o l l, On sequential density estimation, Z.Wahrsch.Verw.Gebiete 36 (1976), 137-151.
- [7] Y.S.C h o w, H.R o b b i n s, On the asymptotic theory of fixed-width segential confidence intervals for the mean, Ann.Math.Statist. 36 (1965), 457-462.
- [8] Y.S.C h o w, H.T e i c h e r, Probability Theory, Springer, New York - Heidelberg - Berlin 1978.
- [9] K.C r a s w e l l, Density estimation in a topological group, Ann.Math.Statist. 36 (1965), 1047-1048.
- [10] M.C s ö r g ö, P.R é v é s z, Strong Approximations in Probability and Statistics, Akademiai Kiado, Budapest 1981.
- [11] H.I.D a v i e s, Strong consistency of a sequential estimator of a probability density function, Bull.Math. Statist. 15 (1973), 49-54.
- [12] H.I.D a v i e s, E.J.W e g m a n, Sequential density estimation, IEEE Trans. Inform. Theory IT-21 (1975), 619-628.
- [13] P.D e h e u v e l s, Sur l'estimation sequentielle de la densite, C.R.Acad.Sci.Paris Ser.A 276 (1973), 1119-1121.
- [14] P.D e h e u v e l s, Conditions necessaires et suffisantes de convergence ponctuelle presque sure et uniforme presque sure des estimateurs de la densite, C.R.Acad.Sci.Paris Ser.A 278 (1974), 1217-1220.
- [15] P.D e h e u v e l s, Estimation séquentielle de la densité, Thèse présentée a l'Université Paris VI, 1974.

-
- [16] P.D e h e u v e l s, Estimation sequentielle de la densite, Contribuciones en Probabilidad y Estadistica Matematica, Esenanza de la Matematica y Analisis. Grindley Ediciones, Granada 1979, 156-168.
- [17] L.D e v r o y e, On the pointwise and integral convergence of recursive kernel estimates of probability densities, Utilitas Math. 15 (1979), 113-128.
- [18] L.D e v r o y e, T.J.W a g n e r, The L_1 convergence of kernel density estimates, Ann.Statist. 7(1979), 1136-1139.
- [19] L.D e v r o y e, T.J.W a g n e r, The strong uniform consistency of kernel density estimates, w Multivariate Analysis V, P.R.Krishnaiah (wyd.) North-Holland 1980, 59-77.
- [20] V.A.E p a n e č n i k o v, Neparametričeskaja ocenka mnogomernoj plotnosti verojatnosti, Teoriya Ver. Prim. 14(1969), 156-161.
- [21] R.H.F a r r e l l, On the lack of a uniformly consistent sequence of estimators of a density function in certain cases, Ann.Math.Statist. 38 (1967), 471-474.
- [22] A.F ö l d e s, Density estimation for dependent sample, Studia Sci.Math.Hungar. 9 (1974), 443-452.
- [23] A.F ö l d e s, P.R é v é s z, A general method for density estimation, Studia Sci.Math.Hungar. 9 (1974), 81-92.
- [24] N.G l i c k, Consistency consitions for probability estimators and integrals of density estimators, Utilitas Math. 6 (1974), 61-74.

-
- [25] L.G y ö r f i, Strong consistent density estimate from ergodic sample, J.Multivariate Anal. 11 (1981), 81-84.
- [26] L.G y o r f i, Z.G y o r f i, I.V a j d a, A strong law of large numbers and some applications. Studia Sci.Math. Hungar. 12 (1977), 233-244.
- [27] P.H a l l, Laws of the iterated logarithm for non-parametric density estimators, Zeit. Wahr. verw. Geb. 56 (1981), 47-61.
- [28] S.A.H a š i m o v, Ravnomernaja srednekvadratičeskaja schodimost' ocenki plotnosti verojatnosti, Sluč. Processy Statist. Vyvody, Izd. "Fan" Uzbetskoj SSR, Taškent 1974, 185-193.
- [29] A.K o z e k, W.W e r t z, Estimators with values in Banach spaces under convex loss and applications to curve estimation problems, Coll.Math.Soc.János Bolyai Vol.32 (1980).
- [30] A.K o z e k, W.W e r t z, On the existence of minimal complete classes of estimators, The First Pannonian Symposium on Mathematical Statistics, P.Révész, L.Schmetterer, V.M.Zolotarev (wyd.), Lecture Notes in Statistics, Vol. 8, Springer, New York - Heidelberg - Berlin 1981, 140-155.
- [31] M.L o é v e, Probability Theory, Third edition, D.van Nostrand Comp., New Jersey 1963.
- [32] M.A.M i r z a h m e d o v, S.A.H a š i m o v, On some properties of estimators of a probability density, Kybernetika (Prague) 9 (1973), 242-250.

-
- [33] E.P a r z e n, On estimation of a probability density function and mode, Ann.Math.Statist. 33 (1962), 1065-1076.
- [34] V.V.P e t r o v, Summary nezavisimych slučajnych veličin, Nauka, Moskva 1972 (istnieje tłum. ang.).
- [35] L.R e j t ö, P.R é v é s z, Density estimation and pattern classification, Problems Control Inform. Theory 2 (1973), 67-80.
- [36] P.R é v é s z, Robbins-Monto procedure in a Hilbert space and its application in the theory of learning processes. I, Studia Sci. Math.Hungar. 8 (1973), 391-398.
- [37] M.R o s e n b l a t t, Remarks on some nonparametric estimates of a density function, Ann.Math.Statist. 27 (1956), 832-837.
- [38] M.R o s e n b l a t t, Curve estimates, Ann.Math.Statist. 42 (1971), 1815-1842.
- [39] L.R u t k o w s k i, Sequential estimates of probability densities by orthogonal series and their application in pattern classification, IEE Trans. Systems Man Cybernet. SMC-10 (1980), 918-920.
- [40] H.S c h e f f e, A useful convergence theorem for probability distributions, Ann.Math.Statist. 18 (1947), 434-438.
- [41] L.S c h m e t t e r e r, Sur l'iteration stochastique. Le calcul des probabilités et ses applications, Coll.Int. Centre National de la Recherche Sci. 87, Paris 1959, 55-67.

-
- [42] L.S c h m e t t e r e r, Sur quelques resultats asymptotiques pour le processus de Robins-Monro, Ann.Sci. Univ.Clermont Math. 12 (1967), 166-176.
- [43] L.S c h m e t t e r e r, From stochastic approximation to the theory of optimization, 11. Steiermärkisches Math.Symposium. Ber.Math.-Statist. Sektion im Forschungszentrum Graz, Graz 1979, Ber.Nr. 127.
- [44] B.S c h n e i d e r, W.W e r t z, Statistical density estimation: a bibliography, Internat.Statist.Rev. 47 (1979), 155-175.
- [45] E.F.S c h u s t e r, Estimation of a probability density function and its derivatives, Ann.Math.Statist. 40 (1969), 1187-1195.
- [46] E.F.S c h u s t e r, On the rate of convergence of an estimate of a functional of a probability density, Scand.Actuar.J. 1974 (1974), 103-107.
- [47] B.W.S i l v e r m a n, Weak and strong uniform consistency of the kernel estimate of a density and its derivatives, Ann. Statist. 6 (1978), 177-184.
- [48] R.C.S r i v a s t a v a, Estimation of probability density function based on random number of observations with applications, Int.Stat.Rev. 41 (1973), 77-86.
- [49] E.L.S t e i n, Singular Integrals and Differentiability Properties of Functions, Princeton University Press, Princeton, New Jersey 1970.
- [50] W.S t u t e, The oscillation behavior of empirical processes, Ann.Probab. 10 (1982), 86-107.

-
- [51] W.S t u t e, A law of the logarithm for kernel estimators, Ann.Probab. 11 (1982), 414-422.
 - [52] W.S t u t e, Sequential fixed-width confidence intervals for a nonparametric density function, Zeit. Wahr. verw. Geb. 62 (1983), 113-123.
 - [53] V.S u s a r l a, G.W a l t e r, Estimation of a multivariate density function using delta sequences, Ann. Statist. 9 (1981), 347-355.
 - [54] H.T e i c h e r, Almost certain behaviour of row sums of double arrays, w Analytical Methods in Probability Theory, D.Dugue, E.Lukacs, V.K.Rohatgi (wyd), Lecture Notes in Math. vol. 861, Springer, New York - Heidelberg - Berlin 1981, 155-165.
 - [55] L.V ă d u v a, Estimatiia unei densitati de repartitie k-dimensională, Stud.Cerc.Math. 14 (1963), 653-660.
 - [56] L.V ă d u v a, Contributii la teoria estimatiilor statistice ale densitatilor de repartitie si aplicatii, Stud. Cerc.Math. 20 (1968), 1207-1276.
 - [57] J.H.V e n t e r, On Dvoretzky stochastic approximation theorems, Ann.Math.Statist. 37 (1966), 1534-1544.
 - [58] G.W a l t e r, J.B l u m, Probability density estimation using delta sequences, Ann.Statist. 7 (1979), 328-340.
 - [59] G.W a t s o n, M.L e a d b e t t e r, On the estimation of the probability density I, Ann.Math.Statist. 34(1963), 480-491.
 - [60] G.W a t s o n, M.L e a d b e t t e r, Hazard analysis I, Biometrika 51 (1964), 175-184.

-
- [61] G.W a t s o n, M.L e a d b e t t e r, Hazard analysis II, Sankhya Ser.A. 26 (1964), 101-116.
- [62] E.J.W e g m a n, H.I.D a v i e s, Remarks on some recursive estimators of a probability density function, Ann.Statist. 7 (1979), 316-327.
- [63] W.W e r t z, Schätzfolgen für Dichten auf topologischen Gruppen, Operations Research-Verf. 14 (1972), 247-257.
- [64] W.W e r t z, Statistical Density Estimation - A Survey, Vandenhoeck & Ruprecht, Göttingen 1978.
- [65] W.W e r t z, non-parametric density estimation in abstract and homogeneous spaces, The First Pannonian Symposium on Mathematical Statistics P.Révész, L.Schmetterer, V.M. Zolotarev (wyd.), Lecture Notes in Statistics, Vol 8, Springer, New York - Heidelberg - Berlin 1981, 290-301.
- [66] W.W e r t z, Recursive and sequential density estimation, w Sequential methods in statistics, R.Zieliński (wyd.) Banach Center Publication vol.16, PWN, Warszawa (w druku).
- [67] P.W h i t t l e, On the smoothing of probability density functions, J.Roy.Statist.Soc.Ser.B 20 (1958), 334-343.
- [68] C.T.W o l v e r t o n, T.J.W a g n e r, Recursive estimates of probability densities, IEEE Trans.Systems Sci. Cybernet. 5 (1969), 246-247.
- [69] H.Y a m a t o, Sequential estimation of a continuous probability density function and mode, Bull. Math. Statist. 14 (1971), 1-12.
- [70] H.Y a m a t o, Some statistical properties of estimators of density and distribution functions, Bull.Math.Statist. 15 (1972), 113-131.

[71] N.N.^yČ e n c o v, Ocenka neizvestnoj plotnosti raspredelenija po nabludenijam, DAN ŠSSR 147, 1(1962), 45-48.

Dodane w korekcje. Spośród prac, w których poczesne miejsce zajmuje problematyka rekurencyjnej estymacji gęstości, a które napisane zostały już po przygotowaniu tego opracowania wymienić trzeba przynajmniej monografię L.Devroye'a i L.Györfiego, Nonparametric Density Estimation. The L₁ View, Wiley, New York 1984.