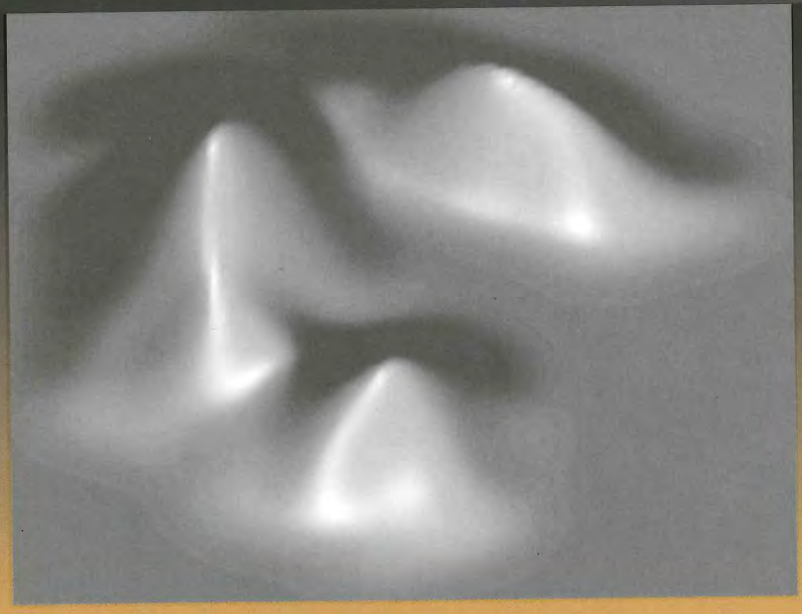
Piotr Kulczycki

estymatory jądrowe w analizie systemowej



Wydawnictwa Naukowo-Techniczne

O Autorze:

Piotr Kulczycki ukończył studia z zakresu automatyki na Wydziale Elektrotechniki, Automatyki i Elektroniki Akademii Górniczo-Hutniczej, a także matematyki stosowanej na Wydziale Matematyki i Fizyki Uniwersytetu Jagiellońskiego. Obecnie pracuje na stanowisku docenta w Instytucie Badań Systemowych Polskiej Akademii Nauk oraz jako profesor nadzwyczajny na Wydziale Inżynierii Elektrycznej i Komputerowej Politechniki Krakowskiej, gdzie jest kierownikiem Zakładu Automatyki. Był wielokrotnym stypendystą rządów Austrii, Belgii, Danii oraz Finlandii. Przebywał również w duńskim Aalborg Universitet w charakterze profesora wizytującego.

Zajmuje się współczesnymi technikami informacyjnymi, zwłaszcza dotyczącymi analizy oraz eksploracji danych, w których są zawarte informacje niepewne i nieprecyzyjne, a także ich zastosowaniami do zagadnień współczesnej analizy systemowej oraz inżynierii sterowania. W swoim dorobku naukowym ma między innymi dwie monografie oraz liczne artykuły w renomowanych czasopismach naukowych o zasięgu międzynarodowym.

Jego hobby to historia starożytna i okresu napoleońskiego oraz turystyka zagraniczna. Interesuje się również muzyką klasyczną i ogrodnictwem.

Dodatkowe informacje na temat zawartego w książce materiału można znaleźć w Internecie pod adresem

http://www.ibspan.waw.pl/kulczycki/EJwAS_05

Wszelkie uwagi na jej temat prosimy kierować pod adresem poczty elektronicznej Autora:

kulczycki@ibspan.waw.pl
kulczycki@pk.edu.pl

estymatory jądrowe w analizie systemowej

Piotr Kulczycki

estymatory jądrowe w analizie systemowej

Wydawnictwa Naukowo-Techniczne
Warszawa

Opiniodawcy:
prof. dr hab. inż. Olgierd Hryniewicz
prof. dr hab. inż. Wojciech Mitkowski

Redaktor *Ewa Zdanowicz*

Okładkę i strony tytułowe projektował *Paweł G. Rubaszewski*
Redaktor techniczny *Anna Szeląg*
Korekta *Zespół*
Przygotowanie do druku *Maciej Jaworski*

Wydanie książki dofinansowane przez Instytut Badań Systemowych PAN

© Copyright by Wydawnictwa Naukowo-Techniczne
Warszawa 2005

All Rights Reserved
Printed in Poland

Utwór w całości ani we fragmentach nie może być powielany ani rozpowszechniany za pomocą urządzeń elektronicznych, mechanicznych, kopiujących, nagrywających i innych, w tym również nie może być umieszczany ani rozpowszechniany w postaci cyfrowej zarówno w Internecie, jak i w sieciach lokalnych bez pisemnej zgody posiadacza praw autorskich.

Wydawnictwa Naukowo-Techniczne
00-048 Warszawa, ul. Mazowiecka 2/4
tel. (0-22) 826 72 71, e-mail: wnt@wnt.pl
www.wnt.pl

ISBN 83-204-3147-6

Spis treści

Przedmowa 7

1. Wstęp 9
   Uwagi bibliograficzne 20

2. Podstawy matematyczne 21
   2.1. Elementy probabilistyki 21
   2.1.1. Podstawowe definicje 22
   2.1.2. Charakterystyki funkcyjne rozkładu prawdopodobieństwa 25
   2.1.3. Charakterystyczne parametry rozkładu prawdopodobieństwa 28
   2.1.4. Niezależność zmiennych losowych 32
   Uwagi bibliograficzne 33
   2.2. Teoria estymacji 33
   2.2.1. Podstawowe definicje 34
   2.2.2. Własności estymatorów 36
   2.2.3. Przykłady estymatorów 38
   Uwagi bibliograficzne 43
   2.3. Rozwiązania równań różniczkowych z nieciągłą prawą stroną 43
   2.3.1. Podstawowe definicje i zależności 44
   2.3.2. Przykład 50
   Uwagi bibliograficzne 54

3. Estymatory jądrowe 55
   3.1. Jądrowy estymator gęstości prawdopodobieństwa 56
   3.1.1. Definicje i interpretacje 56
   3.1.2. Analiza błędu estymacji 59
   3.1.3. Wybór postaci jądra 65
   3.1.4. Transformacja liniowa 72
   3.1.5. Wyznaczanie wartości parametru wygładzania 76
   3.1.6. Modyfikacja parametru wygładzania 90
   3.1.7. Ograniczenie nośnika 94

3.1.8. Współrzędne binarne ... 98
3.1.9. Liczność próby losowej ... 103
3.1.10. Przykłady ... 108
Przykład 3.1 – jednowymiarowa zmienna losowa ... 108
Przykład 3.2 – jednowymiarowa zmienna losowa z ograniczeniem nośnika ... 114
Przykład 3.3 – jednowymiarowa zmienna losowa ze składową binarną ... 115
Przykład 3.4 – wielowymiarowa zmienna losowa z użyciem jądra produktowego ... 123
Przykład 3.5 – wielowymiarowa zmienna losowa z użyciem jądra radialnego ... 128
Uwagi bibliograficzne ... 134

3.2. Jądrowe estymatory innych charakterystyk probabilistycznych ... 135
3.2.1. Jądrowy estymator dystrybuanty ... 135
3.2.2. Jądrowy estymator kwantyla ... 142
3.2.3. Jądrowe estymatory charakterystyk warunkowych ... 146
3.2.4. Przykłady ... 153
Przykład 3.6 – dystrybuanta ... 153
Przykład 3.7 – kwantyl ... 155
Przykład 3.8 – charakterystyki warunkowe ... 157
Uwagi bibliograficzne ... 163

4. Przykładowe zastosowania w analizie systemowej ... 165
Uwagi bibliograficzne ... 167

4.1. Identyfikacja wartości parametrów z niesymetryczną funkcją strat ... 168
4.1.1. Liniowa niesymetryczna funkcja strat ... 168
4.1.2. Kwadratowa niesymetryczna funkcja strat ... 170
4.1.3. Kwadratowa niesymetryczna funkcja strat w przypadku dwuwymiarowym ... 175
4.1.4. Wykorzystanie do zagadnień sterowania optymalnego ... 180
Uwagi bibliograficzne ... 185

4.2. Wykrywanie i klasyfikacja elementów nietypowych (odosobnionych) ... 185
4.2.1. Wykrywanie elementów nietypowych ... 188
4.2.2. Klasyfikacja elementów nietypowych ... 190
4.2.3. Ujęcie lokalne ... 193
Uwagi bibliograficzne ... 195

4.3. Określenie przestrzennego rozkładu popytu na potrzeby planowania systemu telekomunikacyjnego ... 195
4.3.1. Przestrzenny rozkład popytu ... 197
4.3.2. Wyznaczenie optymalnego układu stacji bazowych ... 198
4.3.3. Ujęcie rozmyte ... 202
Uwagi bibliograficzne ... 204

5. Podsumowanie i końcowe komentarze ... 207
Uwagi bibliograficzne ... 215

Bibliografia ... 217

Skorowidz ... 221

Przedmowa

Istotnym problemem nauki i działań praktycznych jest wyznaczanie różnorodnych charakterystyk badanego obiektu na podstawie dostępnych danych eksperymentalnych. I tak na przykład funkcję regresji, opisującą zależność między dwoma skorelowanymi ze sobą wielkościami, określa się, dysponując otrzymanymi doświadczalnie parami tych wielkości. Innym typowym przykładem jest estymacja gęstości prawdopodobieństwa zmiennej losowej na podstawie uzyskanej eksperymentalnie próby.

Klasyczne procedury polegają na arbitralnym założeniu postaci badanych funkcji, a potem wyznaczeniu występujących w nich parametrów. Określa się je mianem metod parametrycznych. Ich zasadniczą własnością jest prostota teoretyczna i obliczeniowa, a także powszechna znajomość i dostępność w literaturze. Obecnie, w miarę rozwoju techniki komputerowej, coraz częściej są wprowadzane metody nieparametryczne, których główną cechą jest brak arbitralnego założenia o postaci rozważanej funkcji. W ujęciu probabilistycznym podstawowym rodzajem tych metod stają się procedury konstruowania estymatorów jądrowych. Chociaż koncepcja estymatorów jądrowych jest prosta, a interpretacja przejrzysta, to praktyczne użycie nie jest jednak możliwe bez zastosowania odpowiedniej techniki komputerowej, co jeszcze w niedalekiej przeszłości znacząco utrudniało badania teoretyczne, a zwłaszcza aplikacyjne.

Estymatory jądrowe nie doczekały się dotychczas opracowania w języku polskim, oprócz krótkich wzmianek w nielicznych podręcznikach. Klasyczne publikacje angielskojęzyczne wydają się obciążone typową wadą nowatorskiej metodyki – wiele przedstawianych w nich zagadnień jest niedopracowanych, co powoduje, że stanowią one raczej materiał do dalszych badań niż gotowe do stosowania algorytmy.

Celem wydania niniejszej monografii jest udostępnienie polskiemu Czytelnikowi pozycji zawierającej tę część przedmiotowego materiału, który został już dopracowany z teoretycznego punktu widzenia i sprawdzony w różnorodnych zastosowaniach. Książka jest przeznaczona dla szerokiego kręgu teoretyków i praktyków zainteresowanych współczesnymi metodami modelowania niepewności, zwłaszcza probabilistyki i logiki rozmytej, a także analizą danych. Materiał jest podany w formie dogodnej do bezpośredniego stosowania, bez szczegółowego wnikania w teoretyczne

aspekty zagadnień. Możliwości aplikacyjne zostały zilustrowane na przykładzie zadań współczesnej analizy systemowej. Wydaje się, że właśnie ta ogólna i interdyscyplinarna dziedzina, analitycznie wspomagająca racjonalne działania w zagadnieniach dotyczących złożonych systemów, najczęściej technicznych, ekonomicznych i społecznych, najbardziej nadaje się do tego celu, biorąc pod uwagę uniwersalne możliwości estymatorów jądrowych.

Książka ta jest podzielona na pięć rozdziałów.

Rozdział pierwszy ma charakter wprowadzający. Została w nim podana motywacja podjęcia tematyki analizy systemowej jako wszechstronnej dziedziny badań, po czym – estymacji nieparametrycznej jako narzędzia do rozwiązywania tego typu zadań.

Rozdział drugi zawiera podstawowe definicje z zakresu probabilistyki (podrozdział 2.1) i teorii estymacji (podrozdział 2.2). W jego skład wchodzi także podrozdział 2.3 traktujący o rozwiązaniach równań różniczkowych z nieciągłą prawą stroną. Mimo że podrozdział ten mógłby być częścią punktu 4.1.4, został tu celowo wyodrębniony, gdyż temat ten – istotny we współczesnej teorii systemów dynamicznych – jest praktycznie nieobecny w literaturze polskojęzycznej.

Rozdział trzeci stanowi główną część niniejszej monografii. W podrozdziale 3.1 została szczegółowo przedstawiona koncepcja jądrowego estymatora gęstości prawdopodobieństwa. Poszczególne punkty zawierają kompletne procedury umożliwiające jego wyznaczenie na podstawie uzyskanej eksperymentalnie próby. Z kolei podrozdział 3.2 jest poświęcony przykładowym koncepcjom uogólnienia tej metody na problemy estymacji dystrybuanty i kwantyla (m.in. mediany), a także charakterystyk warunkowych.

Rozdział czwarty obejmuje kolejno w podrozdziałach 4.1-4.3 opisy konkretnych zastosowań estymatorów jądrowych do zagadnień wyznaczenia wartości parametrów modelu (np. w sterowaniu optymalnym i odpornym), wykrywania stanów nietypowych oraz ich klasyfikacji (np. w diagnostyce technicznej i medycznej), a także oszacowania przestrzennego popytu (np. w zagadnieniach telekomunikacji oraz handlu i usług).

Rozdział piąty jest podsumowaniem zaprezentowanego w tej monografii materiału. Zawiera również dodatkowe komentarze oraz uogólnienia.

Autor pragnie serdecznie podziękować swoim doktorantom, Paniom Małgorzacie Charytanowicz i Karinie Daniel oraz Panom Piotrowi Kowalskiemu, Aleksandrowi Mazgajowi, Cyprianowi Prochotowi i Jackowi Waglowskiemu za prace, których tematyka stanowi podstawę treści rozdziału czwartego, a także za szczegółowe badania oraz konstruktywne dyskusje w zakresie zagadnień przedstawionych w niniejszej książce.

Piotr Kulczycki

Kraków, Warszawa
kwiecień 2005

Rozdział 1

Wstęp

Dokonywanie wyborów i w konsekwencji podejmowanie decyzji jest dla nas czynnością powszednią. Jeśli rzecz dotyczy typowych aspektów codzienności, robimy to rutynowo na podstawie nabytego doświadczenia. Im sytuacja jest bardziej złożona, tym chętniej korzystamy z rad innych osób, a nawet posiłkujemy się prostymi obliczeniami. W przypadku skrajnie skomplikowanych zadań decyzyjnych, odnoszących się do złożonych systemów technicznych, ekonomicznych i społecznych, funkcję doradczą opartą na ścisłych metodach naukowych spełnia **analiza systemowa**. Może być ona użyta wobec każdego problemu, w którym aspekt racjonalnego wyboru jest na tyle wyrazisty, aby opłacało się wykonywać zaawansowane badania analityczne.

Typowymi przykładami zastosowania klasycznej analizy systemowej są złożone decyzje budżetowe i inwestycyjne (m.in. zadania lokalizacji nowych lotnisk, planowania sieci transportowej lub telekomunikacyjnej, budowy systemu zapór wodnych), wyznaczanie kierunku rozwoju gospodarki, a także makroproblemy ekologiczne oraz strategiczne zagadnienia marketingowe i militarne. Analiza systemowa najczęściej nie prowadzi wtedy jednak do podjęcia konkretnej decyzji, lecz jedynie do dostarczenia i odpowiedniego zaprezentowania ścisłych danych ułatwiających tę procedurę lub wręcz do opracowania konkretnych wariantów rozwiązań ze wskazaniem pozytywów i negatywów każdego z nich. W przypadku zadań dotyczących państwa, regionu lub znaczących podmiotów gospodarczych sama decyzja jest bowiem podejmowana z uwzględnieniem uwarunkowań politycznych, społecznych i psychologicznych. Jednak dogłębna analiza problemu na gruncie ścisłych reguł nauki oraz wprowadzenie konkretnych związków ilościowych pozwalają na lepsze zrozumienie istoty procesów zachodzących w złożonych systemach, z licznymi i różnorodnymi powiązaniami jego części składowych.

Gwałtowny rozwój techniki komputerowej, jaki nastąpił w ostatnich dwóch dekadach XX wieku, miał znaczący wpływ na uwarunkowania, a tym samym na przedmiot i metody **współczesnej analizy systemowej**. Charakter tych metod

wyraźnie ewoluował od zmatematyzowanych metod analitycznych w kierunku tak zwanych technik informacyjnych, czyli specjalizowanych procedur przetwarzania informacji. W związku ze znacznym wzrostem zautomatyzowania aparatury pomiarowej łączonej za pomocą sieci komputerowych powszechny dotychczas niedostatek danych paradoksalnie uległ zamianie na ich nadmiar, co wymogło stosowanie marginalnych do niedawna procedur eliminacji informacji małowartościowych i agregację wiedzy zawartej w pozostałej ich części, czyli działań z zakresu tak zwanej inżynierii wiedzy. I wreszcie, ogólna dostępność zaawansowanego sprzętu komputerowego, połączonego w sieci lokalne oraz globalnie za pomocą Internetu, pozwoliła rozpocząć proces upowszechniania procedur analizy systemowej, dzięki możliwości stosowania przyjaznego użytkownikowi oprogramowania, wymagającego jedynie wprowadzenia danych o niespecjalistycznej strukturze, bez konieczności wnikania w często bardzo złożone aspekty stosowanych procedur.

Dzięki tym przeobrażeniom współczesna analiza systemowa objęła wiele nowych zagadnień nauki i działań praktycznych, których przedmiotem są wszechstronne badania rozważanego systemu oraz jego wpływu na otoczenie, a w konsekwencji podjęcie wymaganych decyzji. I tak nadzorowanie złożonych układów sterowania automatycznego w celu wykrycia oraz zdiagnozowania ewentualnych uszkodzeń uważa się za sukcesywne podejmowanie decyzji o stanie technicznym takiego układu i może być ono realizowane za pomocą procedur analizy systemowej. Podobnie można traktować początkową fazę diagnostyki medycznej (w szczególności realizowaną poprzez Internet), polegającą na udzielaniu przez pacjenta odpowiedzi na proste pytania generowane bez udziału lekarza przez tak zwany system ekspercki, a w konsekwencji otrzymaniu wstępnego rozpoznania chorobowego. Procedury analizy systemowej mogą być również stosowane do analizy informacji meteorologicznych oraz sejsmologicznych, w wyniku której dokonuje się preselekcji znacznej ilości danych uzyskiwanych w celu wczesnego ostrzegania przed kataklizmami. Współczesna analiza systemowa ma zatem charakter wysoce interdyscyplinarny.

Obecna **metodyka analizy systemowej** lokuje się na pograniczu matematyki stosowanej, informatyki oraz zaawansowanej inżynierii (zwłaszcza szeroko rozumianej automatyki) i ekonomii. Tak więc wykorzystuje się tutaj ogólną teorię systemów, głównie w odniesieniu do układów dynamicznych, metody modelowania matematycznego, optymalizację i badania operacyjne, ekonometrię, a także techniki informacyjne, inżynierię wiedzy, metody numeryczne i wiele innych dziedzin nauki, odpowiednio je łącząc w interdyscyplinarny konglomerat. Ponieważ w zagadnieniach współczesnej analizy systemowej silnie uwidaczniają się elementy niepewności w postaci nieprzewidywalnej przyszłości, niedokładnych danych i nieprecyzyjnych – często werbalnych – ocen ekspertów, zasadniczego znaczenia nabrały w ostatnim czasie różne sposoby opisu nieokreślono-

ści, zwłaszcza zaawansowane metody probabilistyczne i statystyczne oraz logika rozmyta.

W praktyce każdej z powyższych dyscyplin wszechobecne jest zagadnienie szacowania (lub w ujęciu probabilistycznym – estymacji) funkcji lub parametrów charakteryzujących poszczególne wielkości. Interdyscyplinarność analizy systemowej wymaga – także w przypadku tego typu zadań – uniwersalności stosowanej metodyki, a szybki rozwój techniki komputerowej otwiera całkowicie nowe możliwości teoretyczne i praktyczne. Czynniki te jednoznacznie wskazują na **estymację nieparametryczną** z jej podstawową koncepcją – **estymatorami jądrowymi**. Poniżej zostanie przedstawione intuicyjne wprowadzenie do tego zagadnienia; ścisłe ujęcie jest prezentowane w rozdziale 3 niniejszej monografii. Podmiotem rozważań będzie estymacja gęstości prawdopodobieństwa, jako zagadnienia bliskiego intuicji i łatwego w interpretacji, o fundamentalnym znaczeniu w wielu współczesnych dyscyplinach teoretycznych oraz praktycznych.

Niech zatem dana będzie rzeczywista zmienna losowa X, czyli zmienna przyjmująca z pewnym prawdopodobieństwem wartości ze zbioru liczb rzeczywistych. Podstawową i najbliższą intuicji charakterystyką rozkładu owego prawdopodobieństwa jest jego **gęstość**, oznaczana w dalszej części przez f. Można ją zinterpretować jako taką funkcję, że prawdopodobieństwo tego, iż powyższa zmienna losowa przyjmuje wartości z przedziału $[a, b]$, wynosi $\int_a^b f(x)\,dx$ (por. rys. 1.1). Tak więc, im zmienna losowa częściej przyjmuje wartości z otoczenia liczby x, tym wartość gęstości $f(x)$ jest większa. Naturalne są wymagania, by spełnione były warunki

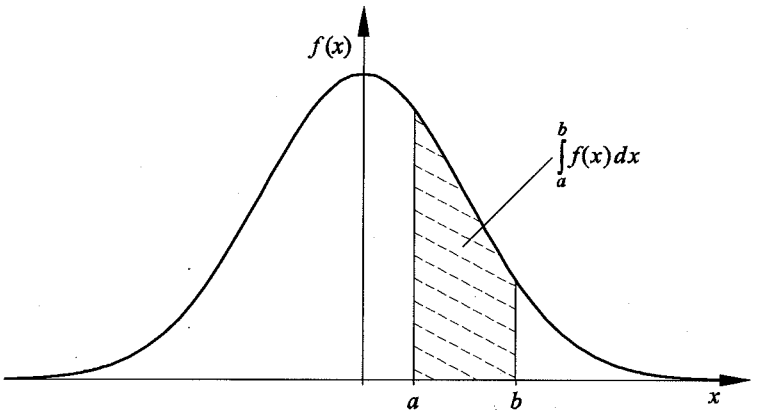
$$f(x) \geq 0 \quad \text{dla każdego } x, \tag{1.1}$$

ponieważ prawdopodobieństwo dowolnego zbioru nie może być ujemne, a także

$$\int_{-\infty}^{\infty} f(x)\,dx = 1, \tag{1.2}$$

jako że prawdopodobieństwo przyjęcia jakiejkolwiek wartości przez zmienną X wynosi 1. Na rysunku 1.1 została pokazana gęstość prawdopodobieństwa klasycznego rozkładu normalnego. Oczywiste jest jednak, że istnieje olbrzymia ilość zjawisk o odmiennych własnościach niż reprezentowane przez ten szczególny rozkład.

Nieznaną funkcję f wyznacza się, określając eksperymentalnie jej oszacowanie, czyli estymator $\hat{f}$ obliczony na podstawie m uzyskanych doświadczalnie wartości $x_1, x_2, ..., x_m$ przyjętych przez zmienną losową X.



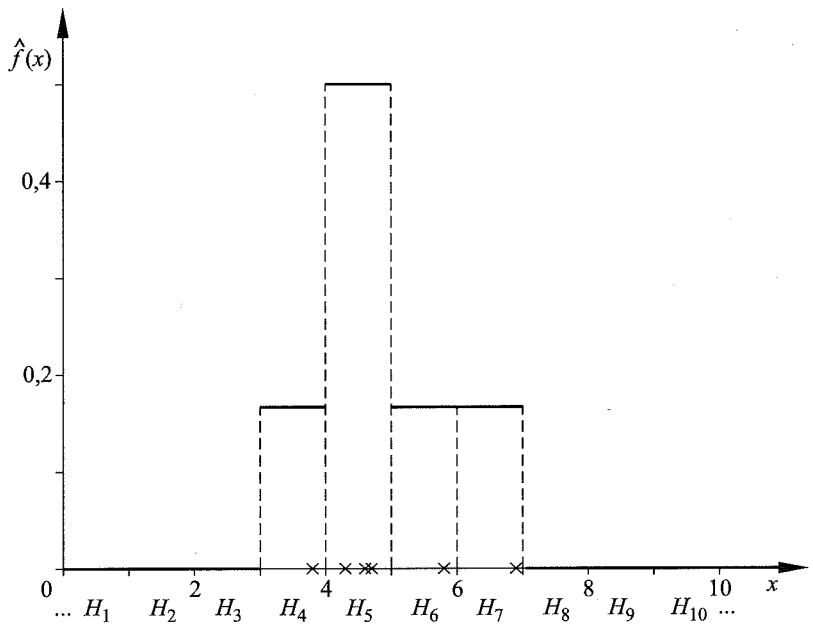
Rys. 1.1. Gęstość prawdopodobieństwa rozkładu normalnego

Klasyczna procedura estymacji polega na arbitralnym wyborze typu rozkładu (co można dodatkowo uwiarygodnić, testując hipotezę o zasadności takiego wyboru) i ustaleniu wartości występujących tam parametrów. Stąd tego rodzaju procedury określa się mianem **metod parametrycznych**. Jednak możliwa do otrzymania tą drogą różnorodność postaci uzyskiwanych gęstości, ograniczona praktycznie do kilkunastu standardowych typów (np. normalny, jednostajny, trójkątny, beta), jest absolutnie nie do zaakceptowania z punktu widzenia złożonych współczesnych zastosowań. Fakt ten spowodował konieczność poszukiwań procedur alternatywnych, niewymagających arbitralnych założeń o typie badanego rozkładu – dla podkreślenia odmienności zostały one nazwane **metodami nieparametrycznymi**.

Trywialnym przykładem należącym do tej grupy metod jest **histogram** (rys. 1.2). Jego koncepcja polega na podziale zbioru liczb rzeczywistych na „przegrody" H_k o jednakowej „szerokości" h, przy czym indeks k jest liczbą całkowitą. Na każdej z owych „przegród" histogram ma wartość stałą, równą liczbie tych wartości próby losowej $x_1, x_2, ..., x_m$, które „wpadły" do danej „przegrody", podzielonej przez mh, czyli

$$\hat{f}(x) = \frac{\#\{x_i \in H_k\}}{mh} \quad \text{dla każdego } x \in H_k \text{ oraz } k \text{ całkowitego,} \tag{1.3}$$

gdzie $\#A$ oznacza liczność zbioru A. Histogram jest obecnie jedynie skutecznym narzędziem ilustracyjnym – nawet laik potrafi bowiem zinterpretować przedstawione w ten sposób wyniki. Brak jest natomiast wiarygodnych metod wyboru wartości parametru h oraz ustalenia położenia środków „przegród", a kształt histogramu okazuje się bardzo wrażliwy na te wielkości. Pochodna histogramu istnieje poza punktami „styku przegród", ale jest tożsamościowo równa zeru, co znacząco utrudnia nawet podstawową analizę teoretyczną.



Rys. 1.2. Histogram

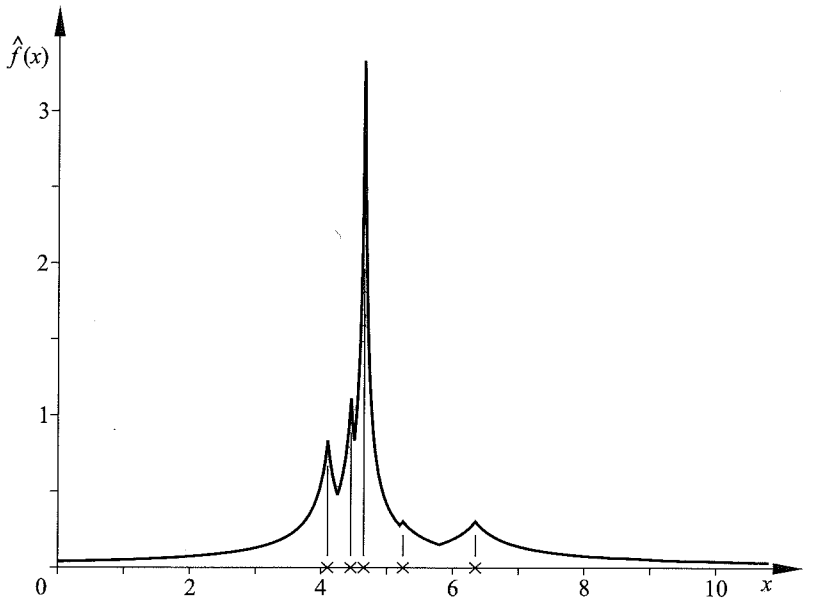
Poszukiwania dalszych metod nieparametrycznych dostarczyły różnorodnych propozycji, aczkolwiek aż do czasu sformułowania koncepcji estymatorów jądrowych żadna z nich nie spełniała w satysfakcjonujący sposób nawet podstawowych uwarunkowań teoretycznych i praktycznych. Jako przykłady zostaną poniżej przedstawione estymator najbliższego sąsiedztwa oraz koncepcja oparta na przekształceniu Fouriera. Warto także wspomnieć o metodach wykorzystujących funkcje sklejane (czyli tzw. *spliny*) oraz teorię falek – używany w nich aparat teoretyczny jest jednak zaawansowany w stopniu uniemożliwiającym ich stosowanie przez osoby bez ścisłego przygotowania matematycznego, co znacząco zmniejsza możliwości aplikacyjne.

Oryginalna koncepcja doprowadziła do konstrukcji **estymatora najbliższego sąsiedztwa** (rys. 1.3). Jego wartość jest dana wzorem

$$\hat{f}(x) = \frac{k-1}{2m\, d_k(x)}, \tag{1.4}$$

gdzie $d_k(x)$ oznacza odległość od argumentu x jego k-tego najbliższego „sąsiada" spośród elementów $x_1, x_2, ..., x_m$, przy czym najczęściej jako k przyjmuje się część całkowitą liczby $\sqrt{m}$. Estymator ten jest więc „sklejeniem" hiperbol, przy czym w miejscach owych „sklejeń" pochodna nie istnieje. Jego wykres ma zatem nieregularny i nienaturalny kształt (por. rys. 1.3). Co ważniejsze, warunek

(1.2) nie jest spełniony, gdyż $\int_{-\infty}^{\tilde{x}_1} \hat{f}(x)\,dx$ oraz $\int_{\tilde{x}_m}^{\infty} \hat{f}(x)\,dx$, gdzie $\tilde{x}_i$ oznacza i-ty pod względem wielkości element zbioru $x_1, x_2, \ldots, x_m$, są proporcjonalne do $\int_{-\infty}^{\tilde{x}_1} 1/(\tilde{x}_k - x)\,dx$ oraz $\int_{\tilde{x}_m}^{\infty} 1/(x - \tilde{x}_{m-k+1})\,dx$, czyli są równe nieskończoności. Nawet gdyby zawęzić rozważania do przedziału ograniczonego, to obliczenie stosownej stałej, gwarantującej spełnienie warunku (1.2), jest zadaniem trudnym do wykonania w praktyce.



Rys. 1.3. Estymator najbliższego sąsiedztwa

Z kolei koncepcja **estymatora Fouriera** wynika wprost z ogólnej teorii przekształcenia Fouriera. Możliwe jest tu określenie estymatora jedynie na ograniczonym przedziale $[a, b]$. Estymator Fouriera jest wówczas zdefiniowany wzorem

$$\hat{f}(x) = \frac{a_0}{2} + \sum_{j=1}^{J} [a_j \cos(j\omega x) + b_j \sin(j\omega x)], \tag{1.5}$$

przy czym J jest odpowiednio ustaloną liczbą naturalną oraz

$$a_j = \frac{2}{(b-a)m} \sum_{i=1}^{m} \cos(j\omega x_i), \tag{1.6}$$

$$b_j = \frac{2}{(b-a)m} \sum_{i=1}^{m} \sin(j\omega x_i), \tag{1.7}$$

$$\omega = \frac{2\pi}{b-a}. \tag{1.8}$$

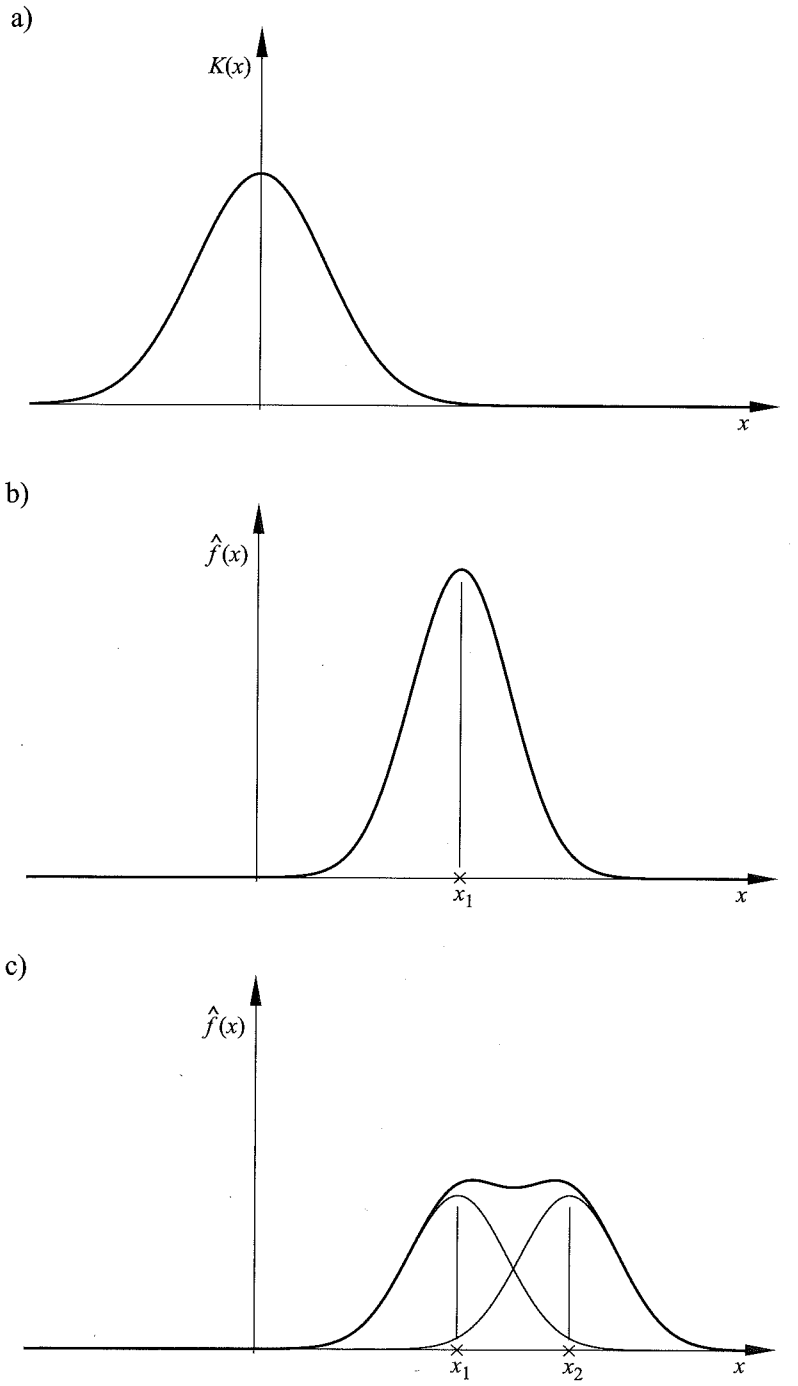
Estymator (1.5) ma pochodną dowolnego rzędu. Co więcej, ponieważ $a_0 = 2/(b-a)$ i $\int_a^b \cos(j\omega x)\,dx = \int_a^b \sin(j\omega x)\,dx = 0$ dla $j = 1, 2, ..., J$, zatem równość (1.2) jest spełniona. Niestety, nie dotyczy to warunku (1.1) – estymator Fouriera może być ujemny w pewnych podprzedziałach dziedziny $[a, b]$.

Obecnie podstawową metodą estymacji nieparametrycznej staje się stosowanie **estymatorów jądrowych**. Koncepcja ich konstrukcji jest naturalna, interpretacje przejrzyste, a postać dogodna do matematycznej analizy. Zostały one zaproponowane na przełomie lat pięćdziesiątych i sześćdziesiątych XX wieku niezależnie przez Rosenblatta i Parzena, ale do lat osiemdziesiątych mogły być przedmiotem zainteresowań badawczych jedynie wąskiej grupy specjalistów. Powszechne badania i przede wszystkim zastosowania estymatorów jądrowych są bowiem niemożliwe bez użycia komputerów o względnie dużej mocy obliczeniowej i możliwości dogodnego przedstawienia wyników na ekranie monitora.

Wyobraźmy sobie hipotetyczne – nierealne z praktycznego punktu widzenia – zadanie estymacji gęstości na podstawie jednej uzyskanej doświadczalnie wartości x_1. Można by w tej sytuacji arbitralnie ustalić nieujemną funkcję K, symetryczną względem zera i mającą w tym punkcie maksimum oraz spełniającą warunek $\int_{-\infty}^{\infty} K(x)\,dx = 1$ (rys. 1.4a), po czym przyjąć wartość estymatora gęstości w postaci

$$\hat{f}(x) = \frac{1}{h} K\left(\frac{x - x_1}{h}\right). \tag{1.9}$$

Wzór ten można zinterpretować następująco. Odjęcie od argumentu x wartości x_1 powoduje przesunięcie maksimum funkcji K z zera do x_1, po czym różnica $x - x_1$ jest dzielona przez dodatni parametr h w celu odpowiedniego wymodelowania kształtu tej funkcji (rys. 1.4b). Zmniejszenie wartości powyższego parametru powoduje bowiem jej „wyszczuplenie", natomiast zwiększenie – „poszerzenie". Całość jest mnożona przez współczynnik $1/h$, aby zapewnić spełnienie warunku (1.2). Funkcja K jest nazywana jądrem, w nawiązaniu do pojęcia jądra operatora całkowego.



Rys. 1.4. Intuicyjna interpretacja koncepcji estymatora jądrowego: a) jądro K, b) estymator (1.9), c) estymator (1.10)

W przypadku dwóch doświadczalnych wartości x_1 i x_2 naturalne byłoby zsumowanie funkcji (1.9) uzyskanych odrębnie dla x_1 oraz x_2 i podzielenie wyniku przez 2 w celu utrzymania warunku (1.2); (por. rys. 1.4c):

$$\hat{f}(x)=\frac{1}{2h}\left[K\left(\frac{x-x_1}{h}\right)+K\left(\frac{x-x_2}{h}\right)\right]. \tag{1.10}$$

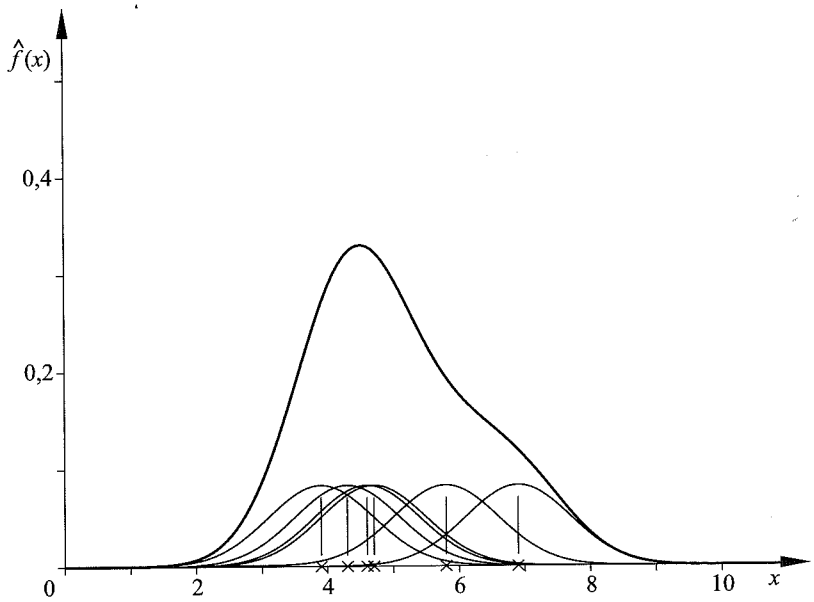
Uogólnienie powyższej koncepcji na m-elementowy zbiór $x_1, x_2, ..., x_m$ prowadzi już wprost do podstawowej postaci definicji estymatora jądrowego (rys. 1.5):

$$\hat{f}(x)=\frac{1}{mh}\sum_{i=1}^{m}K\left(\frac{x-x_i}{h}\right). \tag{1.11}$$

Warunki (1.1) i (1.2) są tu oczywiście spełnione.

Okazuje się, iż estymator jądrowy jest mało wrażliwy na ustaloną postać funkcji K. Dzięki temu istnieje możliwość takiego jej wyboru, aby spełnione były wymagania konkretnego zastosowania. Na przykład, jeżeli korzystne jest, by funkcja $\hat{f}$ miała pochodne dowolnego rzędu, to można przyjąć jądro normalne:

$$K(x)=\frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}. \tag{1.12}$$



Rys. 1.5. Estymator jądrowy

Z kolei, jeśli istnienie pochodnej nie jest wymagane, ale konieczne staje się ograniczenie zbioru, na którym estymator przyjmuje wartości dodatnie, to proponowane jest jądro Epanecznikowa:

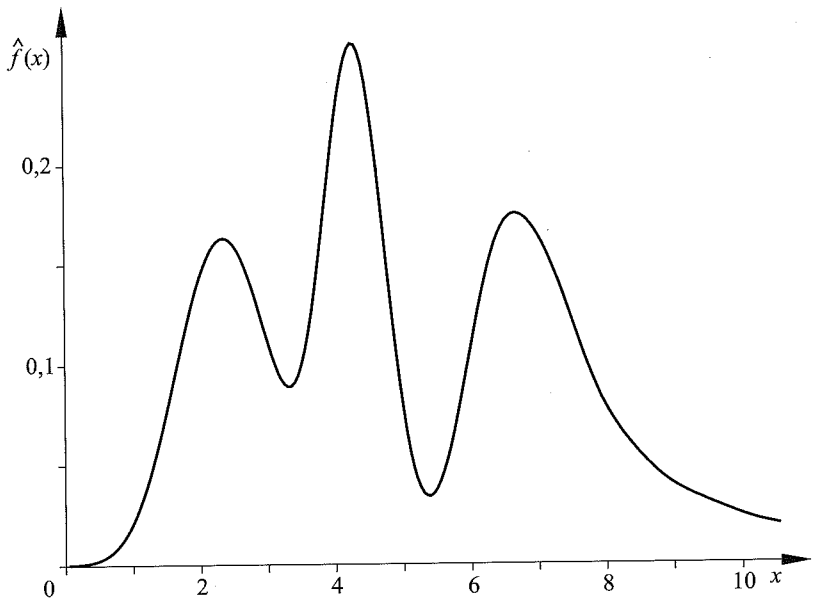
$$K(x)=\begin{cases}\frac{3}{4}(1-x^2), & \text{gdy } -1\le x\le 1,\\ 0 & \text{w przeciwnym przypadku.}\end{cases} \tag{1.13}$$

Podobnie można spełnić większość wymagań przydatnych podczas dalszej analizy rozważanego zagadnienia praktycznego.

Opracowano też wiele dodatkowych procedur, znacząco polepszających aplikacyjne własności estymatorów jądrowych. Część z nich ma ogólny charakter, inne umożliwiają dostosowanie cech estymatora do wymagań konkretnego zastosowania. Koncepcję estymatorów jądrowych nietrudno jest uogólnić na przypadek wielowymiarowych zmiennych losowych. Szczegółowo charakteryzują one wtedy złożone zależności między poszczególnymi współrzędnymi takich zmiennych. Możliwe jest także uwzględnienie współrzędnych binarnych (dwustanowych).

Warto przypomnieć, iż – zgodnie z charakterem metod nieparametrycznych – estymator jądrowy pozwala wyznaczyć gęstość bez arbitralnego ograniczenia do jakiegokolwiek typowego rozkładu probabilistycznego. Umożliwia to określenie własności populacji opisywanej przez badaną zmienną losową, w szczególności rozmieszczenia wartości modalnych (tj. lokalnych maksimów funkcji gęstości f), symetrii związanych z nimi poszczególnych składników, a także cech „ogonów", czyli właściwości funkcji f dla skrajnych wartości argumentu x. Co więcej, informacje te uzyskuje się najczęściej bez dodatkowych, żmudnych i niejednoznacznych procedur testujących. Na rysunku 1.6 jest pokazany estymator jądrowy uzyskany dla przykładowej złożonej struktury danych na podstawie 200-elementowego zbioru $x_1, x_2, \ldots, x_{200}$. Łatwo można wskazać istnienie i rozmieszczenie trzech wartości modalnych, symetrię składników lewego i środkowego oraz niesymetrię prawego. Lewy „ogon" jest wyraźnie „lżejszy" (tj. „cieńszy") od prawego.

Wyznaczenie estymatora jądrowego na podstawie danych otrzymanych z rzeczywistych obiektów technicznych, ekonomicznych lub socjologicznych umożliwia uzyskanie wniosków ilościowych odnoszących się do struktury owych danych, co z kolei może być dogodną podstawą do formułowania interpretacji opisowych. Głównie dotyczy to analizy wartości modalnych – zazwyczaj każda z nich jest bowiem związana z odmiennym aspektem rozważanego zagadnienia, co często pozwala zdekomponować przedmiotowe badania. Jeżeli na przykład jest analizowana zapadalność na określoną chorobę, to istnienie dodatkowej – względem ogólnej, typowej charakterystyki – wartości modalnej umożliwia wyodrębnienie specyficznej w rozważanych warunkach, dodatkowej



Rys. 1.6. Estymator jądrowy dla złożonej struktury danych, przy $m = 200$

przyczyny, co w konsekwencji daje podstawę do podjęcia odpowiedniego przeciwdziałania. Praktyczne wnioski można również otrzymać, rozważając „ogony" estymatora jądrowego. I tak na przykład, jeżeli badana zmienna losowa reprezentuje grubość blachy uzyskanej w procesie walcowania, to „lekkie" ogony wskazują raczej na zwykły rozrzut technologiczny, natomiast „ciężkie" świadczą o zużyciu sprzętu i konieczności wymiany niesprawnych podzespołów. Jeszcze bardziej zróżnicowane wnioski dotyczące struktury danych uzyskanych z rzeczywistych obiektów można otrzymać na podstawie analizy estymatora jądrowego wyznaczonego dla wielowymiarowej zmiennej losowej. Pojawiają się tu bowiem dodatkowe informacje dotyczące wszechstronnych zależności między poszczególnymi współrzędnymi takiej zmiennej.

Koncepcja estymatorów jądrowych została powyżej sformułowana w odniesieniu do zagadnienia wyznaczania gęstości prawdopodobieństwa, z uwagi na wyjątkowo – w tym przypadku – bliską intuicji interpretację. Nietrudno jednak tę koncepcję uogólnić na charakterystyki probabilistyczne odmienne od gęstości, na przykład dystrybuantę lub kwantyl, czy też ujęcie warunkowe stosowane, gdy badany system jest poddany działaniu podlegającego dodatkowym pomiarom czynnika zewnętrznego.

Zagadnienia zaprezentowane pokrótce w niniejszym wstępie, wraz z przykładowymi możliwościami zastosowań estymatorów jądrowych do zagadnień analizy systemowej, będą przedmiotem szczegółowych rozważań w kolejnych rozdziałach niniejszej monografii.

Uwagi bibliograficzne

Przegląd problematyki oraz metod klasycznej i współczesnej analizy systemowej można znaleźć w pracach [1, 65]. Obszerne, ilustracyjne wprowadzenie do zagadnień estymacji nieparametrycznej zostało przedstawione w monografii [62].

Rozdział 2

Podstawy matematyczne

W niniejszym rozdziale zostaną wprowadzone podstawowe pojęcia matematyczne wykorzystywane w dalszej części książki. W podrozdziale 2.1 są przedstawione zasadnicze definicje probabilistyki opartej na teorii miary, a w podrozdziale 2.2 główne pojęcia teorii estymacji. Materiał jest podany w zwięzłej formie, gdyż Czytelnikowi zainteresowanemu metodami nieparametrycznymi rozważane tu zagadnienia są w większości znane, a ponadto istnieje wyjątkowo obszerna, wartościowa literatura z zakresu powyższych dziedzin. Zamieszczenie tych dwóch podrozdziałów ma na celu jedynie uściślenie stosowanej terminologii i notacji, a także podanie kompletu wzorów wykorzystywanych w następnych rozdziałach książki. Z kolei podrozdział 2.3 został poświęcony rozwiązaniom równań różniczkowych z nieciągłą prawą stroną. Temat ten jest jak dotychczas praktycznie nieobecny w literaturze polskojęzycznej, a jego znajomość będzie użyteczna w rozdziale 4.

Symbole $\mathbb{N}$, $\mathbb{R}$ oznaczają w niniejszej monografii, odpowiednio, zbiory liczb naturalnych i rzeczywistych (w szczególności przyjmuje się, iż zero jest liczbą naturalną).

2.1. Elementy probabilistyki

W zagadnieniach współczesnej nauki oraz badań stosowanych coraz powszechniej są uwzględniane – wszechobecne w otaczającej nas rzeczywistości – zjawiska o charakterze przypadkowym (losowym). Co więcej, pojawiają się koncepcje, w których złożone procesy deterministyczne traktuje się jako losowe, z uwagi na brak możliwości ich praktycznej identyfikacji w ujęciu deterministycznym. Część procesów losowych może być stosunkowo łatwo sformalizowana na drodze matematycznej (np. kursy walutowe, temperatura otoczenia, wynik gry w ruletkę), jednak wiele innych w ogóle nie ma reprezentacji liczbo-

wej, a procedurę formalizacji niezwykle utrudnia olbrzymia różnorodność tego typu procesów.

W celu stworzenia jednolitego opisu zjawisk przypadkowych wprowadzono zasadnicze pojęcia zbioru zdarzeń elementarnych i zmiennej losowej. Zbiór zdarzeń elementarnych jest pojęciem pierwotnym probabilistyki, niedefiniowalnym. Interpretuje się go jako zbiór wszystkich możliwych wyników rozważanego zjawiska przypadkowego. W praktycznych rozważaniach korzystniejsze jest jednak zastąpienie poszczególnych (często abstrakcyjnych) zdarzeń elementarnych liczbami rzeczywistymi albo n-wymiarowymi wektorami, które reprezentowałyby owe zdarzenia stosownie do uwarunkowań rozważanego problemu. Spełniająca to zadanie funkcja jest nazywana zmienną losową i stanowi główne pojęcie probabilistyki. Dzięki takiemu ujęciu uciążliwość rozważań i obliczeń może być przeniesiona z mało konkretnego zbioru zdarzeń elementarnych na znacznie dogodniejszą do takich celów przestrzeń $\mathbb{R}$ lub ogólniej $\mathbb{R}^n$.

Zagadnienia te zostaną zaprezentowane poniżej w ścisłej, aczkolwiek skrótowej formie. Obszerny i obrazowy opis można znaleźć w licznych i zróżnicowanych publikacjach, wyszczególnionych w uwagach bibliograficznych na końcu niniejszego podrozdziału.

2.1.1. Podstawowe definicje

Niech Ω będzie zbiorem niepustym, tradycyjnie określanym mianem **zbioru zdarzeń elementarnych**. Ustalona rodzina jego podzbiorów Σ jest nazywana **σ-algebrą**, gdy spełnia następujące warunki:

$$\varnothing \in \Sigma, \tag{2.1}$$

$$V \in \Sigma \Rightarrow \Omega \setminus V \in \Sigma, \tag{2.2}$$

$$(V_i \in \Sigma \text{ dla } i = 1, 2, \ldots) \Rightarrow \bigcup_{i=1}^{\infty} V_i \in \Sigma. \tag{2.3}$$

Z kolei jako **miarę probabilistyczną** definiuje się odwzorowanie $P : \Sigma \to [0, 1]$, jeśli

$$P(\Omega) = 1, \tag{2.4}$$

$$(V_i \in \Sigma \text{ dla } i = 1, 2, \ldots \text{ oraz } V_i \cap V_j = \varnothing \text{ gdy } i \neq j) \Rightarrow$$

$$\Rightarrow P\left(\bigcup_{i=1}^{\infty} V_i\right) = \sum_{i=1}^{\infty} P(V_i). \tag{2.5}$$

Powyższa koncepcja jest naturalnym uogólnieniem klasycznego pojęcia prawdopodobieństwa opartego na zbiorach skończonych i kombinatoryce. Poszczególne elementy σ-algebry Σ reprezentują **zdarzenia**, a wartości miary P są tradycyjnie nazywane **prawdopodobieństwem**.

Zbiór Ω wraz z wyróżnioną w nim σ-algebrą Σ oraz określoną na niej miarą probabilistyczną P stanowi **przestrzeń probabilistyczną** (Ω, Σ, P). Jeśli odwzorowanie $X: \Omega \to \mathbb{R}^n$ jest funkcją mierzalną, czyli gdy przeciwobraz dowolnego zbioru otwartego należy do σ-algebry:

$$X^{-1}(U) \in \Sigma \quad \text{dla każdego zbioru otwartego } U \subset \mathbb{R}^n, \tag{2.6}$$

to nazywa się je **zmienną losową**. Pojęcie to można uściślić jako ***n*-wymiarową zmienną losową**, a więc dla $n = 1$ mamy do czynienia z **jednowymiarową (rzeczywistą) zmienną losową**, natomiast jeśli $n \geq 2$ z **wielowymiarową zmienną losową**.

Rozważana teraz będzie przestrzeń $\mathbb{R}^n$. Najmniejsza (w sensie relacji inkluzji) σ-algebra zawierająca zbiory otwarte nazywana jest **σ-algebrą zbiorów borelowskich** i oznaczana przez $\mathcal{B}^n$, a jej elementy określa się mianem **zbiorów borelowskich**. Rodzina $\mathcal{B}^n$ jest bardzo „bogata" – zbiorami borelowskimi są między innymi poszczególne punkty z przestrzeni $\mathbb{R}^n$, dowolne zbiory otwarte i domknięte oraz ich przeliczalne mnogościowe sumy, różnice i iloczyny. Odwzorowanie $f: \mathbb{R}^{n_1} \to \mathbb{R}^{n_2}$ jest nazywane **funkcją borelowską**, jeżeli przeciwobraz każdego zbioru borelowskiego jest zbiorem borelowskim:

$$f^{-1}(B) \in \mathcal{B}^{n_1} \quad \text{dla każdego} \quad B \in \mathcal{B}^{n_2}. \tag{2.7}$$

Każda funkcja (przedziałami) ciągła jest borelowska. Obłożenie zmiennej losowej X funkcją borelowską f, a zatem złożenie $f \circ X$ również jest zmienną losową. Warunek (2.6) jest równoważny postaci

$$X^{-1}(B) \in \Sigma \quad \text{dla każdego zbioru borelowskiego} \quad B \in \mathcal{B}^n. \tag{2.8}$$

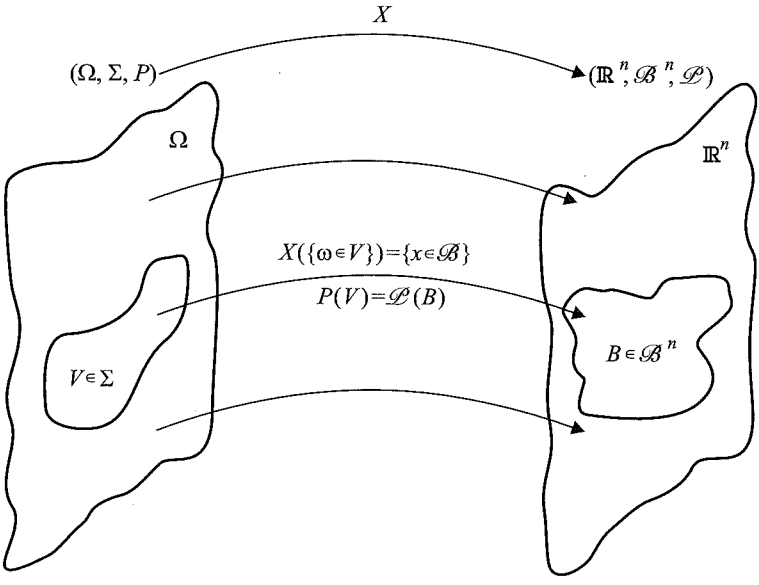
Z kolei n-wymiarowa zmienna losowa X indukuje przestrzeń probabilistyczną $(\mathbb{R}^n, \mathcal{B}^n, \mathcal{P})$, gdzie miara probabilistyczna $\mathcal{P}: \mathcal{B}^n \to [0,1]$ jest zdefiniowana zależnością

$$\mathcal{P}(B) = P(\{\omega \in \Omega : X(\omega) \in B\})$$
$$\text{dla każdego zbioru borelowskiego} \quad B \in \mathcal{B}^n \tag{2.9}$$

i określana mianem **rozkładu prawdopodobieństwa**.

Teraz, po sformułowaniu podstawowych definicji probabilistyki, można uściślić opisową **interpretację** przedstawioną na początku tego podrozdziału. Otóż, wszystkie możliwe wyniki badanego zjawiska przypadkowego (niekoniecznie wyrażalne w postaci liczbowej), tradycyjnie nazywane zdarzeniami elementarnymi, są zgrupowane w zbiorze zdarzeń elementarnych Ω. W zbiorze tym wyróżnia się rodzinę jego podzbiorów Σ, określanych mianem zdarzeń. Każde zdarzenie jest charakteryzowane przez wartość miary probabilistycznej P, reprezentującej prawdopodobieństwo jego występowania w praktyce. Powyższe elementy składają się na przestrzeń probabilistyczną (Ω, Σ, P). Definicje Σ oraz P zawierają warunki zapewniające zgodność podstawowych własności matematycznych z empirycznymi cechami realnych zjawisk przypadkowych.

W zagadnieniach praktycznych ścisłe określenie przestrzeni probabilistycznej (Ω, Σ, P) może okazać się bardzo trudne lub – w ogólnym przypadku – wręcz niewykonalne. Zdarzenia elementarne $\omega \in \Omega$ mogą być bowiem obiektami niemającymi interpretacji liczbowej, na przykład wrażenia estetyczne, samopoczucie, zapach. Zostało zatem wprowadzone pojęcie zmiennej losowej, czyli funkcji, która każdemu zdarzeniu elementarnemu przyporządkowuje wektor n-wymiarowy (w szczególnym przypadku $n = 1$ – liczbę rzeczywistą). Warunek (2.9) – zilustrowany na rysunku 2.1 – sprawia, że informacje, jakie mogłyby być uzyskane z przestrzeni probabilistycznej (Ω, Σ, P), można w praktyce wyznaczyć z przestrzeni indukowanej $(\mathbb{R}^n, \mathscr{B}^n, \mathscr{P})$, znacznie dogodniejszej do rozważań teoretycznych i obliczeń praktycznych.



Rys. 2.1. Ilustracja zależności między przestrzeniami (Ω, Σ, P) oraz $(\mathbb{R}^n, \mathscr{B}^n, \mathscr{P})$

W badaniach aplikacyjnych rozważane zjawisko przypadkowe jest zatem reprezentowane przez zmienną losową, którą charakteryzuje się za pomocą wymiaru n oraz rozkładu prawdopodobieństwa $\mathscr{P}$.

Parametr n, czyli wymiar zmiennej losowej, określa liczbę składowych reprezentujących poszczególne aspekty rozpatrywanego zjawiska opisywanego przez zmienną losową; na przykład dla przypadku zagadnień meteorologicznych mogą nimi być temperatura zewnętrzna, wilgotność powietrza i ciśnienie atmosferyczne. W praktycznych zastosowaniach identyfikacja tego parametru zwykle wynika wprost ze specyfiki badanego zagadnienia, a najczęściej $n = 1$. Jednak w nietrywialnych współczesnych zastosowaniach wybór wartości parametru n stanowi swego rodzaju kompromis między naturalną skłonnością do zwiększania wartości n, co pozwoliłoby uwzględnić dodatkowe aspekty problemu, a koniecznością jej zmniejszania w celu redukcji złożoności zadania.

Rozkład prawdopodobieństwa $\mathscr{P}$ specyfikuje się w praktyce, wyznaczając jego odpowiednie charakterystyki funkcyjne, które umożliwiają później stosunkowo proste obliczenie wartości poszczególnych prawdopodobieństw, a także mogą być podstawą do dalszej wszechstronnej analizy. Podstawowe przykłady stanowią tutaj gęstość prawdopodobieństwa i dystrybuanta. W prostszych zadaniach wystarczające okazuje się podanie jedynie charakterystycznych parametrów rozkładu. Najczęściej są to momenty (m.in. wartość oczekiwana i wariancja) lub kwantyle (m.in. mediana). Pojęcia te zostaną pokrótce przedstawione w kolejnych dwóch punktach.

2.1.2. Charakterystyki funkcyjne rozkładu prawdopodobieństwa

Jeżeli rozkład prawdopodobieństwa $\mathscr{P}$ spełnia warunek

$$\mathscr{P}(B) = 0 \quad \text{dla każdego} \quad B \in \mathscr{B}^n, \quad \text{takiego że} \quad m(B) = 0, \tag{2.10}$$

gdzie m oznacza n-wymiarową miarę Lebesgue'a, to istnieje wtedy gęstość tego rozkładu, zwana **gęstością prawdopodobieństwa**, czyli mierzalna funkcja $f: \mathbb{R}^n \to [0, \infty)$, taka że

$$\mathscr{P}(B) = \int_B f(x)\,dx \quad \text{dla każdego} \quad B \in \mathscr{B}^n. \tag{2.11}$$

Gęstość prawdopodobieństwa dostarcza aparatu matematycznego dogodnego zarówno do teoretycznych rozważań, jak i do praktycznych obliczeń oraz ilustracji wyników. Zgodnie ze wzorem (2.11) gęstość określa wartość miary $\mathscr{P}$ dla każdego n-wymiarowego zbioru borelowskiego, co umożliwia wyznaczenie praw-

dopodobieństwa w zasadzie dowolnego zdarzenia, a niekiedy także wieloaspektową analizę. Interpretacja wykresu funkcji gęstości jest bliska intuicji – im większa jest wartość tej funkcji w danym punkcie, tym większe jest prawdopodobieństwo przyjęcia przez badaną zmienną losową wartości z otoczenia tego punktu. Dzięki temu gęstość prawdopodobieństwa wyjątkowo dobrze nadaje się do ilustrowania otrzymanych wyników i zapewnia ich zrozumienie nawet przez osoby niemające matematycznego doświadczenia.

Wartość, dla której funkcja gęstości prawdopodobieństwa przyjmuje ścisłe maksimum lokalne, nazywa się **wartością modalną**. Jeżeli istnieje jedna taka wartość, to rozkład nazywa się **unimodalnym** (jednomodalnym), a w przeciwnym przypadku – **wielomodalnym**.

Inną istotną charakterystykę gęstości prawdopodobieństwa stanowi jej **nośnik**, czyli domknięcie tych punktów dziedziny, w których funkcja gęstości przyjmuje wartości dodatnie. Jest to zatem najmniejszy (w sensie relacji inkluzji) zbiór domknięty, którego miara probabilistyczna $\mathscr{P}$ wynosi 1. Nośnik gęstości prawdopodobieństwa można zatem interpretować jako zbiór tych wartości, które zmienna losowa przyjmuje z prawdopodobieństwem większym od zera. Dlatego też często kojarzy się go nie tyle z gęstością, co bezpośrednio z samą zmienną losową i nazywa **nośnikiem zmiennej losowej**. Podstawowymi cechami nośnika gęstości są jego **ograniczoność** i **spójność**[1)].

Praktyczne doświadczenie uzyskiwane przy stosowaniu metod probabilistycznych pozwoliło wyróżnić najbardziej typowe postacie rozkładów prawdopodobieństwa. Określa się je przez podanie konkretnych postaci gęstości prawdopodobieństwa. Niektóre typowe rozkłady, wykorzystywane w dalszej części niniejszej monografii, są przedstawione w formie stabelaryzowanej na rysunku 2.2. Nawet gdyby uzupełnić zamieszczony tam zestaw o kilka dalszych pozycji, to możliwość wyboru jest ograniczona w stopniu trudnym do zaakceptowania z punktu widzenia współczesnych problemów aplikacyjnych o złożonej różnorodności modelowanych zjawisk. Obecnie możliwe stało się wprowadzenie metod uniwersalnych, niezależnych od arbitralnych, klasycznych postaci. Ponieważ punktem wyjścia prezentowanej tu koncepcji estymatorów jądrowych jest zadanie estymacji gęstości prawdopodobieństwa, **w dalszej części tej książki przyjmuje się, iż warunek (2.10) jest zawsze spełniony i gęstość prawdopodobieństwa istnieje**.

Pokrewną gęstości charakterystykę funkcyjną rozkładu prawdopodobieństwa stanowi **dystrybuanta**, czyli odwzorowanie $F : \mathbb{R}^n \to [0, 1]$ dane wzorem

---

[1)] Niech X będzie przestrzenią topologiczną. Zbiór $C \subset X$ jest **spójny**, jeżeli przy każdym jego podziale na dwa zbiory niepuste C_1 oraz C_2, czyli gdy $C = C_1 \cup C_2$, spełniony jest warunek $(\overline{C_1} \cap C_2) \cup (C_1 \cap \overline{C_2}) = \varnothing$. Powyższe można zinterpretować tak, iż zbiór spójny składa się tylko z jednej „części".

Rozkład	Gęstość prawdopodobieństwa	Założenia
Normalny standardowy	$f(x)=\frac{1}{\sqrt{2\pi}}\exp\left(-\frac{x^2}{2}\right)$	
Normalny	$f(x)=\frac{1}{\sqrt{2\pi\sigma}}\exp\left(-\frac{(x-m_1)^2}{2\sigma^2}\right)$	$\sigma>0$ $m_1\in\mathbb{R}$
Jednostajny	$f(x)=\begin{cases}\frac{1}{b-a}, & \text{gdy } x\in[a,b]\\ 0, & \text{gdy } x\notin[a,b]\end{cases}$	$a<b$
Trójkątny	$f(x)=\begin{cases}\frac{2}{(b-a)(c-a)}(x-a), & \text{gdy } x\in[a,c]\\ \frac{2}{(b-a)(c-b)}(x-b), & \text{gdy } x\in(c,b]\\ 0, & \text{gdy } x\notin[a,b]\end{cases}$	$a<c<b$
Beta	$f(x)=\begin{cases}\frac{\Gamma(p+q)}{\Gamma(p)+\Gamma(q)}x^{p-1}(1-x)^{q-1}, & \text{gdy } x\in[0,1]\\ 0, & \text{gdy } x\notin[0,1]\end{cases}$	$p>0$ $q>0$
Gamma	$f(x)=\begin{cases}\frac{q^p}{\Gamma(p)}x^{p-1}e^{-qx}, & \text{gdy } x\in(0,\infty)\\ 0, & \text{gdy } x\notin(0,\infty)\end{cases}$	$p>0$ $q>0$
Wykładniczy	$f(x)=\begin{cases}\lambda e^{-\lambda x}, & \text{gdy } x\in[0,\infty)\\ 0, & \text{gdy } x\notin[0,\infty)\end{cases}$	$\lambda>0$
Funkcja Eulera $\Gamma:(0,\infty)\to\mathbb{R}$ jest dana wzorem $\Gamma(x)=\int_0^\infty e^{-y}y^{x-1}\,dy$. Dla liczb postaci $x=0{,}5;\ 1;\ 1{,}5;\ 2;\ \ldots$ jej wartość można łatwo wyznaczyć z zależności rekurencyjnej $\Gamma(x+1)=x\Gamma(x)$, doprowadzając do $\Gamma(0{,}5)=\sqrt{\pi}$ albo $\Gamma(1)=\Gamma(2)=1$.		

Rys. 2.2. Gęstość prawdopodobieństwa typowych rozkładów jednowymiarowych

$$F(x)=\mathcal{P}(\{x: X_1(x)<x_1, X_2(x)<x_2, \ldots, X_n(x)<x_n\}), \tag{2.12}$$

gdzie X_1, X_2, ..., X_n oraz x_1, x_2, ..., x_n oznaczają poszczególne współrzędne zmiennej losowej X oraz wektora x. Prawdziwa jest następująca zależność:

$$F(x)=\int_{-\infty}^{x_1}\int_{-\infty}^{x_2}\ldots\int_{-\infty}^{x_n} f(u)\,du_n\ldots du_2\,du_1, \tag{2.13}$$

przy czym dodatkowo $u_1, u_2, ..., u_n$ oznaczają współrzędne wektora u. W przypadku jednowymiarowej zmiennej losowej wzór ten przyjmuje postać

$$F(x) = \int_{-\infty}^{x} f(u)\,du. \tag{2.14}$$

Dystrybuanta nie stanowi tak dogodnego do obliczeń aparatu matematycznego jak gęstość prawdopodobieństwa. Jej interpretacja nie jest również na tyle bliska intuicji, aby mogła być powszechnie używana do ilustracji wyników. Jednak – w przeciwieństwie do gęstości – dystrybuanta istnieje także w tych punktach, w których warunek (2.10) nie jest spełniony (aczkolwiek jest w nich nieciągła).

2.1.3. Charakterystyczne parametry rozkładu prawdopodobieństwa

Prostszymi od funkcyjnych charakterystykami rozkładu prawdopodobieństwa są jego parametry. Generalnie należą one do dwóch rodzajów, opartych na pojęciach momentów i kwantyli.

Najpierw zostanie rozważona jednowymiarowa zmienna losowa. Niech $k \in \mathbb{N} \setminus \{0\}$ będzie dowolnie ustalone. Zakłada się wtedy, iż

$$\int_{-\infty}^{\infty} x^k f(x)\,dx < \infty, \tag{2.15}$$

gdzie f oznacza konsekwentnie gęstość prawdopodobieństwa. **Momentem zwykłym rzędu *k*** jest nazywana wielkość $m_k \in \mathbb{R}$ zdefiniowana następująco:

$$m_k = \int_{-\infty}^{\infty} x^k f(x)\,dx. \tag{2.16}$$

Szczególne znaczenie ma moment zwykły rzędu pierwszego, czyli **wartość oczekiwana**:

$$E = m_1 = \int_{-\infty}^{\infty} x f(x)\,dx. \tag{2.17}$$

Moment odniesiony do wartości oczekiwanej definiuje się wzorem

$$\mu_k = \int_{-\infty}^{\infty} (x-E)^k f(x)\,dx \tag{2.18}$$

i określa mianem **momentu centralnego rzędu *k***. Szczególną rolę odgrywa w tym przypadku **wariancja** – moment centralny rzędu drugiego:

$$V = \mu_2 = \int_{-\infty}^{\infty} (x - E)^2 f(x)\, dx, \tag{2.19}$$

a także jej pierwiastek kwadratowy, czyli **odchylenie standardowe**:

$$\sigma = \sqrt{V}. \tag{2.20}$$

Wartość oczekiwana wskazuje „średnią ważoną" uzyskiwanych w praktyce wartości zmiennej losowej. Stopień ich „rozrzutu" względem tej średniej reprezentuje odchylenie standardowe (i pośrednio wariancja). Warunek (2.10) zapewnia, iż odchylenie standardowe i wariancja są większe od zera.

Zostaną teraz rozważone momenty dla przypadku wielowymiarowej zmiennej losowej. Niech ustalonych będzie n liczb naturalnych k_1, k_2, ..., k_n, nie wszystkie równe zeru. Zakłada się również, że

$$\int_{\mathbb{R}^n} x_1^{k_1} x_2^{k_2} \ldots x_n^{k_n} f(x)\, dx < \infty, \tag{2.21}$$

gdzie x_1, x_2, ..., x_n oznaczają poszczególne współrzędne wektora x. **Momentem zwykłym rzędu $k_1, k_2, \ldots, k_n$** jest nazywana wielkość $\mu_{k_1, k_2, \ldots, k_n} \in \mathbb{R}$ zdefiniowana wzorem

$$\mu_{k_1, k_2, \ldots, k_n} = \int_{\mathbb{R}^n} x_1^{k_1} x_2^{k_2} \ldots x_n^{k_n} f(x)\, dx. \tag{2.22}$$

W szczególności, jeśli przyjąć $k_i = 1$ dla dowolnie ustalonego i oraz zero dla pozostałych indeksów, to moment taki reprezentuje **wartość oczekiwaną i-tej współrzędnej**:

$$E_i = m_{0,\ldots,0,1,0,\ldots,0} = \int_{\mathbb{R}^n} x_i f(x)\, dx. \tag{2.23}$$

Z kolei wielkość

$$\mu_{k_1, k_2, \ldots, k_n} = \int_{\mathbb{R}^n} (x_1 - E_1)^{k_1} (x_2 - E_2)^{k_2} \ldots (x_n - E_n)^{k_n} f(x)\, dx \tag{2.24}$$

określa się mianem **momentu centralnego rzędu $k_1, k_2, \ldots, k_n$**. W przypadku $k_i = 2$ dla dowolnie ustalonego i oraz zero dla pozostałych indeksów moment ten oznacza **wariancję i-tej współrzędnej**:

$$V_i = \mu_{0,\ldots,0,2,0,\ldots,0} = \int_{\mathbb{R}^n} (x_i - E_i)^2 f(x)\, dx, \tag{2.25}$$

a jej pierwiastek stanowi **odchylenie standardowe *i*-tej współrzędnej**:

$$\sigma_i = \sqrt{V_i}. \tag{2.26}$$

Również tutaj warunek (2.10) implikuje dodatniość V_i oraz σ_i dla każdego $i = 1, 2, ..., n$. Jeżeli natomiast przyjmie się $k_i = 1$ i $k_j = 1$ dla dowolnych $i \neq j$, a także zero dla pozostałych indeksów, to moment taki jest określany mianem **kowariancji *i*-tej oraz *j*-tej współrzędnej**:

$$cov_{i,j} = \mu_{0,...,0,1,0,...,0,1,0,...,0} = \int_{\mathbb{R}^n} (x_i - E_i)(x_j - E_j) f(x)\, dx. \tag{2.27}$$

Wartość $cov_{i,j}$ „przeskalowana" zgodnie ze wzorem

$$\rho_{i,j} = \frac{cov_{i,j}}{\sigma_i \sigma_j} \tag{2.28}$$

stanowi **współczynnik korelacji *i*-tej oraz *j*-tej współrzędnej**. Wartość $\rho_{i,j}$ należy do przedziału [–1, 1]. Najpierw będzie rozważany przypadek $\rho_{i,j} > 0$: jeżeli *i*-ta współrzędna zmiennej losowej przyjmie wartość większą (lub mniejszą) niż jej wartość oczekiwana, to można przypuszczać, iż *j*-ta współrzędna także przyjmie wartość większą (lub odpowiednio mniejszą) od swojej wartości oczekiwanej. Im wartość $\rho_{i,j}$ jest większa, tym przypuszczenie to jest bardziej wiarygodne. W skrajnym przypadku, gdy $\rho_{i,j} = 1$, współrzędne te są liniowo zależne, czyli $X_i = aX_j + b$ z prawdopodobieństwem 1 dla $a, b \in \mathbb{R}$ oraz $a \neq 0$. Analogicznie można interpretować przypadek $\rho_{i,j} < 0$. Jeżeli $\rho_{i,j} \neq 0$, to współrzędne *i*-ta oraz *j*-ta są **skorelowane**. Jeśli natomiast $\rho_{i,j} = 0$, to współrzędne te nazywa się **nieskorelowanymi** – wówczas z relacji między wartością *i*-tej współrzędnej zmiennej losowej a jej wartością oczekiwaną nie można nic domniemywać o analogicznej relacji dotyczącej *j*-tej współrzędnej.

I wreszcie, porównując zależności (2.25) i (2.27), w naturalny sposób można dodatkowo zdefiniować

$$cov_{i,i} = V_i \quad \text{dla } i = 1, 2, ..., n. \tag{2.29}$$

Wtedy macierz *COV* o wymiarze $n \times n$ dana jako

$$COV = [cov_{i,j}], \tag{2.30}$$

gdzie poszczególne elementy $cov_{i,j}$ zostały określone wzorami (2.27) i (2.29) z podstawieniem (2.25), nazywa się **macierzą kowariancji**. Jest ona symetryczna, dodatnio półokreślona, a także dodatnio określona, gdy współrzędne zmiennej losowej nie są liniowo zależne (co w praktyce jest spełnione, gdyż liniowo zależne współrzędne – jako niezawierające użytecznej informacji – należy wy-

eliminować podczas wstępnej fazy rozważań w celu zawsze pożądanego uproszczenia modelu).

Alternatywą wobec przedstawionej powyżej tradycyjnej rodziny parametrów opartych na koncepcji momentów są parametry związane z pojęciem kwantyla.

Najpierw zostanie rozważony przypadek jednowymiarowej zmiennej losowej. Niech dana będzie liczba r należąca do przedziału (0, 1). **Kwantylem rzędu *r*** jest nazywana każda liczba $q \in \mathbb{R}$, taka że

$$\int_{-\infty}^{q} f(x)\,dx = r. \tag{2.31}$$

Spójność nośnika gęstości prawdopodobieństwa implikuje jednoznaczność kwantyla. Kwantyl rzędu 0,5 jest nazywany **medianą** i ma podobne znaczenie jak wartość oczekiwana. Rolę odchylenia standardowego odgrywa wśród kwantyli tak zwane **odchylenie ćwiartkowe**, czyli różnica między kwantylami rzędu 0,75 i 0,25.

W przypadku wielowymiarowej zmiennej losowej, dla dowolnie ustalonego indeksu $i = 1, 2, ..., n$ oraz liczby $r_i \in (0, 1)$, w sposób naturalny definiuje się **kwantyl rzędu r_i *i*-tej współrzędnej**. Jest nim każda liczba $q_i \in \mathbb{R}$ spełniająca równanie

$$\int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} \int_{-\infty}^{q_i} \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} f(x)\,dx = r_i, \tag{2.32}$$

przy czym w powyższym zapisie całka $\int_{-\infty}^{q_i}$ dotyczy i-tej współrzędnej. Analogicznie jak w przypadku jednowymiarowym określa się **medianę** oraz **odchylenie ćwiartkowe *i*-tej współrzędnej**.

Analiza porównawcza momentów i kwantyli nie wykazuje bezwzględnej przewagi jednych nad drugimi, ale wydaje się, iż jej wyniki świadczą jednak na korzyść kwantyli. Lepiej opisują one rzeczywistość, zwłaszcza w przypadkach rozkładów nietypowych, na przykład przy silnie niesymetrycznej funkcji gęstości prawdopodobieństwa. Kwantyle istnieją także wtedy, gdy całki występujące w warunkach (2.15) oraz (2.21) są nieskończone. I wreszcie, szczególnie warta podkreślenia jest możliwość dowolnej modyfikacji rzędu kwantyla, czyli stałej r, w zależności od specyfiki badanego problemu. Rozpatrzmy przykładowo medianę, czyli kwantyl rzędu $r = 0{,}5$. Otóż, jeżeli w konkretnym praktycznym zagadnieniu, przeszacowanie i niedoszacowanie jej wartości implikują odmienne skutki ekonomiczne, to można stosownie zmniejszyć lub powiększyć – względem wartości 0,5 – rząd kwantyla r, tak aby zredukować ten błąd (tj. przeszacowanie lub niedoszacowanie), który implikuje większe straty. Możliwości takich nie ma wartość oczekiwana. Jednakże momenty są powszechnie znane i do-

stępne w literaturze. Wartości momentów są także łatwiejsze do wyznaczenia w praktyce, jednak gwałtowny rozwój metod wspomaganych techniką komputerową sukcesywnie pomniejsza znaczenie tej własności.

2.1.4. Niezależność zmiennych losowych

Rozważane będą n-wymiarowe zmienne losowe X_1, X_2, ..., X_k określone na wspólnej przestrzeni probabilistycznej (Ω, Σ, P). Jeżeli jest spełniony warunek

$$P(\{\omega \in \Omega : X_1(\omega) \in B_1, X_2(\omega) \in B_2, \ldots, X_k(\omega) \in B_k\}) =$$
$$= \prod_{j=1}^{k} P(\{\omega \in \Omega : X_j(\omega) \in B_j\}) \quad \text{dla dowolnych} \quad B_1, B_2, \ldots, B_k \in \mathcal{B}^n, \quad (2.33)$$

to są one nazywane **zmiennymi niezależnymi**. Powyższe można także wyrazić za pomocą charakterystyk rozkładu probabilistycznego. Z uwagi na przyjęte wcześniej założenie o spełnieniu warunku (2.10) definicja (2.33) jest bowiem równoważna zależności

$$f(x_1, x_2, \ldots, x_k) = f_1(x_1) f_2(x_2) \ \ldots \ f_k(x_k)$$
$$\text{dla każdego} \quad x_1, x_2, \ldots, x_k \in \mathbb{R}^n, \quad (2.34)$$

przy czym f oznacza gęstość prawdopodobieństwa zestawienia zmiennych X_1, X_2, ..., X_k, czyli $(k \cdot n)$-wymiarowej zmiennej losowej

$$\begin{bmatrix} X_1 \\ X_2 \\ \vdots \\ X_k \end{bmatrix}, \quad (2.35)$$

a $f_1, f_2, \ldots, f_k$ oznaczają odpowiednio gęstości zmiennych $X_1, X_2, \ldots, X_k$.

Interpretacja niezależności zmiennych losowych wskazuje, iż znajomość wartości, jakie przyjęło dowolne $k-1$ spośród zmiennych $X_1, X_2, \ldots, X_k$, nie wnosi żadnej informacji o wartości przyjętej przez pozostałą.

Jeżeli X jest n-wymiarową zmienną losową, to jej współrzędne można traktować również jako n jednowymiarowych zmiennych. Wtedy interpretacje ich niezależności i przedstawionej wcześniej własności korelacji są podobne. Istotnie, jeżeli jednowymiarowe zmienne losowe są niezależne, to – jako współrzędne zmiennej X – są także nieskorelowane. Relacja przeciwna nie jest ogólnie prawdziwa. Własność niezależności jest zatem mocniejsza od nieskorelowania, ale w przypadku zmiennych zależnych przeważnie wygodniej jest rozważać

korelację, gdyż znak i wartość współczynnika korelacji dostarczają dodatkowych informacji o „kierunku" korelacji i jej intensywności.

Jeżeli poszczególne współrzędne wielowymiarowej zmiennej losowej są niezależne, to macierz kowariancji (2.30) jest diagonalna i ma na przekątnej głównej wariancje poszczególnych współrzędnych.

Uwagi bibliograficzne

Literatura dotycząca podstaw probabilistyki jest liczna i zróżnicowana zarówno w zakresie doboru tematyki, jak i sposobu prezentacji. Szczególnie wartymi rekomendacji są nowoczesne podręczniki [20, 34, 53], przy czym w drugim z nich przeważa ujęcie intuicyjne, a trzeci ma charakter ścisły. Cenne uzupełnienie mogą stanowić klasyczne publikacje [15, 18]. W postaci zmatematyzowanej materiał ten zawiera fundamentalna monografia [5]. Warta polecenia jest także pozycja o charakterze encyklopedycznym [14].

2.2. Teoria estymacji

Sama teoria probabilistyki nie wnosi zbyt wielu elementów oryginalnych, gdyż z teoretycznego punktu widzenia stanowi ona dobrze już opracowany – w ramach klasycznej analizy matematycznej – przypadek miary skończonej. Opinię taką należy jednak diametralnie zmienić, gdy rzecz dotyczy praktycznych zastosowań probabilistyki, głównie w ramach statystyki matematycznej. Wydaje się bowiem, iż nie ma obecnie działu matematyki o większych niż statystyka możliwościach wszechstronnych i różnorodnych aplikacji.

W przeciwieństwie do trywialnych procesów losowych, jak na przykład klasyczny rzut monetą lub kostką, w praktyce najczęściej nie jest możliwe teoretyczne wyznaczenie samego rozkładu prawdopodobieństwa ani jego charakterystyk – funkcyjnych lub parametrów. Dokonuje się tego bezpośrednio na podstawie danych eksperymentalnych uzyskiwanych z badanego obiektu, a ściślej, zmierzonych wartości badanej zmiennej losowej, jakie przyjęła ona dotychczas na skutek wystąpienia konkretnych zdarzeń elementarnych. Powyższe jest przedmiotem badań **teorii estymacji**, jednego z dwóch – obok testowania hipotez – podstawowych działów statystyki matematycznej.

Poniżej będą pokrótce przedstawione podstawowe pojęcia teorii estymacji, przede wszystkim w zakresie, jaki zostanie wykorzystany w dalszej części monografii. Szczegółowy i obrazowy opis można znaleźć w wyjątkowo bogatej i wartościowej literaturze tego przedmiotu, skomentowanej na końcu niniejszego podrozdziału.

2.2.1. Podstawowe definicje

Niech dana będzie przestrzeń probabilistyczna (Ω, Σ, P) i określona na niej n-wymiarowa zmienna losowa X. Jeżeli $X_1, X_2, ..., X_m$ oznaczają zmienne losowe o rozkładzie identycznym jak zmienna X, to stanowią one **próbę losową** odpowiadającą zmiennej X. Zgodnie z ogólną specyfiką pojęć statystyki matematycznej sens powyższej definicji tkwi w jej aspektach aplikacyjnych. Rozważmy więc klasyczny przykład dotyczący badań wzrostu ogółu Polaków. Zbiór zdarzeń elementarnych Ω można utożsamiać ze zbiorem wszystkich obywateli Polski płci męskiej. Zbiór ten jest skończony, a zatem σ-algebra Σ może być zdefiniowana w sposób naturalny jako zbiór wszystkich jego podzbiorów. Z kolei zmienna losowa X reprezentuje cechę będącą przedmiotem badań, a zatem w rozważanym przykładzie przyporządkowuje każdemu Polakowi jego wzrost. Z praktycznego punktu widzenia nie jest możliwe dokonanie pomiaru wzrostu wszystkich Polaków, lecz jedynie ich dostatecznie reprezentatywnego m-elementowego podzbioru i wnioskowanie – na podstawie uzyskanych tym sposobem wartości – o wzroście całej populacji. Wybór owego podzbioru jest losowy, stąd zmienne $X_1, X_2, ..., X_m$ reprezentujące uzyskiwanie poszczególnych wartości zmiennej X, są zmiennymi losowymi i, co oczywiste, o tym samym co ona rozkładzie.

Jeżeli zmienne $X_1, X_2, ..., X_m$ są ponadto niezależne, to próba taka jest nazywana **prostą próbą losową**. Ponieważ w niniejszej monografii nie rozpatruje się innych rodzajów prób, więc **prosta próba losowa będzie nazywana krótko próbą losową**. Liczba $m \in \mathbb{N} \setminus \{0\}$ wyraża **liczność próby losowej**, a sekwencja otrzymanych eksperymentalnie pomiarów

$$X_1(\omega), X_2(\omega), ..., X_m(\omega) \tag{2.36}$$

reprezentuje **wartość próby losowej**. Jeżeli w praktyce są wykorzystywane jedynie konkretne wartości próby, a ich zależność od czynnika losowego nie jest przedmiotem rozważań, to wzór (2.36), dla uproszczenia notacji, przyjmuje się jako

$$x_1, x_2, ..., x_m, \tag{2.37}$$

pomijając w zapisie zależność od czynnika ω[1)].

[1)] Warto w tym miejscu zaznaczyć odmienne ujęcie powyższego zagadnienia, powstałe w ramach tak zwanej **analizy danych**. Otóż punktem wyjścia badań przedmiotowych jest tutaj zbiór doświadczalnie uzyskanych pomiarów (2.37), a cel stanowi otrzymanie na tej podstawie odpowiednich wniosków, lecz dotyczących własności samych danych (2.37), bez nawiązania do nieznanej i przez to abstrakcyjnej zmiennej X. Ujęcie takie, aczkolwiek uproszczone, w praktyce okazuje się nierzadko całkowicie wystarczające, a jako łatwiej zrozumiałe jest powszechniej dostępne i coraz chętniej akceptowane.

Niech $X_1, X_2, ..., X_m$ będzie próbą losową. **Statystyką** nazywa się takie odwzorowanie $S: \underbrace{\mathbb{R}^n \times \mathbb{R}^n \times ... \times \mathbb{R}^n}_{(m \text{ czynników})} \to \mathbb{R}^k$, że funkcja $\mathcal{S}: \Omega \to \mathbb{R}^k$ zdefiniowana wzorem

$$\mathcal{S}(\omega) = S(X_1(\omega), X_2(\omega), ..., X_m(\omega)) \tag{2.38}$$

jest mierzalna, czyli stanowi ona tym samym k-wymiarową zmienną losową. W praktyce często spotyka się $k = 1$ lub $k = n$, a najczęściej $k = n = 1$. Każda funkcja borelowska może być zatem statystyką, jednak podobnie jak w przypadku próby losowej istota powyższej definicji zawiera się w interpretacji: statystyka jest funkcją umożliwiającą wnioskowanie na podstawie uzyskanych wartości próby losowej. Wnioskowanie to ma odmienną formę w ramach różnych działów statystyki matematycznej.

W przypadku jednego z dwóch podstawowych takich działów, mianowicie **testowania hipotez**, celem działania jest stwierdzenie, na podstawie uzyskanej eksperymentalnie próby losowej, czy weryfikowane przypuszczenie jest prawdziwe czy fałszywe. Przedmiotem badań nie jest tu sama wartość statystyki, lecz jedynie fakt, czy wartość ta należy do wyznaczonych wcześniej zbiorów świadczących odpowiednio o prawdziwości lub fałszywości hipotezy. Zagadnienia te będą przedmiotem dalszych rozważań zawartych w podrozdziale 4.2.

Z kolei celem drugiego z podstawowych działów statystyki matematycznej, czyli **teorii estymacji**, jest wyznaczenie analizowanej charakterystyki rozkładu prawdopodobieństwa. Wynik badań stanowi tu wartość statystyki (2.38), obliczona na podstawie uzyskanej eksperymentalnie próby losowej, którą to wartość traktuje się bezpośrednio jako poszukiwane oszacowanie analizowanej wielkości. W przypadku parametrów charakterystycznych reprezentuje ona wartość badanego parametru (gdy $k = 1$) lub wektora parametrów (gdy $k \geq 2$), natomiast w przypadku charakterystyk funkcyjnych – wartość funkcji (wielowymiarowej, gdy $k \geq 2$) w ustalonym punkcie dziedziny. Tak określoną statystykę nazywa się **estymatorem**. A zatem zasadnicze zadanie estymacji polega na wyznaczeniu odpowiedniej formuły estymatora, a następnie – po uzyskaniu wartości próby losowej – obliczeniu wartości tego estymatora, przyjmowanej bezpośrednio jako wartość analizowanej wielkości: parametru lub wartości funkcji dla ustalonego argumentu[1)].

Jeżeli zależność wartości estymatora od próby losowej nie jest bezpośrednim przedmiotem rozważań, to zapis $S(X_1(\omega), X_2(\omega), ..., X_m(\omega))$ upraszcza się do S,

[1)] Przedstawione powyżej ujęcie dotyczy **estymacji punktowej**, podstawowego działu teorii estymacji. Inne koncepcje (np. estymacji przedziałowej, gdzie wyznacza się nie tyle konkretną wartość estymatora, co przedział, do którego z założonym prawdopodobieństwem powinna należeć analizowana wielkość) są pominięte jako niewykorzystywane w dalszej części niniejszej monografii.

pomijając zależność od próby $X_1, X_2, ..., X_m$ (wówczas podobnie jak w przypadku próby losowej, opuszczana jest również zależność od czynnika ω). Co więcej, w praktyce warto stosować dogodną notację, zgodnie z którą w miejsce ogólnego symbolu S podstawia się symbol konkretnej, estymowanej wielkości uzupełnionej „daszkiem". I tak, jeżeli analizowany jest parametr b, to wartość jego estymatora oznacza się jako $\hat{b}$, podobnie jak estymator funkcji f dla argumentu x będzie oznaczany przez $\hat{f}(x)$.

2.2.2. Własności estymatorów

W zależności od uwarunkowań aplikacyjnych w odniesieniu do estymatorów formułuje się różnorodne wymagania normujące ich podstawowe własności.

Rozpatrzmy najpierw przypadek estymacji pojedynczego parametru charakteryzującego rozkład n-wymiarowej zmiennej losowej X. Niech zatem liczba rzeczywista $\hat{b}$ oznacza estymator parametru $b \in \mathbb{R}$, wyznaczony na podstawie m-elementowej próby losowej $X_1, X_2, ..., X_m$. Różnicę jego wartości oczekiwanej $E(\hat{b})$ i „prawdziwej" wartości estymowanego parametru, to znaczy liczbę $E(\hat{b}) - b$, określa się mianem **obciążenia estymatora** $\hat{b}$. Niezerowe obciążenie świadczy o obecności błędu systematycznego w stosowanej metodyce. Dodatnia wartość obciążenia wskazuje na tendencję do przeszacowywania, a ujemna – niedoszacowywania prawdziwej wartości parametru. Jeżeli wartość obciążenia jest równa zeru, czyli gdy zachodzi warunek

$$E(\hat{b}) = b, \tag{2.39}$$

to estymator ten nazywa się **nieobciążonym**. Należy dodać, iż powyższa równość powinna zachodzić dla dowolnej liczności próby losowej, na podstawie której wyznacza się wartość estymatora. Własność nieobciążoności oznacza, że w praktyce wartości estymatora będą „oscylować" wokół „prawdziwej" wartości estymowanego parametru, a ściślej, iż estymator nie będzie jej systematycznie przeszacowywać ani niedoszacowywać.

Tak więc obciążenie estymatora wskazuje, wokół jakiej wartości będą „oscylować" uzyskiwane w praktyce wartości estymatora. O tym, jak duże będą owe „oscylacje", świadczy **wariancja** tego estymatora $V(\hat{b})$. Pożądane jest, aby wariancja dążyła do zera, gdy liczność próby dąży do nieskończoności, czyli aby

$$\lim_{m \to \infty} V(\hat{b}) = 0. \tag{2.40}$$

Jeżeli w miarę wzrostu liczności próby losowej obciążenie estymatora maleje do zera, a zatem gdy jest spełniony warunek

$$\lim_{m\to\infty} E(\hat{b}) = b, \tag{2.41}$$

to estymator ten nazywa się **asymptotycznie nieobciążonym**. Oczywiście własność ta jest słabsza od nieobciążoności[1)].

Jeśli z kolei w miarę zwiększania liczności próby losowej estymator $\hat{b}$ dąży do estymowanej wartości b według prawdopodobieństwa, czyli jeżeli dla każdego $\varepsilon > 0$ prawdziwe jest

$$\lim_{m\to\infty} P(\{\omega \in \Omega : |\hat{b}(X_1(\omega), X_2(\omega), ..., X_m(\omega)) - b| \geq \varepsilon\}) = 0, \tag{2.42}$$

to estymator ten określa się mianem **zgodnego**. Z własności tej wynika, iż w miarę wzrostu liczności próby losowej wartość estymatora jest sukcesywnie bliższa wartości estymowanej. Z aplikacyjnego punktu widzenia oznacza to możliwość osiągania coraz to lepszej dokładności poprzez zebranie dodatkowych danych, aczkolwiek w praktyce wiąże się to najczęściej ze zwiększaniem czasu i kosztu badań. Jeżeli estymator jest asymptotycznie nieobciążony i jego wariancja maleje do zera przy zwiększaniu liczności próby, a zatem gdy jednocześnie spełnione są warunki (2.40) i (2.41), to estymator ten jest także zgodny.

Powyższe pojęcia w sposób naturalny transponuje się na przypadek **estymacji funkcji** $B : \mathbb{R}^n \to \mathbb{R}$, stanowiących charakterystyki funkcyjne rozkładu zmiennej losowej, na przykład gęstość prawdopodobieństwa lub dystrybuantę. Dla ustalonego argumentu x wartość funkcji B jest bowiem traktowana tak jak rozważany wcześniej rzeczywisty parametr b, to znaczy

$$B(x) = b. \tag{2.43}$$

Estymator funkcji może zatem mieć własność (asymptotycznej) nieobciążoności lub zgodności w dowolnym punkcie dziedziny bądź też pewnego jej podzbioru.

W przypadku k-wymiarowego wektora parametrów lub funkcji k-wymiarowej (asymptotyczna) nieobciążoność i zgodność są równoważne spełnieniu przedstawionych powyżej warunków przez wszystkie współrzędne.

W praktyce zgodność – jak przystało na własność asymptotyczną – bywa ważna przy dużych licznościach próby, natomiast przy małych licznościach za

[1)] W powyższym zapisie zastosowano wobec wartości oczekiwanej E oraz wariancji V zapis typu funkcjonałowego $E(\hat{b})$ i $V(\hat{b})$, co pozwala uwidocznić symbol rozważanego aktualnie estymatora. Podobna notacja będzie okazjonalnie stosowana także w dalszej części książki.

istotną uznaje się własność nieobciążoności. Własność asymptotycznej nieobciążoności należy traktować jako minimum wymagane od stosowanych w praktyce estymatorów.

2.2.3. Przykłady estymatorów

Jak było wspomniane, formułę estymatora określa się tak, aby po uzyskaniu wartości próby losowej (2.37) otrzymać bezpośrednio oszacowanie badanego parametru lub charakterystyki funkcyjnej rozkładu dla ustalonego argumentu. W tym punkcie zostaną najpierw przedstawione klasyczne estymatory podstawowych parametrów – wartości oczekiwanej, wariancji i odchylenia standardowego oraz kowariancji. Będą one używane w dalszej części pracy do wyznaczenia współczynników ustalających ostateczną postać estymatorów jądrowych.

Jako pierwszy zostanie rozważony przypadek jednowymiarowej zmiennej losowej. Naturalny **estymator wartości oczekiwanej** (2.17) stanowi średnia arytmetyczna poszczególnych elementów próby losowej (2.37):

$$\hat{E} = \frac{1}{m}\sum_{i=1}^{m} x_i. \tag{2.44}$$

Estymator ten jest nieobciążony i zgodny. Jeśli w rozpatrywanym zagadnieniu wartość oczekiwana E jest znana, to nieobciążony i zgodny **estymator wariancji** (2.19) wyraża się równością

$$\hat{V} = \frac{1}{m}\sum_{i=1}^{m} (x_i - E)^2. \tag{2.45}$$

Jeżeli jednak wartość oczekiwana nie jest znana, to po przyjęciu oczywistego z praktycznego punktu widzenia warunku $m \geq 2$ estymator taki przyjmuje postać

$$\hat{V} = \frac{1}{m-1}\sum_{i=1}^{m} (x_i - \hat{E})^2, \tag{2.46}$$

gdzie wielkość $\hat{E}$ określa wzór (2.44). Równoważna postać tej zależności:

$$\hat{V} = \frac{1}{m-1}\sum_{i=1}^{m} x_i^2 - \frac{1}{m(m-1)}\left(\sum_{i=1}^{m} x_i\right)^2 \tag{2.47}$$

jest nieco dogodniejsza do obliczeń numerycznych, gdyż obie sumy można wyznaczyć w jednej pętli programowej. **Estymator odchylenia standardowego** (2.20) definiuje się w sposób naturalny jako

$$\hat{\sigma} = \sqrt{\hat{V}}, \tag{2.48}$$

przy czym $\hat{V}$ określa zależność (2.45) albo (2.46) lub równoważnie (2.47), odpowiednio, gdy wartość oczekiwana jest znana lub nie. Estymator (2.48) jest asymptotycznie nieobciążony i zgodny.

W przypadku wielowymiarowej zmiennej losowej **estymator wartości oczekiwanej *j*-tej współrzędnej** (2.23) wyznacza się analogicznie do wzoru (2.44), przyjmując do obliczeń j-te współrzędne poszczególnych elementów próby losowej (2.37):

$$\hat{E}_j = \frac{1}{m}\sum_{i=1}^{m} x_{i,j}, \tag{2.49}$$

gdzie $x_{i,j}$ oznacza j-tą współrzędną i-tego elementu próby. Podobnie oblicza się **estymator wariancji *j*-tej współrzędnej** (2.25) w przypadku, gdy jej wartość oczekiwana E_j jest znana:

$$\hat{V}_j = \frac{1}{m}\sum_{i=1}^{m} (x_{i,j} - E_j)^2 \tag{2.50}$$

lub gdy jest szacowana na podstawie wzoru (2.49) przy $m \geq 2$:

$$\hat{V}_j = \frac{1}{m-1}\sum_{i=1}^{m} (x_{i,j} - \hat{E}_j)^2 \tag{2.51}$$

oraz równoważnie

$$\hat{V}_j = \frac{1}{m-1}\sum_{i=1}^{m} x_{i,j}^{\,2} - \frac{1}{m(m-1)}\left(\sum_{i=1}^{m} x_{i,j}\right)^2. \tag{2.52}$$

W konsekwencji **estymator odchylenia standardowego *j*-tej współrzędnej** (2.26) definiuje się jako

$$\hat{\sigma}_j = \sqrt{\hat{V}_j}, \tag{2.53}$$

gdzie $\hat{V}_j$ określa zależność (2.50) albo (2.51) lub równoważnie (2.52), odpowiednio, gdy wartość oczekiwana jest znana lub nie. Z kolei, jeżeli znane są E_{j_1} oraz E_{j_2}, czyli wartości oczekiwane współrzędnych j_1 oraz j_2, to **estymatory elementów macierzy kowariancji** (2.30) przyjmują postać

$$\hat{cov}_{j_1,j_2} = \frac{1}{m}\sum_{i=1}^{m} (x_{i,j_1} - E_{j_1})(x_{i,j_2} - E_{j_2}), \tag{2.54}$$

natomiast gdy wartości oczekiwane są określane równością (2.49) i przy $m \geq 2$, wyrażają się one wzorem

$$\hat{cov}_{j_1,j_2} = \frac{1}{m-1}\sum_{i=1}^{m}(x_{i,j_1} - \hat{E}_{j_1})(x_{i,j_2} - \hat{E}_{j_2}) \tag{2.55}$$

oraz równoważnie

$$\hat{cov}_{j_1,j_2} = \frac{1}{m-1}\sum_{i=1}^{m} x_{i,j_1}x_{i,j_2} - \frac{1}{m(m-1)}\sum_{i=1}^{m} x_{i,j_1}\sum_{i=1}^{m} x_{i,j_2}. \tag{2.56}$$

Zależności (2.50) i (2.54), (2.51) i (2.55), (2.52) i (2.56) są oczywiście parami tożsame, gdy $j_1 = j_2$. Estymatory (2.49)-(2.56) są zgodne. Są także nieobciążone, z wyjątkiem (2.53), który jest jedynie asymptotycznie nieobciążony.

Na zakończenie niniejszego podrozdziału zostaną jeszcze przedstawione koncepcja klasycznych estymatorów kwantyla i procedura rozwiązania problemu estymacji charakterystyk funkcyjnych poprzez sprowadzenie – a ściślej, jego uproszczenie – do zadania estymacji parametrycznej za pomocą arbitralnego ustalenia typu rozkładu prawdopodobieństwa. Celem powyższego jest stworzenie podstaw do porównania metod klasycznych z opartymi na estymatorach jądrowych, które stanowią przedmiot rozważań dalszej części tej monografii.

Rozpatrywana będzie jednowymiarowa zmienna losowa. W przypadku zmiennej wielowymiarowej rozważania są analogiczne – podobnie jak poprzednio dla wartości oczekiwanej i wariancji, są tu brane pod uwagę odpowiednie współrzędne poszczególnych elementów próby losowej.

Niech zatem

$$\tilde{x}_1, \tilde{x}_2, ..., \tilde{x}_m \tag{2.57}$$

oznacza wartość próby losowej (2.37) uporządkowaną rosnąco, czyli tak aby $\tilde{x}_1 \le \tilde{x}_2 \le ... \le \tilde{x}_m$. Ponadto niech założony rząd kwantyla r spełnia warunek

$$0{,}5 \le mr \le m - 0{,}5. \tag{2.58}$$

Najprostszym **klasycznym estymatorem kwantyla *r*-tego rzędu** jest element próby (2.57) o indeksie równym $[mr + 0{,}5]$, a zatem

$$\hat{q} = \tilde{x}_{[mr+0,5]}, \tag{2.59}$$

przy czym $[c]$ oznacza część całkowitą liczby rzeczywistej c. Dla ilustracji rozważmy zadanie estymacji mediany na podstawie trójelementowej próby losowej, czyli $r = 0{,}5$ oraz $m = 3$. Wtedy zgodnie z intuicją

$$\hat{q} = \tilde{x}_{[2]} = \tilde{x}_2, \tag{2.60}$$

a zatem estymatorem mediany jest „środkowy", czyli drugi pod względem wielkości, element próby losowej. Jeżeli jednak zwiększyć o jeden liczność próby losowej, czyli $m = 4$, to nadal

$$\hat{q} = \widetilde{x}_{[2,5]} = \widetilde{x}_2, \tag{2.61}$$

aczkolwiek estymator mediany powinien przecież wzrosnąć i być – na przykład – średnią arytmetyczną drugiego oraz trzeciego pod względem wielkości elementu próby. Tak więc wzór jednoskładnikowy (2.59) modyfikuje się do postaci dwuskładnikowej:

$$\hat{q} = (0{,}5 - mr + [mr + 0{,}5])\widetilde{x}_{[mr+0{,}5]} + (0{,}5 + mr - [mr + 0{,}5])\widetilde{x}_{[mr+1,5]}. \tag{2.62}$$

Łatwo sprawdzić, że jeśli $r = 0{,}5$ oraz $m = 3$, to dwuskładnikowym estymatorem mediany jest nadal

$$\hat{q} = (0{,}5 - 1{,}5 + [2])\widetilde{x}_{[2]} + (0{,}5 + 1{,}5 - [2])\widetilde{x}_{[3]} = \widetilde{x}_{[2]} = \widetilde{x}_2, \tag{2.63}$$

lecz dla $m = 4$ zależność (2.62) implikuje zgodnie z wymaganiem

$$\hat{q} = (0{,}5 - 2 + [2{,}5])\widetilde{x}_{[2,5]} + (0{,}5 + 2 - [2{,}5])\widetilde{x}_{[3,5]} = 0{,}5\widetilde{x}_{[2,5]} + 0{,}5\widetilde{x}_{[3,5]} =$$
$$= 0{,}5\widetilde{x}_2 + 0{,}5\widetilde{x}_3. \tag{2.64}$$

W celu zapewnienia ogólności estymatory (2.59) i (2.62) mogą być uzupełnione w następujący sposób:

$$\hat{q} = \begin{cases} \widetilde{x}_1, & \text{gdy } mr < 0{,}5, \\ \text{według wzoru (2.59) lub (2.62)}, & \text{gdy } 0{,}5 \le mr \le m - 0{,}5, \\ \widetilde{x}_m, & \text{gdy } mr > m - 0{,}5, \end{cases} \tag{2.65}$$

z pominięciem warunku (2.58). Wszystkie wymienione powyżej estymatory, czyli (2.59), (2.62) i (2.65), są asymptotycznie nieobciążone i zgodne.

Zwiększanie ponad dwa liczby składników w definicji estymatora kwantyla okazuje się mało skuteczne, a już ponad cztery wręcz bezefektywne. Korzystniejsze wyniki – zwłaszcza w najtrudniejszych przypadkach niedużej liczności próby losowej i skrajnych (tzn. bliskich 0 lub 1) wartości rzędu kwantyla – uzyskuje się, stosując estymatory jądrowe. Zagadnienie to będzie tematem punktu 3.2.2 i dodatkowo w wersji warunkowej – punktu 3.2.3.

Na koniec zostanie zilustrowana koncepcja rozwiązywania problemu estymacji charakterystyk funkcyjnych z użyciem klasycznych procedur estymacji parametrycznej. Ich zasadnicza koncepcja polega na arbitralnym ustaleniu typu rozkładu badanej zmiennej losowej, a następnie wyznaczeniu wartości występujących tam współczynników za pomocą procedur estymacji parametrycznej.

I tak, ustalenie typu rozkładu (np. normalny, jednostajny, trójkątny, beta) odbywa się tradycyjnie na podstawie ogólnej znajomości badanego zjawiska, oglądu histogramu lub specjalistycznych algorytmów. Tego typu algorytmy polegają – najogólniej rzecz ujmując – na wyróżnieniu reprezentatywnej grupy kilku

potencjalnych typów rozkładów, a następnie wskazaniu jednego z nich, na podstawie badania uzyskanej próby losowej.

Po ustaleniu typu rozkładu są wyznaczane wartości parametrów występujących w jego gęstości prawdopodobieństwa lub innej charakterystyce funkcyjnej. Najbliższą intuicji procedurą jest tradycyjna **metoda momentów**. Polega ona na takim przyjęciu owych parametrów, aby pierwsze momenty określonego w ten sposób rozkładu równały się wartościom otrzymanym na podstawie próby losowej. Załóżmy na przykład, że dla badanej jednowymiarowej zmiennej losowej został w pierwszej fazie arbitralnie ustalony rozkład gamma – jego gęstość jest podana na rysunku 2.2. Wartość oczekiwana i wariancja rozkładu gamma wynoszą odpowiednio

$$E = \frac{p}{q}, \tag{2.66}$$

$$V = \frac{p}{q^2}. \tag{2.67}$$

Rozwiązując układ równań (2.66)-(2.67) względem p oraz q, otrzymuje się

$$p = \frac{E^2}{V}, \tag{2.68}$$

$$q = \frac{E}{V}. \tag{2.69}$$

Jeżeli po uzyskaniu wartości próby losowej, na podstawie zależności (2.44) oraz (2.46) lub (2.47) dysponuje się estymatorami wartości oczekiwanej $\hat{E}$ i wariancji $\hat{V}$, to podstawiając je do wzorów (2.68)-(2.69) w miejsce E oraz V, otrzymuje się estymatory parametrów p oraz q, co ostatecznie ustala estymator poszukiwanej gęstości prawdopodobieństwa.

Metoda momentów ma już obecnie znaczenie głównie ilustracyjne. Korzystniejsze estymatory parametrów uzyskuje się z wykorzystaniem innych klasycznych metod – największej wiarygodności lub najmniejszych kwadratów, a także nowoczesnych metod estymacji bayesowskiej lub minimaksowej. Niezależnie jednak od jakości szacowania samych parametrów podstawową słabością estymacji parametrycznej jest arbitralność założenia o typie rozkładu, w ramach drastycznie ograniczonej w praktyce możliwości wyboru. Wady tej nie mają rozwijające się współcześnie, dzięki postępowi techniki komputerowej, procedury **estymacji nieparametrycznej**. Najczęściej stosowana w tym zakresie koncepcja **estymatorów jądrowych** będzie przedmiotem dalszej części niniejszej książki. Istota tej koncepcji została sformułowana dla bliskiego intuicji zadania

estymacji funkcji gęstości prawdopodobieństwa zmiennej losowej i mogłaby być traktowana jako kontynuacja powyższego przykładu, po zarzuceniu arbitralności założeń dotyczących postaci tej funkcji.

Uwagi bibliograficzne

Statystyka matematyczna, w tym także zagadnienia estymacji, doczekała się niezwykle bogatej literatury przedmiotowej, odpowiednio zróżnicowanej dla odbiorców z odmiennych dziedzin nauki i praktyki. Jako uniwersalne można wymienić prace [20, 26, 53]. W celu poznania podstaw warto natomiast wyróżnić książkę [34], a także podręcznik [22] między innymi z uwagi na jego oryginalną formułę. Monografie [18, 42] reprezentują klasyczne pozycje o ścisłym charakterze. Aspekty związane z numerycznymi metodami obliczeniowymi są interesująco prezentowane w publikacji [6].

2.3. Rozwiązania równań różniczkowych z nieciągłą prawą stroną

Stały wzrost wymagań formułowanych wobec zadań analizy systemowej sprawia, że są w nich coraz częściej wprowadzane nieciągłe modele obiektów i zachodzących tu procesów. Co więcej, powszechna presja na zwiększenie efektywności powoduje, iż nierzadko są podejmowane ekstremalne działania, które mogą w określonych chwilach czasu gwałtownie zmieniać swą wartość. Oba te czynniki coraz częściej skutkują wprowadzeniem do opisu dynamiki badanego systemu równań różniczkowych z nieciągłą prawą stroną. Tymczasem zwykle stosowane klasyczne rozwiązanie równań różniczkowych w takich przypadkach nie istnieje lub niewłaściwie opisuje modelowaną rzeczywistość. Prowadzi to niejednokrotnie do błędów teoretycznych i nieporozumień interpretacyjnych.

W niniejszym podrozdziale są zaprezentowane podstawowe definicje i zależności związane z problematyką rozwiązań równań różniczkowych z nieciągłą prawą stroną. Na końcu jest przedstawiony podsumowujący przykład, obrazowo ilustrujący wprowadzone pojęcia i wskazujący nieporozumienia, jakie może implikować niewłaściwe ich stosowanie. Materiał ten będzie przydatny przy formalizowaniu zagadnienia rozważanego w punkcie 4.1.4. Jeżeli jednak Czytelnik nie jest zainteresowany systemami dynamicznymi, a zwłaszcza specyfiką ich modelowania za pomocą równań różniczkowych, to może pominąć ten podrozdział.

2.3.1. Podstawowe definicje i zależności

Najczęściej stosowanym opisem systemów dynamicznych jest **równanie różniczkowe** postaci

$$\dot{x}(t) = g(x(t), t), \tag{2.70}$$

gdzie $x: T \to \mathbb{R}^n$, $g: \mathbb{R}^n \times T \to \mathbb{R}^n$, przy czym $T \subset \mathbb{R}$ jest przedziałem o niepustym wnętrzu. Równanie takie jest nierozłącznie rozważane wraz z **warunkiem początkowym**

$$x(t_0) = x_0, \tag{2.71}$$

przy czym $t_0 \in T$ oraz $x_0 \in \mathbb{R}^n$ są ustalone. Na potrzeby poszczególnych dyscyplin zależność (2.70) może być zapisana przy użyciu nieco odmiennej notacji. Na przykład w zagadnieniach automatyki, w których powinien być wyraźnie wyeksponowany wpływ sterowania u na stan systemu x, równanie to przyjmuje postać

$$\dot{x}(t) = f(x(t), u(t), t), \tag{2.72}$$

przy $x: T \to \mathbb{R}^n$, $u: T \to \mathbb{R}^m$, $f: \mathbb{R}^n \times \mathbb{R}^m \times T \to \mathbb{R}^n$. Różnica między formułami (2.70) i (2.72) jest pozorna, głównie w sferze interpretacji. Zależność funkcji f od t oraz $u(t)$ została bowiem ujęta we wzorze (2.70) w ramach zależności funkcji g bezpośrednio od t – obecność sterowania nie jest jedynie w przypadku postaci (2.70) *explicite* wyeksponowana.

Klasyczne rozwiązanie równania różniczkowego (2.70)-(2.71) stanowi taka funkcja x, która jest różniczkowalna i spełnia równanie (2.70) dla każdego $t \in T$ (jeśli krańce przedziału T należą do niego, to rozważa się tu odpowiednią pochodną jednostronną), a także spełnia warunek (2.71). Rozwiązanie to jest **jednoznaczne**, jeżeli każde rozwiązanie klasyczne jest funkcją równą mu tożsamościowo. Powyższe definicje są w pełni zgodne z intuicją, przynajmniej w zakresie podstawowych aspektów dotyczących systemów dynamicznych.

Tak określone rozwiązanie bywa użyteczne, gdy funkcja g jest ciągła. Ciągłość ta stanowi bowiem warunek wystarczający istnienia rozwiązania klasycznego i chociaż nie jest warunkiem koniecznym, to założenie ciągłości jest praktycznie nieodłączne rozważaniom nad rozwiązaniami klasycznymi. W przypadku skończonej liczby nieciągłości pierwszego rodzaju[1)] funkcji g względem zmiennej niezależnej t powyższą definicję można jednakże uzupełnić o nieco nieformalną operację tak zwanego **sklejania rozwiązań**. Otóż, przedział T dzieli

[1)] Przy uwarunkowaniach rozważanego zagadnienia, nieciągłość funkcji jest nazywana **nieciągłością pierwszego rodzaju**, jeżeli w punkcie nieciągłości istnieją skończone granice jednostronne. Granice te nie muszą być sobie równe ani równe wartości funkcji w tym punkcie.

się na podprzedziały, w których funkcja g jest ciągła, i w tych podprzedziałach są znajdowane rozwiązania klasyczne, po czym rozwiązania te łączy się („skleja") z zachowaniem ciągłości rozwiązania x (w punktach nieciągłości funkcji g względem zmiennej niezależnej pochodna $\dot{x}$ nie istnieje ze względu na brak równości pochodnych jednostronnych). Powyższa koncepcja ma wyrazistą interpretację fizyczną – uznaje się, że w chwili wystąpienia nieciągłości funkcji g proces opisywany równaniem różniczkowym (2.70) zaczyna się od nowa, z warunkiem początkowym (2.71) będącym lewostronną granicą rozwiązania z poprzedniego obszaru ciągłości. Z matematycznego punktu widzenia postępowanie takie ma bardzo ograniczony zakres stosowalności do przypadku skończonej liczby nieciągłości funkcji g i to jedynie względem zmiennej niezależnej.

Typowym przykładem ilustrującym konieczność wprowadzenia uogólnienia pojęcia rozwiązania klasycznego jest następujące dwuwymiarowe (tzn. dla $n = 2$) równanie różniczkowe, reprezentujące w sposób uproszczony zadanie sterowania w układzie zamkniętym z nieciągłą funkcją regulatora (np. w problemie regulacji minimalnoczasowej):

$$\dot{x}_1(t) = 1, \qquad x_1(0) = x_{01}, \tag{2.73}$$

$$\dot{x}_2(t) = \begin{cases} -1, & \text{gdy } x_2(t) > 0, \\ 1, & \text{gdy } x_2(t) \le 0, \end{cases} \qquad x_2(0) = x_{02}, \tag{2.74}$$

przy czym $x_{01}, x_{02} \in \mathbb{R}$. Jego rozwiązania kończą się po osiągnięciu osi x_1. Gdyby bowiem rozwiązanie dało się przedłużyć, to – jak wynika ze wzoru (2.74) – byłoby ono „wypychane" ponad oś x_1 (ponieważ $\dot{x}_2(t) = 1$, gdy $x_2(t) \le 0$), lecz natychmiast po jej „opuszczeniu" ponownie w nią „wpychane" (ponieważ $\dot{x}_2(t) = -1$, gdy $x_2(t) > 0$), co neguje możliwość istnienia takiego rozwiązania. Warto zauważyć, że w rozważanym powyżej przykładzie funkcja g jest nieciągła na osi x_1.

Koncepcje uogólniające pojęcie klasycznego rozwiązania na potrzeby **równań różniczkowych z nieciągłą prawą stroną**, czyli takich w których funkcja g określona we wzorze (2.70) jest nieciągła, nie doprowadziły do jednolitego ujęcia. Obecnie najczęściej stosowanymi są rozwiązania zaproponowane przez Caratheodory'ego, Filippowa i Krasowskiego. Przed ich zdefiniowaniem zostaną przedstawione pojęcia „prawie wszędzie", funkcji absolutnie ciągłej, powłoki domkniętej wypukłej i zbioru zwartego.

Jeżeli miara tych punktów zbioru $A \subset \mathbb{R}$, w których jakaś własność nie jest spełniona, wynosi zero, to mówi się, że własność ta zachodzi **prawie wszędzie** w A.

Rozważana teraz będzie funkcja $x : [a, b] \to \mathbb{R}^n$. Jest ona nazywana **funkcją absolutnie (bezwzględnie) ciągłą**, jeśli dla każdego $\varepsilon > 0$ istnieje takie $\delta > 0$, że gdy $(a_1, b_1), (a_2, b_2), \ldots, (a_k, b_k)$ są przedziałami zawartymi w dziedzinie

$[a, b]$, rozłącznymi i takimi, że $\sum_{i=1}^{k}(b_i - a_i) \le \delta$, to spełnione jest $\sum_{i=1}^{k}\| x(b_i) - x(a_i)\| \le \varepsilon$. Funkcja absolutnie ciągła jest także ciągła. Co więcej, jest ona różniczkowalna prawie wszędzie w przedziale $[a, b]$. Te dwie własności sprawiają, że funkcja absolutnie ciągła szczególnie dobrze nadaje się do zastosowania w zagadnieniu uogólnienia pojęcia klasycznego rozwiązania równania różniczkowego.

Powłoką domkniętą wypukłą zbioru $C \subset \mathbb{R}^n$, oznaczaną dalej jako conv(C), jest nazywany najmniejszy (w sensie relacji inkluzji) zbiór domknięty i wypukły zawierający zbiór C. Na przykład powłoką domkniętą wypukłą zbioru dwóch różnych punktów jest łączący je odcinek, a zbioru trzech różnych punktów – domknięty trójkąt, którego punkty te są wierzchołkami.

Podzbiór przestrzeni $\mathbb{R}$ jest **zwarty** wtedy i tylko wtedy, gdy jest ograniczony i domknięty.

Zostaną teraz zdefiniowane rozwiązania równania różniczkowego w sensie Caratheodory'ego, Filippowa i Krasowskiego. Funkcja x, absolutnie ciągła na każdym zwartym podprzedziale zbioru T, jest **rozwiązaniem równania różniczkowego** (2.70)-(2.71):

- **w sensie Caratheodory'ego (C-rozwiązaniem)**, jeżeli spełnia równanie (2.70) prawie wszędzie w T, a także spełnia warunek (2.71);
- **w sensie Filippowa (F-rozwiązaniem)**, jeżeli

$$\dot{x}(t) \in F[g](x(t), t) \quad \text{prawie wszędzie w } T, \tag{2.75}$$

a także spełnia warunek (2.71), przy czym operator Filippowa F jest zdefiniowany następująco:

$$F[g](x(t), t) = \bigcap_{r>0} \bigcap_{\substack{Z \subset \mathbb{R}^n : \\ m(Z)=0}} \operatorname{conv}[g((x(t) + rB) \setminus Z, t)]; \tag{2.76}$$

- **w sensie Krasowskiego (K-rozwiązaniem)**, jeżeli

$$\dot{x}(t) \in K[g](x(t), t) \quad \text{prawie wszędzie w } T, \tag{2.77}$$

a także spełnia warunek (2.71), przy czym operator Krasowskiego K jest dany zależnością

$$K[g](x(t), t) = \bigcap_{r>0} \operatorname{conv}[g(x(t) + rB, t)]; \tag{2.78}$$

gdzie B oznacza otwartą kulę jednostkową w przestrzeni $\mathbb{R}^n$, a m stanowi n-wymiarową miarę Lebesgue'a. C-, F- lub K-rozwiązanie jest **jednoznaczne**, jeżeli

każde, odpowiednio, C-, F- lub K-rozwiązanie jest funkcją równą mu tożsamościowo. Przedstawiona poniżej interpretacja tych definicji dostarcza także praktycznych motywacji określonych powyżej ich postaci.

I tak, w przypadku C-rozwiązania pochodna $\dot{x}(t)$ jest zależna (poza zmienną niezależną t) jedynie od aktualnej wartości rozwiązania x. Ten typ rozwiązania stanowi w praktyce jednolite matematycznie uogólnienie „sklejanego" rozwiązania klasycznego.

W przypadku K-rozwiązania są uwzględniane wartości funkcji g nie tylko w $x(t)$, lecz także we wszystkich punktach kuli o środku w tym punkcie i dodatnim, aczkolwiek – ze względu na iloczyn mnogościowy – dowolnie małym promieniu r. Co więcej, powłoka domknięta wypukła uzupełnia tak powstały zbiór wartości funkcji g o wielkości „pośrednie". Interpretując powyższe, można stwierdzić, że K-rozwiązanie, dopuszczając wszystkie punkty z otoczenia wartości $x(t)$, uwzględnia nieuniknione w praktyce błędy pomiarowe.

I wreszcie, dodatkowo wprowadzony w definicji F-rozwiązania iloczyn mnogościowy powoduje, iż z kuli o środku w punkcie $x(t)$ i promieniu r są eliminowane zbiory miary zero, nieistotne z praktycznego punktu widzenia.

Wprost z powyższych definicji wynika, że

A) każde C-rozwiązanie jest K-rozwiązaniem, ponieważ
$g(x(t), t) \in K[g](x(t), t)$;

B) każde F-rozwiązanie jest K-rozwiązaniem, gdyż
$F[g](x(t), t) \subset K[g](x(t), t)$.

Jednoznaczność K-rozwiązań implikuje zatem jednoznaczność C- i F-rozwiązań.

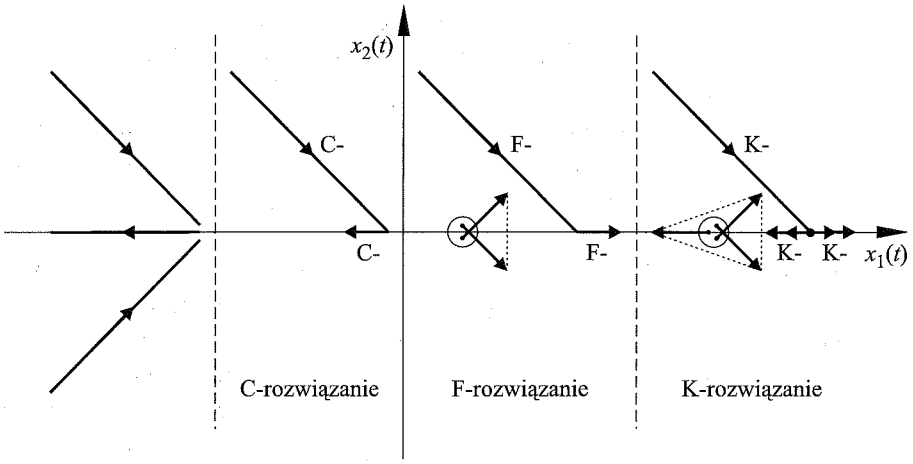
Inne relacje między C-, F- i K-rozwiązaniami ilustruje przykład następującego dwuwymiarowego równania różniczkowego:

$$\dot{x}_1(t) = \begin{cases} 1, & \text{gdy } x_2(t) > 0, \\ -1, & \text{gdy } x_2(t) = 0, \\ 1, & \text{gdy } x_2(t) < 0, \end{cases} \qquad x_1(0) = x_{01}, \tag{2.79}$$

$$\dot{x}_2(t) = \begin{cases} -1, & \text{gdy } x_2(t) > 0, \\ 0, & \text{gdy } x_2(t) = 0, \\ 1, & \text{gdy } x_2(t) < 0, \end{cases} \qquad x_2(0) = x_{02}, \tag{2.80}$$

przy czym $x_{01}, x_{02} \in \mathbb{R}$ (rys. 2.3).

Najpierw zostanie rozpatrzony przypadek, gdy stan początkowy zawiera się w osi x_1. Funkcja $\begin{bmatrix} x_{01} - t \\ 0 \end{bmatrix}$ dla $t \in [0, \infty)$ jest wtedy jednoznacznym C-rozwią-



Rys. 2.3. Rozwiązania równania różniczkowego (2.79)-(2.80)

zaniem. Ponieważ dla każdego $t \in [0, \infty)$ oraz $x(t)$ zawartego w osi x_1 prawdziwa jest równość $F[g](x(t),t) = \left\{ \begin{bmatrix} 1 \\ d \end{bmatrix}, \text{ przy czym } |d| \le 1 \right\}$, więc funkcja $\begin{bmatrix} x_{01}+t \\ 0 \end{bmatrix}$ dla $t \in [0, \infty)$ jest wtedy jednoznacznym F-rozwiązaniem. Z kolei dla każdego $t \in [0, \infty)$ oraz $x(t)$ zawartego w osi x_1 zbiór $K[g](x(t),t)$ jest trójkątem o wierzchołkach $\begin{bmatrix} 1 \\ 1 \end{bmatrix}, \begin{bmatrix} 1 \\ -1 \end{bmatrix}, \begin{bmatrix} -1 \\ 0 \end{bmatrix}$, a zatem absolutnie ciągłe funkcje postaci $\begin{bmatrix} x_1 \\ 0 \end{bmatrix}$, przy czym $x_1(0) = x_{01}$ oraz $|\dot{x}_1(t)| \le 1$ prawie wszędzie w $[0, \infty)$, są K-rozwiązaniami.

Natomiast w przypadku, gdy stan początkowy nie zawiera się w osi x_1, funkcja $\begin{bmatrix} x_{01}+t \\ x_{02}-t\cdot\operatorname{sgn}(x_{02}) \end{bmatrix}$ jest, do chwili przecięcia tej osi, jednoznacznym C-, F- i K-rozwiązaniem. W chwili osiągnięcia osi x_1 rozwiązania te można przedłużyć odpowiednimi rozwiązaniami opisanymi w poprzednim akapicie.

Z powyższego przykładu wynika, że w ogólnym przypadku brak jest zależności między C- i F-rozwiązaniami, czyli kontynuując uwagi sformułowane uprzednio w postaci punktów A-B, można teraz dodać, iż

C) C-rozwiązanie nie musi być F-rozwiązaniem,

D) F-rozwiązanie nie musi być C-rozwiązaniem.

Ponadto K-rozwiązania nierzadko stanowią jednak zbyt „bogatą" – z praktycznego punktu widzenia – klasę. Wszakże dla pewnego rodzaju równań różniczkowych można wskazać zależność między C- i F-rozwiązaniami, co zostanie przedstawione poniżej.

Otóż, funkcja g określona w równaniu (2.70) jest nazywana **funkcją obszarami ciągłą**, jeżeli dla prawie każdego $t \in T$ istnieje rodzina $\{M_{t,i}\}$ (przy czym $i \in I_t$) podzbiorów przestrzeni $\mathbb{R}^n$ takich, że $\mathbb{R}^n = \bigcup_{t \in I_t} M_{t,i}$ i ponadto

$$M_{t,i} \subset \overline{\text{int}(M_{t,i})} \quad \text{dla każdego } i \in I_t, \tag{2.81}$$

a także rodzina $\{g_{t,i}\}$ (przy czym jak poprzednio $i \in I_t$) funkcji $g_{t,i}: \mathbb{R}^n \times T \to \mathbb{R}^n$ ciągłych ze względu na n pierwszych zmiennych oraz takich, że $g \equiv g_{t,i}$ na odpowiednich zbiorach $M_{t,i}$. Przykładem zbioru $M_{t,i}$ spełniającego warunek (2.81) jest kula o dodatnim promieniu, natomiast nie spełnia go linia prosta w przestrzeni $\mathbb{R}^2$, gdyż jej wnętrze jest zbiorem pustym i taka jest również prawa strona relacji (2.81). Warunek funkcji obszarami ciągłej stanowi zatem, że dla prawie każdego t przestrzeń $\mathbb{R}^n$ dekomponuje się na takie zbiory spełniające warunek (2.81), że funkcja g zacieśniona do dowolnego z tych zbiorów da się poszerzyć do funkcji ciągłej względem $x(t)$. Prawdziwe jest więc stwierdzenie, że

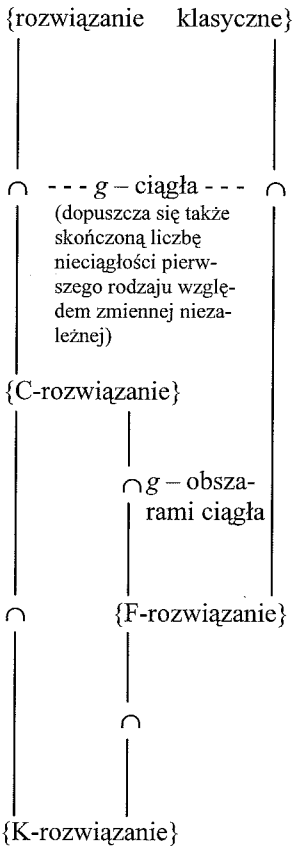
E) jeżeli funkcja g jest obszarami ciągła, to każde C-rozwiązanie jest F-rozwiązaniem.

Załóżmy teraz, że funkcja g jest ciągła, które to założenie – jak zostało wspomniane – bywa nieodłączne rozważaniom nad rozwiązaniami klasycznymi. Ponieważ prawa strona równania (2.70) jest ciągła, więc taką jest również pochodna takiego rozwiązania, czyli $x \in G^1$. Funkcja klasy G^1 na przedziale zwartym jest absolutnie ciągła, zatem rozwiązanie klasyczne jest wtedy C-rozwiązaniem. Z kolei, każda funkcja ciągła jest obszarami ciągła, a więc – w związku z przedstawionym powyżej punktem E – rozwiązanie klasyczne stanowi także F-rozwiązanie. Co więcej, zgodnie z punktem B jest ono również K-rozwiązaniem. Powyższe wnioski pozostają oczywiście prawdziwe także w przypadku, gdy funkcja g ma skończoną liczbę nieciągłości pierwszego rodzaju względem zmiennej niezależnej t, czyli gdy może być stosowana procedura „sklejania" rozwiązań klasycznych. Konkludując:

F) jeżeli funkcja g jest ciągła (dopuszcza się także skończoną liczbę nieciągłości pierwszego rodzaju względem zmiennej niezależnej t), to każde rozwiązanie klasyczne jest C-, F- i K-rozwiązaniem.

Ostatecznie, w praktyce C-, F- i K-rozwiązania są uogólnieniem rozwiązań klasycznych, także z możliwością ich „sklejania".

Przedstawione powyżej **zależności między poszczególnymi typami rozwiązań**, ujęte w postaci punktów A-F, zostały syntetycznie zilustrowane na rysunku 2.4. C-rozwiązania stosuje się najczęściej w przypadku nieciągłości funkcji g względem zmiennej niezależnej t, F-rozwiązania przy nieciągłościach także względem $x(t)$, natomiast K-rozwiązania są uogólnieniem powyższych dwóch typów, aczkolwiek często stanowią zbyt „bogatą" klasę.



Rys. 2.4. Zależności między C-, F-, K- i klasycznymi rozwiązaniami

2.3.2. Przykład

Brak uniwersalnej i naturalnej – z punktu widzenia zastosowań – koncepcji rozwiązania równań różniczkowych z nieciągłą prawą stroną powoduje, iż w wielu pracach problem ten jest niedoprecyzowany lub wręcz pomijany, co może prowadzić do istotnych nieporozumień merytorycznych. W ramach niniejszego przykładu będzie rozważane zagadnienie sterowania optymalnego systemem, którego dynamika jest opisana za pomocą dwuwymiarowego równania różniczkowego

$$\dot{x}_1(t) = x_2(t) + u_1(t), \qquad x_1(0) = x_{01}, \tag{2.82}$$

$$\dot{x}_2(t) = -x_1(t) + u_2(t), \qquad x_2(0) = x_{02}. \tag{2.83}$$

Poszukiwane jest sterowanie, które przeprowadza stan systemu (2.82)-(2.83) ze stanu początkowego takiego, że $\sqrt{x_{01}{}^2 + x_{02}{}^2} \le 2$, do zbioru docelowego $x_k = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$ i minimalizuje wartość wskaźnika jakości J danego wzorem

$$J(u) = \int_{t_0}^{t_k} [(x_1(t)-1)^2 + x_2(t)^2 - 1]^2 [u_1(t)^2 + u_2(t)^2]\, dt, \tag{2.84}$$

gdzie t_k oznacza (swobodny) czas osiągnięcia zbioru docelowego x_k. Postać równania różniczkowego (2.82)-(2.83) nawiązuje do zapisu (2.72) przy oznaczeniach $x(t) = \begin{bmatrix} x_1(t) \\ x_2(t) \end{bmatrix}$ i $u(t) = \begin{bmatrix} u_1(t) \\ u_2(t) \end{bmatrix}$. Zgodnie z terminologią teorii sterowania wartość $x(t)$ reprezentuje stan systemu dynamicznego.

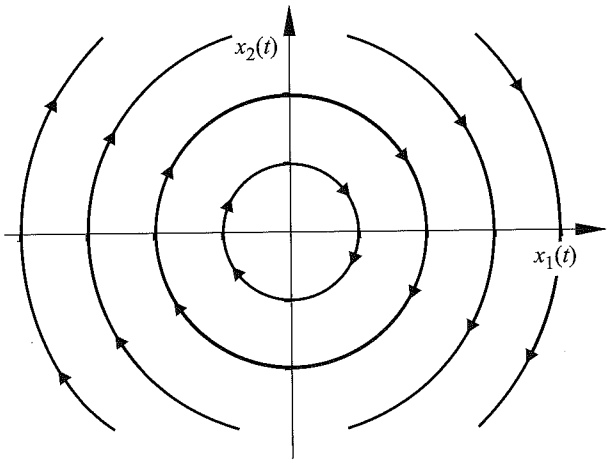
Jeżeli

$$u(t) = \begin{bmatrix} u_1(t) \\ u_2(t) \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \tag{2.85}$$

to zależności (2.82)-(2.83) przyjmują postać

$$\dot{x}_1(t) = x_2(t), \qquad x_1(0) = x_{01}, \tag{2.86}$$

$$\dot{x}_2(t) = -x_1(t), \qquad x_2(0) = x_{02}. \tag{2.87}$$



Rys. 2.5. Rozwiązania równania różniczkowego (2.82)-(2.83) dla sterowania (2.85)

Rozwiązania tego systemu są zilustrowane na rysunku 2.5. Jeżeli natomiast sterowanie zostanie zdefiniowane w układzie zamkniętym wzorem

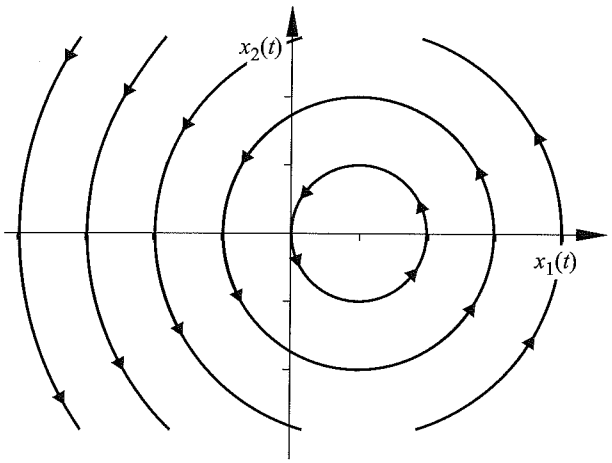
$$u(t)=\begin{bmatrix}u_1(t)\\u_2(t)\end{bmatrix}=\begin{bmatrix}-2x_2(t)\\2x_1(t)-1\end{bmatrix}, \tag{2.88}$$

to zależności (2.82)-(2.83) są dane w postaci

$$\dot{x}_1(t)=-x_2(t), \qquad x_1(0)=x_{01}, \tag{2.89}$$

$$\dot{x}_2(t)=x_1(t)-1, \qquad x_2(0)=x_{02}. \tag{2.90}$$

Otrzymane rozwiązania zostały pokazane na rysunku 2.6.



Rys. 2.6. Rozwiązania równania różniczkowego (2.82)-(2.83) dla sterowania (2.88)

Niech zatem dane będzie następujące sterowanie określone w układzie zamkniętym:

$$u(t)=\begin{bmatrix}u_1(t)\\u_2(t)\end{bmatrix}=\begin{cases}\begin{bmatrix}0\\0\end{bmatrix}, & \text{gdy } x(t)\notin O,\\ \begin{bmatrix}-2x_2(t)\\2x_1(t)-1\end{bmatrix}, & \text{gdy } x(t)\in O,\end{cases} \tag{2.91}$$

przy czym O oznacza okrąg opisany równaniem $[x_1(t)-1]^2+x_2(t)^2=1$, czyli o środku w punkcie $\begin{bmatrix}1\\0\end{bmatrix}$ i promieniu jednostkowym. Jeśli stan systemu nie znajduje się na okręgu O, to jak wynika z rysunku 2.5, występujące w powyższym

wzorze sterowanie $u(t)=\begin{bmatrix}0\\0\end{bmatrix}$ przeprowadza stan do tego okręgu w skończonym czasie. Jeżeli natomiast stan systemu znajduje się na okręgu O, to zgodnie z rysunkiem 2.6 sterowanie $u(t)=\begin{bmatrix}-2x_2(t)\\2x_1(t)-1\end{bmatrix}$ przeprowadza go po tym okręgu do zbioru docelowego również w skończonym czasie. Można stąd sądzić, że wartość wskaźnika jakości (2.84) dla sterowania (2.91) wynosi zero, a ponieważ przyjmuje on jedynie wartości nieujemne, więc ostatecznie powyższe sterowanie minimalizuje wartość rozpatrywanego wskaźnika.

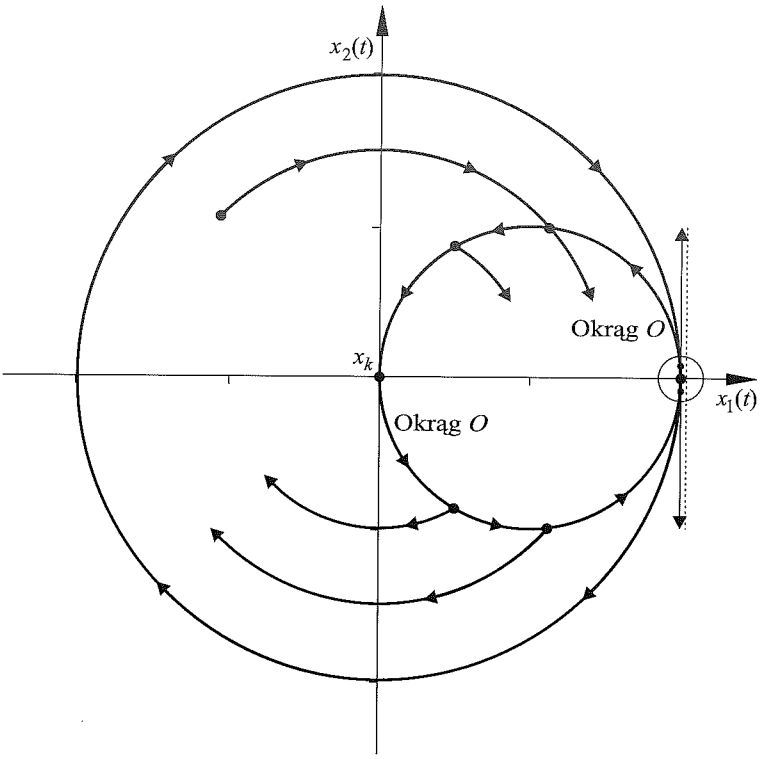
Zostaną teraz szczegółowo rozważone rozwiązania równania różniczkowego (2.86)-(2.87), gdy sterowanie jest dane wzorem (2.91).

Wspomniane wcześniej rozwiązanie, które przeprowadza stan systemu do okręgu O, a następnie po tym okręgu do zbioru docelowego, reprezentuje C-rozwiązanie. Jednak rozwiązanie przeprowadzające stan systemu do okręgu O i przecinające go jest także C-rozwiązaniem, ponieważ rozwiązanie to może nie spełniać równania różniczkowego na zerowej miary podzbiorze przedziału $[t_0, t_k]$. Również z tego powodu, jeżeli stan systemu porusza się po okręgu O, to w każdym momencie może go opuścić. Powyższe elementy mogą się powtarzać w dowolnych konfiguracjach (rys. 2.7).

Na kształt F-rozwiązań nie ma wpływu postać prawej strony równania różniczkowego na zerowej miary podzbiorach przestrzeni $\mathbb{R}^n$, czyli w szczególności na okręgu O, a zatem sterowanie określone wzorem (2.91) jest równoważne w tym przypadku sterowaniu $u(t)=\begin{bmatrix}0\\0\end{bmatrix}$. F-rozwiązania równania różniczkowego (2.82)-(2.83) przy sterowaniu (2.91) są zatem pokazane na rysunku 2.5.

Tak więc w rozważanym przypadku F-rozwiązania są także C-rozwiązaniami (ale zależność przeciwna nie jest prawdziwa). Oczywiście każde C-rozwiązanie jest również K-rozwiązaniem. Jednak postać operatora Krasowskiego w punkcie $\begin{bmatrix}2\\0\end{bmatrix}$, w szczególności stanowiąca, iż zero należy do zbioru jego wartości (por. rys. 2.7), implikuje, że K-rozwiązanie może także „zatrzymać" się w tym punkcie, a potem „ruszyć" w dowolnym momencie. Ponieważ istnieje możliwość powrotu stanu systemu do tego punktu, powyższe może się wielokrotnie powtarzać (rys. 2.7).

Ostatecznie, szczegółowa analiza zagadnienia rozwiązań równania różniczkowego (2.82)-(2.83) wskazuje, że sterowania określonego wzorem (2.91) nie można jednak uznać za optymalne, gdyż generowane przez to sterowanie F-rozwiązania w ogóle nie przeprowadzają stanu systemu do zbioru docelowego, natomiast C- i K-rozwiązania nie są jednoznaczne, a więc wskaźnik jakości (2.84) byłby wtedy funkcją wieloznaczną (wielowartościową). Rozważane zadanie ste-



Rys. 2.7. Rozwiązania równania różniczkowego (2.82)-(2.83) przy sterowaniu (2.91)

rowania optymalnego, pozornie zakończone wyznaczeniem wzoru (2.91), okazało się wymagającym istotnego doprecyzowania.

Przedstawiony powyżej materiał zostanie wykorzystany w punkcie 4.1.4 przy rozważaniach nad wrażliwością układu sterowania optymalnego na niedokładność modelu obiektu, co w konsekwencji umożliwi wyznaczenie – z zastosowaniem estymatorów jądrowych – wartości parametrów tego modelu, minimalizujących straty wynikłe z błędów identyfikacji.

Uwagi bibliograficzne

Teoretyczne aspekty rozważanych w niniejszym podrozdziale zagadnień dotyczących równań różniczkowych z nieciągłą prawą stroną zostały zebrane i zwięźle opisane w przeglądowym artykule [24]. Przykład przedstawiony w punkcie 2.3.2 jest oparty na publikacji [25]. Klasyczną teorię równań różniczkowych zawierają monografie [51] o zmatematyzowanym charakterze oraz [49] eksponująca aspekty praktyczne. Wykorzystywane powyżej elementy analizy matematycznej i funkcjonalnej można znaleźć w podręcznikach [32, 48], a funkcje absolutnie (bezwzględnie) ciągłe są przedstawione w książce [44].

Rozdział 3

Estymatory jądrowe

W niniejszym rozdziale zostanie szczegółowo przedstawiona metodyka konstruowania estymatorów jądrowych. Jej zasadnicza koncepcja wywodzi się z problemu wyznaczania gęstości prawdopodobieństwa zmiennej losowej i temu właśnie tematowi będzie poświęcony – główny w tej książce – podrozdział 3.1. Zostaną w nim przedstawione zarówno podstawowe definicje oraz procedury, jak i szereg dodatkowych algorytmów polepszających własności otrzymanych estymatorów i dopasowujących je do specyfiki badanych problemów. Prezentowaną koncepcję można również wykorzystać w celu otrzymania innych charakterystyk zmiennej losowej, które mogą okazać się przydatne w konkretnych problemach aplikacyjnych. Tego zagadnienia dotyczy podrozdział 3.2, którego przedmiotem są zadania estymacji dystrybuanty, kwantyla i charakterystyk warunkowych.

Prezentowany materiał jest przedstawiony w takiej formie, aby **dostarczyć kompletnych i sprawdzonych procedur**, możliwych do bezpośredniego stosowania przez specjalistów z rozlicznych dziedzin niekoniecznie związanych z matematyką, traktujących statystykę lub analizę danych jako narzędzie różnorodnych i interdyscyplinarnych badań. Dzięki swym zaletom analitycznym oraz przejrzystej interpretacji estymatory jądrowe wydają się szczególnie dogodne do realizacji tak sformułowanego celu.

W dalszej części niniejszej książki **wprowadza się notację, w której** x_1, x_2, ..., x_n **oznaczają kolejne współrzędne** n**-wymiarowego wektora** x:

$$x = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}; \tag{3.1}$$

są one składane odmienną czcionką w celu odróżnienia od elementów próby losowej.

3.1. Jądrowy estymator gęstości prawdopodobieństwa

Gęstość prawdopodobieństwa stanowi naturalną, bliską intuicji i często najdogodniejszą w praktycznych badaniach charakterystykę funkcyjną rozkładu prawdopodobieństwa. Jej interpretacja była przedstawiona w rozdziale 1, a następnie – po sformułowaniu definicji – została uściślona w punkcie 2.1.2. Warto przypomnieć, że z analitycznego punktu widzenia znajomość gęstości pozwala obliczyć prawdopodobieństwo, iż badana zmienna losowa przyjmie wartość z ustalonego zbioru, a sam wykres funkcji może być także wykorzystywany do ilustracji rozpatrywanego zagadnienia – im większa jest bowiem jej wartość w określonym punkcie dziedziny, tym większe staje się prawdopodobieństwo przyjęcia przez badaną zmienną losową wartości bliskich temu punktowi. W praktycznych zadaniach gęstość prawdopodobieństwa najczęściej nie jest znana i musi być wyznaczona na podstawie uzyskanych eksperymentalnie danych.

3.1.1. Definicje i interpretacje

Rozważana będzie n-wymiarowa zmienna losowa X. Zgodnie z wprowadzonym w punkcie 2.1.2 ogólnym założeniem rozkład tej zmiennej ma gęstość prawdopodobieństwa f. Jej estymator będzie wyznaczany na podstawie wartości m-elementowej próby losowej

$$x_1, x_2, \ldots, x_m, \tag{3.2}$$

które są interpretowane jako doświadczalnie uzyskane, w trakcie niezależnych eksperymentów, wartości zmiennej X.

Jądrowy estymator gęstości prawdopodobieństwa $\hat{f}: \mathbb{R}^n \to [0, \infty)$ w podstawowej postaci definiuje się wzorem

$$\hat{f}(x) = \frac{1}{mh^n} \sum_{i=1}^{m} K\left(\frac{x - x_i}{h}\right), \tag{3.3}$$

gdzie $m \in \mathbb{N} \setminus \{0\}$ oznacza liczność próby losowej, dodatnia liczba rzeczywista h jest nazywana **parametrem wygładzania**, a borelowska funkcja $K: \mathbb{R}^n \to [0, \infty)$ spełnia warunek

$$\int_{\mathbb{R}^n} K(x)\, dx = 1 \tag{3.4}$$

i jest określana mianem **jądra**. Zakłada się ponadto, iż funkcja K jest symetryczna względem zera, czyli

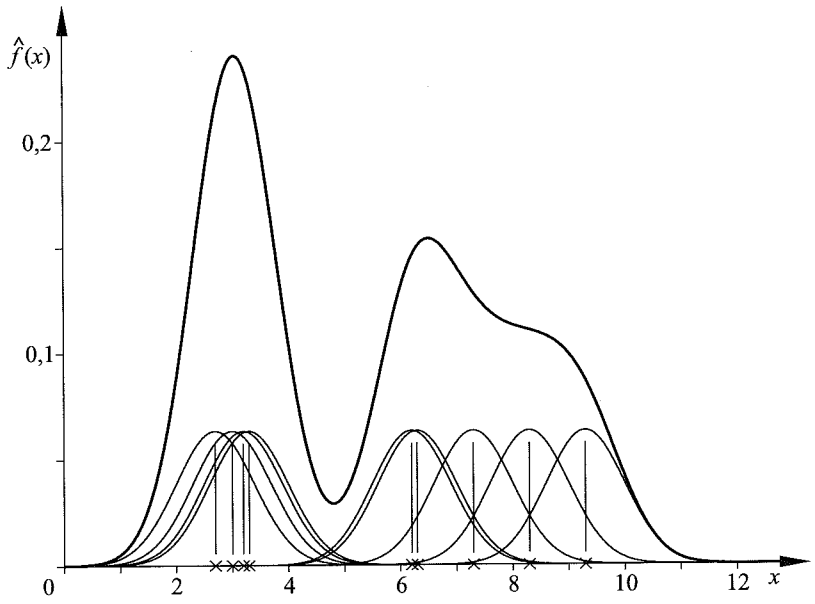
$$K(x) = K(-x) \quad \text{dla każdego} \quad x \in \mathbb{R}^n, \tag{3.5}$$

i ma w tym punkcie (słabe) maksimum globalne:

$$K(0) \geq K(x) \quad \text{dla każdego} \quad x \in \mathbb{R}^n. \tag{3.6}$$

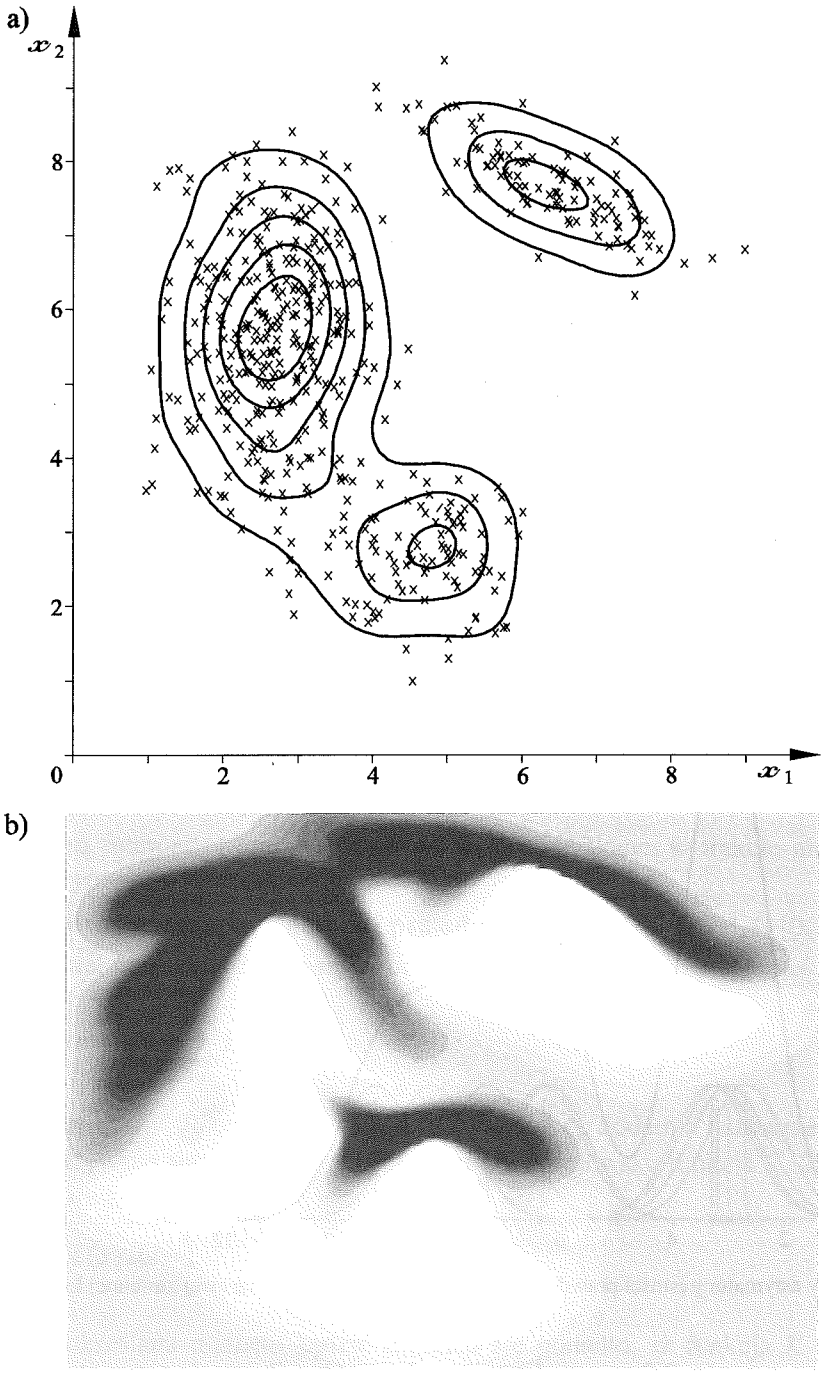
Jej nazwa powstała przez analogię do jądra operatora całkowego. Dogodne w praktycznych zastosowaniach reguły wyboru postaci funkcji K oraz wyznaczania wartości parametru h zostaną szczegółowo przedstawione w dalszej części książki, zwłaszcza w punktach 3.1.3 i 3.1.5.

Interpretacja powyższej definicji została zilustrowana na rysunku 3.1 dla jednowymiarowej zmiennej losowej, to znaczy przy $n = 1$. W przypadku pojedynczego elementu x_i funkcja K, przesunięta o wektor x_i oraz „przeskalowana" współczynnikiem h, reprezentuje oszacowanie rozkładu zmiennej losowej X po uzyskaniu wartości x_i. Dla m niezależnych wartości $x_1, x_2, ..., x_m$ przyjmuje ono postać sumy takich pojedynczych oszacowań. Współczynnik $1/mh^n$ normuje uzyskaną funkcję w celu zagwarantowania warunku $\int_{\mathbb{R}^n} \hat{f}(x)\,dx = 1$, jaki musi spełniać miara probabilistyczna (por. wzór (2.4)).



Rys. 3.1. Jądrowy estymator gęstości prawdopodobieństwa (3.3) jednowymiarowej zmiennej losowej

Z kolei na rysunku 3.2 są przedstawione dwie różne ilustracje estymatora jądrowego dla dwuwymiarowej zmiennej losowej: za pomocą poziomic i jako widok perspektywiczny wykresu funkcji dwuwymiarowej.



Rys. 3.2. Jądrowy estymator gęstości prawdopodobieństwa (3.3) dwuwymiarowej zmiennej losowej: a) poziomice (krzyżykami są zaznaczone wartości próby losowej), b) widok perspektywiczny

Należy podkreślić, że dzięki przyjętej postaci powyższej definicji estymator jądrowy nie jest ograniczony przez arbitralne założenie do żadnego konkretnego typu rozkładu. Fakt ten był już wielokrotnie sygnalizowany jako ogólna cecha estymacji nieparametrycznej. Estymator jądrowy umożliwia oszacowanie gęstości prawdopodobieństwa większości występujących w praktyce rozkładów, aczkolwiek sprawdzenia wymaga dokładność owego oszacowania. Przedstawiona poniżej analiza błędu estymacji nie tylko dostarcza informacji w tym zakresie, ale pozwala również sformułować wiele użytecznych wniosków na temat wyboru postaci funkcji K, wyznaczania wartości parametru h, a także ewentualnego zastosowania dodatkowych procedur, uogólniających podstawową postać (3.3), które umożliwiłyby polepszenie jakości otrzymanego estymatora oraz dostosowanie go do uwarunkowań i wymagań konkretnego zadania aplikacyjnego.

3.1.2. Analiza błędu estymacji

W niniejszym punkcie są przedstawione teoretyczne podstawy stosowania estymatorów jądrowych. Wynikające stąd wnioski praktyczne będą sukcesywnie prezentowane w kolejnych punktach tego podrozdziału. Jeżeli jednak Czytelnik nie jest zainteresowany ich matematycznym uzasadnieniem, to może ten punkt pominąć.

Podstawowy, powszechnie stosowany miernik jakości estymatorów stanowi **kryterium błędu średniokwadratowego**. W przypadku rzeczywistego parametru b i jego estymatora $\hat{b}$ wartością tego kryterium – oznaczanego poniżej przez MSE (*Mean Square Error*) – jest wartość oczekiwana kwadratu błędu estymacji, co zapisuje się także jako

$$\mathrm{MSE} = \mathrm{E}((\hat{b} - b)^2). \tag{3.7}$$

Równoważną, dogodną do interpretacji postać powyższego wzoru stanowi zależność

$$\mathrm{MSE} = [\mathrm{E}(\hat{b}) - b]^2 + V(\hat{b}). \tag{3.8}$$

Wartość błędu średniokwadratowego MSE jest więc sumą kwadratu obciążenia estymatora $\hat{b}$ i jego wariancji. Pierwszy składnik wskazuje zatem, jak odległy od prawdziwej wartości jest „środek" uzyskiwanych w praktyce wartości estymatora, a drugi – jaki jest stopień ich „rozrzutu" względem owego „środka".

W przypadku gdy jest rozważana jakość estymacji gęstości prawdopodobieństwa f, n-wymiarowej zmiennej losowej, kryterium błędu średniokwadratowego (3.7) może być stosowane przy ustalonej wartości argumentu x, czyli

$$\mathrm{MSE}_x = \mathrm{E}([\hat{f}(x) - f(x)]^2) \text{ dla dowolnie ustalonego } x \in \mathbb{R}^n \tag{3.9}$$

lub równoważnie

$$\mathrm{MSE}_x = [\mathrm{E}(\hat{f}(x)) - f(x)]^2 + V(\hat{f}(x)) \text{ dla dowolnie ustalonego } x \in \mathbb{R}^n. \quad (3.10)$$

Globalny wskaźnik jakości estymacji owej gęstości otrzymuje się w wyniku scałkowania powyższych wielkości na przestrzeni $\mathbb{R}^n$. Przyjmuje on wtedy formę **kryterium scałkowanego błędu średniokwadratowego** MISE (*Mean Integrated Square Error*):

$$\mathrm{MISE} = \int_{\mathbb{R}^n} \mathrm{E}([\hat{f}(x) - f(x)]^2)\, dx \quad (3.11)$$

lub równoważnie

$$\mathrm{MISE} = \int_{\mathbb{R}^n} [[\mathrm{E}(\hat{f}(x)) - f(x)]^2 + V(\hat{f}(x))]\, dx. \quad (3.12)$$

Wzór (3.12) stanowi dogodny aparat matematyczny do oceny dokładności estymacji gęstości prawdopodobieństwa f. Wartość scałkowanego błędu średniokwadratowego MISE jest – podobnie jak poprzednio – sumą dwóch składników wynikłych ze scałkowania kwadratu obciążenia estymatora $\hat{f}$ i jego wariancji.

W przypadku estymatorów jądrowych kryterium MISE dostarcza wartościowych wniosków aplikacyjnych, dotyczących zwłaszcza wyboru postaci jądra K oraz wyznaczania wartości parametru wygładzania h. W tym celu są czynione dodatkowe założenia dotyczące funkcji f oraz K.

Najpierw rozważany będzie przypadek jednowymiarowej zmiennej losowej. Zakłada się, że estymowana gęstość f ma ciągłą drugą pochodną

$$f'' \in C^2, \quad (3.13)$$

druga pochodna jest całkowalna z kwadratem:

$$f'' \in \mathrm{L}^2, \quad (3.14)$$

a także iż istnieje $\tilde{x} \geq 0$, takie że funkcja f jest niemalejąca w przedziale $(-\infty, -\tilde{x}]$ i nierosnąca w przedziale $[\tilde{x}, \infty)$, czyli

$$\begin{aligned} \exists\, \tilde{x} \geq 0: \quad & f(x_*) \leq f(x_{**}), \quad \text{gdy} \quad x_* \leq x_{**} \leq -\tilde{x}, \\ & f(x_*) \geq f(x_{**}), \quad \text{gdy} \quad \tilde{x} \leq x_* \leq x_{**}. \end{aligned} \quad (3.15)$$

Borelowska funkcja K powinna oprócz warunków (3.4)-(3.6) spełniać także zależność

$$\int_{-\infty}^{\infty} x^2 K(x)\, dx < \infty. \quad (3.16)$$

W literaturze można także spotkać nieco odmienne wersje powyższych założeń. Z aplikacyjnego punktu widzenia są one praktycznie równoważne.

Niech ponadto dane będą wielkości $U(K)$ i $W(K)$ charakteryzujące jądro K:

$$U(K) = \int_{-\infty}^{\infty} x^2 K(x)\, dx, \tag{3.17}$$

$$W(K) = \int_{-\infty}^{\infty} K(x)^2\, dx \tag{3.18}$$

oraz $Z(f)$ dotycząca estymowanej gęstości f

$$Z(f) = \int_{-\infty}^{\infty} f''(x)^2\, dx. \tag{3.19}$$

Skończona wartość wyrażenia $U(K)$ jest zagwarantowana przez warunek (3.16). Z kolei zależność

$$W(K) = \int_{-\infty}^{\infty} K(x)^2\, dx \le K(0) \int_{-\infty}^{\infty} K(x)\, dx = K(0), \tag{3.20}$$

prawdziwa na podstawie wzorów (3.4) oraz (3.6), stanowi o skończonej wartości wyrażenia $W(K)$. I wreszcie, przyjęte powyżej założenie (3.14) implikuje skończoność $Z(f)$.

Obciążenie estymatora jądrowego (3.3) w dowolnie ustalonym punkcie x wyraża się wzorem

$$\mathrm{E}(\hat{f}(x)) - f(x) = \frac{1}{2} h^2 U(K) f''(x) + o(h^2), \tag{3.21}$$

przy czym funkcja $o(h^2)$ spełnia warunek

$$\lim_{h \to 0} \frac{o(h^2)}{h^2} = 0. \tag{3.22}$$

Jeżeli zatem parametr wygładzania h byłby dobierany tak, aby

$$\lim_{m \to \infty} h = 0, \tag{3.23}$$

to – na podstawie zależności (3.21) – obciążenie estymatora jądrowego będzie dążyć do zera w miarę wzrostu liczności próby losowej, a więc estymator ten będzie asymptotycznie nieobciążony.

Z kolei wariancja estymatora jądrowego (3.3) dla ustalonego x wyraża się wzorem

$$V(\hat{f}(x)) = \frac{1}{mh} W(K) f(x) + o((mh)^{-1}), \tag{3.24}$$

przy czym

$$\lim_{h\to 0}\frac{o((mh)^{-1})}{(mh)^{-1}}=0. \tag{3.25}$$

Jeśli parametr h jest dobierany tak, iż

$$\lim_{m\to\infty}(mh)^{-1}=0, \tag{3.26}$$

czyli równoważnie

$$\lim_{m\to\infty}mh=\infty, \tag{3.27}$$

to – na podstawie zależności (3.24) – wariancja estymatora jądrowego dąży do zera przy wzroście liczności próby losowej.

Warunek (3.23) oznacza, że parametr wygładzania h powinien być dobierany tak, aby w miarę wzrostu liczności próby jego wartość zmniejszała się, dążąc do zera. Warunek (3.27) ogranicza jednak „szybkość" owego zmniejszania. Jednoczesne spełnienie obu tych warunków gwarantuje, że w każdym punkcie dziedziny funkcji gęstości prawdopodobieństwa obciążenie i wariancja estymatora jądrowego (3.3) maleją do zera w miarę zwiększania się liczności próby losowej. Wskutek zmniejszania się obciążenia otrzymywane w praktyce wartości estymatora jądrowego $\hat{f}(x)$ będą „oscylować" wokół wartości znajdującej się coraz bliżej prawdziwej (aczkolwiek nieznanej) wartości funkcji gęstości $f(x)$, a dzięki zmniejszaniu wariancji zmniejszać się także będzie wielkość owych „oscylacji".

Jeżeli więc w dowolnie ustalonym punkcie x są spełnione warunki (3.23) i (3.27), to **estymator jądrowy jest asymptotycznie nieobciążony, a jego wariancja maleje do zera**, gdy zwiększa się liczność próby losowej. Zatem – zgodnie z własnościami estymatorów, przedstawionymi w punkcie 2.2.2 – **estymator jądrowy jest także zgodny**. Jak wynika ze wzoru (3.10) wraz z zależnościami (3.21) i (3.24), w miarę zwiększania liczności próby również **wartość błędu średniokwadratowego $\mathbf{MSE}_x$ maleje wtedy do zera**. Powyższe własności oznaczają, iż zwiększanie próby losowej pozwala na sukcesywne polepszanie dokładności estymacji, aczkolwiek w praktyce proces ten jest ograniczony dostępnością danych.

Warunki (3.23) i (3.27) są łatwe do spełnienia, na przykład przez przyjęcie

$$h=\frac{\alpha}{m^{\beta}}, \tag{3.28}$$

gdzie $\alpha > 0$ oraz $\beta \in (0, 1)$. Są one w zasadzie nieodłącznie związane z wszelkimi rozważaniami nad teorią i zastosowaniami estymatorów jądrowych.

Podstawiając wyrażenia (3.21) i (3.24) do wzoru (3.12), można otrzymać wartość scałkowanego błędu średniokwadratowego MISE, wynoszącą

$$\mathrm{MISE} = \frac{1}{4}h^4 U(K)^2 Z(f) + \frac{1}{mh}W(K) + o(h^4) + o((mh)^{-1}), \tag{3.29}$$

czyli w przybliżeniu, po uwzględnieniu warunków (3.23) i (3.27) w połączeniu z własnościami (3.22) i (3.25):

$$\mathrm{MISE} = \frac{1}{4}h^4 U(K)^2 Z(f) + \frac{1}{mh}W(K). \tag{3.30}$$

Analizując prawą stronę powyższej zależności, warto zauważyć, iż zmniejszanie wartości parametru wygładzania h powoduje co prawda zmniejszenie się wartości pierwszego wyrażenia (związanego z obciążeniem), ale jednocześnie skutkuje zwiększeniem wartości drugiego (związanego z wariancją). I odwrotnie: zwiększanie wartości parametru h implikuje zwiększenie się wartości pierwszego wyrażenia, ale jednocześnie zmniejszenie wartości drugiego. Odzwierciedla to częsty w złożonych zagadnieniach statystycznych problem konieczności znajdowania kompromisu między obciążeniem a wariancją estymatora. Naturalne rozwiązanie tego problemu stanowi ustalenie równowagi, polegającej na przyjęciu pośredniej wartości parametru wygładzania, która minimalizowałaby wartość scałkowanego błędu średniokwadratowego MISE, czyli – zgodnie ze wzorem (3.12) – sumę kwadratu obciążenia i wariancji. Nietrudne obliczenia implikują, iż wartością, przy której wyrażenie (3.30) osiąga minimum, jest

$$h_o = \left(\frac{W(K)}{U(K)^2 Z(f) m} \right)^{1/5}. \tag{3.31}$$

Dla tak określonego parametru wygładzania wartość kryterium MISE wynosi

$$\mathrm{MISE} = c_1 \left(\frac{U(K)^2 W(K)^4 Z(f)}{m^4} \right)^{1/5}, \tag{3.32}$$

gdzie c_1 oznacza pewną dodatnią stałą, nieistotną z punktu widzenia dalszych rozważań. Wzory (3.31)-(3.32) nie mogą być użyte bezpośrednio, gdyż ich prawe strony zależą od podlegającej estymacji, nieznanej funkcji f. Stanowią one jednak dogodną teoretyczną podstawę do opracowania praktycznych reguł wyboru postaci jądra K oraz procedur wyznaczania wartości parametru wygładzania h. Będą one szczegółowo zaprezentowane w dalszej części pracy, zwłaszcza w punktach 3.1.3 i 3.1.5.

Przedstawione powyżej rozważania dotyczące jednowymiarowej zmiennej losowej zostaną teraz uogólnione na przypadek wielowymiarowy. Założenia czynione tu w odniesieniu do estymowanej gęstości prawdopodobieństwa f oraz jądra K przyjmują zróżnicowaną dla poszczególnych wariantów, często złożoną z formalnego punktu widzenia postać i w związku z tym zostaną pominięte. Nie wykraczają one jednak poza naturalne przekształcenie warunków (3.13)-(3.16), zaprezentowanych wcześniej dla przypadku jednowymiarowego.

Wielkości zdefiniowane wzorami (3.17)-(3.19) są tu uogólniane do postaci

$$U(K) = \int_{\mathbb{R}^n} x^{\mathrm{T}} x K(x)\, dx, \tag{3.33}$$

$$W(K) = \int_{\mathbb{R}^n} K(x)^2\, dx, \tag{3.34}$$

$$Z(f) = \int_{\mathbb{R}^n} [\nabla^2 f(x)]^2\, dx, \tag{3.35}$$

przy czym w nawiązaniu do oznaczenia (3.1):

$$\nabla^2 f(x) = \sum_{i=1}^{n} \frac{\partial^2 f(x)}{\partial . x_i^2}. \tag{3.36}$$

Także tu, dla dowolnie ustalonego punktu x, zależność (3.23) zapewnia asymptotyczną nieobciążoność estymatora jądrowego (3.3). Z kolei warunek (3.27) przyjmuje postać

$$\lim_{m\to\infty} mh^n = \infty, \tag{3.37}$$

a jego spełnienie gwarantuje, że w miarę zwiększania liczności próby losowej wariancja tego estymatora dąży do zera. Zachodzenie obu warunków implikuje zatem własność zgodności, a także to, iż wartość błędu średniokwadratowego MSE_x maleje do zera.

Wartość minimalizująca kryterium scałkowanego błędu średniokwadratowego MISE jest dana – z dokładnością do przejścia granicznego wynikłego z warunków (3.27) i (3.37) – następującym wzorem:

$$h_0 = \left(\frac{nW(K)}{U(K)^2 Z(f) m} \right)^{1/(n+4)}, \tag{3.38}$$

a wartość tego kryterium wynosi wtedy

$$\mathrm{MISE} = c_n \left(\frac{U(K)^{2n} W(K)^4 Z(f)^n}{m^4} \right)^{1/(n+4)}, \tag{3.39}$$

przy czym dodatnia stała c_n jest nieistotna dla dalszych rozważań. Warto zwrócić uwagę, że wzory (3.38)-(3.39) stanowią odpowiednio uogólnienie zależności (3.31)-(3.32), gdyby przyjąć $n = 1$.

3.1.3. Wybór postaci jądra

Jak było wspomniane, przedstawiona w poprzednim punkcie analiza błędu estymacji pozwala między innymi sformułować wnioski dotyczące wyboru postaci jądra K.

Najpierw będzie rozważany przypadek jednowymiarowej zmiennej losowej. Z teoretycznego punktu widzenia naturalne jest wymaganie minimalizacji wartości scałkowanego błędu średniokwadratowego (3.32). W przypadku zadania wyboru postaci funkcji K wymaganie to implikuje konieczność takiego jej określenia, aby czynniki prawej strony równości (3.32), zależne od K, były jak najmniejsze. Oznacza to więc żądanie minimalizacji wyrażenia $U(K)^2 W(K)^4$. Jest ono spełnione przez tak zwane **jądro Epanecznikowa**:

$$K(x) = \begin{cases} \frac{3}{4}(1-x^2) & \text{dla} \quad x \in [-1,\ 1], \\ 0 & \text{dla} \quad x \in (-\infty, -1) \cup (1, \infty). \end{cases} \tag{3.40}$$

Wywodzi się ono z ogólnej rodziny jąder, zdefiniowanej zależnością

$$K(x) = \begin{cases} \frac{1}{v_d}(1-x^2)^d & \text{dla} \quad x \in [-1,\ 1], \\ 0 & \text{dla} \quad x \in (-\infty, -1) \cup (1, \infty), \end{cases} \tag{3.41}$$

gdzie $d \in \mathbb{N}$, a v_d jest nietrudną do wyznaczenia, dla konkretnego d, dodatnią stałą daną wzorem $v_d = 2\int_0^1 (1-x^2)^d dx$. W szczególności, gdy $d = 0$, otrzymuje się **jądro jednostajne**

$$K(x) = \begin{cases} \frac{1}{2} & \text{dla} \quad x \in [-1,\ 1], \\ 0 & \text{dla} \quad x \in (-\infty, -1) \cup (1, \infty), \end{cases} \tag{3.42}$$

dla $d = 1$ wzór (3.41) przedstawia jądro Epanecznikowa (3.40), a w przypadku $d = 2$ **jądro dwuwagowe**

$$K(x)=\begin{cases}\frac{15}{16}(1-x^2)^2 & \text{dla} \quad x\in[-1,\,1],\\ 0 & \text{dla} \quad x\in(-\infty,-1)\cup(1,\infty).\end{cases} \tag{3.43}$$

Innymi często stosowanymi postaciami są **jądro normalne**

$$K(x)=\frac{1}{\sqrt{2\pi}}\exp\left(-\frac{x^2}{2}\right) \tag{3.44}$$

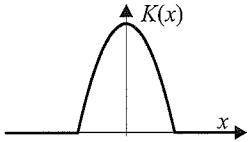
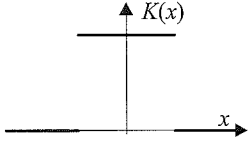
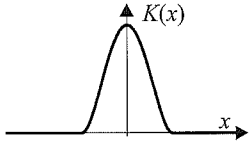
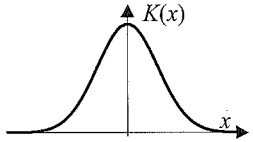
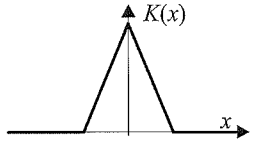
oraz **jądro trójkątne**

$$K(x)=\begin{cases}1-|x| & \text{dla} \quad x\in[-1,\,1],\\ 0 & \text{dla} \quad x\in(-\infty,-1)\cup(1,\infty).\end{cases} \tag{3.45}$$

Ich wzory są oparte na postaciach gęstości prawdopodobieństwa typowych rozkładów (por. rys. 2.2), odpowiednio dostosowanych w celu zapewnienia warunków (3.5) i (3.6), a w przypadku ograniczonego nośnika – ustalenia go w dogodnej formie przedziału [–1, 1] (można zauważyć, że rodzina (3.41) powstała przez modyfikację rozkładu beta).

Zdefiniowane powyżej podstawowe postacie (typy) jąder ujęte są w formie stabelaryzowanej na rysunku 3.3. Zostały tam również przedstawione wartości wyrażenia $U(K)^2W(K)^4$ charakteryzującego jakość danego jądra. Następna kolumna zawiera wartość relatywną tych wielkości względem optymalnego jądra Epanecznikowa. Z kolei – ponieważ liczność próby losowej n występuje we wzorze (3.32) w czwartej potędze, a całe wyrażenie jest podniesione do potęgi 1/5 – w kolejnej kolumnie została uwidoczniona powyższa wartość relatywna podniesiona do potęgi 5/4. I wreszcie, wielkość ta odjęta od jedności i wyrażona w procentach, podana w ostatniej kolumnie, stanowi o **efektywności** danego jądra względem jądra Epanecznikowa. Z praktycznego punktu widzenia wskazuje ona, o ile należy zwiększyć liczność próby, aby uzyskać jakość estymacji otrzymywaną z użyciem optymalnego jądra Epanecznikowa.

Faktem o podstawowym – dla praktycznych zastosowań – znaczeniu jest to, iż efektywności poszczególnych jąder niewiele różnią się między sobą. Najprostsze z uwidocznionych na rysunku 3.3 postaci, mianowicie jądro jednostajne, jest jedynie o 7% mniej efektywne niż optymalne jądro Epanecznikowa. W przypadku innych typów różnica ta jest jeszcze mniejsza, w praktyce wręcz pomijalna. Pojawia się zatem możliwość, aby **przy wyborze funkcji K uwzględniać przede wszystkim własności otrzymanego estymatora**, na przykład klasę regularności, prostotę obliczeniową, ograniczoność nośnika lub też inne cechy istotne z punktu widzenia konkretnego zagadnienia będącego przedmiotem badań.

Typ jądra K	Wykres	$U(K)$	$W(K)$	$U(K)^2W(K)^4$	$\frac{U(K_E)^2W(K_E)^4}{U(K)^2W(K)^4}$	$\left[\frac{U(K_E)^2W(K_E)^4}{U(K)^2W(K)^4}\right]^{\frac{5}{4}}$	Zmniejszenie efektywności względem jądra Epanecznikowa
Epanecznikowa		$\frac{1}{5}$	$\frac{3}{5}$	0,00518	1	1	0%
Jednostajne		$\frac{1}{3}$	$\frac{1}{2}$	0,00694	0,746	0,929	7%
Dwuwagowe		$\frac{1}{7}$	$\frac{5}{7}$	0,00531	0,976	0,994	1%
Normalne		1	$\frac{1}{2\sqrt{\pi}}$	0,00633	0,818	0,951	5%
Trójkątne		$\frac{1}{6}$	$\frac{2}{3}$	0,00549	0,943	0,986	1%

Rys. 3.3. Podstawowe typy jąder oraz odpowiadające im wartość wyrażenia $U(K)^2W(K)^4$ i wielkości pokrewne (K_E oznacza jądro Epanecznikowa)

W szczególności, jądro normalne ma pochodną dowolnego rzędu, lecz jego całka nie jest wyrażalna wzorem analitycznym. W przypadku pozostałych jąder całka jest nietrudna do obliczenia, ale pochodna nie istnieje w całej dziedzinie, a w przypadku jądra jednostajnego jest – w punktach, w których istnieje – stale równa zeru. Powyższe własności będą także prawdziwe dla skonstruowanego z ich wykorzystaniem estymatora jądrowego (3.3).

Nieduża z praktycznego punktu widzenia różnica efektywności poszczególnych typów jąder umożliwia szeroki ich wybór, co pozwala uzyskać dogodne cechy skonstruowanego estymatora. Można przecież proponować dowolną postać funkcji K, także spoza klasycznych typów przedstawionych na rysunku 3.3 – formułowane względem tej funkcji ogólne warunki (3.4)-(3.6) i (3.16) dopuszczają dużą elastyczność. Co więcej, możliwe jest użycie **kombinacji liniowej jąder**. Na przykład jądro dane wzorem

$$K(x)=\frac{9}{10}K_{\mathrm{E}}(x)+\frac{1}{10}K_{\mathrm{N}}(x), \tag{3.46}$$

gdzie K_{E} oznacza jądro Epanecznikowa (3.40) oraz K_{N} jądro normalne (3.44), ma efektywność zbliżoną do – dominującego tu dzięki współczynnikowi 9/10 – jądra Epanecznikowa, lecz z uwagi na „domieszkę" jądra normalnego ma nieograniczony nośnik. Należy zaznaczyć, iż wartości stałych definiujących jądro Epanecznikowa i jądro normalne (np. parametru wygładzania h) powinny być wyznaczane za pomocą oddzielnych obliczeń.

Podsumowując, warto podkreślić, że możliwość znacznej elastyczności przy doborze postaci jądra K jest olbrzymią praktyczną zaletą, ujawniającą się proporcjonalnie do złożoności rozważanego problemu aplikacyjnego.

Zagadnienie wyboru postaci jądra zostanie teraz przeniesione na przypadek wielowymiarowej zmiennej losowej. Stosowane tutaj konstrukcje odwołują się bezpośrednio do koncepcji przedstawionych dla przypadku jednowymiarowego. Niech zatem K oznacza rozważane poniżej jądro n-wymiarowe, podczas gdy – dla uniknięcia kolizji oznaczeń – omawiane wcześniej jądro jednowymiarowe będzie teraz oznaczane jako $\mathscr{K}$.

W praktycznych zastosowaniach stosuje się dwa naturalne ujęcia: radialne i produktowe.

Podstawowym jest **jądro radialne**, dane wzorem

$$K(x)=c_{n,\mathscr{K}}\cdot\mathscr{K}\left(\sqrt{x^{\mathrm{T}}x}\right), \tag{3.47}$$

gdzie $c_{n,\mathscr{K}}$ oznacza dodatnią stałą taką, aby jądro K spełniało warunek (3.4), natomiast przy wprowadzonym wcześniej oznaczeniu (3.1): $x^{\mathrm{T}}x=x_1^2+x_2^2+\ldots+x_n^2$. Podstawowe postacie jąder radialnych są podane w formie stabelaryzowanej na

Typ jądra radialnego	$K(x)$	$W(K)$	$U(K)$	$\frac{W(K)}{U(K)^2}$
Epanecznikowa	$\begin{cases} \frac{n+2}{2V_n}(1-x^{\mathrm{T}}x), & \text{gdy } x^{\mathrm{T}}x \le 1 \\ 0, & \text{gdy } x^{\mathrm{T}}x > 1 \end{cases}$	$\frac{2(n+2)}{V_n(n+4)}$	$\frac{n}{n+4}$	$\frac{2(n+2)(n+4)}{V_n n^2}$
Jednostajne	$\begin{cases} \frac{1}{V_n}, & \text{gdy } x^{\mathrm{T}}x \le 1 \\ 0, & \text{gdy } x^{\mathrm{T}}x > 1 \end{cases}$	$\frac{1}{V_n}$	$\frac{n}{n+2}$	$\frac{(n+2)^2}{V_n n^2}$
Dwuwagowe	$\begin{cases} \frac{(n+2)(n+4)}{8V_n}(1-x^{\mathrm{T}}x)^2, & \text{gdy } x^{\mathrm{T}}x \le 1 \\ 0, & \text{gdy } x^{\mathrm{T}}x > 1 \end{cases}$	$\frac{6(n+2)(n+4)}{V_n(n+6)(n+8)}$	$\frac{n}{n+6}$	$\frac{6(n+2)(n+4)(n+6)}{V_n n^2(n+8)}$
Normalne	$(2\pi)^{-n/2}\exp\left(-\frac{1}{2}x^{\mathrm{T}}x\right)$	$\frac{1}{2^n \pi^{n/2}}$	n	$\frac{1}{2^n \pi^{n/2} n^2}$
Trójkątne	$\begin{cases} \frac{n+1}{V_n}\left(1-\sqrt{x^{\mathrm{T}}x}\right), & \text{gdy } x^{\mathrm{T}}x \le 1 \\ 0, & \text{gdy } x^{\mathrm{T}}x > 1 \end{cases}$	$\frac{2(n+1)}{V_n(n+2)}$	$\frac{n(n+1)}{(n+2)(n+3)}$	$\frac{2(n+2)(n+3)^2}{V_n n^2(n+1)}$

V_n oznacza objętość n-wymiarowej kuli jednostkowej w przestrzeni R^n, daną wzorem $V_n = \begin{cases} \frac{(2\pi)^{n/2}}{2\cdot 4\cdot \ldots \cdot n} & \text{dla } n \text{ parzystych,} \\ \frac{2(2\pi)^{(n-1)/2}}{1\cdot 3\cdot \ldots \cdot n} & \text{dla } n \text{ nieparzystych.} \end{cases}$

Rys. 3.4. Podstawowe typy jąder radialnych oraz odpowiadające im wartości wyrażenia $\frac{W(K)}{U(K)^2}$

rysunku 3.4[1]. Wprost z definicji widać, że na dowolnym okręgu o środku w początku układu współrzędnych wartość jądra radialnego jest stała. Własność tę określa się mianem **symetrii radialnej**. Jeśli przyjąć $,x_2 = ,x_3 = ... = ,x_n = 0$, czyli ograniczyć rozważania do jednowymiarowego przekroju wzdłuż osi x_1, to n-wymiarowe jądro radialne K jest z dokładnością do stałej tożsame z jądrem jednowymiarowym $\mathscr{K}$. Dzięki własności symetrii radialnej cecha ta jest prawdziwa także na przekroju wzdłuż dowolnej prostej przechodzącej przez początek układu współrzędnych – jądro K jest tu proporcjonalne do $\mathscr{K}$.

Z kolei **jądro produktowe** jest zdefiniowane następująco:

$$K(x) = K\left(\begin{bmatrix} ,x_1 \\ ,x_2 \\ \vdots \\ ,x_n \end{bmatrix}\right) = \mathscr{K}(,x_1) \cdot \mathscr{K}(,x_2) \cdot ... \cdot \mathscr{K}(,x_n). \tag{3.48}$$

Wartość n-wymiarowego jądra K jest więc iloczynem (produktem) jednowymiarowych jąder $\mathscr{K}$ dla poszczególnych współrzędnych. Jako występujące w podstawowym wzorze (3.3) wyrażenie związane z parametrem wygładzania należy teraz przyjąć

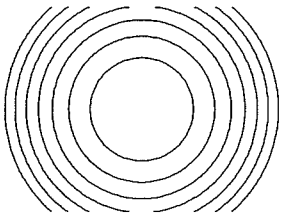
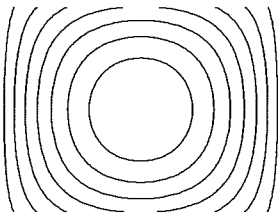
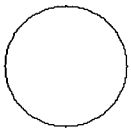
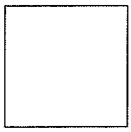
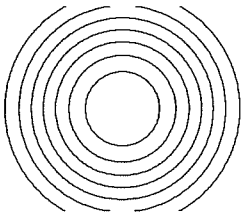
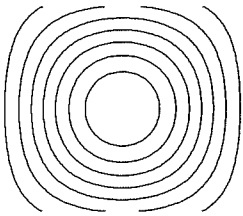
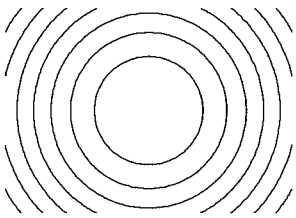
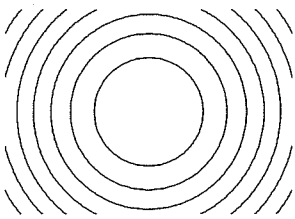
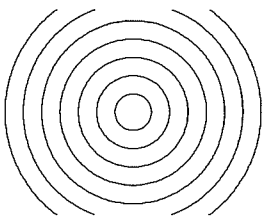
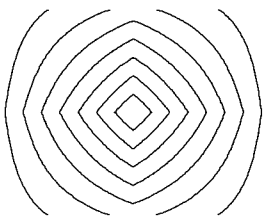
$$h^n = h_1 \cdot h_2 \cdot ... \cdot h_n, \tag{3.49}$$

czyli iloczyn parametrów wygładzania wyznaczonych dla poszczególnych współrzędnych wektora x. Jeśli dowolnie ustalić $n-1$ współrzędnych, to n-wymiarowe jądro produktowe K, rozważane jako funkcja pozostałej (nieustalonej) współrzędnej, jest z dokładnością do stałej tożsame z jądrem jednowymiarowym $\mathscr{K}$. Tak więc na przekroju równoległym do dowolnej osi jądro K jest proporcjonalne do $\mathscr{K}$.

Z teoretycznego punktu widzenia nie ma przeciwwskazań, aby w poszczególnych czynnikach jądra produktowego (3.48) stosować różne jądra jednowymiarowe. Wydaje się jednak, iż postępowanie takie ma nikłe uzasadnienie aplikacyjne, a w praktyce może zwiększyć złożoność procedury i w konsekwencji niebezpieczeństwo pomyłek obliczeniowych i błędów programowania.

Poziomice jądra radialnego (3.47) i produktowego (3.48) dla podstawowych typów jąder jednowymiarowych są dla porównania pokazane na rysunku 3.5. Warto zwrócić uwagę, że w przypadku jądra normalnego poziomice są iden-

[1] W przypadku nietypowej postaci jądra $\mathscr{K}$ może zaistnieć, że $\int_{\mathbb{R}^n} \mathscr{K}\left(\sqrt{x^T x}\right) dx = \infty$. Wówczas wyznaczenie stałej $c_{n,\mathscr{K}}$, a w konsekwencji utworzenie jądra radialnego, nie jest możliwe bez dodatkowych modyfikacji.

Typ jądra K	Radialne	Produktowe
Epanecznikowa		
Jednostajne		
Dwuwagowe		
Normalne		
Trójkątne		

Rys. 3.5. Poziomice podstawowych typów jąder radialnych i produktowych

tyczne, gdyż ze względu na własność $e^{x_*+x_{**}} = e^{x_*}e^{x_{**}}$ jądra radialne i produktowe, w rozważanej dotychczas podstawowej postaci, są tożsame.

Wśród jąder radialnych **najefektywniejsze** – w sensie przedstawionego w punkcie 3.1.2 kryterium scałkowanego błędu średniokwadratowego MISE – jest radialne jądro Epanecznikowa, czyli definiowane wzorem (3.47), gdy jednowymiarowe jądro $\mathscr{K}$ określa zależność (3.40). Także wśród jąder produktowych najefektywniejsze jest produktowe jądro Epanecznikowa dane za pomocą wzorów (3.48) i (3.40).

Z teoretycznego punktu widzenia dla dowolnego jądra jednowymiarowego bardziej efektywne jest jądro radialne niż produktowe. Co więcej, różnica efektywności zwiększa się w miarę wzrostu wymiaru zmiennej losowej. Na przykład dla optymalnego jądra Epanecznikowa wynosi ona 2, 5 i 8% odpowiednio dla wymiarów $n = 2, 3, 4$, a w przypadku mniej efektywnych jąder jednowymiarowych wielkości te są nieco większe. Jednakże z aplikacyjnego punktu widzenia tego rzędu różnice efektywności nie mają istotnego znaczenia. Fakt ten powoduje, iż w praktycznych zastosowaniach generalnie **można zalecać użycie jądra produktowego** jako znacznie dogodniejszego do dalszej analizy. W szczególności, procedury całkowania i różniczkowania jądra produktowego niewiele różnią się od przypadku jednowymiarowego. Dogodność aplikacyjna ma szczególne znaczenie dla tych zagadnień, w których estymator jądrowy stanowi podstawę do dalszych wszechstronnych i złożonych rozważań przedmiotowych, gdyż może nie tylko ułatwić analizę, lecz często wręcz umożliwić uzyskanie konkretnego, odpowiedniego do zastosowań wyniku.

3.1.4. Transformacja liniowa

W przypadku **wielowymiarowej zmiennej losowej, gdy jest używane jądro radialne**, można zastosować prostą i efektywną procedurę dopasowującą kształt jądra do postaci rozkładu tej zmiennej, opartą na koncepcji transformacji liniowej. Stanowi ona przedmiot rozważań niniejszego punktu. Tak więc rozpatrywana teraz będzie wielowymiarowa zmienna losowa, a ściślej n-wymiarowa zmienna X, przy czym $n \geq 2$.

Ostatecznym rezultatem zastosowania procedury **transformacji liniowej** jest uogólnienie definicji jądra radialnego (3.47) do postaci

$$K(x) = \frac{c_{n,\mathscr{K}}}{\sqrt{\det(R)}} \mathscr{K}\left(\sqrt{x^{\mathrm{T}} R^{-1} x}\right), \tag{3.50}$$

gdzie R oznacza macierz nieosobliwą o wymiarze $n \times n$, natomiast znaczenie i wartość stałej $c_{n,\mathscr{K}}$ pozostają niezmienione. Jeżeli R jest macierzą jednostkową, to powyższy wzór staje się równoważny podstawowej definicji (3.47). Załóżmy

również, że macierz ta jest symetryczna. Nie zmniejsza to ogólności, gdyż zawsze można zmienić równoważnie jej postać, przyjmując jako elementy spoza głównej przekątnej średnią arytmetyczną odpowiednich symetrycznych elementów macierzy „wyjściowej", czyli zdefiniować zmodyfikowaną macierz R o wyrazach $r^*_{i,j}$ danych następującym wzorem:

$$r^*_{i,j} = \begin{cases} \dfrac{r_{i,j} + r_{j,i}}{2}, & \text{gdy} \quad i \neq j \\ r_{i,j}, & \text{gdy} \quad i = j \end{cases} \qquad \text{dla } i = 1, 2, ..., n \text{ oraz } j = 1, 2, ..., n, \quad (3.51)$$

gdzie $r_{i,j}$ oraz $r_{j,i}$ oznaczają wspomniane wcześniej elementy macierzy „wyjściowej". Nietrudno zauważyć, że tak określona macierz jest symetryczna, a wartość występującego we wzorze (3.50) wyrażenia $x^T R^{-1} x$ pozostaje niezmieniona.

Postać diagonalną transformacji można otrzymać, przyjmując

$$R = \begin{bmatrix} V_1 & 0 & \cdots & 0 \\ 0 & V_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & 0 \\ 0 & 0 & 0 & V_n \end{bmatrix}, \quad (3.52)$$

przy czym V_1, V_2, ..., V_n oznaczają wariancje (2.25) poszczególnych współrzędnych badanej zmiennej losowej. Z kolei **postać pełną** definiuje się wzorem

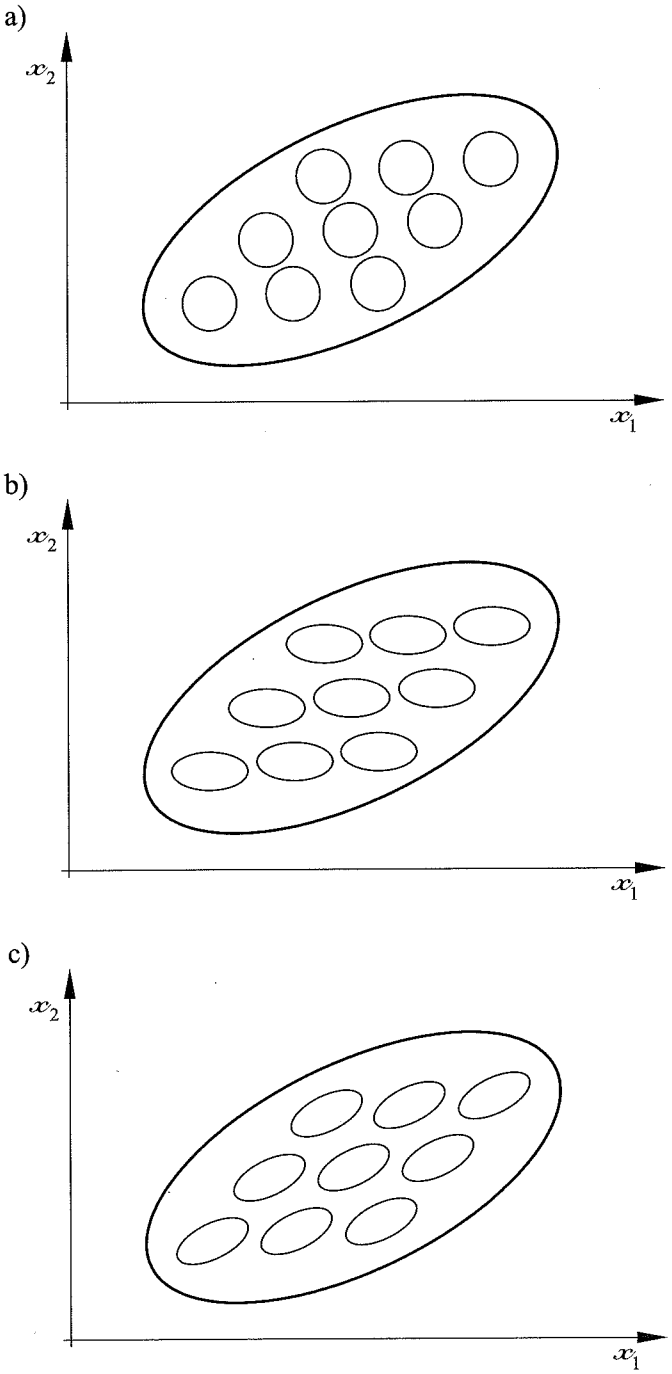
$$R = COV, \quad (3.53)$$

gdzie COV oznacza macierz kowariancji (2.30).

Bez zmniejszania ogólności rozważań można założyć wymaganą nieosobliwość macierzy (3.52)-(3.53). Osobliwość pierwszej z nich oznaczałaby bowiem, iż co najmniej jedna współrzędna badanej zmiennej losowej ma wariancję równą zeru, natomiast osobliwość drugiej mogłaby dodatkowo świadczyć, że co najmniej jedna współrzędna jest liniowo zależna od pozostałych. W obu przypadkach współrzędne takie eliminuje się we wstępnej fazie rozważań, gdyż niepotrzebnie zwiększają one wymiarowość zadania, nie niosąc żadnej informacji probabilistycznej.

Koncepcja transformacji liniowej została zilustrowana dla przypadku dwuwymiarowego na rysunku 3.6. Widoczna w częściach a, b i c, zaznaczona grubszą linią elipsa reprezentuje poziomicę funkcji gęstości prawdopodobieństwa przykładowej zmiennej losowej X o macierzy kowariancji

$$COV = \begin{bmatrix} 4 & 1{,}2 \\ 1{,}2 & 1 \end{bmatrix}. \quad (3.54)$$



Rys. 3.6. Ilustracja transformacji liniowej: a) brak transformacji, b) postać diagonalna (3.52), c) postać pełna (3.53)

W części a tego rysunku zostały także pokazane poziomice jąder dla podstawowej definicji (3.3). Są one okręgami i ich kształt w żaden sposób nie jest dostosowany do poziomicy charakteryzującej rozkład. W przypadku postaci diagonalnej (3.52), zilustrowanej w części b, okręgi te są zastępowane elipsami „rozciągniętymi" równolegle do osi, proporcjonalnie do odchyleń standardowych poszczególnych współrzędnych zmiennej X. I wreszcie, postać pełna (3.53) pozwala odpowiednio ukierunkować owe elipsy (rys. 3.6c).

Z analitycznego punktu widzenia procedura transformacji liniowej polega na przekształceniu przestrzeni wartości zmiennej losowej X według następującej zależności[1]:

$$y = \sqrt{R^{-1}}\, x. \tag{3.55}$$

Można zatem zdefiniować n-wymiarową zmienną losową Y zgodnie ze wzorem $Y \equiv \sqrt{R^{-1}}\, X$. Jeśli jest stosowana postać diagonalna (3.52), to wariancje wszystkich współrzędnych tej zmiennej są jednostkowe. Z kolei w przypadku postaci pełnej (3.53) macierz kowariancji jest jednostkowa, czyli wariancje wszystkich współrzędnych są równe 1, a ponadto współrzędne te są nieskorelowane. Dzięki takiej regularności poziomice poszczególnych symetrycznych radialnie jąder (3.3) są lepiej dopasowane do kształtu rozkładu zmiennej Y, zatem skonstruowany w ten sposób estymator jest efektywniejszy. Po dokonaniu transformacji odwrotnej, czyli podstawieniu wartości (3.55) do wzoru (3.47) zapisanego dla zmiennej y, a także podzieleniu stałej $c_{n,\mathcal{K}}$ przez $\sqrt{\det(R)}$ w celu utrzymania warunku (3.4), otrzymuje się ostatecznie uogólnioną definicję (3.50).

W praktyce, dysponując próbą losową (3.2), należy wyznaczyć estymatory poszczególnych elementów macierzy R. W przypadku postaci diagonalnej można do wzoru (3.50) podstawić bezpośrednio

$$R^{-1} = \begin{bmatrix} \frac{1}{\hat{V}_1} & 0 & \cdots & 0 \\ 0 & \frac{1}{\hat{V}_2} & \cdots & 0 \\ \vdots & \vdots & \ddots & 0 \\ 0 & 0 & 0 & \frac{1}{\hat{V}_n} \end{bmatrix}, \tag{3.56}$$

[1] **Pierwiastek** symetrycznej dodatnio półokreślonej **macierzy** A o wymiarze $n \times n$ definiuje się w sposób naturalny jako taką dodatnio półokreśloną macierz tego wymiaru, która pomnożona przez samą siebie jest równa A. Pierwiastek taki można otrzymać analitycznie, przekształcając macierz A do postaci kanonicznej Jordana, która w tym przypadku jest macierzą diagonalną, obliczając pierwiastki poszczególnych wyrazów na głównej przekątnej i stosując przekształcenie odwrotne. Można również w powyższym celu wykorzystać efektywne procedury numeryczne.

$$\det(R) = \hat{V}_1 \cdot \hat{V}_2 \cdot \ldots \cdot \hat{V}_n, \tag{3.57}$$

gdzie wartości $\hat{V}_1$, $\hat{V}_2$, ..., $\hat{V}_n$ są otrzymywane na podstawie zależności (2.50) albo (2.51) lub równoważnie (2.52). W przypadku postaci pełnej

$$R = [\hat{cov}_{j_1,j_2}], \tag{3.58}$$

przy czym poszczególne elementy $\hat{cov}_{j_1,j_2}$ dla $j_1, j_2 \in \{1, 2, ..., n\}$ oblicza się, odpowiednio, ze wzorów (2.50) i (2.54) albo (2.51) lub równoważnie (2.52) i (2.55) lub równoważnie (2.56). Macierz odwrotną oraz wartość wyznacznika można otrzymać na podstawie powszechnie znanych wzorów, gdy $n = 2, 3$, natomiast dla $n \geq 4$, korzystając z dogodnych procedur numerycznych dostępnych w literaturze przedmiotu.

Na koniec warto zauważyć, że użycie transformacji liniowej byłoby także możliwe wobec jądra produktowego (3.48). Stosowanie postaci diagonalnej (3.52) jest tu jednak zbyteczne, gdyż „skalowanie" poszczególnych współrzędnych zmiennej losowej X odbywa się poprzez indywidualny dobór parametru wygładzania dla kolejnych x_j. Z kolei użycie postaci pełnej (3.53) powoduje, iż zatraca się główną – z praktycznego punktu widzenia – zaletę jądra produktowego: rozdzielenie poszczególnych współrzędnych argumentu x między czynniki $\mathcal{K}(x_j)$.

Ostatecznie, jeżeli dla wielowymiarowej zmiennej losowej jest używane jądro radialne, to jego **definicję (3.47) uogólnia się do postaci (3.50)** w połączeniu z zależnościami (3.52) albo (3.53), co pozwala lepiej dopasować kształt jądra do rozkładu badanej zmiennej. W przypadku postaci diagonalnej do wyznaczenia macierzy R są stosowane wzory (3.56)-(3.57), natomiast w przypadku postaci pełnej – zależność (3.58) z użyciem znanych wzorów algebry liniowej i ewentualnie powszechnie dostępnych, dogodnych metod numerycznych. Prostota procedury przedstawionej w niniejszym punkcie, w połączeniu z naturalnością uzyskiwanych wyników, powoduje, że **w przypadku użycia jądra radialnego stosowanie transformacji liniowej – przynajmniej w postaci diagonalnej – należy zalecać jako wręcz konieczne**.

3.1.5. Wyznaczanie wartości parametru wygładzania

Wartość wprowadzonego we wzorze (3.3) parametru wygładzania h ma zasadnicze znaczenie dla jakości estymatora jądrowego. Istnieją jednakże użyteczne algorytmy umożliwiające wyznaczenie tej wartości na podstawie wartości próby losowej (3.2), przy wykorzystaniu kryterium scałkowanego błędu średniokwadratowego MISE. Algorytmy te będą przedstawione w niniejszym punkcie. Co więcej, zostały opracowane procedury pozwalające zindywidualizować oddziaływanie parametru h, który w podstawowej definicji (3.3) wpływa tak samo

na każde jądro. Służąca temu celowi tak zwana modyfikacja parametru wygładzania stanowi przedmiot rozważań kolejnego punktu – 3.1.6.

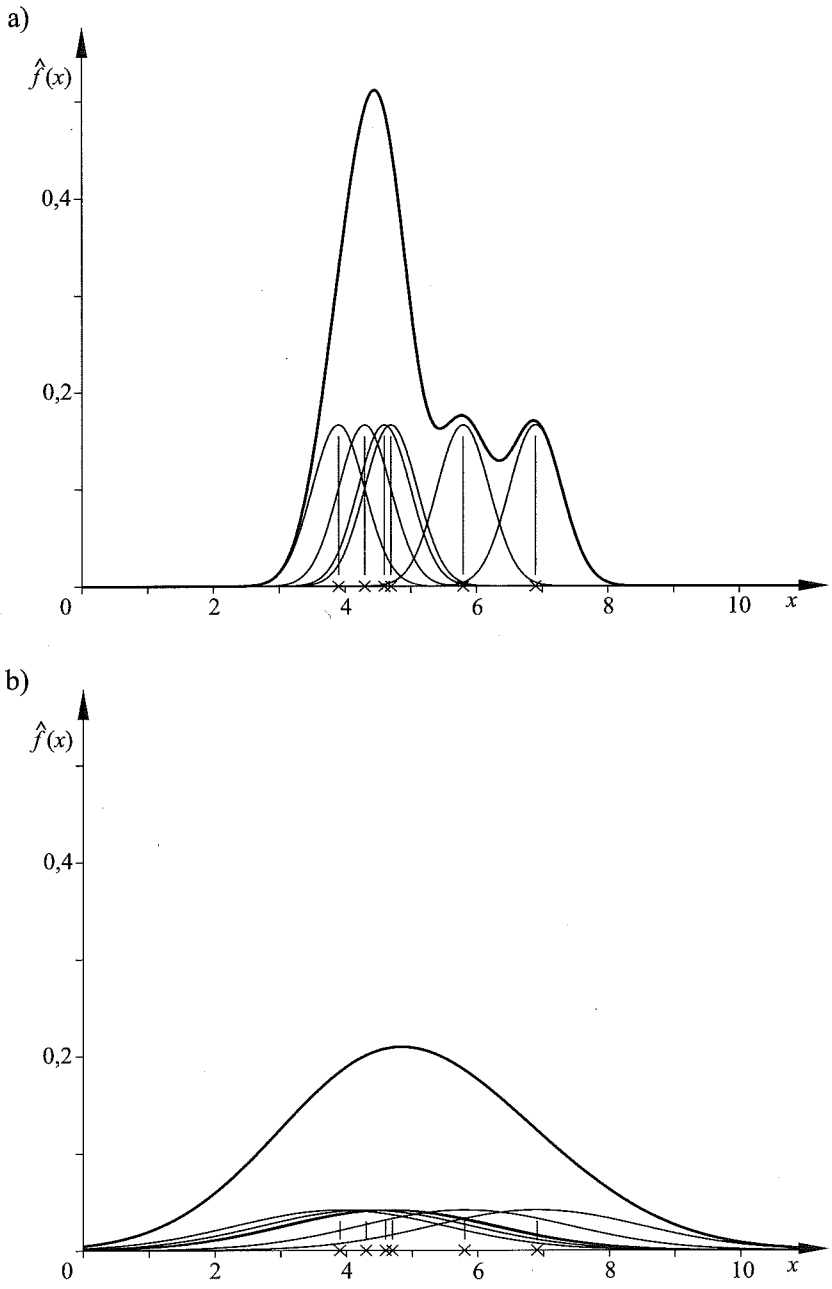
Warto przypomnieć na początku, że gdy zmienna losowa X jest wielowymiarowa i stosuje się jądro produktowe, wówczas parametr wygładzania jest ustalany odrębnie dla każdej współrzędnej, a w konsekwencji względem każdego czynnika $\mathcal{K}(x_j)$, z zastosowaniem procedur właściwych zmiennym jednowymiarowym.

Jak było już wspomniane, w przeciwieństwie do postaci jądra K wartość parametru wygładzania h ma zasadnicze znaczenie dla jakości estymatora jądrowego. Zbyt mała jego wartość powoduje pojawienie się znacznej ilości ekstremów lokalnych estymatora $\hat{f}$, co jest sprzeczne z faktycznymi własnościami realnych populacji. Z kolei zbyt duże wartości parametru h skutkują nadmiernym wygładzeniem tego estymatora, maskującym specyficzne cechy badanego rozkładu. Powyższe ilustruje rysunek 3.7, na którym została przyjęta nadmiernie pomniejszona ($h = 0{,}4$; rys. 3.7a) oraz zbyt powiększona ($h = 1{,}6$; rys. 3.7b) wartość parametru wygładzania. Warto porównać te wyniki z rysunkiem 1.5 – użytą tam wartość $h = 0{,}8$ można uznać za prawidłową.

W powyższym zagadnieniu uwidacznia się także wynikły z interpretacji wzoru (3.8) fakt, iż błąd średniokwadratowy jest sumą kwadratu obciążenia i wariancji estymatora. Zmniejszenie wartości parametru h powoduje bowiem zmniejszenie obciążenia i zwiększenie wariancji; przeciwne skutki są wynikiem zwiększenia wartości tego parametru. Wyznaczenie optymalnej wartości parametru wygładzania można zatem zinterpretować jako przyjęcie pewnego kompromisu między obciążeniem a wariancją estymatora jądrowego.

Stosowane w praktyce algorytmy wyznaczania wartości parametru wygładzania wywodzą się z przedstawionego w punkcie 3.1.2 kryterium scałkowanego błędu średniokwadratowego MISE. W przypadku estymatora jądrowego (3.3) wartość parametru wygładzania, minimalizująca wartość tego kryterium, jest dla jednowymiarowej zmiennej losowej określona zależnością (3.31), uogólnioną w przypadku wielowymiarowym wzorem (3.38). Bezpośrednie stosowanie tych formuł nie jest wszakże możliwe, gdyż występujące tam wyrażenie $Z(f)$ zależy od estymowanej – a więc nieznanej *a priori* – funkcji gęstości f. Wzory te stanowią jednak podstawę dla dogodnych metod suboptymalnych. Poniżej zostaną przedstawione trzy z nich: metoda przybliżona, podstawień i krzyżowego uwiarygodnienia. Wyczerpują one potrzeby aplikacyjne, przy różnych uwarunkowaniach poszczególnych zagadnień praktycznych.

Pierwsza z prezentowanych procedur to **metoda przybliżona**, w której optymalną wartość parametru wygładzania uzyskuje się bezpośrednio ze wzorów (3.31) i (3.38), podstawiając arbitralnie wartość funkcjonału Z otrzymaną dla standardowego rozkładu normalnego. W przypadku jednowymiarowym wartość ta dana jest zależnością



Rys. 3.7. Jądrowy estymator gęstości prawdopodobieństwa (3.3) jednowymiarowej zmiennej losowej dla: a) $h = 0{,}4$, b) $h = 1{,}6$ (por. rys. 1.5 uzyskany dla $h = 0{,}8$)

$$Z(f)=\frac{3}{8\sqrt{\pi}}, \tag{3.59}$$

co w przypadku n-wymiarowym można uogólnić do

$$Z(f)=\frac{n(n+2)}{2^{n+2}\pi^{n/2}}. \tag{3.60}$$

Dla konkretnego wymiaru n i postaci jądra K wyznaczenie wartości występującego we wzorach (3.31) i (3.38) ilorazu $W(K)/U(K)^2$ stanowi jedynie trudność obliczeniową, najczęściej sprowadzającą się do umiejętnego stosowania typowych algorytmów całkowania lub wręcz tablic całek. Na przykład dla jednowymiarowego jądra normalnego (3.44) prawdziwe jest

$$\frac{W(K)}{U(K)^2}=\frac{1}{2\sqrt{\pi}}, \tag{3.61}$$

a dla n-wymiarowego radialnego jądra normalnego:

$$\frac{W(K)}{U(K)^2}=\frac{1}{2^n\pi^{n/2}n^2}. \tag{3.62}$$

Co więcej, w praktycznych zastosowaniach wartość ilorazu $W(K)/U(K)^2$ wyznacza się tylko raz, podczas czynności wstępnych, czyli możliwe jest numeryczne obliczenie tej wartości z dowolną, wymaganą w praktyce dokładnością (a ponieważ i tak jest to metoda przybliżona, więc wymagania te nie są nadmiernie rygorystyczne). W przypadku stosowania transformacji liniowej wartość powyższego ilorazu nie ulega zmianie. Na rysunku 3.4 zostały podane wartości ilorazu $W(K)/U(K)^2$ dla jąder Epanecznikowa, jednostajnego, dwuwagowego, normalnego i trójkątnego.

Dysponując uzyskaną ze wzoru (3.59) lub (3.60) wartością funkcjonału $Z(f)$, a także wartością ilorazu $W(K)/U(K)^2$, można wyznaczyć – korzystając z zależności (3.31) lub (3.38) – wartość parametru wygładzania, która jest optymalna dla standardowego rozkładu normalnego. Aby dostosować ją do dowolnego – niekoniecznie standardowego – rozkładu normalnego, należy w przypadku jednowymiarowej zmiennej losowej (także dla poszczególnych czynników jądra produktowego) pomnożyć ją przez wartość estymatora odchylenia standardowego (2.48) lub odpowiednio (2.53). Z kolei w przypadku wielowymiarowym, gdy stosuje się jądro radialne, zadanie to spełnia zalecana w poprzednim punkcie transformacja liniowa.

Otrzymana za pomocą przedstawionej tu metody przybliżonej wartość parametru wygładzania jest tym użyteczniejsza, im bardziej estymowany rozkład jest zbliżony do normalnego lub ogólniej – trywialnego rozkładu unimodalne-

go. Najczęściej nie są to jednak sytuacje, w których warto stosować estymację nieparametryczną. W praktyce metodę tę należy proponować przede wszystkim jako proste rozwiązanie tymczasowe, w początkowej fazie prac nad programem komputerowym, podczas wstępnej weryfikacji jego poprawności. Jest ona również używana jako pierwsze oszacowanie na potrzeby bardziej złożonych procedur.

Obecnie są stosowane dwie **zaawansowane metody wyznaczania wartości parametru wygładzania**: podstawień oraz krzyżowego uwiarygodnienia. Pierwsza z nich jest zdecydowanie dogodniejsza obliczeniowo, ale praktycznie może być stosowana jedynie w przypadku jednowymiarowym. Z kolei bardziej złożony algorytm drugiej metody nie zależy od wymiaru zmiennej losowej i dlatego znajduje ona zastosowanie głównie do wielowymiarowych zmiennych losowych. Tak więc metody podstawień i krzyżowego uwiarygodnienia stanowią wzajemnie cenne uzupełnienie. W przypadku jednowymiarowym, gdy obie mogą być stosowane, dają wyniki zbliżone (wynikają bowiem z podobnych przesłanek minimalizacji wartości scałkowanego błędu średniokwadratowego MISE), aczkolwiek z niewielką przewagą pierwszej z nich.

Zaprezentowana teraz zostanie **metoda podstawień** (*plug-in*). Będzie zatem rozważana jednowymiarowa zmienna losowa, gdyż – jak zaznaczono powyżej – metoda ta jest przeznaczona do tego przypadku. Najpierw zostanie pokrótce przedstawiona sama koncepcja metody, a potem konkretne wzory umożliwiające jej bezpośrednie stosowanie.

Ponownie, metoda podstawień jest oparta na koncepcji zastosowania zależności (3.31), otrzymanej jako wynik zadania minimalizacji wartości scałkowanego błędu średniokwadratowego MISE. Jak było wspominane przy opisie metody przybliżonej, wyznaczenie dla ustalonego jądra K ilorazu $W(K)/U(K)^2$ nie jest zadaniem trudnym (por. także rys. 3.4 dla $n = 1$). Aby skompletować wielkości występujące we wzorze (3.31), należy także oszacować wartość funkcjonału $Z(f)$ definiowanego zależnością (3.19). Ponieważ w zależności tej jako funkcja podcałkowa występuje druga pochodna (nieznanej przecież) funkcji gęstości, więc w pierwszym kroku metody podstawień jest konstruowany estymator jądrowy owej drugiej pochodnej f''. Aby z kolei wyznaczyć analogicznie wartość parametru wygładzania dla tego estymatora, w drugim kroku konstruuje się estymator jądrowy czwartej pochodnej funkcji gęstości $f^{(4)}$. Po wykonaniu k takich kroków, prowadzących do wyznaczenia estymatora jądrowego funkcji $f^{(2k)}$, w celu ostatecznego zakończenia procedury oblicza się wartość jego parametru wygładzania za pomocą opisanej wcześniej metody przybliżonej. Należy zaznaczyć, że nic nie stoi na przeszkodzie, a wręcz jest wskazane, by do konstrukcji estymatorów jądrowych pochodnych $f'', f^{(4)}, ..., f^{(2k)}$ używać jądra innego niż jądro K stosowane w podstawowym estymatorze (3.3).

Praktyczna realizacja powyższej koncepcji odbywa się oczywiście w odwrotnej kolejności, a zatem najpierw, we wstępnym kroku, jest wyznaczana

metodą przybliżoną wartość parametru wygładzania estymatora jądrowego funkcji $f^{(2k)}$. Pozwala to wyznaczyć na podstawie zależności (3.31) wartość parametru wygładzania kolejno funkcji $f^{(2k-2)}$, $f^{(2k-4)}$ i iteracyjnie aż do f'', co ostatecznie umożliwia oszacowanie wartości funkcjonału $Z(f)$, a tym samym skompletowanie wielkości umożliwiających otrzymanie ze wzoru (3.31) wartości parametru wygładzania dla podstawowej postaci estymatora jądrowego (3.3).

Pozostaje jeszcze ustalić wartość parametru k, czyli liczbę kroków, po których stosuje się już metodę przybliżoną. Z jednej strony k powinno być jak największe, aby zmniejszyć wpływ niedokładności tej metody. Z drugiej strony duże k wymaga estymacji pochodnych wysokiego rzędu, co w każdej dziedzinie nauk stosowanych jest obarczone znacznym błędem, tu będącym dodatkowo skutkiem – stanowiącej kwintesencję rozważań – probabilistycznej nieokreśloności. Konieczny jest zatem wybór kompromisowy. Praktyka wskazuje jednoznacznie na ograniczenie $k \geq 2$, z preferencją dla $k = 2$. Po ustaleniu wartości parametru k nazwę powyższej metody można uściślić jako **metodę podstawień k-tego rzędu**.

Zostaną teraz podane konkretne **wzory umożliwiające bezpośrednie zastosowanie metody podstawień** drugiego rzędu i jednoznaczne wskazówki umożliwiające ewentualne jego zwiększenie.

Niech zatem $k \in \mathbb{N} \setminus \{0\}$ oznacza założony rząd metody i niech będzie dane jądro $\widetilde{K}$ różniczkowalne $(2k+2)$-krotnie, takie że $\widetilde{K}^{(4)}(0)$, $-\widetilde{K}^{(6)}(0)$, $\widetilde{K}^{(8)}(0)$, $-\widetilde{K}^{(10)}(0)$, ..., $(-1)^{k+1}\widetilde{K}^{(2k+2)}(0)$ są dodatnie. W praktyce powszechnie przyjmowane jest tutaj jądro normalne (3.44). Zakłada się dodatkowo, że warunki (3.13)-(3.15) są spełnione nie tylko dla funkcji f, ale także dla jej pochodnych f'', $f^{(4)}$, ..., $f^{(2k+2)}$. Niech ponadto dla pomocniczego parametru $\xi \in \{4, 6, ..., 2k+4\}$ stała $c_\xi \in \mathbb{R}$ przyjmuje wartość

$$c_\xi = \frac{\xi!}{(0{,}5\xi)!\ \sqrt{\pi}\ (2\hat{\sigma})^{\xi+1}} \tag{3.63}$$

oraz dodatkowo dla $h > 0$ niech będzie również dana stała $C_{\xi,h} \in \mathbb{R}$ określona wzorem

$$C_{\xi,h} = \frac{1}{m^2 h^{\xi+1}} \sum_{i=1}^{m} \sum_{j=1}^{m} \widetilde{K}^{(\xi)}\left(\frac{x_i - x_j}{h}\right), \tag{3.64}$$

gdzie $\hat{\sigma}$ oznacza estymator odchylenia standardowego (2.48), a x_i i x_j reprezentują, odpowiednio, i-tą oraz j-tą wartość próby losowej (3.2).

Najpierw będą przedstawione zależności dotyczące stosowania metody podstawień drugiego rzędu. I tak we wstępnym kroku, opartym na metodzie przybli-

żonej, wyznacza się ze wzoru (3.63) wartość parametru c_8. Po podstawieniu $\xi = 8$ wzór ten można uprościć do postaci

$$c_8 = \frac{105}{32\sqrt{\pi}\,\hat{\sigma}^9}. \tag{3.65}$$

Na tej podstawie wykonuje się dwa kroki, zgodnie z założonym rzędem metody, stosując przy tym ich numerację odwrotną do naturalnej. Otóż najpierw, w drugim kroku, zostaje obliczona wartość

$$h_{\mathrm{II}} = \left(\frac{-2\widetilde{K}^{(6)}(0)}{U(\widetilde{K})\, c_8\, m} \right)^{1/9}, \tag{3.66}$$

a potem, w pierwszym kroku, przy wykorzystaniu zależności (3.64), wartość

$$h_{\mathrm{I}} = \left(\frac{-2\widetilde{K}^{(4)}(0)}{U(\widetilde{K})\, C_{6,h_{\mathrm{II}}}\, m} \right)^{1/7}. \tag{3.67}$$

Ostatecznie, otrzymuje się, ponownie stosując zależność (3.64):

$$Z(f) = C_{4,h_{\mathrm{I}}}. \tag{3.68}$$

Wartość ta razem z wartością ilorazu $W(K)/U(K)^2$ (por. także rys. 3.4 dla $n = 1$) tworzą komplet wielkości występujących we wzorze (3.31), umożliwiających oszacowanie optymalnej – w sensie kryterium scałkowanego błędu średniokwadratowego MISE – wartości parametru wygładzania estymatora jądrowego (3.3).

Biorąc pod uwagę wyrazistość koncepcji metody podstawień, uzupełnienie powyższego algorytmu do procedury trzeciego rzędu jest nieskomplikowane. I tak, we wstępnym kroku wyznacza się ze wzoru (3.63) wartość parametru c_{10}:

$$c_{10} = \frac{945}{64\sqrt{\pi}\,\hat{\sigma}^{11}}, \tag{3.69}$$

a stąd w dodatkowo wprowadzonym tu trzecim kroku

$$h_{\mathrm{III}} = \left(\frac{-2\widetilde{K}^{(8)}(0)}{U(\widetilde{K})\, c_{10}\, m} \right)^{1/11} \tag{3.70}$$

oraz w nieco zmienionym drugim kroku

$$h_{\mathrm{II}} = \left(\frac{-2\widetilde{K}^{(6)}(0)}{U(\widetilde{K})\, C_{8,h_{\mathrm{III}}}\, m} \right)^{1/9}. \tag{3.71}$$

Reszta algorytmu, zdefiniowana wzorami (3.67)-(3.68), pozostaje niezmieniona.

Podobnie realizuje się procedurę w przypadku rzędu wyższego niż trzy, chociaż z praktycznego punktu widzenia można wątpić w celowość takiego postępowania.

W powyższym algorytmie jako $\widetilde{K}$ jest powszechnie stosowane jądro normalne (3.44). Wtedy wielkości występujące we wzorach (3.66)-(3.71) wynoszą odpowiednio:

$$U(\widetilde{K}) = 1, \tag{3.72}$$

$$\widetilde{K}^{(4)}(x) = \frac{1}{\sqrt{2\pi}}(x^4 - 6x^2 + 3)\exp\left(-\frac{1}{2}x^2\right), \tag{3.73}$$

$$\widetilde{K}^{(4)}(0) = \frac{3}{\sqrt{2\pi}}, \tag{3.74}$$

$$\widetilde{K}^{(6)}(x) = \frac{1}{\sqrt{2\pi}}(x^6 - 15x^4 + 45x^2 - 15)\exp\left(-\frac{1}{2}x^2\right), \tag{3.75}$$

$$\widetilde{K}^{(6)}(0) = -\frac{15}{\sqrt{2\pi}}, \tag{3.76}$$

$$\widetilde{K}^{(8)}(x) = \frac{1}{\sqrt{2\pi}}(x^8 - 28x^6 + 210x^4 - 420x^2 + 105)\exp\left(-\frac{1}{2}x^2\right), \tag{3.77}$$

$$\widetilde{K}^{(8)}(0) = \frac{105}{\sqrt{2\pi}}. \tag{3.78}$$

Podsumowując, metoda podstawień zapewnia dogodny i szybki algorytm wyznaczania wartości parametru wygładzania estymatora jądrowego (3.3) na podstawie wartości próby losowej (3.2). Może być ona stosowana wobec jednowymiarowej zmiennej losowej, a zatem także w przypadku zmiennej wielowymiarowej, gdy używane jest jądro produktowe – wtedy algorytm obliczeniowy stosuje się n-krotnie, odrębnie dla poszczególnych współrzędnych.

Trzecią prezentowaną tu procedurą wyznaczania wartości parametru wygładzania jest **metoda krzyżowego uwiarygodnienia** (*cross-validation*). Jej podstawową zaletą jest to, że zasadniczo nie zależy ona od wymiaru rozważanej zmiennej losowej. Ponieważ jednak w przypadku jednowymiarowym korzystniejsza bywa metoda podstawień, więc metody krzyżowego uwiarygodnienia używa się w przypadku wielowymiarowym przy stosowaniu jądra radialnego. Jak poprzednio, prezentowana koncepcja jest oparta na zadaniu minimalizacji wartości scałkowanego błędu średniokwadratowego MISE.

Niech zatem dla ustalonej postaci jądra K dana będzie funkcja $g:(0,\infty)\to\mathbb{R}$ zdefiniowana zależnością

$$g(h)=\frac{1}{h^n}\left[\frac{1}{m^2}\sum_{i=1}^{m}\sum_{j=1}^{m}G\left(\frac{x_j-x_i}{h}\right)+\frac{2}{m}K(0)\right], \tag{3.79}$$

przy czym x_i i x_j reprezentują, odpowiednio, i-tą oraz j-tą wartość próby losowej (3.2), a funkcja $G:\mathbb{R}^n\to\mathbb{R}$ jest określona wzorem

$$G(x)=K^{*2}(x)-2K(x), \tag{3.80}$$

gdzie $K^{*2}:\mathbb{R}^n\to[0,\infty)$ oznacza kwadrat splotowy funkcji K, czyli

$$K^{*2}(x)=\int_{\mathbb{R}^n}K(y)\,K(x-y)\,dy. \tag{3.81}$$

Kwadrat splotowy radialnego jądra normalnego jest dany wzorem

$$K^{*2}(x)=(4\pi)^{-n/2}\exp\left(-\frac{1}{4}x^{\mathrm{T}}x\right), \tag{3.82}$$

natomiast dla innych typów jąder do wyznaczenia jego wartości można wykorzystać metody numeryczne oparte na teorii aproksymacji. W tym celu warto zauważyć, iż ze względu na własność symetrii radialnej wystarczy, jeśli dysponuje się funkcją $\widetilde{K}^{*2}:[0,\infty)\to[0,\infty)$, taką że

$$K^{*2}(x)=\widetilde{K}^{*2}\left(\|x\|\right). \tag{3.83}$$

W praktyce odpowiednie przybliżenie powyższej funkcji można otrzymać, zakładając następującą jej postać:

$$\widetilde{K}^{*2}(u)=W(K)\exp(-\beta u^{\alpha})\quad\text{dla}\quad u\in[0,\infty), \tag{3.84}$$

gdy stosowane jednowymiarowe jądro $\mathscr{K}$ ma nieograniczony nośnik, a także

$$\widetilde{K}^{*2}(u)=\begin{cases}\dfrac{W(K)}{1-\exp(-\beta(2\gamma)^{\alpha})}\left[\exp(-\beta u^{\alpha})-\exp(-\beta(2\gamma)^{\alpha})\right] & \text{dla } u\in[0,2\gamma],\\ 0 & \text{dla } u\in(2\gamma,\infty),\end{cases} \tag{3.85}$$

gdy jego nośnik jest ograniczony dodatnią liczbą γ (dla jąder Epanecznikowa, jednostajnego, dwuwagowego i trójkątnego $\gamma=1$; por. rys. 3.4). Parametry α oraz β są nieujemne, a ich wartości można wyznaczyć tak, aby zminimalizować wyrażenie

$$\sum_{j=0}^{J}\left[\widetilde{K}^{*2}(u_j)-K_{\mathrm{NC}}^{*2}\left(\begin{bmatrix}u_j\\0\\\vdots\\0\end{bmatrix}\right)\right]^2, \tag{3.86}$$

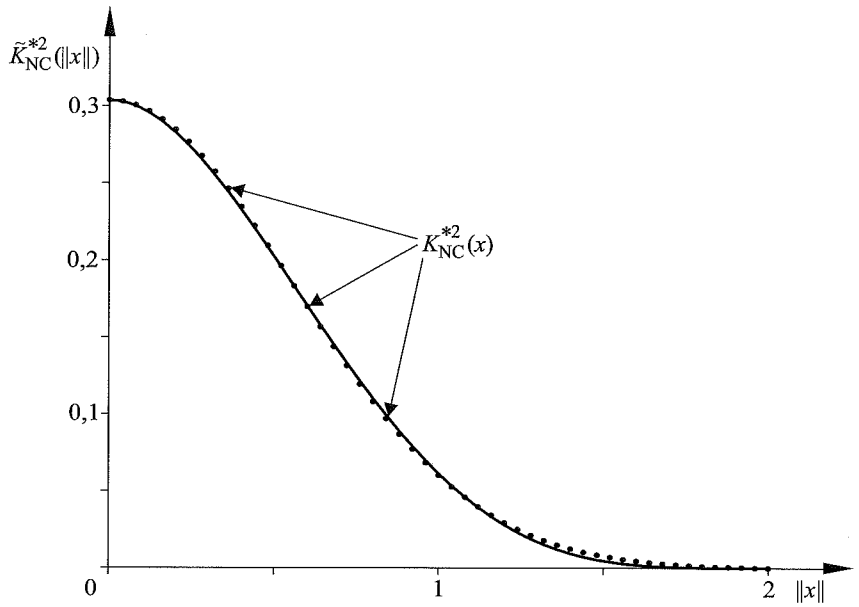
przy czym K_{NC}^{*2} oznacza wartość kwadratu splotowego obliczonego za pomocą metod numerycznego całkowania, a wielkości u_0, u_1, ..., u_J powinny być rozmieszczone równomiernie w przedziale $[0, 4\sqrt{U(\mathscr{K})}]$, gdy stosuje się jądro o nośniku nieograniczonym, lub też w przedziale $[0, 2\gamma]$ dla jąder o nośniku ograniczonym. Wstępnie można przyjąć $J = 50$. Warto również przypomnieć, iż wielkości $U(\mathscr{K})$ oraz $W(K)$ zostały określone zależnościami (3.17) i (3.34), a wzory na ich wartości dla jąder Epanecznikowa, jednostajnego, dwuwagowego i trójkątnego mogą być odczytane odpowiednio z rysunków 3.3 i 3.4.

Wobec sformułowanego powyżej zagadnienia minimalizacji można użyć typowych numerycznych metod optymalizacyjnych lub trywialnie obliczyć najmniejszą wartość wyrażenia (3.86) dla parametrów α i β z założonego zakresu, na przykład dla obu z nich $[0, 3]$ z „krokiem" 0,01, po czym ewentualnie ze zmniejszonym „krokiem", uprzednio stosownie redukując zakresy. Należy zaznaczyć, że po ustaleniu typu jądra powyższe czynności są wykonywane jedynie raz w trakcie całego procesu konstruowania estymatora jądrowego, stąd ich ewentualna czasochłonność obliczeniowa nie stanowi istotnego utrudnienia w praktycznym użyciu powyższej procedury.

Przykładowy wynik dla czterowymiarowego ($n = 4$) jądra Epanecznikowa został pokazany na rysunku 3.8. Warto także zauważyć, że ze względu na równość (3.82) w przypadku jądra normalnego wzór przybliżony (3.84) staje się zależnością dokładną, jeśli przyjąć $\alpha = 2$ oraz $\beta = 1/4$.

Stosując przedstawioną w punkcie 3.1.4 transformację liniową, we wzorach (3.79)-(3.80) powinna być stosowana postać jądra K zgodna z definicją (3.50). W tym przypadku kwadrat splotowy jest zmodyfikowany do postaci $\frac{1}{\sqrt{\det(R)}}K^{*2}(\sqrt{R^{-1}}x)$, gdzie K^{*2} oznacza rozważaną powyżej podstawową postać kwadratu splotowego bez uwzględnienia transformacji liniowej. W szczególności, w przypadku radialnego jądra normalnego:

$$K^{*2}(x)=\frac{1}{\sqrt{\det(R)}}(4\pi)^{-n/2}\exp\left(-\frac{1}{4}x^{\mathrm{T}}R^{-1}x\right). \tag{3.87}$$



Rys. 3.8. Aproksymacja wartości kwadratu splotowego dla czterowymiarowego jądra Epanecznikowa za pomocą wzoru (3.85); $\alpha = 1{,}998$, $\beta = 1{,}613$

Zgodnie z formułą metody krzyżowego uwiarygodnienia za wartość parametru wygładzania przyjmuje się argument h minimalizujący funkcję g daną wzorem (3.79). Warto zwrócić uwagę, że niezależnie od wymiaru n funkcja g jest rzeczywista, czyli o argumencie i wartościach będących liczbami rzeczywistymi. Można zatem stosować różnorodny, opisany w literaturze aparat metod numerycznych przeznaczonych do takiego zadania. W praktycznych przypadkach funkcja g ma przeważnie kształt zbliżony do paraboli o niesymetrycznych „ramionach". W ogólnym przypadku mogą jednak sporadycznie wystąpić lokalne minima, aczkolwiek niezbyt „głębokie", a także „załamanie" w pobliżu zera, skutkujące zależnością $\lim_{h\to 0^+} g(h) = -\infty$. Uwzględniając powyższe uwarunkowania, w celu znalezienia minimum funkcji g można zalecić użycie powszechnie znanej **metody złotego podziału**. We wstępnym kroku określa się tu przedział $[a_0, b_0]$, w którym będą prowadzone poszukiwania, a także wyznacza liczby rzeczywiste

$$p_0 = b_0 - \frac{\sqrt{5}-1}{2}(b_0 - a_0), \tag{3.88}$$

$$q_0 = a_0 + \frac{\sqrt{5}-1}{2}(b_0 - a_0) \tag{3.89}$$

oraz wartości funkcji g w tych punktach:

$$g(p_0), \tag{3.90}$$

$$g(q_0). \tag{3.91}$$

W każdym kolejnym k-tym kroku, dla $k = 1, 2, ..., \mathcal{k}$, przy czym $\mathcal{k} \in \mathbb{N} \setminus \{0\}$ oznacza założoną liczbę kroków, przyjmuje się, że

- jeżeli $g(p_{k-1}) \leq g(q_{k-1})$, to

$$a_k = a_{k-1}, \tag{3.92}$$

$$b_k = q_{k-1}, \tag{3.93}$$

$$p_k = b_k - \frac{\sqrt{5}-1}{2}(b_k - a_k), \tag{3.94}$$

$$q_k = p_{k-1} \tag{3.95}$$

i ponadto wyznacza się wartość funkcji g w punkcie p_k, a zatem

$$g(p_k), \tag{3.96}$$

oraz podstawia obliczoną w poprzednim kroku wartość $g(p_{k-1})$ zamiast $g(q_k)$, czyli

$$g(q_k) = g(p_{k-1}); \tag{3.97}$$

- jeżeli $g(p_{k-1}) > g(q_{k-1})$, to

$$a_k = p_{k-1}, \tag{3.98}$$

$$b_k = b_{k-1}, \tag{3.99}$$

$$p_k = q_{k-1}, \tag{3.100}$$

$$q_k = a_k + \frac{\sqrt{5}-1}{2}(b_k - a_k) \tag{3.101}$$

i ponadto podstawia się obliczoną w poprzednim kroku wartość $g(q_{k-1})$ zamiast $g(p_k)$, czyli

$$g(p_k) = g(q_{k-1}), \tag{3.102}$$

oraz wyznacza wartość funkcji g w punkcie q_k, a zatem

$$g(q_k). \tag{3.103}$$

I wreszcie w ostatnim, k-tym, kroku otrzymuje się poszukiwaną wartość parametru h:

- jeżeli $g(p_k) \leq g(q_k)$, to należy przyjąć

$$h = \frac{a_k + q_k}{2}; \tag{3.104}$$

- jeżeli $g(p_k) > g(q_k)$, to

$$h = \frac{p_k + b_k}{2}. \tag{3.105}$$

Tak więc koncepcja metody złotego podziału polega na wygenerowaniu zstępującego ciągu przedziałów $[a_0, b_0] \supset [a_1, b_1] \supset ... \supset [a_k, b_k]$, przy czym

$$\frac{b_k - a_k}{b_{k-1} - a_{k-1}} = \frac{\sqrt{5}-1}{2} \cong 0{,}618 \quad \text{dla} \quad k = 1, 2, ..., k, \tag{3.106}$$

czyli długość każdego z kolejnych przedziałów stanowi około 62% długości poprzednika. Jeżeli funkcja g ma w przedziale $[a_0, b_0]$ tylko jedno minimum, to jest ono zawarte w każdym z nich, w szczególności w ostatnim $[a_k, b_k]$. Po przyjęciu wartości h zgodnie ze wzorami (3.104)-(3.105) minimum to jest zatem wyznaczane z dokładnością

$$\Delta = \frac{1}{2}(b_k - a_k) = \frac{1}{2}\left(\frac{\sqrt{5}-1}{2}\right)^{k+1}(b_0 - a_0). \tag{3.107}$$

Decyzja o kolejnych redukcjach przedziałów jest podejmowana na podstawie wartości funkcji g, wyznaczanych w 38% i 62%, czyli w przybliżeniu 1/3 i 2/3 długości tych przedziałów. Tak duże zróżnicowanie pozwala w praktyce wyeliminować skutki występowania ewentualnych „płytkich" ekstremów lokalnych funkcji g. Co więcej, jeżeli przyjąć

$$[a_0, b_0] = \left[\frac{h_0}{4}, 4h_0\right], \tag{3.108}$$

gdzie h_0 oznacza wartość parametru wygładzania wyznaczoną metodą przybliżoną, to można także uniknąć skutków ewentualnego „załamania" tej funkcji w sąsiedztwie zera.

Ostatecznie, jeżeli przyjmie się zależność (3.108) i oznaczy wymaganą dokładność wyznaczenia minimum przez Δ, przy czym $\Delta > 0$, to ze wzoru (3.107) można obliczyć gwarantującą tę dokładność minimalną liczbę iteracji:

$$k = \overline{\mathrm{Int}}\left(\frac{\ln\left[\frac{8\Delta}{15\, h_0} \right]}{\ln\left[\frac{\sqrt{5}-1}{2} \right]} \right) - 1, \tag{3.109}$$

gdzie $\overline{\mathrm{Int}}(c)$ oznacza najmniejszą liczbę całkowitą większą lub równą $c \in \mathbb{R}$. W szczególności, dla wystarczającego w praktyce

$$\Delta = h_0/100 \tag{3.110}$$

otrzymuje się

$$k = 10. \tag{3.111}$$

Ponieważ czasochłonne obliczenia wartości funkcji g zdefiniowanej zależnością (3.79) są prowadzone we wstępnym kroku dwukrotnie (wzory (3.90)-(3.91)) i w każdym następnym jednokrotnie (wzór (3.96) albo (3.103)), zatem w przypadku przyjęcia zależności (3.108) i (3.110) obliczenia te będą wykonywane 12-krotnie.

We wstępnej fazie projektowania, a także w przypadku pojawienia się wątpliwości dotyczących poprawności otrzymywanego na podstawie powyższej procedury wyniku, warto porównać go z rezultatem sukcesywnego obliczania wartości funkcji (3.79) w przedziale $[h_0/100, 4h_0]$ z krokiem $h_0/100$ oraz prostego wyznaczenia na ich podstawie odpowiedniego minimum, co zabezpiecza przed potencjalnymi ekstremami lokalnymi i ewentualnym „załamaniem" funkcji g w pobliżu zera. Postępowanie takie oznacza jednak ponad 30-krotne zwiększenie czasochłonności obliczeń.

Podsumowując: w niniejszym punkcie zostały przedstawione trzy kompletne, dogodne do bezpośrednich zastosowań procedury wyznaczania wartości parametru wygładzania estymatora jądrowego (3.3). Ich uwarunkowania wzajemnie się uzupełniają. Metoda przybliżona, bardzo prosta, przeznaczona do stosowania w przypadku jedno- i wielowymiarowym, aczkolwiek niegwarantująca dokładności wymaganej w wielu zastosowaniach, może być wykorzystywana jako szybkie rozwiązanie tymczasowe we wstępnej fazie konstrukcji estymatora. Z kolei metoda podstawień jest dokładna, stosunkowo szybka, ale bywa stosowana jedynie w przypadku jednowymiarowym lub wielowymiarowym, gdy używa się jądra produktowego. Nieco mniej korzystne własności ma metoda krzyżowego uwiarygodnienia, może jednakże być przydatna w dowolnym przypadku, także w wielowymiarowym, gdy jest używane jądro radialne.

Na podstawie powyższego podsumowania można również sformułować kolejną przesłankę dotyczącą wyboru w przypadku wielowymiarowym jądra produktowego lub radialnego. Otóż, możliwość korzystania, w przypadku stosowa-

nia tego pierwszego, z prostej i jednocześnie efektywnej metody podstawień, zamiast konieczności użycia dla jądra radialnego złożonej metody krzyżowego uwiarygodnienia, jest następnym argumentem przemawiającym za preferowaniem jądra produktowego w działaniach praktycznych.

W wyniku stosowania każdej z przedstawionych tu metod otrzymuje się wartości parametru wygładzania spełniające warunek (3.23), a także (3.27) lub w przypadku wielowymiarowym (3.37).

Prezentowane metody wyznaczania wartości parametru wygładzania są wygodne w praktycznym użyciu przede wszystkim dlatego, że ich algorytmy są kompletne i nie wymagają od projektanta żmudnych badań przedmiotowych. Jednak procedury te zostały arbitralnie oparte na zagadnieniu minimalizacji błędu średniokwadratowego, ze względu na wszechstronne korzyści wynikające z własności przestrzeni Hilberta L^2. W praktycznych zadaniach często brak jest jednak przesłanek, aby twierdzić, że właśnie to kryterium (podobnie zresztą jak każde inne) jest odpowiednie z punktu widzenia wymagań konkretnego zastosowania. Z tego powodu, przy konstruowaniu estymatora jądrowego w danym problemie aplikacyjnym, można polecać dokonanie przez projektanta końcowej, wizualnej oceny tego estymatora, a w konsekwencji **ewentualnej korekty wartości parametru wygładzania opartej na przesłankach intuicyjnych**. Praktyka wskazuje, iż wstępnie warto zaproponować nieznaczne, kilku- lub kilkunastoprocentowe, zmniejszenie wartości parametru wygładzania, zwłaszcza w przypadku rozkładów wielomodalnych, niesymetrycznych i przy relatywnie małych licznościach próby losowej. Podstawą analizy powinna być interpretacja porównawcza efektów widocznych na rysunkach 1.5 i 3.7. W przypadku dwuwymiarowej zmiennej losowej można posłużyć się programem komputerowym do wizualizacji wartości funkcji dwóch zmiennych, a w przypadku wymiarów większych niż dwa dokonywać obserwacji na przekrojach, to znaczy przy reprezentatywnie ustalonych wartościach pozostałych współrzędnych. Co więcej, taka analiza daje możliwość wykrycia ewentualnych błędów programowania oraz merytorycznych. W najstarszych zastosowaniach estymatorów jądrowych, z powodu ówczesnego braku sprawdzonych metod analitycznych, heurystyczna procedura wizualna była wręcz zalecanym, podstawowym narzędziem aplikacyjnym. Obecnie końcową ewentualną korektę wartości parametru wygładzania na podstawie intuicyjnej analizy czynionej przez projektanta należy traktować jako cenne uzupełnienie bezdusznych metod analitycznych.

3.1.6. Modyfikacja parametru wygładzania

Wartość parametru wygładzania h, wprowadzonego w definicji (3.3), jest taka sama dla każdego jądra K. Wpływ tej wartości na jego kształt był rozważany w poprzednim punkcie i został zilustrowany na rysunku 3.7 – mała wartość po-

woduje „wyszczuplenie", natomiast duża „rozciągnięcie" wykresu funkcji K. Jeżeli zatem możliwe jest zindywidualizowanie wpływu parametru wygładzania na poszczególne jądra, to w rejonach „zagęszczenia" elementów próby losowej wartość ta powinna być zmniejszona (co pozwoli lepiej ukazać specyficzne cechy rozkładu), w przeciwieństwie do obszarów, w których elementy te są „rzadkie", gdzie powinna być zwiększona (co spowoduje dodatkowe wygładzenie „ogonów"). Służąca temu celowi tak zwana modyfikacja parametru wygładzania stanowi przedmiot rozważań niniejszego punktu.

Konstrukcję estymatora jądrowego z **modyfikacją parametru wygładzania** można zrealizować w następujący sposób:

- wyznaczany jest estymator jądrowy $\hat{f}$ zgodnie z przedstawioną wcześniej metodyką;
- określane są **parametry modyfikujące** $s_i > 0$ $(i = 1, 2, ..., m)$ postaci

$$s_i = \left(\frac{\hat{f}(x_i)}{\tilde{s}} \right)^{-c}, \tag{3.112}$$

gdzie $c \geq 0$, a $\tilde{s}$ jest średnią geometryczną liczb $\hat{f}(x_1)$, $\hat{f}(x_2)$, ..., $\hat{f}(x_m)$ daną jako

$$\tilde{s} = \exp\left(\frac{1}{m} \sum_{i=1}^{m} \ln(\hat{f}(x_i)) \right); \tag{3.113}$$

- ostatecznie definiuje się estymator jądrowy ze zmodyfikowanym parametrem wygładzania za pomocą wzoru

$$\hat{f}(x) = \frac{1}{mh^n} \sum_{i=1}^{m} \frac{1}{s_i^n} K\left(\frac{x - x_i}{hs_i} \right). \tag{3.114}$$

Powyższa procedura jest taka sama dla przypadku jedno- i wielowymiarowego (w pierwszym przypadku należy przyjąć we wzorze (3.114) $n = 1$). Definicja (3.114) stanowi uogólnienie podstawowej postaci (3.3). Jeżeli bowiem $c = 0$, to $s_i \equiv 1$ i zależności te są równoważne.

Koncepcja modyfikacji polega na pomnożeniu wartości parametru h, jednolitej dla każdego jądra, przez parametr modyfikujący s_i wyznaczony indywidualnie dla każdego z nich. Wartość tego parametru ustala się na podstawie wzoru (3.112), gdzie jest porównywana wartość estymatora jądrowego bez modyfikacji, obliczona w punkcie i-tego elementu próby losowej, ze średnią wartością tego estymatora dla wszystkich elementów próby losowej. Jeżeli zatem wartość

estymatora jądrowego na i-tym elemencie próby losowej jest równa średniej, to ze wzoru (3.112) otrzymuje się $s_i = 1$ i na podstawie zależności (3.114) parametr wygładzania h nie jest modyfikowany. Z kolei, gdy wartość estymatora na i-tym elemencie próby losowej jest mniejsza od średniej, to $s_i > 1$ i parametr wygładzania jest zwiększany, czyli i-te jądro ulega „rozciągnięciu". I wreszcie, jeśli wartość estymatora na i-tym elemencie próby losowej jest większa od średniej, to $s_i < 1$, a zatem parametr wygładzania jest zmniejszany, czyli i-te jądro ulega „wyszczupleniu". W rezultacie – zgodnie z założeniami – obszary, w których estymator przyjmuje małą wartość, zostają dodatkowo „wygładzone", w przeciwieństwie do fragmentów jego dużych wartości, gdzie polepsza się charakteryzacja specyficznych cech rozkładu (rys. 3.9).

Parametr c decyduje o intensywności modyfikacji. Wartość $c = 0$ implikuje jej brak, natomiast wraz ze wzrostem wartości tego parametru intensywność zwiększa się. W praktyce najczęściej stosuje się $c \in [0, 1]$. Rozważania wynikające z kryterium średniokwadratowego wskazują na wartość

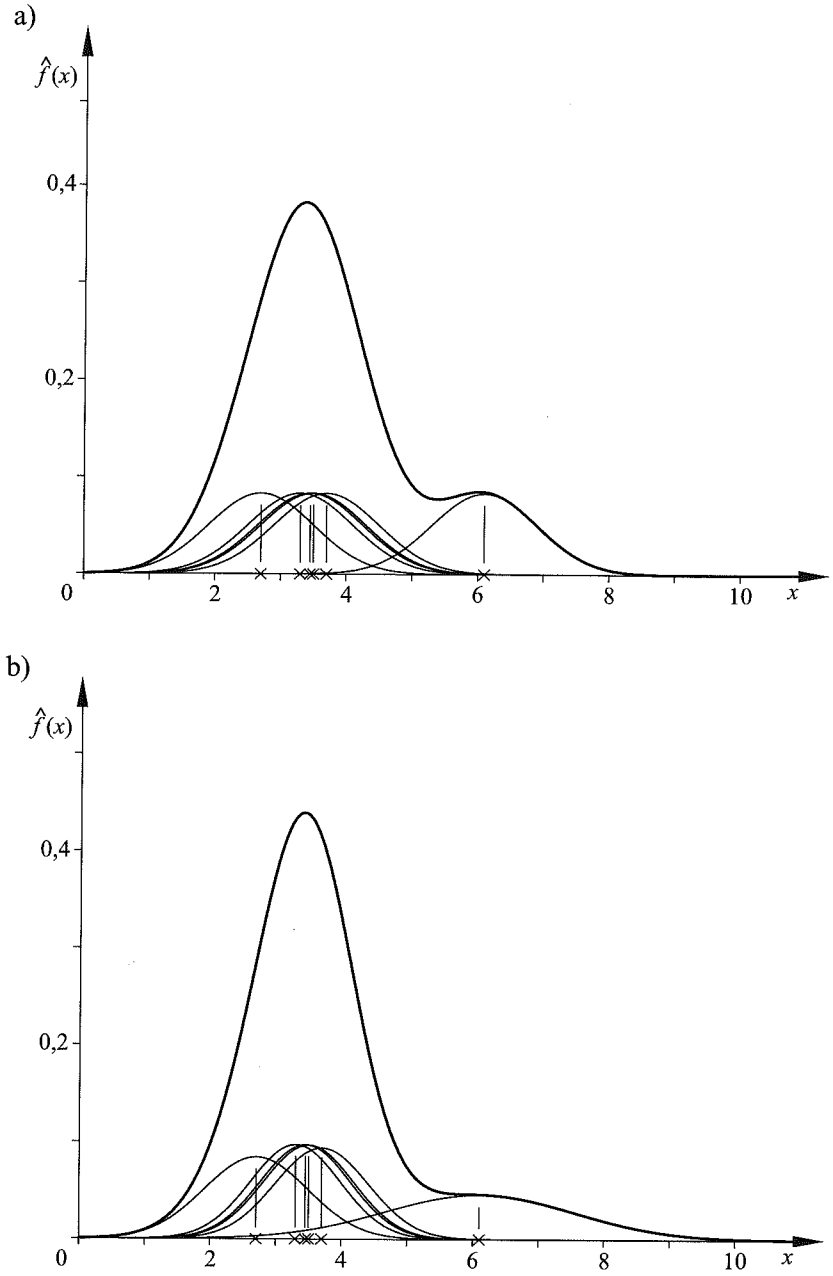
$$c = \frac{1}{2} \tag{3.115}$$

i jest ona polecana do stosowania w zadaniach aplikacyjnych niepoprzedzanych szczegółowymi badaniami statystycznymi.

Przedstawiona powyżej procedura modyfikacji parametru wygładzania ma wyrazistą, bliską intuicji interpretację. W praktycznych zastosowaniach skonstruowane z jej użyciem estymatory jądrowe mają jednak szereg dodatkowych, **pozytywnych własności**. Przede wszystkim są one odporniejsze na niewłaściwy dobór wartości parametru wygładzania h. Biorąc pod uwagę, iż zagadnienie to stanowi istotny problem podczas konstruowania estymatora jądrowego, powyższa cecha jest wyjątkowo korzystna dla praktyków niezainteresowanych wgłębianiem się w statystyczne aspekty zagadnienia.

Co więcej, stosowanie procedury modyfikacji zwiększa efektywność poszczególnych typów jąder względem optymalnego jądra Epanecznikowa, podobnie jak w przypadku wielowymiarowym zwiększa efektywność jądra produktowego względem radialnego. Ponownie, właściwość ta jest niezwykle cenna dla praktyków, gdyż podkreśla możliwość preferowania tych typów jąder, które są korzystne dla dalszej analizy w konkretnych zadaniach aplikacyjnych.

Podsumowując: powyższe istotne zalety, jakimi charakteryzują się estymatory jądrowe po zastosowaniu procedury modyfikacji parametru wygładzania, w połączeniu z jej prostotą obliczeniową, powodują, iż **stosowanie tej procedury należy w praktyce zalecać wręcz jako obligatoryjne**, ze szczególnym uzasadnieniem wobec przypadków, gdy są stosowane typy jąder o mniejszej efektywności, a w przypadku wielowymiarowym – jądro produktowe.



Rys. 3.9. Jądrowy estymator gęstości prawdopodobieństwa: a) bez modyfikacji parametru wygładzania (3.3), b) z modyfikacją parametru wygładzania (3.114)

3.1.7. Ograniczenie nośnika

Oprócz rozwiązań generalnie polepszających własności estymatora jądrowego, jak przedstawione wcześniej transformacja liniowa i modyfikacja parametru wygładzania, mogą być także stosowane dodatkowe, używane opcjonalnie procedury, dostosowujące własności estymatora do indywidualnych cech badanego zjawiska przypadkowego. W kolejnych dwóch punktach zostaną przedstawione najważniejsze z nich, pozwalające uwzględnić wymaganie ograniczenia nośnika estymatora i występowanie składowych binarnych.

W praktycznych zagadnieniach wartości zmiennej losowej X mogą reprezentować różnorodne wielkości, na przykład fizyczne lub ekonomiczne. Wiele z nich, jeśli ma być poprawnie i ściśle interpretowanych, powinno należeć jedynie do ograniczonych podzbiorów. Przykładowo wzrost człowieka lub czas wykonywania operacji technologicznej może być wyłącznie dodatni. Tymczasem nawet prawidłowo skonstruowany estymator jądrowy najczęściej przyjmuje wartości różne od zera także poza tymi podzbiorami. Jest to nieuniknione w przypadku użycia jądra o nieograniczonym nośniku, ale nawet w przypadku stosowania funkcji K o ograniczonym nośniku jądra odpowiadające skrajnym elementom próby losowej mogą częściowo „wykraczać" poza owe podzbiory. W większości praktycznych zadań nie ma to wpływu na poprawność ostatecznego wyniku, zwłaszcza że wartości, jakie przyjmuje estymator poza owymi ograniczonymi podzbiorami, są najczęściej znikome, nawet sumarycznie nieprzewyższające błędów wynikłych z probabilistycznej nieokreśloności zjawisk oraz niedokładności metrologicznej. Nietrudno jednak wskazać zagadnienia, w których sytuacja taka nie może być akceptowana. I tak na przykład w przypadku prezentacji wyników wobec gremium, któremu obce są zagadnienia związane z naturalnymi w praktyce przybliżeniami, odkrycie co prawda małego, ale niezerowego prawdopodobieństwa ujemnego wzrostu człowieka, mogłoby zdezawuować przedstawione wyniki, nawet gdyby były całkowicie poprawne w granicach akceptowalnego błędu.

W niniejszym punkcie zostanie opisana procedura ograniczająca nośnik estymatora jądrowego. Jako pierwszy będzie rozpatrywany przypadek jednowymiarowej zmiennej losowej.

Najpierw zostanie szczegółowo rozważony **przypadek ograniczenia lewostronnego**, czyli gdy spełniony ma być warunek $\hat{f}(x) = 0$ dla każdego $x < x_*$, przy dowolnie ustalonym $x_* \in \mathbb{R}$. Wymagane jest zatem, aby nośnik estymatora jądrowego był zawarty w przedziale $[x_*, \infty)$.

Istota proponowanej procedury sprowadza się do dokonania symetrycznego „odbicia" względem ograniczenia x_* tej części każdego jądra, która znajduje się poza przedziałem $[x_*, \infty)$. Tak sformułowana koncepcja została zilustrowana w kolejnych częściach rysunku 3.10. W części a jest pokazane jądro K odpowia-

dające i-tej wartości próby losowej x_i, dla dowolnie ustalonego $i \in \{1, 2, ..., m\}$. Fragment tego jądra, który znajduje się poza przedziałem $[x_*, \infty)$, został w części b „odbity" symetrycznie względem ograniczenia x_*, a w części c potraktowany jako fragment jądra odpowiadającego symetrycznemu „odbiciu" elementu x_i względem ograniczenia x_*, to znaczy wartości $2x_* - x_i$.

Z analitycznego punktu widzenia, jeśli dla dowolnie ustalonego $x_* \in \mathbb{R}$ dana jest funkcja charakterystyczna przedziału $[x_*, \infty)$, czyli

$$\chi_{[x_*,\infty)}(x) = \begin{cases} 1, & \text{gdy} \quad x \geq x_*, \\ 0, & \text{gdy} \quad x < x_*, \end{cases} \tag{3.116}$$

to zilustrowaną na rysunku 3.10 sumę dwóch jąder, odpowiadających elementom x_i oraz $2x_* - x_i$, można zapisać jako

$$K\left(\frac{x - x_i}{h}\right) + K\left(\frac{x - (2x_* - x_i)}{h}\right) =$$

$$= K\left(\frac{x - x_i}{h}\right) + K\left(\frac{x + x_i - 2x_*}{h}\right) \quad \text{dla dowolnego } i = 1, 2, ..., m. \tag{3.117}$$

Podstawowa definicja estymatora jądrowego (3.3) przyjmuje wtedy postać

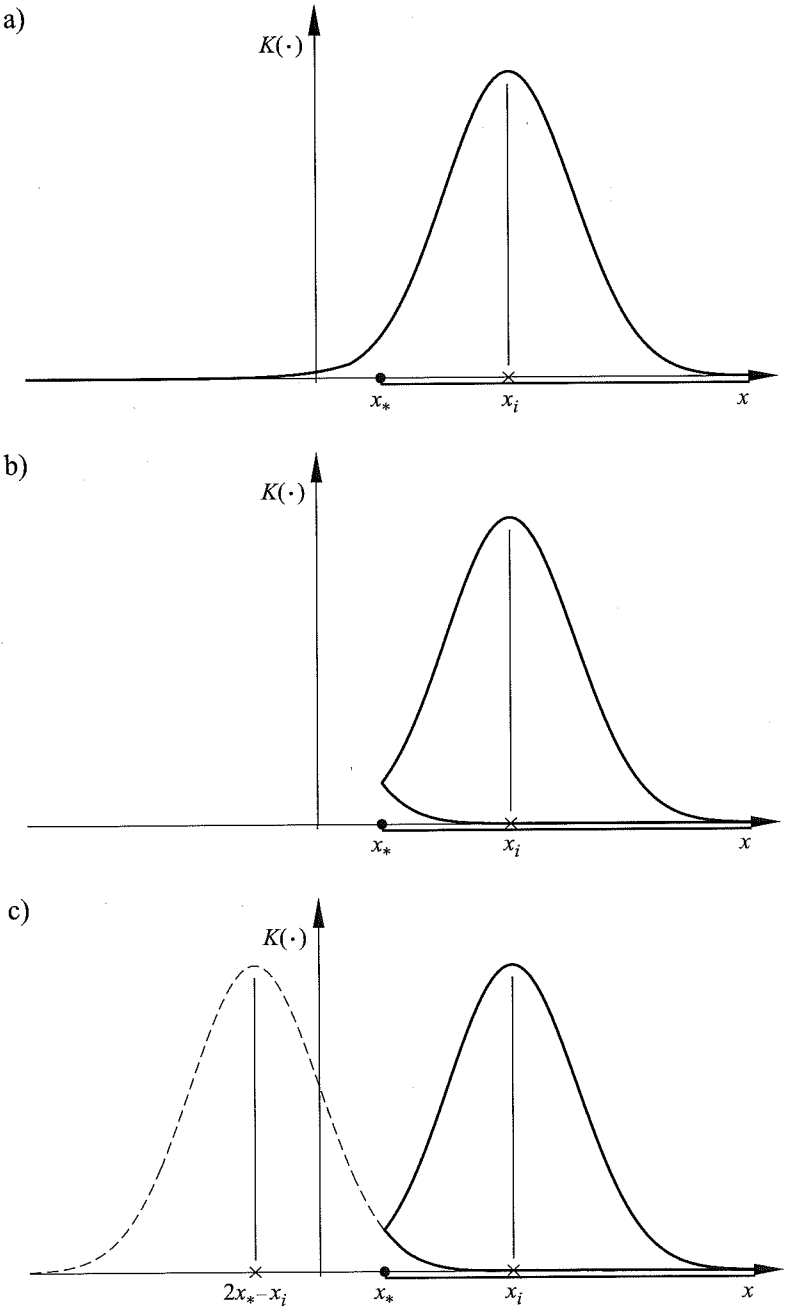
$$\hat{f}(x) = \frac{1}{mh} \sum_{i=1}^{m} \chi_{[x_*,\infty)}(x) \left[K\left(\frac{x - x_i}{h}\right) + K\left(\frac{x + x_i - 2x_*}{h}\right) \right], \tag{3.118}$$

natomiast po wprowadzeniu modyfikacji parametru wygładzania (3.114) zależność tę można przedstawić w postaci

$$\hat{f}(x) = \frac{1}{mh} \sum_{i=1}^{m} \frac{1}{s_i} \chi_{[x_*,\infty)}(x) \left[K\left(\frac{x - x_i}{hs_i}\right) + K\left(\frac{x + x_i - 2x_*}{hs_i}\right) \right]. \tag{3.119}$$

Estymatory (3.118)-(3.119) spełniają warunek $\hat{f}(x) = 0$ dla $x < x_*$, tak więc ich nośnik zawiera się w założonym przedziale $[x_*, \infty)$.

Estymator jądrowy wyznaczony zgodnie z przedstawioną wyżej koncepcją ma podstawową, szczególnie dogodną własność: fragmenty poszczególnych jąder „usuwane" poza założonym ograniczeniem są „wstawiane" do jego wnętrza w bezpośrednim sąsiedztwie owego ograniczenia, a więc w zakresie akceptowalnego w praktyce błędu (por. rys. 3.10). Skutkiem tego jest wszakże nieciągłość funkcji $\hat{f}$ na ograniczeniu x_*, także w przypadku stosowania ciągłego jądra K.



Rys. 3.10. Lewostronne ograniczenie nośnika estymatora jądrowego: a) pojedyncze jądro K odpowiadające elementowi x_i, b) jądro K z symetrycznym „odbiciem" jego fragmentu spoza przedziału $[x_*, \infty)$, c) symetryczne „odbicie" jako fragment jądra K odpowiadającego symetrycznemu „odbiciu" elementu x_i względem ograniczenia x_*

Jednak gdy funkcja K jest różniczkowalna, to $\lim_{x\to x_*^-} \hat{f}'(x) = \lim_{x\to x_*^+} \hat{f}'(x) = 0$. Jeżeli zatem uzupełnić definicję o $\hat{f}'(x_*) = 0$, to pochodna estymatora jądrowego jest ciągła na ograniczeniu x_*. Ostatecznie, wybór jądra K, takiego że jego pochodna jest ciągła, implikuje, że taka staje się też pochodna estymatora $\hat{f}'$. Analogiczne rozumowanie dotyczy dowolnej pochodnej wyższego rzędu.

Podobnie można rozważyć **przypadek ograniczenia prawostronnego** nośnika estymatora jądrowego, to znaczy gdy jest wymagany warunek $\hat{f}(x) = 0$ dla każdego $x > x^*$, przy dowolnie ustalonym $x^* \in \mathbb{R}$. Nośnik estymatora powinien się zatem zawierać w przedziale $(-\infty, x^*]$. Otóż, jeśli $\chi_{(-\infty,x^*]}$ oznacza funkcję charakterystyczną przedziału $(-\infty, x^*]$, czyli

$$\chi_{(-\infty,x^*]}(x) = \begin{cases} 1, & \text{gdy } x \le x^*, \\ 0, & \text{gdy } x > x^*, \end{cases} \tag{3.120}$$

to podstawowa definicja estymatora jądrowego (3.3) zostaje zmieniona na

$$\hat{f}(x) = \frac{1}{mh}\sum_{i=1}^{m} \chi_{(-\infty,x^*]}(x)\left[K\left(\frac{x-x_i}{h}\right) + K\left(\frac{x+x_i-2x^*}{h}\right)\right], \tag{3.121}$$

a po wprowadzeniu modyfikacji parametru wygładzania wzór (3.114) przyjmuje postać

$$\hat{f}(x) = \frac{1}{mh}\sum_{i=1}^{m} \frac{1}{s_i}\chi_{(-\infty,x^*]}(x)\left[K\left(\frac{x-x_i}{hs_i}\right) + K\left(\frac{x+x_i-2x^*}{hs_i}\right)\right]. \tag{3.122}$$

Powyższe estymatory spełniają warunek $\hat{f}(x) = 0$ dla każdego $x > x^*$, czyli – zgodnie z założeniem – ich nośnik zawiera się w przedziale $(-\infty, x^*]$. Nie są one ciągłe na ograniczeniu x^*, ale po uzupełnieniu dowolnej pochodnej w tym punkcie o wartość zero, pochodna ta staje się ciągła. Jeżeli zatem funkcja K ma ciągłą pochodną dowolnego rzędu, to pochodną taką ma także estymator $\hat{f}$.

Na koniec będzie skomentowany przypadek wielowymiarowej zmiennej losowej. Jeżeli jest stosowane jądro produktowe (3.48), to powyższej procedury można użyć bezpośrednio wobec poszczególnych jednowymiarowych czynników $\mathcal{K}(\cdot, x_j)$ tego typu jądra. Także w przypadku jądra radialnego (3.47) lub (3.50) koncepcję tę można wykorzystać względem poszczególnych współrzędnych, na których wymagane są ograniczenia.

3.1.8. Współrzędne binarne

W wielu problemach aplikacyjnych występują **wielkości binarne** (dwustanowe), to znaczy przyjmujące jedynie dwie wartości, symbolicznie oznaczane dalej jako 0 oraz 1. Na przykład istotną informacją w bazie danych medycznych jest płeć pacjenta, cecha o charakterze dwuwartościowym: „kobieta albo mężczyzna". Poza wielkościami mającymi ze swej natury postać binarną często stosuje się uproszczenia do takiej postaci opisu nawet bardziej złożonych zjawisk, lecz o pomniejszym wpływie na ostateczny wynik. Dzieje się tak nie tylko ze względu na powszechne w zagadnieniach praktycznych dążenie do uproszczenia modelu, lecz także z uwagi na częsty brak możliwości ścisłego ustalenia reprezentatywnej miary. Przykładowo, opinia, czy pacjent jest dobrze odżywiony, mogłaby być mało wiarygodna w przypadku kwalifikacji bardziej złożonej niż dwustanowa „prawidłowo lub nieprawidłowo", zwłaszcza gdyby oceny dokonywały osoby o różnym punkcie widzenia, odnoszące to zagadnienie do odmiennych warunków środowiskowych.

Prezentowana metodyka konstruowania estymatorów jądrowych umożliwia uwzględnienie specyfiki wielkości binarnych. Korzystną cechą przedstawionej poniżej koncepcji jest to, że jeśli do rozważanej zmiennej losowej X zostają wprowadzone współrzędne tego typu, to złożoność obliczeniowa wzrasta nieznacznie w zależności od ich liczby. W skomplikowanych problemach pozwala to uwzględnić wielkości o pomniejszym znaczeniu dla rozważanego zagadnienia, po uproszczeniu ich do postaci dwuwartościowej, na przykład „wpływ pozytywny albo negatywny" lub „wystąpienie albo brak danego czynnika".

Rozważana teraz będzie pomocniczo k-wymiarowa binarna zmienna losowa $Y: \Omega \to \{0, 1\}^k$. Wartościami, jakie może ona przyjąć, są k-wymiarowe wektory o współrzędnych binarnych:

$$\begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}, \begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ \vdots \\ 0 \end{bmatrix}, \ldots, \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix}. \tag{3.123}$$

Jest ich 2^k, a zatem skończona liczba. Podstawowa analiza takiej zmiennej stanowi zadanie klasycznego rachunku prawdopodobieństwa opartego na kombinatoryce. Rozkład zmiennej losowej Y jest charakteryzowany przez 2^k prawdopodobieństw wystąpienia każdego z wektorów (3.123). Klasyczna metoda jego estymacji polega na uzyskaniu ze zmiennej Y próby losowej i określeniu prawdopodobieństwa każdego wektora jako ilorazu liczby jego wystąpień w próbie do jej liczności. Taka metodyka wymaga jednak znacznej liczności próby lo-

sowej, co w praktyce stanowi często o wzroście kosztu prowadzonych tym sposobem badań. W zagadnieniach aplikacyjnych o prawdopodobieństwie wystąpienia dowolnego wektora (3.123) często można także wnioskować na podstawie obserwacji wektorów do niego „podobnych". Im owe „podobieństwo" jest większe, tym wnioskowanie takie jest bardziej wiarygodne. Powyższa koncepcja może być realizowana przez estymatory jądrowe w opisanej poniżej wersji binarnej.

Niech $\{0, 1\}^k$ oznacza zbiór wszystkich k-wymiarowych wektorów o wartościach binarnych. Dla ujednolicenia notacji zostanie wprowadzona funkcja $p: \{0, 1\}^k \to [0, \infty)$, przyporządkowująca każdemu k-wymiarowemu wektorowi o wartościach binarnych prawdopodobieństwo jego wystąpienia[1)]. Estymator jądrowy wprowadzonej w ten sposób funkcji p, charakteryzującej rozkład binarnej zmiennej losowej Y, wyznacza się na podstawie wartości uzyskanej z niej doświadczalnie m-elementowej próby losowej

$$y_1, y_2, \ldots, y_m \tag{3.124}$$

i jest on definiowany wzorem

$$\hat{p}(y) = \frac{1}{m} \sum_{i=1}^{m} L(y, y_i), \tag{3.125}$$

gdzie funkcję $L: \{0, 1\}^k \times \{0, 1\}^k \to [0, 1]$ określa zależność

$$L(y, y_i) = \delta^{k - d(y, y_i)} (1 - \delta)^{d(y, y_i)}, \tag{3.126}$$

przy czym $0{,}5 \le \delta \le 1$, a funkcja $d: \{0, 1\}^k \times \{0, 1\}^k \to \mathbb{N}$ jest dana jako

$$d(y, y_i) = (y - y_i)^{\mathrm{T}} (y - y_i); \tag{3.127}$$

w powyższym zapisie przyjmuje się $0^0 = 1$, co gwarantuje ciągłość funkcji L względem parametru δ dla dowolnie ustalonej wartości funkcji d.

Wartość funkcji d jest równa liczbie tych współrzędnych wektorów y oraz y_i, na których wektory te są różne, a zatem wielkość $k - d(y, y_i)$ wskazuje liczbę współrzędnych, na których wektory te są równe, i reprezentuje wspomniane wcześniej „podobieństwo" wektorów binarnych.

Funkcja L pełni rolę, jaką w definicji (3.3) odgrywało jądro K. Nietrudno pokazać, że

[1)] Funkcja ta odgrywa rolę zdefiniowanej w punkcie 2.1.2 gęstości prawdopodobieństwa f, jednakże ze względu na niespełnienie warunku (2.10) nie może być użyta *explicite* zgodnie ze wzorem (2.11).

$$\sum_{y\in\{0,1\}^k} L(y, y_i) = 1 \quad \text{dla dowolnych } \delta \in [0{,}5;\ 1] \text{ oraz } y_i \in \{0,\ 1\}^k \text{ przy } i = 1, 2, \ldots, m. \tag{3.128}$$

Im wektor y jest bardziej „podobny" do wartości próby losowej y_i, tym wartość $L(y, y_i)$ jest większa. Funkcja L jest zatem nazywana **jądrem binarnym**.

Z kolei parametr δ ma analogiczne znaczenie dla jądra L jak parametr h dla jądra K z definicji (3.3). Niech wartość próby losowej y_i będzie dowolnie ustalona. W przypadku $\delta = 1$ ze wzoru (3.126) otrzymuje się

$$L(y, y_i) = \begin{cases} 1 & \text{dla} \quad y = y_i, \\ 0 & \text{dla} \quad y \neq y_i, \end{cases} \tag{3.129}$$

a zatem funkcja L jest „skoncentrowana" na y_i. W drugim przypadku granicznym, to znaczy gdy $\delta = 0{,}5$, wartość funkcji L jest stała i wynosi $1/2^k$ dla dowolnego $y \in \{0,\ 1\}^k$, jest zatem niezależna od wartości próby y_i. Dla wartości $\delta \in (0{,}5;\ 1)$ występują przypadki pośrednie: im wartość tego parametru jest bliższa 1, tym wartości funkcji $L(\cdot, y_i)$ bardziej wzrastają w otoczeniu y_i, natomiast im jest ona bliższa 0,5, tym bardziej funkcja ta upodobnia się do funkcji stałej.

Przenosząc te rozważania na funkcję $\hat{p}$ zdefiniowaną wzorem (3.125), można zauważyć, iż w przypadku $\delta = 0{,}5$ powyższa funkcja jest stała o wartości $1/2^k$, a zatem jest niezależna od uzyskanej doświadczalnie próby losowej (3.124). Z kolei, jeśli $\delta = 1$, to wartości $\hat{p}(y)$ są tożsame ze wspomnianym klasycznym ujęciem prawdopodobieństwa i są równe ilorazowi liczby wystąpień wektora y w próbie losowej do jej liczności. W przypadkach pośrednich $\delta \in (0{,}5;\ 1)$ na wartość $\hat{p}(y)$ mają wpływ wszystkie elementy uzyskanej doświadczalnie próby losowej, przy czym im bardziej dany element próby losowej jest „podobny" do y, tym wpływ ten jest większy. O stopniu owego wpływu stanowi wartość parametru δ – im wartość ta jest większa, tym wpływ staje się mniejszy. Własności funkcji L dla $\delta \to 0{,}5^+$ przypominają zatem własności jądra K, gdy $h \to 0^+$, podobnie jak $\delta \to 1^-$ oraz $h \to \infty$. Przez analogię, stała δ jest zatem określana mianem **binarnego parametru wygładzania**.

W praktyce wartość binarnego parametru wygładzania wyznacza się, minimalizując funkcję $g : [0{,}5;\ 1] \to \mathbb{R}$ zdefiniowaną wzorem

$$g(\delta) = -\prod_{i=1}^{m} \hat{p}_{-i}(y_i), \tag{3.130}$$

gdzie $\hat{p}_{-i}$ dla $i = 1, 2, \ldots, m$ oznacza estymator (3.125) wyznaczony na podstawie próby losowej $y_1, y_2, \ldots, y_{i-1}, y_{i+1}, \ldots, y_m$, a zatem próby (3.124) z pominięciem i-tego elementu. Procedurę wyznaczenia minimum funkcji g można

zrealizować, wykorzystując algorytm opisany wzorami (3.88)-(3.105), przyjmując we wstępnym kroku

$$a_0 = 0{,}5, \tag{3.131}$$

$$b_0 = 1 \tag{3.132}$$

i zastępując w ostatnim, k-tym, kroku wzory (3.104)-(3.105) odpowiednio przez:

- jeżeli $g(p_k) \leq g(q_k)$, to należy przyjąć

$$\delta = \frac{a_k + q_k}{2}; \tag{3.133}$$

- jeżeli $g(p_k) > g(q_k)$, to

$$\delta = \frac{p_k + b_k}{2}. \tag{3.134}$$

Minimum jest wtedy wyznaczane z dokładnością

$$\Delta = \frac{1}{2}(b_k - a_k) = \frac{1}{2}\left(\frac{\sqrt{5}-1}{2}\right)^{k+1}(b_0 - a_0) = \frac{1}{4}\left(\frac{\sqrt{5}-1}{2}\right)^{k+1}. \tag{3.135}$$

Oznaczając – jak poprzednio – żądaną dokładność przez $\Delta > 0$, można stąd obliczyć wymaganą liczbę iteracji

$$k = \overline{\mathrm{Int}}\left(\frac{\ln[4\Delta]}{\ln\left[\frac{\sqrt{5}-1}{2}\right]}\right) - 1, \tag{3.136}$$

gdzie $\overline{\mathrm{Int}}(c)$ oznacza najmniejszą liczbę całkowitą większą lub równą $c \in \mathbb{R}$. W szczególności dla wystarczającego w praktyce

$$\Delta = 0{,}01 \tag{3.137}$$

otrzymuje się

$$k = 6. \tag{3.138}$$

Czasochłonne obliczenia wartości funkcji g, zdefiniowanej wzorem (3.130), będą zatem wykonywane 8-krotnie.

Na koniec zostanie rozważona $(n+k)$-wymiarowa zmienna losowa $Z \equiv \begin{bmatrix} X \\ Y \end{bmatrix}$, będąca zestawieniem rozważanej dotychczas w niniejszej książce n-wymiarowej zmiennej X oraz wprowadzonej powyżej k-wymiarowej binarnej zmiennej Y, z zachowaniem przyjętych dla nich oznaczeń. Jeżeli estymator jądrowy $\hat{f}: \mathbb{R}^n \times \{0,1\}^k \to [0,\infty)$ zmiennej losowej Z jest wyznaczany na podstawie m-elementowej próby losowej

$$\begin{bmatrix} x_1 \\ y_1 \end{bmatrix}, \begin{bmatrix} x_2 \\ y_2 \end{bmatrix}, \ldots, \begin{bmatrix} x_m \\ y_m \end{bmatrix}, \tag{3.139}$$

to w podstawowej postaci definiuje się go wzorem

$$(\hat{f},\hat{p})\left(\begin{bmatrix} x \\ y \end{bmatrix}\right) = \frac{1}{mh^n}\sum_{i=1}^{m} K\left(\frac{x-x_i}{h}\right) L(y,y_i), \tag{3.140}$$

natomiast w przypadku stosowania modyfikacji parametru wygładzania ma postać

$$(\hat{f},\hat{p})\left(\begin{bmatrix} x \\ y \end{bmatrix}\right) = \frac{1}{mh^n}\sum_{i=1}^{m} \frac{1}{s_i^n} K\left(\frac{x-x_i}{hs_i}\right) L(y,y_i). \tag{3.141}$$

Należy tu jednak zaznaczyć, że w praktycznych zastosowaniach estymatorów jądrowych podstawowa informacja jest zawarta w składowych rozpatrywanej dotychczas zmiennej X, a składowe binarne zmiennej Y stanowią jedynie ich uzupełnienie. Dlatego wydaje się, iż dla spójności zapisu warto traktować powyższe jako estymator jądrowy zmiennej losowej X (jak to było dotychczas rozważane) z dodatkiem składowych binarnych reprezentowanych przez zmienną Y, zapisując wzory (3.140)-(3.141) w postaci

$$\hat{f}\left(\begin{bmatrix} x \\ y \end{bmatrix}\right) = \frac{1}{mh^n}\sum_{i=1}^{m} K\left(\frac{x-x_i}{h}\right) L(y,y_i) \tag{3.142}$$

oraz

$$\hat{f}\left(\begin{bmatrix} x \\ y \end{bmatrix}\right) = \frac{1}{mh^n}\sum_{i=1}^{m} \frac{1}{s_i^n} K\left(\frac{x-x_i}{hs_i}\right) L(y,y_i), \tag{3.143}$$

czyli pomijając w zapisie lewych stron symbol $\hat{p}$. Ostatecznie, zależności (3.142)-(3.143) dostarczają definicji **estymatora jądrowego z dodatkowym uwzględ-**

nieniem składowych binarnych, odpowiednio, w postaci podstawowej i w przypadku stosowania modyfikacji parametru wygładzania. Warto powtórzyć, iż w praktycznych zastosowaniach, a zwłaszcza w najbardziej złożonych zadaniach, procedura taka może zapewnić znaczące korzyści, pozwalając uwzględnić nie tylko wielkości będące ze swej natury binarnymi, ale także wielkości o odmiennym charakterze, lecz uproszczone do postaci binarnej, które ze względu na pomniejsze dla zagadnienia znaczenie musiałyby być pominięte. Należy jednak przestrzec przed zbyt pochopnym włączaniem do modelu wielkości binarnych, zwłaszcza pojedynczej, czyli przy $k = 1$. Jeżeli bowiem rozkłady zmiennej X zbyt różnią się między sobą dla wartości binarnych 0 oraz 1, to korzystniejsze może okazać się uwzględnienie dwóch niezależnych modeli odrębnie dla tych wartości, biorąc nawet pod uwagę to, że liczności prób losowych użytych do konstrukcji każdego z nich byłyby o około połowę mniejsze. Pozytywne aspekty wprowadzenia wielkości binarnych wyraziście uwidaczniają się przy większej liczbie współrzędnych o tym charakterze. Ograniczenie stanowią tu jednak wymagania dotyczące liczności próby losowej, wzrastające wraz z powiększaniem wymiaru k. Zagadnienie to będzie tematem rozważań przedstawionych w następnym punkcie.

3.1.9. Liczność próby losowej

Ostatnią wielkością wymagającą uściślenia jest liczność próby losowej m. Wartość tego parametru ma znaczący wpływ na jakość otrzymanego estymatora jądrowego. Z teoretycznego punktu widzenia, ze względu na własności asymptotycznej nieobciążoności i zgodności oraz zmniejszanie się wariancji estymatora i jego błędu średniokwadratowego, wskazana jest jak największa liczność próby losowej. Jednakże w praktyce konieczny staje się kompromis, w którym byłyby także uwzględnione aspekty praktyczne, a zwłaszcza dostępność i koszty uzyskiwania danych oraz ograniczenia czasowe obliczeń. Wymagana liczność próby zależy przede wszystkim od wymiaru badanej zmiennej losowej, a także od własności jej rozkładu, zwłaszcza od liczby wartości modalnych oraz symetrii poszczególnych, związanych z tymi wartościami składników. Ponadto w zagadnieniach aplikacyjnych pewien wpływ może mieć skorelowanie poszczególnych elementów próby losowej, gdyż założenie, iż próba ta jest prosta, nie zawsze jest w praktyce spełnione. Zagadnienia te będą kolejno rozważane w niniejszym punkcie.

W przypadku stosowania podstawowej postaci estymatora jądrowego (3.3) można jako liczność próby, konieczną do zapewnienia 10-procentowej dokładności w punkcie zero dla n-wymiarowego standardowego rozkładu normalnego, przyjąć w przybliżeniu wartość 4^n. Ze względu na szczególną regularność tego rozkładu oraz mało rygorystyczny charakter powyższego kry-

terium wydaje się, że wartość ta stanowi bezwzględne minimum. Widać jednak, iż **wraz ze zwiększaniem wymiaru *n* wymagana liczność próby gwałtownie rośnie**. Uwidacznia się tu zjawisko powszechnie występujące podczas statystycznej analizy wielowymiarowych zmiennych losowych, które otrzymało nawet swoistą nazwę „przekleństwa wielowymiarowości". Jest ono wszakże całkowicie uzasadnione, jeśli uwzględnić mnogość i różnorodność informacji o wszechstronnych zależnościach między poszczególnymi współrzędnymi wielowymiarowej zmiennej, zawartych w estymatorze jądrowym. Charakteryzuje on bowiem nie tylko rozkłady poszczególnych (pojedynczych) współrzędnych, ale także złożone zależności między nimi, zarówno bilateralne (między każdą parą współrzędnych), jak i między ich grupami (przykładowo, określając w przypadku pięciowymiarowym prawdopodobieństwo, że wartości dwóch pierwszych współrzędnych badanej zmiennej należą do pewnego zbioru z przestrzeni $\mathbb{R}^2$, gdy wartości trzech kolejnych należą do innego podzbioru $\mathbb{R}^3$).

Ocena stopnia trudności estymacji poszczególnych rozkładów jest możliwa na podstawie kryterium scałkowanego błędu średniokwadratowego MISE, w szczególności wzoru na jego wartość (3.32), uogólnionego dla przypadku wielowymiarowego do postaci (3.39). Wskazuje na to występujące tam wyrażenie $Z(f)$, zdefiniowane wzorami (3.19) i (3.35)-(3.36). W przeciwieństwie jednak do rozważanych wcześniej czynników związanych z postacią jądra, których wartość mogła podlegać optymalizacji poprzez odpowiedni wybór postaci funkcji K, wartość wyrażenia $Z(f)$ jest determinowana przez badaną rzeczywistość i jedyne możliwe w tym zakresie działanie stanowi odpowiednie dopasowanie wartości parametru m, czyli liczności posiadanej próby losowej.

Najpierw zostanie rozpatrzony przypadek jednowymiarowy. I tak, w nawiązaniu do wzoru (3.32), będzie rozważana wartość czynnika $[Z(f)/m^4]^{1/5}$, czyli

$$\left[\frac{\int_{-\infty}^{\infty} f''(x)^2\,dx}{m^4}\right]^{\frac{1}{5}}. \tag{3.144}$$

Wynikające stąd wnioski są korzystne z aplikacyjnego punktu widzenia. Jeżeli bowiem dla jednego z ewentualnie porównywanych rozkładów wyrażenie $\int_{-\infty}^{\infty} f''(x)^2\,dx$ byłoby w-krotnie większe, to aby otrzymać tę samą wartość błędu estymacji, liczność próby losowej należy zwiększyć tylko $\sqrt[4]{w}$ -krotnie. Co więcej, gdy nie jest to możliwe, wówczas wartość błędu powiększy się jedynie $\sqrt[5]{w}$ -krotnie.

Rozpatrzmy teraz możliwe rozkłady zmiennej losowej „przeskalowane" tak, aby ich wariancja była jednostkowa (w przypadku estymatora jądrowego funkcję tę spełnia optymalny dobór wartości parametru wygładzania). Wyrażenie $\int_{-\infty}^{\infty} f''(x)^2 dx$ przyjmuje wtedy najmniejszą wartość dla rozkładu beta z parametrami $p = 4$, $q = 4$ (por. rys. 2.2). Rozkład ten okazuje się zatem najdogodniejszy do szacowania za pomocą estymatorów jądrowych. Wykres jego gęstości prawdopodobieństwa przypomina wykres gęstości rozkładu normalnego, aczkolwiek nośnik jest ograniczony. Generalnie rozkład beta przy $p = 4$, $q = 4$ jest unimodalny oraz symetryczny, i ogólnie rozkłady tego typu są najdogodniejsze do estymacji za pomocą estymatorów jądrowych. Wymagane liczności próby losowej nie różnią się znacząco w klasie rozkładów o tych własnościach: na przykład w przypadku rozkładu normalnego różnica ta wynosi 9%. Uzasadnia to użycie powszechnie znanego rozkładu normalnego w celu sformułowania, na początku niniejszych rozważań, wstępnego oszacowania minimalnej liczności próby losowej.

Wielomodalność badanego rozkładu skutkuje wzrostem wymagań dotyczących liczności próby. W przypadku takiego rozkładu funkcja f'' naprzemiennie bowiem rośnie i maleje, co znacząco zwiększa wartość wyrażenia $\int_{-\infty}^{\infty} f''(x)^2 dx$. Fakt ten można inaczej zinterpretować, zauważając, że każdy ze składników związanych z poszczególnymi wartościami modalnymi jest jak gdyby estymowany jedynie na podstawie części próby losowej. Generalnie dwumodalność implikuje konieczność prawie dwukrotnego powiększenia liczności próby, jednak dalsze zwiększanie liczby wartości modalnych powoduje już wolniejszy niż wprost proporcjonalny wzrost wymagań w tym zakresie.

Podobną analizę można przeprowadzić w odniesieniu do własności **symetrii** badanego rozkładu. W przypadku rozkładów wielomodalnych własność ta nie jest tu jednak rozumiana w aspekcie globalnym, lecz odrębnie jako symetria poszczególnych składników związanych z wartościami modalnymi. Trudności wynikłe z faktu estymacji niesymetrycznej funkcji $\hat{f}$ za pomocą kombinacji liniowej symetrycznych z założenia funkcji K są intuicyjnie zrozumiałe. Ze względu na złożony i często lokalny charakter tego zjawiska wielkość koniecznego zwiększenia liczności próby losowej trudno tu wyrazić w sposób ścisły, ale w praktyce niesymetria rozkładu rzadko wymaga jej wzrostu większego niż dwukrotny.

W punkcie 2.2.1 zostało założone, że rozważane w tej książce próby losowe są proste, a zatem ich poszczególne elementy pozostają niezależne. W praktyce oznacza to niezależność eksperymentów, na podstawie których wartości te zostały uzyskane. Metodykę konstruowania estymatorów jądrowych można jednak stosować również w przypadku, gdy wykorzystana do tego celu próba losowa

nie jest prosta, a jej elementy są **skorelowane**, aczkolwiek wymaga to między innymi odpowiedniego zwiększenia liczności tej próby. Wymaganie takie jest intuicyjnie oczywiste, gdyż w takiej sytuacji każdy element wnosi mniej informacji, ponieważ jej część jest już zawarta w skorelowanych z nim elementach próby. Najczęściej zdarza się to, gdy wartości próby losowej są otrzymywane sekwencyjnie w miarę upływu czasu, a każdy element jest w pewnym stopniu związany ze swym poprzednikiem. Typowy przykład reprezentują tu dane meteorologiczne. Korelacja rzadko wymaga znaczącego zwiększenia liczności próby losowej, w praktyce wielkość ta wynosi najczęściej 10-30%.

Dla wielowymiarowej zmiennej losowej wzór (3.144) uogólnia się, uwzględniając zależność (3.39) wraz z (3.35)-(3.36), do postaci

$$\left[\frac{\int_{-\infty}^{\infty}\left(\sum_{i=1}^{n}\frac{\partial^2 f(x)}{\partial . x_i^2}\right)^2 dx}{m^{4/n}}\right]^{\frac{n}{n+4}}. \tag{3.145}$$

W przypadku w-krotnego zwiększenia licznika powyższego wyrażenia, w celu uzyskania tego samego błędu estymacji należy liczność próby losowej zwiększyć $\sqrt[4/n]{w}$-krotnie. Gdyby to nie nastąpiło, wartość tego błędu zwiększy się $\sqrt[(n+4)/n]{w}$-krotnie. Powyższe wyniki są zatem bardziej wymagające niż w opisywanym wzorem (3.144) przypadku jednowymiarowej zmiennej losowej. Na przykład gdyby przyjąć $w = 2$, to dla $n = 1$, $n = 2$, $n = 4$ wartość próby należy zwiększyć, odpowiednio, o 18, 41, 100%, a w przypadku niemożności spełnienia tego warunku wartość błędu wzrosłaby, odpowiednio, o 15, 26, 41%. Pozostała część rozważań dotyczących rozkładów wielomodalnych, niesymetrycznych, a także skorelowanych wartości próby losowej jest taka sama, jak przedstawiono wcześniej dla jednowymiarowej zmiennej losowej.

I na koniec, w przypadku obecności k **współrzędnych binarnych** liczność próby, wynikającą z opisanej powyższej analizy, należy w praktyce dodatkowo pomnożyć przez $(3/2)^k$.

Oprócz przedstawionych aspektów istnieje wiele innych pomniejszych czynników, które mogą implikować potrzebę zwiększenia liczności próby, na przykład mniejsza efektywność użytego – z uwagi na wymagania aplikacyjne – jądra. Jednakże zaprezentowane dodatkowe procedury, jak transformacja liniowa, a zwłaszcza modyfikacja parametru wygładzania, umożliwiają istotną redukcję skutków powyższego. Co więcej, w poszczególnych przypadkach korzystne mogą być także proponowane opcjonalne procedury, na przykład ograniczenia nośnika estymatora jądrowego. Warto bowiem zauważyć, że skrajnie niesymetryczne rozkłady, jak na przykład wykładniczy (por. rys. 2.2), są związane z ograniczeniem nośnika badanej zmiennej i zastosowanie tej procedury w znaczącym stop-

niu redukuje negatywne skutki niesymetrii. Mnogość i złożoność powyższych zjawisk nie pozwalają na skonstruowanie uniwersalnej formuły umożliwiającej wyznaczenie wymaganej liczności próby losowej. Nie miałaby ona zresztą praktycznego znaczenia, gdyż w zagadnieniach aplikacyjnych trudno jest arbitralnie ustalić żądaną jakość modelu. Ocena jakości nierzadko wymaga wszechstronnych badań i rekurencyjnej modyfikacji modelu w celu ostatecznego ustalenia kompromisu między jego dokładnością a kosztami uzyskiwania danych oraz pożądaną w dalszych fazach analizy prostotą i szybkością obliczeniową.

Podsumowując: w niniejszym punkcie została przedstawiona analiza pozwalająca na wstępne oszacowanie liczności próby losowej wymaganej w celu wyznaczenia estymatora jądrowego. Dla konkretnego wymiaru n jako wyjściową wielkość można przyjąć wartość 4^n, ewentualnie mnożąc ją przez współczynnik wynikający z obecności k zmiennych binarnych $(3/2)^k$. Z kolei uzyskany wynik mnoży się przez heurystycznie określane współczynniki związane z koniecznością polepszenia jakości estymacji, wielomodalnością i niesymetrią rozkładu oraz korelacją elementów próby losowej – w praktyce iloczyn tych współczynników wynosi najczęściej 3-10, w ekstremalnych przypadkach nawet 100. Dla jednowymiarowej zmiennej losowej wymagana liczność próby wynosi w praktyce 20-50, odpowiednio zwiększając się wraz ze wzrostem wymiaru zmiennej. Jednak dzięki współczesnej technice komputerowej, nawet w szczególnie złożonych zagadnieniach wielowymiarowych i wielomodalnych, nie musi to obecnie stanowić istotnej przeszkody aplikacyjnej. Zawsze należy przecież pamiętać o znaczących korzyściach wynikających ze stosowania metod estymacji nieparametrycznej, w tym estymatorów jądrowych. Umożliwiają one bowiem identyfikację charakterystyki występujących w zadaniach praktycznych rozkładów, aczkolwiek wymagają stosownej liczności próby, adekwatnej do mnogości i wszechstronności zawartej informacji, którą można w zróżnicowany sposób wykorzystać w dalszej analizie na potrzeby konkretnych zagadnień aplikacyjnych.

Przedstawiony powyżej punkt widzenia jest typowy dla tradycyjnej statystyki matematycznej. Celem estymacji jest tu możliwie jak najdokładniejsze scharakteryzowanie, na podstawie próby, abstrakcyjnej zmiennej losowej opisującej badany fragment rzeczywistości. Odmienny punkt widzenia jest prezentowany w ramach naturalnej koncepcji analizy danych. Główny zamysł stanowi tu uzyskanie z próby losowej możliwie jak najbardziej użytecznych dla danego zagadnienia informacji. Ponieważ nie ma tu żadnego abstrakcyjnego nawet wzorca, więc zamiast dokładności rozważa się raczej adekwatność, a przede wszystkim potencjalną przydatność modelu otrzymanego z posiadanej próby. W tym przypadku formułowane w niniejszym punkcie wymagania dotyczące niezbędnej liczności próby losowej mają znaczenie drugorzędne – podstawowym zagadnieniem staje się praktyczna użyteczność otrzymanego wyniku.

3.1.10. Przykłady

Niniejszy punkt zawiera pięć przykładów ilustrujących aspekty obliczeniowe przedstawionej uprzednio metodyki konstruowania estymatorów jądrowych. Wszechstronne możliwości aplikacyjne powodują, iż estymatory te mogą być stosowane przez specjalistów z dziedzin niezwiązanych ściśle z matematyką. W tych przypadkach cenna wydaje się możliwość sprawdzenia poprawności interpretacji poszczególnych procedur i wzorów, które będą zilustrowane na konkretnych przykładach obliczeniowych. W celu zapewnienia przejrzystości **przykłady te zostaną skonstruowane dla prób co najwyżej pięcioelementowych, absolutnie nieakceptowalnych w praktyce**. Aplikacyjna użyteczność estymatorów jądrowych jest bowiem nieodłącznie związana z zastosowaniem systemów komputerowych, ale przy stosowanych w takich przypadkach licznościach próby przedstawione obliczenia byłyby całkowicie nieczytelne. Warto w tym miejscu przypomnieć treść punktu 3.1.9, a jako przykład uzasadniony z punktu widzenia uwarunkowań aplikacyjnych przywołać estymator jądrowy pokazany we wstępie na rysunku 1.6 dla złożonej struktury danych, a uzyskany na podstawie 200-elementowej próby losowej.

Poniżej zostanie przedstawionych pięć przykładów obejmujących wszystkie opisane w niniejszej monografii procedury, w których rozważane będą kolejno:

- jednowymiarowa zmienna losowa,
- jednowymiarowa zmienna losowa z ograniczeniem nośnika,
- jednowymiarowa zmienna losowa ze składową binarną,
- wielowymiarowa zmienna losowa z użyciem jądra produktowego,
- wielowymiarowa zmienna losowa z użyciem jądra radialnego.

Przykład 3.1 – jednowymiarowa zmienna losowa

W niniejszym przykładzie będzie rozważana jednowymiarowa zmienna losowa X ($n = 1$) i uzyskana z niej trójelementowa próba losowa ($m = 3$) o wartościach

$$x_1 = 3, \quad x_2 = 3{,}3, \quad x_3 = 5. \tag{3.146}$$

Jak wielokrotnie było podkreślane, wybór jądra, czyli postaci funkcji K, może być dokonany arbitralnie, zgodnie z wymaganiami późniejszej analizy w rozważanym zagadnieniu praktycznym. Nie mając konkretnych wskazań lub podczas wstępnych czynności wygodnie jest wybrać jądro normalne (3.44), powtórzone poniżej w postaci

$$K_N(x) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right). \tag{3.147}$$

Jądro to będzie stosowane w niniejszym przykładzie.

W przypadku rozważanej tu jednowymiarowej zmiennej losowej do wyznaczenia wartości parametru wygładzania h jest zalecana metoda podstawień. Zgodnie z ogólnymi wskazaniami założono drugi jej rząd. Najpierw, na podstawie wzoru (2.47), oblicza się wartość estymatora wariancji:

$$\hat{V} = \frac{1}{3-1}(3^2 + 3{,}3^2 + 5^2) - \frac{1}{3 \cdot 2}(3 + 3{,}3 + 5)^2 = 1{,}163, \tag{3.148}$$

a po podstawieniu do zależności (2.48) – estymatora odchylenia standardowego:

$$\hat{\sigma} = \sqrt{1{,}163} = 1{,}078. \tag{3.149}$$

Przyjmując – jak to jest powszechnie czynione w metodzie podstawień – iż do estymacji kolejnych pochodnych jest stosowane jądro normalne (3.44) zapisane poniżej zgodnie z wprowadzonymi uprzednio oznaczeniami jako

$$\widetilde{K}_{\mathrm{N}}(x) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right), \tag{3.150}$$

otrzymuje się z zależności (3.65), że

$$c_8 = \frac{105}{32 \cdot \sqrt{\pi} \cdot 1{,}078^9} = 0{,}942, \tag{3.151}$$

a podstawiając powyższy wynik wraz z równościami (3.72) i (3.76) do wzoru (3.66):

$$h_{\mathrm{II}} = \left(\frac{-2 \cdot -\frac{15}{\sqrt{2\pi}}}{1 \cdot 0{,}942 \cdot 3} \right)^{1/9} = 1{,}174. \tag{3.152}$$

Na podstawie zależności (3.64) można stąd obliczyć

$$\begin{aligned} C_{6;\,1{,}174} = \frac{1}{3^2 \cdot 1{,}174^7} \Bigg[& \widetilde{K}^{(6)}\left(\frac{3-3}{1{,}174}\right) + \widetilde{K}^{(6)}\left(\frac{3-3{,}3}{1{,}174}\right) + \widetilde{K}^{(6)}\left(\frac{3-5}{1{,}174}\right) + \\ & + \widetilde{K}^{(6)}\left(\frac{3{,}3-3}{1{,}174}\right) + \widetilde{K}^{(6)}\left(\frac{3{,}3-3{,}3}{1{,}174}\right) + \widetilde{K}^{(6)}\left(\frac{3{,}3-5}{1{,}174}\right) + \\ & + \widetilde{K}^{(6)}\left(\frac{5-3}{1{,}174}\right) + \widetilde{K}^{(6)}\left(\frac{5-3{,}3}{1{,}174}\right) + \widetilde{K}^{(6)}\left(\frac{5-5}{1{,}174}\right) \Bigg], \end{aligned} \tag{3.153}$$

a po uwzględnieniu postaci funkcji $\widetilde{K}^{(6)}$ określonej poprzez (3.75):

$$C_{6;\,1,174} = 0{,}036\,(-5{,}984 - 4{,}682 + 1{,}281 - 4{,}682 - 5{,}984 + 3{,}164 + \\ + 1{,}281 + 3{,}164 - 5{,}984) = -0{,}663. \tag{3.154}$$

Podstawiając powyższe wraz z równościami (3.72) i (3.74) do wzoru (3.67), uzyskuje się

$$h_1 = \left(\frac{-2 \cdot \dfrac{3}{\sqrt{2\pi}}}{1 \cdot -0{,}663 \cdot 3} \right)^{1/7} = 1{,}027 \tag{3.155}$$

i ponownie na podstawie zależności (3.64) można otrzymać

$$C_{4;\,1,027} = \frac{1}{3^2 \cdot 1{,}027^5} \left[\widetilde{K}^{(4)}\left(\frac{3-3}{1{,}027}\right) + \widetilde{K}^{(4)}\left(\frac{3-3{,}3}{1{,}027}\right) + \widetilde{K}^{(4)}\left(\frac{3-5}{1{,}027}\right) + \right. \\ + \widetilde{K}^{(4)}\left(\frac{3{,}3-3}{1{,}027}\right) + \widetilde{K}^{(4)}\left(\frac{3{,}3-3{,}3}{1{,}027}\right) + \widetilde{K}^{(4)}\left(\frac{3{,}3-5}{1{,}027}\right) + \\ \left. + \widetilde{K}^{(4)}\left(\frac{5-3}{1{,}027}\right) + \widetilde{K}^{(4)}\left(\frac{5-3{,}3}{1{,}027}\right) + \widetilde{K}^{(4)}\left(\frac{5-5}{1{,}027}\right) \right], \tag{3.156}$$

a po uściśleniu postaci funkcji $\widetilde{K}^{(4)}$ dzięki równaniu (3.73):

$$C_{4;\,1,027} = 0{,}092\,(1{,}197 + 0{,}954 - 0{,}322 + 0{,}954 + 1{,}197 - 0{,}601 - \\ - 0{,}322 - 0{,}601 + 1{,}197) = 0{,}336. \tag{3.157}$$

Wreszcie, korzystając z zależności (3.68), dostajemy

$$Z(f) = C_{4;\,1,027} = 0{,}336, \tag{3.158}$$

co wraz z wartością

$$\frac{W(K_N)}{U(K_N)^2} = 0{,}354, \tag{3.159}$$

otrzymaną z równości (3.61) dla jądra normalnego K_N, pozwala skompletować wielkości wymagane we wzorze (3.31) i uzyskać poszukiwaną wartość parametru wygładzania

$$h = \left(\frac{0{,}354}{0{,}336 \cdot 3} \right)^{1/5} = 0{,}811. \tag{3.160}$$

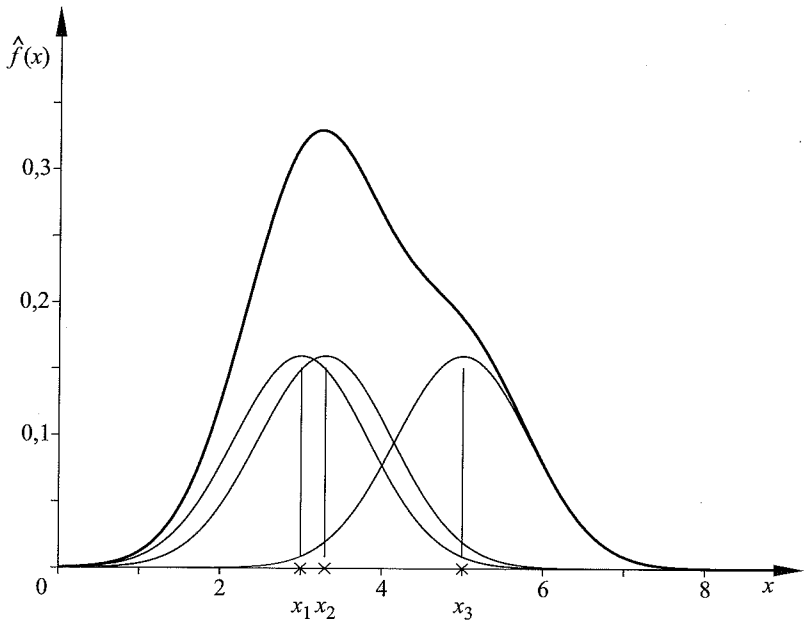
Korzystając z zależności określających jądro K (3.147) i parametr wygładzania (3.160), można zapisać podstawową postać estymatora jądrowego (3.3):

$$\hat{f}(x)=\frac{1}{3\cdot 0{,}811}\left[\frac{1}{\sqrt{2\pi}}\exp\left(-\frac{1}{2}\left(\frac{x-3}{0{,}811}\right)^2\right)+\frac{1}{\sqrt{2\pi}}\exp\left(-\frac{1}{2}\left(\frac{x-3{,}3}{0{,}811}\right)^2\right)+\right.$$
$$\left.+\frac{1}{\sqrt{2\pi}}\exp\left(-\frac{1}{2}\left(\frac{x-5}{0{,}811}\right)^2\right)\right] \tag{3.161}$$

i ostatecznie po wymnożeniu

$$\hat{f}(x)=0{,}160\left[\exp\left(-\frac{1}{2}\left(\frac{x-3}{0{,}811}\right)^2\right)+\exp\left(-\frac{1}{2}\left(\frac{x-3{,}3}{0{,}811}\right)^2\right)+\right.$$
$$\left.+\exp\left(-\frac{1}{2}\left(\frac{x-5}{0{,}811}\right)^2\right)\right]. \tag{3.162}$$

Ilustrację powyższego wyniku stanowi rysunek 3.11.



Rys. 3.11. Estymator jądrowy (3.162) – postać podstawowa

Procedura modyfikacji parametru wygładzania jest realizowana na podstawie algorytmu zdefiniowanego wzorami (3.112)-(3.115). Dysponując obliczoną powyżej podstawową postacią estymatora jądrowego (3.162), można wyznaczyć jego wartości dla poszczególnych elementów próby losowej (3.146):

$$\hat{f}(x_1) = \hat{f}(3) = 0{,}160\left[\exp\left(-\frac{1}{2}\left(\frac{3-3}{0{,}811}\right)^2\right) + \exp\left(-\frac{1}{2}\left(\frac{3-3{,}3}{0{,}811}\right)^2\right) + \right.$$
$$\left. + \exp\left(-\frac{1}{2}\left(\frac{3-5}{0{,}811}\right)^2\right)\right] = 0{,}317, \tag{3.163}$$

$$\hat{f}(x_2) = \hat{f}(3{,}3) = 0{,}160\left[\exp\left(-\frac{1}{2}\left(\frac{3{,}3-3}{0{,}811}\right)^2\right) + \exp\left(-\frac{1}{2}\left(\frac{3{,}3-3{,}3}{0{,}811}\right)^2\right) + \right.$$
$$\left. + \exp\left(-\frac{1}{2}\left(\frac{3{,}3-5}{0{,}811}\right)^2\right)\right] = 0{,}327, \tag{3.164}$$

$$\hat{f}(x_3) = \hat{f}(5) = 0{,}160\left[\exp\left(-\frac{1}{2}\left(\frac{5-3}{0{,}811}\right)^2\right) + \exp\left(-\frac{1}{2}\left(\frac{5-3{,}3}{0{,}811}\right)^2\right) + \right.$$
$$\left. + \exp\left(-\frac{1}{2}\left(\frac{5-5}{0{,}811}\right)^2\right)\right] = 0{,}185, \tag{3.165}$$

a – na podstawie zależności (3.113) – ich średnią geometryczną

$$\tilde{s} = \exp\left(\frac{1}{3}[\ln(0{,}317) + \ln(0{,}327) + \ln(0{,}185)]\right) = 0{,}268. \tag{3.166}$$

Wartości parametrów modyfikujących są obliczane ze wzoru (3.112) z podstawieniem równości (3.115) i wynoszą zatem

$$s_1 = \left(\frac{0{,}317}{0{,}268}\right)^{-1/2} = 0{,}919, \tag{3.167}$$

$$s_2 = \left(\frac{0{,}327}{0{,}268}\right)^{-1/2} = 0{,}905, \tag{3.168}$$

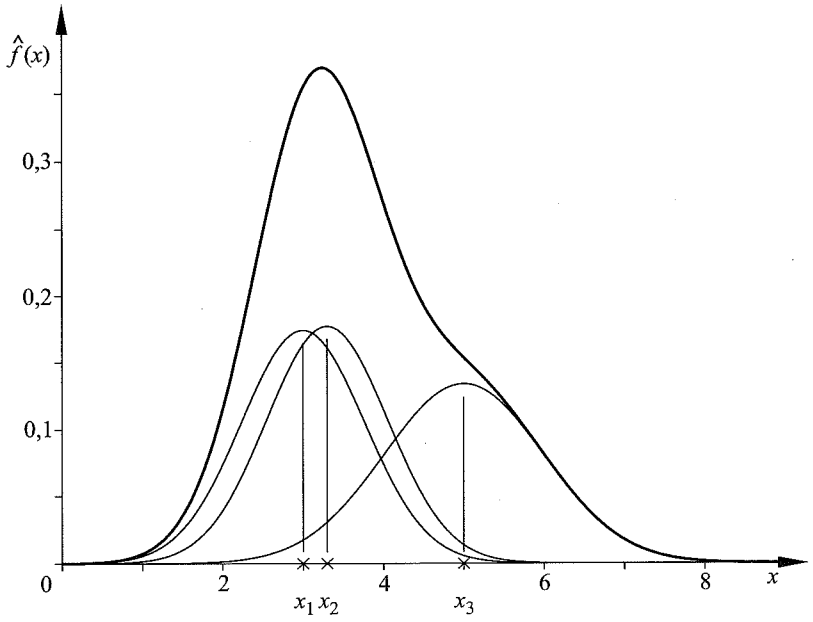
$$s_3 = \left(\frac{0{,}185}{0{,}268}\right)^{-1/2} = 1{,}204. \tag{3.169}$$

Estymator jądrowy z modyfikacją parametru wygładzania (3.114) jest więc określony jako

$$\hat{f}(x)=\frac{1}{3\cdot 0{,}811}\left[\frac{1}{0{,}919\sqrt{2\pi}}\exp\left(-\frac{1}{2}\left(\frac{x-3}{0{,}811\cdot 0{,}919}\right)^2\right)+\right.$$
$$+\frac{1}{0{,}905\sqrt{2\pi}}\exp\left(-\frac{1}{2}\left(\frac{x-3{,}3}{0{,}811\cdot 0{,}905}\right)^2\right)+$$
$$\left.+\frac{1}{1{,}204\sqrt{2\pi}}\exp\left(-\frac{1}{2}\left(\frac{x-5}{0{,}811\cdot 1{,}204}\right)^2\right)\right], \tag{3.170}$$

a po dokonaniu stosownych operacji arytmetycznych:

$$\hat{f}(x)=0{,}178\exp\left(-\frac{1}{2}\left(\frac{x-3}{0{,}745}\right)^2\right)+0{,}181\exp\left(-\frac{1}{2}\left(\frac{x-3{,}3}{0{,}734}\right)^2\right)+$$
$$+0{,}136\exp\left(-\frac{1}{2}\left(\frac{x-5}{0{,}976}\right)^2\right). \tag{3.171}$$



Rys. 3.12. Estymator jądrowy (3.171) – z modyfikacją parametru wygładzania

Powyższą postać należy traktować jako fundamentalną na potrzeby praktycznych zastosowań. Została ona zilustrowana na rysunku 3.12, w szczególności dla porównania z rysunkiem 3.11. Można bowiem zauważyć, iż jądra odpowiadające dwóm pierwszym elementom próby losowej zostały „wyszczuplone", co pozwala lepiej scharakteryzować badany rozkład w otoczeniu wartości modalnej, natomiast jądro odpowiadające nieco oddalonemu trzeciemu elementowi jest „spłaszczone", zgodnie ze specyfiką „ogonów" rozkładu probabilistycznego. Przy używanych w praktyce, istotnie liczniejszych próbach, tego typu efekty są znacznie intensywniejsze.

Przykład 3.2 – jednowymiarowa zmienna losowa z ograniczeniem nośnika

W niniejszym przykładzie będzie rozważana – podobnie jak w przykładzie 3.1 – jednowymiarowa zmienna losowa X i uzyskana z niej trójelementowa próba (3.146), aczkolwiek przy dodatkowym założeniu, iż nośnik tej zmiennej jest lewostronnie ograniczony wartością 1,5, a zatem jest zawarty w przedziale $[1{,}5;\ \infty)$.

Z uwagi na zbieżność uwarunkowań niniejszego zagadnienia z tematem przykładu 3.1 założenia i obliczenia prowadzące do konstrukcji estymatora jądrowego z modyfikacją parametru wygładzania (3.171) są identyczne. Wobec przyjętego teraz dodatkowo warunku ograniczenia nośnika do przedziału $[1{,}5;\ \infty)$ zostanie wprowadzone oznaczenie $x_* = 1{,}5$, po czym – w konsekwencji wzoru (3.116) – zależność

$$\chi_{[1,5;\infty)}(x) = \begin{cases} 1, & \text{gdy} \quad x \geq 1{,}5, \\ 0, & \text{gdy} \quad x < 1{,}5, \end{cases} \tag{3.172}$$

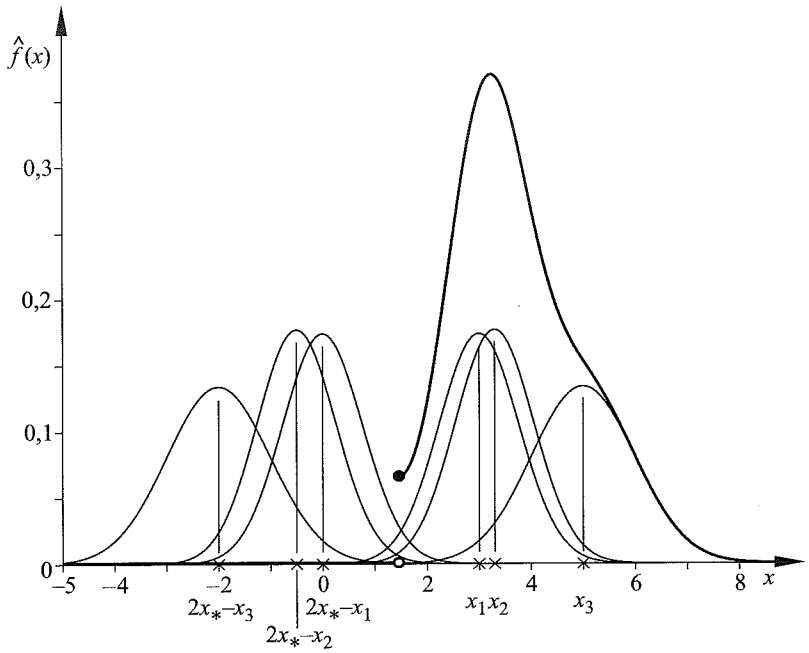
określająca funkcję charakterystyczną tego przedziału. Dysponując równością (3.171), po zastosowaniu wzoru (3.119) uzyskuje się ostatecznie następującą postać estymatora jądrowego z ograniczeniem nośnika:

$$\hat{f}(x) = \chi_{[1,5;\infty)}(x)\left[0{,}178 \exp\left(-\frac{1}{2}\left(\frac{x-3}{0{,}745} \right)^2 \right) + 0{,}178 \exp\left(-\frac{1}{2}\left(\frac{x}{0{,}745} \right)^2 \right) + \right.$$

$$+ 0{,}181 \exp\left(-\frac{1}{2}\left(\frac{x-3{,}3}{0{,}734} \right)^2 \right) + 0{,}181 \exp\left(-\frac{1}{2}\left(\frac{x+0{,}3}{0{,}734} \right)^2 \right) +$$

$$\left. + 0{,}136 \exp\left(-\frac{1}{2}\left(\frac{x-5}{0{,}976} \right)^2 \right) + 0{,}136 \exp\left(-\frac{1}{2}\left(\frac{x+2}{0{,}976} \right)^2 \right) \right]. \tag{3.173}$$

Jego wykres jest pokazany na rysunku 3.13.



Rys. 3.13. Estymator jądrowy (3.173) – z ograniczeniem nośnika

Przykład 3.3 – jednowymiarowa zmienna losowa ze składową binarną

W niniejszym przykładzie zostanie rozpatrzona – podobnie jak w przykładach 3.1 i 3.2 – jednowymiarowa zmienna losowa X ($n = 1$), lecz tym razem zestawiona z jednowymiarową zmienną binarną Y ($k = 1$). Niech ponadto dana będzie, uzyskana dla stanowiącej ich zestawienie dwuwymiarowej zmiennej losowej $Z \equiv \begin{bmatrix} X \\ Y \end{bmatrix}$, pięcioelementowa próba losowa ($m = 5$) o wartościach

$$z_1 = \begin{bmatrix} x_1 \\ y_1 \end{bmatrix} = \begin{bmatrix} 4 \\ 1 \end{bmatrix}, \quad z_2 = \begin{bmatrix} x_2 \\ y_2 \end{bmatrix} = \begin{bmatrix} 5 \\ 1 \end{bmatrix}, \quad z_3 = \begin{bmatrix} x_3 \\ y_3 \end{bmatrix} = \begin{bmatrix} 5{,}3 \\ 1 \end{bmatrix},$$

$$z_4 = \begin{bmatrix} x_4 \\ y_4 \end{bmatrix} = \begin{bmatrix} 7 \\ 0 \end{bmatrix}, \quad z_5 = \begin{bmatrix} x_5 \\ y_5 \end{bmatrix} = \begin{bmatrix} 8 \\ 1 \end{bmatrix}. \tag{3.174}$$

Najpierw należy wyznaczyć estymator jądrowy dla zmiennej losowej X, stanowiącej pierwszą współrzędną zmiennej Z. Powtarzając procedurę z przykładu 3.1 wobec próby losowej

$$x_1 = 4, \quad x_2 = 5, \quad x_3 = 5{,}3, \quad x_4 = 7, \quad x_5 = 8, \tag{3.175}$$

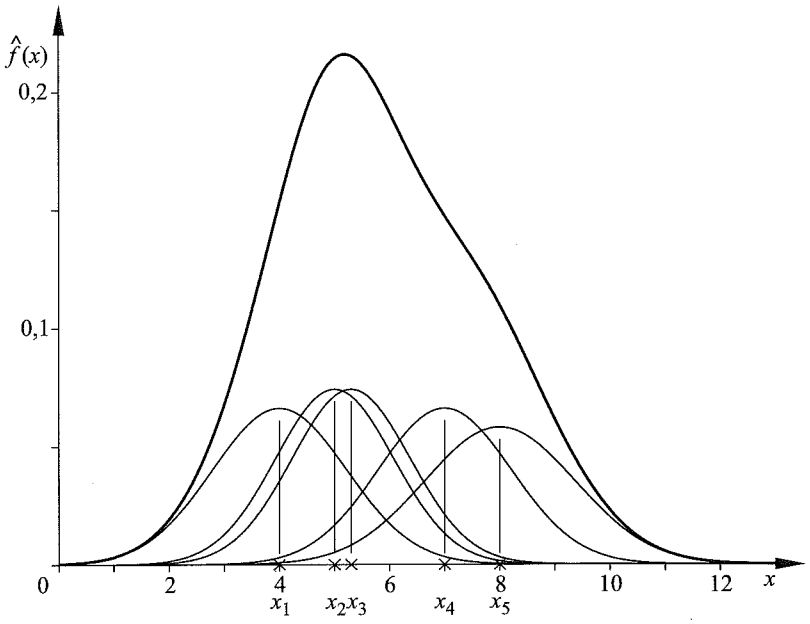
otrzymuje się następujące wartości parametrów: $h = 1{,}150$, $s_1 = 1{,}028$, $s_2 = 0{,}953$, $s_3 = 0{,}879$, $s_4 = 0{,}997$, $s_5 = 1{,}165$, i w rezultacie odpowiednik wzoru (3.171):

$$\hat{f}(x) = 0{,}067 \exp\left(-\frac{1}{2}\left(\frac{x-4}{1{,}182}\right)^2\right) + 0{,}073 \exp\left(-\frac{1}{2}\left(\frac{x-5}{1{,}096}\right)^2\right) +$$

$$+ 0{,}079 \exp\left(-\frac{1}{2}\left(\frac{x-5{,}3}{1{,}011}\right)^2\right) + 0{,}070 \exp\left(-\frac{1}{2}\left(\frac{x-7}{1{,}147}\right)^2\right) +$$

$$+ 0{,}060 \exp\left(-\frac{1}{2}\left(\frac{x-8}{1{,}340}\right)^2\right). \tag{3.176}$$

Wykres powyższej funkcji jest przedstawiony na rysunku 3.14.



Rys. 3.14. Estymator jądrowy (3.176) – bez uwzględnienia składowej binarnej

Zostanie teraz określony estymator jądrowy dla zmiennej binarnej Y, stanowiącej drugą współrzędną zmiennej losowej Z. Rozważana zatem będzie próba losowa

$$y_1 = 1, \quad y_2 = 1, \quad y_3 = 1, \quad y_4 = 0, \quad y_5 = 1. \tag{3.177}$$

Najpierw wyznacza się wartość binarnego parametru wygładzania δ. Poszczególne czynniki funkcji g zdefiniowanej równością (3.130) przyjmują tu – zgodnie ze wzorami (3.125)-(3.126) – postać

$$\hat{p}_{-1}(y_1)=\frac{1}{4}[\delta^{1-d(y_1,y_2)}(1-\delta)^{d(y_1,y_2)}+\delta^{1-d(y_1,y_3)}(1-\delta)^{d(y_1,y_3)}+$$
$$+\delta^{1-d(y_1,y_4)}(1-\delta)^{d(y_1,y_4)}+\delta^{1-d(y_1,y_5)}(1-\delta)^{d(y_1,y_5)}], \tag{3.178}$$

$$\hat{p}_{-2}(y_2)=\frac{1}{4}[\delta^{1-d(y_2,y_1)}(1-\delta)^{d(y_2,y_1)}+\delta^{1-d(y_2,y_3)}(1-\delta)^{d(y_2,y_3)}+$$
$$+\delta^{1-d(y_2,y_4)}(1-\delta)^{d(y_2,y_4)}+\delta^{1-d(y_2,y_5)}(1-\delta)^{d(y_2,y_5)}], \tag{3.179}$$

$$\hat{p}_{-3}(y_3)=\frac{1}{4}[\delta^{1-d(y_3,y_1)}(1-\delta)^{d(y_3,y_1)}+\delta^{1-d(y_3,y_2)}(1-\delta)^{d(y_3,y_2)}+$$
$$+\delta^{1-d(y_3,y_4)}(1-\delta)^{d(y_3,y_4)}+\delta^{1-d(y_3,y_5)}(1-\delta)^{d(y_3,y_5)}], \tag{3.180}$$

$$\hat{p}_{-4}(y_4)=\frac{1}{4}[\delta^{1-d(y_4,y_1)}(1-\delta)^{d(y_4,y_1)}+\delta^{1-d(y_4,y_2)}(1-\delta)^{d(y_4,y_2)}+$$
$$+\delta^{1-d(y_4,y_3)}(1-\delta)^{d(y_4,y_3)}+\delta^{1-d(y_4,y_5)}(1-\delta)^{d(y_4,y_5)}], \tag{3.181}$$

$$\hat{p}_{-5}(y_5)=\frac{1}{4}[\delta^{1-d(y_5,y_1)}(1-\delta)^{d(y_5,y_1)}+\delta^{1-d(y_5,y_2)}(1-\delta)^{d(y_5,y_2)}+$$
$$+\delta^{1-d(y_5,y_3)}(1-\delta)^{d(y_5,y_3)}+\delta^{1-d(y_5,y_4)}(1-\delta)^{d(y_5,y_4)}]. \tag{3.182}$$

Uwzględniając, iż na podstawie zależności (3.127):

$$d(y_1,y_2)=d(y_2,y_1)=0, \tag{3.183}$$
$$d(y_1,y_3)=d(y_3,y_1)=0, \tag{3.184}$$
$$d(y_1,y_4)=d(y_4,y_1)=1, \tag{3.185}$$
$$d(y_1,y_5)=d(y_5,y_1)=0, \tag{3.186}$$
$$d(y_2,y_3)=d(y_3,y_2)=0, \tag{3.187}$$
$$d(y_2,y_4)=d(y_4,y_2)=1, \tag{3.188}$$
$$d(y_2,y_5)=d(y_5,y_2)=0, \tag{3.189}$$
$$d(y_3,y_4)=d(y_4,y_3)=1, \tag{3.190}$$
$$d(y_3,y_5)=d(y_5,y_3)=0, \tag{3.191}$$
$$d(y_4,y_5)=d(y_5,y_4)=1, \tag{3.192}$$

otrzymuje się stąd

$$\hat{p}_{-1}(y_1) = \frac{1}{4}[(1-\delta)+\delta+\delta+\delta] = \frac{1}{4}[1+2\delta], \tag{3.193}$$

$$\hat{p}_{-2}(y_2) = \frac{1}{4}[(1-\delta)+\delta+\delta+\delta] = \frac{1}{4}[1+2\delta], \tag{3.194}$$

$$\hat{p}_{-3}(y_3) = \frac{1}{4}[(1-\delta)+\delta+\delta+\delta] = \frac{1}{4}[1+2\delta], \tag{3.195}$$

$$\hat{p}_{-4}(y_4) = \frac{1}{4}[(1-\delta)+(1-\delta)+(1-\delta)+(1-\delta)] = 1-\delta, \tag{3.196}$$

$$\hat{p}_{-5}(y_5) = \frac{1}{4}[(1-\delta)+\delta+\delta+\delta] = \frac{1}{4}[1+2\delta], \tag{3.197}$$

a zatem sama funkcja g jest ostatecznie dana wzorem

$$g(\delta) = -\frac{1}{256}(1+2\delta)^4(1-\delta). \tag{3.198}$$

Do wyznaczenia wartości, dla której przyjmuje ona minimum na przedziale [0,5; 1], zostanie użyty algorytm określony wzorami (3.131)-(3.132) i (3.88)-(3.103) oraz (3.133)-(3.134). Przyjmując zatem $a_0 = 0{,}5$ i $b_0 = 1$, otrzymuje się w kroku wstępnym $p_0 = 0{,}691$ i $q_0 = 0{,}809$ oraz $g(p_0) = -0{,}038858$ i $g(q_0) = -0{,}035049$. Zakładając dokładność $\Delta = 0{,}01$, dzięki równościom (3.136)-(3.138) wiadomo, że konieczne będzie wykonanie 6 iteracji. I tak

(1) dla $k = 1$: ponieważ $g(p_0) < g(q_0)$, więc na podstawie zależności (3.92)-(3.97) uzyskuje się

$$a_1 = a_0 = 0{,}5, \tag{3.199}$$

$$b_1 = q_0 = 0{,}809, \tag{3.200}$$

$$p_1 = b_1 - \frac{\sqrt{5}-1}{2}(b_1 - a_1) = 0{,}618, \tag{3.201}$$

$$q_1 = p_0 = 0{,}691, \tag{3.202}$$

$$g(p_1) = -\frac{1}{256}(1+2p_1)^4(1-p_1) = -0{,}037300, \tag{3.203}$$

$$g(q_1) = g(p_0) = -0{,}038858; \tag{3.204}$$

(2) dla $k = 2$: ponieważ $g(p_1) > g(q_1)$, więc na podstawie zależności (3.98)-(3.103) otrzymuje się

$$a_2 = p_1 = 0{,}618, \tag{3.205}$$

$$b_2 = b_1 = 0{,}809, \tag{3.206}$$

$$p_2 = q_1 = 0{,}691, \tag{3.207}$$

$$q_2 = a_2 + \frac{\sqrt{5}-1}{2}(b_2 - a_2) = 0{,}736, \tag{3.208}$$

$$g(p_2) = g(q_1) = -0{,}038858, \tag{3.209}$$

$$g(q_2) = -\frac{1}{256}(1+2q_2)^4(1-q_2) = -0{,}038509; \tag{3.210}$$

(3) dla $k = 3$: ponieważ $g(p_2) < g(q_2)$, więc na podstawie zależności (3.92)-(3.97) uzyskuje się

$$a_3 = a_2 = 0{,}618, \tag{3.211}$$

$$b_3 = q_2 = 0{,}736, \tag{3.212}$$

$$p_3 = b_3 - \frac{\sqrt{5}-1}{2}(b_3 - a_3) = 0{,}663, \tag{3.213}$$

$$q_3 = p_2 = 0{,}691, \tag{3.214}$$

$$g(p_3) = -\frac{1}{256}(1+2p_3)^4(1-p_3) = -0{,}038533, \tag{3.215}$$

$$g(q_3) = g(p_2) = -0{,}038856; \tag{3.216}$$

(4) dla $k = 4$: ponieważ $g(p_3) > g(q_3)$, więc na podstawie zależności (3.98)-(3.103) otrzymuje się

$$a_4 = p_3 = 0{,}663, \tag{3.217}$$

$$b_4 = b_3 = 0{,}736, \tag{3.218}$$

$$p_4 = q_3 = 0{,}691, \tag{3.219}$$

$$q_4 = a_4 + \frac{\sqrt{5}-1}{2}(b_4 - a_4) = 0{,}708, \tag{3.220}$$

$$g(p_4) = g(q_3) = -0{,}038856, \tag{3.221}$$

$$g(q_4) = -\frac{1}{256}(1+2q_4)^4(1-q_4) = -0{,}038862; \tag{3.222}$$

(5) dla $k = 5$: ponieważ $g(p_4) > g(q_4)$, więc ponownie na podstawie zależności (3.98)-(3.103) uzyskuje się

$$a_5 = p_4 = 0{,}691, \tag{3.223}$$

$$b_5 = b_4 = 0{,}736, \tag{3.224}$$

$$p_5 = q_4 = 0{,}708, \tag{3.225}$$

$$q_5 = a_5 + \frac{\sqrt{5}-1}{2}(b_5 - a_5) = 0{,}719, \tag{3.226}$$

$$g(p_5) = g(q_4) = -0{,}038862, \tag{3.227}$$

$$g(q_5) = -\frac{1}{256}(1+2q_5)^4(1-q_5) = -0{,}038778; \tag{3.228}$$

(6) dla $k = 6$: ponieważ $g(p_5) < g(q_5)$, więc na podstawie zależności (3.92)-(3.97) otrzymuje się

$$a_6 = a_5 = 0{,}691, \tag{3.229}$$

$$b_6 = q_5 = 0{,}719, \tag{3.230}$$

$$p_6 = b_6 - \frac{\sqrt{5}-1}{2}(b_6 - a_6) = 0{,}702, \tag{3.231}$$

$$q_6 = p_5 = 0{,}708, \tag{3.232}$$

$$g(p_6) = -\frac{1}{256}(1+2p_6)^4(1-p_6) = -0{,}038879, \tag{3.233}$$

$$g(q_6) = g(p_5) = -0{,}038778. \tag{3.234}$$

I ostatecznie, ponieważ $g(p_6) \leq g(q_6)$, więc korzystając z równości (3.133), uzyskuje się

$$\delta = \frac{a_6 + q_6}{2} = 0{,}700. \tag{3.235}$$

A teraz, uwzględniając, iż $y_1 = y_2 = y_3 = y_5 = 1$ i $y_4 = 0$ oraz że y może przyjąć jedynie wartość 0 albo 1, ze wzoru (3.126) otrzymuje się

$$L(y, y_1) = L(y, y_2) = L(y, y_3) = L(y, y_5) = 0{,}7^y\ 0{,}3^{1-y}, \tag{3.236}$$

$$L(y, y_4) = 0{,}7^{1-y}\ 0{,}3^y, \tag{3.237}$$

czyli na podstawie definicji (3.125) można stąd uzyskać estymator jądrowy dla zmiennej binarnej Y:

$$\hat{p}(y) = \frac{4}{5} 0{,}7^{y}\, 0{,}3^{1-y} + \frac{1}{5} 0{,}7^{1-y}\, 0{,}3^{y}. \tag{3.238}$$

I wreszcie, łącząc zależności (3.176) oraz (3.236)-(3.237), otrzymujemy ostatecznie na podstawie definicji (3.143) estymator jądrowy zmiennej losowej będącej zestawieniem zmiennych X i Y:

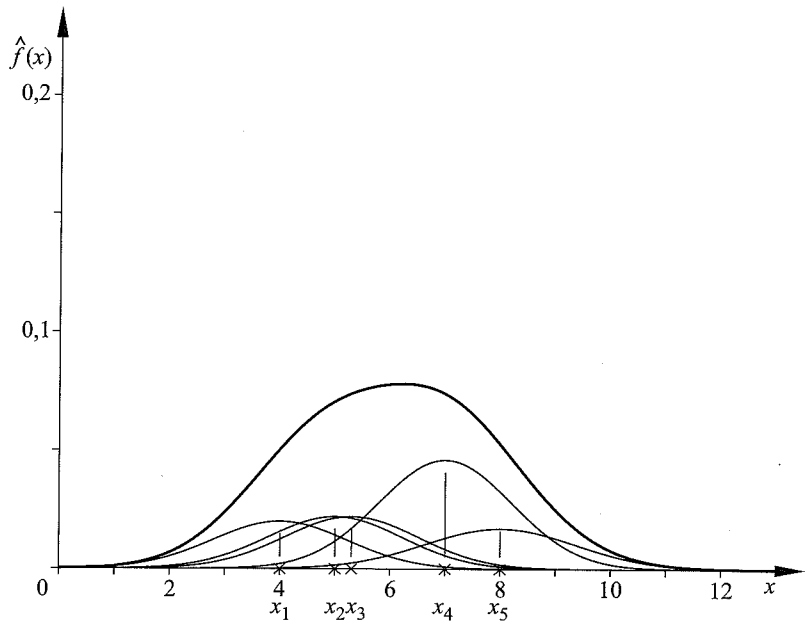
$$\begin{aligned}\hat{f}\left(\begin{bmatrix} x \\ y \end{bmatrix}\right) &= 0{,}067 \exp\left(-\frac{1}{2}\left(\frac{x-4}{1{,}182}\right)^2\right) 0{,}7^{y} 0{,}3^{1-y} + \\ &+ 0{,}073 \exp\left(-\frac{1}{2}\left(\frac{x-5}{1{,}096}\right)^2\right) 0{,}7^{y} 0{,}3^{1-y} + 0{,}079 \exp\left(-\frac{1}{2}\left(\frac{x-5{,}3}{1{,}011}\right)^2\right) 0{,}7^{y} 0{,}3^{1-y} + \\ &+ 0{,}070 \exp\left(-\frac{1}{2}\left(\frac{x-7}{1{,}147}\right)^2\right) 0{,}7^{1-y} 0{,}3^{y} + 0{,}060 \exp\left(-\frac{1}{2}\left(\frac{x-8}{1{,}340}\right)^2\right) 0{,}7^{y} 0{,}3^{1-y}.\end{aligned} \tag{3.239}$$

W szczególności, dla $y = 0$:

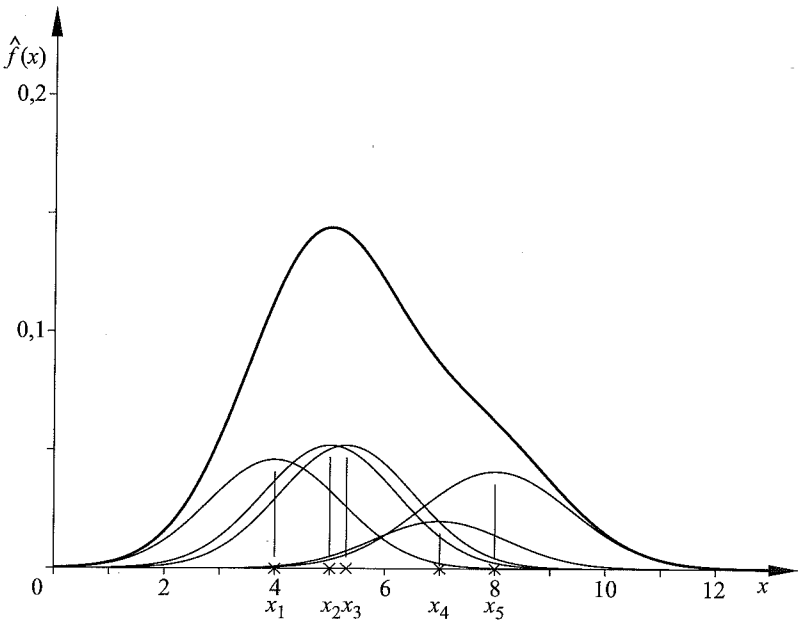
$$\begin{aligned}\hat{f}\left(\begin{bmatrix} x \\ 0 \end{bmatrix}\right) &= 0{,}047 \exp\left(-\frac{1}{2}\left(\frac{x-4}{1{,}182}\right)^2\right) + 0{,}051 \exp\left(-\frac{1}{2}\left(\frac{x-5}{1{,}096}\right)^2\right) + \\ &+ 0{,}055 \exp\left(-\frac{1}{2}\left(\frac{x-5{,}3}{1{,}011}\right)^2\right) + 0{,}021 \exp\left(-\frac{1}{2}\left(\frac{x-7}{1{,}147}\right)^2\right) + \\ &+ 0{,}042 \exp\left(-\frac{1}{2}\left(\frac{x-8}{1{,}340}\right)^2\right),\end{aligned} \tag{3.240}$$

natomiast dla $y = 1$:

$$\begin{aligned}\hat{f}\left(\begin{bmatrix} x \\ 1 \end{bmatrix}\right) &= 0{,}020 \exp\left(-\frac{1}{2}\left(\frac{x-4}{1{,}182}\right)^2\right) + 0{,}022 \exp\left(-\frac{1}{2}\left(\frac{x-5}{1{,}096}\right)^2\right) + \\ &+ 0{,}024 \exp\left(-\frac{1}{2}\left(\frac{x-5{,}3}{1{,}011}\right)^2\right) + 0{,}049 \exp\left(-\frac{1}{2}\left(\frac{x-7}{1{,}147}\right)^2\right) + \\ &+ 0{,}018 \exp\left(-\frac{1}{2}\left(\frac{x-8}{1{,}340}\right)^2\right).\end{aligned} \tag{3.241}$$



Rys. 3.15. Estymator jądrowy (3.240) – dla ustalonej wartości zmiennej binarnej $y = 0$



Rys. 3.16. Estymator jądrowy (3.241) – dla ustalonej wartości zmiennej binarnej $y = 1$

Funkcje (3.240)-(3.241) są pokazane kolejno na rysunkach 3.15 i 3.16. Wyniki te warto porównać z estymatorem (3.176) zilustrowanym na rysunku 3.14, otrzymanym dla zmiennej losowej X, czyli przy pominięciu informacji reprezentowanej przez zmienną binarną Y. Otóż, widać, iż w przypadku funkcji (3.240) z rysunku 3.15, uzyskanej przy ustalonej wartości zmiennej binarnej $y = 0$, „wzmocnieniu" uległo jądro czwarte – odpowiadające z_4, czyli temu elementowi próby losowej (3.174), dla którego druga współrzędna $y_4 = 0$. Podobnie jest w przypadku funkcji (3.241) z rysunku 3.16, otrzymanej przy ustalonej wartości $y = 1$: „wzmocnieniu" uległy jądra przyporządkowane pozostałym elementom próby (3.174), których drugie współrzędne $y_1 = y_2 = y_3 = y_5 = 1$. Tak więc dla ustalonej wartości zmiennej binarnej następuje zwiększenie znaczenia tych jąder, które odpowiadają owej ustalonej wartości, a wpływ pozostałych ulega zmniejszeniu, aczkolwiek nie całkowitej likwidacji.

Przykład 3.4 – wielowymiarowa zmienna losowa z użyciem jądra produktowego

W niniejszym przykładzie jest rozpatrywana dwuwymiarowa zmienna losowa X ($n = 2$), a do konstrukcji estymatora jądrowego zostanie zastosowane jądro produktowe (3.48). Niech więc dana będzie uzyskana dla tej zmiennej trójelementowa próba losowa ($m = 3$) o wartościach

$$x_1 = \begin{bmatrix} 3 \\ 6 \end{bmatrix}, \quad x_2 = \begin{bmatrix} 3{,}3 \\ 7 \end{bmatrix}, \quad x_3 = \begin{bmatrix} 5 \\ 4 \end{bmatrix}. \tag{3.242}$$

Rozważając próbę losową złożoną z pierwszych współrzędnych powyższej próby, czyli

$$x_1 = 3, \quad x_2 = 3{,}3, \quad x_3 = 5, \tag{3.243}$$

po przyjęciu jako $\mathscr{K}$ (jednowymiarowego) jądra normalnego (3.44), w przedstawionym wcześniej przykładzie 3.1 uzyskano wartość parametru wygładzania

$$h_1 = 0{,}811. \tag{3.244}$$

Postępując analogicznie dla próby losowej składającej się z drugich współrzędnych próby (3.242), to znaczy

$$x_1 = 6, \quad x_2 = 7, \quad x_3 = 4, \tag{3.245}$$

można otrzymać

$$h_2 = 1{,}412. \tag{3.246}$$

Uzyskany w ten sposób dla jądra produktowego (3.48) estymator jądrowy w podstawowej postaci (3.3) jest dany wzorem

$$\hat{f}(x)=\hat{f}\left(\begin{bmatrix}x_1\\x_2\end{bmatrix}\right)=\frac{1}{3\cdot 0{,}811\cdot 1{,}412}\left[\frac{1}{2\pi}\exp\left(-\frac{1}{2}\left(\frac{x_1-3}{0{,}811}\right)^2\right)\exp\left(-\frac{1}{2}\left(\frac{x_2-6}{1{,}412}\right)^2\right)+\right.$$

$$+\frac{1}{2\pi}\exp\left(-\frac{1}{2}\left(\frac{x_1-3{,}3}{0{,}811}\right)^2\right)\exp\left(-\frac{1}{2}\left(\frac{x_2-7}{1{,}412}\right)^2\right)+$$

$$\left.+\frac{1}{2\pi}\exp\left(-\frac{1}{2}\left(\frac{x_1-5}{0{,}811}\right)^2\right)\exp\left(-\frac{1}{2}\left(\frac{x_2-4}{1{,}412}\right)^2\right)\right], \tag{3.247}$$

czyli po dokonaniu stosownych działań algebraicznych

$$\hat{f}(x)=\hat{f}\left(\begin{bmatrix}x_1\\x_2\end{bmatrix}\right)=0{,}046\left[\exp\left(-\frac{1}{2}\left[\left(\frac{x_1-3}{0{,}811}\right)^2+\left(\frac{x_2-6}{1{,}412}\right)^2\right]\right)+\right.$$

$$+\exp\left(-\frac{1}{2}\left[\left(\frac{x_1-3{,}3}{0{,}811}\right)^2+\left(\frac{x_2-7}{1{,}412}\right)^2\right]\right)+$$

$$\left.+\exp\left(-\frac{1}{2}\left[\left(\frac{x_1-5}{0{,}811}\right)^2+\left(\frac{x_2-4}{1{,}412}\right)^2\right]\right)\right]. \tag{3.248}$$

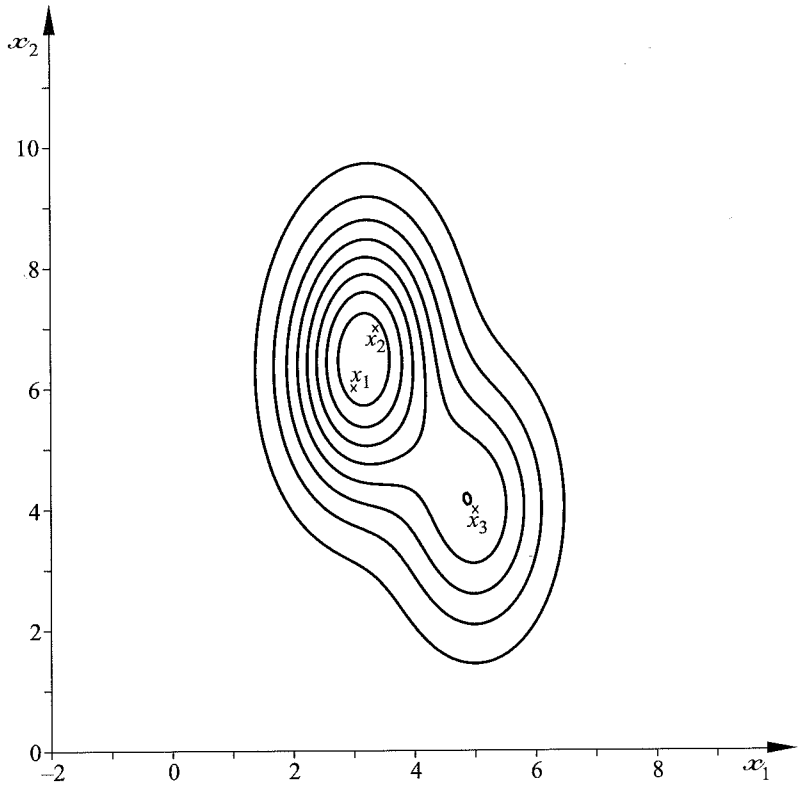
Estymator ten został zilustrowany na rysunku 3.17.

Procedura modyfikacji parametru wygładzania jest realizowana na podstawie algorytmu (3.112)-(3.115). Dysponując obliczoną powyżej podstawową postacią estymatora jądrowego (3.248), można zatem wyznaczyć jego wartości dla poszczególnych elementów próby losowej (3.242):

$$\hat{f}(x_1)=\hat{f}\left(\begin{bmatrix}3\\6\end{bmatrix}\right)=0{,}046\left[\exp\left(-\frac{1}{2}\left[\left(\frac{3-3}{0{,}811}\right)^2+\left(\frac{6-6}{1{,}412}\right)^2\right]\right)+\right.$$

$$+\exp\left(-\frac{1}{2}\left[\left(\frac{3-3{,}3}{0{,}811}\right)^2+\left(\frac{6-7}{1{,}412}\right)^2\right]\right)+$$

$$\left.+\exp\left(-\frac{1}{2}\left[\left(\frac{3-5}{0{,}811}\right)^2+\left(\frac{6-4}{1{,}412}\right)^2\right]\right)\right]=0{,}081, \tag{3.249}$$



Rys. 3.17. Poziomice estymatora jądrowego (3.248) – z użyciem jądra produktowego, postać podstawowa

$$\hat{f}(x_2) = \hat{f}\left(\begin{bmatrix} 3{,}3 \\ 7 \end{bmatrix}\right) = 0{,}046\left[\exp\left(-\frac{1}{2}\left[\left(\frac{3{,}3-3}{0{,}811}\right)^2 + \left(\frac{7-6}{1{,}412}\right)^2\right]\right) + \right.$$

$$+\exp\left(-\frac{1}{2}\left[\left(\frac{3{,}3-3{,}3}{0{,}811}\right)^2 + \left(\frac{7-7}{1{,}412}\right)^2\right]\right) +$$

$$\left. +\exp\left(-\frac{1}{2}\left[\left(\frac{3{,}3-5}{0{,}811}\right)^2 + \left(\frac{7-4}{1{,}412}\right)^2\right]\right)\right] = 0{,}080, \qquad (3.250)$$

$$\hat{f}(x_3)=\hat{f}\left(\begin{bmatrix}5\\4\end{bmatrix}\right)=0{,}046\left[\exp\left(-\frac{1}{2}\left[\left(\frac{5-3}{0{,}811}\right)^2+\left(\frac{4-6}{1{,}412}\right)^2\right]\right)+\right.$$
$$+\exp\left(-\frac{1}{2}\left[\left(\frac{5-3{,}3}{0{,}811}\right)^2+\left(\frac{4-7}{1{,}412}\right)^2\right]\right)+$$
$$\left.+\exp\left(-\frac{1}{2}\left[\left(\frac{5-5}{0{,}811}\right)^2+\left(\frac{4-4}{1{,}412}\right)^2\right]\right)\right]=0{,}047, \tag{3.251}$$

ich średnią geometryczną

$$\tilde{s}=\exp\left(\frac{1}{3}[\ln(0{,}081)+\ln(0{,}080)+\ln(0{,}047)]\right)=0{,}067 \tag{3.252}$$

i wreszcie wartości parametrów modyfikujących

$$s_1=\left(\frac{0{,}081}{0{,}067}\right)^{-1/2}=0{,}896, \tag{3.253}$$

$$s_2=\left(\frac{0{,}080}{0{,}067}\right)^{-1/2}=0{,}915, \tag{3.254}$$

$$s_3=\left(\frac{0{,}047}{0{,}067}\right)^{-1/2}=1{,}194. \tag{3.255}$$

Estymator jądrowy z modyfikacją parametru wygładzania jest więc określony jako

$$\hat{f}(x)=\hat{f}\left(\begin{bmatrix}x_1\\x_2\end{bmatrix}\right)=$$
$$=0{,}046\left[\frac{1}{0{,}896^2}\exp\left(-\frac{1}{2}\left[\left(\frac{x_1-3}{0{,}811\cdot 0{,}896}\right)^2+\left(\frac{x_2-6}{1{,}412\cdot 0{,}896}\right)^2\right]\right)+\right.$$
$$+\frac{1}{0{,}915^2}\exp\left(-\frac{1}{2}\left[\left(\frac{x_1-3{,}3}{0{,}811\cdot 0{,}915}\right)^2+\left(\frac{x_2-7}{1{,}412\cdot 0{,}915}\right)^2\right]\right)+$$
$$\left.+\frac{1}{1{,}194^2}\exp\left(-\frac{1}{2}\left[\left(\frac{x_1-5}{0{,}811\cdot 1{,}194}\right)^2+\left(\frac{x_2-4}{1{,}412\cdot 1{,}194}\right)^2\right]\right)\right] \tag{3.256}$$

i ostatecznie po wykonaniu stosownych działań:

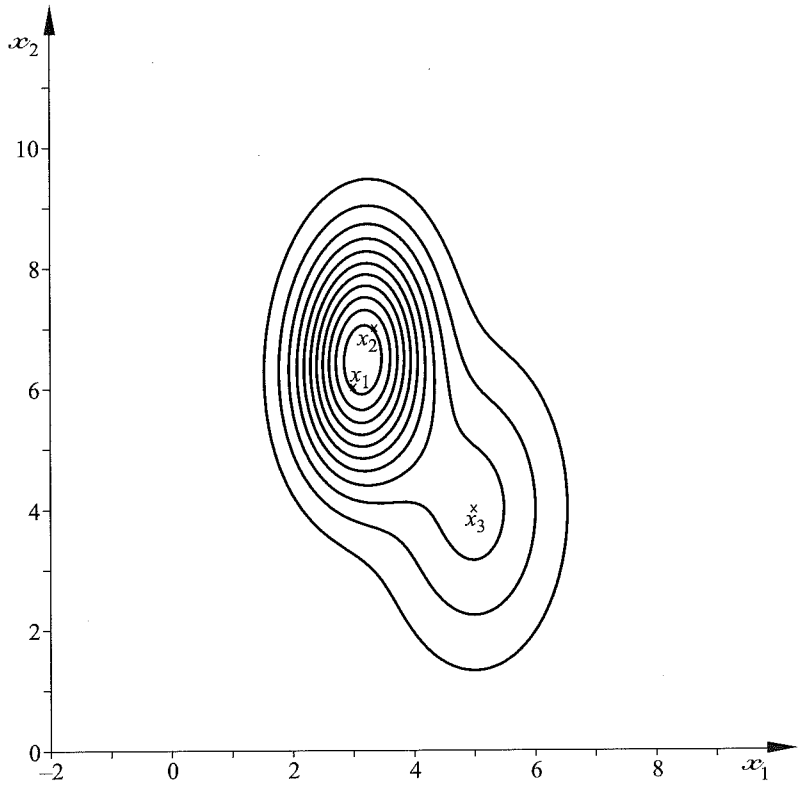
$$\hat{f}(x) = \hat{f}\left(\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}\right) = 0{,}057 \exp\left(-\frac{1}{2}\left[\left(\frac{x_1 - 3}{0{,}727}\right)^2 + \left(\frac{x_2 - 6}{1{,}265}\right)^2\right]\right) +$$

$$+ 0{,}055 \exp\left(-\frac{1}{2}\left[\left(\frac{x_1 - 3{,}3}{0{,}742}\right)^2 + \left(\frac{x_2 - 7}{1{,}292}\right)^2\right]\right) +$$

$$+ 0{,}032 \exp\left(-\frac{1}{2}\left[\left(\frac{x_1 - 5}{0{,}968}\right)^2 + \left(\frac{x_2 - 4}{1{,}686}\right)^2\right]\right)\Bigg]. \qquad (3.257)$$

Ilustrację powyższego wyniku stanowi rysunek 3.18.

Porównując rysunki 3.17 i 3.18, ponownie można zauważyć pozytywny wynik działania procedury modyfikacji parametru wygładzania. Na drugim z nich



Rys. 3.18. Poziomice estymatora jądrowego (3.257) – z użyciem jądra produktowego, z modyfikacją parametru wygładzania

poziomice wokół trzeciego, oddalonego elementu próby losowej są wyraźnie rzadsze, co świadczy o „wygładzeniu" tego fragmentu estymatora. Przeciwnie, poziomice w otoczeniu wartości modalnej generowanej przez dwa pierwsze elementy próby losowej są zdecydowanie gęstsze, co dowodzi dodatkowego „wyszczuplenia" tej jego części.

Przykład 3.5 – wielowymiarowa zmienna losowa z użyciem jądra radialnego

W niniejszym przykładzie będzie rozpatrywana – podobnie jak poprzednio – dwuwymiarowa zmienna losowa X i uzyskana z niej próba (3.242), przy czym do konstrukcji estymatora jądrowego zostanie tutaj zastosowane jądro radialne z uwzględnieniem transformacji liniowej (3.50). Jak uprzednio, jądro $\mathscr{K}$ przyjęto w postaci (jednowymiarowego) jądra normalnego (3.44).

Poniższe obliczenia są przedstawione z zastosowaniem pełnej postaci macierzy transformacji liniowej (3.53) – przypadek postaci diagonalnej (3.52) będzie skomentowany na końcu przykładu. A zatem najpierw zostanie wyznaczona macierz transformacji liniowej. Uwzględniając zależność (2.30) z podstawieniem (2.25), (2.27) i (2.29) oraz stosując odpowiednio wzory estymatorów (2.52) i (2.56), można kolejno otrzymać elementy estymatora macierzy kowariancji:

$$\hat{cov}_{1,1} = \hat{V}_1 = \frac{1}{2}(3^2 + 3{,}3^2 + 5^2) - \frac{1}{3 \cdot 2}(3 + 3{,}3 + 5)^2 = 1{,}163, \tag{3.258}$$

$$\hat{cov}_{1,2} = \hat{cov}_{2,1} = \frac{1}{2}(3 \cdot 6 + 3{,}3 \cdot 7 + 5 \cdot 4) - \frac{1}{3 \cdot 2}(3 + 3{,}3 + 5)(6 + 7 + 4) =$$

$$= -1{,}467, \tag{3.259}$$

$$\hat{cov}_{2,2} = \hat{V}_2 = \frac{1}{2}(6^2 + 7^2 + 4^2) - \frac{1}{3 \cdot 2}(6 + 7 + 4)^2 = 2{,}333. \tag{3.260}$$

Pełna postać macierzy transformacji liniowej (3.58) jest zatem dana wzorem

$$R = \begin{bmatrix} 1{,}163 & -1{,}467 \\ -1{,}467 & 2{,}333 \end{bmatrix}; \tag{3.261}$$

łatwo stąd otrzymać

$$\det(R) = 0{,}561, \tag{3.262}$$

$$R^{-1} = \begin{bmatrix} 4{,}157 & 2{,}614 \\ 2{,}614 & 2{,}072 \end{bmatrix}. \tag{3.263}$$

Korzystając z zależności (3.50) i dodatkowo odczytując z rysunku 3.4 dla $n = 2$ stałą $c_{n,\mathscr{K}}$, można napisać równanie radialnego jądra normalnego:

$$K(x)=\frac{1}{\sqrt{0{,}561}\ 2\pi}\exp\left(-\frac{1}{2}x^{\mathrm{T}}\begin{bmatrix}4{,}157 & 2{,}614\\ 2{,}614 & 2{,}072\end{bmatrix}x\right). \tag{3.264}$$

Wartość parametru wygładzania estymatora jądrowego zostanie wyznaczona z wykorzystaniem metody krzyżowego uwiarygodnienia. Zgodnie z zależnością (3.87) kwadrat splotowy powyższego jądra jest dany wzorem

$$K^{*2}(x)=\frac{1}{\sqrt{0{,}561}\ 4\pi}\exp\left(-\frac{1}{4}x^{\mathrm{T}}\begin{bmatrix}4{,}157 & 2{,}614\\ 2{,}614 & 2{,}072\end{bmatrix}x\right). \tag{3.265}$$

Po wykonaniu stosownych działań zależności (3.264)-(3.265) przyjmują odpowiednio postać

$$K(x)=K\left(\begin{bmatrix}x_1\\ x_2\end{bmatrix}\right)=0{,}212\exp(-2{,}078.x_1^2+2{,}614.x_1.x_2-1{,}036.x_2^2), \tag{3.266}$$

$$K^{*2}(x)=K\left(\begin{bmatrix}x_1\\ x_2\end{bmatrix}\right)=0{,}106\exp(-1{,}039.x_1^2+1{,}307.x_1.x_2-0{,}518.x_2^2). \tag{3.267}$$

Funkcja g określona wzorami (3.79)-(3.81) jest zatem dana następującą równością:

$$g(h)=\frac{1}{h^2}\left[\frac{1}{3^2}\sum_{i=1}^{3}\sum_{j=1}^{3}\left(K^{*2}\left(\frac{x_j-x_i}{h}\right)-2K\left(\frac{x_j-x_i}{h}\right)\right)+\frac{2}{3}K(0)\right], \tag{3.268}$$

czyli uwzględniając zależności (3.266)-(3.267) i wartości próby losowej (3.242), otrzymuje się

$$\begin{aligned}g(h)=\frac{1}{h^2}\Bigg[&0{,}012\exp(0)-2\cdot 0{,}024\exp(0)+\\&+0{,}012\exp\left(\frac{-1{,}039(3{,}3-3)^2+1{,}307(3{,}3-3)(7-6)-0{,}518(7-6)^2}{h^2}\right)-\\&-2\cdot 0{,}024\exp\left(\frac{-2{,}078(3{,}3-3)^2+2{,}614(3{,}3-3)(7-6)-1{,}036(7-6)^2}{h^2}\right)+\\&+0{,}012\exp\left(\frac{-1{,}039(5-3)^2+1{,}307(5-3)(4-6)-0{,}518(4-6)^2}{h^2}\right)-\\&-2\cdot 0{,}024\exp\left(\frac{-2{,}078(5-3)^2+2{,}614(5-3)(4-6)-1{,}036(4-6)^2}{h^2}\right)+\end{aligned}$$

cdn.

cd.

$$+0{,}012\exp\left(\frac{-1{,}039(3-3{,}3)^2+1{,}307(3-3{,}3)(6-7)-0{,}518(6-7)^2}{h^2}\right)-$$

$$-2\cdot 0{,}024\exp\left(\frac{-2{,}078(3-3{,}3)^2+2{,}614(3-3{,}3)(6-7)-1{,}036(6-7)^2}{h^2}\right)+$$

$$+0{,}012\exp(0)-2\cdot 0{,}024\exp(0)+$$

$$+0{,}012\exp\left(\frac{-1{,}039(5-3{,}3)^2+1{,}307(5-3{,}3)(4-7)-0{,}518(4-7)^2}{h^2}\right)-$$

$$-2\cdot 0{,}024\exp\left(\frac{-2{,}078(5-3{,}3)^2+2{,}614(5-3{,}3)(4-7)-1{,}036(4-7)^2}{h^2}\right)+$$

$$+0{,}012\exp\left(\frac{-1{,}039(3-5)^2+1{,}307(3-5)(6-4)-0{,}518(6-4)^2}{h^2}\right)-$$

$$-2\cdot 0{,}024\exp\left(\frac{-2{,}078(3-5)^2+2{,}614(3-5)(6-4)-1{,}036(6-4)^2}{h^2}\right)+$$

$$+0{,}012\exp\left(\frac{-1{,}039(3{,}3-5)^2+1{,}307(3{,}3-5)(7-6)-0{,}518(7-4)^2}{h^2}\right)-$$

$$-2\cdot 0{,}024\exp\left(\frac{-2{,}078(3{,}3-5)^2+2{,}614(3{,}3-5)(7-4)-1{,}036(7-4)^2}{h^2}\right)+$$

$$+0{,}012\exp(0)-2\cdot 0{,}024\exp(0)+0{,}141\Bigg], \tag{3.269}$$

a więc ostatecznie po wymnożeniu i pogrupowaniu składników:

$$g(h)=\frac{1}{h^2}\Bigg[0{,}033+0{,}012\exp\left(\frac{-0{,}219}{h^2}\right)-0{,}096\exp\left(\frac{-0{,}439}{h^2}\right)+$$

$$+0{,}012\exp\left(\frac{-16{,}684}{h^2}\right)-0{,}096\exp\left(\frac{-33{,}368}{h^2}\right)+$$

$$+0{,}012\exp\left(\frac{-16{,}684}{h^2}\right)-0{,}096\exp\left(\frac{-19{,}773}{h^2}\right)\Bigg]. \tag{3.270}$$

Minimum powyższej funkcji będzie poszukiwane na zbiorze określonym zależnością (3.108). W tym celu należy najpierw obliczyć wartość parametru wy-

gładzania za pomocą metody przybliżonej. I tak, podstawiając $n = 2$ do wzorów (3.60) i (3.62), otrzymuje się, odpowiednio:

$$Z(f) = \frac{2 \cdot 4}{2^4 \pi} = 0{,}159, \tag{3.271}$$

$$\frac{W(K)}{U(K)^2} = \frac{1}{2^2 \pi 2^2} = 0{,}020, \tag{3.272}$$

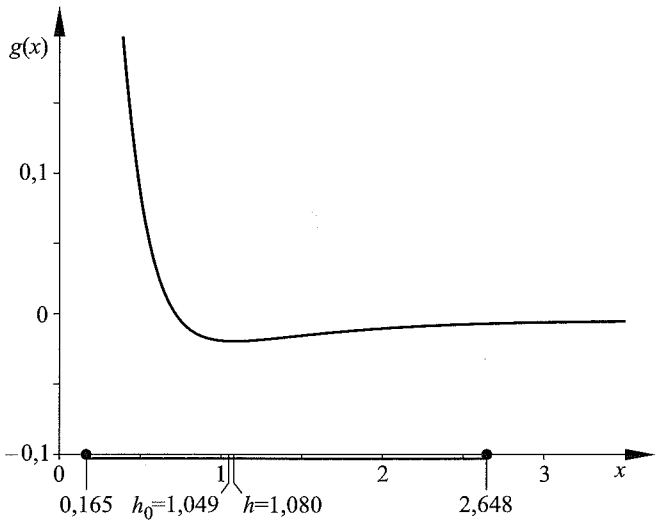
a stąd na podstawie równości (3.38):

$$h_0 = \left(\frac{2 \cdot 0{,}020}{0{,}159 \cdot 3} \right)^{1/6} = 0{,}662. \tag{3.273}$$

Zgodnie z zależnością (3.108) jako zakres poszukiwań należy zatem przyjąć przedział

$$\left[\frac{h_0}{4}, 4h_0 \right] = [0{,}165; 2{,}648]. \tag{3.274}$$

Do wyznaczenia minimum funkcji g warto użyć algorytmu opartego na metodzie złotego podziału (3.88)-(3.105), analogicznie jak było to pokazane w przykładzie 3.3. Jednak gdy w konkretnym zadaniu aplikacyjnym nie ma istotnych ograniczeń dotyczących czasu obliczeń, można wyznaczyć owe minimum, uzy-



Rys. 3.19. Funkcja g określona wzorem (3.270) wraz z wielkościami (3.273)-(3.275)

skując kolejno wartości tej funkcji w całym przedziale (3.274) z krokiem równym wymaganej dokładności, ewentualnie nawet w $[h_0/100, 4h_0]$ analizując jej wykres dla uniknięcia potencjalnego „załamania" się funkcji g w pobliżu zera. Na rysunku 3.19 jest pokazany otrzymany w ten sposób wykres funkcji (3.270) na przedziale (3.274). Minimum wyznaczone z dokładnością 0,001 występuje dla

$$h = 1{,}080. \tag{3.275}$$

Z uwagi na wartości próby losowej (3.242) oraz uzyskaną postać jądra (3.266) podstawowa postać estymatora jądrowego (3.3) wyraża się zatem wzorem

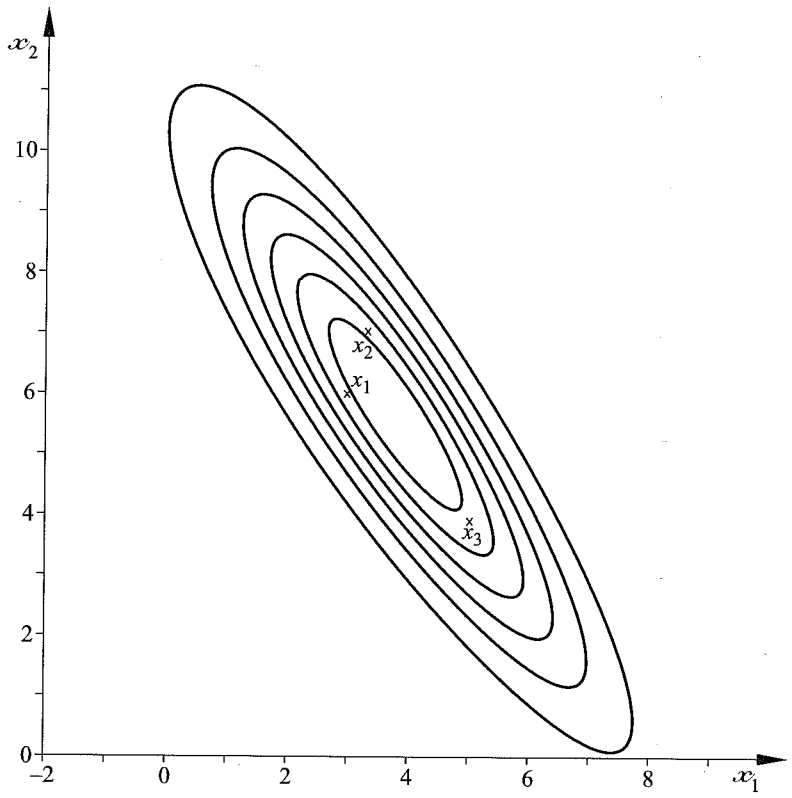
$$\begin{aligned}\hat{f}(x) = \hat{f}\left(\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}\right) = \frac{1}{3 \cdot 1{,}080^2} \cdot \\ \cdot \left[0{,}212 \exp\left(-2{,}078 \left(\frac{x_1 - 3}{1{,}080} \right)^2 - 2{,}614 \left(\frac{x_1 - 3}{1{,}080} \right)\left(\frac{x_2 - 6}{1{,}080} \right) - 1{,}036 \left(\frac{x_2 - 6}{1{,}080} \right)^2 \right) + \right. \\ + 0{,}212 \exp\left(-2{,}078 \left(\frac{x_1 - 3{,}3}{1{,}080} \right)^2 - 2{,}614 \left(\frac{x_1 - 3{,}3}{1{,}080} \right)\left(\frac{x_2 - 7}{1{,}080} \right) - 1{,}036 \left(\frac{x_2 - 7}{1{,}080} \right)^2 \right) + \\ \left. + 0{,}212 \exp\left(-2{,}078 \left(\frac{x_1 - 5}{1{,}080} \right)^2 - 2{,}614 \left(\frac{x_1 - 5}{1{,}080} \right)\left(\frac{x_2 - 4}{1{,}080} \right) - 1{,}036 \left(\frac{x_2 - 4}{1{,}080} \right)^2 \right) \right]\end{aligned} \tag{3.276}$$

i ostatecznie

$$\begin{aligned}\hat{f}(x) = \hat{f}\left(\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}\right) = \\ = 0{,}056 \left[\exp\left(-\frac{1}{2} \left[\left(\frac{x_1 - 3}{0{,}530} \right)^2 + 0{,}892 \left(\frac{x_1 - 3}{0{,}530} \right)\left(\frac{x_2 - 6}{0{,}751} \right) + \left(\frac{x_2 - 6}{0{,}751} \right)^2 \right] \right) + \right. \\ + \exp\left(-\frac{1}{2} \left[\left(\frac{x_1 - 3{,}3}{0{,}530} \right)^2 + 0{,}892 \left(\frac{x_1 - 3{,}3}{0{,}530} \right)\left(\frac{x_2 - 7}{0{,}751} \right) + \left(\frac{x_2 - 7}{0{,}751} \right)^2 \right] \right) + \\ \left. + \exp\left(-\frac{1}{2} \left[\left(\frac{x_1 - 5}{0{,}530} \right)^2 + 0{,}892 \left(\frac{x_1 - 5}{0{,}530} \right)\left(\frac{x_2 - 4}{0{,}751} \right) + \left(\frac{x_2 - 4}{0{,}751} \right)^2 \right] \right) \right].\end{aligned} \tag{3.277}$$

Powyższy estymator został zilustrowany na rysunku 3.20. Warto również porównać uzyskaną postać estymatora z zależnością (3.248) stanowiącą jej analo-

gię dla jądra produktowego. W przypadku jądra radialnego występujące powyżej stałe 0,530 i 0,751, będące odpowiednikami parametrów wygładzania 0,811 i 1,412, są mniejsze ze względu na dodatkowy wpływ pojawiającego się we wzorze (3.277) wyrazu mieszanego $x_1 x_2$.



Rys. 3.20. Poziomice estymatora jądrowego (3.277) – z użyciem jądra radialnego

Procedura modyfikacji parametru wygładzania może być zrealizowana w przypadku jądra radialnego identycznie, jak było to przedstawione w poprzednim przykładzie 3.4 dla jądra produktowego z wykorzystaniem zależności (3.249)-(3.257).

Należy również zaznaczyć, że gdyby we wzorze (3.259) przyjąć $\hat{cov}_{1,2} = \hat{cov}_{2,1} = 0$, to macierz transformacji liniowej przyjęłaby postać diagonalną (3.52), a dalsza część procedury byłaby analogiczna do przedstawionej powyżej w niniejszym przykładzie dla postaci pełnej. Jeśli jednak stosowane jest jądro normalne, to postępowanie takie nie ma praktycznego uzasadnienia. Dzięki własności funkcji eksponencjalnej $e^{x^* + x^{**}} = e^{x^*} e^{x^{**}}$ jądro radialne z diagonalną postacią macierzy transformacji ma bowiem postać taką samą jak jądro produkto-

we, a ponieważ procedura konstrukcji estymatora z jądrem produktowym jest znacznie dogodniejsza, zatem użycie jądra radialnego byłoby w tym przypadku bezcelowe.

Interpretując uzyskaną postać estymatora (3.277), a zwłaszcza ilustrację jego wykresu zawartą na rysunku 3.20, można zauważyć właściwe dopasowanie kierunku „rozciągnięcia" poziomic do kierunku „rozciągnięcia" elementów próby losowej. Zostało to uzyskane w wyniku użycia pełnej postaci macierzy transformacji liniowej. Co więcej, wszystkie elementy próby losowej leżą na tej samej poziomicy. Otrzymana regularność uwiarygodnia i stanowi cenną ilustrację poprawności metodycznej przedstawionego algorytmu. Sytuacja taka może wystąpić jedynie przy próbie trójelementowej, gdyż pełna postać macierzy transformacji liniowej ma właśnie trzy stopnie swobody – trzy niezależne parametry[1)]. Jednakże na skutek ulokowania wszystkich elementów próby losowej na jednej poziomicy konieczne stało się pominięcie procedury modyfikacji parametru wygładzania, gdyż jej naturalnym rezultatem byłby w tym przypadku trywialny wynik $s_1 = s_2 = s_3 = 1$. Nie zmniejsza to jednak praktycznie użyteczności powyższego przykładu, gdyż w przypadku jądra radialnego procedura modyfikacji byłaby identyczna z przedstawioną w przykładzie 3.4 dla jądra produktowego. Przykład 3.5 został zatem rozważony z użyciem próby trójelementowej nie tylko w celu uproszczenia wieloczłonowych w tym przypadku wzorów, ale przede wszystkim w celu zilustrowania wspomnianej powyżej, obrazowej regularności.

Uwagi bibliograficzne

Fumdamentalnymi podręcznikami dotyczącymi statystycznych estymatorów jądrowych, w których można znaleźć przedstawione w tym podrozdziale podstawowe definicje, wzory i algorytmy z klasycznego zakresu tego materiału, są pionierska publikacja [62] oraz praca [68]. Druga z nich jest lepiej uporządkowana i zawiera wyniki nowszych badań, ale pierwsza wydaje się bardziej przydatna dla praktyków polegających w większym stopniu na intuicji. Z kolei pozycje [10, 50, 54, 59, 60] mogą stanowić ich cenne uzupełnienie. W literaturze polskojęzycznej występują jedynie krótkie notki w nielicznych podręcznikach, na przykład [13, 20, 36]. Obecnie w czasopismach naukowych ukazuje się znaczna ilość artykułów, zwłaszcza z zakresu zastosowań statystyki, przedstawiających nowe, aczkolwiek często fragmentaryczne i niesprawdzone jeszcze koncepcje.

[1)] Stanowi to obrazową analogię do klasycznej własności z zakresu interpolacji wielomianowej. Otóż, przy dowolnie zadanych n wartościach funkcji, możliwe jest „przeprowadzenie" przez nie wielomianu stopnia $n-1$, czyli – uwzględniając wyraz wolny – mającego n współczynników. W szczególności przez trzy punkty możliwe jest „przeprowadzenie" trójmianu kwadratowego $ax^2 + bx + c$, a zatem wielomianu o trzech współczynnikach, czyli trzech stopniach swobody.

Informacje z zakresu wzmiankowanych w niniejszym podrozdziale aspektów matematycznych i obliczeniowych można uzupełnić, korzystając z licznych monografii przedmiotowych. I tak, z zakresu rachunku całkowego warto skorzystać z klasycznego podręcznika [16], posiłkując się również tablicami [52, 58]. Podstawowe algorytmy numeryczne są prezentowane w pracach [9, 57, 63] i dodatkowo z dziedziny zagadnień algebry liniowej w książce [30]. Procedurom znajdowania minimum funkcji zostały poświęcone publikacje [17, 61].

Wiele przedstawionych w niniejszym podrozdziale aspektów stanowi przedmiot szczegółowych badań prowadzonych w ramach prac [8, 46, 55, 66].

3.2. Jądrowe estymatory innych charakterystyk probabilistycznych

Rozważana w poprzednim podrozdziale gęstość prawdopodobieństwa dostarcza użytecznego i uniwersalnego opisu badanego zjawiska losowego. Jednak poza bezpośrednim wykorzystaniem do badań statystycznych, analizy danych lub wizualizacji wyników jej estymator może być także podstawą do wyznaczenia innych charakterystyk rozkładu prawdopodobieństwa, zarówno funkcyjnych, jak i jego parametrów, z dokładnością często przewyższającą metody klasyczne. Jako przykład pierwszego z powyższych zagadnień zostanie zaprezentowane zadanie estymacji dystrybuanty (punkt 3.2.1), a jako przykład drugiego – estymacja kwantyla (punkt 3.2.2). W kolejnej części niniejszego podrozdziału (punkt 3.2.3) będzie wprowadzona koncepcja jądrowego estymatora charakterystyk warunkowych, uściślona wobec uprzednio rozpatrywanych: gęstości prawdopodobieństwa, dystrybuanty i kwantyla.

3.2.1. Jądrowy estymator dystrybuanty

Najpierw będzie rozważany przypadek jednowymiarowej zmiennej losowej. Dla przypomnienia materiału przedstawionego w punkcie 2.1.2, dystrybuantę zmiennej losowej o rozkładzie $\mathscr{P}$ stanowi funkcja $F : \mathbb{R} \to [0, 1]$ dana wzorem $F(x) = \mathscr{P}((-\infty, x))$. Jej wartość dla argumentu x jest zatem równa prawdopodobieństwu, że rozważana zmienna losowa przyjmuje wartości mniejsze od x. Jeżeli – jak założono generalnie w niniejszej monografii – istnieje gęstość prawdopodobieństwa f, to między funkcjami F oraz f zachodzi związek

$$F(x) = \int_{-\infty}^{x} f(u)\, du. \tag{3.278}$$

Niech więc dana będzie jednowymiarowa zmienna losowa X, której rozkład ma gęstość prawdopodobieństwa spełniającą warunki (3.13)-(3.15), a także uzyskana z niej m-elementowa próba losowa

$$x_1, x_2, \ldots, x_m. \tag{3.279}$$

Stosując metodykę przedstawioną w podrozdziale 3.1, można wyznaczyć jądrowy estymator gęstości prawdopodobieństwa $\hat{f}$, po czym – na podstawie wzoru (2.278) – określić jądrowy estymator dystrybuanty $\hat{F}: \mathbb{R} \to [0,1]$ zgodnie z naturalną w tej sytuacji zależnością

$$\hat{F}(x) = \int_{-\infty}^{x} \hat{f}(u)\, du. \tag{3.280}$$

Po podstawieniu do powyższej równości podstawowej postaci jądrowego estymatora gęstości prawdopodobieństwa (3.3) otrzymuje się

$$\hat{F}(x) = \frac{1}{mh} \sum_{i=1}^{m} \int_{-\infty}^{x} K\left(\frac{u - x_i}{h}\right) du. \tag{3.281}$$

Oznaczając jako $I: \mathbb{R} \to [0, 1]$ funkcję pierwotną jądra K, czyli

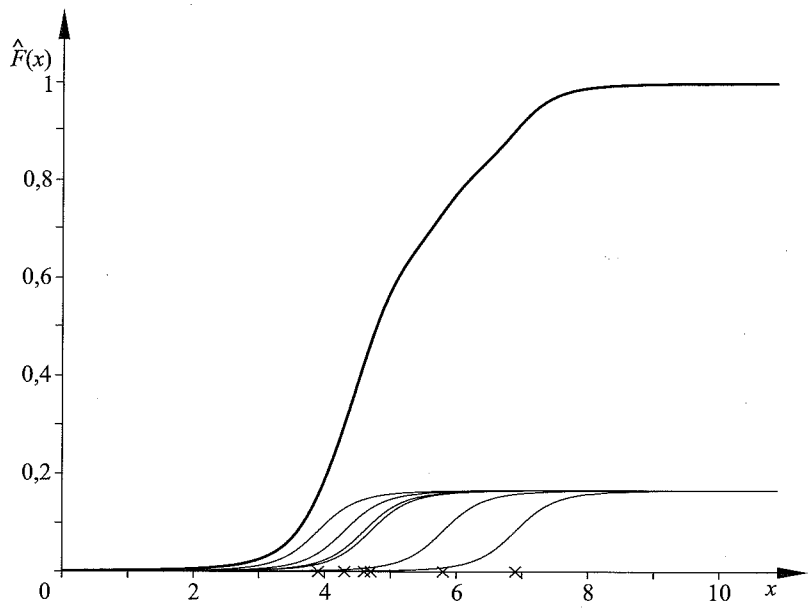
$$I(x) = \int_{-\infty}^{x} K(u)\, du, \tag{3.282}$$

i uwzględniając pochodną funkcji wewnętrznej, można zdefiniować ostatecznie podstawową postać **jądrowego estymatora dystrybuanty** następującym wzorem:

$$\hat{F}(x) = \frac{1}{m} \sum_{i=1}^{m} I\left(\frac{x - x_i}{h}\right). \tag{3.283}$$

Powyższy estymator został pokazany na rysunku 3.21. Funkcje I oraz, w konsekwencji, $\hat{F}$ są ciągłe i niemalejące. Jeżeli stosuje się jądro K o wartościach dodatnich, to funkcja I, a więc także $\hat{F}$, jest (ściśle) rosnąca.

Również w przypadku estymatora dystrybuanty wartość parametru wygładzania h można wyznaczyć w praktyce zgodnie z procedurami podanymi w punkcie 3.1.5, opartymi na kryterium scałkowanego błędu średniokwadratowego MISE. Przy wyborze postaci jądra K – poza specyficznymi uwarunkowaniami badanego zagadnienia aplikacyjnego – podstawowe staje się teraz jednakże wymaganie, aby funkcja pierwotna I mogła być wyrażona dogodnym wzorem analitycznym. Warunek ten spełnia większość stosowanych w praktyce typów jąder,



Rys. 3.21. Jądrowy estymator dystrybuanty (3.283); por. także rys. 1.5

oprócz normalnego (3.44). W szczególności, nawiązując do definicji (3.40)-(3.45), w przypadku optymalnego w sensie kryterium scałkowanego błędu średniokwadratowego MISE **jądra Epanecznikowa**

$$I(x)=\begin{cases}0 & \text{dla} \quad x\in(-\infty,\,-1),\\ \dfrac{1}{4}(-x^3+3x+2) & \text{dla} \quad x\in[-1,\,1],\\ 1 & \text{dla} \quad x\in(1,\,\infty),\end{cases} \tag{3.284}$$

dla **jądra jednostajnego**

$$I(x)=\begin{cases}0 & \text{dla} \quad x\in(-\infty,\,-1),\\ \dfrac{1}{2}(x+1) & \text{dla} \quad x\in[-1,\,1],\\ 1 & \text{dla} \quad x\in(1,\,\infty),\end{cases} \tag{3.285}$$

dla **jądra dwuwagowego**

$$I(x)=\begin{cases}0 & \text{dla} \quad x\in(-\infty,\,-1),\\ \dfrac{1}{16}(3x^5-10x^3+15x+8) & \text{dla} \quad x\in[-1,\,1],\\ 1 & \text{dla} \quad x\in(1,\,\infty)\end{cases} \tag{3.286}$$

oraz **jądra trójkątnego**:

$$I(x)=\begin{cases}0 & \text{dla} \quad x\in(-\infty,\,-1),\\ -\frac{1}{2}\operatorname{sgn}(x)x^2+x+\frac{1}{2} & \text{dla} \quad x\in[-1,\,1],\\ 1 & \text{dla} \quad x\in(1,\,\infty).\end{cases} \tag{3.287}$$

Jeżeli jest wymagane jądro o dodatnich wartościach, to warto wtedy zaproponować **jądro Cauchy'ego**:

$$K(x)=\frac{2}{\pi(x^2+1)^2}, \tag{3.288}$$

którego funkcja pierwotna wyraża się wzorem

$$I(x)=\frac{1}{\pi}\left[\frac{x}{x^2+1}+\operatorname{arctg}(x)+\frac{\pi}{2}\right]. \tag{3.289}$$

Wartości funkcjonałów (3.17)-(3.18), używanych w procedurach wyznaczania wartości parametru wygładzania, wynoszą w tym przypadku

$$U(K)=1, \tag{3.290}$$

$$W(K)=\frac{5}{4\pi}. \tag{3.291}$$

Efektywność jądra Cauchy'ego można zwiększyć, stosując **kombinację liniową** – analogiczną do zależności (3.46) – postaci

$$K(x)=\frac{9}{10}K_{\mathrm{E}}(x)+\frac{1}{10}K_{\mathrm{C}}(x), \tag{3.292}$$

gdzie K_{E} oznacza jądro Epanecznikowa (3.40), a K_{C} jądro Cauchy'ego (3.288). Dzięki współczynnikowi 9/10 powyższa kombinacja ma efektywność zbliżoną do efektywności dominującego w niej optymalnego jądra Epanecznikowa, lecz „domieszka" jądra Cauchy'ego zapewnia dodatniość wartości $K(x)$. Oczywiście funkcja pierwotna jądra (3.292) wyraża się wzorem

$$I(x)=\frac{9}{10}I_{\mathrm{E}}(x)+\frac{1}{10}I_{\mathrm{C}}(x), \tag{3.293}$$

gdzie funkcje pierwotne I_{E} oraz I_{C} dane są, odpowiednio, wzorami (3.284) i (3.289). Wartość parametru wygładzania powinna być wyznaczana osobno dla jąder Epanecznikowa i Cauchy'ego. W przypadku stosowania kombinacji liniowej jąder taka odrębność postępowania dotyczy również dalszych, przedstawianych poniżej aspektów.

Wobec jądrowego estymatora dystrybuanty możliwe jest stosowanie dodatkowych procedur, proponowanych w podrozdziale 3.1 względem estymatora gęstości prawdopodobieństwa. I tak, jeśli do wzoru (3.280) podstawić zależność (3.114), to **jądrowy estymator dystrybuanty z modyfikacją parametru wygładzania** przyjmuje postać

$$\hat{F}(x)=\frac{1}{m}\sum_{i=1}^{m} I\left(\frac{x-x_i}{hs_i}\right), \tag{3.294}$$

przy czym parametry s_1, s_2, ..., s_m powinny być wyznaczone zgodnie z algorytmem (3.112)-(3.115). Podobnie jak poprzednio, modyfikacja parametru wygładzania dodatkowo „spłaszcza" kształt estymatora w obszarach małego „zagęszczenia" elementów próby losowej, w przeciwieństwie do otoczenia wartości modalnych, gdzie następuje jego „wyostrzenie", co pozwala lepiej oddać specyfikę realnych rozkładów w obu powyższych przypadkach. Wyjątkowo korzystne – z praktycznego punktu widzenia – są tu również zmniejszenie wrażliwości tego typu estymatora na nieprecyzyjny dobór wartości parametru wygładzania i redukcja różnicy efektywności poszczególnych typów jąder względem optymalnego jądra Epanecznikowa. Ostatecznie, podobnie jak w przypadku estymatora gęstości prawdopodobieństwa, także dla jądrowego estymatora dystrybuanty stosowanie wyjątkowo korzystnej i jednocześnie relatywnie prostej obliczeniowo modyfikacji parametru wygładzania należy w zagadnieniach aplikacyjnych zalecać wręcz jako **obligatoryjne**. Z tego też powodu dalsze wzory w niniejszym podrozdziale będą zapisywane w postaci uwzględniającej użycie tej procedury. Warto pamiętać, iż postać podstawową uzyskuje się, przyjmując $s_1 = s_2 = ... = s_m = 1$.

Przedstawioną powyżej koncepcję jądrowego estymatora dystrybuanty można także uzupełnić w celu uwzględnienia **ograniczenia nośnika** badanej zmiennej losowej.

Najpierw rozpatrzmy przypadek ograniczenia lewostronnego do wartości $x_* \in \mathbb{R}$. Ponownie, niech $\chi_{[x_*,\infty)}$ oznacza funkcję charakterystyczną przedziału $[x_*, \infty)$, określoną równością (3.116). Ze wzoru (3.282) wynika wówczas, iż

$$\int_{-\infty}^{x} \chi_{[x_*,\infty)} K(u)\,du = \chi_{[x_*,\infty)}(x)[I(x)-I(x_*)]. \tag{3.295}$$

Nawiązując do zależności (3.119), jądrowy estymator dystrybuanty jest wtedy dany jako

$$\hat{F}(x)=\frac{1}{m}\sum_{i=1}^{m}\chi_{[x_*,\infty)}(x)\left[I\left(\frac{x-x_i}{hs_i}\right)-I\left(\frac{x_*-x_i}{hs_i}\right)+I\left(\frac{x+x_i-2x_*}{hs_i}\right)-I\left(\frac{x_i-x_*}{hs_i}\right)\right]. \tag{3.296}$$

Spełnia on w oczywisty sposób warunek $\hat{F}(x_*) = 0$, a zatem nośnik zmiennej losowej zawiera się w przedziale $[x_*, \infty)$. Funkcja $\hat{F}$ jest ciągła, także na ograniczeniu x_*.

Podobnie można rozważyć przypadek prawostronnego ograniczenia nośnika zmiennej losowej do wartości $x^* \in \mathbb{R}$. Oznaczając przez $\chi_{(-\infty,x^*]}$ funkcję charakterystyczną przedziału $(-\infty, x^*]$ określoną zależnością (3.120), z równości (3.282) otrzymuje się, iż

$$\int_{-\infty}^{x} \chi_{(-\infty,x^*]} K(u)\, du = \begin{cases} I(x), & \text{gdy} \quad x \le x^*, \\ I(x^*), & \text{gdy} \quad x > x^*. \end{cases} \tag{3.297}$$

Na podstawie wzoru (3.122) jądrowy estymator dystrybuanty jest wówczas dany jako

$$\hat{F}(x) = [1 - \chi_{(-\infty,x^*]}(x)] + \frac{1}{m}\sum_{i=1}^{m} \chi_{(-\infty,x^*]}(x) \left[I\left(\frac{x - x_i}{hs_i}\right) + I\left(\frac{x + x_i - 2x^*}{hs_i}\right)\right]. \tag{3.298}$$

Ponieważ, w konsekwencji założenia (3.5) stanowiącego o symetrii funkcji K względem zera, prawdziwe jest

$$I\left(\frac{x^* - x_i}{hs_i}\right) + I\left(\frac{x_i - x^*}{hs_i}\right) = 1 \tag{3.299}$$

(por. także rys. 3.10), więc estymator ten spełnia warunek $\hat{F}(x^*) = 1$, co implikuje, że nośnik zmiennej losowej zawiera się w założonym przedziale $(-\infty, x^*]$. Funkcja $\hat{F}$ jest ciągła, również na ograniczeniu x^*.

Koncepcja jądrowego estymatora dystrybuanty zostanie teraz poszerzona na **przypadek wielowymiarowy**. Otóż, rozważając n-wymiarową zmienną losową X i dysponując estymatorem gęstości prawdopodobieństwa $\hat{f}$, można na podstawie zależności (2.13) zdefiniować jądrowy estymator dystrybuanty jako funkcję $\hat{F}: \mathbb{R}^n \to [0,1]$ określoną wzorem

$$\hat{F}(x) = \hat{F}\left(\begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}\right) = \int_{-\infty}^{x_1}\int_{-\infty}^{x_2} \cdots \int_{-\infty}^{x_n} \hat{f}\left(\begin{bmatrix} u_1 \\ u_2 \\ \vdots \\ u_n \end{bmatrix}\right) du_n \ldots du_2\, du_1. \tag{3.300}$$

Jeżeli zatem $I: \mathbb{R}^n \to [0, 1]$ oznacza funkcję pierwotną jądra K, czyli

$$I(x) = I\left(\begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}\right) = \int_{-\infty}^{x_1}\int_{-\infty}^{x_2}\cdots\int_{-\infty}^{x_n} K\left(\begin{bmatrix} u_1 \\ u_2 \\ \vdots \\ u_n \end{bmatrix}\right) du_n \ldots du_2\, du_1, \tag{3.301}$$

to jądrowy estymator dystrybuanty jest dany zależnością

$$\hat{F}(x) = \frac{1}{m}\sum_{i=1}^{m} I\left(\frac{x - x_i}{hs_i}\right), \tag{3.302}$$

będącą odpowiednikiem wzoru (3.294). Poza wymienionymi wcześniej – w trakcie analizy jednowymiarowej zmiennej losowej – korzyściami uzyskiwanymi dzięki stosowaniu procedury modyfikacji, w przypadku wielowymiarowym należy dodatkowo zaznaczyć zmniejszenie różnicy efektywności między jądrem radialnym i produktowym. Własność ta jest wyjątkowo użyteczna z praktycznego punktu widzenia, gdyż w zagadnieniu estymacji dystrybuanty dla przypadku wielowymiarowego szczególnie preferowane jest jądro produktowe. Wtedy to bowiem zależność (3.301) przyjmuje dogodną postać

$$I(x) = I\left(\begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}\right) = \mathcal{J}(x_1)\cdot\mathcal{J}(x_2)\cdot\ldots\cdot\mathcal{J}(x_n), \tag{3.303}$$

gdzie $\mathcal{J}$ oznacza funkcję pierwotną jednowymiarowego jądra $\mathcal{K}$. Wobec jej poszczególnych czynników można również stosować w naturalny sposób procedury ograniczenia nośnika, jak było to przedstawione wcześniej dla przypadku jednowymiarowej zmiennej losowej.

Należy jeszcze uściślić, iż gdyby w specyficznym zadaniu pojawiła się konieczność stosowania wobec estymatora (3.302) **radialnego jądra Cauchy'ego**, to w celu zapewnienia warunku (3.16) należy uogólnić definicję (3.288) do postaci

$$K(x) = K\left(\begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}\right) = \frac{1}{c_n}\frac{1}{(1 + x_1^2 + x_2^2 + \ldots + x_n^2)^k}, \tag{3.304}$$

gdzie $k = [(n + 4)/2]$, przy czym $[a]$ stanowi część całkowitą liczby a, natomiast

$$c_n = nV_n \int_0^{\infty} \frac{u^{n-1}}{(1+u^2)^k} du, \tag{3.305}$$

gdy V_n oznacza objętość n-wymiarowej kuli jednostkowej w przestrzeni $\mathbb{R}^n$. Wzór określający jej wartość został podany na rysunku 3.4. Z kolei dla ustalonego wymiaru n, wyrażenie $\int_0^{\infty} u^{n-1}/(1+u^2)^k du$ może być łatwo wyznaczone na podstawie wzorów rekurencyjnych, które znajdują się w tablicach całek. Na koniec warto porównać treść powyższego akapitu z przypisem zamieszczonym w punkcie 3.1.3.

3.2.2. Jądrowy estymator kwantyla

Dysponując jądrowym estymatorem gęstości prawdopodobieństwa, można otrzymać nie tylko inne charakterystyki funkcyjne, ale także parametry charakterystyczne rozkładu. Należy jednak zaznaczyć, że ze względu na mnogość klasycznych, prostych i powszechnie znanych metod ich uzyskiwania takie postępowanie jest zalecane jedynie w nietypowych przypadkach lub przy niekorzystnych uwarunkowaniach problemu.

Zagadnienie wyznaczania parametrów charakterystycznych za pomocą estymatorów jądrowych zostanie poniżej przedstawione na przykładzie estymatora kwantyla jednowymiarowej zmiennej losowej, zdefiniowanego równaniem (2.31). Pośrednio obejmuje to także przypadek kwantyla wielowymiarowej zmiennej losowej, gdyż – jak wynika z definicji (2.32) – w celu jego oszacowania należy przeprowadzić procedurę właściwą jednowymiarowej zmiennej losowej wobec dowolnie ustalonej współrzędnej badanej zmiennej.

Rozważmy zatem jednowymiarową zmienną losową o rozkładzie $\mathcal{P}$ mającym gęstość prawdopodobieństwa. Niech będzie także ustalone $r \in (0, 1)$. Dla przypomnienia: kwantylem rzędu r jest taka liczba rzeczywista q, że prawdopodobieństwo, iż badana zmienna losowa przyjmie wartość mniejszą lub równą q, wynosi r, czyli spełniony jest warunek

$$\mathcal{P}\big((-\infty, q]\big) = r. \tag{3.306}$$

W przypadku gdy rozkład ma gęstość prawdopodobieństwa, wówczas jak wynika wprost z definicji (2.11):

$$\mathcal{P}\big(\{x\}\big) = 0 \quad \text{dla dowolnego} \quad x \in \mathbb{R} \tag{3.307}$$

i, na podstawie zależności (2.12) stosowanej przy $n = 1$, równanie (3.306) można zapisać w postaci

$$F(q) = r. \tag{3.308}$$

Jeżeli zatem gęstość prawdopodobieństwa rozważanej zmiennej losowej spełnia warunki (3.13)-(3.15) i dana jest uzyskana z niej m-elementowa próba losowa

$$x_1, x_2, \ldots, x_m, \tag{3.309}$$

to stosując metodykę opisaną w punkcie 3.2.1, można wyznaczyć jądrowy estymator dystrybuanty $\hat{F}$, po czym – uwzględniając wzór (3.308) – zdefiniować **jądrowy estymator kwantyla** $\hat{q} \in \mathbb{R}$ jako rozwiązanie równania

$$\hat{F}(\hat{q}) = r. \tag{3.310}$$

Jego istnienie jest zapewnione dzięki ciągłości funkcji $\hat{F}$. Jeżeli z kolei do konstrukcji jądrowego estymatora dystrybuanty używa się jądra K o dodatnich wartościach, to funkcja ta jest (ściśle) rosnąca i rozwiązanie powyższego równania jest wtedy jedyne, co stanowi ostatecznie o istnieniu i jednoznaczności zdefiniowanego wzorem (3.310) jądrowego estymatora kwantyla.

W celu znalezienia rozwiązania równania (3.310) można zastosować dogodne algorytmy numeryczne. I tak na przykład w przypadku klasycznej **metody (algorytmu) Newtona**, zakładając iż funkcja K jest dodatnia, jeżeli zapisać powyższe równanie w równoważnej postaci

$$\hat{F}(x) - r = 0 \tag{3.311}$$

i traktować lewą stronę jako funkcję

$$L(x) = \hat{F}(x) - r, \tag{3.312}$$

to z uwagi na zależność (3.278) jej pochodna wynosi

$$L'(x) = \hat{f}(x). \tag{3.313}$$

Dysponując skonstruowanymi na podstawie próby losowej (3.309) jądrowymi estymatorami gęstości prawdopodobieństwa $\hat{f}$ oraz dystrybuanty $\hat{F}$, jądrowy estymator kwantyla można wyznaczyć w praktyce jako granicę ciągu $\{\hat{q}_k\}_{k=0}^{\infty}$ danego wzorami

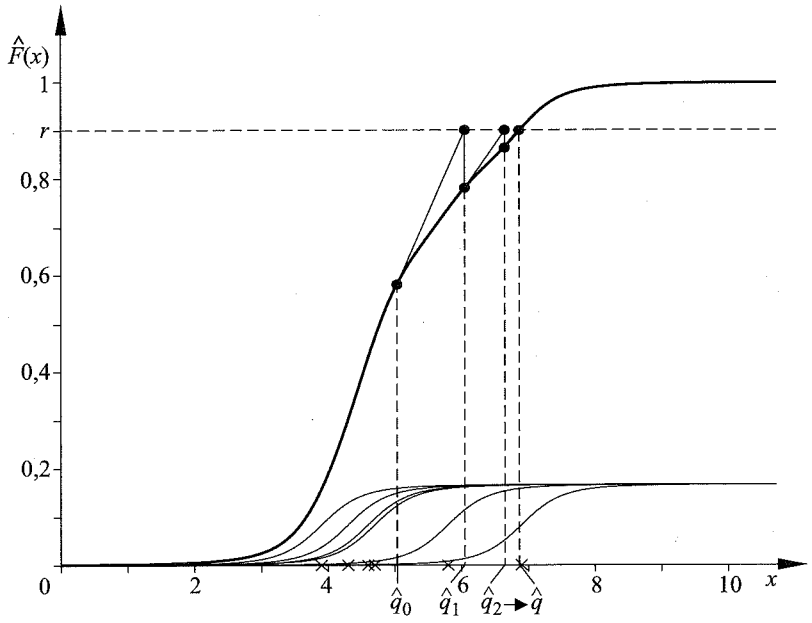
$$\hat{q}_0 = \frac{1}{m} \sum_{i=1}^{m} x_i, \tag{3.314}$$

$$\hat{q}_{k+1} = \hat{q}_k - \frac{\hat{F}(\hat{q}_k) - r}{\hat{f}(\hat{q}_k)} \quad \text{dla } k = 0, 1, \ldots, \tag{3.315}$$

przy warunku stopu postaci

$$\left|\hat{q}_{k+1}-\hat{q}_k\right| \le \frac{\hat{\sigma}}{100}, \tag{3.316}$$

gdzie $\hat{\sigma}$ oznacza estymator odchylenia standardowego (2.48). W praktyce algorytm (3.314)-(3.316) wymaga kilku, najczęściej 5-7, iteracji. Typowy przykład jest pokazany na rysunku 3.22.



Rys. 3.22. Jądrowy estymator kwantyla (3.310) – rozwiązanie z wykorzystaniem algorytmu (3.314)-(3.316)

W ogólnym przypadku powyższy algorytm nie musi być jednak globalnie zbieżny. W celu wyeliminowania ewentualnej – występującej sporadycznie przy bardzo małych licznościach próby losowej – rozbieżności zależność (3.315) należy zmodyfikować do postaci

$$\hat{q}_{k+1} = \begin{cases} \hat{q}_k - \dfrac{\hat{F}(\hat{q}_k)-r}{\hat{f}(\hat{q}_k)}, & \text{gdy} \quad \left|\dfrac{\hat{F}(\hat{q}_k)-r}{\hat{f}(\hat{q}_k)}\right| \le \dfrac{\hat{\sigma}}{10}, \\ \hat{q}_k - \dfrac{\hat{\sigma}}{10}\operatorname{sgn}(\hat{F}(\hat{q}_k)-r), & \text{gdy} \quad \left|\dfrac{\hat{F}(\hat{q}_k)-r}{\hat{f}(\hat{q}_k)}\right| > \dfrac{\hat{\sigma}}{10}, \end{cases}$$

$$\text{dla } k = 1, 2, \ldots, \tag{3.317}$$

ograniczając w ten sposób krok poszczególnych iteracji wartością $\hat{\sigma}/10$, co umożliwia osiągnięcie obszaru lokalnej zbieżności algorytmu Newtona. Liczba iteracji może wzrosnąć w tym przypadku nawet o 5-20, ale przy stosowanych w praktyce licznościach prób losowych rozbieżność algorytmu Newtona jest w rozważanym tu zagadnieniu na tyle mało prawdopodobna, iż wzór (3.317) należy raczej traktować jako opcjonalny element zapewnienia „bezpieczeństwa" obliczeń, zwłaszcza przy bardzo złożonych programach, gdy wykrycie przyczyn anomalii bywa niezwykle żmudne, a ich skutki mogą być wyjątkowo dotkliwe.

W procesie konstruowania estymatorów $\hat{f}$ oraz $\hat{F}$ można również wykorzystać dodatkowe, opisane wcześniej procedury. Ponownie, szczególnie poleca się stosowanie **modyfikacji parametru wygładzania**. Możliwe jest też zwiększenie efektywności poprzez użycie **kombinacji liniowej jąder** (3.292), a także wprowadzenie warunku **ograniczenia nośnika** zmiennej losowej. Należy jednak pamiętać, aby wobec obu estymatorów, $\hat{f}$ oraz $\hat{F}$, stosować zarówno te same procedury, jak i identyczne typy jąder oraz wartości parametrów wygładzania i modyfikujących. Estymator $\hat{f}$ nie jest tu bowiem traktowany jako integralne narzędzie modelowania rzeczywistości, lecz jedynie pochodna funkcji $\hat{F}$, i temu celowi musi być podporządkowany.

W wielu zagadnieniach aplikacyjnych jądrowy estymator kwantyla (3.310) pozwala uzyskać wyniki dokładniejsze, niż zapewniają to klasyczne estymatory tej wielkości, na przykład przedstawiony w punkcie 2.2.3 estymator (2.65). Jednak ze względu na relatywnie dużą złożoność estymator jądrowy powinien być w praktyce stosowany jedynie przy nietypowych lub niekorzystnych uwarunkowaniach, na przykład dla małej liczności próby losowej i/lub skrajnych wartości rzędu kwantyla, czyli gdy r jest bliskie 0 albo 1. Porównanie koncepcji estymatora klasycznego (2.65) z jądrowym (3.310) można bowiem zobrazować w następujący sposób. W przypadku pierwszego z nich, aby uzyskać odpowiednią dokładność oszacowania, „po obu stronach" szacowanej wartości kwantyla powinna się znaleźć pewna liczba elementów próby losowej. Gdyby przyjąć przykładowo, iż jest nią 5, to przy rzędzie $r = 0{,}9$ otrzymuje się już znaczną liczność próby $m = 50$. Natomiast w przypadku drugiego z nich wpływ na poszukiwaną wartość estymatora mają elementy próby losowej ulokowane „po obu jego stronach" – wszystkie elementy próby mają zatem swój udział przynajmniej dzięki „ogonom" przyporządkowanych im jąder. Przy prawidłowo skonstruowanym estymatorze jądrowym zwiększa to możliwości wykorzystania danych dostępnych w posiadanej próbie losowej.

3.2.3. Jądrowe estymatory charakterystyk warunkowych

W licznych zagadnieniach praktycznych informacja zawarta w próbie losowej może być znacząco uściślona, jeśli w opisie rzeczywistości zostaną uwzględnione dodatkowe wielkości mające bezpośredni wpływ na badany system, po czym – w trakcie procedury aplikacyjnej – podane zostaną wartości, jakie w konkretnej sytuacji wielkości te przyjmują. Przykładem może być tu sekwencja pomiarów temperatury zewnętrznej, otrzymywanych w ustalonym miejscu o różnych porach dnia i roku, w perspektywie wielu lat. Jeśli pominąć datę i godzinę uzyskiwania poszczególnych pomiarów, to użyteczność takiej próby jest minimalna, gdyż dane z różnych pór dnia i roku są nierozróżnialne i podczas wnioskowania statystycznego musiałyby być traktowane identycznie. Informacja zawarta w takiej próbie może być znacząco uściślona, gdyby uznać datę i godzinę za tak zwane zmienne warunkujące. Wówczas, w konkretnej sytuacji badanego zagadnienia praktycznego, podając ustalone (np. aktualne) wartości tych zmiennych, selekcjonowane są z próby adekwatne pomiary temperatury odpowiadające właśnie tym wartościom. Jednak w trywialnym ujęciu oznacza to pomiary z danej godziny i z danego dnia, czyli uzyskiwane raz w roku – z natury rzeczy jest ich niewiele i wnioskowanie statystyczne na podstawie tak nielicznej próby losowej byłoby mało wiarygodne. Metodyka konstruowania estymatorów jądrowych może dostarczyć użytecznego dla praktycznych zastosowań aparatu matematycznego do sformalizowania powyższego zagadnienia. Brane są tu pod uwagę wszystkie elementy próby losowej, ale im zostały one uzyskane przy „bliższych" owej ustalonej (aktualnej) wartości zmiennych warunkujących, tym mają większe znaczenie. Zmienne warunkujące mogą mieć także charakter binarny. Przykładowo, w przypadku zagadnień współczesnej medycyny większość danych może dotyczyć całej populacji, zarówno kobiet, jak i mężczyzn, ale rozważając konkretnego pacjenta, można uściślić informację, traktując płeć jako zmienną warunkującą – wtedy dane przyporządkowane tej właśnie płci będą mieć większe znaczenie.

Rozważana teraz będzie n-wymiarowa zmienna losowa X, a także n^*-wymiarowa zmienna W, zwana dalej **warunkującą zmienną losową**. Ich zestawienie $XW = \begin{bmatrix} X \\ W \end{bmatrix}$ jest więc $(n + n^*)$-wymiarową zmienną losową. Na podstawie przyjętego w punkcie 2.1.2 założenia o istnieniu gęstości prawdopodobieństwa rozważanych rozkładów niech $f_{XW} : \mathbb{R}^{n+n^*} \to [0, \infty)$ i $f_W : \mathbb{R}^{n^*} \to [0, \infty)$ oznaczają, odpowiednio, gęstości zmiennych XW oraz W. Niech także dana będzie tak zwana **wartość warunkująca**, czyli ustalona wartość warunkującej zmiennej losowej $w^* \in \mathbb{R}^{n^*}$, taka że

$$f_W(w^*) > 0. \tag{3.318}$$

Wtedy funkcja $f_{X|W=w^*} : \mathbb{R}^n \to [0, \infty)$ określona wzorem

$$f_{X|W=w^*}(x) = \frac{f_{XW}(x, w^*)}{f_W(w^*)} \quad \text{dla każdego } x \in \mathbb{R}^n \tag{3.319}$$

stanowi **warunkową gęstość prawdopodobieństwa** zmiennej losowej X przy wartości warunkującej w^*. Jest to odpowiednik znanej definicji z zakresu klasycznego rachunku prawdopodobieństwa:

$$P(A \mid B) = \frac{P(A \cap B)}{P(B)} \quad \text{dla } A, B \in \Sigma, \text{ gdy } P(B) \neq 0. \tag{3.320}$$

Warunkową gęstość prawdopodobieństwa $f_{X|W=w^*}$ można zatem traktować jako „podstawową" gęstość przedstawioną w punkcie 2.1.2, której postać została w zagadnieniu praktycznym uściślona poprzez podanie w^* – konkretnej wartości, jaką w danej sytuacji przyjęła zmienna warunkująca W.

Załóżmy zatem, że dana jest próba losowa

$$\begin{bmatrix} x_1 \\ w_1 \end{bmatrix}, \begin{bmatrix} x_2 \\ w_2 \end{bmatrix}, \ldots, \begin{bmatrix} x_m \\ w_m \end{bmatrix} \tag{3.321}$$

otrzymana ze zmiennej $XW = \begin{bmatrix} X \\ W \end{bmatrix}$. Jej poszczególne elementy są interpretowane jako wartości x_i uzyskane ze zmiennej losowej X, gdy zmienna warunkująca W przyjmowała odpowiednio wartości w_i, przy $i = 1, 2, \ldots, m$. Próbę (3.321) można w naturalny sposób zdekomponować na dwie próby losowe: złożoną z pierwszych n współrzędnych

$$x_1, x_2, \ldots, x_m, \tag{3.322}$$

odpowiadającą zmiennej X, oraz składającą się z ostatnich n^* współrzędnych

$$w_1, w_2, \ldots, w_m, \tag{3.323}$$

przyporządkowaną zmiennej W. Korzystając z metodyki przedstawionej w podrozdziale 3.1, na podstawie próby (3.321) można wtedy wyznaczyć $\hat{f}_{XW}$, czyli jądrowy estymator gęstości prawdopodobieństwa zmiennej losowej XW, a na podstawie próby (3.323) można uzyskać $\hat{f}_W$ – jądrowy estymator gęstości zmiennej warunkującej W. **Jądrowy estymator warunkowej gęstości prawdo-**

podobieństwa jest definiowany, w naturalnej konsekwencji zależności (3.319), jako funkcja $\hat{f}_{X|W=w^*}: \mathbb{R}^n \to [0, \infty)$ dana wzorem

$$\hat{f}_{X|W=w^*}(x) = \frac{\hat{f}_{XW}(x, w^*)}{\hat{f}_W(w^*)}. \tag{3.324}$$

Jeżeli do konstrukcji estymatora $\hat{f}_W$ stosuje się jądro o dodatnich wartościach, to implikowane przez nierówność (3.318) wymaganie

$$\hat{f}_W(w^*) > 0 \tag{3.325}$$

jest spełnione dla dowolnego $w^* \in \mathbb{R}^{n^*}$.

Formuła jądrowego estymatora warunkowej gęstości prawdopodobieństwa staje się szczególnie dogodna, jeśli do konstrukcji estymatorów $\hat{f}_{XW}$ i $\hat{f}_W$ używa się jąder produktowych przy założeniu tego samego typu jądra jednowymiarowego $\mathcal{K}$ o dodatnich wartościach. Wzór (3.324) można wówczas uściślić do postaci

$$\hat{f}_{X|W=w^*}(x) = \frac{\frac{1}{h_1 h_2 \ldots h_n} \sum_{i=1}^{m} \frac{1}{s_i^{n+n^*}} \mathcal{K}\left(\frac{x_1 - x_{i,1}}{h_1 s_i}\right) \mathcal{K}\left(\frac{x_2 - x_{i,2}}{h_2 s_i}\right) \ldots \mathcal{K}\left(\frac{x_n - x_{i,n}}{h_n s_i}\right) \mathcal{K}\left(\frac{w_1^* - w_{i,1}}{h_1^* s_i}\right) \mathcal{K}\left(\frac{w_2^* - w_{i,2}}{h_2^* s_i}\right) \ldots \mathcal{K}\left(\frac{w_{n^*}^* - w_{i,n^*}}{h_{n^*}^* s_i}\right)}{\sum_{i=1}^{m} \frac{1}{s_i^{n^*}} \mathcal{K}\left(\frac{w_1^* - w_{i,1}}{h_1^* s_i}\right) \mathcal{K}\left(\frac{w_2^* - w_{i,2}}{h_2^* s_i}\right) \ldots \mathcal{K}\left(\frac{w_{n^*}^* - w_{i,n^*}}{h_{n^*}^* s_i}\right)}, \tag{3.326}$$

przy czym $h_1, h_2, \ldots, h_n$ oraz $h_1^*, h_2^*, \ldots, h_{n^*}^*$ reprezentują, odpowiednio, parametry wygładzania przyporządkowane poszczególnym współrzędnym zmiennych losowych X oraz W, $s_1, s_2, \ldots, s_m$ stanowią parametry modyfikujące wyznaczone dla zmiennej XW, a poszczególne współrzędne wektorów x, w^* i $\begin{bmatrix} x_i \\ w_i \end{bmatrix}$ są oznaczone w sposób następujący:

$$x = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}, \tag{3.327}$$

$$w^* = \begin{bmatrix} w_1^* \\ w_2^* \\ \vdots \\ w_{n^*}^* \end{bmatrix}, \tag{3.328}$$

$$\begin{bmatrix} x_i \\ w_i \end{bmatrix} = \begin{bmatrix} x_{i,1} \\ x_{i,2} \\ \vdots \\ x_{i,n} \\ w_{i,1} \\ w_{i,2} \\ \vdots \\ w_{i,n^*} \end{bmatrix} \quad \text{dla } i = 1, 2, ..., m. \tag{3.329}$$

Zdefiniujmy tak zwane **parametry warunkujące** d_i dla $i = 1, 2, ..., m$ następującym wzorem:

$$d_i = \frac{1}{s_i^{n^*}} \mathscr{K}\left(\frac{w_1^* - w_{i,1}}{h_1^* s_i}\right) \mathscr{K}\left(\frac{w_2^* - w_{i,2}}{h_2^* s_i}\right) \ldots \mathscr{K}\left(\frac{w_{n^*}^* - w_{i,n^*}}{h_{n^*}^* s_i}\right). \tag{3.330}$$

Ponieważ zostało założone, iż jądro $\mathscr{K}$ przyjmuje wartości dodatnie, więc powyższe parametry są większe od zera. Wówczas jądrowy estymator warunkowej gęstości prawdopodobieństwa (3.326) można przedstawić w równoważnej postaci

$$\hat{f}_{X|W=w^*}(x) = \\ = \frac{1}{h_1 h_2 ... h_n \sum_{i=1}^{m} d_i} \sum_{i=1}^{m} \frac{d_i}{s_i^n} \mathscr{K}\left(\frac{x_1 - x_{i,1}}{h_1 s_i}\right) \mathscr{K}\left(\frac{x_2 - x_{i,2}}{h_2 s_i}\right) \ldots \mathscr{K}\left(\frac{x_2 - x_{i,n}}{h_n s_i}\right). \tag{3.331}$$

Wartość parametru d_i charakteryzuje „odległość" przyjętej wartości warunkującej w^* od wartości zmiennej warunkującej, przy której został uzyskany i-ty element próby losowej. Wówczas estymator (3.331) można zinterpretować jako kombinację liniową jąder przyporządkowanych poszczególnym elementom próby losowej odpowiadającej zmiennej X, przy czym współczynniki powyższej

kombinacji charakteryzują reprezentatywność tych elementów dla przyjętej wartości warunkującej w^*. Czynnik $\sum_{i=1}^{m} d_i$ normuje wartość estymatora w celu zapewnienia warunku $\int_{\mathbb{R}^n} \hat{f}_{X|W=w^*}(x)\,dx = 1$.

Dysponując jądrowym estymatorem warunkowej gęstości prawdopodobieństwa, można otrzymać inne warunkowe charakterystyki funkcyjne oraz parametry charakterystyczne rozkładu badanej zmiennej losowej. Kontynuując zakres rozważań poprzednich punktów, jako przykłady zostaną rozpatrzone jądrowe estymatory warunkowej dystrybuanty i kwantyla.

I tak warunkową dystrybuantę $F_{X|W=w^*} : \mathbb{R}^n \to [0,1]$ można, w nawiązaniu do wzoru (3.300), określić zależnością

$$F_{X|W=w^*}(x) = F_{X|W=w^*}\left(\begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}\right) = \int_{-\infty}^{x_1}\int_{-\infty}^{x_2}\cdots\int_{-\infty}^{x_n} f_{X|W=w^*}\left(\begin{bmatrix} u_1 \\ u_2 \\ \vdots \\ u_n \end{bmatrix}\right) du_n \ldots du_2\, du_1, \tag{3.332}$$

a następnie zdefiniować **jądrowy estymator warunkowej dystrybuanty** $\hat{F}_{X|W=w^*} : \mathbb{R}^n \to [0,1]$ w naturalnej postaci

$$\hat{F}_{X|W=w^*}(x) = \hat{F}_{X|W=w^*}\left(\begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}\right) = \int_{-\infty}^{x_1}\int_{-\infty}^{x_2}\cdots\int_{-\infty}^{x_n} \hat{f}_{X|W=w^*}\left(\begin{bmatrix} u_1 \\ u_2 \\ \vdots \\ u_n \end{bmatrix}\right) du_n \ldots du_2\, du_1. \tag{3.333}$$

Jeżeli stosuje się jądra produktowe (3.48) przy tym samym typie jądra jednowymiarowego $\mathcal{K}$ o dodatnich wartościach, to

$$\hat{F}_{X|W=w^*}(x) = \frac{\sum_{i=1}^{m} \frac{1}{s_i^{n^*}} \mathcal{J}\left(\frac{x_1 - x_{i,1}}{h_1 s_i}\right)\mathcal{J}\left(\frac{x_2 - x_{i,2}}{h_2 s_i}\right)\ldots\mathcal{J}\left(\frac{x_n - x_{i,n}}{h_n s_i}\right)\mathcal{K}\left(\frac{w_1^* - w_{i,1}}{h_1^* s_i}\right)\mathcal{K}\left(\frac{w_2^* - w_{i,2}}{h_2^* s_i}\right)\ldots\mathcal{K}\left(\frac{w_{n^*}^* - w_{i,n^*}}{h_{n^*}^* s_i}\right)}{\sum_{i=1}^{m} \frac{1}{s_i^{n^*}} \mathcal{K}\left(\frac{w_1^* - w_{i,1}}{h_1^* s_i}\right)\mathcal{K}\left(\frac{w_2^* - w_{i,2}}{h_2^* s_i}\right)\ldots\mathcal{K}\left(\frac{w_{n^*}^* - w_{i,n^*}}{h_{n^*}^* s_i}\right)}, \tag{3.334}$$

gdzie poszczególne oznaczenia są zgodne z podanymi za pomocą wzorów (3.327)-(3.329) oraz dodatkowo $\mathscr{T}$ jest funkcją pierwotną jądra $\mathscr{K}$ (por. wzór (3.282)). W szczególności, po wprowadzeniu parametrów warunkujących d_i, zdefiniowanych zależnością (3.330), jądrowy estymator warunkowej dystrybuanty (3.334) można przedstawić w równoważnej postaci

$$\hat{F}_{X|W=w^*}(x) = \frac{1}{\sum_{i=1}^{m} d_i} \sum_{i=1}^{m} d_i \, \mathscr{T}\left(\frac{x_1 - x_{i,1}}{h_1 s_i}\right) \mathscr{T}\left(\frac{x_2 - x_{i,2}}{h_2 s_i}\right) \ldots \mathscr{T}\left(\frac{x_2 - x_{i,n}}{h_n s_i}\right). \tag{3.335}$$

Podobnie jak poprzednio w przypadku gęstości prawdopodobieństwa, estymator (3.335) jest kombinacją liniową funkcji pierwotnych kolejnych jąder przyporządkowanych poszczególnym elementom próby losowej odpowiadającej zmiennej X, przy czym współczynniki tej kombinacji charakteryzują reprezentatywność tych elementów dla przyjętej wartości warunkującej w^*.

Rozważmy teraz przypadek, gdy X jest jednowymiarową zmienną losową. Niech będzie także dana liczba $r \in (0, 1)$. Wówczas w naturalny sposób przenosi się zależność (3.308), definiując warunkowy kwantyl rzędu r jako taką liczbę rzeczywistą $q_{X|W=w^*}$, która stanowi rozwiązanie równania

$$F_{X|W=w^*}(q_{X|W=w^*}) = r. \tag{3.336}$$

Jądrowy estymator warunkowego kwantyla rzędu *r*, oznaczany dalej przez $\hat{q}_{X|W=w^*} \in \mathbb{R}$, można wtedy określić jako rozwiązanie następującego równania:

$$\hat{F}_{X|W=w^*}(\hat{q}_{X|W=w^*}) = r. \tag{3.337}$$

Stosowanie do konstrukcji estymatora $\hat{F}_{X|W=w^*}$ jądra o dodatnich wartościach zapewnia istnienie i jednoznaczność powyższego rozwiązania. Co więcej, dysponując także jądrowym estymatorem warunkowej gęstości prawdopodobieństwa $\hat{f}_{X|W=w^*}$, wartość powyższego rozwiązania można wyznaczyć numerycznie z użyciem metody Newtona jako granicę ciągu $\{q_k^*\}_{k=0}^{\infty}$ zdefiniowanego wzorami

$$\hat{q}_0^* = \frac{1}{m} \sum_{i=1}^{m} x_i, \tag{3.338}$$

$$\hat{q}_{k+1}^* = \hat{q}_k^* - \frac{r - \hat{F}_{X|W=w^*}(\hat{q}_k^*)}{\hat{f}_{X|W=w^*}(\hat{q}_k^*)} \quad \text{dla } k = 0, 1, \ldots \tag{3.339}$$

z warunkiem stopu

$$|\hat{q}^*_{k+1} - \hat{q}^*_k| \leq \frac{\hat{\sigma}}{100}, \tag{3.340}$$

gdzie $\hat{\sigma}$ oznacza estymator odchylenia standardowego zmiennej losowej X, który można obliczyć, stosując wzór (2.48) wobec próby (3.322). W przypadku rozbieżności powyższego algorytmu równość (3.339) należy zastąpić analogiczną do (3.317) zależnością

$$\hat{q}^*_{k+1} = \begin{cases} \hat{q}^*_k - \dfrac{\hat{F}_{X|W=w^*}(\hat{q}^*_k) - r}{\hat{f}_{X|W=w^*}(\hat{q}^*_k)}, & \text{gdy} \quad \left| \dfrac{\hat{F}_{X|W=w^*}(\hat{q}^*_k) - r}{\hat{f}_{X|W=w^*}(\hat{q}^*_k)} \right| \leq \dfrac{\hat{\sigma}}{10}, \\ \hat{q}^*_k - \dfrac{\hat{\sigma}}{10} \operatorname{sgn}(\hat{F}_{X|W=w^*}(\hat{q}^*_k) - r), & \text{gdy} \quad \left| \dfrac{\hat{F}_{X|W=w^*}(\hat{q}^*_k) - r}{\hat{f}_{X|W=w^*}(\hat{q}^*_k)} \right| > \dfrac{\hat{\sigma}}{10}, \end{cases}$$

$$\text{dla } k = 0, 1, \ldots, \tag{3.341}$$

ograniczając krok poszczególnych iteracji w celu osiągnięcia obszaru jego lokalnej zbieżności.

Przedstawiona tu metodyka obejmuje praktycznie również zagadnienie estymacji kwantyla warunkowego n-wymiarowej zmiennej losowej, gdyż w takim przypadku należy przeprowadzić powyższą procedurę, właściwą jednowymiarowej zmiennej losowej, wobec dowolnie ustalonej współrzędnej zmiennej losowej.

W trakcie konstruowania estymatorów $\hat{f}_{X|W=w^*}$ oraz $\hat{F}_{X|W=w^*}$ (także na potrzeby kwantyla warunkowego $\hat{q}_{X|W=w^*}$) można stosować **dodatkowe procedury** przedstawione w podrozdziale 3.1. Modyfikacja parametru wygładzania jest nadal w praktyce zalecana jako obligatoryjna. Z teoretycznego punktu widzenia pewne korzyści są związane ze stosowaniem – w celu zwiększenia efektywności – kombinacji liniowej jąder. Możliwość uwzględnienia ograniczeń nośnika poszczególnych współrzędnych zmiennych losowych X oraz W jest oczywista. W praktycznych zastosowaniach można także wprowadzić do zmiennej warunkującej W współrzędne binarne. W przypadku estymacji warunkowej gęstości prawdopodobieństwa współrzędne tego typu mogą także dotyczyć zmiennej X. Należy ponowić zastrzeżenie, aby na potrzeby konstrukcji jądrowego estymatora warunkowego kwantyla stosować wobec $\hat{f}_{X|W=w^*}$ oraz $\hat{F}_{X|W=w^*}$ te same procedury, identyczne typy jąder i wartości parametrów.

3.2.4. Przykłady

Niniejszy punkt zawiera trzy przykłady ilustrujące wyznaczanie poszczególnych rodzajów estymatorów jądrowych rozważanych wcześniej w tym podrozdziale. I tak rozpatrzone będą kolejno:

- jądrowy estymator dystrybuanty,
- jądrowy estymator kwantyla,
- jądrowe estymatory charakterystyk warunkowych: gęstości prawdopodobieństwa, dystrybuanty i kwantyla.

Podobnie jak w podrozdziale 3.1, w celu zapewnienia przejrzystości algorytmów obliczeniowych przykłady te zostaną przedstawione dla prób trójelementowych, czyli o liczności absolutnie nieakceptowalnej w praktyce. Wszelkie podstawowe aspekty, dotyczące wyznaczania jądrowego estymatora gęstości prawdopodobieństwa oraz ogólnych procedur związanych z użyciem tego typu estymatorów, zostaną tu znacznie ograniczone, gdyż były szczegółowo przedstawione w ramach przykładów 3.1-3.5 zamieszczonych w punkcie 3.1.10.

Przykład 3.6 – dystrybuanta

W nawiązaniu do przykładu 3.1, niech dana będzie jednowymiarowa zmienna losowa X i uzyskana z niej trójelementowa próba

$$x_1 = 3, \quad x_2 = 3{,}3, \quad x_3 = 5. \tag{3.342}$$

Rozważmy zadanie wyznaczenia jądrowego estymatora dystrybuanty, gdy do jego konstrukcji jest używane jądro Cauchy'ego (3.288).

Wartość parametru wygładzania zostanie określona za pomocą metody podstawień drugiego rzędu. Do obliczenia wartości $Z(f)$ powszechnie stosuje się tu jądro normalne i dlatego pierwsza faza procedury jest identyczna z początkiem przykładu 3.1, aż do zależności (3.158). Podstawiając do równości (3.31) uzyskany tam wynik, a także odczytane ze wzorów (3.290)-(3.291) wartości funkcjonałów U i W dla przyjętego teraz jądra Cauchy'ego, można otrzymać

$$h = \left(\frac{0{,}399}{1^2 \cdot 0{,}336 \cdot 3} \right)^{1/5} = 0{,}831. \tag{3.343}$$

Na potrzeby procedury modyfikacji parametru wygładzania (3.112)-(3.115) należy najpierw wyznaczyć podstawową postać jądrowego estymatora gęstości prawdopodobieństwa z zastosowaniem jądra Cauchy'ego, czyli tego samego typu, które będzie używane do konstrukcji estymatora dystrybuanty. I tak, korzystając z zależności (3.288) i (3.343), na podstawie definicji (3.3) otrzymuje się

$$\hat{f}(x)=0{,}255\left(\frac{1}{\left[\left(\frac{x-3}{0{,}831}\right)^2+1\right]^2}+\frac{1}{\left[\left(\frac{x-3{,}3}{0{,}831}\right)^2+1\right]^2}+\frac{1}{\left[\left(\frac{x-5}{0{,}831}\right)^2+1\right]^2}\right). \tag{3.344}$$

Podobnie jak w przykładzie 3.1, należy teraz obliczyć kolejno: wartości estymatora (3.344) dla poszczególnych elementów próby losowej $\hat{f}(x_1)=0{,}460$, $\hat{f}(x_2)=0{,}464$, $\hat{f}(x_3)=0{,}270$, ich średnią geometryczną $\tilde{s}=0{,}386$, a stąd wartości parametrów modyfikujących

$$s_1=0{,}916, \tag{3.345}$$

$$s_2=0{,}912, \tag{3.346}$$

$$s_3=1{,}196. \tag{3.347}$$

Podstawiając funkcję pierwotną jądra Cauchy'ego (3.289) oraz wartości parametrów (3.343) i (3.345)-(3.347) do wzoru określającego jądrowy estymator dystrybuanty (3.294), uzyskuje się następującą jego postać:

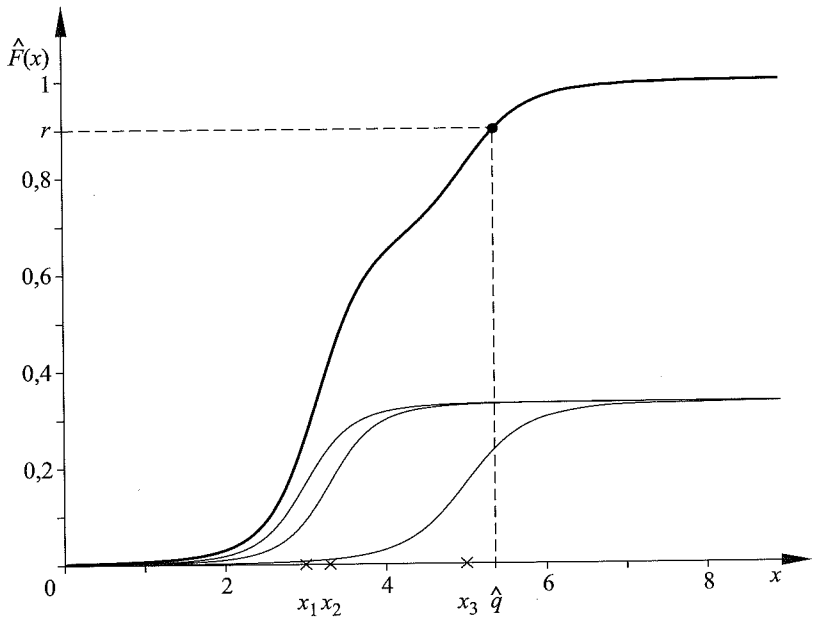
$$\begin{aligned}\hat{F}(x)=\frac{1}{3}\Bigg[&\frac{1}{\pi}\left(\frac{\frac{x-3}{0{,}831\cdot 0{,}916}}{\left(\frac{x-3}{0{,}831\cdot 0{,}916}\right)^2+1}+\operatorname{arctg}\left(\frac{x-3}{0{,}831\cdot 0{,}916}\right)+\frac{\pi}{2}\right)+\\&+\frac{1}{\pi}\left(\frac{\frac{x-3{,}3}{0{,}831\cdot 0{,}912}}{\left(\frac{x-3{,}3}{0{,}831\cdot 0{,}912}\right)^2+1}+\operatorname{arctg}\left(\frac{x-3{,}3}{0{,}831\cdot 0{,}912}\right)+\frac{\pi}{2}\right)+\\&+\frac{1}{\pi}\left(\frac{\frac{x-5}{0{,}831\cdot 1{,}196}}{\left(\frac{x-5}{0{,}831\cdot 1{,}196}\right)^2+1}+\operatorname{arctg}\left(\frac{x-5}{0{,}831\cdot 1{,}196}\right)+\frac{\pi}{2}\right)\Bigg]\end{aligned} \tag{3.348}$$

i stąd po wykonaniu odpowiednich operacji

$$\hat{F}(x) = 0{,}106\left[\frac{\frac{x-3}{0{,}761}}{\left(\frac{x-3}{0{,}761}\right)^2+1} + \operatorname{arctg}\left(\frac{x-3}{0{,}761}\right) + \frac{\frac{x-3{,}3}{0{,}758}}{\left(\frac{x-3{,}3}{0{,}758}\right)^2+1} + \operatorname{arctg}\left(\frac{x-3{,}3}{0{,}758}\right) + \right.$$

$$\left. + \frac{\frac{x-5}{0{,}994}}{\left(\frac{x-5}{0{,}994}\right)^2+1} + \operatorname{arctg}\left(\frac{x-5}{0{,}994}\right) + 4{,}712\right]. \qquad (3.349)$$

Otrzymany w ten sposób jądrowy estymator dystrybuanty jest pokazany na rysunku 3.23.



Rys. 3.23. Jądrowy estymator dystrybuanty (3.349); por. także rys. 3.12 (zaznaczono również jądrowy estymator kwantyla rzędu $r = 0{,}9$ z przykładu 3.7)

Przykład 3.7 – kwantyl

Kontynuując przedmiot przykładów 3.1 i 3.6, rozpatrzmy jednowymiarową zmienną losową X i uzyskaną z niej trójelementową próbę (3.342). W niniejszym przykładzie będzie wyznaczony jądrowy estymator kwantyla rzędu $r = 0{,}9$, z użyciem jądra Cauchy'ego (3.288). Przy powyższych założeniach w przykła-

dzie 3.6 zostały obliczone kolejno: wartości parametru wygładzania (3.343) i parametrów modyfikujących (3.345)-(3.347) oraz jądrowy estymator dystrybuanty (3.349).

A teraz, postępując analogicznie jak w przykładzie 3.1, aczkolwiek dla przyjętego tu jądra Cauchy'ego (3.288) oraz przy wartościach parametrów (3.343) i (3.345)-(3.347), uzyskuje się ze wzoru (3.114) równanie jądrowego estymatora gęstości prawdopodobieństwa z modyfikacją parametru wygładzania:

$$\hat{f}(x)=\frac{0{,}279}{\left[\left(\frac{x-3}{0{,}761}\right)^2+1\right]^2}+\frac{0{,}280}{\left[\left(\frac{x-3{,}3}{0{,}758}\right)^2+1\right]^2}+\frac{0{,}214}{\left[\left(\frac{x-5}{0{,}994}\right)^2+1\right]^2} \tag{3.350}$$

(można zauważyć, że podstawowa postać tego estymatora (3.3) została już wyznaczona w poprzednim przykładzie za pomocą zależności (3.344)).

Z kolei, również w ramach przykładu 3.1, otrzymano we wzorze (3.149):

$$\hat{\sigma}=1{,}078, \tag{3.351}$$

co pozwala obliczyć wartość prawej strony warunku stopu algorytmu Newtona (3.316):

$$0{,}01\cdot\hat{\sigma}=0{,}01\cdot 1{,}078=0{,}011. \tag{3.352}$$

Dysponując estymatorami dystrybuanty (3.349) i gęstości prawdopodobieństwa (3.350), należy teraz zastosować algorytm Newtona (3.314)-(3.316). Otrzymuje się wówczas kolejno:

$$\hat{q}_0=\frac{1}{3}(3+3{,}3+5)=3{,}767, \tag{3.353}$$

$$\hat{q}_1=3{,}767-\frac{\hat{F}(3{,}767)-0{,}9}{\hat{f}(3{,}767)}=4{,}989 \quad \text{oraz} \quad |\hat{q}_1-\hat{q}_0|=1{,}222, \tag{3.354}$$

$$\hat{q}_2=4{,}989-\frac{\hat{F}(4{,}989)-0{,}9}{\hat{f}(4{,}989)}=5{,}335 \quad \text{oraz} \quad |\hat{q}_2-\hat{q}_1|=0{,}346, \tag{3.355}$$

$$\hat{q}_3=5{,}335-\frac{\hat{F}(5{,}335)-0{,}9}{\hat{f}(5{,}335)}=5{,}370 \quad \text{oraz} \quad |\hat{q}_3-\hat{q}_2|=0{,}034, \tag{3.356}$$

$$\hat{q}_4=5{,}370-\frac{\hat{F}(5{,}370)-0{,}9}{\hat{f}(5{,}370)}=5{,}371 \quad \text{oraz} \quad |\hat{q}_4-\hat{q}_3|=0{,}001. \tag{3.357}$$

Porównując sukcesywnie wartości poszczególnych kroków $|\hat{q}_1 - \hat{q}_0|$, $|\hat{q}_2 - \hat{q}_1|$, $|\hat{q}_3 - \hat{q}_2|$ i $|\hat{q}_4 - \hat{q}_3|$ z wartością wyznaczoną we wzorze (3.352), można stwierdzić, iż warunek stopu (3.316) jest spełniony dopiero w przypadku ostatniego z nich. Tak więc jako jądrowy estymator kwantyla należy przyjąć czwarty element ciągu $\hat{q}_4$, czyli ostatecznie

$$\hat{q} = 5{,}371. \tag{3.358}$$

Wartość ta została zaznaczona na rysunku 3.23.

Przykład 3.8 – charakterystyki warunkowe

W niniejszym przykładzie będzie rozważana jednowymiarowa zmienna losowa X, a także jednowymiarowa zmienna warunkująca W, czyli $n = 1$ oraz $n^* = 1$. Załóżmy ponadto, że w chwili gdy zmienna warunkująca W przyjmowała wartość 6, otrzymana wartość zmiennej losowej X wynosiła 3, a następnie odpowiednio: 7 oraz 3,3 i wreszcie 4 oraz 5. Wyniki te można zapisać w postaci następującej trójelementowej próby:

$$\begin{bmatrix} x_1 \\ w_1 \end{bmatrix} = \begin{bmatrix} 3 \\ 6 \end{bmatrix}, \quad \begin{bmatrix} x_2 \\ w_2 \end{bmatrix} = \begin{bmatrix} 3{,}3 \\ 7 \end{bmatrix}, \quad \begin{bmatrix} x_3 \\ w_3 \end{bmatrix} = \begin{bmatrix} 5 \\ 4 \end{bmatrix}. \tag{3.359}$$

Powyższą próbę losową można zdekomponować na dwie próby: składającą się z pierwszych współrzędnych

$$x_1 = 3, \quad x_2 = 3{,}3, \quad x_3 = 5, \tag{3.360}$$

a także złożoną z drugich współrzędnych

$$w_1 = 6, \quad w_2 = 7, \quad w_3 = 4. \tag{3.361}$$

Przedmiot niniejszego przykładu stanowią charakterystyki warunkowe zmiennej losowej X – gęstość prawdopodobieństwa, dystrybuanta i kwantyl – wyznaczone przy wartości warunkującej

$$w^* = 4{,}5. \tag{3.362}$$

Najpierw będzie rozważany **jądrowy estymator warunkowej gęstości prawdopodobieństwa**. Użyte tu zostanie jądro produktowe (3.48) z jednowymiarowym jądrem normalnym (3.44).

I tak, w przykładzie 3.4 uzyskano dla prób (3.360)-(3.361) wartości parametrów wygładzania dane wzorami (3.244) i (3.246), które poniżej będą oznaczone – zgodnie z nieco odmienną notacją użytą w odniesieniu do charakterystyk warunkowych – jako

$$h = 0{,}811, \tag{3.363}$$

$$h^* = 1{,}412. \tag{3.364}$$

Zostały tam także obliczone wartości parametrów modyfikujących (3.253)-(3.255).

Stosując równanie jądra normalnego (3.44) oraz wartości (3.361)-(3.362), (3.364) i (3.253)-(3.255), na podstawie definicji (3.330) można uzyskać parametry warunkujące

$$d_1 = \frac{1}{0{,}896}\frac{1}{\sqrt{2\pi}}\exp\left(-\frac{1}{2}\left(\frac{4{,}5-6}{1{,}412\cdot 0{,}896}\right)^2\right) = 0{,}220, \tag{3.365}$$

$$d_2 = \frac{1}{0{,}915}\frac{1}{\sqrt{2\pi}}\exp\left(-\frac{1}{2}\left(\frac{4{,}5-7}{1{,}412\cdot 0{,}915}\right)^2\right) = 0{,}067, \tag{3.366}$$

$$d_3 = \frac{1}{1{,}194}\frac{1}{\sqrt{2\pi}}\exp\left(-\frac{1}{2}\left(\frac{4{,}5-4}{1{,}412\cdot 1{,}194}\right)^2\right) = 0{,}320. \tag{3.367}$$

Stąd, podstawiając wartości (3.363), (3.253)-(3.255) oraz (3.365)-(3.367) do wzoru (3.331), otrzymuje się

$$\hat{f}_{X|W=4,5}(x) =$$

$$= \frac{1}{0{,}811\cdot(0{,}220+0{,}067+0{,}320)}\left[\frac{0{,}220}{0{,}896}\frac{1}{\sqrt{2\pi}}\exp\left(-\frac{1}{2}\left(\frac{x-3}{0{,}811\cdot 0{,}896}\right)^2\right)+\right.$$

$$\left.+\frac{0{,}067}{0{,}915}\frac{1}{\sqrt{2\pi}}\exp\left(-\frac{1}{2}\left(\frac{x-3{,}3}{0{,}811\cdot 0{,}915}\right)^2\right)+\frac{0{,}320}{1{,}194}\frac{1}{\sqrt{2\pi}}\exp\left(-\frac{1}{2}\left(\frac{x-5}{0{,}811\cdot 1{,}194}\right)^2\right)\right], \tag{3.368}$$

czyli ostatecznie

$$\hat{f}_{X|W=4,5}(x) = 0{,}199\exp\left(-\frac{1}{2}\left(\frac{x-3}{0{,}727}\right)^2\right)+0{,}059\exp\left(-\frac{1}{2}\left(\frac{x-3{,}3}{0{,}742}\right)^2\right)+$$

$$+0{,}217\exp\left(-\frac{1}{2}\left(\frac{x-5}{0{,}968}\right)^2\right). \tag{3.369}$$

Wyznaczony powyżej jądrowy estymator warunkowej gęstości prawdopodobieństwa zmiennej losowej X przy wartości warunkującej (3.362) jest pokazany na rysunku 3.24. W stosunku do gęstości bezwarunkowej z rysunku 3.12 widać zwiększony wpływ jądra przyporządkowanego trzeciemu elementowi próby losowej, który został uzyskany przy warunku równym 4, czyli najbliższym założonej wartości warunkującej 4,5. Osłabieniu uległo natomiast znaczenie jądra związanego z drugim elementem próby, uzyskanym przy warunku 7 jako najdalszym powyższej wartości warunkującej.

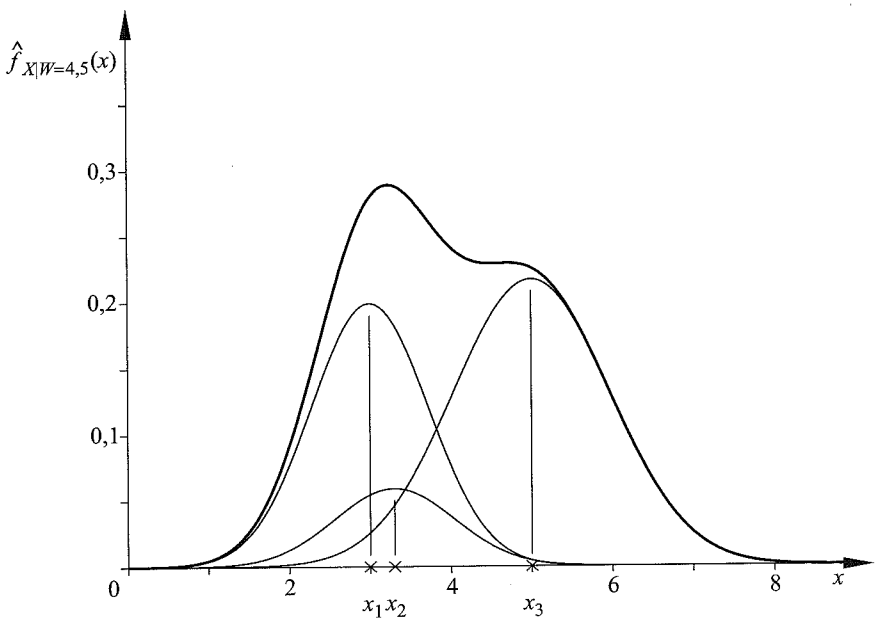
Kontynuując temat niniejszego przykładu, zostanie teraz wyznaczony **jądrowy estymator warunkowej dystrybuanty**. Z powodu braku analitycznej zależności na funkcję pierwotną jądra normalnego, które było dotychczas stosowane wobec gęstości prawdopodobieństwa, poniżej zostało przyjęte jądro produktowe (3.48) z jednowymiarowym jądrem Cauchy'ego (3.288).

Rozważając próbę losową (3.360), w przykładzie 3.6 – gdzie także było stosowane jądro Cauchy'ego – wyznaczono za pomocą wzoru (3.343) wartość parametru wygładzania

$$h = 0{,}831. \tag{3.370}$$

Postępując identycznie wobec próby (3.361), otrzymuje się

$$h^* = 1{,}447. \tag{3.371}$$



Rys. 3.24. Jądrowy estymator warunkowej gęstości prawdopodobieństwa (3.369); por. także rys. 3.12

Na potrzeby modyfikacji parametru wygładzania należy teraz wyznaczyć podstawową postać jądrowego estymatora gęstości prawdopodobieństwa (3.3) dla jądra produktowego (3.48) z jednowymiarowym jądrem Cauchy'ego (3.288), która po odpowiednich podstawieniach i obliczeniach jest dana wzorem

$$\hat{f}\left(\begin{bmatrix} x \\ w \end{bmatrix}\right) = 0{,}112\left[\frac{1}{\left[\left(\left(\frac{x-3}{0{,}831}\right)^2+1\right)\left(\left(\frac{w-6}{1{,}447}\right)^2+1\right)\right]^2} + \right.$$

$$\left. + \frac{1}{\left[\left(\left(\frac{x-3{,}3}{0{,}831}\right)^2+1\right)\left(\left(\frac{w-7}{1{,}447}\right)^2+1\right)\right]^2} + \frac{1}{\left[\left(\left(\frac{x-5}{0{,}831}\right)^2+1\right)\left(\left(\frac{w-4}{1{,}447}\right)^2+1\right)\right]^2}\right]. \tag{3.372}$$

Realizując procedurę modyfikacji parametru wygładzania (3.112)-(3.115), podobnie jak w przykładzie 3.4 obliczyć można kolejno: wartości estymatora (3.372) dla poszczególnych elementów próby losowej $\hat{f}\left(\begin{bmatrix} x_1 \\ w_1 \end{bmatrix}\right) = 0{,}153$, $\hat{f}\left(\begin{bmatrix} x_2 \\ w_2 \end{bmatrix}\right) = 0{,}152$, $\hat{f}\left(\begin{bmatrix} x_3 \\ w_3 \end{bmatrix}\right) = 0{,}112$, ich średnią geometryczną $\tilde{s} = 0{,}138$ i wreszcie wartości parametrów modyfikujących $s_1 = 0{,}950$, $s_2 = 0{,}953$ i $s_3 = 1{,}110$. Stosując je razem z definicją jądra Cauchy'ego (3.288) oraz wielkościami (3.361)-(3.362) i (3.371), otrzymuje się następujące parametry warunkujące (3.330):

$$d_1 = \frac{1}{0{,}950}\frac{2}{\pi\left[\left(\frac{4{,}5-6}{1{,}447\cdot 0{,}950}\right)^2+1\right]^2} = 0{,}134, \tag{3.373}$$

$$d_2 = \frac{1}{0{,}953}\frac{2}{\pi\left[\left(\frac{4{,}5-7}{1{,}447\cdot 0{,}953}\right)^2+1\right]^2} = 0{,}036, \tag{3.374}$$

$$d_3 = \frac{1}{1{,}110}\frac{2}{\pi\left[\left(\frac{4{,}5-4}{1{,}447\cdot 1{,}110}\right)^2+1\right]^2} = 0{,}477. \tag{3.375}$$

Z kolei, podstawiając do wzoru (3.335) równanie opisujące funkcję pierwotną (3.289), wartość parametru wygładzania (3.370) oraz obliczone powyżej parametry modyfikujące i warunkujące, można ostatecznie uzyskać

$$\hat{F}_{X|W=4,5}(x) = \frac{1}{0,134+0,036+0,477} \cdot$$

$$\cdot \left[0,134 \frac{1}{\pi} \left(\frac{\frac{x-3}{0,831 \cdot 0,950}}{\left(\frac{x-3}{0,831 \cdot 0,950} \right)^2 + 1} + \operatorname{arctg}\left(\frac{x-3}{0,831 \cdot 0,950} \right) + \frac{\pi}{2} \right) + \right.$$

$$+ 0,036 \frac{1}{\pi} \left(\frac{\frac{x-3,3}{0,831 \cdot 0,953}}{\left(\frac{x-3,3}{0,831 \cdot 0,953} \right)^2 + 1} + \operatorname{arctg}\left(\frac{x-3,3}{0,831 \cdot 0,953} \right) + \frac{\pi}{2} \right) +$$

$$\left. + 0,477 \frac{1}{\pi} \left(\frac{\frac{x-5}{0,831 \cdot 1,110}}{\left(\frac{x-5}{0,831 \cdot 1,110} \right)^2 + 1} + \operatorname{arctg}\left(\frac{x-5}{0,831 \cdot 1,110} \right) + \frac{\pi}{2} \right) \right] \tag{3.376}$$

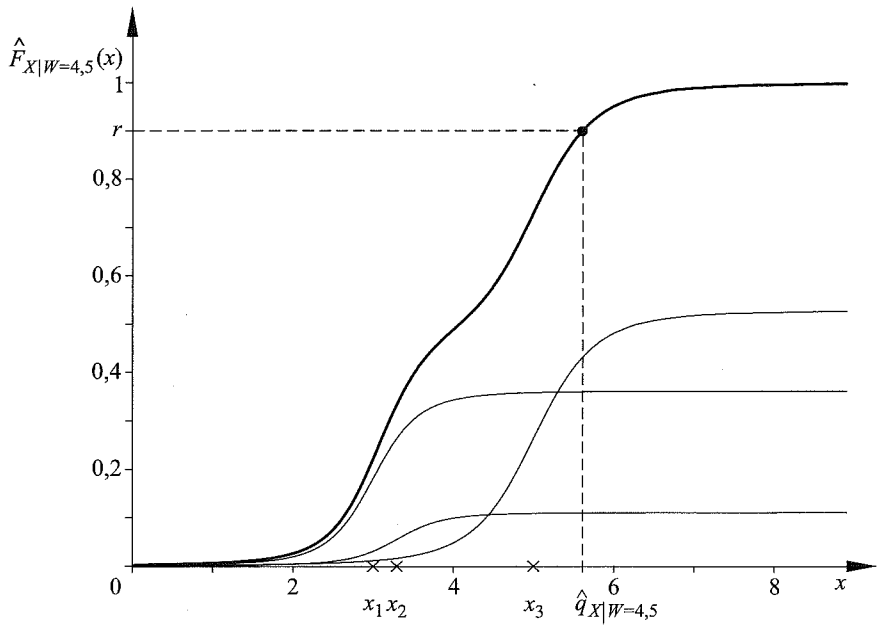
i równoważnie

$$\hat{F}_{X|W=4,5}(x) = 0,166 \left(\frac{\frac{x-3}{0,789}}{\left(\frac{x-3}{0,789} \right)^2 + 1} + \operatorname{arctg}\left(\frac{x-3}{0,789} \right) + \frac{\pi}{2} \right) +$$

$$+ 0,018 \left(\frac{\frac{x-3,3}{0,792}}{\left(\frac{x-3,3}{0,792} \right)^2 + 1} + \operatorname{arctg}\left(\frac{x-3,3}{0,792} \right) + \frac{\pi}{2} \right) +$$

$$+ 0,235 \left(\frac{\frac{x-5}{0,922}}{\left(\frac{x-5}{0,922} \right)^2 + 1} + \operatorname{arctg}\left(\frac{x-5}{0,922} \right) + \frac{\pi}{2} \right) \Bigg]. \tag{3.377}$$

Wyznaczony powyżej jądrowy estymator warunkowej dystrybuanty zmiennej losowej X przy wartości warunkującej (3.362) jest pokazany na rysunku 3.25. Podobnie jak podczas interpretacji wzoru (3.369), w porównaniu z rysunkiem 3.23 ilustrującym dystrybuantę bezwarunkową widać wyraźnie zwiększony wpływ funkcji pierwotnej jądra przyporządkowanego trzeciemu elementowi próby losowej, głównie kosztem owej funkcji związanej z drugim elementem próby.



Rys. 3.25. Jądrowy estymator warunkowej dystrybuanty (3.377); por. także rys. 3.23 (zaznaczono również jądrowy estymator warunkowego kwantyla rzędu $r = 0,9$; por. także rys. 3.23)

I wreszcie, obliczony teraz będzie **jądrowy estymator warunkowego kwantyla** rzędu $r = 0,9$. Jak poprzednio, zostało przyjęte jądro produktowe (3.48) z jednowymiarowym jądrem Cauchy'ego (3.288).

I tak, estymator warunkowej dystrybuanty $\hat{F}_{X|W=4,5}$ jest dany wzorem (3.377). Analogicznie jak na początku niniejszego przykładu wyznaczyć można estymator warunkowej gęstości prawdopodobieństwa (z użyciem jądra Cauchy'ego, dla identycznych parametrów jak w przypadku estymatora dystrybuanty), który przyjmuje następującą postać:

$$\hat{f}_{X|W=4,5}(x) = \frac{0,167}{\left[\left(\frac{x-3}{0,789}\right)^2 + 1\right]^2} + \frac{0,045}{\left[\left(\frac{x-3,3}{0,792}\right)^2 + 1\right]^2} + \frac{0,509}{\left[\left(\frac{x-5}{0,922}\right)^2 + 1\right]^2}. \quad (3.378)$$

Do obliczenia wartości estymatora warunkowego kwantyla można wtedy zastosować numeryczny algorytm Newtona, podobnie jak uczyniono to w przykładzie 3.7. W swej podstawowej postaci (3.338)-(3.340) okazuje się on jednak rozbieżny, jako że w wyniku jego pierwszych iteracji otrzymuje się kolejno $\hat{q}_0^* = 3{,}767$, $\hat{q}_1^* = 8{,}443$, $\hat{q}_2^* = -33{,}391$, $\hat{q}_3^* = 4170796$, ... Konieczne staje się zatem ograniczenie kroku, realizowane poprzez zastąpienie wzoru (3.340) zależnością (3.341). Po wykonaniu 21 iteracji (z czego 17 odbywało się z ograniczonym krokiem, po których algorytm znalazł się w obszarze lokalnej zbieżności) obliczony został – podobnie jak w przykładzie 3.7 – jądrowy estymator warunkowego kwantyla rzędu $r = 0{,}9$:

$$\hat{q}_{X|W=4,5} = 5{,}712. \tag{3.379}$$

Jest on zaznaczony na rysunku 3.25. Ze względu na widoczne tam zwiększenie wpływu jądra przyporządkowanego trzeciemu elementowi próby losowej ($x_3 = 5$) kosztem drugiego ($x_2 = 3{,}3$, a więc mniejszego niż x_3), wartość (3.379) jest większa od określonego za pomocą wzoru (3.358) kwantyla bezwarunkowego $\hat{q} = 5{,}371$.

Uwagi bibliograficzne

Podstawowe informacje dotyczące dystrybuanty, kwantyla i rozkładów warunkowych można znaleźć w podręcznikach [20, 34, 53], klasycznych monografiach [15, 18], a w formie zmatematyzowanej w książce [5]. Pomocna przy wyznaczaniu funkcji pierwotnych nietypowych jąder oraz całki występującej we wzorze (3.305) może okazać się praca [16], a przede wszystkim tablice [52, 58]. Dodatkowe wiadomości na temat numerycznej metody Newtona znajdują się w publikacjach [9, 57, 63].

Podobnie jak w poprzednim podrozdziale, wiele prezentowanych tu aspektów stanowi przedmiot szczegółowych badań prowadzonych w ramach prac [8, 46, 55, 66].

Rozdział 4

Przykładowe zastosowania w analizie systemowej

W niniejszym rozdziale zostaną zaprezentowane trzy przykładowe zastosowania estymatorów jądrowych w zadaniach aplikacyjnych z zakresu szeroko rozumianej analizy systemowej.

Najpierw, w podrozdziale 4.1 będzie rozważane zasadnicze zagadnienie identyfikacji – problem wyznaczania wartości parametrów modelu. Przedstawiona tu procedura jest przeznaczona dla tych przypadków, w których dodatnie i ujemne błędy estymacji skutkują odmiennymi zjawiskami, a w sensie ilościowym implikują różne straty.

Treść podrozdziału 4.2 stanowi opis systemu wykrywania i klasyfikacji elementów nietypowych. Pojawienie się takiego elementu wśród otrzymanych pomiarów może przykładowo świadczyć o zaistnieniu awarii urządzenia przemysłowego lub stanu patologicznego w organizmie pacjenta. Dodatkowo będzie tu rozważone zagadnienie lokalnego ujęcia transformacji liniowej estymatora jądrowego.

I wreszcie, w podrozdziale 4.3 estymatory jądrowe zostaną użyte do oszacowania przestrzennego rozkładu popytu na usługi bezprzewodowej transmisji danych, głównie związane z zapewnieniem dostępu do Internetu w aglomeracjach miejskich. Dotychczas stosowana w niniejszej monografii postać estymatorów jądrowych zostanie uzupełniona o elementy logiki rozmytej.

Zgodnie z przedmiotem współczesnej analizy systemowej powyższe zagadnienia reprezentują zadania aplikacyjne związane z szeroko rozumianym podejmowaniem decyzji w problemach praktycznych o różnorodnej specyfice. Poniżej zostaną przedstawione klasyczne elementy **statystycznej teorii decyzji**, które będą stanowić matematyczną podstawę rozważań zawartych w dalszej części tego rozdziału.

Zasadnicze zadanie statystycznej teorii decyzji polega na wyodrębnieniu ze zbioru wszystkich możliwych do podjęcia decyzji jednego, najkorzystniejszego w danym zagadnieniu elementu, jedynie na podstawie probabilistycznych infor-

macji o otaczającej rzeczywistości (naturze), w szczególności bez znajomości jej faktycznego stanu. Podstawowa formalizacja powyższej teorii opiera się na trzech pojęciach: zbiorze możliwych do podjęcia decyzji, zbiorze stanów natury oraz funkcji strat.

Zbiór możliwych do podjęcia decyzji $\mathcal{D}$ reprezentuje wszystkie decyzje, jakie ewentualnie mogą być podjęte po przeanalizowaniu danego zagadnienia. Zbiór ten nie może być oczywiście pusty, a w przypadku zbioru jednoelementowego problem decyzyjny staje się trywialny. Z kolei w przypadku zbioru dwuelementowego zadanie sprowadza się do konieczności wyboru w rodzaju: „zaakceptować lub odrzucić", „uznać za poprawny lub wadliwy" i temu podobne. Często zbiór możliwych do podjęcia decyzji jest tożsamy ze zbiorem liczb rzeczywistych, a zatem $\mathcal{D} = \mathbb{R}$. Podjęcie decyzji polega wówczas na wskazaniu konkretnej liczby, najkorzystniejszej z punktu widzenia rozpatrywanego zadania.

Zbiór stanów natury $\mathcal{X}$ reprezentuje z kolei wszystkie stany, jakie badany obiekt, proces lub zjawisko – ogólnie określane tutaj mianem „natury" – może przyjąć. Należy jednak podkreślić, że w chwili podejmowania decyzji faktycznie występujący stan nie jest znany. Zbiór $\mathcal{X}$ nie może być pusty. Jeśli zbiór stanów natury jest równy zbiorowi liczb rzeczywistych, czyli $\mathcal{X} = \mathbb{R}$, to przypadek taki oznacza, iż na potrzeby rozważanego zadania „natura" jest charakteryzowana przez pojedynczą liczbę (np. parametr), aczkolwiek o wartości nieznanej w chwili podejmowania decyzji.

Poniżej przyjmuje się zbiór stanów natury w postaci $\mathcal{X} = \mathbb{R}^n$ przy ustalonym $n \in \mathbb{N} \setminus \{0\}$. Poszczególne elementy tego zbioru są interpretowane jako wartości n-wymiarowej zmiennej losowej X o rozkładzie $\mathcal{P}$[1]. Zgodnie z przyjętym w punkcie 2.1.2 ogólnym założeniem rozkład ten ma gęstość f. Stanowić ona zatem będzie charakterystykę opisującą nieokreśloność natury.

I wreszcie, ostatnim elementem podstawowej formalizacji statystycznej teorii decyzji jest **funkcja strat** $l : \mathcal{D} \times \mathcal{X} \to \mathbb{R}$, która każdej parze $(d, x) \in \mathcal{D} \times \mathcal{X}$ przyporządkowuje wartość strat wynikłych z podjęcia decyzji d, podczas gdy w rzeczywistości występuje hipotetyczny stan x.

W praktycznych zagadnieniach klasyczne procedury, pozwalające wyodrębnić ze zbioru możliwych do podjęcia decyzji $\mathcal{D}$ element najkorzystniejszy z punktu widzenia rozważanego problemu, stanowią reguły minimaksowa i bayesowska. Ich zasadnicza koncepcja zostanie przedstawiona poniżej.

I tak, **funkcją strat minimaksowych** jest nazywane odwzorowanie $l_m : \mathcal{D} \to \mathbb{R}$ zdefiniowane wzorem

[1] Nawiązując do terminologii wprowadzonej w punkcie 2.1.1, oznacza to, iż zmienna losowa X indukuje przestrzeń probabilistyczną $(\mathcal{X}, \mathcal{B}^n, \mathcal{P})$.

$$l_m(d) = \sup_{x \in \mathrm{supp}(f)} l(d, x), \tag{4.1}$$

gdzie $\mathrm{supp}(f)$ oznacza nośnik gęstości f. Z kolei, jeżeli dla każdego $d \in \mathcal{D}$ istnieje całka $\int_{\mathbb{R}^n} l(d, x) f(x)\, dx$, to odwzorowanie $l_b : \mathcal{D} \to \mathbb{R}$ dane jako

$$l_b(d) = \int_{\mathbb{R}^n} l(d, x) f(x)\, dx \tag{4.2}$$

określa się mianem **funkcji strat bayesowskich**. Wtedy też każdy element $d_m \in \mathcal{D}$, taki że

$$l_m(d_m) = \inf_{d \in \mathcal{D}} l_m(d), \tag{4.3}$$

nosi nazwę **decyzji minimaksowej** i analogicznie dowolne $d_b \in \mathcal{D}$ spełniające warunek

$$l_b(d_b) = \inf_{d \in \mathcal{D}} l_b(d) \tag{4.4}$$

jest **decyzją bayesowską**, a powyższe procedury są nazywane, odpowiednio, **regułami minimaksową i bayesowską**.

Tak więc, jak wynika wprost z definicji (4.1), funkcja strat minimaksowych przyporządkowuje każdej możliwej do podjęcia decyzji największą wartość straty, jaka w praktyce może nastąpić po podjęciu tej decyzji. Jest to zatem ujęcie skrajnie pesymistyczne, zakładające wystąpienie najgorszego scenariusza. Zgodnie ze wzorem (4.3) reguła minimaksowa minimalizuje tak rozumiane straty.

Z kolei funkcja strat bayesowskich (4.2) przyporządkowuje każdej możliwej decyzji wartość oczekiwaną strat po podjęciu tej decyzji. Reguła bayesowska, która zgodnie ze wzorem (4.4) polega na przyjęciu elementu minimalizującego tę wartość, ma zatem charakter racjonalny.

Uwagi bibliograficzne

Ilustracyjną formułę prezentacji statystycznej teorii decyzji znaleźć można w podręczniku [43], natomiast w postaci matematycznie ścisłej zagadnienie to jest przedstawione w rozdziale 3 książki [4].

4.1. Identyfikacja wartości parametrów z niesymetryczną funkcją strat

Podstawowe znaczenie w wielu zagadnieniach współczesnej nauki i działań praktycznych odgrywa właściwa identyfikacja parametrów użytych do modelowania badanej rzeczywistości, na przykład obiektu, procesu lub zjawiska. Z oczywistych przyczyn metrologicznych, a czasem także technologicznych i ekonomicznych, nie jest jednak możliwy idealny pomiar ich wartości. W wielu zadaniach praktycznych niedoszacowanie wartości parametru implikuje odmienne skutki niż jego przeszacowanie, zarówno o jakościowym, jak i ilościowym charakterze. Zagadnienie identyfikacji tej wartości dla przypadku, gdy wynikłe stąd straty można opisać za pomocą funkcji liniowej lub kwadratowej i – co należy szczególnie podkreślić – przy różnych współczynnikach charakteryzujących błędy ujemne i dodatnie, będzie tematem niniejszego podrozdziału.

Poniżej zostanie zaproponowany algorytm oparty na bayesowskiej regule decyzyjnej, co pozwala uzyskać minimalną wartość oczekiwaną potencjalnych strat. Rozważone będą kolejno przypadki liniowej niesymetrycznej funkcji strat (punkt 4.1.1), kwadratowej niesymetrycznej funkcji strat dla pojedynczego parametru (punkt 4.1.2) i wektora identyfikowanych parametrów (punkt 4.1.3). Procedury przedstawione zostaną w postaci kompletnej, tak aby ich praktyczna implementacja nie wymagała od użytkownika prowadzenia dodatkowych badań i obliczeń.

4.1.1. Liniowa niesymetryczna funkcja strat

Rozważany parametr, którego wartość podlega oszacowaniu, będzie oznaczany przez x. W niniejszym punkcie jest rozpatrywany przypadek jednowymiarowy, a zatem $x \in \mathbb{R}$. Zakłada się, że dysponujemy uzyskanymi metrologicznie pomiarami tego parametru

$$x_1, x_2, \ldots, x_m, \tag{4.5}$$

traktowanymi jako m-elementowa próba losowa. Zgodnie z metodyką estymacji punktowej przyjmuje się, że pomiary te są sumą „prawdziwej" (aczkolwiek nieznanej) wartości badanego parametru oraz zakłóceń losowych różnorodnego pochodzenia. Celem badań jest wyznaczenie wartości estymatora tego parametru (oznaczanego dalej przez $\hat{x}$), która szacowałaby powyższą „prawdziwą" wartość parametru, najkorzystniej z punktu widzenia rozpatrywanego problemu.

W powyższym celu użyte zostaną przedstawione na początku niniejszego rozdziału elementy statystycznej teorii decyzji, a w szczególności reguła bayesowska. Aby dostosować się do stosowanej tu formalizacji, zbiory możliwych

do podjęcia decyzji i stanów natury zostaną przyjęte w postaci $\mathcal{D} = \mathbb{R}$ oraz $\mathcal{X} = \mathbb{R}$. Utożsamiając podjęcie decyzji z wyznaczeniem wartości estymatora, można także oznaczyć $d = \hat{x}$. Podobnie, wartości szacowanego parametru x będą traktowane jako wartości jednowymiarowej zmiennej losowej X (czyli $X(\omega) = x$), o rozkładzie mającym gęstość prawdopodobieństwa f. Gęstość ta reprezentuje więc nieokreśloność natury – w tym przypadku owego szacowanego parametru. I wreszcie, elementy zbiorów $\mathcal{D}$ oraz $\mathcal{X}$ będą oznaczane, odpowiednio, jako $\hat{x}$ oraz x, a zatem $\hat{x} \in \mathcal{D}$ i $x \in \mathcal{X}$.

Zgodnie z przyjętymi uwarunkowaniami rozpatrywanego w tym punkcie zagadnienia **funkcja strat** zostanie przyjęta w postaci **liniowej niesymetrycznej**:

$$l(\hat{x}, x) = \begin{cases} -p(\hat{x} - x), & \text{gdy } \hat{x} - x < 0, \\ 0, & \text{gdy } \hat{x} - x = 0, \\ q(\hat{x} - x), & \text{gdy } \hat{x} - x > 0, \end{cases} \tag{4.6}$$

przy czym współczynniki p i q są dodatnie oraz – co należy wyraźnie stwierdzić – mogą być różne. A zatem funkcja strat bayesowskich (4.2) wynosi w tym przypadku

$$l_b(\hat{x}) = -p \int_{\hat{x}}^{\infty} (\hat{x} - x) f(x)\, dx + q \int_{-\infty}^{\hat{x}} (\hat{x} - x) f(x)\, dx. \tag{4.7}$$

Nietrudno pokazać, iż funkcja ta osiąga swe minimum dla tej wartości, która stanowi rozwiązanie następującego równania z argumentem $\hat{x}$:

$$\int_{-\infty}^{\hat{x}} f(x)\, dx = \frac{p}{p+q}, \tag{4.8}$$

czyli jest kwantylem rzędu $p/(p+q)$ [1]. Jeżeli zatem gęstość f jest szacowana z użyciem estymatora jądrowego $\hat{f}$, to powyższe równanie przyjmuje postać

$$\int_{-\infty}^{\hat{x}} \hat{f}(x)\, dx = \frac{p}{p+q}, \tag{4.9}$$

a więc procedura wyznaczenia wartości poszukiwanego tu estymatora $\hat{x}$ jest tożsama z algorytmem obliczenia wartości jądrowego estymatora kwantyla rzędu $p/(p+q)$, przedstawionym szczegółowo w punkcie 3.2.2. Warto również zwrócić uwagę, że

[1] W przypadku symetrycznej funkcji strat, czyli gdy $p = q$, staje się ona medianą.

$$\frac{p}{p+q} = \frac{\frac{p}{q}}{\frac{p}{q}+1}, \tag{4.10}$$

z czego wynika, że nie jest konieczna osobna identyfikacja parametrów p oraz q, lecz jedynie ich ilorazu.

W wielu praktycznych zagadnieniach identyfikowany parametr może być istotnie zależny od czynnika zewnętrznego, w tym sensie iż podanie odpowiedniej (np. aktualnej) wartości tego czynnika znacząco uściśla stosowany model (np. w zagadnieniach technicznych jest nim często temperatura). Warto wówczas wykorzystać koncepcję jądrowego estymatora warunkowego kwantyla, przedstawioną pod koniec punktu 3.2.3. Zmienna warunkująca może także zawierać wielkości binarne (punkt 3.1.8).

W przypadku gdy zbiór wartości, jakie przyjmuje estymowany parametr, jest ograniczony od dołu, a wartość ułamka $p/(p+q)$ bliska zeru (w szczególności mniejsza od 0,1), wtedy należy przypuszczać, że wartość kwantyla jest zbliżona do wartości ograniczającej i wprowadzenie dolnego ograniczenia nośnika estymatora jądrowego może uściślić uzyskiwany wynik. Symetryczna sytuacja występuje dla ograniczenia górnego przy wartości ułamka $p/(p+q)$ zbliżonej do 1, w szczególności większej od 0,9. Stosowne procedury zostały przedstawione w punkcie 3.1.7.

Ostatecznie, przyjmując konkretną wartość ilorazu p/q oraz dysponując pomiarami (4.5), na podstawie przedstawionego powyżej materiału można uzyskać kompletny algorytm umożliwiający obliczenie wartości identyfikowanego parametru, optymalnej w sensie bayesowskiej reguły decyzyjnej dla liniowej niesymetrycznej funkcji strat (4.6). Uwzględniając materiał punktu 3.2.2 oraz ewentualnie 3.1.7, 3.1.8 i 3.2.3, podana została pełna procedura niezbędna do jego bezpośredniej implementacji numerycznej.

4.1.2. Kwadratowa niesymetryczna funkcja strat

Zagadnienie rozpatrywane w poprzednim punkcie będzie kontynuowane poniżej, aczkolwiek liniowa niesymetryczna **funkcja strat** (4.6) zostanie tu zastąpiona postacią **kwadratową niesymetryczną**:

$$l(\hat{x},x) = \begin{cases} p(\hat{x}-x)^2, & \text{gdy} \quad \hat{x}-x<0, \\ 0, & \text{gdy} \quad \hat{x}-x=0, \\ q(\hat{x}-x)^2, & \text{gdy} \quad \hat{x}-x>0, \end{cases} \tag{4.11}$$

gdzie współczynniki p oraz q są dodatnie i – ponownie – mogą być różne. Funkcja strat bayesowskich (4.2) przyjmuje w tym przypadku postać

$$l_b(\hat{x}) = p\int_{\hat{x}}^{\infty}(\hat{x}-x)^2 f(x)\,dx + q\int_{-\infty}^{\hat{x}}(\hat{x}-x)^2 f(x)\,dx \tag{4.12}$$

i osiąga swe minimum dla wartości stanowiącej rozwiązanie następującego równania z argumentem $\hat{x}$:

$$(p-q)\int_{-\infty}^{\hat{x}}(\hat{x}-x)f(x)\,dx - p\int_{-\infty}^{\infty}(\hat{x}-x)f(x)\,dx = 0. \tag{4.13}$$

Rozwiązanie to istnieje i jest jedyne. Dzieląc powyższe kryterium obustronnie przez q, uzyskuje się jego równoważną formułę

$$\left(\frac{p}{q}-1\right)\int_{-\infty}^{\hat{x}}(\hat{x}-x)f(x)\,dx = \frac{p}{q}\int_{-\infty}^{\infty}(\hat{x}-x)f(x)\,dx, \tag{4.14}$$

z której wynika, że nie jest konieczna osobna identyfikacja parametrów p oraz q, lecz jedynie ich ilorazu[1)].

Jeżeli do oszacowania gęstości f zostanie użyty estymator jądrowy $\hat{f}$, to równanie (4.13) przyjmuje postać

$$(p-q)\int_{-\infty}^{\hat{x}}(\hat{x}-x)\hat{f}(x)\,dx - p\int_{-\infty}^{\infty}(\hat{x}-x)\hat{f}(x)\,dx = 0. \tag{4.15}$$

W celu znalezienia jego rozwiązania można skonstruować efektywny algorytm numeryczny. Niech zatem dane będzie jednowymiarowe jądro, podczas gdy opisująca go funkcja K niech będzie ciągła i przyjmuje wartości dodatnie. Przy dowolnie ustalonym $i = 1, 2, ..., m$ definiuje się funkcje $U_i : \mathbb{R} \to \mathbb{R}$ i $V_i : \mathbb{R} \to \mathbb{R}$ określone wzorami

$$U_i(\hat{x}) = \frac{1}{hs_i}\int_{-\infty}^{\hat{x}} K\left(\frac{y-x_i}{hs_i}\right)dy, \tag{4.16}$$

$$V_i(\hat{x}) = \frac{1}{hs_i}\int_{-\infty}^{\hat{x}} yK\left(\frac{y-x_i}{hs_i}\right)dy. \tag{4.17}$$

[1)] W przypadku symetrycznej funkcji strat, czyli gdy $p = q$, zależność ta redukuje się do postaci $\hat{x} = \int_{-\infty}^{\infty} x f(x)\,dx$, a zatem rozwiązaniem kryterium (4.13) jest wartość oczekiwana (2.17).

Doboru postaci jądra K należy dokonać tak, aby odwzorowania dane zależnościami $I(x)=\int_{-\infty}^{x} K(y)\,dy$ i $J(x)=\int_{-\infty}^{x} yK(y)\,dy$ mogły być wyrażone wzorami analitycznymi, co skutkuje podobną własnością dotyczącą funkcji U_i oraz V_i.

W przypadku gdy wartość oczekiwaną w sposób naturalny szacuje się za pomocą średniej arytmetycznej wartości próby losowej, równanie (4.15) jest równoważne postaci

$$\sum_{i=1}^{m}[(p-q)(\hat{x}U_i(\hat{x})-V_i(\hat{x}))+px_i]-p\hat{x}m=0. \tag{4.18}$$

Jeżeli oznaczyć lewą stronę powyższego równania jako $L(\hat{x})$, to stosując wynikającą wprost z zależności (4.16)-(4.17) równość $V_i'(\hat{x})=\hat{x}U_i'(\hat{x})$, można napisać wzór na wartość pochodnej tej funkcji:

$$L'(\hat{x})=\sum_{i=1}^{m}[(p-q)U_i(\hat{x})]-pm. \tag{4.19}$$

Ponieważ, z uwagi na założoną dodatniość jądra, prawdziwa jest relacja $U_i(\hat{x})\in(0,1)$, więc także $L'(\hat{x})\in(-pm-qm,0)$. Wynika stąd, że pochodna ta jest różna od zera.

Rozwiązanie kryterium (4.15) może być w tej sytuacji wyznaczone numerycznie z użyciem metody Newtona jako granica ciągu $\{\hat{x}_k\}_{k=0}^{\infty}$ zdefiniowanego wzorami

$$\hat{x}_0=\frac{1}{m}\sum_{i=1}^{m}x_i, \tag{4.20}$$

$$\hat{x}_{k+1}=\hat{x}_k-\frac{L(\hat{x}_k)}{L'(\hat{x}_k)} \quad \text{dla } k=0,\,1,\,..., \tag{4.21}$$

przy czym funkcje L oraz L' określone są zależnościami (4.18)-(4.19), a warunek stopu przyjmuje postać

$$\left|\hat{x}_{k+1}-\hat{x}_k\right|\le\frac{\hat{\sigma}}{100}, \tag{4.22}$$

gdzie $\hat{\sigma}$ oznacza estymator odchylenia standardowego (2.48) wyznaczony na podstawie próby (4.5). Podobnie jak w przypadku procedury (3.314)-(3.316), wykorzystywanej do estymacji kwantyla, także powyższy algorytm nie musi być globalnie zbieżny i w celu wyeliminowania ewentualnej rozbieżności wzór (4.21) można zmodyfikować do postaci

$$\hat{x}_{k+1} = \begin{cases} \hat{x}_k - \dfrac{L(\hat{x}_k)}{L'(\hat{x}_k)}, & \text{gdy } \left|\dfrac{L(\hat{x}_k)}{L'(\hat{x}_k)}\right| \le \dfrac{\hat{\sigma}}{10}, \\ \hat{x}_k - \dfrac{\hat{\sigma}}{10}\operatorname{sgn}\left(\dfrac{L(\hat{x}_k)}{L'(\hat{x}_k)}\right), & \text{gdy } \left|\dfrac{L(\hat{x}_k)}{L'(\hat{x}_k)}\right| > \dfrac{\hat{\sigma}}{10}, \end{cases} \quad \text{dla } k = 0, 1, \ldots, \qquad (4.23)$$

co umożliwia w tym przypadku osiągnięcie obszaru lokalnej zbieżności algorytmu Newtona.

Jak było wielokrotnie wspominane, pozytywną – z punktu widzenia potencjalnych zastosowań – cechę estymatorów jądrowych stanowi możliwość wyboru postaci jądra w zależności od uwarunkowań konkretnego praktycznego zadania. Poza typowymi wymaganiami, sformułowanymi w punktach 3.1.1-3.1.2 wobec funkcji K, w powyższych rozważaniach zostały poczynione w tym zakresie następujące założenia:

A) funkcja K jest ciągła oraz dodatnia,

B) odwzorowania dane zależnościami $I(x) = \int_{-\infty}^{x} K(y)\,dy$ i $J(x) = \int_{-\infty}^{x} yK(y)\,dy$ powinny być wyrażane wzorami analitycznymi.

Optymalne jądro Epanecznikowa (3.40) nie spełnia warunku A, natomiast najczęściej stosowane jądro normalne (3.44) – warunku B. Można w tej sytuacji zaproponować jądro Cauchy'ego (3.288) lub ogólniej, w celu poprawienia efektywności, kombinację liniową jąder Epanecznikowa i Cauchy'ego określoną wzorem

$$K(x) = \alpha K_{\mathrm{E}}(x) + (1-\alpha)K_{\mathrm{C}}(x), \qquad (4.24)$$

przy czym $\alpha \in [0, 1)$, a K_{E} oznacza jądro Epanecznikowa (3.40) oraz K_{C} – jądro Cauchy'ego (3.288). Wartość $\alpha = 0$ oznacza nieco mniej efektywny – ze statystycznego punktu widzenia – ale prostszy w zastosowaniach przypadek wyłącznego użycia jądra Cauchy'ego. Jeżeli natomiast $\alpha = 0{,}9$, to jądro (4.24) ma efektywność bliską optymalnej i jest dodatnie, aczkolwiek taka wartość parametru α nie jest na tyle bliska jedności, aby w praktyce wprowadzić do obliczeń trudności wynikłe na przykład z występowania przybliżeń, nieodłącznie związanych przecież ze stosowaniem metod numerycznych. W ogólnym znaczeniu zagadnienie to było rozważane w punkcie 3.1.3 (por. wzór (3.46)).

Dla dowolnie ustalonego $i = 1, 2, \ldots, m$ funkcja U_i zdefiniowana wzorem (4.16) jest zatem kombinacją stosownych funkcji $U_{i,\mathrm{E}}$ oraz $U_{i,\mathrm{C}}$, odpowiednio dla jąder Epanecznikowa i Cauchy'ego:

$$U_i(\hat{x}) = \alpha U_{i,\mathrm{E}}(\hat{x}) + (1-\alpha)U_{i,\mathrm{C}}(\hat{x}), \qquad (4.25)$$

przy czym

$$U_{i,\mathrm{E}}(\hat{x}) = \begin{cases} 0, & \text{gdy} \quad \hat{x} < x_i - hs_i, \\ \dfrac{-\hat{x}^3 + 3x_i\hat{x}^2 + 3(h^2 s_i^2 - x_i^2)\hat{x} + x_i^3 + 2h^3 s_i^3 - 3h^2 s_i^2 x_i}{4(hs_i)^3}, & \\ & \text{gdy} \quad x_i - hs_i \le \hat{x} \le x_i + hs_i, \\ 1, & \text{gdy} \quad \hat{x} > x_i + hs_i, \end{cases} \tag{4.26}$$

$$U_{i,\mathrm{C}}(\hat{x}) = \frac{1}{\pi}\operatorname{arctg}\left(\frac{\hat{x} - x_i}{hs_i}\right) + \frac{\dfrac{\hat{x} - x_i}{hs_i}}{\pi\left[1 + \left(\dfrac{\hat{x} - x_i}{hs_i}\right)^2\right]} + \frac{1}{2}. \tag{4.27}$$

Analogicznie można zapisać w przypadku funkcji V_i danej przez (4.17):

$$V_i(\hat{x}) = \alpha V_{i,\mathrm{E}}(\hat{x}) + (1-\alpha)V_{i,\mathrm{C}}(\hat{x}), \tag{4.28}$$

gdzie

$$V_{i,\mathrm{E}}(\hat{x}) = \begin{cases} 0, & \text{gdy} \quad \hat{x} < x_i - hs_i, \\ \dfrac{-3\hat{x}^4 + 8x_i\hat{x}^3 + 6(h^2 s_i^2 - x_i^2)\hat{x}^2 + x_i^4 - 6h^2 s_i^2 x_i^2 + 8h^3 s_i^3 x_i - 3h^4 s_i^4}{16(hs_i)^3}, & \\ & \text{gdy} \quad x_i - hs_i \le \hat{x} \le x_i + hs_i, \\ x_i, & \text{gdy} \quad \hat{x} > x_i + hs_i, \end{cases} \tag{4.29}$$

$$V_{i,\mathrm{C}}(\hat{x}) = x_i\left(\frac{1}{\pi}\operatorname{arctg}\left(\frac{\hat{x} - x_i}{hs_i}\right) + \frac{\dfrac{\hat{x} - x_i}{hs_i}}{\pi\left[1 + \left(\dfrac{\hat{x} - x_i}{hs_i}\right)^2\right]} + \frac{1}{2}\right) - \frac{hs_i}{\pi\left[1 + \left(\dfrac{\hat{x} - x_i}{hs_i}\right)^2\right]}. \tag{4.30}$$

W przypadku gdy $\alpha \neq 0$, parametr wygładzania h oraz parametry modyfikujące s_1, s_2, ..., s_m powinny być obliczone odrębnie dla jąder Epanecznikowa i Cauchy'ego, zgodnie z procedurami przedstawionymi w punktach 3.1.5 oraz 3.1.6. W szczególności, dla przypadku jednowymiarowej zmiennej losowej, do wyznaczania parametru wygładzania jest tam zalecana metoda podstawień rzędu drugiego. Wartość ilorazu $W(K)/U(K)^2$ dla jądra Epanecznikowa zawiera rysunek 3.3 (z podstawieniem $n = 1$), natomiast dla jądra Cauchy'ego (3.288) może być uzyskana na podstawie wzorów (3.290)-(3.291). Wyczerpuje to ostatecznie wszystkie wielkości niezbędne do wyznaczenia wartości identyfikowanego parametru w rozpatrywanym tu przypadku kwadratowej niesymetrycznej funkcji strat.

Jeżeli identyfikowany parametr jest istotnie zależny od czynnika zewnętrznego, to można wtedy zastosować ujęcie warunkowe, analogicznie do procedury przedstawionej w punkcie 3.2.3 dla zagadnienia estymacji kwantyla, uwzględniając ewentualnie składowe binarne (punkt 3.1.8). Z kolei w przypadku gdy wartości estymowanego parametru są ograniczone od dołu (lub od góry), a wartość ułamka p/q jest mniejsza niż 0,1 (lub odpowiednio większa niż 10), wówczas należy przypuszczać, że wartość identyfikowanego parametru jest zbliżona do wartości ograniczającej i wprowadzenie lewostronnego (lub odpowiednio prawostronnego) ograniczenia nośnika estymatora jądrowego może uściślić uzyskiwany wynik. Stosowne procedury zostały opisane w punkcie 3.1.7.

Podsumowując: powyżej został przedstawiony kompletny algorytm, pozwalający obliczyć wartość identyfikowanego parametru, optymalnego w sensie bayesowskiej reguły decyzyjnej dla kwadratowej niesymetrycznej funkcji strat (4.11), z ewentualnym uwzględnieniem ograniczenia nośnika i ujęcia warunkowego, także w obecności składowych binarnych.

4.1.3. Kwadratowa niesymetryczna funkcja strat w przypadku dwuwymiarowym

Koncepcja przedstawiona w poprzednim punkcie zostanie teraz uogólniona do zagadnienia dwuwymiarowego, czyli do przypadku gdy oszacowaniu są poddawane dwa parametry. Sama idea może być uogólniona na większą ich ilość, aczkolwiek kosztem naturalnego w takiej sytuacji zwiększenia złożoności problemu.

Niech zatem szacowane parametry $x_1, x_2 \in \mathbb{R} \times \mathbb{R}$ będą traktowane jako dwuwymiarowy wektor $x = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$, podobnie jak ich estymatory $\hat{x} = \begin{bmatrix} \hat{x}_1 \\ \hat{x}_2 \end{bmatrix}$. Kontynuując notację wprowadzoną na początku rozdziału 3, x_1 oraz x_2 oznaczają

współrzędne wektora x i są one składane odmienną czcionką dla odróżnienia od elementów próby losowej. Nawiązując do formalizacji statystycznej teorii decyzji, można zapisać, iż w tym przypadku $\mathcal{D} = \mathbb{R}^2$ oraz $\mathcal{X} = \mathbb{R}^2$, a także $\hat{x} \in \mathcal{D}$ i $x \in \mathcal{X}$. Podjęcie decyzji również jest tu tożsame z wyznaczeniem wartości estymatora, czyli $d = \begin{bmatrix} d_1 \\ d_2 \end{bmatrix} = \begin{bmatrix} \hat{x}_1 \\ \hat{x}_2 \end{bmatrix}$, natomiast wartości szacowanego parametru są traktowane jako wartości dwuwymiarowej zmiennej losowej, a zatem $X(\omega) = \begin{bmatrix} X_1(\omega) \\ X_2(\omega) \end{bmatrix} = \begin{bmatrix} .x_1 \\ .x_2 \end{bmatrix}$, o rozkładzie mającym gęstość $f: \mathbb{R}^2 \to [0, \infty)$, reprezentującą nieokreśloność szacowanych parametrów.

Zakłada się, iż składowe powyższej zmiennej losowej, czyli X_1 i X_2, są niezależne, aczkolwiek straty wynikłe z błędów estymacji parametrów $.x_1$ i $.x_2$ są wzajemnie skorelowane. Jako ilustrację takiej sytuacji nietrudno wyobrazić sobie, że szacowane parametry opisują wielkości pochodzące z różnych niezależnych podukładów, ale skutki wynikłe z błędów ich estymacji w ramach całego systemu są skorelowane, gdyż mogą się częściowo znosić lub przeciwnie – kumulować. **Dwuwymiarowa, kwadratowa niesymetryczna funkcja strat** będzie zatem dana następującą zależnością:

$$l\left(\begin{bmatrix} \hat{x}_1 \\ \hat{x}_2 \end{bmatrix}, \begin{bmatrix} .x_1 \\ .x_2 \end{bmatrix}\right) = \begin{cases} a_p(\hat{x}_1 - .x_1)^2 + a_{pg}(\hat{x}_1 - .x_1)(\hat{x}_2 - .x_2) + a_g(\hat{x}_2 - .x_2)^2, \\ \qquad \text{gdy} \quad \hat{x}_1 - .x_1 \geq 0 \ \text{ i } \ \hat{x}_2 - .x_2 \geq 0, \\ a_l(\hat{x}_1 - .x_1)^2 + a_{lg}(\hat{x}_1 - .x_1)(\hat{x}_2 - .x_2) + a_g(\hat{x}_2 - .x_2)^2, \\ \qquad \text{gdy} \quad \hat{x}_1 - .x_1 \leq 0 \ \text{ i } \ \hat{x}_2 - .x_2 \geq 0, \\ a_l(\hat{x}_1 - .x_1)^2 + a_{ld}(\hat{x}_1 - .x_1)(\hat{x}_2 - .x_2) + a_d(\hat{x}_2 - .x_2)^2, \\ \qquad \text{gdy} \quad \hat{x}_1 - .x_1 \leq 0 \ \text{ i } \ \hat{x}_2 - .x_2 \leq 0, \\ a_p(\hat{x}_1 - .x_1)^2 + a_{pd}(\hat{x}_1 - .x_1)(\hat{x}_2 - .x_2) + a_d(\hat{x}_2 - .x_2)^2, \\ \qquad \text{gdy} \quad \hat{x}_1 - .x_1 \geq 0 \ \text{ i } \ \hat{x}_2 - .x_2 \leq 0, \end{cases} \tag{4.31}$$

przy czym a_p, a_l, a_g, $a_d > 0$, a_{pg}, $a_{ld} \geq 0$ oraz a_{lg}, $a_{pd} \leq 0$. Współczynniki mieszane a_{pg}, a_{lg}, a_{ld}, a_{pd} reprezentują wzajemną korelację błędów estymacji obu parametrów. Warto także zauważyć, iż w przypadku gdy są one równe zeru, wtedy problem redukuje się do dwóch oddzielnych zagadnień rozważanych w poprzednim punkcie.

Ze względu na założoną niezależność zmiennych losowych X_1 i X_2, reprezentujących szacowane parametry, gęstość prawdopodobieństwa f opisująca ich

nieokreśloność może być przedstawiona jako iloczyn jednowymiarowych gęstości $f_1 : \mathbb{R} \to [0, \infty)$ oraz $f_2 : \mathbb{R} \to [0, \infty)$ odpowiadających poszczególnym składowym (por. wzór (2.34)), czyli

$$f(x) = f\left(\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}\right) = f_1(x_1) f_2(x_2). \tag{4.32}$$

Po uwzględnieniu zależności (4.31)-(4.32) funkcja strat bayesowskich (4.2) osiąga minimum w punkcie stanowiącym rozwiązanie następującego układu dwóch równań z argumentami $\hat{x}_1$ i $\hat{x}_2$:

$$\begin{aligned}
&\int_{-\infty}^{\hat{x}_1} f_1(x_1)\, dx_1 \left[(a_{pg} + a_{ld} - a_{lg} - a_{pd}) \left(\hat{x}_2 \int_{-\infty}^{\hat{x}_2} f_2(x_2)\, dx_2 - \int_{-\infty}^{\hat{x}_2} x_2 f_2(x_2)\, dx_2 \right) + \right. \\
&\left. + (a_{pd} - a_{ld}) \left(\hat{x}_2 - \int_{-\infty}^{\infty} x_2 f_2(x_2)\, dx_2 \right) \right] + 2a_l \left(\hat{x}_1 - \int_{-\infty}^{\infty} x_1 f_1(x_1)\, dx_1 \right) + \\
&+ 2(a_p - a_l) \left(\hat{x}_1 \int_{-\infty}^{\hat{x}_1} f_1(x_1)\, dx_1 - \int_{-\infty}^{\hat{x}_1} x_1 f_1(x_1)\, dx_1 \right) + a_{ld} \left(\hat{x}_2 - \int_{-\infty}^{\infty} x_2 f_2(x_2)\, dx_2 \right) + \\
&+ (a_{lg} - a_{ld}) \left(\hat{x}_2 \int_{-\infty}^{\hat{x}_2} f_2(x_2)\, dx_2 - \int_{-\infty}^{\hat{x}_2} x_2 f_2(x_2)\, dx_2 \right) = 0
\end{aligned} \tag{4.33}$$

oraz

$$\begin{aligned}
&\int_{-\infty}^{\hat{x}_2} f_2(x_2)\, dx_2 \left[(a_{pg} + a_{ld} - a_{lg} - a_{pd}) \left(\hat{x}_1 \int_{-\infty}^{\hat{x}_1} f_1(x_1)\, dx_1 - \int_{-\infty}^{\hat{x}_1} x_1 f_1(x_1)\, dx_1 \right) + \right. \\
&\left. + (a_{lg} - a_{ld}) \left(\hat{x}_1 - \int_{-\infty}^{\infty} x_1 f_1(x_1)\, dx_1 \right) \right] + 2a_d \left(\hat{x}_2 - \int_{-\infty}^{\infty} x_2 f_2(x_2)\, dx_2 \right) + \\
&+ 2(a_g - a_d) \left(\hat{x}_2 \int_{-\infty}^{\hat{x}_2} f_2(x_2)\, dx_2 - \int_{-\infty}^{\hat{x}_2} x_2 f_2(x_2)\, dx_2 \right) + a_{ld} \left(\hat{x}_1 - \int_{-\infty}^{\infty} x_1 f_1(x_1)\, dx_1 \right) + \\
&+ (a_{pd} - a_{ld}) \left(\hat{x}_1 \int_{-\infty}^{\hat{x}_1} f_1(x_1)\, dx_1 - \int_{-\infty}^{\hat{x}_1} x_1 f_1(x_1)\, dx_1 \right) = 0.
\end{aligned} \tag{4.34}$$

Rozwiązanie to istnieje i jest jedyne. Powyższy układ stanowi dwuwymiarowy odpowiednik kryterium (4.13).

Do wyznaczenia gęstości f_1 oraz f_2 zostaną zastosowane estymatory jądrowe, odpowiednio, $\hat{f}_1$ oraz $\hat{f}_2$, z użyciem jądra K spełniającego warunki A-B sformułowane w punkcie 4.1.2. Rozwiązanie układu równań (4.33)-(4.34) może być wówczas obliczone za pomocą efektywnych metod numerycznych. W tym celu, przy dowolnie ustalonym $i = 1, 2, ..., m$, zostaną określone funkcje $U_{i,1} : \mathbb{R} \to \mathbb{R}$, $U_{i,2} : \mathbb{R} \to \mathbb{R}$, $V_{i,1} : \mathbb{R} \to \mathbb{R}$ oraz $V_{i,2} : \mathbb{R} \to \mathbb{R}$ dane wzorami (4.16)-(4.17), gdzie jako $\hat{x}$, x_i, h, s_i należy podstawić odpowiednio wielkości związane z pierwszą i drugą współrzędną, czyli $\hat{x}_1$ $x_{i,1}$, h_1, $s_{i,1}$ w przypadku funkcji $U_{i,1}$ oraz $V_{i,1}$, a także $\hat{x}_2$, $x_{i,2}$, h_2, $s_{i,2}$ dla $U_{i,2}$ i $V_{i,2}$. Po wprowadzeniu powyższych oznaczeń kryteria (4.33)-(4.34) można zapisać w następującej postaci:

$$\sum_{i=1}^{m}\Big[U_{i,1}(\hat{x}_1)\big[(a_{pg}+a_{ld}-a_{lg}-a_{pd})\big(\hat{x}_2U_{i,2}(\hat{x}_2)-V_{i,2}(\hat{x}_2)\big)+(a_{pd}-a_{ld})(\hat{x}_2-x_{i,2})\big]+$$
$$+2a_l(\hat{x}_1-x_{i,1})+2(a_p-a_l)\big(\hat{x}_1U_{i,1}(\hat{x}_1)-V_{i,1}(\hat{x}_1)\big)+a_{ld}(\hat{x}_2-x_{i,2})+$$
$$+(a_{lg}-a_{ld})\big(\hat{x}_2U_{i,2}(\hat{x}_2)-V_{i,2}(\hat{x}_2)\big)\Big]=0, \tag{4.35}$$

$$\sum_{i=1}^{m}\Big[U_{i,2}(\hat{x}_2)\big[(a_{pg}+a_{ld}-a_{lg}-a_{pd})\big(\hat{x}_1U_{i,1}(\hat{x}_1)-V_{i,1}(\hat{x}_1)\big)+(a_{lg}-a_{ld})(\hat{x}_1-x_{i,1})\big]+$$
$$+2a_d(\hat{x}_2-x_{i,2})+2(a_g-a_d)\big(\hat{x}_2U_{i,2}(\hat{x}_2)-V_{i,2}(\hat{x}_2)\big)+a_{ld}(\hat{x}_1-x_{i,1})+$$
$$+(a_{pd}-a_{ld})\big(\hat{x}_1U_{i,1}(\hat{x}_1)-V_{i,1}(\hat{x}_1)\big)\Big]=0. \tag{4.36}$$

Jeżeli oznaczyć lewe strony powyższych równań, odpowiednio, przez $L_1(\hat{x}_1,\hat{x}_2)$ oraz $L_2(\hat{x}_1,\hat{x}_2)$, to wartości pochodnych cząstkowych funkcji L_1 i L_2 względem, odpowiednio, $\hat{x}_1$ i $\hat{x}_2$ wynoszą

$$\frac{\partial L_1(\hat{x}_1,\hat{x}_2)}{\partial \hat{x}_1}=\sum_{i=1}^{m}\Bigg[\frac{1}{h_1s_{i,1}}K\Big(\frac{\hat{x}_1-x_{i,1}}{h_1s_{i,1}}\Big)\Big[(a_{pg}+a_{ld}-a_{lg}-a_{pd})\big(\hat{x}_2U_{i,2}(\hat{x}_2)-V_{i,2}(\hat{x}_2)\big)+$$
$$+(a_{pd}-a_{ld})(\hat{x}_2-x_{i,2})\Big]+2(a_p-a_l)U_{i,1}(\hat{x}_1)+2a_l\Bigg], \tag{4.37}$$

$$\frac{\partial L_2(\hat{x}_1,\hat{x}_2)}{\partial \hat{x}_2}=\sum_{i=1}^{m}\Bigg[\frac{1}{h_2s_{i,2}}K\Big(\frac{\hat{x}_2-x_{i,2}}{h_2s_{i,2}}\Big)\Big[(a_{pg}+a_{ld}-a_{lg}-a_{pd})\big(\hat{x}_1U_{i,1}(\hat{x}_1)-V_{i,1}(\hat{x}_1)\big)+$$
$$+(a_{lg}-a_{ld})(\hat{x}_1-x_{i,1})\Big]+2(a_g-a_d)U_{i,2}(\hat{x}_2)+2a_d\Bigg]. \tag{4.38}$$

Wtedy rozwiązanie układu równań (4.33)-(4.34) może być wyznaczone z użyciem metody Newtona jako granica dwuwymiarowego ciągu $\{\hat{x}_k\}_{k=0}^{\infty} = \left\{ \begin{matrix} \hat{x}_{k,1} \\ \hat{x}_{k,2} \end{matrix} \right\}_{k=0}^{\infty}$ zdefiniowanego wzorami

$$\hat{x}_{0,1} = \frac{1}{m} \sum_{i=1}^{m} x_{i,1}, \tag{4.39}$$

$$\hat{x}_{0,2} = \frac{1}{m} \sum_{i=1}^{m} x_{i,2}, \tag{4.40}$$

$$\hat{x}_{k+1,1} = \hat{x}_{k,1} - \frac{L_1(\hat{x}_{k,1}, \hat{x}_{k,2})}{\frac{\partial L_1(\hat{x}_{k,1}, \hat{x}_{k,2})}{\partial \hat{x}_1}} \quad \text{dla } k = 0, 1, ..., \tag{4.41}$$

$$\hat{x}_{k+1,2} = \hat{x}_{k,2} - \frac{L_2(\hat{x}_{k,1}, \hat{x}_{k,2})}{\frac{\partial L_2(\hat{x}_{k,1}, \hat{x}_{k,2})}{\partial \hat{x}_2}} \quad \text{dla } k = 0, 1, ..., \tag{4.42}$$

przy czym występujące powyżej wielkości są określone zależnościami (4.35)-(4.38), a warunek stopu jest koniunkcją poniższych nierówności:

$$| \hat{x}_{k,1} - \hat{x}_{k-1,1} | \leq \frac{\hat{\sigma}_1}{100}, \tag{4.43}$$

$$| \hat{x}_{k,2} - \hat{x}_{k-1,2} | \leq \frac{\hat{\sigma}_2}{100}, \tag{4.44}$$

gdzie $\hat{\sigma}_1$ i $\hat{\sigma}_2$ oznaczają estymatory odchyleń standardowych (2.53) poszczególnych współrzędnych. Dla zapewnienia globalnej zbieżności powyższego algorytmu możliwa jest stosowna modyfikacja wzorów (4.41)-(4.42) w celu uzyskania postaci analogicznej do zależności (4.23), z podstawieniem w miejsce $\hat{x}_{k+1}$, $\hat{x}_k$, L, L', $\hat{\sigma}$ odpowiednio $\hat{x}_{k+1,1}$, $\hat{x}_{k,1}$, L_1, $\frac{\partial L_1}{\partial \hat{x}_1}$, $\hat{\sigma}_1$ oraz $\hat{x}_{k+1,2}$, $\hat{x}_{k,2}$, L_2, $\frac{\partial L_2}{\partial \hat{x}_2}$, $\hat{\sigma}_2$.

Podobnie jak w poprzednich punktach, jako jądro K ponownie proponuje się kombinację liniową jąder Epanecznikowa i Cauchy'ego (4.24). Wymagane jest tu wyznaczenie wartości parametru wygładzania oraz parametrów modyfikują-

cych, osobno dla obu współrzędnych, a w przypadku $\alpha \neq 0$, także odrębnie dla jąder Epanecznikowa i Cauchy'ego. Odpowiednie wzory na wartości funkcji $U_{i,1}$, $U_{i,2}$ oraz $V_{i,1}$ i $V_{i,2}$ są podane jako zależności (4.25)-(4.30), które należy teraz stosować odpowiednio dla współrzędnej pierwszej i drugiej. Podobnie jak w poprzednich punktach, można uwzględnić ograniczenie nośnika poszczególnych współrzędnych (punkt 3.1.7) oraz ujęcie warunkowe (punkt 3.2.3), także w obecności składowych binarnych (3.1.8).

4.1.4. Wykorzystanie do zagadnień sterowania optymalnego

Procedura zaprezentowana w niniejszym podrozdziale ma charakter uniwersalny i może być stosowana do różnorodnych problemów z zakresu nauki i praktyki. Jako ilustracja jej użycia zostaną poniżej rozważone zagadnienia z zakresu klasycznego sterowania optymalnego. Dzięki występującemu tu wskaźnikowi jakości zagadnienia takie stają się bowiem dogodne do przejrzystych interpretacji. Materiał zawarty w tym punkcie nawiązuje do treści podrozdziału 2.3.

Ostatnią uniwersalną koncepcją, jaka powstała w dziedzinie teorii sterowania, było sterowanie optymalne. Konstruowane na tej podstawie systemy okazały się wszakże w praktyce bardzo wrażliwe na niedokładności modelu i zakłócenia pojawiające się w rzeczywistych obiektach, co stanowiło ostatecznie główne ograniczenie ich zastosowań. Jednak same wynikłe stąd koncepcje stały się dogodną podstawą do tworzenia struktur, w których wrażliwość taka byłaby znacząco zredukowana. W szczególności, występujący tu wskaźnik jakości sterowania może stanowić również o jakości identyfikacji, umożliwiając stworzenie procedury optymalnej estymacji wartości parametrów modelu, istotnie poprawiających własności otrzymanego układu.

W ramach **pierwszego przykładu** (dotyczącego przypadku liniowej niesymetrycznej funkcji strat – punkt 4.1.1) niech dany będzie system dynamiczny opisywany równaniem różniczkowym

$$\begin{bmatrix} \dot{x}_1(t) \\ \dot{x}_2(t) \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} x_1(t) \\ x_2(t) \end{bmatrix} + \begin{bmatrix} 0 \\ \frac{1}{m} \end{bmatrix} u(t), \qquad \begin{bmatrix} x_1(0) \\ x_2(0) \end{bmatrix} = \begin{bmatrix} x_{01} \\ x_{02} \end{bmatrix}, \tag{4.45}$$

przy czym $m > 0$ i $x_{01}, x_{02} \in \mathbb{R}$. Reprezentuje on masę m poddaną działaniu siły zgodnie z drugą zasadą dynamiki Newtona. Wówczas x_1, x_2 oraz u oznaczają, odpowiednio, położenie i prędkość masy oraz siłę traktowaną jako sterowanie. System taki stanowi podstawę większości badań z zakresu robotyki, prowadzących w konsekwencji do znacznie bardziej złożonych modeli, o specyfice właściwej dla rozpatrywanego problemu.

Rozważmy zagadnienie minimalnoczasowe, w którym wskaźnikiem jakości sterowania J jest czas osiągnięcia zbioru docelowego, a zatem

$$J(u) = t_k, \tag{4.46}$$

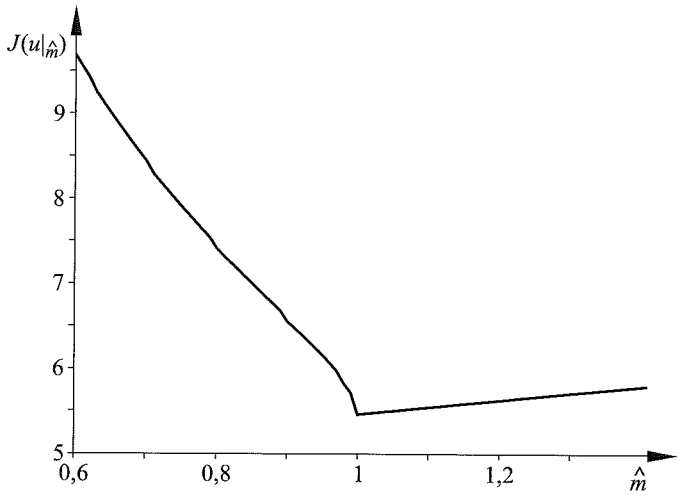
gdzie t_k oznacza (swobodny) czas osiągnięcia zbioru docelowego. Podstawowa postać takiego zadania polega na sprowadzeniu stanu systemu do początku układu współrzędnych, w minimalnym i skończonym czasie, przy czym zakłada się, iż wartości sterowania są ograniczone. Fundamentalne znaczenie dla zjawisk zachodzących w systemie sterowania ma właściwa identyfikacja wartości parametru m. Sterowanie jest przecież określane na podstawie wartości estymatora $\hat{m}$, w praktyce różnej od wartości parametru m występującego w obiekcie.

W czysto hipotetycznym przypadku $\hat{m} = m$, a zatem gdy wartość estymatora tego parametru jest równa jego rzeczywistej wartości, proces ma regularny charakter. Stan systemu osiąga początek układu współrzędnych w minimalnym i skończonym czasie. Równanie różniczkowe (4.45) ma w tym przypadku C-rozwiązania (por. punkt 2.3.1).

Z kolei, jeśli estymator jest przeszacowany (czyli $\hat{m} < m$), to w systemie pojawiają się przeregulowania – jego stan oscyluje wokół początku układu współrzędnych i osiąga go w skończonym czasie, aczkolwiek większym od minimalnego. Równanie różniczkowe (4.45) także wtedy ma C-rozwiązania.

I wreszcie, w przypadku niedoszacowania (tzn. gdy $\hat{m} > m$) stan systemu porusza się wzdłuż tak zwanej trajektorii poślizgowej i ostatecznie osiąga początek układu współrzędnych w skończonym czasie, ponownie większym od minimalnego. Równanie różniczkowe (4.45) nie ma tu jednak C-rozwiązań, lecz F-rozwiązania (por. punkt 2.3.1).

Jeżeli zatem do analizy równania różniczkowego (4.45) zostaną użyte K-rozwiązania (por. ponownie punkt 2.3.1) stanowiące uogólnienie zarówno C-, jak i F-rozwiązań, to możliwe jest formalnie jednolite zestawienie powyższych trzech przypadków. Na rysunku 4.1 został pokazany wykres wskaźnika jakości J w funkcji wartości estymatora $\hat{m}$ przy przykładowo ustalonym $m = 1$. Widać, iż wzrost wartości tego wskaźnika jest w przybliżeniu wprost proporcjonalny do błędu estymacji $|\hat{m} - m|$, aczkolwiek o odmiennych współczynnikach dla błędów dodatnich i ujemnych. A zatem do wyznaczenia wartości estymatora $\hat{m}$ warto użyć procedury proponowanej w punkcie 4.1.1 z użyciem liniowej niesymetrycznej funkcji strat (4.6). Korzyści uzyskiwane ze stosowania powyższej procedury, w porównaniu z estymatorem klasycznie otrzymywanym jako średnia arytmetyczna z pozyskanych pomiarów, wynoszą do 15% wartości wskaźnika jakości.



Rys. 4.1. Wartość wskaźnika jakości J otrzymana dla różnych wartości estymatora $\hat{m}$, przy $m = 1$

Jako **drugi przykład** (dotyczący przypadku kwadratowej niesymetrycznej funkcji strat – punkt 4.1.2) rozważone zostanie zagadnienie sterowania optymalnego z kwadratowym wskaźnikiem jakości. Obiektem niech będzie tu następujący typowy system, którego dynamika jest opisana równaniem różniczkowym z macierzą stanu w postaci kanonicznej Jordana:

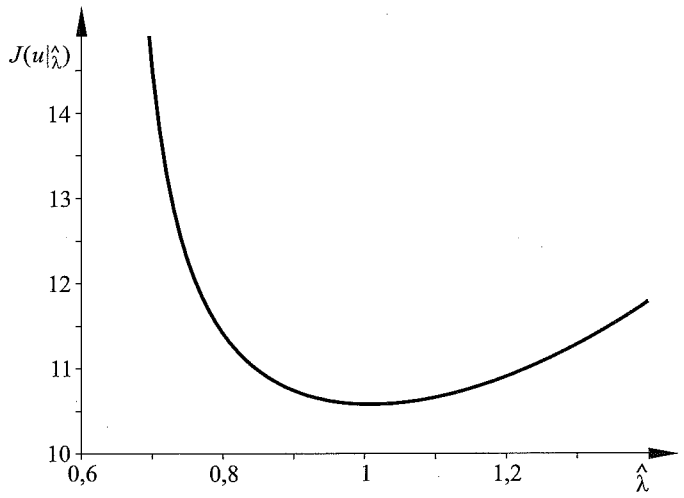
$$\begin{bmatrix} \dot{x}_1(t) \\ \dot{x}_2(t) \end{bmatrix} = \begin{bmatrix} \lambda & 1 \\ 0 & \lambda \end{bmatrix} \begin{bmatrix} x_1(t) \\ x_2(t) \end{bmatrix} + \begin{bmatrix} 0 \\ \lambda \end{bmatrix} u(t), \qquad \begin{bmatrix} x_1(0) \\ x_2(0) \end{bmatrix} = \begin{bmatrix} x_{01} \\ x_{02} \end{bmatrix}, \tag{4.47}$$

przy czym $x_{01}, x_{02} \in \mathbb{R}$ oraz $\lambda \in \mathbb{R} \setminus \{0\}$. Niech ponadto $\hat{\lambda} \in \mathbb{R} \setminus \{0\}$ reprezentuje estymator parametru λ. Regulator optymalny jest określony na podstawie wartości $\hat{\lambda}$, w praktyce różnej od wartości λ występującej w obiekcie. Dla prostoty ilustracji przyjęto przypadek nieskończonego czasu końcowego z jednostkowymi macierzami (parametrem) podcałkowej funkcji wskaźnika jakości, czyli podlegający minimalizacji wskaźnik J przyjmuje tu postać

$$J(u) = \int_0^\infty x_1^2(t) + x_2^2(t) + u^2(t)\, dt. \tag{4.48}$$

Otrzymane dla poszczególnych $\hat{\lambda}$ wartości wskaźnika J, przy przykładowo ustalonym $\lambda = 1$, zostały pokazane na rysunku 4.2. Widać, iż wzrost wskaźnika jakości w zależności od błędu estymacji można z dużą dokładnością aproksymować kwadratową niesymetryczną funkcją strat postaci (4.11), a do wyznaczenia wartości estymatora zastosować procedurę szczegółowo przedstawioną

w punkcie 4.1.2. Korzyści uzyskiwane względem klasycznie używanej średniej arytmetycznej wynosiły nawet do 50% kwadratowego wskaźnika jakości. Dodatkowo nastąpiło istotne zwiększenie zapasu stabilności.



Rys. 4.2. Wartość wskaźnika jakości J otrzymana dla różnych wartości estymatora $\hat{\lambda}$, przy $\lambda = 1$

I wreszcie w ramach **trzeciego przykładu** (dotyczącego przypadku dwuwymiarowej kwadratowej niesymetrycznej funkcji strat – punkt 4.1.3) zostanie rozważony system dynamiczny powstały poprzez uwzględnienie w układzie (4.47) inercji elementu wykonawczego, a zatem zastąpienie sterowania u dodatkową współrzędną x_3 i dodanie trzeciego równania różniczkowego w postaci

$$\dot{x}_3(t) = -Tx_3(t) + u(t), \tag{4.49}$$

przy stałej inercji $T > 0$. Łącząc zatem wzory (4.47) oraz (4.49), otrzymuje się

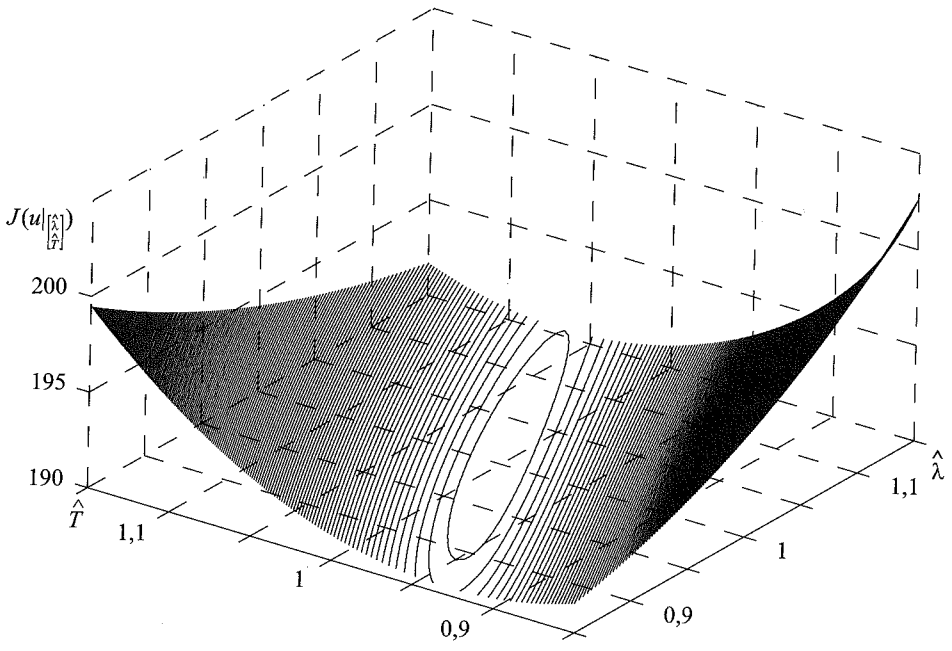
$$\begin{bmatrix} \dot{x}_1(t) \\ \dot{x}_2(t) \\ \dot{x}_3(t) \end{bmatrix} = \begin{bmatrix} \lambda & 1 & 0 \\ 0 & \lambda & \lambda \\ 0 & 0 & -T \end{bmatrix} \begin{bmatrix} x_1(t) \\ x_2(t) \\ x_3(t) \end{bmatrix} + \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} u(t), \qquad \begin{bmatrix} x_1(0) \\ x_2(0) \\ x_3(0) \end{bmatrix} = \begin{bmatrix} x_{01} \\ x_{02} \\ x_{03} \end{bmatrix}, \tag{4.50}$$

przy czym $x_{01}, x_{02}, x_{03} \in \mathbb{R}$. Estymacji podlegają zatem parametry $\begin{bmatrix} \lambda \\ T \end{bmatrix}$ pochodzące od odmiennych podukładów: λ z samego obiektu oraz T charakteryzujące element wykonawczy, a więc są one niezależne. Nadal rozważane będzie zadanie sterowania optymalnego z nieskończonym czasem końcowym przy jednostkowych macierzach (parametrze) funkcji podcałkowej, czyli dla następującego podlegającego minimalizacji wskaźnika jakości J:

$$J(u) = \int_0^{\infty} x_1^2(t) + x_2^2(t) + x_3^2(t) + u^2(t)\, dt. \tag{4.51}$$

Regulator optymalny jest określony na podstawie wartości estymatorów $\begin{bmatrix} \hat{\lambda} \\ \hat{T} \end{bmatrix}$, które różnią się od rzeczywistych wartości parametrów $\begin{bmatrix} \lambda \\ T \end{bmatrix}$.

Na rysunku 4.3 został pokazany przestrzenny wykres charakteryzujący wykres wskaźnika jakości sterowania w funkcji wartości estymatorów $\begin{bmatrix} \hat{\lambda} \\ \hat{T} \end{bmatrix}$, przy przykładowo ustalonym $\begin{bmatrix} \lambda \\ T \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$. Można go aproksymować dwuwymiarową, kwadratową niesymetryczną funkcją strat (4.31), po czym do wyznaczenia wartości estymatora wykorzystać procedurę z punktu 4.1.3. Uzyskiwane korzyści względem metod klasycznych opartych na średniej arytmetycznej pomiarów były podobne jak wspomniane wyżej dla przypadku jednowymiarowego.



Rys. 4.3. Wartość wskaźnika jakości J otrzymana dla różnych wartości estymatorów $\begin{bmatrix} \hat{\lambda} \\ \hat{T} \end{bmatrix}$, przy $\begin{bmatrix} \lambda \\ T \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$

Przedstawiona w niniejszym podrozdziale procedura identyfikacji wartości parametrów ma charakter uniwersalny. Podobne uwarunkowania, jak określone powyżej dla problemów sterowania optymalnego, można bowiem wskazać dla wielu zagadnień nawet spoza zakresu szeroko rozumianej inżynierii. Otóż, według **teorii Kahnemana**, psychologa, laureata Nagrody Nobla w dziedzinie ekonomii, zachowania gospodarcze nie są do końca racjonalne. Zgodnie z jego teorią człowiek silnie reaguje na ekstremalne bodźce i jest skłonny do przeceniania strat i niedoceniania zysków. Ta obawa przed dużymi stratami pozwala przetrwać zwierzęciu w przyrodzie, jednak w gospodarce prowadzi do irracjonalnego lęku przed zmianami. Gdyby zatem preferencje psychologiczne przeciętnego człowieka opisać za pomocą aparatu matematycznego podobnego do wskaźnika jakości sterowania optymalnego, to postać owych preferencji staje się podobna do formuły kwadratowej niesymetrycznej (4.11). Odmienny stosunek człowieka do strat i zysków jest tu reprezentowany przez niesymetrię (tj. $p \neq q$), a obawa przed zjawiskami ekstremalnymi – poprzez postać kwadratową.

Uwagi bibliograficzne

Niniejszy podrozdział ujmuje wyniki badań prowadzonych przez autora wraz z Małgorzatą Charytanowicz (dotyczących liniowej niesymetrycznej funkcji strat) i Aleksandrem Mazgajem (w zakresie zagadnień związanych z funkcją kwadratową niesymetryczną) w ramach Ich prac doktorskich [8, 46]. Powyższy tekst zawiera fragmenty tych rozpraw oraz wspólnych publikacji [37, 38].

Ogólne zagadnienia modelowania matematycznego są prezentowane w książkach [23, 45].

Klasycznym podręcznikiem z zakresu sterowania optymalnego jest praca [2] – materiał przedstawiony w punkcie 4.1.4 szczególnie nawiązuje do zawartych w niej podrozdziałów 7.2, 9.5, a pośrednio również 7.4. W tej dziedzinie oraz pokrewnych tematach warto także polecić publikacje [7, 28, 29, 31, 33, 41, 47, 69]. Przykład szczegółowej analizy wrażliwości układu regulacji automatycznej można znaleźć w monografii [35].

4.2. Wykrywanie i klasyfikacja elementów nietypowych (odosobnionych)

W licznych zagadnieniach nauki i działań praktycznych pojawia się zadanie **wykrycia elementu nietypowego** o własnościach istotnie odmiennych od tych, jakie mają pozostałe elementy z rozpatrywanego zbioru. Na przykład w problemach analizy danych powyższe może świadczyć o istotnym błędzie popełnionym w procesie metrologicznym albo we wstępnej fazie gromadzenia lub prze-

twarzania uzyskanych informacji. W konsekwencji wykrycie takiego elementu pozwala na dokonanie odpowiedniej korekty lub stanowi wręcz przesłankę do wyeliminowania go z posiadanego zbioru danych. Z kolei w innych dziedzinach współczesnych nauk stosowanych obecność tego typu elementu może świadczyć – przykładowo – o uszkodzeniu nadzorowanego urządzenia technicznego albo pojawieniu się stanów patologicznych w organizmie badanego pacjenta. Umożliwia to nadzorowanie złożonych obiektów technicznych w trakcie wykonywania właściwych im zadań produkcyjnych lub też – w zagadnieniach medycznych – stanowi cenne narzędzie wstępnej fazy badań o charakterze masowym, zwłaszcza profilaktycznych. Zgodnie z nazewnictwem statystyki matematycznej elementy znacząco różniące się od pozostałych elementów próby losowej są określane mianem **elementów odosobnionych**. W przypadku wykrycia takiego elementu dodatkowe korzyści można odnieść poprzez identyfikację charakteru wykrytej odmienności, czyli jego ewentualną przynależność do ustalonej klasy. Stanowi to przykładową postać tak zwanego **zadania klasyfikacji**.

Mogący służyć powyższym celom aparat matematyczny wykształcił się w ramach **testowania hipotez** – drugiego, obok teorii estymacji, podstawowego działu statystyki. Poniżej będą przedstawione pokrótce te pojęcia z powyższego zakresu, które zostaną wykorzystane później do zaprojektowania procedury wykrywania i klasyfikacji elementów nietypowych, opartej na metodyce konstruowania estymatorów jądrowych.

Hipotezą statystyczną jest nazywane dowolne przypuszczenie weryfikowane na podstawie wartości próby losowej. Regułę postępowania, która wszystkim możliwym wartościom próby przyporządkowuje decyzję o przyjęciu albo odrzuceniu weryfikowanej hipotezy (lub w ogólnym przypadku także o niepodjęciu żadnej z tych decyzji), nazywa się **testem statystycznym**, a samą procedurę – testowaniem hipotezy statystycznej. Podstawę takiego testu stanowi statystyka S (por. punkt 2.2.1), która w zależności od tego, czy testowana hipoteza jest prawdziwa czy fałszywa, przyjmuje z możliwie dużym prawdopodobieństwem wartości należące do rozłącznych zbiorów: **zbioru odrzucenia (krytycznego)** $A \subset \mathbb{R}^k$, w przypadku gdy testowana hipoteza jest fałszywa, albo zbioru przyjęcia $B \subset \mathbb{R}^k$, jeśli hipoteza jest prawdziwa. Zbiór przyjęcia określa się przeważnie jako dopełnienie zbioru krytycznego:

$$B = \mathbb{R}^k \setminus A. \tag{4.52}$$

W podstawowym przypadku $k = 1$ i wówczas jeżeli o fałszywości testowanej hipotezy świadczą małe wartości statystyki testowej, zbiór krytyczny przyjmuje się w postaci $A = (-\infty, a]$, czyli jako **lewostronny zbiór krytyczny**, w przeciwnym razie jako **prawostronny zbiór krytyczny** $A = [a, \infty)$, przy czym liczba a jest nazywana **wartością krytyczną**. Ostatecznie, jeżeli wartości $x_1, x_2, \ldots, x_m$ stanowią próbę losową, to testowana hipoteza zostaje odrzucona, gdy

$$S(x_1, x_2, \ldots, x_m) \in A, \tag{4.53}$$

lub przyjęta, jeśli

$$S(x_1, x_2, \ldots, x_m) \in B. \tag{4.54}$$

W ogólnym przypadku możliwe staje się również niepodjęcie żadnej z powyższych decyzji, co najczęściej implikuje dalsze badania przedmiotowe.

Poza hipotezą testowaną H można także wyróżnić hipotezę alternatywną H_A, przyjmowaną za prawdziwą w przypadku odrzucenia hipotezy H. Typowa hipoteza alternatywna jest zaprzeczeniem hipotezy testowanej: $H_A = \sim H$.

Jak wynika z definicji hipotezy statystycznej, nie można wykazać z całą pewnością jej prawdziwości albo fałszywości, gdyż podstawą wnioskowania jest tu próba losowa. Możliwe staje się zatem podjęcie poprawnej decyzji albo też popełnienie jednego z dwóch błędów: odrzucenia hipotezy prawdziwej, czyli tak zwany **błąd pierwszego rodzaju**, lub przyjęcia hipotezy fałszywej, co z kolei stanowi **błąd drugiego rodzaju**.

Jeżeli założyć, że postać statystyki testowej i zbioru krytycznego oraz liczność próby są ustalone, to prawdopodobieństwa błędów zależą już bezpośrednio od przyjętej wartości krytycznej, przy czym na ogół ewentualna jej zmiana, dokonana w celu zmniejszenia prawdopodobieństwa jednego rodzaju błędu, powoduje bezpośrednio zwiększenie prawdopodobieństwa drugiego. W praktyce często stosuje się tak zwane **testy istotności**, w których wymaga się, aby prawdopodobieństwo błędu pierwszego rodzaju było równe założonemu **poziomowi istotności** α, natomiast nie kontroluje się bezpośrednio prawdopodobieństwa błędu drugiego rodzaju. W wyniku takiego postępowania można odrzucić testowaną hipotezę, lecz często nie podejmuje się decyzji o jej przyjęciu, gdyż wówczas zachodziłoby niebezpieczeństwo popełnienia niekontrolowanego tu błędu drugiego rodzaju. Jednakże w niektórych zastosowaniach sam brak przyjęcia hipotezy, a więc odstąpienie od decyzji i w konsekwencji zaniechanie jakiegokolwiek działania jest całkowicie wystarczające z aplikacyjnego punktu widzenia. (Warto zauważyć, że w przypadku stosowania testów istotności określenie zbioru przyjęcia oraz hipotezy alternatywnej często może być czynnością zbędną – nie podejmuje się bowiem decyzji o przyjęciu testowanej hipotezy, a zależne od hipotezy alternatywnej prawdopodobieństwo błędu drugiego rodzaju jest poza zakresem rozważań.)

Duże znaczenie w procesie projektowania testu statystycznego ma założona **wartość poziomu istotności** $\alpha \in (0, 1)$. Wybór jest właściwie arbitralny, aczkolwiek powinien on uwzględniać ekonomiczne konsekwencje ewentualnych błędów pierwszego i drugiego rodzaju. Przeważnie są używane wartości 0,01, 0,02, 0,05 (jako podstawowa) i 0,1. W trakcie stosowania testu możliwa jest adaptacja algorytmu polegająca na zmianach wartości tego parametru.

Powyższa, klasyczna koncepcja testowania hipotez jest zgodna z **teorią Neymana-Pearsona**. Bardziej wyrafinowana **teoria Walda** oparta została na statystycznej teorii decyzji przedstawionej na początku niniejszego rozdziału. W typowym przypadku zakłada się dwuelementowy zbiór stanów natury – są one charakteryzowane, odpowiednio, przez hipotezy testowaną i alternatywną. Z kolei przyjęcie hipotezy testowanej albo alternatywnej jest uznawane za dwie możliwe do podjęcia decyzje. Poprzez odpowiednie zdefiniowanie postaci funkcji strat są tu dodatkowo uwzględniane konsekwencje ewentualnych błędów pierwszego i drugiego rodzaju.

W dalszej części tego podrozdziału metodyka wyznaczania estymatorów jądrowych zostanie wykorzystana do skonstruowania procedury przeznaczonej do wykrywania i klasyfikacji elementów nietypowych (odosobnionych). I tak, w punkcie 4.2.1 będzie przedstawiony test umożliwiający ich wykrywanie, a w punkcie 4.2.2 – klasyfikację. Koncepcje te zostaną uzupełnione w kolejnym punkcie 4.2.3 o wersję lokalną, różnicującą znaczenie poszczególnych elementów próby losowej odpowiednio do ich ulokowania względem – stanowiącego tu punkt odniesienia – testowanego elementu.

4.2.1. Wykrywanie elementów nietypowych

Poniżej będzie rozważane zadanie wykrywania elementów nietypowych. Ponieważ z uwagi na uwarunkowania aplikacyjne nie przyjmuje się tutaj żadnych założeń dotyczących zbioru (wzorca) wszystkich możliwych elementów tego typu, co implikuje znaczną nieokreśloność ewentualnej hipotezy alternatywnej oraz prawdopodobieństwa błędu drugiego rodzaju, zaproponowany zostanie test istotności, zgodnie z zasadami teorii Neymana-Pearsona.

Niech zatem dana będzie n-wymiarowa zmienna losowa X oraz uzyskana z niej próba losowa

$$x_1, x_2, \ldots, x_m. \tag{4.55}$$

Jej elementy są traktowane w badanym zagadnieniu jako **typowe**. Parametr $\alpha \in (0, 1)$ oznacza założony poziom istotności. Testowana będzie **hipoteza** stanowiąca, że ustalony wektor

$$\tilde{x} \in \mathbb{R}^n \tag{4.56}$$

jest elementem typowym.

Stosowaną w tym celu **statystykę** $S: \mathbb{R}^n \to [0, \infty)$ można zdefiniować zależnością

$$S(\tilde{x}) = \hat{f}(\tilde{x}), \tag{4.57}$$

przy czym $\hat{f}$ oznacza jądrowy estymator gęstości prawdopodobieństwa zmiennej losowej X, określony dla próby (4.55). Jego konstrukcja została przedstawio-

na w podrozdziale 3.1. Poza modyfikacją parametru wygładzania (punkt 3.1.6) zalecane jest tu użycie jądra będącego kombinacją liniową jąder Epanecznikowa oraz normalnego (3.46), a w przypadku wielowymiarowym opartego na nim jądra radialnego (3.49) z transformacją liniową w pełnej postaci (3.53) – punkt 3.1.4. Należy podkreślić, iż postać statystyki (4.57) wraz z własnościami estymatorów jądrowych pozwala na identyfikację ewentualnych różnic dotyczących nie tylko wartości wektorów (4.55) i (4.56), ale przede wszystkim złożonych relacji występujących między ich współrzędnymi.

Wartość statystyki (4.57) można interpretować jako prawdopodobieństwo wystąpienia wartości z otoczenia wektora (4.56), w warunkach charakteryzowanych przez próbę (4.55) złożoną z elementów traktowanych jako typowe. O odrzuceniu testowanej hipotezy, stanowiącym, iż wektor (4.56) takowy nie jest, świadczą zatem małe wartości $S(\widetilde{x})$. **Zbiór krytyczny** zostanie więc zdefiniowany w postaci lewostronnej

$$A = (-\infty, a], \tag{4.58}$$

gdzie **wartość krytyczna** a jest dana jako estymator kwantyla $\hat{q}$ rzędu równego przyjętemu poziomowi istotności

$$r = \alpha \tag{4.59}$$

zmiennej losowej $Z \equiv \hat{f} \circ X$, czyli wyznaczony dla próby $z_1, z_2, ..., z_m$ o elementach określonych wzorami

$$\begin{aligned} z_1 &= \hat{f}(x_1), \\ z_2 &= \hat{f}(x_2), \\ &\vdots \\ z_m &= \hat{f}(x_m). \end{aligned} \tag{4.60}$$

Podsumowując powyższe, zapiszmy

$$a = \hat{q}. \tag{4.61}$$

Do wyznaczenia wartości estymatora kwantyla $\hat{q}$ rzędu (4.59) na podstawie próby losowej (4.60) można wykorzystać algorytm przedstawiony w punkcie 3.2.2. Z uwagi na to, iż rząd kwantyla jest bliski zeru, a wartości zmiennej losowej Z są ograniczone do zbioru liczb nieujemnych, szczególnie zalecane jest tu użycie procedury ograniczenia nośnika do przedziału $[0, \infty)$, przedstawionej w punkcie 3.1.7.

Po wyznaczeniu funkcji $\hat{f}$ i liczby $\hat{q}$, a więc w konsekwencji postaci statystyki (4.57) oraz – po podstawieniu (4.61) – zbioru krytycznego (4.58), uzyskuje

się komplet wielkości niezbędnych do stosowania projektowanego testu. I tak, po obliczeniu wartości statystyki dla testowanego elementu, czyli wielkości $S(\tilde{x})$, otrzymujemy ostatecznie, że jeżeli zachodzi warunek

$$S(\tilde{x}) \in A, \tag{4.62}$$

to należy odrzucić hipotezę stanowiącą, iż element ten jest typowy, a zatem uznać go za nietypowy. W zadaniach praktycznych najczęściej skutkuje to podjęciem stosownych czynności właściwych konkretnemu zagadnieniu aplikacyjnemu. W szczególności możliwe jest wówczas sprawdzenie, czy element ten należy do pewnych wyróżnionych klas elementów nietypowych – zagadnienie to będzie przedmiotem rozważań przedstawionych w następnym punkcie.

W przypadku istnienia czynnika zewnętrznego mającego bezpośredni wpływ na badane zagadnienie, w tym sensie iż podanie konkretnej (np. aktualnej) wartości tego czynnika mogłoby znacząco uściślić będące w dyspozycji dane, warto wykorzystać w powyższej koncepcji ujęcie warunkowe, zastępując funkcję $\hat{f}$ jej postacią warunkową $\hat{f}_{X|W=w^*}$ szczegółowo przedstawioną w punkcie 3.2.3. Zarówno zmienna losowa X, jak i zmienna warunkująca W mogą zawierać współrzędne binarne, zgodnie z koncepcją wprowadzoną w punkcie 3.1.8, a wobec współrzędnych zmiennej X można stosować procedurę ograniczenia nośnika (punkt 3.1.7).

4.2.2. Klasyfikacja elementów nietypowych

W wielu praktycznych zagadnieniach dodatkowe korzyści uzyskuje się przez testowanie przynależności wykrytego elementu nietypowego do ustalonej klasy takich elementów, z wykorzystaniem probabilistycznej reprezentacji powyższej klasy w postaci pobranej z niej próby losowej. Zakres potencjalnych zastosowań wzmiankowanych procedur stanowi obszerne spektrum aplikacyjnych problemów, od złożonych zagadnień diagnostyki technicznej i medycznej aż po proste przypadki identyfikacji trywialnych błędów powodujących odmienność wykrytego elementu. Ilustracyjnym przykładem jest tu omyłkowe podanie wartości pomiaru w innych niż założone jednostkach fizycznych – w takich przypadkach możliwa jest nawet automatyczna korekta powstałej nieprawidłowości, czyli swoista „naprawa" elementu nietypowego.

Kontynuując oznaczenia poprzedniego punktu, niech X oznacza n-wymiarową zmienną losową, a

$$x_1, x_2, \ldots, x_m \tag{4.63}$$

uzyskaną z niej m-elementową próbę losową, złożoną z elementów traktowanych jako **typowe**. Niech ponadto będzie ustalony pewien **podzbiór elementów**

nietypowych, określany dalej mianem **klasy** C, a także m^*-elementowa próba losowa

$$x_1^*, x_2^*, ..., x_{m^*}^* \tag{4.64}$$

otrzymana z tej zmiennej jako reprezentacja powyższej klasy. Testowana będzie **hipoteza** stanowiąca, że ustalony wektor

$$\tilde{x} \in \mathbb{R}^n \tag{4.65}$$

należy do klasy C, przeciw **hipotezie alternatywnej**, iż jest elementem typowym.

Do rozwiązania powyższego zadania zostanie wykorzystana wzmiankowana we wstępie do podrozdziału 4.2 teoria Walda, w nawiązaniu do statystycznej teorii decyzji przedstawionej na początku rozdziału 4. Tak więc zakłada się dwuelementowy zbiór stanów natury $\mathcal{X} = \{n^*, n_*\}$, przy czym

$$n^* \text{ – oznacza, że element (4.65) należy do klasy } C, \tag{4.66}$$

$$n_* \text{ – oznacza, że element (4.65) jest typowy.} \tag{4.67}$$

Podobnie, zbiór możliwych do podjęcia decyzji przyjmuje postać $\mathcal{D} = \{d^*, d_*\}$, gdzie

$$d^* \text{ – oznacza przyjęcie hipotezy, iż element (4.65) należy do klasy } C, \tag{4.68}$$

$$d_* \text{ – oznacza przyjęcie hipotezy, iż element (4.65) jest typowy.} \tag{4.69}$$

Funkcja strat jest dana w naturalnej postaci

$$l(n, d) = \begin{cases} 0, & \text{gdy występuje stan } n^*, \text{ a podjęto decyzję } d^*, \\ 0, & \text{gdy występuje stan } n_*, \text{ a podjęto decyzję } d_*, \\ p, & \text{gdy występuje stan } n^*, \text{ a podjęto decyzję } d_*, \\ q, & \text{gdy występuje stan } n_*, \text{ a podjęto decyzję } d^*, \end{cases} \tag{4.70}$$

gdzie $n \in \{n^*, n_*\}$ oraz $d \in \{d^*, d_*\}$, natomiast dodatnie parametry p oraz q reprezentują straty, jakimi skutkują błędy – odpowiednio – pierwszego i drugiego rodzaju, a zatem charakteryzujące straty uznania elementu z klasy C za typowy oraz straty wynikłe z błędu przeciwnego, czyli przyporządkowania do tej klasy elementu typowego.

Proponowana teraz **statystyka** $S^* : \mathbb{R}^n \to [0, \infty)$ zostanie przyjęta w następującej postaci:

$$S^*(\tilde{x}) = \hat{f}^*(\tilde{x}), \tag{4.71}$$

gdzie $\hat{f}^*$ oznacza jądrowy estymator gęstości prawdopodobieństwa otrzymany na podstawie próby losowej (4.64). Prawdziwe pozostają tu wszelkie uwagi poczynione wcześniej wobec statystyki S, sformułowane poniżej zależności (4.57).

Za podjęciem decyzji d^* świadczą zatem duże wartości statystyki S^*, a więc powinno to nastąpić, gdy

$$S^*(\tilde{x}) \in A^*, \tag{4.72}$$

gdzie

$$A^* = [a^*, \infty). \tag{4.73}$$

Stanowi to o uznaniu, iż testowany element (4.65) należy do klasy C, reprezentowanej przez próbę losową (4.64).

W celu wyznaczenia **wartości krytycznej** a^* zostanie poniżej użyta bayesowska reguła decyzyjna, przedstawiona na początku rozdziału 4. Nietrudna analiza wskazuje, że optymalna w sensie tego kryterium wartość krytyczna jest rozwiązaniem następującego równania z argumentem a^*:

$$\hat{F}(a^*) + \frac{p}{q}\hat{F}^*(a^*) = 1. \tag{4.74}$$

gdzie $\hat{F}^*$ oznacza jądrowy estymator zmiennej losowej $Z \equiv \hat{f}^* \circ X$, gdy występuje stan n^*, natomiast $\hat{F}$ – jądrowy estymator tej zmiennej przy założeniu stanu n_*. Estymatory te są zatem otrzymywane, odpowiednio, na podstawie prób $z_1^*, z_2^*, \ldots, z_{m^*}^*$ oraz $z_1, z_2, \ldots, z_m$, o elementach danych zależnościami

$$\begin{aligned} z_1^* &= \hat{f}^*(x_1^*), \\ z_2^* &= \hat{f}^*(x_2^*), \\ &\vdots \\ z_{m^*}^* &= \hat{f}^*(x_{m^*}^*) \end{aligned} \tag{4.75}$$

oraz

$$\begin{aligned} z_1 &= \hat{f}^*(x_1), \\ z_2 &= \hat{f}^*(x_2), \\ &\vdots \\ z_m &= \hat{f}^*(x_m). \end{aligned} \tag{4.76}$$

Sama procedura wyznaczania jądrowego estymatora dystrybuanty została przedstawiona w punkcie 3.2.1. W rozważanym powyżej zadaniu musi być ona stosowana odrębnie wobec estymatorów $\hat{F}^*$ i $\hat{F}$.

Równanie (4.74) ma rozwiązanie i w przypadku stosowania jądra o dodatnich własnościach rozwiązanie to jest jedyne. Obliczenie jego wartości może nastąpić z użyciem algorytmu Newtona, analogicznie jak w punkcie 3.2.2 dokonano tego wobec kwantyla (por. procedurę (3.314)-(3.316)).

Z postaci kryterium (4.74) ponownie widać, że nie jest konieczne odrębne określenie wartości parametrów p oraz q, lecz jedynie ich ilorazu.

Podobnie jak w przypadku zagadnienia wykrywania elementów nietypowych, możliwe jest zastosowanie ujęcia warunkowego, polegającego na zastąpieniu funkcji $\hat{f}^*$ jej postacią warunkową $\hat{f}^*_{X|W=w^*}$ (punkt 3.2.3). Zarówno zmienna losowa X, jak i zmienna warunkująca W mogą zawierać współrzędne binarne (punkt 3.1.8), a wobec współrzędnych zmiennej X można stosować procedurę ograniczenia nośnika (punkt 3.1.7).

4.2.3. Ujęcie lokalne

W przypadku wielowymiarowym, gdy jest stosowane jądro radialne z transformacją liniową określoną macierzą w postaci pełnej (3.53), czyli gdy za macierz transformacji R przyjmuje się macierz kowariancji COV (por. punkt 3.1.4), możliwe jest uwzględnienie **lokalnych własności** zmiennej losowej odniesionych do położenia testowanego elementu, przez odpowiednie preferowanie tych elementów próby losowej, które są mu bliższe.

Najpierw będzie rozważane zagadnienie wyznaczania jądrowego estymatora gęstości prawdopodobieństwa $\hat{f}$ na potrzeby statystyki (4.57). Wprowadźmy zatem następującą wielkość:

$$d = \max_{i=1,2,\ldots,m} \left\| R^{-1} \frac{\tilde{x} - x_i}{hs_i} \right\|. \tag{4.77}$$

Liczba d stanowi zatem maksymalną odległość testowanego elementu $\tilde{x}$ od poszczególnych elementów próby losowej (4.55), po uwzględnieniu „zmian skali" wprowadzanych przez macierz R i parametry h oraz s_i. Bez zmniejszania ogólności rozważań można założyć, że $d > 0$. Zerowa wartość oznaczałaby bowiem przypadek trywialny, gdy $x_1 = x_2 = \ldots = x_m = \tilde{x}$, co stanowi, iż testowany element jest taki sam jak wszystkie inne, a zatem z pewnością nie można go uznać za nietypowy. Niech ponadto

$$d_i = K_E\left(\frac{1}{d}R^{-1}\frac{\tilde{x}-x_i}{hs_i}\right) \quad \text{dla } i = 1, 2, ..., m, \tag{4.78}$$

gdzie K_E oznacza n-wymiarowe radialne jądro Epanecznikowa, czyli określone zależnością (3.47) z podstawieniem (3.40). Oczywiste jest, iż $d_i \geq 0$. Można także założyć, że nie wszystkie parametry d_i są równe zeru, co z praktycznego punktu widzenia nie zmniejsza ogólności rozważań. Gdyby bowiem $d_i \equiv 0$, to elementy $x_1, x_2, ..., x_m$ byłyby ulokowane na okręgu o środku w punkcie $\tilde{x}$ i promieniu d, czyli w zbiorze zerowej miary Lebesgue'a – wobec przyjętego założenia (2.10) prawdopodobieństwo takiego zdarzenia wynosi zero.

Analogicznie do definicji estymatorów elementów macierzy kowariancji (2.54)-(2.55) określić można lokalne ich postacie $\hat{cov}^{loc}_{j_1,j_2}$ dla $j_1, j_2 \in \{1, 2, ..., n\}$. I tak, jeżeli znane są E_{j_1} oraz E_{j_2}, czyli wartości oczekiwane współrzędnych j_1 oraz j_2, to

$$\hat{cov}^{loc}_{j_1,j_2} = \frac{1}{\sum_{i=1}^{m} d_i}\sum_{i=1}^{m} d_i(x_{i,j_1} - E_{j_1})(x_{i,j_2} - E_{j_2}), \tag{4.79}$$

natomiast w przypadku gdy wartości oczekiwane są określone zależnością (2.49) i przy $m \geq 2$:

$$\hat{cov}^{loc}_{j_1,j_2} = \frac{m}{(m-1)\sum_{i=1}^{m} d_i}\sum_{i=1}^{m} d_i(x_{i,j_1} - \hat{E}_{j_1})(x_{i,j_2} - \hat{E}_{j_2}). \tag{4.80}$$

Biorąc pod uwagę postać parametrów d_i zdefiniowanych zależnością (4.78), wpływ poszczególnych elementów próby losowej na wartości powyższych wyrażeń jest proporcjonalny do tego, jak blisko dany element jest położony względem testowanego elementu.

Identyczne wzory można przedstawić w przypadku jądrowego estymatora gęstości prawdopodobieństwa $\hat{f}^*$ wyznaczanego dla statystyki (4.71) na podstawie próby losowej (4.64).

Ostatecznie, nawiązując do definicji (3.58), można następująco określić **lokalne ujęcie transformacji liniowej** w postaci pełnej:

$$R = [\hat{cov}^{loc}_{j_1,j_2}], \tag{4.81}$$

przy czym elementy $\hat{cov}^{loc}_{j_1,j_2}$ dla $j_1, j_2 \in \{1, 2, ..., n\}$ są dane wzorami (4.79) albo (4.80). W przypadkach zmiennych losowych o specyficznych rozkładach,

na przykład będących sumą dwóch składników o znacząco różnych wartościach oczekiwanych i istotnie odmiennych macierzach kowariancji, może to stanowić o wymiernych korzyściach w postaci lepszego dopasowania estymatora do rzeczywistych uwarunkowań w otoczeniu konkretnego, będącego przedmiotem badań elementu $\widetilde{x}$.

Warto zwrócić uwagę nie tylko na samą przedstawioną powyżej procedurę, lecz także na jej **ogólną ideę** uwzględnienia lokalnych własności badanej zmiennej losowej. Może być ona bowiem stosowana w wielu aspektach konstrukcji estymatora jądrowego, na przykład przy ustalaniu wartości parametru wygładzania. Najczęściej wymaga to jednak indywidualnej i wszechstronnej analizy, gdyż w większości zagadnień aplikacyjnych brak jest tak wyrazistego punktu odniesienia, jaki stanowił w rozpatrywanym tu zagadnieniu testowany element.

Uwagi bibliograficzne

Niniejszy podrozdział ujmuje wyniki przedstawione w monografii [36], a także badania prowadzone przez autora wraz z Cyprianem Prochotem w ramach Jego pracy doktorskiej [55] – w tym zakresie powyższy tekst zawiera fragmenty wspólnej publikacji [39].

Klasyczny podręcznik dotyczący elementów odosobnionych reprezentuje pozycja [3].

Testowanie hipotez statystycznych ma wyjątkowo bogatą literaturę przedmiotową. Szczególnie warta polecenia jest książka [22]. Praca [12] stanowi obszerne kompendium najczęściej stosowanych testów. W podręczniku [43] jest zaprezentowane zagadnienie testowania hipotez z wykorzystaniem elementów statystycznej teorii decyzji.

Metodyka związana z diagnostyką techniczną ulega obecnie gwałtownym zmianom ze względu na szybki rozwój techniki komputerowej. Reprezentatywny przegląd współczesnego stanu wiedzy na ten temat można znaleźć w pracy [11]. Z kolei interesująca – z punktu widzenia analizy systemowej – koncepcja właściwa diagnostyce medycznej, polegająca na automatycznym rozumieniu rozpoznawanych obrazów, została zaproponowana w monografii [64].

4.3. Określenie przestrzennego rozkładu popytu na potrzeby planowania systemu telekomunikacyjnego

Obserwowany w ostatnich latach gwałtowny rozwój zastosowań systemów informatycznych, w tym spektakularne zainteresowanie dostępem do Internetu, spowodował znaczny wzrost znaczenia usług **bezprzewodowej szerokopasmowej transmisji danych**. Wykorzystywany przez operatorów telekomunika-

cyjnych do tego celu **system LMDS** (*Local Multipoint Distribution System*) umożliwia połączenie węzła sieci operatorskiej z budynkami, w których są zlokalizowani odbiorcy usług, bez konieczności budowy kosztownej infrastruktury kablowej. Dane są przekazywane między rozlokowanymi na terenie aglomeracji miejskiej stacjami bazowymi, które drogą radiową obsługują stałe połączenia z wieloma stacjami abonenckimi, znajdującymi się w zasięgu skutecznej transmisji ich urządzeń nadawczo-odbiorczych. Stacje abonenckie transmitują z kolei, za pomocą sieci lokalnych, owe dane do użytkowników w obrębie danej lokalizacji.

Istotnym czynnikiem, często decydującym o ekonomicznej zasadności stosowania systemu LMDS, jest **wyznaczenie miejsc zainstalowania stacji bazowych** w taki sposób, aby w granicach dostępnych środków inwestycyjnych osiągnąć największy zysk. Zagadnienie planowania rozmieszczenia stacji bazowych systemu LMDS nie jest zadaniem łatwym z uwagi na konieczność uwzględnienia w procesie planistycznym szeregu uwarunkowań natury technicznej oraz ekonomicznej. Do podstawowych ograniczeń technicznych można zaliczyć zasięg urządzeń nadawczo-odbiorczych stacji bazowej, a także ich maksymalną przepustowość. Dla zapewnienia transmisji jest również wymagana linia widoku pomiędzy antenami stacji bazowej oraz abonenckiej. Ze względu na nierównomierne ukształtowanie terenu oraz występowanie przeszkód, takich jak wysokie budynki, w obrębie teoretycznego zasięgu mogą istnieć obszary „niewidoczne", dla których realizacja połączeń nie jest możliwa. Z tego też powodu istnieje ograniczona ilość miejsc, które ze względu na swoją wysokość są dobrze widoczne, stanowiąc potencjalne lokalizacje stacji bazowych. Obok powyższych ograniczeń technicznych fundamentalny problem stwarza planistom także oszacowanie popytu na usługi transmisji danych, które w praktyce wykonuje się na podstawie często mało dokładnych i niekompletnych informacji, dotyczących potencjalnych odbiorców usług zlokalizowanych na danym terenie. Proces decyzyjny jest dodatkowo trudniejszy, gdy dopuszcza się horyzont wieloletni, zwłaszcza w warunkach niestacjonarności danych.

Tak więc zadanie planowania wymaga dokonania wyboru tych spośród możliwych lokalizacji stacji bazowych, które zapewnią łącznie maksymalny zysk z realizacji usług, przy czym liczbę stacji ogranicza wielkość środków inwestycyjnych. W niniejszym podrozdziale zostanie zaprezentowany algorytm **wyznaczania optymalnego układu stacji bazowych systemu LMDS**. Estymatory jądrowe będą zastosowane do oszacowania przestrzennego rozkładu popytu na usługi transmisji danych. Ze względu na naturalną nieprecyzyjność ocen przyszłego popytu zostaną tu dodatkowo wprowadzone elementy logiki rozmytej. Prezentowana metoda ma charakter uniwersalny i może być łatwo zaadaptowana do zadań pokrewnych, na przykład rozplanowania punktów obsługi klienta na terenie aglomeracji miejskich. Znaczna różnorodność

poruszanych problemów powoduje, że zagadnienie to zostanie przedstawione jedynie w zarysie, z wyeksponowaniem nietypowego tu zastosowania estymatorów jądrowych.

4.3.1. Przestrzenny rozkład popytu

Przestrzenny rozkład popytu na usługi transmisji danych ma – w dosłownym znaczeniu – charakter punktowy, związany z poszczególnymi potencjalnymi klientami. W tej postaci jest on jednak praktycznie nieidentyfikowalny na obszarze aglomeracji miejskiej. W celu wyznaczenia użytecznej charakterystyki powyższego rozkładu zostanie poniżej zaadaptowana koncepcja estymatorów jądrowych. Opierając się na najbardziej reprezentatywnych obiektach, umożliwiają one uzyskanie ciągłego i uśrednionego modelu przestrzennego rozkładu popytu, dzięki czemu jest on dogodny do praktycznych zadań optymalizacji.

Funkcja $\hat{f}$, otrzymana z wykorzystaniem przedstawionej w podrozdziale 3.1 metodyki, będzie charakteryzować **przestrzenny rozkład popytu na usługi transmisji danych** w obrębie rozpatrywanego obszaru działania systemu LMDS. Przyjmuje się zatem wymiar $n = 2$, a poszczególne współrzędne punktu $x = \begin{bmatrix} ,x_1 \\ ,x_2 \end{bmatrix}$ oznaczają odpowiednio: $,x_1$ – jego długość oraz $,x_2$ – szerokość geograficzną. Kolejne elementy próby losowej $x_1, x_2, \ldots, x_m$ reprezentują poszczególne lokalizacje głównych obiektów abonenckich ze zgromadzonej bazy danych. Ze względu na dogodność dalszych analitycznych obliczeń zostało przyjęte jądro radialne postaci

$$K(x) = K\left(\begin{bmatrix} ,x_1 \\ ,x_2 \end{bmatrix}\right) = \frac{1}{\pi} \frac{1}{(1 + ,x_1^2 + ,x_2^2)^2}. \tag{4.82}$$

Zmniejszenie wykładnika w powyższym mianowniku, względem typowej postaci Cauchy'ego (3.304), ma na celu ułatwienie dalszych obliczeń analitycznych. Aczkolwiek skutkiem tego warunek (3.16) nie jest spełniony, to w przypadku rozpatrywanego tu zagadnienia interpretacja wykresu funkcji $\hat{f}$ jest na tyle wyrazista, iż dobór wartości parametru wygładzania h korzystnie jest przeprowadzić na podstawie wizualnej oceny wykresu otrzymanej funkcji $\hat{f}$, zgodnie z sugestią sformułowaną na końcu punktu 3.1.5. Zastosowana została procedura modyfikacji parametru wygładzania z wyznaczonymi na podstawie algorytmu (3.112)-(3.115) parametrami $s_1, s_2, \ldots, s_m$ (por. punkt 3.1.6). Pominięta jest natomiast transformacja liniowa (punkt 3.1.4), ze względu na brak uzasadnienia in-

terpretacyjnego w tym konkretnym, nietypowym zastosowaniu. Rozważana będzie zatem następująca postać użytego tu estymatora jądrowego:

$$\hat{f}(x) = \frac{1}{h^2 \sum_{i=1}^{m} w_i} \sum_{i=1}^{m} \frac{w_i}{s_i^2} K\left(\frac{x - x_i}{h s_i}\right), \tag{4.83}$$

przy czym jądro K jest dane wzorem (4.82), a wprowadzone powyżej dodatnie parametry w_i dla $i = 1, 2, ..., m$ są przypisane poszczególnym jądrom i mogą być interpretowane jako potencjalny popyt związany z i-tym obiektem abonenckim. W przypadku gdy $w_i \equiv 1$, zależność ta jest równoważna definicji (3.114) przy $n = 2$.

Wartość całki funkcji $\hat{f}$ reprezentuje zatem potencjalny popyt na usługi transmisji danych, jaki może wystąpić na rozważanym obszarze. Ponieważ poszukiwana będzie wartość argumentu, dla której jego realizacja jest maksymalna, więc możliwe staje się pominięcie we wzorze (4.83) stałego czynnika $1/(h^2 \sum_{i=1}^{m} w_i)$, jako niewpływającego na wynik zadania optymalizacyjnego, co implikuje ostatecznie:

$$\hat{f}(x) = \sum_{i=1}^{m} \frac{w_i}{s_i^2} K\left(\frac{x - x_i}{h s_i}\right). \tag{4.84}$$

Koncepcja estymatorów jądrowych została zatem zastosowana powyżej do oszacowania przestrzennego rozkładu potencjalnego popytu na usługi transmisji danych. Ich użycie umożliwi uzyskanie ciągłej charakterystyki takiego rozkładu na podstawie posiadanej bazy danych o charakterze punktowym, gdzie poszczególne obiekty abonenckie są charakteryzowane poprzez ich położenie geograficzne x_i oraz przypisane im parametry w_i reprezentujące popyt. Ze względu na własności uśredniające estymatorów jądrowych możliwe jest korzystanie tu z uproszczonej bazy danych, zawierającej jedynie lokalizacje głównych budynków abonenckich, i uwzględnienie w odpowiadających im parametrach w_i zapotrzebowania sąsiednich, pomniejszych obiektów. Dzięki temu znacząco zmniejszony zostanie wpływ niedokładnej lub niepełnej identyfikacji potencjalnych lokalizacji abonenckich na otrzymany wynik, a także obniży się koszt związany z jej przeprowadzeniem.

4.3.2. Wyznaczenie optymalnego układu stacji bazowych

Dysponując określoną w poprzednim punkcie funkcją (4.84) charakteryzującą przestrzenny rozkład popytu na usługi transmisji danych, kolejnym potencjal-

nym lokalizacjom stacji bazowych można przypisać wielkość wynikającą z całkowania tej funkcji w obrębie zasięgu zainstalowanych tam radiowych urządzeń nadawczo-odbiorczych. Z kolei w przypadku układu stacji bazowych całka dla całego obszaru pokrytego zasięgami poszczególnych urządzeń układu określa sumaryczny popyt, stanowiąc jednocześnie kryterium oceny jakości tego układu.

Niech dany będzie układ k potencjalnych lokalizacji stacji bazowych w miejscach $\tilde{x}_j$ przy $j = 1, 2, ..., k$. Wprowadzone zostaną następujące oznaczenia:

$$Q_j = \int_{C_j} \hat{f}(x)\,dx, \tag{4.85}$$

$$Q_{j_1,j_2,...,j_n} = \int_{C_{j1} \cap C_{j2} \cap ... \cap C_{jn}} \hat{f}(x)\,dx, \tag{4.86}$$

gdzie C_j oznacza j-te koło o środku w $\tilde{x}_j$ i dodatnim promieniu r_j (reprezentującym zasięg urządzenia nadawczo-odbiorczego przypisanego j-tej lokalizacji), a także $j_1, j_2, ..., j_n \in \{1, 2, ..., k\}$ są różne, przy czym $2 \le n \le k$. **Sumaryczny popyt**, charakteryzujący jakość rozważanego układu stacji bazowych, jest określony zależnością

$$Q = \int_{C_1 \cup C_2 \cup ... \cup C_n} \hat{f}(x)\,dx = \sum_{j=1}^{k} Q_j - \sum_{\{j_1,j_2\}} Q_{j_1,j_2} + \sum_{\{j_1,j_2,j_3\}} Q_{j_1,j_2,j_3} + ... + (-1)^k Q_{1,2,...,k}. \tag{4.87}$$

Uwzględniając postać powyższego wzoru, obliczenie wartości wyrażeń (4.85) i (4.86) wyczerpuje procedurę pozwalającą na wyznaczenie popytu na usługi transmisji danych dla ustalonego układu stacji bazowych, co stanowi ocenę jakości tego układu.

Ze względu na wybór jądra w postaci (4.82) możliwe jest określenie analitycznej zależności na całkę funkcji pojedynczego jądra K_i z parametrami h, s_i i w_i, po kole C_j o promieniu r_j i odległości $d_{i,j}$ pomiędzy środkiem tego koła a środkiem jądra, dla $i = 1, 2, ..., m$ oraz $j = 1, 2, ..., k$. Wyraża się ona wzorem

$$Q_j = \sum_{i=1}^{m} w_i \left(\frac{r_j^2 - d_{i,j}^2 - h^2 s_i^2}{\sqrt{r_j^4 + 2(h^2 s_i^2 - d_{i,j}^2) r_j^2 + (h^2 s_i^2 + d_{i,j}^2)^2}} + 1 \right), \tag{4.88}$$

przy czym ponownie została pominięta dodatnia stała jako niewpływająca na wynik zadania optymalizacyjnego.

Z kolei analityczne wyznaczenie wartości całki po zbiorze będącym przecięciem dowolnej liczby kół stanowi w praktyce zadanie niewykonalne. Możliwe

jest jednak skonstruowanie procedury przybliżonej. I tak, przecięcie dwóch kół aproksymuje się kołem zastępczym, dla którego można już użyć wzoru (4.88). Przy zagwarantowaniu jednakowej powierzchni koła i soczewki oraz właściwym ulokowaniu środka owego koła, różnica między wartościami funkcji (4.84) na obszarach soczewki i koła nie jest bowiem duża, a błąd mającego charakter uśredniający całkowania po ich polach – stosunkowo niewielki. Warto zauważyć, iż największe wartości tego błędu występują wtedy, gdy soczewka jest silnie spłaszczona, czyli gdy jej powierzchnia, a zatem również wartość wyrażenia Q_{j_1,j_2} jest relatywnie mała. Procedura ta może być łatwo uogólniona na drodze rekurencyjnej dla przecięcia dowolnej liczby kół większej niż dwa.

Powyższe kończy podstawowy algorytm obliczeń niezbędny do zastosowania wzoru (4.87), umożliwiający scharakteryzowanie jakości danego układu stacji bazowych. Zostanie on teraz uzupełniony w celu uwzględnienia obszarów „niewidocznych" i ograniczonej przepustowości stacji bazowych.

Aby uwzględnić występowanie **obszarów „niewidocznych"**, w których transmisja nie jest możliwa ze względu na nierównomierne ukształtowanie terenu oraz występowanie różnorodnych przeszkód, od wartości danej wzorem (4.87) należy odjąć wartość całki po odpowiednich polach. W praktyce projektowania sieci teletransmisyjnych najczęściej aproksymuje się obszary niewidoczne prostymi figurami geometrycznymi, które z kolei przybliża się kołami lub ogólniej – sumą kół. Przy takim założeniu przedstawiony poprzednio algorytm umożliwia łatwe obliczenie całki funkcji $\hat{f}$ po obszarach „niewidocznych" – analogicznie jak przedstawia to wzór (4.87) – a następnie odjęcie tej wartości od podstawowej postaci wskaźnika Q.

Nietrudno także uzupełnić proponowaną powyżej metodę o procedurę pozwalającą uwzględnić **ograniczoną przepustowość** poszczególnych stacji bazowych. I tak, zbiór będący sumą obszarów leżących w zasięgu rozważanego układu stacji bazowych jest podzielony przez okręgi, stanowiące granice zasięgów poszczególnych urządzeń nadawczo-odbiorczych, na skończoną liczbę podzbiorów, oznaczanych dalej jako Z_j dla $j = 1, 2, ..., J$. Korzystając z przedstawionego wcześniej algorytmu, dla każdego z nich można obliczyć przybliżoną wartość całki $\int_{Z_j} \hat{f}(x)\,dx$. Wskazanie, jaka część popytu każdego z obszarów Z_j powinna być obsłużona przez poszczególne stacje bazowe, tak aby zaspokoić możliwie jak najwięcej popytu przy uwzględnieniu ograniczonych przepustowości urządzeń nadawczo-odbiorczych, staje się wtedy typowym zadaniem programowania liniowego, możliwym do rozwiązania z użyciem klasycznej metody sympleks.

Niniejszym została skompletowana **procedura wyznaczania wskaźnika jakości** ustalonego układu stacji bazowych, uwzględniająca wszystkie sformu-

łowane na początku niniejszego podrozdziału uwarunkowania. Warto zwrócić uwagę, iż podstawową, wielokrotnie wykorzystywaną w powyższej procedurze własność stanowiła dogodna formuła estymatora jądrowego, będącego kombinacją liniową funkcji o postaci możliwej do wyboru zgodnie z uwarunkowaniami konkretnego zagadnienia aplikacyjnego, w przypadku rozważanego tu problemu – możliwości całkowania. Dysponując określonym powyżej wskaźnikiem jakości ustalonego układu stacji bazowych, można przystąpić do wyznaczenia optymalnego ich układu. W tym celu zostaną użyte metody właściwe badaniom operacyjnym.

W rozważanym ujęciu przyjęto możliwość wyboru w każdej z potencjalnych lokalizacji jednej spośród p możliwych wersji sprzętowych urządzeń nadawczo-odbiorczych (przy czym $p \in \mathbb{N} \setminus \{0\}$), różniących się parametrami użytkowymi. Niech zatem dany będzie k-wymiarowy wektor decyzyjny

$$[g_1, g_2, ..., g_k]. \tag{4.89}$$

Jego poszczególne współrzędne reprezentują potencjalne lokalizacje stacji bazowych i przyjmują wartości $g_j \in \{0, 1, ..., p\}$ dla $j = 1, 2, ..., k$. Jeżeli j-ta współrzędna tego wektora wynosi zero, oznacza to, iż w j-tej lokalizacji nie zakłada się zamontowania urządzeń nadawczo-odbiorczych, natomiast gdy współrzędna ta przyjmuje wartość q z zakresu $1, 2, ..., p$, to w j-tej lokalizacji przyjmuje się q-tą wersję sprzętową tych urządzeń. Poszukiwane będzie maksimum wyrażenia

$$\max_{g_1, g_2, ..., g_k} Q([g_1, g_2, ..., g_k]) \tag{4.90}$$

przy ograniczeniu

$$\sum_{j=1}^{k} c_{g_j} \leq \widetilde{C}, \tag{4.91}$$

gdzie $Q([g_1, g_2, ..., g_k])$ oznacza wartość funkcji (4.87) dla układu urządzeń nadawczo-odbiorczych umiejscowionych zgodnie z wartościami wektora decyzyjnego $[g_1, g_2, ..., g_k]$, dodatnie liczby c_q reprezentują koszt zakupu i montażu urządzeń q-tej wersji urządzeń nadawczo-odbiorczych, a $\widetilde{C}$ określa maksymalną wielkość dostępnych środków finansowych.

Powyższy problem optymalizacyjny można łatwo rozwiązać, stosując klasyczne metody badań operacyjnych. Niech zatem dane będzie k-poziomowe drzewo decyzyjne, przy czym jego poszczególne poziomy reprezentują kolejne potencjalne lokalizacje stacji bazowych. Rozwiązanie rozważanego problemu polega na wyznaczeniu ścieżki od węzła poziomu pierwszego do węzła poziomu

k-tego, opisanej za pomocą wektora $[g_1, g_2, ..., g_k]$, dla którego funkcja Q osiąga maksimum i spełnione jest ograniczenie (4.91). Do rozwiązania tak sformułowanego zadania może być wykorzystana klasyczna metoda podziału i ograniczeń.

Powyższe zadanie nietrudno jest uogólnić na przypadek **planowania z kilkuletnim horyzontem**, również przy niestacjonarności danych. I tak, jeżeli inwestycja jest rozpatrywana w obrębie $T \in \mathbb{N} \setminus \{0, 1\}$ lat, to wektor decyzyjny (4.89) może być uogólniony do postaci

$$[g_{1,t=1}, g_{2,t=1}, ..., g_{k_1,t=1}, g_{1,t=2}, g_{2,t=2}, ..., g_{k_2,t=2}, ..., g_{1,t=T}, g_{2,t=T}, ..., g_{k_T,t=T}], \tag{4.92}$$

gdzie parametr $t = 1, 2, ..., T$ charakteryzuje poszczególne rozważane lata. W zapisie tym dopuszcza się także, aby liczba potencjalnych lokalizacji stacji bazowych $k_1, k_2, ..., k_T$ była zmienna w czasie. Ograniczenie (4.91) przyjmuje wówczas postać T niezależnych warunków dotyczących poszczególnych lat, a wskaźnik jakości (4.87) staje się kombinacją liniową kolejnych, odpowiadających im składników. Przy określaniu powyższych wielkości dane charakteryzujące poszczególne lata mogą być zmienne.

4.3.3. Ujęcie rozmyte

Występujące we wzorze (4.84) współczynniki w_i reprezentują zapotrzebowanie na usługi teletransmisyjne przypisane poszczególnym lokalizacjom stacji abonenckich. Ich wartość szacowana jest głównie poprzez analizę rodzaju poszczególnych odbiorców usług, przy uwzględnieniu ich przynależności do określonych grup konsumenckich. Identyfikacja odbywa się w praktyce na podstawie oceny eksperta, wyrażanej werbalnie, opartej często jedynie na przesłankach intuicyjnych. W związku z powyższym do opisu przewidywanego w stacji abonenckiej zapotrzebowania na usługi teletransmisyjne zostaną użyte elementy **logiki rozmytej**. Jako specyfikę rozważanego tu zadania należy wziąć pod uwagę znaczną ilość koniecznych do zidentyfikowania i późniejszej analizy liczb rozmytych (równej liczbie stacji abonenckich), a także, iż rzadko – wręcz incydentalnie – niektóre współczynniki w_i mogą być deterministyczne, dzięki podpisanym wcześniej umowom. Szczególnie dogodne w tej sytuacji są **liczby rozmyte typu L-R**, których funkcja przynależności będzie zapisana w następującej postaci:

$$\mu_{(w_i,\alpha_i,\beta_i)}(x) = \begin{cases} L\left(\dfrac{w_i - x}{\alpha_i}\right), & \text{gdy} \quad x \le w_i, \\ R\left(\dfrac{x - w_i}{\beta_i}\right), & \text{gdy} \quad x > w_i, \end{cases} \tag{4.93}$$

przy czym w_i, α_i, $\beta_i > 0$, a funkcje $L : \mathbb{R} \to [0, 1]$ oraz $R : \mathbb{R} \to [0, 1]$ są symetryczne względem zera, przyjmują tam wartość 1 i są niemalejące na przedziale $(-\infty, 0]$. Parametr w_i można zinterpretować jako wartość maksymalną, natomiast α_i i β_i decydują, odpowiednio, o lewostronnym i prawostronnym skupieniu wokół tej wartości.

Liczbę rozmytą $\mathcal{A}$ typu L-R można więc utożsamiać z trzema parametrami, co oznaczane będzie jako $\mathcal{A} = (w, \alpha, \beta)$, i w konsekwencji proces jej identyfikacji wymaga określenia jedynie tych, bliskich intuicyjnej interpretacji wartości. Działania algebraiczne na liczbach rozmytych typu L-R są zdefiniowane następującymi zależnościami:

$$\mathcal{A} + \mathcal{B} = (w_{\mathcal{A}}, \alpha_{\mathcal{A}}, \beta_{\mathcal{A}}) + (w_{\mathcal{B}}, \alpha_{\mathcal{B}}, \beta_{\mathcal{B}}) = (w_{\mathcal{A}} + w_{\mathcal{B}}, \alpha_{\mathcal{A}} + \alpha_{\mathcal{B}}, \beta_{\mathcal{A}} + \beta_{\mathcal{B}}), \tag{4.94}$$

$$\mathcal{A} - \mathcal{B} = (w_{\mathcal{A}}, \alpha_{\mathcal{A}}, \beta_{\mathcal{A}}) - (w_{\mathcal{B}}, \alpha_{\mathcal{B}}, \beta_{\mathcal{B}}) = (w_{\mathcal{A}} - w_{\mathcal{B}}, \alpha_{\mathcal{A}} + \alpha_{\mathcal{B}}, \beta_{\mathcal{A}} + \beta_{\mathcal{B}}), \tag{4.95}$$

$$c \cdot \mathcal{A} = (cw_{\mathcal{A}}, c\alpha_{\mathcal{A}}, c\beta_{\mathcal{A}}), \tag{4.96}$$

przy czym $\mathcal{A}$ i $\mathcal{B}$ oznaczają liczby rozmyte, natomiast c – dodatnią liczbę rzeczywistą. Jeżeli przyjąć notację, w której liczba rzeczywista a jest zapisywana w postaci trzech parametrów $a = (a, 0, 0)$, to działania te mogą być uogólnione na dodawanie i odejmowanie liczby rozmytej i rzeczywistej. Co więcej, wzory (4.94)-(4.96) prawidłowo wyrażają wtedy także działania na dwóch liczbach rzeczywistych. Oznacza to zatem, że liczba rozmyta typu L-R w powyższym zakresie jest uogólnieniem liczby rzeczywistej. Na potrzeby przedstawionej tu procedury przyjęto

$$L(x) = R(x) = \begin{cases} 1 - x^2, & \text{gdy} \quad x \in [-1, 1], \\ 0, & \text{gdy} \quad x \in (-\infty, -1) \cup (1, \infty). \end{cases} \tag{4.97}$$

Ostatecznie, dla każdej z m lokalizacji budynków abonenckich wprowadzony we wzorach (4.83)-(4.84) współczynnik w_i, reprezentujący popyt na usługi transmisji danych, został uogólniony do dogodnej w identyfikacji i obliczeniach trójparametrycznej liczby rozmytej, oznaczanej poniżej jako $\mathcal{W}_i = (w_i, \alpha_i, \beta_i)$, przy czym $w_i - \alpha_i \geq 0$ dla każdego $i = 1, 2, ..., m$. W szczególnym przypadku $\mathcal{W}_i = (w_i, 0, 0)$ może reprezentować liczbę rzeczywistą (nierozmytą) w_i.

Jak wynika ze wzorów (4.85)-(4.88), wskaźnik jakości rozważanego układu stacji bazowych ma wówczas postać kombinacji liniowej trójparametrycznych liczb rozmytych $\mathcal{W}_i$, a zatem – z uwagi na wzory (4.94)-(4.96) – jego wartość staje się także trójparametryczną liczbą rozmytą, oznaczaną dalej jako $\mathcal{Q}$. W celu umożliwienia porównywania jakości poszczególnych układów stacji bazowych

zostaną użyte elementy **teorii preferencji rozmytych**. Funkcja preferencji $\mathcal{R}$ liczby rozmytej $\mathcal{Q}$ o ograniczonym nośniku funkcji przynależności będzie przyjęta w następującej postaci wynikłej z praktyki podejmowania decyzji:

$$\mathcal{R}(\mathcal{Q}) = \delta \frac{\int_{\min \operatorname{supp} \mu_{\mathcal{Q}}}^{\max \operatorname{supp} \mu_{\mathcal{Q}}} x\, \mu_{\mathcal{Q}}(x)\, dx}{\int_{\min \operatorname{supp} \mu_{\mathcal{Q}}}^{\max \operatorname{supp} \mu_{\mathcal{Q}}} \mu_{\mathcal{Q}}(x)\, dx} + (1-\delta) \min \operatorname{supp} \mu_{\mathcal{Q}}, \tag{4.98}$$

gdzie $\delta \in [0, 1]$, $\mu_{\mathcal{Q}}$ oznacza funkcję przynależności liczby rozmytej $\mathcal{Q}$, a $\operatorname{supp} \mu_{\mathcal{Q}}$ – jej nośnik. Wartość funkcji przynależności jest zatem kombinacją liniową z wagami δ i $1-\delta$ wartości średniej liczby rozmytej oraz wartości minimalnej jej nośnika. Wartość średnia odpowiada bayesowskiej regule decyzyjnej i wyraża działanie „realistyczne", natomiast wartość minimalna nośnika funkcji przynależności wynika z reguły minimaksu i reprezentuje działanie „pesymistyczne". Parametr δ stanowi zatem o strategii firmy w zakresie od „realistycznej" – zakładającej średni poziom przewidywanego popytu – dla $\delta = 1$, do „pesymistycznej" – przyjmującej najmniejszy przewidywany popyt – dla $\delta = 0$.

W przypadku liczby trójparametrycznej $\mathcal{Q} = (q, \alpha, \beta)$ dla funkcji L i R danych zależnością (4.97) – wartość funkcji preferencji (4.98) wyraża się wzorem

$$\mathcal{R}(\mathcal{Q}) = \delta\left(q + \frac{3(\beta-\alpha)}{8}\right) + (1-\delta)(q-\alpha) = q - \alpha + \delta\left(\frac{5\alpha+3\beta}{8}\right). \tag{4.99}$$

Jeżeli rozważane są dwa układy stacji bazowych scharakteryzowanych przez rozmyte wskaźniki jakości, to za „lepszy" należy uznać ten, dla którego wartość powyższej funkcji jest większa.

Ponownie warto zwrócić uwagę, że wprowadzenie do modelu nieprecyzyjności w postaci elementów logiki rozmytej i dostosowanie do ich obecności opracowywanej metody stało się możliwe w dużym stopniu dzięki dogodnej formule estymatora jądrowego. Wskaźnik jakości pozostał bowiem w klasie liczb rozmytych typu L-R, założonej wcześniej wobec współczynników w_i.

Uwagi bibliograficzne

Niniejszy podrozdział ujmuje wyniki badań prowadzonych przez autora wraz z Jackiem Waglowskim w ramach Jego pracy doktorskiej [66]. Powyższy tekst zawiera fragmenty tej rozprawy oraz wspólnej publikacji [40].

Obecnie nie istnieje uniwersalna metodyka umożliwiająca rozwiązanie rozważanego tu problemu planowania rozmieszczenia stacji bazowych systemu

LMDS. W praktyce stosuje się procedury heurystyczne, w dużym stopniu oparte na intuicji, lub adaptuje metody z pokrewnych dziedzin. Wiele z nich, wraz z odpowiednią literaturą przedmiotową, przedstawiono w pracy [56].

Poruszone zagadnienia z zakresu badań operacyjnych i programowania liniowego można znaleźć w podręcznikach [21, 67]. Teoria preferencji rozmytych jest przedmiotem książki [19]. Szczególnie wartą polecenia pozycję z zakresu logiki rozmytej stanowi klasyczna w tej dziedzinie monografia [27].

Rozdział 5

Podsumowanie i końcowe komentarze

W niniejszej książce została przedstawiona i szczegółowo rozważona koncepcja **statystycznych estymatorów jądrowych**. Dzięki możliwościom i powszechności współczesnej techniki komputerowej ten rodzaj estymatorów stał się obecnie najczęściej stosowanym narzędziem estymacji nieparametrycznej, umożliwiającym wyznaczanie charakterystyk zmiennych losowych bez arbitralnych założeń dotyczących typu ich rozkładu.

W podrozdziale 3.1 przedstawiono kompendium metodyki dotyczącej estymatorów jądrowych najczęściej stosowanej charakterystyki funkcyjnej rozkładu zmiennej losowej – gęstości prawdopodobieństwa. Sformułowane zostały definicja podstawowej postaci tego typu estymatora (punkt 3.1.1) i procedury określania występujących tam wielkości: wyboru postaci jądra (punkt 3.1.3) oraz wyznaczania wartości parametru wygładzania (punkt 3.1.4). Formuła ta została rozszerzona o koncepcję transformacji liniowej, użytecznej w przypadku wielowymiarowym, gdy stosuje się jądro radialne (punkt 3.1.5), a także procedurę modyfikacji parametru wygładzania (punkt 3.1.6), mających ogólny wpływ na polepszenie własności otrzymanego estymatora. Zaproponowano także procedury uzupełniające, pozwalające na ograniczenie jego nośnika (punkt 3.1.7) i uwzględnienie wielkości binarnych (punkt 3.1.8), co w praktycznych zastosowaniach nierzadko umożliwia lepsze dostosowanie modelu do badanej rzeczywistości. Powyższe uogólnienia i uzupełnienia mogą być dowolnie łączone, zgodnie z uwarunkowaniami konkretnego problemu aplikacyjnego.

Co więcej, ze względu na sugestywną formułę i interpretację estymatorów jądrowych możliwe jest również dokonanie dalszych ewentualnych modyfikacji, zgodnie z wymaganiami rozważanego zagadnienia. Jako przykłady przedstawiono ujęcie lokalne (punkt 4.2.3) oraz wprowadzenie współczynników charakteryzujących znaczenie poszczególnych elementów próby losowej, które mogą być także liczbami rozmytymi (punkt 4.3.3), co w konsekwencji pozwala na uwzględnienie dodatkowego aspektu nieprecyzyjności ocen.

Wyrazista i dogodna z analitycznego punktu widzenia koncepcja jądrowego estymatora gęstości prawdopodobieństwa umożliwia ponadto uzyskanie innych charakterystyk rozkładu zmiennej losowej, zarówno funkcyjnych, jak i jego parametrów. Jako przykłady zostały rozważone estymatory jądrowe dystrybuanty (punkt 3.2.1) oraz kwantyla (punkt 3.2.2). Powyższa koncepcja może być także wykorzystana do różnorodnych odmiennych zagadnień estymacji, między innymi do wyznaczania charakterystyk warunkowych (punkt 3.2.3).

Materiał rozdziału 3 niniejszej monografii **dostarcza pełnego zestawu procedur** pozwalających wyznaczyć estymatory jądrowe w przedstawionym powyżej zakresie. Poprzez odwołania do treści rozdziału 2 można także otrzymać wzory wszystkich potrzebnych do tego celu klasycznych estymatorów. Ponieważ publikacja ta jest przeznaczona dla szerokiego kręgu odbiorców, o różnym stopniu przygotowania z zakresu statystyki oraz metod analitycznych i numerycznych, więc jako pomocnicze algorytmy obliczeniowe zostały zaproponowane procedury powszechnie znane oraz łatwe do zaimplementowania, jak na przykład metoda złotego podziału lub algorytm Newtona, i to w ich podstawowych wersjach. Jednak w przypadku specyficznych uwarunkowań problemu, połączonych z możliwościami badawczymi zespołu projektującego, zalecane jest wykorzystanie dostępnych w literaturze modyfikacji lub wręcz użycie całkowicie odmiennych metod i algorytmów. Podobnie sugeruje się sukcesywną weryfikację uzyskiwanych wyników za pomocą doraźnych sprawdzianów dostosowanych do konkretnej sytuacji. Przede wszystkim polecić tu można ogląd wykresu projektowanego estymatora, łatwy do zrealizowania w przypadku jedno- i dwuwymiarowym, a w pozostałych przypadkach możliwy „na przekrojach", czyli po ustaleniu wartości współrzędnych poza jedną lub dwoma. Nic nie zastąpi bowiem ludzkiej inteligencji i wyczucia, dopasowujących ogólne wyniki analitycznych kryteriów optymalizacyjnych do wymagań konkretnego zadania.

Ograniczenia w praktycznych zastosowaniach estymatorów jądrowych wynikają głównie z trzech przesłanek. Podstawową stanowi sama struktura estymatorów tego typu, w szczególności konieczność szacowania rozkładów niesymetrycznych z użyciem symetrycznych jąder, a także ewentualna wielomodalność badanego rozkładu przy silnie zróżnicowanych składnikach, czemu nie sprzyja jednolita wartość parametrów wygładzania i modyfikujących oraz macierzy transformacji liniowej. Drugą przesłanką jest stopień odporności przedstawionej metodyki na poczynione na potrzeby poszczególnych jej procedur założenia, nie zawsze spełnione w praktyce (por. przykładowo warunek (3.13)). I wreszcie, jako najważniejsze ograniczenie stosowalności estymatorów jądrowych należy uznać wymagania dotyczące liczności próby losowej, szczególnie drastyczne w przypadku wielowymiarowych zmiennych losowych. Nawet przy zapewnieniu odpowiednio licznej próby problem ten może się uwidocznić poprzez wynikłe stąd trudności obliczeniowe, istotne zwłaszcza w systemach pracujących

w czasie rzeczywistym, a więc bezwzględnie wymagających krótkiego czasu obliczeń. Szczęśliwie w miarę rozwoju techniki komputerowej oraz aparatury do pomiaru, transmisji i gromadzenia danych powstałe stąd utrudnienia będą ulegać naturalnej redukcji.

Zakres zastosowań estymatorów jądrowych może być znacząco poszerzany poza rozważane w niniejszej monografii charakterystyki probabilistyczne. Poniżej zostanie pokrótce przedstawionych pięć przykładów o sukcesywnie zwiększającej się odmienności od prezentowanych wcześniej zagadnień.

Jako pierwszy, w dużym stopniu naturalny przykład rozważona zostanie często używana w teorii odnowy i niezawodności funkcja $r: \mathbb{R} \to [0, \infty)$ zwana **intensywnością awarii**, zdefiniowana wzorem

$$r(x) = \frac{f(x)}{1 - F(x)} \quad \text{dla każdego } x, \text{ takiego że } F(x) \neq 1, \tag{5.1}$$

gdzie f oraz F oznaczają, odpowiednio, gęstość prawdopodobieństwa i dystrybuantę jednowymiarowej zmiennej losowej charakteryzującej bezawaryjny czas pracy danego urządzenia. Wartość funkcji r dla argumentu x stanowi zatem o prawdopodobieństwie wystąpienia awarii w chwili x, pod warunkiem że urządzenie to nie uległo uszkodzeniu do tego czasu. Dysponując jądrowymi estymatorami gęstości $\hat{f}$ i dystrybuanty $\hat{F}$, skonstruowanymi z użyciem jądra o dodatnich wartościach, nietrudno jest wyznaczyć jądrowy estymator intensywności awarii zgodnie z zależnością

$$\hat{r}(x) = \frac{\hat{f}(x)}{1 - \hat{F}(x)}. \tag{5.2}$$

Nieco bardziej złożonym, drugim przykładem jest estymacja **funkcjonałów gęstości prawdopodobieństwa** n-wymiarowej zmiennej losowej, wykorzystywanych w różnych nietypowych problemach statystyki, między innymi całki kwadratu funkcji f: $\int_{\mathbb{R}^n} f^2(x)\,dx$. Oczywiście możliwe jest numeryczne całkowanie powyższej funkcji, ale nietrudne przekształcenia pozwalają wyznaczyć dogodny wzór określający jądrowy estymator tego funkcjonału, w podstawowej postaci dany jako

$$\int_{\mathbb{R}^n} \hat{f}^2(x)\,dx = \frac{1}{m^2 h^n} \sum_{i=1}^{m} \sum_{j=1}^{m} K^{*2}\left(\frac{x_i - x_j}{h}\right), \tag{5.3}$$

gdzie K^{*2} oznacza kwadrat splotowy n-wymiarowego jądra K, a x_i i x_j stanowią i-ty oraz j-ty element dostępnej próby losowej.

Sama idea estymatorów jądrowych może być również użyta do polepszania własności powszechnie znanych i stosowanych estymatorów. W celu zilustrowania tego aspektu w ramach trzeciego przykładu zostanie rozpatrzony jądrowy estymator **gęstości spektralnej**, czyli mierzalnej funkcji $f\colon [0, \infty) \to [0, \infty)$, takiej że wartość całki $\int_a^b f(\omega)\,d\omega$ interpretuje się jako łączną moc – a ściślej wariancję – składowych okresowych badanego sygnału o częstotliwościach ω z przedziału $[a, b]$. Niech zatem próba losowa $x_1, x_2, \ldots, x_m$ składa się z elementów uzyskiwanych kolejno w równych odstępach czasu. Najczęściej stosowane estymatory gęstości spektralnej są definiowane jako funkcje $\hat{f}:[0,\pi] \to \mathbb{R}$ następującej postaci:

$$\hat{f}(\omega) = \frac{1}{2\pi} \sum_{i=-m+1}^{m-1} c_i \left(1 - \frac{|i|}{m}\right) g(|i|) \cos(i\omega), \tag{5.4}$$

przy czym $c_i \in \mathbb{R}$ oraz

$$g(i) = \frac{1}{m-i} \sum_{j=1}^{m-i} x_j x_{j+i} - \left(\frac{1}{m} \sum_{j=1}^{m} x_j\right)^2 \quad \text{dla } i = 1, 2, \ldots, m-1. \tag{5.5}$$

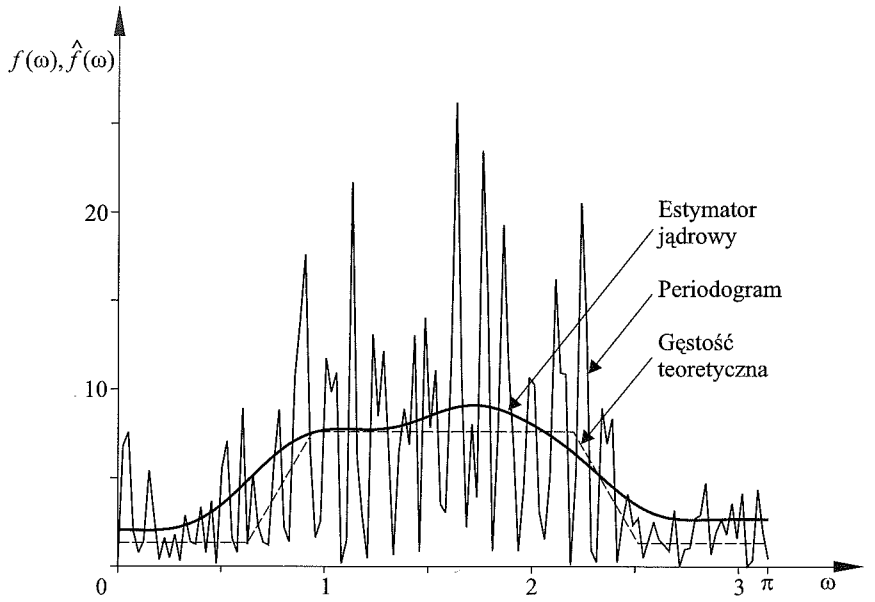
Jeżeli $c_i \equiv 1$, to powyższe wzory reprezentują najprostszą postać estymatora, zwaną **periodogramem** (por. rys. 5.1). Jego wykres ma kształt bardzo nieregularny i może przyjmować wartości ujemne, co pozostaje w jawnej sprzeczności z własnościami gęstości spektralnej. Aby wyeliminować powyższe niedogodności, można wprowadzić koncepcję jądrowego estymatora gęstości spektralnej, powstałą z zastosowania wobec periodogramu zasadniczej idei estymatorów jądrowych. Przekształcenia oparte na transformacji Fouriera implikują, iż współczynniki c_i przyjmują postać

$$c_i = \widetilde{K}\left(\frac{i}{h}\right), \tag{5.6}$$

gdzie $\widetilde{K}:\mathbb{R} \to [0,\infty)$ jest odwrotną transformatą Fouriera pewnego jądra K, a dodatnia stała h odgrywa rolę parametru wygładzania. Korzystne własności charakteryzują funkcję $\widetilde{K}$ postaci

$$\widetilde{K}(x) = \begin{cases} 1 - 6x^2 + 6|x|^3, & \text{gdy } |x| \le 0{,}5, \\ 2(1-|x|)^3, & \text{gdy } 0{,}5 \le |x| \le 1, \\ 0, & \text{gdy } |x| \ge 1. \end{cases} \tag{5.7}$$

Wartości otrzymanego powyżej jądrowego estymatora gęstości spektralnej danego wzorami (5.4)-(5.7) są nieujemne, a wykres odpowiednio wygładzony poprzez stosowny dobór wartości parametru h. Ilustrację powyższej koncepcji stanowi rysunek 5.1.



Rys. 5.1. Jądrowy estymator gęstości spektralnej (zaznaczono także periodogram i gęstość teoretyczną)

Koncepcja estymatorów jądrowych może być także wykorzystana do wyznaczania charakterystyk probabilistycznych zasadniczo odmiennych od prezentowanych dotychczas w ramach niniejszej monografii. Jako czwarty przykład zostanie rozpatrzone powszechnie znane zagadnienie **regresji liniowej**. Niech zatem dane będą dwie jednowymiarowe zmienne losowe X i Y oraz uzyskana z ich zestawienia $\begin{bmatrix} Y \\ X \end{bmatrix}$ próba losowa

$$\begin{bmatrix} y_1 \\ x_1 \end{bmatrix}, \begin{bmatrix} y_2 \\ x_2 \end{bmatrix}, \ldots, \begin{bmatrix} y_m \\ x_m \end{bmatrix}. \tag{5.8}$$

Przedmiotem rozważań jest tu funkcja $f: \mathbb{R} \to \mathbb{R}$, która najlepiej charakteryzowałaby ewentualną zależność między zmiennymi losowymi X oraz Y. Jej estymator $\hat{f}: \mathbb{R} \to \mathbb{R}$ wyznacza się tak, aby zminimalizować wyrażenie $\sum_{i=1}^{m} [y_i - \hat{f}(x_i)]^2$.

Można w tym miejscu poczynić dwie dodatkowe uwagi. Otóż, poniższe wyniki pozostałyby niezmienione, gdyby przyjąć, iż wartości $x_1, x_2, ..., x_m$ nie są eksperymentalnie uzyskanymi wartościami zmiennej losowej X, lecz arbitralnie ustalonymi wielkościami, dla których otrzymano kolejne wartości zmiennej losowej Y, czyli $y_1, y_2, ..., y_m$. I ponadto, poszczególnym elementom próby losowej (5.8) można przyporządkować nieujemne wagi $w_1, w_2, ..., w_m$, charakteryzujące ich znaczenie. Na przykład, jeśli model ma być użyty do problemu prognozowania, to warto przyporządkować takie wagi proporcjonalnie do aktualności poszczególnych elementów owej próby. Ze względu na spójność notacji wagi te zapisuje się jako odpowiednie elementy głównej przekątnej (diagonali) macierzy $\mathcal{W}$ o wymiarze $m \times m$, czyli

$$\mathcal{W} = \mathrm{diag}(w_1, w_2, \ldots, w_m). \tag{5.9}$$

Tradycyjne, parametryczne metody wyznaczania funkcji regresji wymagają arbitralnego ustalenia postaci funkcji f, a w konsekwencji także $\hat{f}$, po czym wyznaczenia występujących tu współczynników. Najczęściej przyjmowana jest postać liniowa tych funkcji i wówczas otrzymuje się

$$\hat{f}(x) = e[\mathcal{X}(x)^{\mathrm{T}} \mathcal{W} \mathcal{X}(x)]^{-1} \mathcal{X}(x)^{\mathrm{T}} \mathcal{W} y, \tag{5.10}$$

gdzie

$$e = [1, 0], \tag{5.11}$$

$$\mathcal{X}(x) = \begin{bmatrix} 1 & x_1 - x \\ 1 & x_2 - x \\ \vdots & \vdots \\ 1 & x_m - x \end{bmatrix}, \tag{5.12}$$

$$y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_m \end{bmatrix}, \tag{5.13}$$

przy założeniu nieosobliwości macierzy $\mathcal{X}(x)^{\mathrm{T}} \mathcal{W} \mathcal{X}(x)$.

Koncepcja jądrowego estymatora funkcji regresji polega na uzależnieniu współczynników w_i od zmiennej niezależnej x i – w podstawowej postaci – określenie ich zależnością

$$w_i(x) = \frac{1}{h} K\left(\frac{x - x_i}{h}\right). \tag{5.14}$$

Tak więc poszczególnym elementom próby losowej (5.8) są przypisane wagi o wartości tym większej, im współrzędna x_i jest bliższa argumentowi x. Macierz (5.9) przyjmuje wówczas postać

$$\mathcal{W}(x) = \operatorname{diag}\left(\frac{1}{h}K\left(\frac{x-x_1}{h}\right), \frac{1}{h}K\left(\frac{x-x_2}{h}\right), \ldots, \frac{1}{h}K\left(\frac{x-x_m}{h}\right)\right). \tag{5.15}$$

Co więcej, możliwe jest wtedy wyrażenie wzoru (5.10) w dogodnej formie analitycznej

$$\hat{f}(x) = \frac{1}{mh}\sum_{i=1}^{m} \frac{[g_2(x) + g_1(x)(x-x_i)]K\left(\frac{x-x_i}{h}\right)y_i}{g_2(x)g_0(x) - g_1^2(x)}, \tag{5.16}$$

gdzie

$$g_k(x) = \frac{1}{mh}\sum_{i=1}^{m}(x_i - x)^k K\left(\frac{x-x_i}{h}\right) \quad \text{dla } k = 0, 1, 2. \tag{5.17}$$

Rysunek 5.2 stanowi ilustrację powyższej koncepcji. Pokazane są tam przykładowa estymowana funkcja $f\colon [0, 1] \to \mathbb{R}$ dana wzorem $f(x) = -0{,}8\cos(13x+1) - -8\exp(-0{,}2(2{,}6x-1{,}6)^2) + 9{,}23$, elementy próby losowej (5.8) uzyskane przez dodanie do poszczególnych wartości $f(x_i)$ składowej losowej, a także klasyczna funkcja regresji (5.10) oraz jej jądrowy estymator (5.16)-(5.17). Widać, iż ten ostatni nie jest ograniczony żadnym arbitralnym założeniem dotyczącym postaci funkcji $\hat{f}$.

I wreszcie, ostatni – piąty – przykład stanowi ciekawe, nietypowe zastosowanie estymatorów jądrowych do tak zwanych **metod bootstrapowych**. Ogólny zamysł opiera się tu na znanym fakcie, że jeżeli m^*-elementową próbę losową

$$x_1^*, x_2^*, \ldots, x_{m^*}^* \tag{5.18}$$

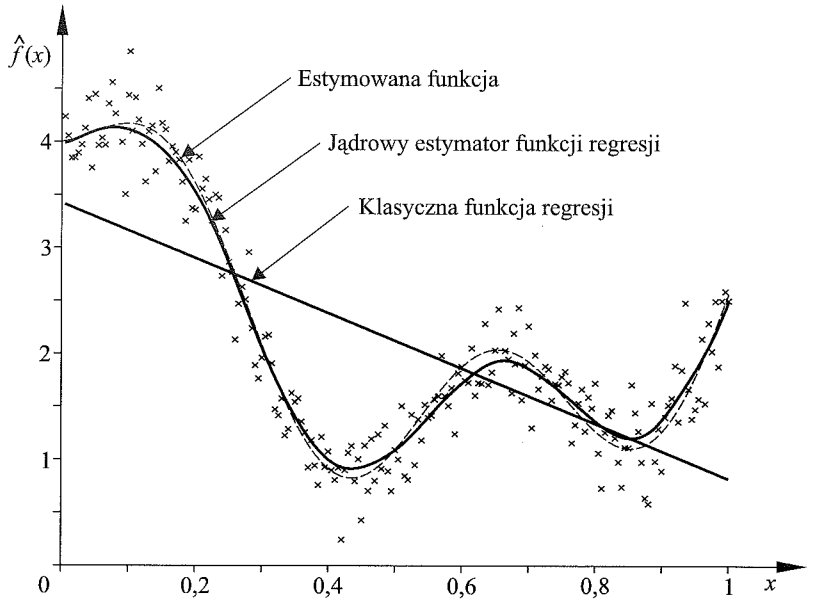
otrzymano z generatora liczb pseudolosowych o rozkładzie jednostajnym, a odwracalna funkcja F jest dystrybuantą pewnego rozkładu, to można uznać, iż próba

$$F^{-1}(x_1^*), F^{-1}(x_2^*), \ldots, F^{-1}(x_{m^*}^*) \tag{5.19}$$

pochodzi z generatora o rozkładzie scharakteryzowanym przez dystrybuantę F.

Niech zatem dana będzie jednowymiarowa zmienna losowa X o rozkładzie mającym dystrybuantę F, a także uzyskana z niej m-elementowa próba losowa

$$x_1, x_2, \ldots, x_m. \tag{5.20}$$



Rys. 5.2. Jądrowy estymator funkcji regresji (zaznaczono również klasyczną funkcję regresji i estymowaną funkcję f oraz, krzyżykami, wartości próby losowej)

Na jej podstawie można otrzymać jądrowy estymator dystrybuanty $\hat{F}$. Jeśli do jego konstrukcji jest używane jądro o dodatnich wartościach, to istnieje funkcja odwrotna $\hat{F}^{-1}$. Otrzymując z generatora liczb pseudolosowych o rozkładzie jednostajnym próbę losową (5.18), uzyskuje się m^*-elementową próbę

$$\hat{F}^{-1}(x_1^*),\ \hat{F}^{-1}(x_2^*),\ \ldots,\ \hat{F}^{-1}(x_{m^*}^*) \tag{5.21}$$

o rozkładzie mającym dystrybuantę F, oczywiście „z dokładnością" jej oszacowania estymatorem jądrowym $\hat{F}$. Jeżeli zatem m^* jest znacząco większe niż m, to liczność dostępnej próby (5.21) ulega istotnemu zwiększeniu względem (5.20). Ograniczenie stanowi tu jednak potęgowanie się błędu wynikłego z oszacowania dystrybuanty F estymatorem jądrowym $\hat{F}$. W praktyce stosowanie powyższej metody pozwala na uzyskanie korzyści porównywalnych nawet z kilkakrotnym zwiększeniem liczności dostępnej próby losowej (5.20). Nawiązując do materiału prezentowanego w niniejszej książce, wielkość $\hat{F}^{-1}(x_i^*)$ dla $i = 1, 2, \ldots, m^*$ można wyznaczyć jako wartość jądrowego estymatora kwantyla rzędu x_i^*.

Na koniec warto wspomnieć, iż estymatory jądrowe są coraz częściej używane do konstrukcji specjalistycznych algorytmów z zakresu analizy danych, w szczególności klastryfikacji (łączenia elementów próby losowej w grupy/klasy

o podobnych cechach) i klasyfikacji (określania przynależności ustalonych elementów do wyróżnionych grup/klas, czyli w podobnym, aczkolwiek ogólniejszym sensie, niż przedstawiono to w punkcie 4.2.2 w odniesieniu do elementów nietypowych).

Podsumowując, metodyka konstruowania estymatorów jądrowych dostarcza nie tylko aparatu matematycznego umożliwiającego uzyskanie standardowych charakterystyk rozkładu zmiennych losowych, ale może być także wykorzystana do otrzymania wielu innych wielkości stosowanych w różnorodnych zagadnieniach praktycznych. Sama idea tego typu estymatorów często bywa użyteczna przy polepszeniu własności tradycyjnych elementów klasycznej statystyki matematycznej oraz konstruowaniu specjalistycznych procedur właściwych współczesnej statystyce i analizie danych. Dogodna formuła estymatora jądrowego, określona jako kombinacja liniowa odpowiednich funkcji, w połączeniu z możliwością swobodnego wyboru ich postaci, umożliwia szerokie spektrum aplikacji, również spoza zakresu związanego z probabilistyką, w szczególności w tych zagadnieniach, gdzie wymagane jest oszacowanie gęstości miary skończonej (tj. takiej, iż miara całej przestrzeni nie jest równa nieskończoności).

W rozdziale 4 niniejszej monografii zostały przedstawione trzy przykładowe **zastosowania estymatorów jądrowych do zagadnień współczesnej analizy systemowej**. Najpierw przedmiotem rozważań był klasyczny problem identyfikacji parametrów modelu, aczkolwiek dla przypadku gdy jego przeszacowanie i niedoszacowanie implikują istotnie odmienne skutki w rozważanym systemie. Następnie rozpatrywano zadanie wykrywania oraz klasyfikacji elementów nietypowych (odosobnionych) za pomocą zaprojektowanych do tego celu testów statystycznych. Jako ostatni zaprojektowano algorytm wyznaczania optymalnego układu stacji bazowych bezprzewodowego systemu transmisji danych, na podstawie przestrzennej charakterystyki rozkładu popytu w obecności nieprecyzyjnych jego ocen. Specyfika powyższych zagadnień wydaje się być typowa dla współczesnej analizy systemowej. Cechuje ją bowiem interdyscyplinarność przedmiotu badań, wszechobecność różnorodnych sposobów opisu nieokreśloności i wreszcie, podporządkowanie stosowanych metod nowoczesnym technikom informacyjnym.

Uzupełnienia oraz errata tekstu niniejszej monografii zostaną udostępnione pod adresem internetowym http://www.ibspan.waw.pl/kulczycki/EJwAS_05. Autor będzie wdzięczny za wszelkie **sugestie i uwagi** przekazywane na adres: kulczycki@ibspan.waw.pl lub kulczycki@pk.edu.pl.

Uwagi bibliograficzne

Rozliczne aspekty związane z teorią i zastosowaniami estymatorów jądrowych są dostępne w klasycznych monografiach [62, 68], a także w zróżnicowanej

pod względem tematyki i sposobu prezentacji literaturze [10, 50, 54, 59, 60] oraz krótkich wzmiankach znajdujących się w polskojęzycznych pozycjach [13, 20, 36].

Radykalne zmiany, jakie ostatnio nastąpiły zarówno w samym przedmiocie, jak i metodyce analizy systemowej, można ocenić, porównując opublikowaną 20 lat temu książkę [1] z jej współczesnym odpowiednikiem [65].

Bibliografia

[1] *Analiza systemowa – podstawy i metodologia.* Red. W. Findeisen. PWN, Warszawa 1985.
[2] Athans M., Falb P.L.: *Sterowanie optymalne.* WNT, Warszawa 1969.
[3] Barnett V., Lewis T.: *Outliers in Statistical Data.* Wiley, Chichester 1994.
[4] Bartoszewicz J.: *Wykłady ze statystyki matematycznej.* PWN, Warszawa 1989.
[5] Billingsley P.: *Prawdopodobieństwo i miara.* PWN, Warszawa 1987.
[6] Brandt S.: *Analiza danych.* PWN, Warszawa 1998.
[7] Bubnicki Z.: *Teoria i algorytmy sterowania.* PWN, Warszawa 2002.
[8] Charytanowicz M.: Bayesowska determinizacja informacji nieprecyzyjnej w zagadnieniach medycznych. Praca doktorska. Instytut Badań Systemowych PAN, Warszawa 2005.
[9] Dahlquist G., Bjork A.: *Metody numeryczne.* PWN, Warszawa 1983.
[10] Devroe L., Gyorfi L.: *Nonparametric Density Estimation: the L_1 View.* Wiley, New York 1985.
[11] *Diagnostyka procesów.* Red. J. Korbicz, J.M. Kościelny, Z. Kowalczuk, W. Cholewa. WNT, Warszawa 2002.
[12] Domański C.: *Testy statystyczne.* PWE, Warszawa 1990.
[13] Domański C., Pruska K.: *Nieklasyczne metody statystyczne.* PWE, Warszawa 2000.
[14] *Encyklopedia dla wszystkich – matematyka.* WNT, Warszawa 2000.
[15] Feller W.: *Wstęp do rachunku prawdopodobieństwa.* PWN, Warszawa 1980 (tom I), 1978 (tom II).
[16] Fichtenholz G.M.: *Rachunek różniczkowy i całkowy.* PWN, Warszawa 1978.
[17] Findeisen W., Szymanowski J., Wierzbicki A.: *Teoria i metody obliczeniowe optymalizacji.* PWN, Warszawa 1980.
[18] Fisz M.: *Rachunek prawdopodobieństwa i statystyka matematyczna.* PWN, Warszawa 1969.
[19] Fodor J., Roubens M.: *Fuzzy Preference Modelling and Multicriteria Decision Support.* Kluwer, Dordrecht 1994.
[20] Gajek L., Kałuszka M.: *Wnioskowanie statystyczne.* WNT, Warszawa 2000.
[21] Garfinkel R.S., Nemhauser G.L.: *Programowanie całkowitoliczbowe.* PWN, Warszawa 1978.
[22] Greń J.: *Statystyka matematyczna.* PWN, Warszawa 1987.
[23] Gutenbaum J.: *Modelowanie matematyczne systemów.* EXIT, Warszawa 2003.
[24] Hajek O.: Discontinuous Differential Equations, I-II. *Journal of Differential Equations*, 1979, Vol. 32, s. 149–185.
[25] Hermes H.: Discontinuous Vector Fields and Feedback Control. W: *Differential Equations and Dynamical Systems.* Red. J.K. Halle, J.P. LaSalle, Academic Press, New York 1967, s. 155–165.

[26] Hryniewicz O.: *Wykłady ze statystyki dla studentów informatycznych technik zarządzania*. Wydawnictwo WSISiZ, Warszawa 2004.
[27] Kacprzyk J.: *Zbiory rozmyte w analizie systemowej*. PWN, Warszawa 1986.
[28] Kacprzyk J.: *Wieloetapowe sterowanie rozmyte*. WNT, Warszawa 2001.
[29] Kaczorek T.: *Teoria sterowania*. PWN, Warszawa 1977 (tom I), 1981 (tom II).
[30] Kiełbasiński A., Schwetlick H.: *Numeryczna algebra liniowa*. WNT, Warszawa 1992.
[31] Klamka J.: *Controllability of Dynamical Systems*. PWN, Warszawa / Kluwer, Dordrecht 1991.
[32] Kołodziej W.: *Analiza matematyczna*. PWN, Warszawa 1978.
[33] Kozłowski K., Dutkiewicz P., Wróblewski W.: *Modelowanie i sterowanie robotów*. PWN, Warszawa 2003.
[34] Koronacki J., Mielniczuk J.: *Statystyka*. WNT, Warszawa 2001.
[35] Kulczycki P.: *Czasowooptymalne sterowanie stochastyczne nieciągłym układem dynamicznym*. Wydawnictwa PK, Kraków 1992.
[36] Kulczycki P.: *Wykrywanie uszkodzeń w systemach zautomatyzowanych metodami statystycznymi*. Alfa, Warszawa 1998.
[37] Kulczycki P., Charytanowicz M.: Bayes Sharpening of Imprecise Information. *Applied Mathematics and Computer Science*, 2005, Vol. 15 (w druku).
[38] Kulczycki P., Mazgaj A.: An Algorithm for Bayes Parameter Identification with Quadratic Nonsymmetrical Loss Function. *Control and Cybernetics*, 2005, Vol. 34 (w druku).
[39] Kulczycki P., Prochot C.: Wykrywanie elementów odosobnionych za pomocą metod estymacji nieparametrycznej. W: *Badania operacyjne i systemowe, podejmowanie decyzji – podstawy teoretyczne i zastosowania*. Red. R. Kulikowski, J. Kacprzyk, R. Słowiński, EXIT, Warszawa 2004, s. 313–328.
[40] Kulczycki P., Wagłowski J.: On the Application of Statistical Kernel Estimators for the Demand-Based Design of a Wireless Data Transmission System. *Control and Cybernetics*, 2005, Vol. 34 (w druku).
[41] Kulikowski R.: *Sterowanie w wielkich systemach*. WNT, Warszawa 1974.
[42] Lehmann E.L.: *Teoria estymacji punktowej*. PWN, Warszawa 1991.
[43] Lindgren B.W.: *Elementy teorii decyzji*. PWN, Warszawa 1977.
[44] Łojasiewicz S.: *Wstęp do teorii funkcji rzeczywistych*. PWN, Warszawa 1973.
[45] Mańczak K., Nahorski Z.: *Komputerowa identyfikacja obiektów dynamicznych*. PWN, Warszawa 1983.
[46] Mazgaj A.: Modelowanie niepewności parametrów obiektu dla potrzeb sterowania optymalnego. Praca doktorska. Akademia Górniczo-Hutnicza, Wydział Elektrotechniki, Automatyki, Informatyki i Elektroniki, Kraków 2005.
[47] Mitkowski W.: *Stabilizacja systemów dynamicznych*. WNT, Warszawa 1991.
[48] Musielak J.: *Wstęp do analizy funkcjonalnej*. PWN, Warszawa 1976.
[49] Muszyński J., Myszkis A.D.: *Równania różniczkowe zwyczajne*. PWN, Warszawa 1984.
[50] Nadaraya E.A.: *Nonoparametric Estimation of Probability Densities and Regression Curves*. Kluwer, Dordrecht 1989.
[51] Pelczar A., Szarski J.: *Wstęp do teorii równań różniczkowych*. PWN, Warszawa 1987 (tom I), 1989 (tom II).
[52] Piłat B., Wasilewski M.J.: *Tablice całek*. WNT, Warszawa 1985.
[53] Plucińska A., Pluciński E.: *Probabilistyka*. WNT, Warszawa 2000.
[54] Prakasa Rao B.L.S.: *Nonparametric Functional Estimation*. Academic Press, New York 1983.
[55] Prochot C.: Identyfikacja stanów nietypowych w oparciu o metodykę estymatorów jądrowych. Praca doktorska. Instytut Badań Systemowych PAN, Warszawa (w przygotowaniu).

[56] *Radio Network Planning and Optimization for UMTS*. Red. J. Laiho, A. Wacker, T. Novosad, Wiley, New York 2001.
[57] Ralston A.: *Wstęp do analizy numerycznej*. PWN, Warszawa 1983.
[58] Ryżik I.M., Gradsztejn I.S.: *Tablice całek, sum, szeregów i iloczynów*. PWN, Warszawa 1964.
[59] Scholkopf B., Smola A.J.: *Learning with Kernels*. MIT Press, Cambridge 2002.
[60] Scott D.W.: *Multivariate Density Estimation: Theory, Practice, and Visualization*. Wiley, New York 1992.
[61] Seidler J., Badach A., Molisz W.: *Metody rozwiązywania zadań optymalizacji*. WNT, Warszawa 1980.
[62] Silverman B.W.: *Density Estimation for Statistics and Data Analysis*. Chapman and Hall, London 1986.
[63] Stoer J., Bulirsch R.: *Wstęp do analizy numerycznej*. PWN, Warszawa 1987.
[64] Tadeusiewicz R., Ogiela M.R.: *Medical Image Understanding Technology*. Springer, Berlin 2004.
[65] *Techniki informacyjne w badaniach systemowych*. Red. P. Kulczycki, O. Hryniewicz, J. Kacprzyk (ukaże się nakładem WNT w 2006 r.).
[66] Waglowski J.: Metoda planowania optymalnego układu stacji bazowych bezprzewodowego systemu transmisji danych LMDS. Praca doktorska. Instytut Badań Systemowych PAN, Warszawa 2005.
[67] Wagner H.M.: *Badania operacyjne*. PWE, Warszawa 1980.
[68] Wand M.P., Jones M.C.: *Kernel Smoothing*. Champan and Hall, London 1995.
[69] Wierzbicki P.: *Modele i wrażliwość układów sterowania*. WNT, Warszawa 1977.

Skorowidz

σ-algebra 22
– zbiorów borelowskich 23

Analiza systemowa 9-11

Bezprzewodowy system transmisji danych LMDS 195-205
binarny parametr wygładzania 100
błąd drugiego rodzaju 187
– pierwszego rodzaju 187

Charakterystyki funkcyjne rozkładu prawdopodobieństwa 25-28

Decyzja bayesowska 167
– minimaksowa 167
dystrybuanta 26-28

Efektywność jądra 66
element nietypowy (odosobniony) 185-186
estymacja nieparametryczna 11-19
– parametryczna 12, 41-42
estymator 35
– asymptotycznie nieobciążony 37
– Fouriera 14-15
– jądrowy 15-19, 55-163
– – całki kwadratu gęstości prawdopodobieństwa 209
– – dystrybuanty 135-142
– – – warunkowej 150-152
– – funkcji regresji 211-213
– – gęstości prawdopodobieństwa 56-59
– – – – warunkowej 147-150, 152
estymator jądrowy gęstości spektralnej 210-211
– – kwantyla 142-145
– – – warunkowego 151-152
– –, ujęcie lokalne 193-195
– –, – rozmyte 202-204
– kowariancji 39-40
– kwantyla, klasyczny 40-41
– najbliższego sąsiedztwa 13-14
– nieobciążony 36
– odchylenia standardowego 38-39
– wariancji 38-39
– wartości oczekiwanej 38-39
– zgodny 37

Funkcja absolutnie (bezwzględnie) ciągła 45-46
– borelowska 23
– intensywności awarii 209
– mierzalna 23
– obszarami ciągła 49
– pierwotna jądra Cauchy'ego 138
– – – dwuwagowego 137
– – – Epanecznikowa 137
– – – jednostajnego 137
– – – trójkątnego 138
– sklejana (*spline*) 13
– strat 166
– – bayesowskich 167
– –, kwadratowa niesymetryczna 170-171, 176
– –, liniowa niesymetryczna 169
– – minimaksowych 166-167

Gęstość prawdopodobieństwa 11-12, 25-26
– spektralna 210

Hipoteza alternatywna 187
– statystyczna 186
histogram 12-13

Identyfikacja wartości parametru 168-185

Jądro 56
– binarne 99-100
– Cauchy'ego 138, 141-142
– dwuwagowe 65-66
– Epanecznikowa 65
– jednostajne 65
– normalne 66
– produktowe 70
– radialne 68-70
– trójkątne 66
jądrowy estymator dystrybuanty 135-142
– – funkcji regresji 211-213
– – gęstości prawdopodobieństwa 56-59
– – – spektralnej 210-211
– – kwantyla 142-145
– – warunkowego kwantyla 151-152
– – warunkowej dystrybuanty 150-152
– – – gęstości prawdopodobieństwa 147-150, 152
jednoznaczność rozwiązania równania różniczkowego 44, 46-47

Klasyczne rozwiązanie równania różniczkowego 44
klasyczny estymator kwantyla 40-41
klasyfikacja elementów nietypowych 186, 190-193
kombinacja liniowa jąder 68, 138
kowariancja 30-31
kryterium błędu średniokwadratowego MSE 59-60
– scałkowanego błędu średniokwadratowego MISE 60
kwadratowa niesymetryczna funkcja strat 170-171, 176
kwantyl 31

Lewostronne ograniczenie nośnika estymatora jądrowego 94-97
liczby rozmyte typu L-R 202-203
liczność próby losowej 34
liczność próby losowej na potrzeby estymatora jądrowego 103-107
liniowa niesymetryczna funkcja strat 169
liniowo zależne współrzędne zmiennej losowej 30
lokalne ujęcie estymatora jądrowego 193--195
– – transformacji liniowej 194-195

Macierz kowariancji 30-31
mediana 31
metoda krzyżowego uwiarygodnienia (*cross--validation*) 83-89
– (algorytm) Newtona 143-145, 151-152, 156-157, 163, 172-173, 179
– podstawień (*plug-in*) 80-83
– przybliżona 77-80
– złotego podziału 86-89
metody bootstrapowe 213-214
miara probabilistyczna 22
modyfikacja parametru wygładzania 90-93
moment centralny 28-29
– zwykły 28-29

Nieciągłość pierwszego rodzaju 44
niezależne zmienne losowe 32-33
nośnik gęstości prawdopodobieństwa 26
– zmiennej losowej 26

Obciążenie estymatora 36
odchylenie ćwiartkowe 31
– standardowe 29-30
ograniczenie nośnika estymatora jądrowego 94-97
– – – – lewostronne 94-97
– – – – prawostronne 97

Parametr wygładzania 56
– –, binarny 100
parametry charakterystyczne rozkładu prawdopodobieństwa 28-32
– modyfikujące 91
– warunkujące 149
periodogram 210
pierwiastek macierzy 75
postać diagonalna transformacji liniowej 73
– pełna transformacji liniowej 73
powłoka domknięta wypukła 46
poziom istotności 187

prawdopodobieństwo 23
prawostronne ograniczenie nośnika estymatora jądrowego 97
prosta próba losowa 34
próba losowa 34
– –, prosta 34
przestrzeń probabilistyczna 23

Reguła bayesowska 167
– minimaksowa 167
rozkład prawdopodobieństwa 23
– unimodalny (jednomodalny) 26
– wielomodalny 26
rozmyte ujęcie estymatora jądrowego 202--204
rozwiązanie równania różniczkowego 43-54
– – –, klasyczne 44
– – – w sensie Caratheodory'ego (C-rozwiązanie) 46, 181
– – – – Filippowa (F-rozwiązanie) 46, 181
– – – – Krasowskiego (K-rozwiązanie) 46, 181
równanie różniczkowe 44
– – z nieciągłą prawą stroną 45

Sklejanie rozwiązań 44-45
statystyka 35
sterowanie minimalnoczasowe 181-182
– optymalne 50-54, 180-184
– – z kwadratowym wskaźnikiem jakości 182-184
symetria radialna 70

Teoria decyzji 165-167
– falek 13
– Kahnemana 185
– Neymana-Pearsona 186-188
– preferencji rozmytych 203-204
– Walda 188
test istotności 187
– statystyczny 186
testowanie hipotez statystycznych 186-188
transformacja liniowa 72-76
– –, postać diagonalna 73
transformacja liniowa, postać pełna 73
– –, ujęcie lokalne 194-195

Wariancja 28-29
– estymatora 36
wartość krytyczna 184
– modalna 26
– oczekiwana 28-29
– próby losowej 34
– warunkująca 146-147
warunek początkowy równania różniczkowego 44
warunkująca zmienna losowa 146-147
własność prawie wszędzie 45
współczynnik korelacji 30
współrzędne binarne 98-103
– nieskorelowane 30
– skorelowane 30
– zmiennej losowej, liniowo zależne 30
wybór postaci jądra 65-72
wykrywanie elementów nietypowych 185--186, 188-190
wyznaczenie wartości binarnego parametru wygładzania 100-101
– – parametru wygładzania 76-90

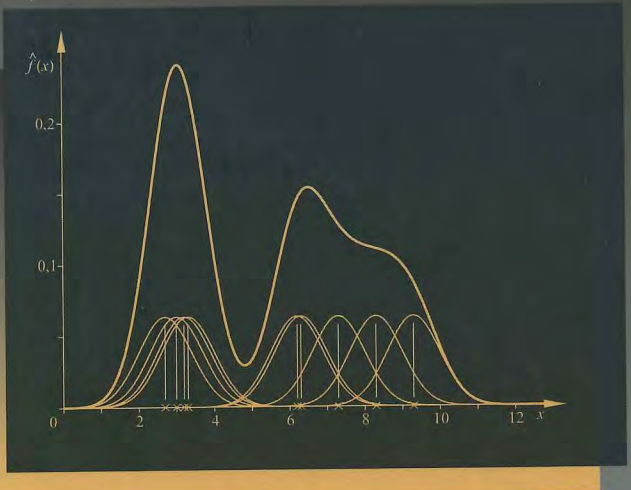
Zbiór borelowski 23
– krytyczny 186
– – lewostronny 186
– – prawostronny 186
– możliwych do podjęcia decyzji 166
– odrzucenia (krytyczny) 186
– przyjęcia 186
– spójny 26
– stanów natury 166
– zdarzeń elementarnych 22
zdarzenie 23
zmienna losowa 23
– – jednowymiarowa (rzeczywista) 23
– – *n*-wymiarowa 23
– – warunkująca 146
– – wielowymiarowa 23
zwarty podzbiór $\mathbb{R}$ 46

WYDAWNICTWA NAUKOWO-TECHNICZNE
ul. Mazowiecka 2/4, 00-048 Warszawa
tel. (0-22) 826 72 71 do 79
Dział Marketingu i Sprzedaży
tel. (0-22) 827 56 87, fax (0-22) 826 82 93
e-mail: marketing@wnt.pl

WNT. Warszawa 2005. Wyd. I
Ark. wyd. 14,0. Ark. druk. 14,0
Symbol Et/84091/IBS PAN
Poznańskie Zakłady Graficzne SA

W książce przedstawiono nieparametryczną metodę statystycznych estymatorów jądrowych, umożliwiającą wyznaczenie i obrazową ilustrację charakterystyk rozkładu zmiennej losowej, bez informacji o jego przynależności do ustalonej klasy rozkładów. Chociaż koncepcja związana z estymatorami jądrowymi jest naturalna i prosta, a interpretacja przejrzysta, to praktyczne zastosowania są niemożliwe bez użycia odpowiedniej techniki komputerowej, co jeszcze niedawno istotnie utrudniało badania teoretyczne, a zwłaszcza aplikacyjne.

Jak dotychczas brak jest na polskim rynku pozycji poświęconej tej tematyce. Niniejsza książka wypełnia tę lukę. Jest przeznaczona dla osób stosujących metody statystyczne oraz analizę danych do rozwiązywania różnych problemów współczesnej nauki i praktyki. Znajdą w niej przede wszystkim kompletne procedury umożliwiające uzyskanie ostatecznego wyniku bez konieczności wnikania w teoretyczne aspekty zagadnienia. Prezentowany materiał jest zilustrowany zadaniami aplikacyjnymi z zakresu szeroko rozumianej analizy systemowej, a także licznymi przykładami obliczeniowymi. Ułatwia to zrozumienie i użycie przedstawionej metody również przez osoby bez ścisłego przygotowania matematycznego, traktujące statystykę i analizę danych jako narzędzia do swych badań z zakresu inżynierii, ekonometrii, socjologii, biomedycyny i innych pokrewnych dziedzin.



Cena 39,- zł
ISBN 83-204-3147-6
9 788320 431476