POLITECHNIKA LUBELSKA

MONOGRAFIE

Matematyczne miscellanea
Tom 2

redakcja
Izolda Gorgol

Lublin 2018

Matematyczne miscellanea

Tom 2

Monografie – Politechnika Lubelska

Politechnika Lubelska
Wydział Podstaw Techniki
ul. Nadbystrzycka 38
20-618 Lublin

Matematyczne miscellanea

Tom 2

redakcja
Izolda Gorgol

Politechnika Lubelska
Lublin 2018

Recenzenci:
prof. dr hab. Witold Rzymowski (rozdziały 1,3,4,5)
dr hab. Józef Waniurski, prof. Politechniki Lubelskiej (rozdział 2)

Redakcja naukowa i techniczna: Izolda Gorgol

Publikacja wydana za zgodą Rektora Politechniki Lubelskiej

ISBN: 978-83-7947-311-3

Wydawca: Politechnika Lubelska
ul. Nadbystrzycka 38D, 20-618 Lublin
Realizacja: Biblioteka Politechniki Lubelskiej
Ośrodek ds. Wydawnictw i Biblioteki Cyfrowej
ul. Nadbystrzycka 36A, 20-618 Lublin
tel. (81) 538-46-59, email: wydawca@pollub.pl
www.biblioteka.pollub.pl
Druk: TOP Agencja Reklamowa Agnieszka Łuczak
www.agencjatop.pl

Elektroniczna wersja książki dostępna w Bibliotece Cyfrowej PL www.bc.pollub.pl
Nakład: 100 egz.

Spis treści

Przedmowa

Niniejsza publikacja powstała w ramach obchodów dziesięciolecia powstania Wydziału Podstaw Techniki. Jubileusz ten zbiegł się z dziesiątą rocznicą utworzenia kierunku *matematyka* w Politechnice Lubelskiej, który jest prowadzony w tym właśnie Wydziale. Kształcenie na kierunku *matematyka* na uczelni technicznej to szczególne wyzwanie. Nie tylko trzeba zachować standardy tego kierunku, ale należy również uwzględnić bardzo ważną pozycję matematyki w nowoczesnym kształceniu inżyniera. Minione dziesięć lat to okres wytężonej pracy nad przygotowaniem kierunku, ciągłą modyfikacją programów w celu dostosowania ich do wymogów formalnych, ale i wymogów stawianych przez rozwijającą się wiedzę i gospodarkę. To czas nawiązywania nowych kontaktów naukowych, budowy nowych profili zespołów naukowych, uwzględniających potrzeby rozwoju dydaktycznego kierunku. To również okres wystarczająco długi, by podjąć próbę oceny tych dokonań.

Niniejsza monografia jest cząstkową próbą prezentacji zarówno osiągnięć pracowników Katedry Matematyki Stosowanej, jak i pokazania efektów prac innych pracowników naukowych związanych bezpośrednio lub pośrednio z kierunkiem *matematyka*, tak z Politechniki Lubelskiej, jak i z innych ośrodków naukowych.

W pierwszym rozdziale dokonano przeglądu znanych rezultatów wynikających z badania izooptyk, tzn. krzywych o tej własności, że z każdego ich punktu widać zadane krzywe pod tym samym kątem.

Drugi rozdział prezentuje zagadnienia dotyczące zbiorów Koebego, które są związane ze zbiorami wartości przyjmowanych przez funkcje analityczne. Wyznaczono zbiory i funkcje ekstremalne, dzięki którym znaleziono zbiory Koebego dla klas $\mathscr{X}^*$ oraz $\mathscr{Y}^*$ złożonych z funkcji gwiaździstych i wypukłych odpowiednio

w kierunku osi rzeczywistej i w kierunku osi urojonej. Przedstawiono również wyniki dla pewnych klas związanych z rodzinami $\mathscr{X}^*$ i $\mathscr{Y}^*$.

W rozdziale trzecim zawarto krótkie omówienie teorii jąder reprodukujących, przedstawiono pewne ich zastosowania oraz zaprezentowano przykłady przestrzeni Hilberta z jądrem reprodukującym.

W czwartym rozdziale podjęto temat rzetelności w badaniach naukowych. Omówiono podstawowe błędy, które można napotkać w różnego rodzaju opracowaniach wykorzystujących narzędzia statystyki matematycznej. Opisano źródła tych błędów oraz wskazano sposoby ich unikania.

Rozdział piąty traktuje o aktualnych tendencjach w obszarze badań nad prawami wielkich liczb z uwzględnieniem różnych struktur zależności ciągów zmiennych losowych. Rozważania te wynikają z zastosowań praw wielkich liczb w statystyce, teorii niezawodności czy analizie ryzyka.

To tylko niewielki wycinek zainteresowań naukowych osób związanych z kierunkiem *matematyka* prowadzonym w Politechnice Lubelskiej. Kolejne zostaną przedstawiaone w następnej monografii.

Redaktor

Izooptyki

Waldemar Cieślak[1]
Witold Mozgawa[2]

Streszczenie

Pojęcie izooptyki jest znane od ponad trzystu lat. Mimo to istnieje niewiele prac napisanych na ten temat do lat dziewięćdziesiątych XX wieku. W opracowaniu przedstawimy wyniki z teorii izooptyk otrzymane po roku 1990, których inspiracją były prace [2, 4].

Abstract

The notion of an isoptic curve has been known for over three hundred years. Nevertheless, there are not many papers written up to the nineties of the twentieth century. We will present the results from theory of isoptics obtained after 1990, which were inspired by the papers [2, 4].

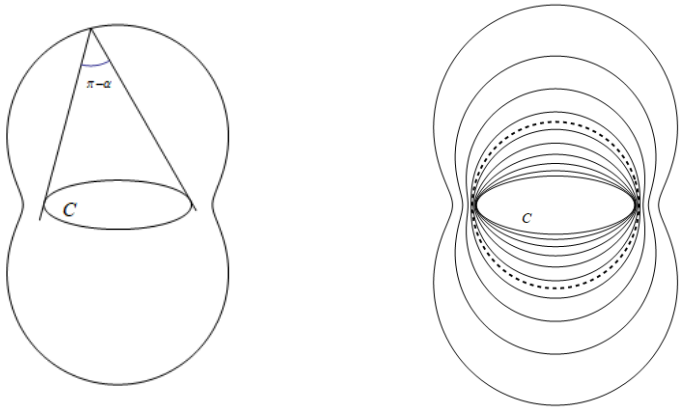
Słowa kluczowe: izooptyka, wzory całkowe, funkcja podparcia, owal, zamknięta krzywa ściśle wypukła, środek Steinera, sekantooptyka, ewolutoida, jeż, hiperpowierzchnia izooptyczna

[1] Katedra Matematyki Stosowanej; Politechnika Lubelska
[2] Instytut Matematyki; Uniwersytet Marii Curie-Skłodowskiej
e-mail: mozgawa@poczta.umcs.lublin.pl

1 Wstęp

Przypomnijmy, że zamkniętą, regularną, prostą, dodatnio zorientowaną krzywą płaską C klasy C^2 o dodatniej krzywiźnie nazywamy owalem. Czasem w tym opracowaniu zamiast owalu będziemy rozważać zamkniętą krzywą ściśle wypukłą, ale wtedy ten fakt wyraźnie podkreślimy.

α-izooptyką owalu C nazywamy miejsce geometryczne punktów, z których owal C jest widziany pod ustalonym kątem $\pi - \alpha$. Termin „izooptyka" pochodzi od dwóch greckich słów „isos"=„równy" i „optikos"=„wzrok". Na poniższym rysunku z lewej strony przedstawiono α-izooptykę elipsy, zaś z prawej strony kilka α-izooptyk elipsy dla różnych kątów α.



Rysunek 1: Po lewej stronie α-izooptyka elipsy; po prawej stronie kilka izooptyk elipsy, linią przerywaną zaznaczona jest izooptyka dla kąta prostego

Izooptyki były studiowane już w 1704 r. przez P. de la Hire (cf. [23]) i w 1877 r. przez M. Chaslesa w [3]. W 1884 r. badał je Ch. Taylor w [59], zaś w 1919 r. H. Hilton i R. Colomb w pracy [27]. W 1947 roku H. W. Richmond w [53] przedstawił ówczesny stan wiedzy dotyczącej izooptyk.

Współcześnie J. W. Green w [25], S. Matsuura w [33] oraz W. Wunderlich w [60] oraz w [61] opisywali krzywe, dla których izooptyką jest okrąg lub elipsa. W. Wunderlich w pracy [62] podał zastosowania izooptyk w inżynierii, przy badaniu mechanizmów krzywkowych.

Bardzo interesująca jest hipoteza Klamkina podana w pracy [28] i jej rozwiązanie podane przez J. C. C. Nitschego w [49].

Praca [2] napisana przez K. Benko, W. Cieślaka, S. Góździa i W. Mozgawę zapoczątkowała długą serię prac o izooptykach, których tematyka jest daleka od wyczerpania i ma kontynuatorów, cf. [11–16, 18–22, 29, 30, 47, 50, 55–58]. Poniżej omówimy głównie prace związane z autorami tego opracowania.

2 Izooptyki i ich własności

W pracy [2] rozważono globalne własności izooptyk owali. Podano tam warunek konieczny i dostateczny na wypukłość izooptyk. Ponadto wprowadzono pojęcie kąta granicznego jako kresu górnego zbioru $(0, \beta)$ takiego, że dla każdego kąta $\alpha < \beta$ α-izooptyka jest wypukła. Wyznaczenie takiego kąta w terminach geometrii owalu jest nadal problemem otwartym i pewne informacje na ten temat są zawarte właśnie w [2] oraz w [35]. Jeśli $z(s)$, $s \in \langle 0, L\rangle$ jest naturalną parametryzacją owalu C, $T(s)$ oznacza wektor styczny do C w punkcie $z(s)$, zaś $k(s)$ krzywiznę krzywej w tym punkcie, to dla ustalonego $a \in (0, L)$ istnieje liczba $\alpha \in (0, 2\pi)$ taka, że $T(a) = e^{i\alpha}T(0)$. Rozważono równanie różniczkowe $\varphi' = \frac{k}{k\circ\varphi}$ z warunkiem początkowym $\varphi(0) = a$. Wtedy dla każdego s mamy $T \circ \varphi(s) = e^{i\alpha}T(s)$ i zbiór punktów przecięcia prostych stycznych do owalu w punktach $z(s)$ oraz $z(\varphi(s))$ jest α-izooptyką C_α, której parametryzację opisano w pracy. Jeśli $q(s) = z(s) - z(\varphi(s))$, to wykazano, że dla każdego s zachodzi wzór

$$\frac{|z_\alpha(s) - z(s)|}{\sin\alpha_1(s)} = \frac{|z_\alpha(s)|}{\sin\alpha_2(s)} = \frac{|q(s)|}{\sin\alpha},$$

który nazywamy twierdzeniem sinusów i który w dalszych pracach odgrywa rolę jednego z narzędzi teorii izooptyk. Pracę kończą proste, wykonane przy pomocy

techniki izooptyk, dowody dwóch, znanych w literaturze, wzorów całkowych typu Croftona dla owali, mianowicie

$$\iint\limits_{\mathrm{Ext}\,C} \frac{\sin\omega}{t_1 t_2} dxdy = 2\pi^2, \qquad \iint\limits_{\mathrm{Ext}\,C} \frac{\sin\omega}{t_1 t_2}(R_1+R_2)dxdy = 2\pi L,$$

gdzie t_1 i t_2 są długościami odcinków prostych stycznych do C od punktu na izooptyce do ich punktów styczności z C, ω jest kątem między tymi odcinkami, L obwodem owalu, $\mathrm{Ext}\,C$ zewnętrzem obszaru ograniczonego przez C, a R promieniem krzywizny w odpowiednim punkcie.

Metodę tę zastosowano w kolejnych pracach do wyznaczenia nowych i nietrywialnych wzorów całkowych typu Croftona.

W pracy [4] zauważono, że w badaniu izooptyk płaskiej, zamkniętej krzywej ściśle wypukłej C bardzo wygodna jest jej parametryzacja $z(t) = p(t)e^{it} + p'(t)ie^{it}$ za pomocą narzędzia pochodzącego z analizy wypukłej, czyli funkcji podparcia $p(t)$, [54]. Funkcja podparcia jest odległością prostej podpierającej do krzywej od pewnego punktu wewnętrznego O tej krzywej, który przyjmujemy za początek układu współrzędnych. Wtedy α-izooptyką C_α krzywej C jest zbiór punktów przecięcia dwóch prostych podpierających w punktach $z(t)$ oraz $z(t+\pi)$, które się przecinają pod kątem $\pi - \alpha$. Wykorzystując parametryzację krzywej C i parametryzację izooptyki C_α

$$z_\alpha(t) = p(t)e^{it} + \left\{-p(t)\operatorname{ctg}\alpha + \frac{1}{\sin\alpha}p(t+\alpha)\right\} ie^{it}, \quad t \in [0, 2\pi],$$

dowiedziono wyników z pracy [2] przy osłabionych założeniach. Wykazano, że

$$\frac{|q|}{\sin\alpha} = \frac{|z_\alpha(t) - z(t)|}{\sin\alpha_1} = \frac{|z_\alpha(t) - z(t+\alpha)|}{\sin\alpha_2},$$

gdzie $q(t) = z(t) - z(t+\alpha)$, zaś α_1 jest kątem pomiędzy prostą podparcia krzywej w punkcie $z(t)$ a styczną do izooptyki z punkcie $z_\alpha(t)$, zaś α_2 jest kątem pomiędzy prostą podparcia krzywej w punkcie $z(t+\alpha)$ a styczną do izooptyki z punkcie $z_\alpha(t)$. Ten wzór jest wyżej wspomnianym twierdzeniem sinusów, którego uogólnienia podano w wielu różnych późniejszych pracach. Pokazano, że każda izooptyka jest zbiorem gwiaździstym, zaś przy wykorzystaniu wektora

$q(t) = z(t) - z(t+\alpha)$, że α-izooptyka jest krzywą wypukłą wtedy i tylko wtedy, gdy $\det(q, q') \leqslant 2|q|^2$ dla każdego parametru t. Za pomocą techniki izooptyk podano warunki, przy których dana z góry krzywa zamknięta jest izooptyką pewnego owalu oraz wykazano dwa nowe wzory całkowe typu Croftona

$$\iint\limits_{\mathrm{Ext}\, C} \frac{\sin^2\omega}{t_1} = \pi L, \qquad \iint\limits_{\mathrm{Ext}\, C} \frac{\sin^2\omega}{t_2} = \pi L,$$

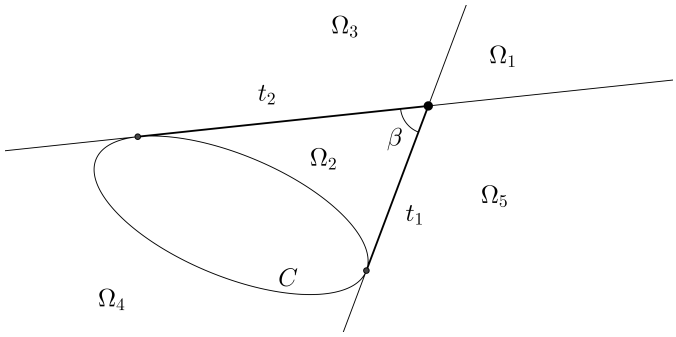
gdzie t_1 i t_2 są długościami odcinków prostych podparcia od punktu na izooptyce aż do punktów podparcia krzywej ściśle wypukłej C, ω jest kątem między tymi odcinkami, L obwodem owalu, a $\mathrm{Ext}\, C$ zewnętrzem krzywej C.

Warto podkreślić, że w pracy magisterskiej [26] dla płaskiej, zamkniętej krzywej ściśle wypukłej C udowodniono następujące wzory całkowe E. Czubera ([17]) oraz M. W. Croftona ([9, 10]), za pomocą metod z pracy [4], istotnie różnych od metody Czubera:

$$\iint \frac{\sin\omega}{t_1 t_2} dxdy = \frac{1}{2}\beta^2, \qquad \iint\limits_{\Omega_2} \frac{\sin\omega}{t_1 t_2} dxdy = \frac{1}{2}(\pi - \beta)^2,$$

$$\iint\limits_{\Omega_3} \frac{\sin\omega}{t_1 t_2} dxdy = \iint\limits_{\Omega_5} \frac{\sin\omega}{t_1 t_2} dxdy = \frac{1}{2}(\pi^2 - \beta^2), \qquad \iint\limits_{\Omega_4} \frac{\sin\omega}{t_1 t_2} dxdy = \pi\beta + \frac{\pi^2}{2}.$$

Sens geometryczny oznaczeń w powyższych wzorach wyjaśnia rysunek 2.



Rysunek 2: Ilustracja oznaczeń we wzorach Czubera

Ponieważ zbiory $\Omega_1 \ldots, \Omega_5$ pokrywają całą płaszczyznę, więc dodając je stronami otrzymujemy natychmiast wzór Croftona (2.11) z monografii [54].

W pracy [5], która jest dalszym ciągiem pracy [4], kontynuowano badania izooptyk. Zbadano w niej własności pierścieni utworzonych przez płaską, zamkniętą, prostą i ściśle wypukłą krzywą C o obwodzie L i jej izooptykę C_α. Dla tego pierścienia CC_α wykazano wzór typu Croftona

$$\iint\limits_{CC_\alpha} \frac{dxdy}{t_1(x,y)} = L \operatorname{tg} \frac{\alpha}{2},$$

gdzie $t_1(x,y)$ oznacza odległość pomiędzy punktem (x,y) pierścienia CC_α a punktem podparcia krzywej C wyznaczonym przez pierwszą, ze względu na orientację krzywej C, prostą podparcia przechodzącą przez punkt (x,y). Dalej podano twierdzenie opisujące pochodną pola pierścienia jako funkcji kąta α i twierdzenie o stycznych do izooptyki, które mówi, że jeśli $z'_\alpha(t)$ jest wektorem stycznym do α-izooptyki w punkcie t, a $q(t) = z(t) - z(t+\alpha)$, to dla każdego parametru t

$$\measuredangle(z'_\alpha(t), z'_\alpha(t+\tau)) + \measuredangle(q(t), q(t+\tau)) = 2\tau,$$

gdzie τ jest dowolną liczbą z przedziału $(0, 2\pi)$. Jeśli C jest ściśle wypukłą krzywą o stałej szerokości d, to w parametryzacji rozważanej w pracy odległość pomiędzy punktami izooptyki $z_\alpha(t)$ i $z_\alpha(t+\pi)$ jest stała i równa $d/\cos\frac{\alpha}{2}$. Twierdzenie odwrotne jest prawdziwe, jeśli kąty α i π są liniowo niezależne nad ciałem liczb wymiernych. Pracę kończy prezentacja pewnych równań różniczkowych dla obiektów związanych z izooptykami.

Izooptyki owali można traktować jak pewną ewolucję krzywej, dlatego jest interesujące zbadanie własności takiej ewolucji. Jednym z nierozstrzygniętych pytań jest pytanie o wypukłość izooptyk przy zmianie kąta $\pi - \alpha$. W pracy [35] podano następujący warunek dostateczny na to, by izooptyka owalu C była wypukła.

Twierdzenie 2.1 ([35]). *Jeśli dla każdego punktu $p \in C$ i dla każdej cięciwy m wychodzącej z punktu p długość rzutu wektora krzywizny owalu C w punkcie p na m jest mniejsza niż $|q|$, to wszystkie izooptyki owalu C są wypukłe.*

Prace [39] i [24] poświęcone są teorii izooptyk rozet. W pracy [39] wprowadzono pojęcie zorientowanej funkcji podparcia, która jest właściwym narzędziem do badania geometrii rozet. Niech rozeta $C\colon z = z(s)$ o indeksie j będzie krzywą klasy C^2, sparametryzowaną naturalnie. Ustalamy punkt $z(s_0)$ tak, by prosta styczna w tym punkcie była prostopadła do osi Ox. Dla dowolnego punktu $z(s)$ określamy wektor $e^{it} = \cos t + i \sin t$ jako jednostkowy zewnętrzny wektor normalny do krzywej C w tym punkcie, gdzie $t \in \langle 0, 2\pi j\rangle$ jest zorientowanym kątem pomiędzy dodatnim kierunkiem osi Ox i wektorem e^{it}. Zorientowaną odległość $p(t)$ początku układu współrzędnych od prostej stycznej do krzywej C w punkcie $z(s)$ określamy następująco: jeśli wektor e^{it} leży w półpłaszczyźnie, która zawiera początek układu, to przypisujemy odległości pomiędzy początkiem układu współrzędnych a prostą styczną w punkcie $z(s)$ znak minus, w przeciwnym przypadku $p(t)$ jest zwykłą odległością pomiędzy początkiem układu współrzędnych a prostą styczną w punkcie. Niech $z(t')$, $t' > t$ będzie najbliższym w sensie parametryzacji punktem takim, że proste styczne w punktach $z(t)$ i $z(t')$ tworzą kąt $\pi - \alpha$. Miejsce geometryczne przecięć powyżej określonych stycznych nazywamy α-izooptyką rozety C. W pracy wykazano, że izooptyka rozety ma taki sam indeks jak rozeta, i choć liczby samoprzecięć obu krzywych mogą się różnić, to ich liczby Whitney'a są identyczne. Następnie wykazano twierdzenie o stycznych do izooptyki, które mówi, że dla każdego ciągu liczb dodatnich $\tau_1, \ldots, \tau_n$ takich, że $\tau_1 + \cdots + \tau_n < 2\pi j$, jeśli przyjmiemy $t_1 = t, t_2 = t + \tau_1, \ldots, t_{n+1} = t + \tau_1 + \cdots + \tau_n$, to dla każdego parametru t mamy

$$\sum_{i=1}^{n} \measuredangle(z'_\alpha(t_i), z'_\alpha(t_{i+1})) + \sum_{i=1}^{n} \measuredangle(q(t-i), q(t_{i+1})) = 2(\tau_1 + \ldots + \tau_n).$$

W pracy [24] uogólniono pojęcie izooptyk dla rozet w następujący sposób: jeśli C jest rozetą o indeksie j, to istnieje $2j$ wartości parametru t takich, że $0 \leqslant t_0 < t_1 < \ldots < t_{2j-1} \leqslant 2\pi j$ i takich, że proste styczne do rozety C w punktach $z(t_0)$, $z(t_1)$, …, $z(t_{2j-1})$ są równoległe do kierunku k tworzącego dodatnio zorientowany kąt α ze styczną w punkcie $z(t)$. Zbiór punktów przecięcia prostych stycznych w punktach $z(t)$ i $z(t_k)$, $k = 1, \ldots, 2j-1$, nazywamy izooptyką typu (k, α) i ozna-

czamy $C_{k,\alpha}$. Z wcześniejszych prac o izooptykach wiadomo, że przez każdy punkt zewnętrza owalu izooptyki przechodzi dokładnie jedna izooptyka, zaś w przypadku izooptyk rozet sytuacja jest zupełnie inna. Izooptyki rozety, nawet tego samego typu, mogą się wzajemnie przecinać, a nawet pokrywać na pewnych zbiorach. Jednakże, jeśli $n \neq k$ lub $\alpha \neq \beta$ dla $k, n \in \{1, \ldots, 2j-1\}$, to $C_{k,\alpha} \neq C_{n,\beta}$. Następnie wykazano, że każda izooptyka $C_{k,\alpha}$ ma taki sam indeks jak rozeta C oraz rozważono środki Steinera obu krzywych i dowiedziono twierdzenia sinusów dla rozet i twierdzenia odwrotnego do twierdzenia sinusów dla rozet.

Praca [36] zawiera rozważania o izooptykach płaskich, otwartych krzywych wypukłych bez asymptot klasy C^2. W pracy opisano konstrukcję izooptyk dla takiej krzywej C i wykazano, między innymi, że przez dowolny punkt zewnętrza krzywej przechodzi dokładnie jedna α-izooptyka, jednakże jej parametryzacja przez funkcję podparcia $p(t)$ jest określona dla $t \in (0, \pi - \alpha)$. Dla takich izooptyk dowiedziono odpowiedniej wersji twierdzenia sinusów oraz wzoru całkowego typu Croftona

$$\iint\limits_{\operatorname{Ext} C} \frac{\sin\omega}{t_1 t_2} dxdy = \frac{\pi^2}{2},$$

gdzie t_1 i t_2 są długościami odcinków prostych stycznych do krzywej od punktu na izooptyce aż do punktów styczności z krzywą C, ω jest kątem między tymi odcinkami, a $\operatorname{Ext} C$ zewnętrzem obszaru ograniczonego przez C. Warto podkreślić, że nietrywialne rozwinięcie tej tematyki oraz tematyki izooptyk rozet z poprzedniego akapitu są zawarte w pracach D. Szałkowskiego, [56–58], gdzie autor wprowadza i bada pojęcie izooptyki k-tego rzędu rozety otwartej oraz par rozet otwartych. Nowym zjawiskiem jest fakt, że izooptyki rozet otwartych oraz par rozet otwartych mogą mieć punkty osobliwe, których geometryczne własności autor także opisuje.

W pracy [37] wprowadzono i rozważono pojęcie izooptyki pierwszego i drugiego rodzaju dla zagnieżdżonych par ściśle wypukłych krzywych. Podano szereg nietrywialnych, nowych wzorów całkowych typu Cauchy'ego-Croftona. Niech C_1 i C_2 będą ściśle wypukłymi krzywymi, gdzie krzywa C_2 leży we wnętrzu krzywej C_1, zaś początek układu współrzędnych O leży wewnątrz C_2. Załóżmy, że krzywe są dane równaniami $z_i(t) = p_i(t)e^{it} + p_i'(t)ie^{it}$, dla $i = 1, 2$, gdzie p_1, p_2 są

odpowiednimi funkcjami podparcia krzywych C_1 i C_2. Rozważamy prostą podparcia k_1 do krzywej C_1 w punkcie $z_1(t)$ i „dalszą" prostą podparcia k_2' krzywej C_2 równoległą do k_1. Obracamy prostą k_2' w kierunku zgodnym z ruchem wskazówek zegara do takiego położenia k_2, że proste podparcia k_1 i k_2 tworzą kąt α, $\alpha \in (0, \pi)$. Krzywa $z_\alpha(t)$ utworzona przez punkty przecięcia prostych tych podpierających jest nazywana α-izooptyką pierwszego rodzaju i oznaczana przez C_α. Jeśli obrócimy prostą k_2' w kierunku przeciwnym do ruchu wskazówek zegara o kąt α, to w przecięciu otrzymamy punkty krzywej $\tilde{z}_\alpha(t)$, którą nazywamy α-izooptyką drugiego rodzaju. Warto podkreślić, że przy tej definicji istnieją dokładnie dwie izooptyki każdego rodzaju przechodzące przez każdy punkt zewnętrza krzywej C_1 i że izooptyki są krzywymi regularnymi. Dowiedziono odpowiednich wersji twierdzenia sinusów dla obu typów izooptyk. Jeśli $q(t) = z_1(t) - z_2(t+\alpha)$, to wykazano, że izooptyka pierwszego (drugiego) rodzaju jest wypukła wtedy i tylko wtedy, gdy $\left|\frac{d}{dt}\left(\frac{q(t)}{|q(t)|}\right)\right| < 2$ dla każdego t. Problem wypukłości izooptyk jest rozważony w następującym twierdzeniu.

Twierdzenie 2.2 ([37]). *Izooptyka C_α jest krzywą wypukłą wtedy i tylko wtedy, gdy dla każdego t podwojona długość wektora $q(t)$ jest większa niż długość rzutu na kierunek wektora $q(t)$ wektora krzywizny krzywej C_1 w punkcie t i algebraicznej długości rzutu wektora krzywizny krzywej C_2 w punkcie $t+\alpha$ na kierunek wektora $q(t)$.*

To twierdzenie pozwala sprawdzić lokalną wypukłość izooptyki, jeśli znamy jedynie punkt, w którym badamy izooptykę i nie potrzebujemy nawet znać jej równania. W przypadku krzywych zagnieżdżonych rozważanie wzorów całkowych typu Croftona jest utrudnione, gdyż istnieją po dwie izooptyki każdego rodzaju, które się przecinają. Niech $\omega(t)$ będzie kątem pomiędzy prostą podpierającą krzywą C_1 w punkcie $z_1(t)$ i odcinkiem $z_1(t)z_2(t+\alpha)$. Wtedy odwzorowanie $F(\alpha, t) = z_\alpha(t)$ jest dyfeomorfizmem zbioru $\{(\alpha, t) \in (0, \pi) \times (0, 2\pi) \colon \omega(t) < \alpha < \pi\}$ na zewnętrze krzywej C_1 minus pewien zbiór miary zero, co oznacza, że izooptyki sparametryzowane przez punkty tego zbioru są rozłączne. Wykorzystując ten dyfeomorfizm wykazano prawdziwość kilku wzorów całkowych typu Croftona. W tym celu dla

każdego punktu $(x,y) \in \operatorname{Ext} C_1$, gdzie $\operatorname{Ext} C_1$ jest zewnętrzem krzywej C_1 rozważono cztery proste podparcia do krzywych C_1 i C_2, których kolejne odcinki od (x,y) do punktów podparcia oznaczamy przez l_1, m_1, m_2, l_2, gdzie l_1, l_2 podpierają krzywą C_1, a m_1, m_2 krzywą C_2. Wtedy otrzymujemy dwa wzory postaci

$$\iint\limits_{\operatorname{Ext} C_1} \left(\frac{\sin(l_2, m_2)}{l_2 m_2} + \frac{\sin(l_2, m_1)}{l_1 m_2} \right) dxdy = 2\pi^2,$$

$$\iint\limits_{\operatorname{Ext} C_1} \left(\frac{\sin(l_1, m_1)}{l_1 m_1} + \frac{\sin(l_2, m_1)}{l_2 m_1} \right) dxdy = 2\pi^2,$$

które są bardziej ogólne niż wzór (2.10) z monografii [54], Jeśli $k_1, \hat{k}_1, k_2$ są krzywiznami krzywych C_1 i C_2 w punktach styczności $z_1, \hat{z}_1, z_2$ z odcinkami l_1, l_2 i m_1, a kąt α jest kątem pomiędzy l_1 a m_1, a γ kątem pomiędzy m_1 a l_2 oraz $\beta = \pi - \gamma$, to

$$\iint\limits_{\operatorname{Ext} C_1} \frac{\sin\alpha}{l_2 m_2} \left(\frac{1}{k_1} + \frac{1}{k_2} \right) + \frac{\sin\alpha}{l_1 m_2} \left(\frac{1}{\hat{k}_1} + \frac{1}{k_2} \right) \Bigg) dxdy = \pi(L_1 + L_2)$$

oraz

$$\iint\limits_{\operatorname{Ext} C_1} \frac{\sin\alpha}{l_2 m_2} \frac{1}{k_1} \frac{1}{k_2} + \frac{\sin\gamma}{l_1 m_2} \frac{1}{k_1} \frac{1}{\hat{k}_2} \Bigg) dxdy = \frac{1}{2} L_1 L_2,$$

gdzie L_1 i L_2 są obwodami rozważanych krzywych. Powyższe wzory redukują się do znanych wzorów, gdy $C_1 = C_2$. Ponadto w pracy [37] dowiedziono twierdzenia całkowego dla pierścienia. Ponieważ izooptyki rozważane w tej pracy mogą się przecinać, to należy ograniczyć rozważania do odpowiednich kątów. Niech $\omega_M = \max_{t \in <0,2\pi>} \omega(t)$ i $\omega_m = \min_{t \in <0,2\pi>} \omega(t)$. Wtedy dla $\beta_2 > \beta_1 > \omega_M$ (lub dla $\omega_m > \beta_2 > \beta_1$) izooptyki C_{β_2} i C_{β_1} nie przecinają się. Dla ustalonych kątów β_1 i β_2, takich że $\beta_2 > \beta_1 > \omega_M$ rozważono pierścień $C_{\beta_1} C_{\beta_2}$ i dowiedziono wzoru

$$\iint\limits_{C_{\beta_1} C_{\beta_2}} \frac{1}{l_1} dxdy = L_1(\operatorname{ctg}\beta_1 - \operatorname{ctg}\beta_2) + L_2 \left(\frac{1}{\sin\beta_2} - \frac{1}{\sin\beta_1} \right).$$

Podobnie wykazano, że

$$\iint\limits_{C_{\beta_1}C_{\beta_2}} \frac{1}{m_2}dxdy = L_1\left(\frac{1}{\sin\beta_2} - \frac{1}{\sin\beta_1}\right) - L_2(\operatorname{ctg}\beta_2 - \operatorname{ctg}\beta_1).$$

Dodając oba wzory otrzymano

$$\iint\limits_{C_{\beta_1}C_{\beta_2}} \left(\frac{1}{l_1} + \frac{1}{m_2}\right)dxdy = (L_1 + L_2)\left(\operatorname{tg}\frac{\beta_2}{2} - \operatorname{tg}\frac{\beta_1}{2}\right).$$

Kilka kolejnych wzorów typu Cauchy'ego-Croftona podano w pracy [43], w której wykorzystano pewną specjalną parametryzację izooptyk owalu C. Dowiedziono następującego wzoru typu Croftona

$$\iint\limits_{\operatorname{Ext}C} \frac{\sin^2\omega}{h^2}\left(\frac{R_1}{t_1} + \frac{R_2}{t_2}\right)dxdy = 2\pi^2,$$

gdzie t_1 i t_2 są długościami odcinków prostych stycznych do krzywej od punktu (x, y) aż do punktów styczności z krzywą C, ω jest kątem między tymi odcinkami, R_1 i R_2 promieniami krzywizny w punktach styczności odcinków t_1 i t_2 z krzywą C, h jest odległością tych punktów styczności, a $\operatorname{Ext}C$ zewnętrzem krzywej C. Podano też inną postać tego wzoru, a następnie rozważono pewną powierzchnię $\widetilde{C}$ stowarzyszoną z owalem $C\colon z = z(t)$ sparametryzowanym za pomocą funkcji podparcia. Jeśli tak jak wcześniej $q(t,\alpha) = z(t) - z(t+\alpha)$, to powierzchnia jest dana przez jedną parametryzację $r(t,\alpha) = (q(t,\alpha),\alpha)$, dla $t \in (0, 2\pi)$, $\alpha \in (0,\pi)$. Przez $A(\widetilde{C})$ oznaczamy pole tej powierzchni, a objętość wyznaczoną przez tę powierzchnię i zawartą pomiędzy płaszczyznami $z = 0$ i $z = \pi$ przez $V(\widetilde{C})$. Wtedy przy powyższych oznaczeniach

$$\iint\limits_{\operatorname{Ext}C} \sin^2\omega\left(\frac{R_1}{t_1} + \frac{R_2}{t_2}\right)dxdy = 2V(\widetilde{C}),$$

oraz

$$\iint\limits_{\operatorname{Ext}C} \frac{\sin\omega}{t_1t_2}\sqrt{R_2^2R_1^2\sin^2\omega + R_1^2 + R_2^2 - 2R_1R_2\cos\omega}dxdy = A(\widetilde{C}).$$

Praca [44] prezentuje kolejny wzór całkowy typu Cauchy'ego-Croftona. W tej pracy rozważono izooptyki wewnętrzne wprowadzone w pracy [41]. Zakładamy, że owal C jest sparametryzowany za pomocą funkcji podparcia względem punktu Steinera O owalu oraz że układ współrzędnych ma początek w O i jest tak ułożony, że prosta styczna w punkcie $z(0)$ jest prostopadła do osi Ox. Wtedy $P(t) = -\frac{\det(z(t),q(t))}{|q(t)|}$ jest odległością od punktu O do prostej $d(t)$ przechodzącej przez punkty $z(t)$ i $z(t+\alpha), \alpha \in (0,\pi)$.

Obwiednię prostych $d(t)$ nazywamy izooptyką wewnętrzną a jej parametryzacja $C_\alpha\colon\ z = z^\alpha_{iso}(t)$ otrzymana przy pomocy funkcji $P(t)$ i wyjściowego parametru t jest zbyt skomplikowana, by ją tu przytaczać. Kąt

$$\alpha_0 = \sup\left\{\alpha\colon\ z^\alpha_{\mathrm{iso}} \text{ jest wypukła i } (0,\alpha) \text{ jest odcinkiem}\right\},$$

jest nazywany kątem granicznym dla izooptyki wewnętrznej. Dla tak określonego kąta rozważono pierścień CC_{α_0} ograniczony wyjściowym owalem i „graniczną" izooptyką wewnętrzną C_{α_0}. Dla punktu (x,y) leżącego wewnątrz tego pierścienia istnieje więc dokładnie jedna izooptyka wewnętrzna przechodząca przez ten punkt. Przyjmijmy, że odcinek q leżący wewnątrz owalu C styczny do izooptyki wewnętrznej w tym punkcie jest podzielony punktem styczności na dwa odcinki o długościach t_1 i t_2, że styczne do owalu C w końcach tego odcinka tworzą z q kąty α_1 i α_2 oraz że krzywizny w końcach tego odcinka są równe κ_1 i κ_2. Jeśli przez κ_3 oznaczymy krzywiznę izooptyki wewnętrznej w punkcie (x,y), to mamy następujący wzór całkowy

$$\iint\limits_{CC_{\alpha_0}} \frac{\kappa_1\cdot\kappa_2\cdot\kappa_3}{\sin\alpha_1\cdot\sin\alpha_2}\cdot(t_1+t_2)\cdot dxdy = 2\pi\cdot\alpha_0.$$

Praca [48] podaje jawne wzory na izooptyki trójkąta i prezentuje szereg ich własności. Praca ta ma ścisły związek ze wspomnianymi wyżej pracami [16,29,30,50], o izooptykach krzywych. Niech $\alpha \in (0,\pi)$ będzie ustalonym kątem i niech $z_k,\ k = 1,2,3,$ oznaczają wierzchołki dodatnio zorientowanego trójkąta T na płaszczyźnie $\mathbb{C}$. Dla $k = 1,2,3$ wykorzystano następujące oznaczenia - jeśli $\overrightarrow{a_k} = \overrightarrow{z_{k+1}z_{k+2}}$

jest zorientowanym bokiem trójkąta T, to a_k oznacza jego długość a β_k jest kątem trójkąta T odpowiadającym z_k i leżącym naprzeciwko boku $\overrightarrow{a_k}$. Zakładamy, że $\beta_1 \geqslant \beta_2 \geqslant \beta_3$ lub równoważnie $a_1 \geqslant a_2 \geqslant a_3$. Z elementarnej geometrii wiadomo, że α-izooptyka trójkąta T jest sumą co najmniej 3 i co najwyżej 6 łuków okręgów, jeden łuk na każdy bok trójkąta T, a jeśli $\alpha > \pi - \beta_k$, to mamy dodatkowy łuk nad wierzchołkiem z_k. Pokazano, że α-izooptyka trójkąta jest krzywą wypukłą, jeśli $\alpha \leqslant (\pi - \max\{\beta_1, \beta_2, \beta_3\})/2$. Następnie rozważono własności funkcji długości α-izooptyki jako funkcji kąta α i dowiedziono twierdzenia, że ta funkcja jest ciągła, ściśle rosnąca, ale nie jest wypukła. Rozważono własności funkcji pola powierzchni α-izooptyki jako funkcji kąta α i wykazano, że ta funkcja jest różniczkowalna, ściśle rosnąca i wypukła.

Warto podkreślić, że takie twierdzenie nie jest prawdziwe dla dowolnej krzywej ściśle wypukłej i w pracy [47] M. Michalska pokazała przykłady owali, dla których funkcja pola powierzchni izooptyki nie jest wypukła i sformułowała pewien dostateczny warunek na to, by ta funkcja była wypukła.

W ostatnim rozdziale pracy, dla dowolnego owalu C sparametryzowanego za pomocą funkcji podparcia, rozważono dodatnio zorientowane trójkąty T utworzone przez $z(t)$, $z(t+\alpha)$ i dowolny punkt ξ z otwartego kąta

$$\measuredangle(\overrightarrow{z_\alpha(t)z(t+\alpha)}, \overrightarrow{z_\alpha(t)z(t)})$$

i fragment jego izooptyki $C_{\alpha,t}$, gdzie $t \in [0, 2\pi]$, który jest jednocześnie fragmentem okręgu

$$\left| z - \frac{(z(t+\alpha)+z(t)) + i\operatorname{ctg}\alpha(z(t+\alpha)-z(t))}{2} \right| = \frac{|z(t+\alpha)-z(t)|}{2\sin\alpha}.$$

Wykazano, że obwiednią rodziny

$$\mathcal{F}_\alpha = \{C_{\alpha,t}, t \in [0, 2\pi]\}$$

jest α-izooptyka owalu C. Rozważania w tej pracy mają związek z pracą [32] o klasycznej teorii światła w $\mathbb{R}^d$, $d \geqslant 2$.

Wśród omawianych tutaj prac wyróżnia się ogólna i abstrakcyjna praca [6], w której podano daleko idące uogólnienie izooptyk i szereg ich własności rozważając pewną rodzinę regularnych, płaskich, ściśle wypukłych krzywych i rodzinę specjalnego typu równań różniczkowych stowarzyszonych z ciągłymi, okresowymi funkcjami o wartościach rzeczywistych.

W pracy rozważono rodzinę $\mathcal{F}$ ciągłych funkcji $a\colon \mathbb{R} \to \mathbb{R}$ o wartościach dodatnich i o okresie 2π. Dla każdej funkcji $a \in \mathcal{F}$ określono funkcję rosnącą $A\colon \mathbb{R} \to \mathbb{R}$ wzorem

$$A(t) = \int_0^t a(s)\, ds.$$

Ustalono funkcję $a \in \mathcal{F}$ i stowarzyszone równanie różniczkowe

$$\varphi' = \frac{a}{a \circ \varphi}$$

z warunkiem początkowym $\varphi(0) = m$, gdzie $m > 0$. Wtedy funkcja

$$\varphi(t, m) = A^{-1}\left(A(t) + A(m)\right)$$

jest rozwiązaniem tego równania różniczkowego, które spełnia warunek początkowy

$$\varphi(0, m) = m.$$

W dalszym ciągu pracy ograniczono rozważania do rozwiązań $\varphi(\cdot, m)$ dla $m > 0$ takich, że

$$t < \varphi(t, m) < t + \pi.$$

Lewa strona nierówności jest zawsze spełniona, a prawa strona jest spełniona wtedy i tylko wtedy, gdy $m \in (0, M)$, gdzie $M = \min_{t \in [0, 2\pi]} A^{-1}(A(t+\pi) - A(t))$. Dla regularnej, ściśle wypukłej, zamkniętej krzywej płaskiej C sparametryzowanej przez funkcję podparcia $p(t)$ i ustalonego rozwiązania $\varphi(\cdot, m)$ powyższego równania określono zamkniętą krzywą $C(m)$ jako miejsce geometryczne przecięcia się prostych stycznych do C w punktach t i $\varphi(t, m)$, której równanie parametryczne

jest postaci

$$w(t,m) = p(t)e^{it} + \frac{p(\varphi(t,m)) - p(t)\cos(\varphi(t,m)-t)}{\sin(\varphi(t,m)-t)} ie^{it}.$$

Taką krzywą nazywano w pracy φ-optyką. Przyjmując $Q(t) = z(t) - z(\varphi(t))$ znaleziono bardzo ogólny wzór całkowy

$$\int_0^{2\pi} \frac{\det(Q,Q')}{|Q|^2} dt = 2\pi.$$

Dla $\alpha(t) = \varphi(t) - t$ znaleziono następujące uogólnienie wzoru Bianchiego-Auerbacha z pracy [1]

$$\frac{\alpha'(t)}{2} = \frac{[R(t) + (1+\alpha'(t))R(t+\alpha(t))]\sin\frac{\alpha(t)}{2}}{|Q(0)| + \int_0^t [R(s+\alpha(s))(1+\alpha'(s)) - R(s)]\cos\frac{\alpha(s)}{2}} - 1,$$

gdzie $R(t)$ jest promieniem krzywizny krzywej C w punkcie t. Następnie autorzy rozważyli krzywą C, której promień krzywizny $R = p + p''$ i jest dodatnią funkcją ciągłą oraz odwzorowanie $\varphi\colon \mathbb{R} \to \mathbb{R}$ takie, że

$$\int_t^{\varphi(t)} R(s)\, ds = \text{const},$$

gdzie stała jest mniejsza niż długość $\sigma_{\min}$ najkrótszego łuku łączącego punkty $z(t)$ i $z(t+\pi)$ krzywej C. Wtedy mamy

$$\varphi' = \frac{R}{R \circ \varphi},$$

czyli φ jest rozwiązaniem rozważanego równania różniczkowego, gdy $a = R$. W ten sposób otrzymano nietrywialną ilustrację powyższej teorii i krzywe wyznaczające powyższe φ nazwano w pracy łukowo optycznymi krzywymi. Ta rodzina krzywych jest blisko związana z problemem Ulama dotyczącym równowagi ciała pływającego i opisuje generujące krzywe jednorodnego cylindra pływającego horyzontalnie

w wodzie, tak by był on w równowadze zgodnie z prawem Archimedesa. Dla tej rodziny krzywych powyższy wzór Bianchiego-Auerbacha redukuje się do wzoru $\frac{\alpha'}{2} = \frac{2}{|Q(0)|} R \sin\frac{\alpha}{2} - 1$, użytego przez H. Auerbacha do przedstawienia rozwiązania problemu Ulama w [1]. Jeśli $\Gamma = |w - z|$, $\Delta = |w - z \circ \varphi|$, zaś $\mathcal{A}(t)$ oznacza pole fragmentu wnętrza krzywej C zawarte pomiędzy cięciwą łączącą punkty $z(t)$ i $z(\varphi(t))$ a zorientowanym łukiem krzywej od $z(t)$ do $z(\varphi(t))$, to autorzy wykazują następującą geometryczną charakteryzację tej rodziny.

Twierdzenie 2.3 ([6]). *W rodzinie krzywych łukowo optycznych następujące warunki są równoważne:*

$$\mathcal{A} \equiv \text{const},$$
$$|Q| \equiv \text{const},$$
$$\Gamma \equiv \Delta.$$

Inną rodziną, która określa φ-optyki są zwykłe izooptyki, o których wyżej była wielokrotnie mowa i wtedy $\varphi(t) = t + \alpha$, dla pewnego $\alpha \in (0, \pi)$. W ramach tej rodziny rozważono takie krzywe, dla których $\Gamma = \Delta$ i z uogólnionego równania Bianchiego-Auerbacha otrzymano pewien układ równań, który ma rozwiązanie wtedy i tylko wtedy, gdy $\operatorname{tg}\frac{\alpha}{2} - \frac{1}{n}\operatorname{tg}\frac{n\alpha}{2} = 0$. Pokazano, że dla $n > 3$ istnieje co najmniej jedno rozwiązanie $\alpha \in (0, \pi)$. Dla krzywej, której funkcja podparcia jest postaci $p(t) = \frac{c_0}{2} + c_1 \cos t + d_1 \sin t + c_k \cos kt + d_k \sin kt$, $k \geqslant 4$, gdzie $p(t) + \ddot{p}(t) > 0$, zaś α jest niezerowym rozwiązaniem powyższego równania dla $n = k$, zachodzi równość $\Gamma \equiv \Delta$. Niestety autorom nie udało się udowodnić, że są to jedyne krzywe o takiej własności, ale mimo tego, jest to interesujący wynik, gdyż E. Raphaël, J.-M. di Meglio, M. Berger, E. Calabi w pracy [52] badali warunki pozycji równowagi długich pryzmatycznych cząsteczek, których przekrojem jest zbiór wypukły na granicy faz ciecz-ciecz, przy założeniu braku grawitacji, i wykazali, że warunek $\Gamma = \Delta$ charakteryzuje takie położenie. Wyznaczona w pracy powyższa rodzina krzywych wypukłych zawiera nieskończenie wiele zbiorów wypukłych, dla których cząsteczki typu rozważanego przez powyższych autorów są w równowadze w każdym położeniu na granicy faz ciecz-ciecz. Tak więc, teoria

φ-optyk może mieć znaczenie praktyczne lub być wykorzystana w chemii teoretycznej. Na koniec pracy pokazano, że w przypadku izooptyk krzywych C takich, że $p+p'' > 0$ następujące warunki są równoważne:

$$\mathcal{P} \equiv \text{const},$$
$$Q \parallel w',$$
$$\Gamma \equiv \Delta,$$

gdzie $\mathcal{P}(t)$ jest polem zawartym pomiędzy zorientowanym łukiem krzywej od $z(t)$ do $z(\varphi(t))$ a stycznymi w punktach $z(t)$ i $z(t+\alpha)$.

3 Sekantooptyki

Naturalnym uogólnieniem izooptyk są sekantooptyki. Niech C będzie owalem sparametryzowanym przy użyciu funkcji podparcia, zaś $\beta \in [0,\pi)$, $\gamma \in [0,\pi-\beta)$ i $\alpha \in (\beta+\gamma,\pi)$ ustalonymi kątami. Przez $l_1(t)$ oznaczamy prostą styczną do owalu C w punkcie $z(t)$. Prostą $l_1(t)$ obracamy dookoła punktu $z(t)$ o kąt $-\beta$. Tak otrzymaną sieczną owalu C oznaczamy przez $s_1(t)$. Podobnie, niech $l_2(t) = l_1(t+\alpha-\beta-\gamma)$ oznacza prostą styczną do owalu C w punkcie $z(t+\alpha-\beta-\gamma)$. Przez $s_2(t)$ oznaczamy sieczną owalu C otrzymaną przez obrót prostej $l_2(t)$ wokół punktu styczności o kąt γ. Wówczas zakładamy, że $s_1(t)$ i $s_2(t)$ przecinają się tworząc ustalony wcześniej kąt α. Zbiór punktów $z_{\alpha,\beta,\gamma}(t)$ przecięcia się siecznych $s_1(t)$ i $s_2(t)$ owalu C dla $t \in [0,2\pi]$ tworzy krzywą, którą nazywamy sekantooptyką $C_{\alpha,\beta,\gamma}$ owalu C.

Sekantooptyki w pewnym stopniu zachowują własności izooptyk owali, ale okazuje się, że także mają silne związki z ewolutoidami danej krzywej, z izooptykami dla par krzywych oraz z jeżami zdefiniowanymi i badanymi przez R. Langevina, G. Levitta i H. Rosenberga w pracy [31] i badanych ostatnio przez Y. Martineza-Maure'a. Jeżem nazywamy krzywą Γ, którą można sparametryzować za pomocą równania

$$z(t) = \psi(t)e^{it} + \psi'(t)ie^{it},$$

gdzie $\psi(t) = h(\cos t, \sin t)$ oraz $h \in C^2(\mathbb{S}^1, \mathbb{R})$. Funkcja $\psi(t)$ jest nazywana funkcją podparcia jeża Γ.

W pracy [42] rozważono parę jeży

$$\Gamma_1 \colon \; z_1(t) = \psi_1(t)e^{it} + \psi_1'(t)ie^{it}$$

oraz

$$\Gamma_2 \colon \; z_2(t) = \psi_2(t)e^{it} + \psi_2'(t)ie^{it}$$

danych przez ich funkcje podparcia $\psi_1(t) = h_1(\cos t, \sin t)$ i $\psi_2(t) = h_2(\cos t, \sin t)$, względem dowolnie wybranego początku układu O. Warto podkreślić, że nie zakłada się tutaj, tak jak w pracy [37], że jeże są zagnieżdżone. Niech $l(t)$ będzie prostą styczną do Γ_1 w punkcie $z_1(t)$, a $m(t)$ będzie prostą styczną do Γ_2 w punkcie $z_2(t)$. Wtedy dla danego kąta $\alpha \in (0, \pi)$, proste $l(t)$ oraz $m(t+\alpha)$ tworzą kąt α. Miejsce geometryczne $z_\alpha^{\Gamma_1\Gamma_2}(t)$ przecięć prostych stycznych $l(t)$ i $m(t+\alpha)$ dla $t \in [0, 2\pi]$ tworzy krzywą, którą nazywamy α- izooptyką $C_\alpha^{\Gamma_1\Gamma_2}$ pary Γ_1 i Γ_2. Następnie przypomniano klasyczną definicję ewolutoidy, która mówi, że ewolutoidą o kącie α pewnej krzywej regularnej $F \colon f = f(s)$ sparametryzowanej parametrem s nazywamy obwiednię prostych tworzących ustalony kąt δ z prostą normalną w punkcie $f(s)$. Wykazano, że każda ewolutoida owalu jest jeżem i że każda sekantooptyka owalu jest izooptyką pary odpowiednio dobranych ewolutoid. Następnie wykorzystując wprowadzoną teorię sekantooptyk i ich własności podano kilka uogólnień wzorów całkowych wyprowadzonych wcześniej przy pomocy teorii izooptyk. Wybierając odcinki t_1 i t_2 jak przy konstrukcji sekantooptyki $C_\omega^{\Gamma_{-\beta}\Gamma_\gamma}$ otrzymano dla owalu C wzory całkowe

$$\iint\limits_{\operatorname{Ext} C} \frac{\sin\omega}{t_1 t_2} dxdy = 2\pi^2 - 2\pi(\beta+\gamma),$$

$$\iint\limits_{\operatorname{Ext} C} \frac{\sin^2\omega}{t_1} dxdy = L_{\Gamma_{-\beta}}(\pi - (\beta+\gamma)) + L_{\Gamma_\gamma}\sin(\beta+\gamma),$$

$$\iint\limits_{\operatorname{Ext} C} \frac{\sin^2\omega}{t_2} dxdy = L_{\Gamma_\gamma}(\pi - (\beta+\gamma)) + L_{\Gamma_{-\beta}}\sin(\beta+\gamma),$$

gdzie $L_{\Gamma_{-\beta}}$ i L_{Γ_γ} są algebraicznymi długościami jeży $\Gamma_{-\beta}$ i Γ_γ. Na koniec pracy uogólniono do sekantooptyk wyżej omówione twierdzenia o wypukłości izooptyk.

Twierdzenie 3.1 ([42]). *Sekantooptyka $C_{\alpha,\beta,\gamma}$ jest wypukła wtedy i tylko wtedy, gdy suma rzutów wektorów krzywizny ewolutoidy $\Gamma_{-\beta}$ w punkcie $z^{-\beta}(t-\beta)$ i ewolutoidy Γ_γ w punkcie $z^\gamma(t+\alpha-\beta)$ na kierunek wektora Q jest mniejsza niż $2|Q(t)|$, gdzie $Q(t) = -M(t)ie^{t+\alpha-\beta} - L(t)ie^{i(t-\beta)}$, dla pewnych określonych w pracy funkcji $L(t)$ i $M(t)$.*

Twierdzenie 3.2 ([42]). *Jeśli suma długości rzutów wektorów krzywizny ewolutoidy $\Gamma_{-\beta}$ w punkcie $z^{-\beta}(t)$ iewolutoidy Γ_γ w punkcie $z^\gamma(t+\alpha)$ na kierunek wektora*

$$z^{-\beta}(t)z^\gamma(t+\alpha)$$

jest mniejsza niż $2|z^{-\beta}(t)z^\gamma(t+\alpha)|$ dla każdego $t \in [0,2\pi]$ i $\alpha \in (\beta+\gamma,\pi)$, to sekantooptyki $C_{\alpha,\beta,\gamma}$ owalu C są wypukłe dla każdego kąta $\alpha \in (\beta+\gamma,\pi)$.

Kontynuując badanie sekantooptyk M. Skrzypiec i W. Mozgawa podali w pracy [45] pewne własności sekantooptyk dla owali o stałej szerokości, twierdzenie o stycznych do sekantooptyki i twierdzenie odwrotne do twierdzenia sinusów dla sekantooptyk. Wykazano, że jeśli C jest owalem o stałej szerokości d, to w parametryzacji rozważanej w tej pracy odległość pomiędzy punktami $z_{\alpha,\beta,\gamma}(t)$ i $z_{\alpha,\beta,\gamma}(t+\pi)$ sekantooptyki $C_{\alpha,\beta,\gamma}$ jest stała i równa

$$\frac{d}{\sin\alpha}\sqrt{\cos^2\beta+\cos^2\gamma-2\cos\alpha\cos\beta\cos\gamma}.$$

Pewna wersja twierdzenia odwrotnego jest prawdziwa, gdyż pokazano, że jeśli kąty $\alpha-2\beta$ i π są liniowo niezależne nad ciałem liczb wymiernych oraz odległość pomiędzy punktami $z_{\alpha,\beta,\beta}(t)$ oraz $z_{\alpha,\beta,\beta}(t+\pi)$ na sekantooptyce $C_{\alpha,\beta,\beta}$ jest stała, to C jest owalem o stałej szerokości. Dalej dla ustalonego kąta $\tau \in (0,2\pi)$ wprowadzając oznaczenie $h^\tau(t)$ na funkcję $h(t+\tau)$ pokazano, że dla sekantooptyki $C_{\alpha,\beta,\gamma}$ owalu C zachodzi następująca relacja, zwana twierdzeniem o stycznych,

$$\measuredangle(z'_{\alpha,\beta,\gamma}, z'^{\tau}_{\alpha,\beta,\gamma}) + \measuredangle(Q,Q^\tau) = 2\tau,$$

gdzie wektor Q, którego precyzyjna definicja jest trudna do przytoczenia, pełni rolę wektora q w teorii izooptyk. Twierdzenie sinusów dla sekantooptyk udowodniła M. Skrzypiec w pracy [55]. W pracy [45] autorzy dowiedli następującego twierdzenia odwrotnego do twierdzenia sinusów.

Twierdzenie 3.3 ([45])**.** *Niech* $C\colon z(t) = p(t)e^{it} + \dot{p}(t)ie^{it}$ *będzie owalem, którego funkcja podparcia* p *względem początku układu współrzędnych jest klasy* C^3 *i niech* Γ *będzie regularną krzywą Jordana. Załóżmy, że:*

1. *dana jest różniczkowalna funkcja* $\varphi : \mathbb{R} \mapsto \mathbb{R}$*;*
2. *sieczną* $s_1(t)$ *otrzymujemy przez obrót prostej stycznej do owalu* C *przechodzącej przez punkt* $z(t)$ *dookoła punktu styczności o kąt* $-\beta$*, gdzie* $\beta \in [0, \pi)$*;*
3. *sieczną* $s_2(t)$ *otrzymujemy przez obrót prostej stycznej do owalu* C *przechodzącej przez punkt* $z(\varphi(t))$ *dookoła punktu styczności o kąt* γ*, gdzie* $\gamma \in [0, \pi - \beta)$*;*
4. *sieczne* $s_1(t)$ *i* $s_2(t)$ *przecinają się tworząc kąt* $\alpha(t) \in (\beta + \gamma, \pi)$ *dla każdego* $t \in [0, 2\pi]$ *w pewnym punkcie na krzywej* Γ *i że dla każdego* $t \in [0, 2\pi]$ *zachodzi następująca tożsamość zwana wzorem sinusów*

$$\frac{|z(t) - z_\Gamma(t)| + R(t)\sin\beta}{\sin\alpha_1(t)} = \frac{|z(\varphi(t)) - z_\Gamma(t)| + R(\varphi(t))\sin\gamma}{\sin\alpha_2(t)} = \\ = \frac{|(z(t) - R(t)\sin\beta\, ie^{t-\beta}) - (z(\varphi(t)) - R(\varphi(t))\sin\gamma\, ie^{\varphi(t)+\gamma})|}{\sin\alpha(t)},$$

   gdzie $\alpha_1(t)$ *i* $\alpha_2(t)$ *są kątami, które tworzy prosta styczna do krzywej* Γ*, odpowiednio, z siecznymi* $s_1(t)$ *i* $s_2(t)$*, a* R *jest promieniem krzywizny owalu* K*.*

Wtedy Γ *jest sekantooptyką* $C_{\alpha,\beta,\gamma}$ *owalu* C*.*

Praca [46] prezentuje pewien wzór całkowy typu Cauchy'ego-Croftona dla pierścienia $CC_{a,\beta,\gamma}$ ograniczonego owalem C i jego sekantooptyką $C_{a,\beta,\gamma}$ oraz rozważania dotyczące własności pola powierzchni tego pierścienia.

Niech $\beta \in [0,\pi)$ i $\gamma \in [0,\pi-\beta)$ oraz $a \in (\beta+\gamma,\pi)$. Niech $(x,y) \in CC_{a,\beta,\gamma}$ i niech $t_1(x,y)$ oznacza odległość pomiędzy punktem (x,y) a punktem $(x_1,y_1) \in C$ takim, że prosta otrzymana przez obrót prostej podparcia krzywej C w (x_1,y_1) o kąt $-\beta$ przechodzi przez punkt (x,y). Wtedy

$$\iint\limits_{CC_{a,\beta,\gamma}} \frac{dxdy}{t_1} = L_C\left(\frac{\cos\gamma-\cos\beta\cos a}{\sin a}-\sin\beta\right),$$

gdzie L_C oznacza obwód owalu C.

Jeśli $A_{\beta,\gamma}(\alpha)$ oznacza pole powierzchni ograniczone przez sekantooptykę jako funkcję zmiennej α, to w pracy [46] przedstawiono tę funkcję jako pewną całkę z parametrem i wykazano, że ta funkcja spełnia równanie różniczkowe

$$A'_{\beta,\gamma}(\alpha)\sin\alpha+2A_{\beta,\gamma}(\alpha)\cos\alpha = G(\alpha),$$

gdzie $\alpha \in (\beta+\gamma,\pi)$, zaś $G(\alpha)$ jest pewną funkcją wyznaczoną w pracy. Ponadto dla $\beta \neq 0$ lub $\gamma \neq 0$ oszacowano:

$$0 \leqslant A'_{\beta,\gamma}((\beta+\gamma)^+) \leqslant L_C \max_{t\in[0,2\pi]} R(t)\frac{\sin\beta\sin\gamma}{\sin(\beta+\gamma)}.$$

Pozostaje problemem otwartym zbadanie własności funkcji obwodu i pola powierzchni ograniczonej sekantooptyką jako funkcji kąta α tak, jak to zrobiono w pracy [48] dla izooptyk trójkątów.

4 Zastosowania izooptyk w pewnych problemach teorii krzywych płaskich

4.1 Zastosowanie izooptyk w teorii krzywych o stałej szerokości

Przedwcześnie zmarły matematyk A. P. Mellish w pracy [34], która miała duży wpływ na wiele prac z teorii krzywych płaskich, podał szereg warunków równoważnych temu, by krzywa była o stałej szerokości. W pracy [40], korzystając z teorii

izooptyk budowanej w powyżej opisanych pracach, podano daleko idące uogólnienie tego twierdzenia, które ma związek z geometrią przestrzeni Minkowskiego, geometrią rzutową i wyżej wspomnianą teorią jeży.

Tak jak wcześniej, niech $C\colon z(t) = p(t)e^{it} + p'(t)ie^{it}$ będzie owalem sparametryzowanym za pomocą funkcji podparcia $p(t)$. Jego α-izooptyka C_α ma wtedy parametryzację

$$z_\alpha(t) = p(t)e^{it} + \left(-p(t)\operatorname{ctg}\alpha + \frac{1}{\sin\alpha}p(t+\alpha)\right)ie^{it}.$$

Jeśli przyjmiemy, że $q(t,\alpha) = z(t) - z(t+\alpha)$, to jej krzywizna wyraża się wzorem

$$k_\alpha(t) = \frac{\sin\alpha}{|q(t,\alpha)|^3}\left(2|q(t,\alpha)|^2 - [q(t,\alpha), q'(t,\alpha)]\right),$$

gdzie $[\ ,\ ]$ oznacza wyznacznik od swoich argumentów. Z definicji izooptyka owalu C jest miejscem geometrycznym punktów, z których owal C jest widziany pod ustalonym kątem $\pi - \alpha$. Z drugiej strony owal o stałej szerokości a może być swobodnie obracany w pasie o szerokości a, co oznacza, że równoległe styczne do tego owalu przecinają się „w nieskończoności" dając „π-izooptykę". Niestety rzutowe pojęcie prostej w nieskończoności nie jest tu odpowiednim obiektem i dlatego wprowadzono tu inny obiekt, który jest związany z owalem C i jego α-izooptyką. Sinus-krzywizną owalu C nazywano funkcję

$$\kappa_\alpha(t) = \frac{2|q(t,\alpha)|^2 - [q(t,\alpha), q'(t,\alpha)]}{|q(t,\alpha)|^3}$$

i w ten sposób owal, dla którego $\kappa_\alpha \equiv const$ jest nazywany owalem o stałej α-szerokości. W ten sposób można rozważać π-izooptyki i stałą π-szerokość, którą nazywano stałą szerokością w nieskończoności i pokazano, że owal C jest krzywą o stałej szerokości w nieskończoności wtedy i tylko wtedy, gdy jest on owalem o stałej szerokości (klasycznej). I jako wniosek otrzymano, że dla $0 < \alpha \leqslant \pi$ owal jest o stałej α-szerokości wtedy i tylko wtedy, gdy jego α-izooptyka jest okręgiem. Jeśli owal C jest sparametryzowany za pomocą funkcji podparcia, to jego środek

Steinera wyznacza się wzorem

$$S = \frac{1}{2\pi} \int_0^{2\pi} z(t)dt.$$

W pracy przyjęto, że jest to początek układu współrzędnych. Jest problemem otwartym, czy środek Steinera izooptyki pokrywa się ze środkiem Steinera wyjściowego owalu, ale w pracy [40] udowodniono, że jeśli dla owalu C jego α-izooptyka dla pewnego kąta $\alpha \in (0, \pi)$ jest okręgiem, to jego środek jest środkiem Steinera owalu C. Z postaci wektora $q(t, \alpha)$ widać, że nie może on być wykorzystany do uogólnienia pojęcia szerokości. W dalszej części pracy wprowadzono więc funkcję $d_\alpha(t)$, która ma związek z szerokością owalu, a którą nazwano α-rozprzestrzenieniem owalu C w punkcie t. W tym celu określono wektor $Q(t, \alpha)$ pomiędzy rzutami początku układu współrzędnych na proste podparcia w punktach $z(t)$ oraz $z(t + \alpha)$ owalu C. Wtedy $d_\alpha(t) = |Q(t, \alpha)| = \sqrt{p(t)^2 + p^2(t + \alpha) - 2p(t)p(t + \alpha)\cos\alpha}$, gdzie $p(t)$ jest funkcją podparcia owalu. Dla $\alpha = \pi$ mamy, że $d_\pi(t) = p(t) + p(t + \pi)$ jest zwykłą szerokością owalu w punkcie t. Z owalem C stowarzyszono naturalnego jeża, czyli krzywą H_α określoną równaniem parametrycznym $H_\alpha(t) = P(t)e^{it} + P'(t)ie^{it}$, której funkcja podparcia $P(t)$ jest dana wzorem $P(t) = \frac{|Q(t,\alpha)|}{2}$. Wtedy dowiedziono głównego twierdzenia pracy.

Twierdzenie 4.1 ([40])**.** *Dla ustalonego kąta $\alpha \in (0, \pi)$ stwierdzenia:*

(i) *owal jest o stałej α-szerokości;*

(ii) *owal jest o stałym α-rozprzestrzenieniu;*

(iii) *każdy wektor $q(t, \alpha)$ jest równoległy do wektora $Q(t, \alpha)$;*

(iv) *jeśli α_1 i α_2 są kątami pomiędzy wektorem $q(t, \alpha)$ a prostymi podpierającymi owal w punktach $z(t)$ i $z(t + \alpha)$, to wyrażenie*

$$\frac{1}{|q(t,\alpha)|^2}\left(2|q(t,\alpha)| - \left(\frac{\sin\alpha_1}{k(t+\alpha)} + \frac{\sin\alpha_1}{k(t)}\right)\right) \tag{4.1}$$

jest stałe, gdzie k oznacza krzywiznę owalu;

są równoważne.

Ponadto, dla wszystkich krzywych o tej samej stałej α-szerokości b stowarzyszone jeże H_α mają tę samą długość L daną wzorem

$$L = \pi b.$$

Dla $\alpha = \pi$ wszystkie stwierdzenia tego twierdzenia redukują się do odpowiednich stwierdzeń pierwszego twierdzenia z pracy Mellisha [34].

4.2 Zastosowanie izooptyk w poryzmie Ponceleta

W tym rozdziale przedstawimy zastosowanie izooptyk do rozwiązania pewnego problemu związanego z poryzmem Ponceleta. W tym celu krótko opiszemy twierdzenie Ponceleta o zamknięciu [51].

Niech E_1 i E_2 będą dwiema takimi elipsami, że elipsa E_1 zawiera w swoim wnętrzu elipsę E_2 oraz niech P_1 będzie punktem na E_1, zaś s_1 prostą styczną do E_2 wyprowadzoną z punktu P_1. Wychodząc z pary (P_1, s_1) konstruujemy transwersalną Ponceleta, to jest ciąg $(P_1, s_1, P_2, s_2, P_3, s_3, \ldots)$ taki, że $P_i \in E_1$, zaś s_i jest prostą styczną do elipsy E_2 oraz $P_{i+1} = s_i \cap s_{i+1}$. Mówimy, że transwersalna Ponceleta zamyka się po n krokach, jeśli $P_{n+1} = P_1$.

W latach 1813-1814 Jean Victor Poncelet dowiódł twierdzenia zwanego twierdzeniem o zamknięciu, które zostało opublikowane po jego powrocie z niewoli w Rosji.

Twierdzenie 4.2 ([51]). *Jeśli E_1 i E_2 są dwiema takimi elipsami, że elipsa E_1 zawiera w swoim wnętrzu elipsę E_2 oraz transwersalna Ponceleta wychodząca z punktu $P_1 \in E_1$ zamyka się po n krokach, to transwersalna Ponceleta wychodząca z każdego innego punktu elipsy E_1 zamyka sie po n krokach.*

Twierdzenie Ponceleta często jest nazywane w literaturze poryzmem Ponceleta lub własnością poryzmu Ponceleta.

Tutaj będziemy tak nazywać własność, która polega na tym, że jeśli pewien fakt zachodzi w jednym punkcie, to również zachodzi we wszystkich punktach rozważanego zbioru.

W pracy [41] rozważono poryzm Ponceleta dla pary takich dwóch zagnieżdżonych owali C_1, C_2, że owal C_2 leży wewnątrz owalu C_1, dla których, analogicznie jak powyżej dla elips, rozważono odpowiednią transwersalną Ponceleta. Otrzymano w ten sposób tak zwany bilard barowy, lub bilard z grzybkiem, z powodu podobieństwa do pewnej formy bilardu, w którą grano kilkadziesiąt lat temu. W tym bilardzie gracze zdobywali punkty wbijając kule do otworów unikając potrącenia „grzybka" stojącego na środku stołu. Teraz rolę „grzybka" gra wewnętrzny owal, który w każdym ruchu musi być „potrącony" przez kolejną styczną. Dla danego owalu C rozważono pytanie, czy dla niego istnieją takie dwa owale, jeden owal zawarty w nim C_{in} i drugi owal zawierający go C_{out}, że dla każdej z par (C, C_{in}) i (C, C_{out}) zachodzi własność poryzmu Ponceleta. W tym celu rozważono owal C w parametryzacji danej przez jego funkcję podparcia wraz z jego izooptyką C_α i wykazano, że jeśli $\kappa(t)$ oznacza krzywiznę owalu C a $\kappa_\alpha(t)$ krzywiznę izooptyki C_α, to funkcja

$$k_\alpha(t) = \begin{cases} \kappa_\alpha(t), & 0 < \alpha < \pi \\ \kappa(t), & \alpha = 0 \end{cases}$$

jest prawostronnie ciągła względem α. Dlatego istnieje takie $\varepsilon > 0$, że $k_\alpha(t) > 0$, gdy $0 < \alpha < \varepsilon$ i $t \in \mathbb{R}$. Ponadto, jeśli $\alpha = \frac{2\pi}{n}$ jest takim kątem, że $0 < \alpha < \varepsilon$ dla powyższego ε, to dla bilardu barowego, dla którego wyjściowy owal C jest owalem C_2 a jego α-izooptyka jest owalem C_1 transwersalna Ponceleta zamyka się po n odbiciach każdego t i tworzy n-kąt z przecięciami lub bez. Tak więc, dla każdego owalu C istnieje taki zawierający go owal $C_{out} = C_1$, że odpowiadający tej parze bilard barowy ma własność poryzmu Ponceleta.

W drugiej części pracy rozważono problem istnienia owalu wewnątrz owalu C, dla którego zachodzi własność poryzmu Ponceleta. W tym przypadku założono, że owal C jest sparametryzowany swoją funkcją podparcia $z(t) = p(t)e^{it} + p'(t)ie^{it}$, a środek układu współrzędnych jest wybrany w środku Steinera owalu. Rozważono proste $l(t)$ przechodzące przez punkty $z(t)$ i $z(t+\alpha)$, dla każdego $t \in [0, 2\pi]$ i wyprowadzono równanie obwiedni rodziny prostych $\{l(t) \colon t \in \mathbb{R}\}$ przy ustalonym α. Dokonując odpowiedniej zmiany parametru zapisano obwiednię w terminach wyj-

ściowego parametru t. Otrzymano w ten sposób zamkniętą krzywą $z_{iso}(t)$, która może mieć ostrza i samoprzecięcia i którą nazywano izooptyką wewnętrzną. Pokazano, że

$$\lim_{\alpha\to 0} z_{izo}(t) = z(t),$$
$$\lim_{\alpha\to 0} z'_{izo}(t) = z'(t),$$
$$\lim_{\alpha\to 0} z''_{izo}(t) = z''(t),$$

dla każdego t oraz że funkcja

$$k_{izo}(t) = \begin{cases} \kappa_{izo}(t), & 0 < \alpha < \pi, \\ \kappa(t), & \alpha = 0, \end{cases}$$

gdzie $\kappa_{izo}(t)$ oznacza krzywiznę izooptyki wewnętrznej $z_{izo}(t)$, jest ciągła z prawej względem α. Wtedy istnieje taki $\varepsilon > 0$, że $k_{izo}(t) > 0$, gdy $0 < \alpha < \varepsilon$ i $t \in \mathbb{R}$. Ponadto, jeśli $\alpha = \frac{2\pi}{n}$ jest takim kątem, że $0 < \alpha < \varepsilon$ dla powyżej wybranego ε, to dla bilardu barowego utworzonego przez wyjściowy owal C i jego wewnętrzną izooptykę $z_{izo}(t)$ dla kąta α wziętą za owal C_2 transwersalna Ponceleta zamyka się po n odbiciach dla każdego t i tworzy n-kąt z samoprzecięciami lub bez. Stąd dla każdego owalu C istnieje taki owal zawarty w C, że wyznaczony przez tę parę bilard barowy ma własność poryzmu Ponceleta.

5 Pojęcie izooptyki w innych przestrzeniach

Analogiczne problemy dotyczące izooptyk jak w pracy [4] były badane przez G. Csimę i J. Szirmaia na płaszczyznach hiperbolicznej i eliptycznej w pracach [12,14,15]. W artykule *Isoptic surfaces of polyhedra* [16] podano algorytm i odpowiedni program komputerowy do wyznaczania izooptycznych powierzchni dla dowolnego wypukłego wielościanu w trójwymiarowej przestrzeni euklidesowej. Wyznaczono powierzchnie izooptyczne dla brył platońskich i pewnych półforemnych wielościanów Archimedesa oraz zilustrowane je graficznie dla pewnych kątów.

W pracy [13] podano naturalne przedłużenie pojęcia izooptyki dla n-wymiarowej przestrzeni euklidesowej, które nazwano izooptycznymi hiperpowierzchniami. Podano tam algorytm do wyznaczania izooptycznych hiperpowierzchni, w szczególności wyznaczono równanie izooptycznej hiperpowierzchni dla prostokąta. Ponadto podano zastosowania izooptycznych hiperpowierzchni, np. do projektowania stadionów, teatrów, kin, gdyż jest istotne by te miejsca były tak zbudowane, żeby z każdego punktu na widowni można było widzieć boisko, czy scenę pod tym samym kątem przestrzennym.

Th. Dana-Picard, G. Mann, N. Zehavi w pracach [18–20] przedstawili izooptyki krzywych stożkowych jako przekroje pewnych torusów z samoprzecięciami, co otwiera nowe perspektywy w badaniu izooptyk stożkowych.

Literatura

[1] H. Auerbach, Sur un problème de M. Ulam concernant lñéquilibre des corps flottant, *Stud. Math.*, 7, 1938, 121–142.

[2] K. Benko, W. Cieślak, S. Góźdź, W. Mozgawa, On isoptic curves, *An. Ştiinţ. Univ. "Al. I. Cuza"*, 36, 1990, 47–54.

[3] M. Chasles, Aperçu historique sur l'origine et le dévéloppement des méthodes en géométrie, Paris, 1875, Darboux Bull., 9, 1877, 97–98.

[4] W. Cieślak, A. Miernowski, W. Mozgawa, Isoptics of a strictly convex curve, Global differential geometry and global analysis, Proc. Conf., Berlin/Ger. 1990, Lect. Notes Math. 1481, 1991, 28–35.

[5] W. Cieślak, A. Miernowski, W. Mozgawa, Isoptics of a closed strictly convex curve II, *Rend. Sem. Mat. Univ. Padova*, 96, 1996, 37–49.

[6] W. Cieślak, A. Miernowski, W. Mozgawa, φ-optics and generalized BianchiŐAuerbach equation, *Journal of Geometry and Physics*, 62, 2012, 2337–2345.

[7] W. Cieślak, H. Martini, W. Mozgawa, On rotation index of bar billiards and Ponceletñs porism, *Bull. Belg. Math. Soc. Simon Stevin*, 20, 2, 2013, 287–300.

[8] W. Cieślak, W. Mozgawa, On rosettes and almost rosettes, *Geom. Dedicata*, 24, 1987, 221–228.

[9] M. W. Crofton, On the theory of local probability, *Phil. Trans. Roy. Soc. London*, 158, 1868, 181–199.

[10] M. W. Crofton, Probability, in Encyclopaedia Britannica, 9th ed., Vol. 19, 1885, 768–788.

[11] G. Csima, J. Szirmai, Isoptic curves of conic sections in constant curvature geometries, *Math. Commun.*, 19, 2, 2014, 277–290.

[12] G. Csima, J. Szirmai, Isoptic curves in the hyperbolic plane, *Stud. Univ. Žilina, Math. Ser*. 24, 1, 2010, 15–22.

[13] G. Csima, J. Szirmai, On the isoptic hypersurfaces in the nn-dimensional Euclidean space. *KoG*, 17, 2013, 53–57.

[14] G. Csima, J. Szirmai, Isoptic curves to parabolas in the hyperbolic plane, *Pollac Periodica*, 7/1/1, 2012, 55–64.

[15] G. Csima, J. Szirmai, Isoptic curves of generalized conic sections in the hyperbolic plane, wysłane do druku, 2015.

[16] G. Csima, J. Szirmai, Isoptic surfaces of polyhedra, *Computer Aided Geometric Design*, 47, 2016, 55–60.

[17] E. Czuber, *Geometrische Wahrscheinlichkeiten und Mittelwerte*, Leipzig, Teubner, 1884.

[18] Th. Dana-Picard, G. Mann, N. Zehavi, From conic intersections to toric intersections: the case of the isoptic curves of an ellipse, *The Montana Mathematical Enthusiast*, 9, 1, 2011, 59–76.

[19] Th. Dana-Picard, G. Mann, N. Zehavi, What can be learnt when changing the point of view on conic sections? (in Hebrew), *Aleh*, 38, 2007, 22–31.

[20] Th. Dana-Picard, N. Zehavi, G. Mann, Bisoptic curves of hyperbolas, *International Journal of Mathematical Education in Science and Technology*, 45, 5, 2014, 762-781.

[21] Th. Dana-Picard, A. Naiman, Isoptics of Fermat Curves, wysłane do *Mathematics in Computer Science*, 2017.

[22] Th. Dana-Picard, Automated study of isoptic curves of an astroid, wysłane do *Journal of Symbolic Computation*, 2018.

[23] R. Ferréol, Courbe isoptique, Mathcurve–Encyclopédie des formes mathématiques remarquables,

https://www.mathcurve.com/courbes2d/isoptic/isoptic.shtml+

[24] S. Góźdź, A. Miernowski, W. Mozgawa, Sine theorem for rosettes, *An. Ştiinţ. Univ. Al. I. Cuza, Iaşi*, XLIV, 1998, 385–394.

[25] J. W. Green, Sets subtending a constant angle on a circle, *Duke Math. J.*, 17, 1950, 263–267.

[26] D. Haraszczuk, Izooptyki i ich zastosowania, Politechnika Lubelska, WPT, Lublin, 2015.

[27] H. Hilton, R. E. Colomb, On orthoptic and isoptic loci, *American J.* 39, 1917, 86–94.

[28] M. S. Klamkin, Conjectured Isoptic Characterization of a Circle, *The American Mathematical Monthly* 95, 9, 1988, 845.

[29] R. Kunkli, I. Papp, M. Hoffmann, Isoptics of Bézier curves, *Comput. Aided Geom. Des.*, 30, No. 1, 2013, 78–84.

[30] A. Kurusa, T. Ódor, Isoptic characterization of spheres, *J. Geom.*, 106, No. 1, 2015, 63–73.

[31] R. Langevin, G. Levitt i H. Rosenberg, *Hérissons et multihérissons* (*Envellopes paramétrées par leur application de Gauss*), Singularities, Warsaw 1985, 245–253, Banach Center Publ. 20, 1988, PWN Warsaw.

[32] H. Martini, A contribution to the light field theory, *Beitr. Algebra Geom.*, 30, 1990, 193–201.

[33] S. Matsuura, On non-convex curves of constant angle, Komatsu, Hikosaburo (ed.), Functional analysis and related topics, 1991. Proceedings of the international conference in memory of Professor Kôsaku Yosida held at RIMS, Kyoto University, Japan, July 29-Aug. 2, 1991. Berlin: Springer-Verlag. Lect. Notes Math. 1540, 1993, 251–268.

[34] A. P. Mellish, Notes on differential geometry, *Ann. Math.*, 2, 32, 1931, 181–190.

[35] A. Miernowski, W. Mozgawa, On some geometric condition for convexity of isoptics, *Rend. Sem. Mat. Univ. Pol. Torino*, 55, 2, 1997, 93–98.

[36] A. Miernowski, W. Mozgawa, Isoptics of open, convex curves and Crofton-type formulas, *Istanbul Üniversitesi Fen Fakültesi Matematik Dergisi*, 57-58, 1998-1999, 33–39.

[37] A. Miernowski, W. Mozgawa, Isoptics of pairs of nested closed strictly convex curves and Crofton-type formulas, *Beitr. Algebra Geom.*, 42, 2001, 281–288.

[38] A. Miernowski, W. Mozgawa, On the curves of constant relative width, *Rend. Sem. Mat. Univ. Padova*, 107, 2002, 57–65.

[39] A. Miernowski, W. Mozgawa, Isoptics of rosettes and rosettes of constant width, *Note di Matematica*, 15, 1995, 203–213.

[40] W. Mozgawa, Mellish theorem for generalized constant width curves, *Aequationes Math.*, 89, 4, 2015, 1095–1105.

[41] W. Mozgawa, Bar billiards and Ponceletñs porism, *Rend. Semin. Mat. Univ. Padova* 120, 2008, 157–166.

[42] W. Mozgawa, M. Skrzypiec, Secantoptics as isoptics of pairs of evolutoids, *Bull. Belg. Math. Soc. Simon Stevin*, 16, 2009, 435Ő-445.

[43] W. Mozgawa, Integral formulas related to ovals, *Beitr. Algebra Geom.*, 50, No. 2, 2009, 555–561.

[44] H. Martini, W. Mozgawa An integral formula related to inner isoptics, *Rend. Semin. Mat. Univ. Padova*, 125, 2011, 39–49.

[45] W. Mozgawa, M. Skrzypiec, Some properties of secantoptics of ovals, *Beitr. Algebra Geom.*, 53, 2012, 261–272.

[46] W. Mozgawa, M. Skrzypiec, Integral formula for secantoptics and its application, *Annales UMCS*, 66, 2012, 49–62.

[47] M. Michalska, A sufficient condition for the convexity of the area of an isoptic curve of an oval, *Rend. Sem. Mat. Univ. Padova*, 110, 2003, 161–169.

[48] M. Michalska, W. Mozgawa, α-isoptics of a triangle and their connection to α-isoptic of an oval, *Rend. Semin. Mat. Univ. Padova*, 133, 2015, 159–172.

[49] J. C. C. Nitsche, Isoptic Characterization of a Circle (Proof of a Conjecture of M. S. Klamkin), *The American Mathematical Monthly*, Vol. 97, No. 1, 1990, 45–47.

[50] B. Odehnal, Equioptic curves of conic sections, *J. Geom. Graph.*, 14, No. 1, 2010, 29–43.

[51] J. V. Poncelet, *Traité des propriétés projectives des figures: ouvrage utile a qui s'occupent des applications de la géométrie descriptive et d'opérations*

géométriques sur le terrain, Vols. 1-2, 2nd ed. Paris: Gauthier-Villars, 1865–66.

[52] E. Raphaël, J.-M. di Meglio, M. Berger, E. Calabi, Convex particles at interfaces, *J. Phys. I France*, 2, 5, 1992, 571–579.

[53] H. W. Richmond, On isoptic families of curves , *Edinburgh Math. Notes*, 36, 1947, 18–21.

[54] L. Santaló, *Integral geometry and geometric probability. With a foreword by Mark Kac*, Cambridge Mathematical Library. Cambridge: Cambridge University Press, 2004.

[55] M. Skrzypiec, A note on secantoptics, *Beiträge Algebra Geom.*, 49 no. 1, 2008, 205–215.

[56] D. Szałkowski, Singular points of isoptics of open rosettes, *An. Ştiinţ. Univ. Al. I. Cuza Iaşi, Ser. Nouã, Mat.*, 60, No. 1, 2014, 85-98.

[57] D. Szałkowski, Isoptics of open rosettes. II, *An. Ştiinţ. Univ. Al. I. Cuza, Iaşi, Ser. Nouã, Mat.*, 53, No. 1, 2007, 167–176 .

[58] D. Szałkowski, Isoptics of open rosettes, *Ann. Univ. Mariae Curie-Skłodowska*, Sect. A 59, 2005, 119–128.

[59] C. Taylor, Note of a theory of orthoptic and isoptic loci, *Lond. R. S. Proc.*, 37, 1884, 138–141.

[60] W. Wunderlich, Kurven mit isopticher Ellipse, *Monathsch. Math.*, 75, 1971, 346–362.

[61] W. Wunderlich, Kurven mit isoptischem Kreis, *Aequat. Math.*, 6, 1971, 71–81.

[62] W. Wunderlich, Contributions to the geometry of cam mechanisms with oscillating followers, *Journal of Mechanisms*, 6, 1, 1971, 1–20.

Zbiory Koebego dla rodzin funkcji z ustalonym kierunkiem wypukłości

Paweł Zaprawa[1]

Streszczenie

W pracy zaprezentowane zostaną problemy związane z wyznaczaniem zbiorów Koebego dla wybranych rodzin funkcji analitycznych zmiennej zespolonej. Głównym narzędziem wykorzystywanym w tych badaniach jest zasada podporządkowania. Wyznaczone zostaną zbiory i funkcje ekstremalne, dzięki którym znalezione zostaną zbiory Koebego dla klas $\mathcal{X}^*$ oraz $\mathcal{Y}^*$ złożonych z funkcji gwiaździstych i wypukłych odpowiednio w kierunku osi rzeczywistej i w kierunku osi urojonej. Przedstawione zostaną także wyniki dla pewnych klas związanych z rodzinami $\mathcal{X}^*$ i $\mathcal{Y}^*$.

Abstract

In this paper we discuss the problem of determination of the Koebe sets for selected families of analytic functions of a complex variable. The main tool used in this research is the Lindelöf principle. We derive the extremal sets and the extremal functions and, in a consequence, the Koebe sets for two classes: $\mathcal{X}^*$ and $\mathcal{Y}^*$ consisting of functions that are starlike and convex in the direction of the real axis and the imaginary axis, respectively. Moreover, the results for some related classes are presented.

Słowa kluczowe: zbiory Koebego, funkcje gwiaździste, funkcje wypukłe w ustalonym kierunku

[1]Zakład Matematyki; Wydział Mechaniczny; Politechnika Lubelska
e-mail: p.zaprawa@pollub.pl

1 Wstęp

Zagadnienia związane ze zbiorami wartości przyjmowanych przez funkcje analityczne są głęboko zakorzenione w geometrycznej teorii funkcji analitycznych. Należą, obok problemów współczynnikowych, do klasycznych problemów pojawiających się już w początkowym okresie rozwoju tej teorii. Źródłem tych zagadnień są prace powstałe w pierwszych dwóch dekadach XX wieku.

W 1907 Koebe [13] sformułował hipotezę dotyczącą nieprzyjmowanych wartości przez funkcje analityczne i jednolistne, mające w kole jednostkowym $\Delta \equiv \{z \in \mathbb{C} : |z| < 1\}$ rozwinięcie w szereg Taylora postaci

$$f(z) = z + a_2 z^2 + a_3 z^3 + \dots\ ,\ z \in \Delta. \tag{1.1}$$

Klasę takich funkcji oznaczamy przez $\mathcal{S}$. Mianowicie, Koebe przewidywał, że moduł nieprzyjętej wartości musi być niemniejszy niż 1/4. Dowód tego faktu podał w 1916 Bieberbach [1] w następującym twierdzeniu, zwanym często twierdzeniem Koebego-Bieberbacha.

Twierdzenie 1.1. *Jeśli $f \in \mathcal{S}$, to $\Delta_{1/4} \subset f(\Delta)$.*

Symbol Δ_r oznacza zbiór $\{z \in \mathbb{C} : |z| < r\}$.

Bieberbach w dowodzie wykorzystał operator $f \mapsto df(z)/(d - f(z))$, który przy założeniach powyższego twierdzenia zachowuje jednolistność, oraz udowodniony właśnie fakt, że $|a_2| \leqslant 2$.

W tym samym czasie publikowane są pierwsze twierdzenia dotyczące oszacowania wyrażeń $|f(z)|$ i $|f'(z)|$, gdy f należy do ustalonej klasy funkcji. Autorem wyników dla klasy $\mathcal{S}$ był Koebe. Wykazał on następujące twierdzenie.

Twierdzenie 1.2. *Jeśli $f \in \mathcal{S}$, to dla $|z| < 1$,*

$$\frac{|z|}{(1+|z|)^2} \leqslant |f(z)| \leqslant \frac{|z|}{(1-|z|)^2} \quad \textit{oraz} \quad \frac{1-|z|}{(1+|z|)^3} \leqslant |f'(z)| \leqslant \frac{1+|z|}{(1-|z|)^3}.$$

Także w dowodzie ww. faktów (zwanych twierdzeniami Koebego o pokryciu i zniekształceniu) wykorzystuje się oszacowanie $|a_2| \leqslant 2$ dla funkcji należących

do $\mathcal{S}$. Jak widać, twierdzenie 1.2 daje również informacje o wartościach przyjmowanych, a także nieprzyjmowanych, przez funkcje jednolistne. Biorąc supremum wyrażenia $|z|/(1+|z|)^2$ po wszystkich $z \in \Delta$ otrzymujemy stałą 1/4 występującą w tezie twierdzenia 1.1.

2 Podstawowe definicje, fakty i oznaczenia

Oznaczmy symbolem $\mathcal{A}$ klasę wszystkich funkcji analitycznych w kole Δ postaci (1.1). Rozwinięcie (1.1) oznacza, że dla $f \in \mathcal{A}$ mamy

$$f(0) = 0,\ f'(0) = 1. \tag{2.1}$$

Dla ustalonej klasy $A \subset \mathcal{A}$ przyjmujemy następujące określenie.

Definicja 2.1. *Zbiór $K_A \equiv \bigcap_{f\in A} f(\Delta)$ nazywamy zbiorem Koebego klasy A.*

Tak zdefiniowane pojęcie zbioru Koebego pojawia się po raz pierwszy w pracy [14] Krzyża i Reade'a z 1967.

Twierdzenie 1.1 można zapisać wykorzystując definicję zbioru Koebego oraz tzw. niezmienniczość klasy $\mathcal{S}$ względem obrotu w następujący sposób.

Twierdzenie 2.1. $K_{\mathcal{S}} = \Delta_{1/4}$.

Własność niezmienniczości klasy $A \subset \mathcal{A}$ względem obrotu oznacza, że dla dowolnego $\varphi \in \mathbb{R}$ mamy

$$f \in A \Leftrightarrow f_\varphi \in A, \tag{2.2}$$

gdzie $f_\varphi(z) = e^{-i\varphi} f(ze^{i\varphi})$. Łatwo można uzasadnić, że jeśli klasa A posiada własność (2.2), to zbiorem Koebego takiej klasy jest koło Δ_m, gdzie

$$m = \lim_{r\to 1^-} \min\{|f(z)| : f \in A, |z| = r\}.$$

Niezależnie od tego, czy zbiór Koebego jest kołem czy też nie, promień największego koła Δ_r zawartego w K_A nazywamy stałą Koebego dla klasy A. Jeśli

zatem w_0 jest nieprzyjmowaną wartością przez funkcje klasy A, to moduł liczby w_0 musi być niemniejszy od stałej Koebego tej klasy.

Wprost z definicji zbioru Koebego wynika następujący lemat.

Lemat 2.1. *Jeśli* $A_1, A_2 \subset \mathcal{A}$ *i* $A_1 \subset A_2$*, to* $K_{A_2} \subset K_{A_1}$.

Funkcjami ekstremalnymi w twierdzeniu 2.1 są funkcje postaci $f(z) = \frac{z}{(1-ze^{i\theta})^2}$, które należą również do klasy $\mathcal{S}^*$ funkcji gwiaździstych względem punktu 0 i unormowanych warunkiem (2.1). Przypomnijmy, że funkcja f jest gwiaździsta względem punktu w_0, jeśli zbiór $f(\Delta)$ jest gwiaździsty względem tego punktu. Z kolei, o zbiorze D mówimy, że jest gwiaździsty względem punktu w_0, jeśli dla każdego punktu w należącego do wnętrza zbioru D, odcinek $[w_0, w]$ jest zawarty w tym zbiorze.

Zatem z faktu, że $\mathcal{S}^* \subset \mathcal{S}$ i z lematu 2.1 wynika następujące twierdzenie.

Twierdzenie 2.2. $K_{\mathcal{S}^*} = \Delta_{1/4}$.

Problematyka wyznaczania zbiorów Koebego dla różnych klas funkcji analitycznych stała się jednym z najważniejszych zagadnień rozważanych w II połowie XX wieku przez matematyków zajmujących się geometryczną teorią funkcji analitycznych. Warto w tym miejscu wspomnieć o pracach McGregora (1964), Merkesa i Scotta (1966), Brannana i Kirwana (1969), Goodmana (1977, 1979), Goodmana i Saffa (1979), Bshouty'ego, Hengartnera i Schobera (1980), (patrz m.in. [2–4, 18, 19]). Badania nad tymi problemami w zakresie funkcji harmonicznych prowadzili Clunie i Sheil-Small (1984) oraz Mateljević (2007), a w zakresie funkcji quasikonforemnych - Juve (1954) oraz Ikoma (1970).

Należy zauważyć bardzo istotny wkład w badania nad zbiorami Koebego matematyków związanych z lubelskim ośrodkiem naukowym. Niezwykle ważne są przede wszystkim wyniki Krzyża (wspólne prace ze Złotkiewiczem (1971, [15]) i Reade (1987, [14]), Złotkiewicza (wraz z Reade (1971, [20])) oraz Lewandowskiego (we współpracy z Miazgą i Szynalem (1975, [16] oraz 1980, [17])). Warto wspomnieć także o pracach Goduli (1986), Koczana (1989), Sobczak-Kneć (2007), a w ostatnich latach Koczana i Zaprawy (patrz m.in. [6–12, 21–23]).

Jeden ze wspomnianych powyżej rezultatów będzie wykorzystany w dalszej części tej pracy. Wynik ten został uzyskany przez Reade'a i Złotkiewicza.

Oznaczmy przez $\mathcal{K}(1)$ zbiór funkcji $f \in \mathcal{A}$, które są jednolistne i wypukłe w kierunku osi rzeczywistej. O funkcji f mówimy, że jest wypukła w kierunku osi rzeczywistej, jeśli przecięcie dowolnej prostej równoległej do osi rzeczywistej i $f(\Delta)$ jest zbiorem spójnym (zbiorem pustym, odcinkiem lub półprostą zawartą w tej prostej, całą prostą). Zastępując oś rzeczywistą osią urojoną otrzymamy klasę $\mathcal{K}(i)$ funkcji jednolistnych i wypukłych w kierunku osi urojonej. Dla $\mathcal{K}(1)$ mamy następujące twierdzenie.

Twierdzenie 2.3 ([20]). $K_{\mathcal{K}(1)} = \{w \in \mathbb{C} : 8|w|(|w| + |\operatorname{Re} w|) < 1\}$.

Zbiór $K_{\mathcal{K}(1)}$ jest symetryczny względem obu osi płaszczyzny zespolonej. Równanie biegunowe brzegu tego zbioru w prawej półpłaszczyźnie jest postaci $w = \varrho(\varphi)e^{i\varphi\pi/2}$, $\varphi \in [-1, 1]$, gdzie $\varrho(\varphi) = \frac{1}{4\cos(\pi\varphi/4)}$. Mamy ponadto poniższy wniosek.

Wniosek 2.1. $K_{\mathcal{K}(i)} = iK_{\mathcal{K}(1)}$.

Ważną rolę przy wyznaczaniu zbiorów Koebego odgrywa jednolistność funkcji danej klasy. Pozwala ona na wykorzystanie jako podstawowego narzędzia zasady podporządkowania.

Niech f i F będą funkcjami analitycznymi w Δ. Mówimy, że funkcja f jest podporządkowana funkcji F (piszemy: $f \prec F$), jeśli istnieje funkcja analityczna ω taka, że $\omega(0) = 0$, $|\omega(z)| < 1$ i $f(z) = F(\omega(z))$ dla $z \in \Delta$. Zasadą podporządkowania, lub też zasadą Lindelöfa, nazywamy następujące twierdzenie.

Twierdzenie 2.4. *Jeśli $f \prec F$, to*

1. $f(0) = F(0)$,
2. $|f'(0)| \leqslant |F'(0)|$,
3. $f(\Delta_r) \subset F(\Delta_r)$ *dla każdego* $r \in (0, 1]$.

Jeśli dodatkowo funkcja F jest jednolistna, to podporządkowanie $f \prec F$ jest równoważne temu, że $f(\Delta) \subset F(\Delta)$.

W dalszej części tej pracy przedstawiona zostanie metoda wyznaczania zbiorów Koebego przy wykorzystaniu podporządkowania dla klasy złożonej z funkcji, które są jednocześnie gwiaździste i wypukłe w kierunku osi rzeczywistej. Aby uprościć notację, zamiast symbolu $\mathcal{K}(1) \cap \mathcal{S}^*$ do oznaczenia tej klasy będziemy używali symbolu $\mathcal{X}^*$. Zastępując w powyższym określeniu wypukłość w kierunku osi rzeczywistej wypukłością w kierunku osi urojonej otrzymamy klasę $\mathcal{K}(1) \cap \mathcal{S}^*$ oznaczaną w skrócie $\mathcal{Y}^*$. Oczywiście funkcje należące do $\mathcal{X}^*$ lub $\mathcal{Y}^*$ są jednolistne.

Przydatne będą następujące oznaczenia: $(D)^c = \{z : z \notin D\}$, $\overline{D} = \{\overline{z} : z \in D\}$, $\lambda D = \{\lambda z : z \in D\}$.

3 Zbiory ekstremalne klasy $\mathcal{X}^*$

Udowodnimy najpierw, następujący lemat.

Lemat 3.1. *Jeśli $f \in \mathcal{X}^*$, to $\overline{f(\overline{z})}$, $-f(-z)$, $-\overline{f(-\overline{z})}$ też należą do $\mathcal{X}^*$.*

Dowód. Oznaczmy przez D obraz koła Δ przez funkcję $f \in \mathcal{X}^*$. Wtedy, oznaczając przez f_1, f_2, f_3 funkcje z tezy lematu, mamy

$$f_1(\Delta) = \overline{D},\ f_2(\Delta) = -D,\ f_3(\Delta) = -\overline{D}.$$

Ponieważ powyższe zbiory są symetryczne do D względem, odpowiednio, osi rzeczywistej, punktu 0, osi urojonej, zatem są one zarówno gwiaździste, jak i wypukłe w kierunku osi rzeczywistej. □

Wnioskujemy stąd, że zbiór Koebego klasy $\mathcal{X}^*$ jest symetryczny względem obu osi płaszczyzny zespolonej. Rzeczywiście, jeśli w jest punktem brzegowym zbioru $K_{\mathcal{X}^*}$, to istnieje funkcja $f \in \mathcal{X}^*$, dla której $w \in \partial f(\Delta)$. O takiej funkcji mówimy, że jest ekstremalna w problemie wyznaczania zbiorów Koebego. Na mocy lematu 3.1, punkty $\overline{w}$, $-w$, $-\overline{w}$ także należą do brzegu zbioru $K_{\mathcal{X}^*}$, co daje nam wspomnianą wcześniej symetrię. Oznacza to, że przy wyznaczaniu brzegu zbioru Koebego dla

$\mathcal{X}^*$ wystarczy ograniczyć się do tej części brzegu, który leży w pierwszej ćwiartce płaszczyzny zespolonej.

Wyznaczymy teraz zbiory ekstremalne, tj. takie zbiory E, które są obrazami koła Δ przez funkcje klasy $\mathcal{X}^*$, a choćby jeden punkt brzegu zbioru E jest zarazem punktem brzegowym zbioru $K_{\mathcal{X}^*}$.

Niech $f \in \mathcal{X}^*$ i $w_0 \notin f(\Delta)$, przy czym $w_0 = \varrho e^{i\varphi\pi/2}$, $\varrho > 0$, $\varphi \in [0,1]$. Z gwiaździstości funkcji f wynika, że półprosta $\{w_0 t : t \geqslant 1\}$ jest rozłączna z $f(\Delta)$. Z wypukłości w kierunku osi rzeczywistej wynika, że jedna z półprostych poziomych wychodzących z w_0 też jest rozłączna z $f(\Delta)$. Z powyższego otrzymujemy, że zbiory ekstremalne są dopełnieniami sektorów postaci

$$E_1 = \{w \in \mathbb{C} \,:\, 0 \leqslant \arg(w - w_0) \leqslant \arg w_0\}$$

lub

$$E_2 = \{w \in \mathbb{C} \,:\, \arg w_0 \leqslant \arg(w - w_0) \leqslant \pi\}.$$

W dalszej kolejności wyznaczymy funkcje klasy $\mathcal{X}^*$ odwzorowujące Δ na zbiory $(E_j)^c$, $j = 1,2$, to jest dopełnienia sektorów o miarach $\varphi\pi/2$ i $(1-\varphi/2)\pi$, $\varphi \in [0,1]$. Oba te zbiory można uzyskać z dopełnienia zbioru

$$\{w \in \mathbb{C} \,:\, |\arg(w-a)| \leqslant \varphi\pi/4\},\ a = 1/(4-\varphi),$$

stosując odpowiednio dobrane przesunięcie i obrót (w przypadku zbioru $(E_2)^c$ trzeba dodatkowo zastąpić $\varphi/2$ przez $1-\varphi/2$). Oznacza to, że funkcje ekstremalne można uzyskać składając funkcję

$$f(z) = a\left[1 - \left(\frac{1-z}{1+z}\right)^{2-\varphi/2}\right],\ a = \frac{1}{4-\varphi} \tag{3.1}$$

z transformacją Möbiusa

$$f_t(z) = \frac{f\left(\frac{z-it}{1+itz}\right) - f(-it)}{(1-t^2)f'(-it)},\ t \in (-1,1) \tag{3.2}$$

i obrotem o kąt θ, z odpowiednio dobranymi parametrami t, θ.

Niech $\varphi \in [0,1]$ i niech f, f_t będą dane wzorami (3.1) i (3.2). Mamy więc

$$f'(-it) = \left(\frac{1+it}{1-it}\right)^{1-\varphi/2} \frac{1}{(1-it)^2},$$

skąd

$$f_t(z) = \frac{(1-it)^2}{1-t^2}\left(\frac{1-it}{1+it}\right)^{1-\varphi/2}\left[f\left(\frac{z-it}{1+itz}\right)-a\right]+a_t,$$

gdzie $a_t = a\frac{1+t^2}{1-t^2}$.

Ponieważ $f(\Delta) = \mathbb{C}\setminus\{w\in\mathbb{C} : |\arg(w-a)| \leqslant \varphi\pi/4\}$, zatem $(f_t(\Delta))^c$ jest sektorem o kącie $\varphi\pi/2$ i wierzchołku a_t leżącym na dodatniej półosi rzeczywistej. Ramiona l_1 i l_2 tego sektora tworzą z osią rzeczywistą kąty

$$\theta_1\pi = -\varphi\pi/4+2\arg(1-it)+(1-\varphi/2)\arg\left(\frac{1-it}{1+it}\right) = -\varphi\pi/4-(4-\varphi)\operatorname{arc\,tg} t$$

oraz

$$\theta_2\pi = \varphi\pi/4+2\arg(1-it)+(1-\varphi/2)\arg\left(\frac{1-it}{1+it}\right) = \varphi\pi/4-(4-\varphi)\operatorname{arc\,tg} t.$$

Należy teraz obrócić ten sektor o kąt $\varphi\pi/2$, aby dopełnienie tego obróconego zbioru było zbiorem gwiaździstym względem punktu 0 i wypukłym w kierunku osi rzeczywistej. Po takim obrocie otrzymamy sektor

$$\{w\in\mathbb{C} : 0 \leqslant \arg(w-w_0)| \leqslant \varphi\pi/2\},\ w_0 = a_t e^{i\varphi\pi/2}.$$

Aby uzyskać pożądany kierunek ramion sektora, musi zachodzić warunek $\theta_2 = 0$, czyli

$$t = t(\varphi) = \operatorname{tg}\left(\frac{\pi}{4}\cdot\frac{\varphi}{4-\varphi}\right). \tag{3.3}$$

Funkcją realizującą opisany obrót jest

$$F_\varphi(z) = e^{i\varphi\pi/2} f_{t(\varphi)}(ze^{-i\varphi\pi/2}), \quad z\in\Delta. \tag{3.4}$$

Z nierówności $0 \leqslant \varphi \leqslant 1$ i (3.3) wynika, że $t \in [0, 2-\sqrt{3}]$.

Z powyższego rozumowania wnioskujemy, że $F_\varphi(\Delta) = (E_1)^c$, przy czym $w_0 = A(\varphi)e^{i\varphi\pi/2}$, gdzie $A(\varphi) = a_{t(\varphi)}$. Mamy zatem

$$A(\varphi) = \frac{1}{(4-\varphi)\cos\left(\frac{\varphi\pi}{8-2\varphi}\right)},\ \varphi \in [0,1]. \tag{3.5}$$

W szczególnym przypadku, gdy $\varphi = 0$ (wówczas $t = 0$), mamy $f(z) = \frac{z}{(1+z)^2}$ i w konsekwencji $F_0(z) = \frac{z}{(1+z)^2}$.

W sposób podobny do przedstawionego powyżej można wyznaczyć funkcję jednolistną odwzorowującą Δ na dopełnienie zbioru E_2. Funkcja ta jest postaci

$$G_\varphi(z) = e^{i\varphi\pi/2} g_{t(\varphi)}(ze^{-i\varphi\pi/2}), \quad z \in \Delta, \tag{3.6}$$

gdzie

$$g_t(z) = \frac{g\left(\frac{z-it}{1+itz}\right) - g(-it)}{(1-t^2)g'(-it)},\ t \in (-1,1), \tag{3.7}$$

$$g(z) = b\left[1 - \left(\frac{1-z}{1+z}\right)^{1+\varphi/2}\right],\ b = \frac{1}{2+\varphi}. \tag{3.8}$$

Mamy zatem

$$G_\varphi(\Delta) = \mathbb{C} \setminus \{w \in \mathbb{C}\ : \varphi\pi/2 \leqslant \arg(w - w_0)| \leqslant \pi\},\ w_0 = B(\varphi)e^{i\varphi\pi/2},$$

gdzie

$$B(\varphi) = \frac{1}{(2+\varphi)\cos\left(\frac{(2-\varphi)\pi}{4+2\varphi}\right)},\ \varphi \in (0,1], \tag{3.9}$$

przy czym t i φ związane są zależnością

$$t = t(\varphi) = -\operatorname{tg}\left(\frac{\pi}{4} \cdot \frac{2-\varphi}{2+\varphi}\right). \tag{3.10}$$

Ze względu na to, że 0 musi być punktem wewnętrznym zbioru $G_\varphi(\Delta)$ przyjmujemy, że $\varphi \in (0,1]$. Uwzględniając wzór (3.10), otrzymujemy, że $t \in (-1, \sqrt{3}-2]$.

4 Zbiór Koebego dla klasy $\mathcal{X}^*$

Rozważmy teraz funkcje $A(\varphi)$ i $B(\varphi)$ odpowiadające krzywym zakreślonym przez zmieniające się wierzchołki sektorów opisanych w poprzednim rozdziale. Zauważmy, że obie funkcje $[0,1] \ni \varphi \mapsto 4-2\varphi$ oraz $[0,1] \ni \varphi \mapsto \cos\left(\frac{\varphi\pi}{4-2\varphi}\right)$ są malejące. Zatem $A(\varphi)$ jest w przedziale $[0,1]$ funkcją rosnącą oraz

$$\max\{A(\varphi) : \varphi \in [0,1]\} = A(1) = 2\sqrt{3}/9.$$

Z drugiej strony mamy $B(\varphi) = A(2-\varphi)$, a zatem

$$\min\{B(\varphi) : \varphi \in (0,1]\} = B(1) = A(1).$$

Oznacza to, że

$$A(\theta) \leqslant B(\psi) \quad \text{dla każdego } \theta \in [0,1], \psi \in (0,1]. \tag{4.1}$$

Można teraz zdefiniować dwa obszary H_1 i H_2 leżące w pierwszej ćwiartce płaszczyzny zespolonej. Obszar H_1 jest ograniczony krzywą $A(\varphi)e^{i\varphi\pi/2}$, $\varphi \in [0,1]$ oraz odcinkami $[0,1/4]$ i $[0,2\sqrt{3}i/9]$, zaś obszar H_2 jest zbiorem nieograniczonym, którego brzeg stanowi dodatnia półoś rzeczywista, odcinek $[0,2\sqrt{3}i/9]$ i krzywa $B(\varphi)e^{i\varphi\pi/2}$, $\varphi \in (0,1]$. W poprzednim rozdziale wykazliśmy, że

$$H_1 = \bigcap_{\varphi\in[0,1]} F_\varphi(\Delta) \cap \{w \in \mathbb{C} \, : \operatorname{Re} w > 0, \operatorname{Im} w > 0\} \tag{4.2}$$

oraz

$$H_2 = \bigcap_{\varphi\in(0,1]} G_\varphi(\Delta) \cap \{w \in \mathbb{C} \, : \operatorname{Re} w > 0, \operatorname{Im} w > 0\}. \tag{4.3}$$

Co więcej, z (4.1) wynika, że

$$H_1 \subset H_2. \tag{4.4}$$

Twierdzenie 4.1. *Zbiór Koebego klasy $\mathcal{X}^*$ jest obszarem ograniczonym i symetrycznym względem obu osi płaszczyzny zespolonej. Brzeg tego zbioru w pierwszej ćwiartce płaszczyzny zespolonej jest postaci $w = A(\varphi)e^{i\varphi\pi/2}$, $\varphi \in [0,1]$, gdzie $A(\varphi)$ jest dane wzorem (3.5).*

Dowód. Niech w_0 będzie dowolnym punktem brzegowym zbioru Koebego klasy $\mathcal{X}^*$, $\arg w_0 = \varphi\pi/2$, $\varphi \in [0,1]$. Oznacza to, że istnieje funkcja $f \in \mathcal{X}^*$ taka, że $w_0 \in \partial f(\Delta)$. Gwiaździstość oraz wypukłość w kierunku osi rzeczywistej funkcji f powodują, że

$$f(\Delta) \subset F_\varphi(\Delta) \quad \text{lub} \quad f(\Delta) \subset G_\varphi(\Delta),$$

gdzie $\varphi \in [0,1]$ (w przypadku funkcji G_φ wykluczamy 0).

Z jednolistności funkcji F_φ, G_φ oraz z tego, że $f(0) = F_\varphi(0) = G_\varphi(0) = 0$ wynika, że

$$f \prec F_\varphi \quad \text{lub} \quad f \prec G_\varphi.$$

Oznacza to, że mamy

$$1 = f'(0) \leqslant F'_\varphi(0) = 1 \quad \text{lub} \quad 1 = f'(0) \leqslant G'_\varphi(0) = 1,$$

co w konsekwencji daje

$$f = F_\varphi \quad \text{lub} \quad f = G_\varphi.$$

Powyższe rozumowanie prowadzi do stwierdzenia, że w celu wyznaczenia zbioru Koebego wystarczy rozważać jedynie funkcje F_φ, G_φ dane wzorami (3.4) i (3.6). Stąd i z (4.4) mamy

$$K_{\mathcal{X}^*} \cap \{w \in \mathbb{C} \,:\, \operatorname{Re} w > 0, \operatorname{Im} w > 0\} = H_1. \tag{4.5}$$

Co więcej, punkty $w_1 = 1/4$ oraz $w_2 = 2\sqrt{3}i/9$ należą do zbioru $K_{\mathcal{X}^*}$. Mianowicie, dla funkcji $F_0(z) = \frac{z}{(1+z)^2}$ mamy $F_0(\Delta) = \mathbb{C} \setminus [1/4, \infty)$, zaś dla funkcji F_1 jest $F_1(\Delta) = \mathbb{C} \setminus \{w \in \mathbb{C} \,:\, \operatorname{Re} w \geqslant 0, \operatorname{Im} w \geqslant 2\sqrt{3}/9\}$.

Z powyższych faktów i symetrii zbioru Koebego względem obu osi płaszczyzny zespolonej wynika teza. □

Wniosek 4.1. Jeśli $f \in \mathcal{X}^*$, to $\Delta_{1/4} \subset f(\Delta)$.

Inaczej mówiąc, stałą Koebego klasy $\mathcal{X}^*$ jest 1/4.

5 Zbiory Koebego dla klas związanych z $\mathcal{X}^*$

Zauważmy najpierw, że zbiór Koebego dla klasy $\mathcal{Y}^*$ można w łatwy sposób otrzymać ze zbioru $K_{\mathcal{X}^*}$. Rzeczywiście, jeśli funkcja $f \in \mathcal{A}$ jest gwiaździsta i wypukła w kierunku osi rzeczywistej, to funkcja $if(z/i)$ też jest gwiaździsta, a ponadto wypukła w kierunku osi urojonej. W konsekwencji, jeśli E jest zbiorem ekstremalnym dla klasy $\mathcal{X}^*$, to zbiór iE jest zbiorem ekstremalnym dla klasy $\mathcal{Y}^*$. Stąd

Twierdzenie 5.1.

$$K_{\mathcal{Y}^*} = iK_{\mathcal{X}^*}.$$

Na mocy definicji klas $\mathcal{X}^*$ i $\mathcal{Y}^*$ mamy $\mathcal{X}^* \subset \mathcal{S}^*$ oraz $\mathcal{Y}^* \subset \mathcal{S}^*$. Zatem, zgodnie z lematem 2.1,

$$K_{\mathcal{S}^*} \subset K_{\mathcal{X}^*} \quad \text{oraz} \quad K_{\mathcal{S}^*} \subset K_{\mathcal{Y}^*}.$$

Łatwo sprawdzić, że powyższe zawierania mają miejsce gdyż, na mocy twierdzenia 2.2, $K_{\mathcal{S}^*}$ jest kołem o promieniu 1/4, a z drugiej strony

$$1/4 = A(0) = \min\{A(\varphi) : \varphi \in [0,1]\}.$$

Podobnie, z definicji klas $\mathcal{X}^*$ i $\mathcal{Y}^*$ wynika, że rodziny te są podzbiorami klas $\mathcal{K}(1)$ i $\mathcal{K}(i)$, odpowiednio. Zatem, znów na mocy lematu 2.1, mamy

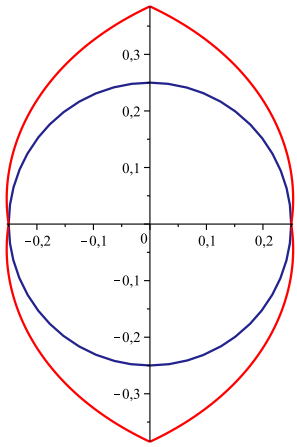
$$K_{\mathcal{K}(1)} \subset K_{\mathcal{X}^*} \quad \text{oraz} \quad K_{\mathcal{K}(i)} \subset K_{\mathcal{Y}^*}.$$

Powyższe relacje pomiędzy zbiorami Koebego dla klas $\mathcal{X}^*$, $\mathcal{S}^*$ i $\mathcal{K}(1)$ pokazują Rysunki 1 i 2.

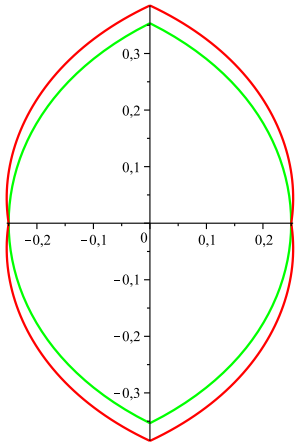
Rozważmy teraz rodzinę $\mathcal{X}^* \cap \mathcal{Y}^*$ złożoną z funkcji gwiaździstych, które są jednocześnie wypukłe w kierunku osi rzeczywistej i osi urojonej. Lemat 3.1 gwarantuje, że zbiór Koebego tej klasy też jest symetryczny względem obu osi płaszczyzny zespolonej. Z kolei, z lematu 2.1 wiemy, że

$$K_{\mathcal{X}^*} \subset K_{\mathcal{X}^* \cap \mathcal{Y}^*} \quad \text{oraz} \quad K_{\mathcal{Y}^*} \subset K_{\mathcal{X}^* \cap \mathcal{Y}^*}.$$

W celu wyznaczenia zbioru $K_{\mathcal{X}^* \cap \mathcal{Y}^*}$ przeanalizujemy zbiory ekstremalne dla klasy $\mathcal{X}^* \cap \mathcal{Y}^*$.



Rysunek 1: Zbiory Koebego: $K_{\mathcal{S}^*}$ wewnątrz i $K_{\mathcal{X}^*}$ na zewnątrz



Rysunek 2: Zbiory Koebego: $K_{\mathcal{K}(1)}$ wewnątrz i $K_{\mathcal{X}^*}$ na zewnątrz

Niech $f \in \mathcal{X}^* \cap \mathcal{Y}^*$ i $w_0 \notin f(\Delta)$, przy czym $w_0 = \varrho e^{i\pi\varphi/2}$, $\varrho > 0$, $\varphi \in [0,1]$. Z gwiaździstości funkcji f wynika, że półprosta $\{w_0 t : t \geqslant 1\}$ jest rozłączna z $f(\Delta)$. Na podstawie wypukłości w kierunku osi rzeczywistej wnioskujemy, że jedna z półprostych poziomych wychodzących z w_0 jest rozłączna z $f(\Delta)$. Wypukłość w kierunku osi urojonej powoduje, że z $f(\Delta)$ jest rozłączna jedna z półprostych pionowych wychodzących z w_0. Łącząc powyższe spostrzeżenia otrzymujemy, że zbiory ekstremalne mogą być dopełnieniami sektorów postaci

$$\begin{aligned} D_1 &= \{w \in \mathbb{C} : 0 \leqslant \arg(w - w_0) \leqslant \pi/2\} \\ D_2 &= \{w \in \mathbb{C} : \arg w_0 \leqslant \arg(w - w_0) \leqslant \pi\} \\ D_3 &= \{w \in \mathbb{C} : -\pi/2 \leqslant \arg(w - w_0) \leqslant \arg w_0\} \\ D_4 &= \{w \in \mathbb{C} : -\pi/2 \leqslant \arg(w - w_0) \leqslant \pi\}. \end{aligned}$$

Niech więc

$$h(z) = \frac{1}{3}\left[1 - \left(\frac{1-z}{1+z}\right)^{3/2}\right] \tag{5.1}$$

oraz

$$h_t(z) = \frac{h\left(\frac{z-it}{1+itz}\right) - h(-it)}{(1-t^2)h'(-it)},\ t \in (-1,1). \tag{5.2}$$

Zbiór $(h_t(\Delta))^c$ jest sektorem, którego wierzchołek leży na dodatniej półosi rzeczywistej, a ramiona tworzą z tą półosią kąty o miarach $-\pi/4 - 3\operatorname{arctg} t$ oraz $\pi/4 - 3\operatorname{arctg} t$. Jeśli teraz obrócimy ten zbiór o kąt $\varphi\pi/2 = \pi/4 + 3\operatorname{arctg} t$, to otrzymamy zbiór postaci D_1. Mamy zatem

$$t = t(\varphi) = \operatorname{tg}\left(\frac{\pi}{12} \cdot (2\varphi - 1)\right). \tag{5.3}$$

Zatem funkcja

$$H_\varphi(z) = e^{i\varphi\pi/2} h_t(\varphi)(z e^{-i\varphi\pi/2}) \quad , \quad z \in \Delta, \tag{5.4}$$

odwzorowuje koło Δ na dopełnienie zbioru D_1, przy czym wierzchołkiem tego sektora jest $w_0 = C(\varphi)e^{i\varphi\pi/2}$, gdzie

$$C(\varphi) = \frac{1}{3\cos(\varphi\pi/3 - \pi/6)},\ \varphi \in [0,1]. \tag{5.5}$$

Wykażemy teraz, że dla każdej funkcji f_j odwzorowującej koło Δ na zbiór $(D_j)^c$, $j = 2,3,4$ mamy $w_0 = k \cdot C(\varphi)e^{i\varphi\pi/2}$, $k > 1$.

Przypuśćmy jednak, że istnieje $f_4 \in \mathcal{X}^* \cap \mathcal{Y}^*$ taka, że $f_4(\Delta) = \mathbb{C} \setminus D_4$, gdzie $w_0 = k \cdot C(\varphi)e^{i\varphi\pi/2}$, $k \in (0,1]$. Z faktu, że $(D_4)^c \subset (D_1)^c$ wynika podporządkowanie $f_4 \prec f_1$. Ale wtedy, z zasady podporządkowania, otrzymujemy $1 = f_4'(0) < f_1'(0) = 1$, czyli sprzeczność.

Przypuśćmy teraz, że istnieje $f_2 \in \mathcal{X}^* \cap \mathcal{Y}^*$ taka, że $f_2(\Delta) = \mathbb{C} \setminus D_2$, przy czym wierzchołkiem sektora D_2 jest punkt $w_0 = k \cdot C(\varphi)e^{i\varphi\pi/2}$, $k \in (0,1]$. W tym przypadku funkcja $\tilde{f}_2(z) = -\overline{f_2(-\overline{z})}$ także należy do $\mathcal{X}^* \cap \mathcal{Y}^*$ i przekształca Δ na zbiór $(-\overline{D}_2)^c$, który jest zawarty w $(D_1)^c$. Oznacza to, że $\tilde{f}_2 \prec f_1$, co prowadzi do sprzeczności podobnie jak w przypadku funkcji f_4. Analogiczne rozumowanie może być przeprowadzone również dla funkcji f_3.

Z powyższego wynika, że spośród zbiorów $(D_j)^c$, $j = 1,2,3,4$ najmniejszą co do modułu nieprzyjmowaną wartością na półprostej o kierunku $e^{i\pi\varphi/2}$ jest $C(\varphi)$ postaci (5.5). Innymi słowy, zbiór Koebego klasy $\mathcal{X}^* \cap \mathcal{Y}^*$ wyznaczają zbiory $(D_1)^c$, a brzeg tego zbioru można wyrazić przy pomocy funkcji $C(\varphi)$. Mamy więc

Twierdzenie 5.2. *Zbiór Koebego klasy $\mathcal{X}^* \cap \mathcal{Y}^*$ jest obszarem ograniczonym i symetrycznym względem obu osi płaszczyzny zespolonej. Brzeg tego zbioru w pierwszej ćwiartce płaszczyzny zespolonej jest postaci $w = C(\varphi)e^{i\varphi\pi/2}$, $\varphi \in [0,1]$, gdzie $C(\varphi)$ jest dane wzorem (5.5).*

Analizując monotoniczność funkcji $C(\varphi)$, $\varphi \in [0,1]$ otrzymujemy następujący wniosek.

Wniosek 5.1. Jeśli $f \in \mathcal{X}^* \cap \mathcal{Y}^*$, to $\Delta_{1/3} \subset f(\Delta)$.

Na koniec warto porównać otrzymany zbiór $K_{\mathcal{X}^* \cap \mathcal{Y}^*}$ ze zbiorem Koebego dla klasy $\mathcal{K}(1) \cap \mathcal{K}(i)$. W klasie tej nie żądamy gwiaździstości funkcji, jak to się dzieje w przypadku funkcji klasy $\mathcal{X}^* \cap \mathcal{Y}^*$.

Klasa $\mathcal{K}(1) \cap \mathcal{K}(i)$ jest szczególnym przypadkiem rodziny funkcji analitycznych, które są wypukłe w dwóch ustalonych kierunkach $e^{i\pi\beta/2}$ oraz $e^{i\pi\gamma/2}$, gdzie $-1 \leqslant \beta \leqslant \gamma \leqslant 1$. Zbiór wszystkich takich funkcji oznaczany jest symbolem $K_{\beta,\gamma}$.

Klasa ta badana była w pracy [7]. Zgodnie z powyższym określeniem, $\mathcal{K}(1) \cap \mathcal{K}(i) = K_{0,1}$.

Niech f będzie funkcją należącą do $\mathcal{K}(1) \cap \mathcal{K}(i)$ i niech $w_0 \notin f(\Delta)$, przy czym $w_0 = \varrho e^{i\pi\varphi/2}$, $\varrho > 0$, $\varphi \in [0,1]$. Z wypukłości tej funkcji w kierunku obu osi płaszczyzny zespolonej otrzymujemy, że jedna z półprostych $\{w_0 + t : t \geqslant 0\}$ lub $\{w_0 - t : t \geqslant 0\}$ jest rozłączna z $f(\Delta)$ oraz jedna z półprostych $\{w_0 + ti : t \geqslant 0\}$ lub $\{w_0 - ti : t \geqslant 0\}$ jest rozłączna z $f(\Delta)$. Zatem zbiór $f(\Delta)$ zawarty jest w jednym ze zbiorów postaci $(C_j)^c$, $j = 1,2,3,4$, gdzie

$$\begin{aligned} C_1 &= \{w \in \mathbb{C} \,:\, 0 \leqslant \arg(w - w_0) \leqslant \pi/2\} \\ C_2 &= \{w \in \mathbb{C} \,:\, \pi/2 \leqslant \arg(w - w_0) \leqslant \pi\} \\ C_3 &= \{w \in \mathbb{C} \,:\, -\pi/2 \leqslant \arg(w - w_0) \leqslant 0\} \\ C_4 &= \{w \in \mathbb{C} \,:\, -\pi/2 \leqslant \arg(w - w_0) \leqslant \pi\}. \end{aligned}$$

Zauważmy, że $C_1 = D_1$ oraz $C_4 = D_4$, gdzie D_1 i D_4 są zbiorami ekstremalnymi dla klasy $\mathcal{X}^* \cap \mathcal{Y}^*$. W takim sam sposób jak dla klasy $\mathcal{X}^* \cap \mathcal{Y}^*$, można pokazać, że

$$K_{\mathcal{K}(1)\cap\mathcal{K}(i)} \cap \{w \in \mathbb{C} \,:\, \operatorname{Re} w > 0, \operatorname{Im} w > 0\} = \left(\bigcap_{\varphi\in[0,1]} H_\varphi(\Delta) \right) \cap \{w \in \mathbb{C} \,:\, \operatorname{Re} w > 0, \operatorname{Im} w > 0\},$$

gdzie H_φ jest postaci (5.4). Okazuje się więc, że pomimo iż

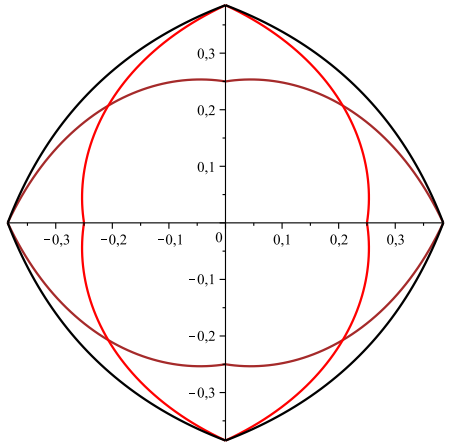
$$\mathcal{X}^* \cap \mathcal{Y}^* \subsetneqq \mathcal{K}(1) \cap \mathcal{K}(i),$$

to prawdziwe jest poniższe twierdzenie.

Twierdzenie 5.3.

$$K_{\mathcal{K}(1)\cap\mathcal{K}(i)} = K_{\mathcal{X}^*\cap\mathcal{Y}^*}.$$

Zbiory Koebego dla klas $\mathcal{X}^*$, $\mathcal{Y}^*$ i $\mathcal{X}^* \cap \mathcal{Y}^*$ (a zatem także dla klasy $\mathcal{K}(1) \cap \mathcal{K}(i)$) zaprezentowane zostały na Rysunku 3.



Rysunek 3: Zbiory Koebego: $K_{\mathcal{X}^*}$, $K_{\mathcal{Y}^*}$ wewnątrz i $K_{\mathcal{K}(1)\cap\mathcal{K}(i)}$ na zewnątrz

6 Uwagi dotyczące klas funkcji o współczynnikach rzeczywistych

W badaniach zbiorów Koebego często pojawiają się klasy złożone z funkcji, których współczynniki są liczbami rzeczywistymi. Takie dodatkowe założenie skutkuje symetrią zbiorów $f(\Delta)$ względem osi rzeczywistej. Ten fakt należy uwzględniać przy poszukiwaniu zbiorów ekstremalnych danej klasy i powoduje zwykle dużo większe trudności w wyznaczaniu zbiorów Koebego.

Opisany problem zilustrowany zostanie na przykładzie klasy $\mathcal{K}_{\mathbb{R}}(1)\cap\mathcal{K}_{\mathbb{R}}(i)$ złożonej z funkcji $f\in\mathcal{K}(1)\cap\mathcal{K}(i)$ mających wszystkie współczynniki rzeczywiste. Wykorzystując geometryczne własności klasy łatwo zauważyć, że zbiorami ekstremalnymi mogą być jedynie dopełnienia następujących zbiorów:

$$B_1 = C_1\cup\overline{C}_1,\ B_2 = C_2\cup\overline{C}_2,\ B_4 = C_4\cup\overline{C}_4,$$

gdzie C_j, $j=1,2,3,4$ zostały opisane w podrozdziale 5 (przyjmujemy, że wspólne wierzchołki w_0 tych sektorów mają argument w zakresie $[0,\pi)$). Zauważmy, że zbiór $(C_3\cup\overline{C}_3)^c$ jest półpłaszczyzną $\{w\in\mathbb{C}:\operatorname{Re}w<\operatorname{Re}w_0\}$. Jest on szczególnym

przypadkiem zbioru $(C_1 \cup \overline{C}_1)^c$ przy założeniu, że $\varphi = 0$. Z tego powodu nie musi być osobno uwzględniany w rozważaniach.

Funkcje jednolistne, które odwzorowują koło Δ na zbiory ekstremalne $(B_1)^c$, $(B_2)^c$, $(B_4)^c$ można wyznaczyć na podstawie wzorów Schwarza-Christoffela, gdyż zbiory te są uogólnionymi wielokątami. Ponadto, w przypadku dwóch pierwszych zbiorów, które są czworokątami, kąty wewnętrzne mają miary 0, $3\pi/2$, $-\pi$, $3\pi/2$, zaś dla uogólnionego trójkąta $(B_4)^c$ miary wewnętrzne są równe $\pi/2$, 0, $\pi/2$.

Argumentując w sposób podobny jak w podrozdziale 5 można stwierdzić, że punkty brzegowe zbioru Koebego dla klasy $\mathcal{K}_{\mathbb{R}}(1) \cap \mathcal{K}_{\mathbb{R}}(i)$ leżące w pierwszej ćwiartce płaszczyzny zespolonej są generowane jedynie przez zbiory postaci B_1. Funkcja odwzorowująca koło Δ na $(B_1)^c$ dana jest wzorem

$$J_\varphi(z) = \int_0^z \frac{\sqrt{(1-\zeta e^{-i\varphi})(1-\zeta e^{i\varphi})}}{(1-\zeta)(1+\zeta)^2}\, d\zeta,$$

gdzie gałąź pierwiastka zespolonego jest dobrana tak, by $\sqrt{1} = 1$. Zauważmy, że w granicznych przypadkach otrzymujemy $J_0(z) = \frac{z}{1+z}$ oraz $J_\pi(z) = \frac{1}{2}\log\frac{1+z}{1-z}$.

Wierzchołki sektorów B_1 leżą na krzywej $D([0,\pi))$, gdzie

$$D(\varphi) = J_\varphi(e^{i\varphi}) = e^{i\varphi}\int_0^1 \frac{\sqrt{(1-t)(1-te^{2i\varphi})}}{(1-te^{i\varphi})(1+te^{i\varphi})^2}\, dt. \tag{6.1}$$

Analizując kształt zbiorów $(B_1)^c$ i wykorzystując zasadę podporządkowania można również stwierdzić, że funkcja $[0,\pi) \mapsto \mathrm{Re}(D(\varphi))$ jest malejąca. Ponieważ $D(0) = 1/2$ i $\lim_{\varphi\to\pi}\mathrm{Re}(D(\varphi)) = -\infty$, więc istnieje $\varphi_0 \in (0,\pi)$ takie, że $\mathrm{Re}(D(\varphi_0)) = 0$. Oznaczmy zatem przez H_3 zbiór domknięty leżący w pierwszej ćwiartce płaszczyzny zespolonej ograniczony przez krzywe $D([0,\varphi_0)]$ oraz odcinki $[0,1/2]$ i $[0, D(\varphi_0)]$ oraz oznaczmy przez H wnętrze zbioru $H_3 \cup \overline{H}_3 \cup (-H_3) \cup (-\overline{H}_3)$. Otrzymujemy zatem następujące twierdzenie.

Twierdzenie 6.1.

$$K_{\mathcal{K}_{\mathbb{R}}(1)\cap\mathcal{K}_{\mathbb{R}}(i)} = H.$$

Wynik równoważny powyższemu, choć zaprezentowany w innej postaci, uzyskany został również w pracy [5].

Jak widać, wyznaczanie zbiorów Koebego dla klas funkcji o współczynnikach rzeczywistych wiąże się nie tylko z poszukiwaniem odwzorowań koła na bardziej skomplikowane zbiory ekstremalne. Również opis krzywej brzegowej bardzo często wyraża się przy pomocy funkcji nieelementarnych. Z tego powodu, zbiory Koebego np. dla klas $\mathcal{X}^*_{\mathbb{R}}$ czy $\mathcal{Y}^*_{\mathbb{R}}$ nie zostały jak dotąd wyznaczone.

Literatura

[1] L. Bieberbach, Über die Koeffizienten derjenigen Potenzreihen, welche eine schlichte Abbildung des Einheitskreises vermitteln, *Berl. Ber.*, 1916, 940–955.

[2] A.W. Goodman, The domain covered by a typically real function, *Proc. Am. Math. Soc.*, 64, 1977, 233–237.

[3] A.W. Goodman, Koebe domains for certain families, *Proc. Am. Math. Soc.*, 74, 1979, 87–94.

[4] A.W. Goodman, E.B. Saff, On univalent functions convex in one direction, *Proc. Am. Math. Soc.*, 73, 1979, 183–187.

[5] M. Gregorczyk, L. Koczan, Koebe domains for the class of typically real functions that are convex in two directions, przyjęte do druku w: *J. Appl. Anal.*

[6] L. Koczan, Typically real functions convex in the direction of the real axis, *Ann. Univ. Mariae Curie-Skłodowska*, 43, 1989, 23–29.

[7] L. Koczan, P. Zaprawa, On covering problems in the class of typically real functions, *Ann. Univ. Mariae Curie-Sklodowska*, 59, 2005, 51–65.

[8] L. Koczan, P. Zaprawa, Koebe domains for the classes of functions with ranges included in given sets, *J. Appl. Anal.*, 14, 2008, 43–52.

[9] L. Koczan, P. Zaprawa, On functions convex in the direction of the real axis with real coefficients, *Bull. Belg. Math. Soc. - Simon Stevin*, 18, 2011, 321–335.

[10] L. Koczan, P. Zaprawa, On functions convex in the direction of the real axis with a fixed second coefficient, *Appl. Math. Comput.*, 217, 2011, 6644–6651.

[11] L. Koczan, P. Zaprawa, Covering problems for functions n-fold symmetric and convex in the direction of the real axis, *Appl. Math. Comput.*, 219, 2012, 947–958.

[12] L. Koczan, P. Zaprawa, Koebe sets for the class of functions convex in two directions, *Turk. J. Math.*, 41, 2017, 282–292.

[13] P. Koebe, Über die Uniformisierung beliebiger analytischer Kurve, *Gött. Nachr.*, 2, 1907, 191–210.

[14] J. Krzyż, M.O. Reade, Koebe domains for certain classes of analytic functions, *J. Anal. Math.*, 18, 1967, 185–195.

[15] J. Krzyż, E. Złotkiewicz, Koebe sets for univalent functions with two preassigned values, *Ann. Acad. Sci. Fenn.*, 487, 1971, 12 p.

[16] Z. Lewandowski, J. Miazga, The Koebe domain for the class of typically-real functions under Montel's normalization, *Bull. Acad. Pol. Sci.*, 28, 1980, 465–470.

[17] Z. Lewandowski, J. Miazga, J. Szynal, Koebe domains for univalent functions with real coefficients under Montel's normalization, *Ann. Pol. Math.*, 30, 1975, 333–336.

[18] M.T. McGregor, On three classes of univalent functions with real coefficients, *J. Lond. Math. Soc.*, 39, 1964, 43–50.

[19] E.P. Merkes, W.T. Scott, Covering theorems for univalent functions, *Mich. Math. J.*, 13, 1966, 41–47.

[20] M.O. Reade, E. Złotkiewicz, On univalent functions with two preassigned values, *Proc. Am. Math. Soc.*, 30, 1971, 539–544.

[21] M. Sobczak-Kneć, Koebe domains for certain subclasses of starlike functions, *Ann. Univ. Mariae Curie-Skłodowska*, 61, 2007, 129–135.

[22] M. Sobczak-Kneć, P. Zaprawa, Covering domains for classes of functions with real coefficients, *Complex Var. Elliptic Equ.*, 52, 2007, 519–535.

[23] P. Zaprawa, Koebe sets for certain classes of circularly symmetric functions, *C. R., Math., Acad. Sci. Paris*, 354, 2016, 245–252.

Przestrzenie Hilberta z jądrem reprodukującym

Renata Rososzczuk[1]

Streszczenie

Praca zawiera krótkie omówienie teorii jąder reprodukujących. Podajemy przykłady przestrzeni Hilberta z jądrem reprodukującym. Również przedstawiamy pewne zastosowania jąder reprodukujących.

Abstract

The paper contains basis and general theory of reproducing kernels. Some examples of reproducing kernel spaces are included. We also present some applications of reproducing kernel.

Słowa kluczowe: jądra reprodukujące, przestrzenie Hardy'ego, ważone przestrzenie Bergmana, przestrzenie niezmiennicze, twierdzenie typu Beurlinga.

1 Wstęp

Niech E będzie dowolnym zbiorem oraz niech H będzie przestrzenią Hilberta złożoną z funkcji zespolonych określonych w zbiorze E. Mówimy, że H posiada własność reprodukowania, jeżeli istnieje taka rodzina

$$\{K_z\}_{z\in E} \subset H,$$

[1]Katedra Matematyki Stosowanej; Wydział Podstaw Techniki; Politechnika Lubelska; e-mail: r.rososzczuk@pollub.pl

że dla dowolnego $z \in E$ i dla dowolnego $f \in H$

$$f(z) = \langle f, K_z \rangle.$$

Funkcję $K : E \times E \to \mathbb{C}$ daną przez

$$K(z,w) \stackrel{\text{def}}{=} \langle K_w, K_z \rangle, \quad z, w \in E$$

nazywamy jądrem reprodukującym dla przestrzeni H, a parę (H, K) nazywamy przestrzenią z jądrem reprodukującym (na E) [24].

Jądro reprodukujące K istnieje dla przestrzeni Hilberta H wtedy i tylko wtedy, gdy dla każdego $w \in E$, odwzorowanie

$$\phi_w : f \to f(w), \quad f \in H$$

jest liniowym i ograniczonym funkcjonałem na H.

Jeśli $\{e_n\}$ jest dowolną bazą ortonormalną dla H, to dla wszystkich z i w zachodzi wzór

$$K(z,w) = \sum_{n=0}^{\infty} e_n(z)\overline{e_n(w)}. \tag{1.1}$$

Własność reprodukowania oraz wzór (1.1) na jądro reprodukujące w przypadku gdy $\{e_n\}$ jest bazą ortonormalną złożoną z wektorów własnych pojawiły się po raz pierwszy w pracy Stanisława Zaremby, profesora Uniwersytetu Jagiellońskiego w 1907 roku [25, 26]. Były to niewątpliwie epokowe i niezwykle interesujące odkrycia (chociaż początkowo nie przykładano do nich zbyt dużej uwagi). Czytelnikowi zainteresowanemu historią jąder reprodukujących polecam zapoznanie z pracami [21, rozdział 1] oraz [25].

Obecnie można wymienić wiele dziedzin, w których własność reprodukowania znajduje zastosowania. „Są to: analiza zespolona (w tym interpolacja à la Pick Nevanlinna), (cząstkowe) równania różniczkowe, aproksymacja (w szczególności funkcje gięte), wielomiany ortogonalne, teoria operatorów, teoria systemów, problemy odwrotne, *sampling theory and time-frequency analysis*, rachunek prawdopodobieństwa (w tym procesy stochastyczne) i statystyka, *learning theory*, stany

kontrahentne (tu jesteśmy już na poziomie optyki kwantowej), a także nieco dalej od matematyki, tomografia (rezonans magnetyczny) czy też system informacji geograficznej (GIS)" [25].

Celem niniejszej pracy jest między innymi zaprezentowanie pewnych zastosowań jąder reprodukujących w analizie zespolonej oraz analizie funkcjonalnej. Wybór omówionych zastosowań jąder reprodukujących odzwierciedla w znaczny sposób upodobania autorki.

Ponadto zamierzeniem autorki było zaprezentowanie w sposób przystępny dla osób mających szeroką wiedzę i umiejętności z informatyki, otwartego problemu dotyczącego pewnych własności niezmienniczych podprzestrzeni ważonych przestrzeni Bergmana A^2_α, $\alpha > -1$, a dokładniej prawdziwości twierdzenia typu Beurlinga dla $\alpha \in (1, 4)$. Dla α będącego liczbą całkowitą dodatnią większą lub równą od 4 problem ten można sprowadzić do ustalenia liczby zer pewnego wielomianu w kole jednostkowym. W tym celu pomocne okazuje się zastosowanie programów takich jak Maxima, Mathematica czy Matlab (zob. np. [15]).

W podrozdziale drugim zaprezentowano podstawowe pojęcia i fakty z analizy funkcjonalnej, których znajomość jest niezbędna do zrozumienienia teorii przestrzeni Hilberta z jądrem reprodukującym omówionej w kolejnych podrozdziałach.

W podrozdziale trzecim przedstawiamy podstawowe i najważniejsze fakty dotyczące przestrzeni Hilberta z jądrem reprodukującym. Między innymi wyjaśniamy kiedy istnieje jądro reprodukujące dla przestrzeni Hilberta, kiedy dana funkcja jest jądrem reprodukującym dla pewnej przestrzeni Hilberta oraz podajemy wzór na jądro reprodukujące w terminach bazy ortonormalnej dla przestrzeni Hilberta.

W podrozdziale czwartym omawiamy pewne przestrzenie funkcji analitycznych z jądrem reprodukującym. Znajdujemy wzory na jądra reprodukujące dla przestrzeni Hardy'ego H^2 oraz przestrzeni Bergmana A^2 w kole jednostkowym.

W podrozdziale piątym podajemy przykłady zastosowania teorii jąder reprodukujących do rozwiązywania problemów ekstremalnych oraz wykorzystanie własności jąder reprodukujących do wykazania, że dla $\alpha \geqslant 4$ twierdzenie typu Beurlinga jest fałszywe dla ważonej przestrzeni Bergmana A^2_α.

2 Podstawowe pojęcia i fakty

Przypomnimy podstawowe pojęcia i fakty (zob. np. [6, 7, 16, 19]).

2.1 Ośrodkowe przestrzenie Hilberta

Celem tego paragrafu jest m.in. podanie definicji ośrodkowej przestrzeni Hilberta.

Definicja 2.1. *Przestrzenią unitarną* (*nad ciałem liczb zespolonych* $\mathbb{C}$) *nazywamy parę* $(H, \langle\cdot,\cdot\rangle)$, *gdzie* H *jest przestrzenią liniową, zaś* $\langle\cdot,\cdot\rangle : H \times H \to \mathbb{C}$ *jest iloczynem skalarnym, tzn. funkcją spełniającą dla dowolnych* $f, g, h \in H$ *oraz dla dowolnego* $c \in \mathbb{C}$ *następujące warunki:*

1. $\langle f, f\rangle \geqslant 0$ *oraz* $\langle f, f\rangle = 0 \Leftrightarrow f = 0$
2. $\langle f, g\rangle = \overline{\langle g, f\rangle}$
3. $\langle cf, g\rangle = c\langle f, g\rangle$
4. $\langle f + g, h\rangle = \langle f, h\rangle + \langle g, h\rangle$

Z warunków (2) i (3) wynika, że

$$\langle f, cg\rangle = \overline{\langle cg, f\rangle} = \overline{c}\overline{\langle g, f\rangle} = \overline{c}\langle f, g\rangle.$$

Korzystając z własności (2) i (4), otrzymujemy

$$\langle f, g + h\rangle = \overline{\langle g + h, f\rangle} = \overline{\langle g, f\rangle} + \overline{\langle h, f\rangle} = \langle f, g\rangle + \langle f, h\rangle.$$

Lemat 2.1. ***Nierówność Schwarza.*** *Niech* $(H, \langle\cdot,\cdot\rangle)$ *będzie przestrzenią unitarną. Dla dowolnych* $f, g \in H$ *mamy*

$$|\langle f, g\rangle|^2 \leqslant \langle f, f\rangle\langle g, g\rangle.$$

W przestrzeni unitarnej $(H, \langle \cdot, \cdot \rangle)$ definiujemy funkcję

$$f \to \|f\|, \quad f \in H$$

wzorem

$$\|f\| = \sqrt{\langle f, f \rangle}.$$

Z nierówności Schwarza (zob. np. [7]) wynika, że funkcja ta jest normą, tzn. spełnia dla dowolnych $f, g \in H$ oraz dla dowolnego $c \in \mathbb{C}$ następujące warunki:

1. $\|f\| \geqslant 0$.
2. $\|f\| = 0 \Leftrightarrow f = 0$
3. $\|cf\| = |c|\|f\|$
4. $\|f + g\| \leqslant \|f\| + \|g\|$.

Parę $(H, \|\cdot\|)$ nazywamy przestrzenią unormowaną. W każdej przestrzeni unitarnej można więc określić normę. Jednak istnieją przestrzenie unormowane, w których nie można wprowadzić iloczynu skalarnego.

Ciąg elementów $\{f_n\}$ przestrzeni unormowanej $(H, \|\cdot\|)$ jest zbieżny w tej przestrzeni (zbieżny w normie) do elementu $f \in H$,, jeżeli $\lim_{n\to\infty} \|f_n - f\| = 0$. Piszemy wtedy: $\lim_{n\to\infty} f_n = f$.

Mówimy, że ciąg $\{f_n\}$ elementów przestrzeni $(H, \|\cdot\|)$ spełnia warunek Cauchy'ego (jest ciągiem Cauchy'ego), jeżeli

$$\bigwedge_{\varepsilon>0} \bigvee_{n_0 \in \mathbb{N}} \bigwedge_{n,m \geqslant n_0} \|f_n - f_m\| < \varepsilon.$$

W przestrzeni unitarnej definiujemy zbieżność ciągu oraz ciąg Cauchy'ego w taki sam sposób jak w przestrzeni unormowanej.

Każdy ciąg zbieżny jest ciągiem Cauchy'ego (ale nie koniecznie na odwrót). Przestrzeń H, w której każdy ciąg elementów spełniający warunek Cauchy'ego jest w tej przestrzeni zbieżny do pewnego elementu $f_0 \in H$, nazywamy *przestrzenią zupełną.* Zupełną przestrzeń unitarną nazywamy *przetrzenią Hilberta*, natomiast zupełną przestrzeń unormowaną nazywamy *przestrzenia Banacha*. Oczywiście każda

przestrzeń Hilberta jest przestrzenią Banacha, ale nie każda przestrzeń Banacha jest przestrzenią Hilberta.

Definicja 2.2. *Kulą otwartą w przestrzeni unormowanej* $(H, \|\cdot\|)$ *o środku w punkcie* f_0 *i promieniu* $r > 0$ *nazywamy zbiór*

$$B(f_0, r) = \{f \in H : \|f - f_0\| < r\}.$$

Kulą domkniętą w przestrzeni unormowanej $(H, \|\cdot\|)$ *o środku w punkcie* f_0 *i promieniu* $r > 0$ *nazywamy zbiór*

$$\overline{B(f_0, r)} = \{f \in H : \|f - f_0\| \leqslant r\}.$$

Definicja 2.3. *Zbiór* $X \subset H$*, gdzie* $(H, \|\cdot\|)$ *jest przestrzenią unormowaną, nazywamy otwartym, gdy dla każdego elementu* $f_0 \in X$ *istnieje kula* $B(f_0, r)$ *zawarta w* X*. Dopełnienie* X' *zbioru otwartego* X *nazywamy zbiorem domkniętym w* $(H, \|\cdot\|)$*.*

Twierdzenie 2.1. *Zbiór* $X \subset H$ *jest domknięty w* $(H, \|\cdot\|)$ *wtedy i tylko wtedy, gdy dla każdego ciągu elementów* $f_n \in X$ *mamy*

$$\|f_n - f_0\| \underset{n\to\infty}{\longrightarrow} 0,\ f_0 \in H \quad \Rightarrow \quad f_0 \in X.$$

Definicja 2.4. *Domknięciem* $\overline{X}$ *zbioru* $X \subset H$ *w przestrzeni unormowanej* $(H, \|\cdot\|)$ *nazywamy iloczyn* $\bigcap Y$ *wszystkich zbiorów domkniętych* Y *zawierających zbiór* X*.*

Domknięcie dowolnego zbioru jest zbiorem domkniętym. Łatwo można sprawdzić, że domknięta podprzestrzeń przestrzeni Banacha jest przestrzenią zupełną.

Definicja 2.5. *Zbiór* $X \subset H$ *w przestrzeni unormowanej* $(H, \|\cdot\|)$ *nazywamy zwartym, gdy każdy ciąg* $\{f_n\}$ *elementów zbioru* X *zawiera podciąg* $\{f_{n_k}\}$ *zbieżny w* $(H, \|\cdot\|)$ *do pewnego elementu* $f_0 \in X$*.*

Każdy zbiór zwarty w $(H, \|\cdot\|)$ jest domknięty.

Definicja 2.6. *Niech $(H, \|\cdot\|)$ będzie przestrzenią unormowaną. Zbiór $X \subset H$ nazywamy gęstym w zbiorze $Y \subset H$, gdy $Y \subset \overline{X}$. Zbiór $X \subset H$ nazywamy gęstym w przestrzeni $(H, \|\cdot\|)$, albo krócej gęstym, jeśli $\overline{X} = H$. Jeżeli istnieje w H zbiór X przeliczalny (lub skończony) i gęsty w $(H, \|\cdot\|)$, to zbiór X nazywamy ośrodkiem w $(H, \|\cdot\|)$, a przestrzeń H nazywamy przestrzenią ośrodkową.*

Ze względu na twierdzenie Schmidta o ortoganalizacji (zob. np. [16, 8.10]) będziemy rozważać wyłącznie układy oraz bazy ortonormalne.

Definicja 2.7. *Ciąg elementów $\{e_n\}_{n=0}^{\infty}$ przestrzeni Hilberta H nazywany bazą ortonormalną dla H, jeśli posiada następujące własności:*

1. *wektory z $\{e_n\}_{n=0}^{\infty}$ są wzajemnie ortogonalne, tzn. dla dowolnych nieujemnych liczb całkowitych n, m takich że $n \neq m$ mamy $\langle e_n, e_m \rangle = 0$.*
2. *każdy element e_n jest wektorem jednostkowym, czyli $\|e_n\| = 1, n = 0, 1, 2, \ldots$.*
3. *dla dowolnego $f \in H$, szereg $\sum_{n=0}^{\infty} \langle f, e_n \rangle e_n$ jest zbieżny w normie do f.*

Jeżeli ciąg $\{e_n\}$ spełnia (tylko) warunki (1) i (2), to nazywamy go układem ortonormalnym.

Twierdzenie 2.2. (*zob np. [7, 4.13], [6, rozdział 10]*) *Niech $\{e_n\}_{n=0}^{\infty}$ będzie układem ortonormalnym w przestrzeni Hilberta H. Wtedy następujące warunki są równoważne:*

1. *dla dowolnego $f \in H$, szereg $\sum_{n=0}^{\infty} \langle f, e_n \rangle e_n$ jest zbieżny w normie do f.*
2. *układ $\{e_n\}_{n=0}^{\infty}$ jest zupełny, tzn. warunek $\langle f, e_n \rangle = 0$, dla $n = 0, 1, 2, \ldots$ pociąga równość $f = 0$.*
3. *układ $\{e_n\}_{n=0}^{\infty}$ jest zamknięty, tzn. dla każdego $f \in H$ istnieje taki ciąg kombinacji liniowych $f_n = \sum_{k=0}^{n} a_{k,n} e_k$, że $\|f_n - f\| \underset{n \to \infty}{\longrightarrow} 0$.*
4. *dla każdego $f \in H$ ma miejsce równość Parsevala $\|f\|^2 = \sum_{n=0}^{\infty} |\langle f, e_n \rangle|^2$.*

Twierdzenie 2.3. *[16, 12.2] Przestrzeń unormowana $(H, \|\cdot\|)$ z przeliczalną bazą $\{e_n\}$ jest ośrodkowa, przy czym ośrodkiem w $(H, \|\cdot\|)$ jest zbiór wszystkich wymiernych skończonych kombinacji liniowych elementów $e_1, e_2, \ldots$.*

Twierdzenia 2.3 nie można odwrócić, bo nie każda ośrodkowa przestrzeń Banacha ma przeliczalną bazę. Problem został postawiony przez Banacha i uzyskał negatywne rozwiązanie w 1972 roku (P. Enflö). Ważną własnością ośrodkowej przestrzeni Hilberta jest istnienie przeliczalnej (lub skończonej) bazy ortonormalnej (zob. np. [16, II §8 oraz II §12], [28]).

Przykład 2.1. Przykładem ośrodkowej przestrzeni Hilberta jest przestrzeń n-wymiarowych ciągów liczb zepolonych $a = (a_1, \ldots, a_n)$ z iloczynem skalarnym

$$\langle a, b\rangle = \sum_{k=1}^{n} a_k \overline{b_k}, \text{ gdzie } a = (a_1, \ldots, a_n) \text{ oraz } b = (b_1, \ldots, b_n).$$

Bazą ortonormalną w tej przestrzeni jest układ wektorów $\{e_1, e_2, \ldots, e_n\}$, gdzie

$$\begin{array}{c} e_1 = (1, 0, \ldots, 0) \\ e_2 = (0, 1, \ldots, 0) \\ \ldots\ldots\ldots\ldots \\ e_n = (0, 0, \ldots, 1). \end{array} .$$

2.2 Ciągłe operatory liniowe

Niech H_1, H_2 będą przestrzeniami unormowanymi, z normami odpowiednio $\|\cdot\|_{H_1}, \|\cdot\|_{H_2}$. Odwzorowanie $T : H_1 \to H_2$ nazywamy *operatorem liniowym*, jeśli

$$T(af + bg) = aTf + bTg$$

dla dowolnych $f, g \in H_1$ oraz $a, b \in \mathbb{C}$. Jeżeli $H_2 = \mathbb{C}$, to operator liniowy $T : H_1 \to \mathbb{C}$ nazywamy *funkcjonałem liniowym*.

Odwzorowanie $T : H_1 \to H_2$ nazywamy *ograniczonym,* jeżeli istnieje taka liczba $M > 0$, że dla każdego $f \in H_1$ zachodzi nierówność

$$\|Tf\|_{H_2} \leqslant M\|f\|_{H_1}.$$

Twierdzenie 2.4. *[6, twierdzenie 12.1] Dla operatora liniowego $T : H_1 \to H_2$ następujące warunki są równoważne:*

1. *T jest ciągły,*
2. *T jest ciągły w 0,*
3. *T jest ograniczony.*

3 Ogólna teoria przestrzeni Hilberta z jądrem reprodukującym

W całym tym podrozdziale E będzie zbiorem niepustym oraz $(H, \langle \cdot, \cdot \rangle)$ będzie przestrzenią Hilberta złożoną z funkcji $f : E \to \mathbb{C}$. Funkcję $K : E \times E \to \mathbb{C}$ nazywamy jądrem.

Definicja 3.1. *Funkcję $K : E \times E \to \mathbb{C}$ nazywamy jądrem reprodukującym dla przestrzeni H wtedy i tylko wtedy, gdy spełnia następujące dwa warunki:*

1. *dla dowolnego $w \in E$, $K_w(z) = K(z, w)$ należy do H jako funkcja zmiennej z,*
2. *dla dowolnego $w \in E$ oraz dla dowolnego $f \in H$*

$$f(w) = \langle f, K_w \rangle.$$

Z powyższej definicji wynikają natychmiast następujące własności jądra reprodukującego:

1. $K(z, z) \geqslant 0$, dla każdego $z \in E$,
2. $K(z, w) = \overline{K(w, z)}$, dla dowolnych $z, w \in E$,
3. $|K(z, w)|^2 \leqslant K(z, z)K(w, w)$, dla dowolnych $z, w \in E$.

Dowód (1): Dla każdego $z \in E$ mamy

$$K(z,z) = K_z(z) = \langle K_z, K_z \rangle = \|K_z\|^2 \geqslant 0.$$

Dowód (2): Korzystając z własności jądra reprodukującego K, otrzymujemy

$$\begin{aligned}\overline{K(w,z)} = \overline{K_z(w)} = \overline{\langle K_z, K_w \rangle} = \langle K_w, K_z \rangle \\ = K_w(z) = K(z,w).\end{aligned}$$

Dowód (3): Ponieważ $\|K_z\|^2 = K_z(z)$, to korzystając z nierówności Schwarza dostajemy

$$|K_w(z)|^2 = |\langle K_w, K_z \rangle| \leqslant \|K_w\|^2 \|K_z\|^2 = K_z(z) K_w(w) = K(z,z)K(w,w).$$

Podamy teraz prosty przykład przestrzeni Hilberta z jądrem reprodukującym.

Przykład 3.1. Rozpatrzmy dowolną skończenie wymiarową przestrzeń Hilberta $(H, \langle \cdot, \cdot \rangle)$ określoną w E. (Może być to np. przestrzeń podana w przykładzie 2.1.) Niech $\{e_1, e_2 \ldots, e_n\}$ będzie bazą ortonormalną dla H. Zdefiniujmy funkcję K wzorem

$$K(z,w) = \sum_{i=1}^{n} e_i(z) \overline{e_i(w)}.$$

Wtedy dla dowolnego $w \in E$, funkcja

$$K_w(\cdot) = K(\cdot, w) = \sum_{i=1}^{n} \overline{e_i(w)} e_i(\cdot)$$

należy do przestrzeni H. Ponadto dla dowolnej funkcji

$$\varphi(\cdot) = \sum_{i=1}^{n} \lambda_i e_i(\cdot)$$

należącej do H oraz dla dowolnego $w \in E$ mamy

$$\begin{aligned}\langle \varphi, K(\cdot, w)\rangle &= \left\langle \sum_{i=1}^{n} \lambda_i e_i(\cdot), \sum_{j=1}^{n} \overline{e_j(w)} e_j(\cdot) \right\rangle \\ &= \sum_{i=1}^{n}\sum_{j=1}^{n} \lambda_i \overline{\overline{e_j(w)}} \langle e_i, e_j\rangle \\ &= \sum_{i=1}^{n} \lambda_i e_i(w) = \varphi(w).\end{aligned}$$

Zatem każda skończenie wymiarowa przestrzeń Hilberta ma jądro reprodukujące. Ponadto, co za chwilę wykażemy, jest ono jedyne.

Twierdzenie 3.1. *Jądro reprodukujące K jest jednoznacznie wyznaczone przez przestrzeń Hilberta $(H, \langle\cdot,\cdot\rangle)$.*

Dowód. Z własności jąder reprodukujących K i K^*, patrz punkt 2 definicji 3.1, wynika, że dla dowolnego $w \in E$

$$\begin{aligned}\|K_w - K_w^*\|^2 &= \langle K_w - K_w^*, K_w - K_w^*\rangle = \langle K_w - K_w^*, K_w\rangle - \langle K_w - K_w^*, K_w^*\rangle \\ &= (K_w - K_w^*)(w) - (K_w - K_w^*)(w) = 0.\end{aligned}$$

□

Skoro już wiemy, że jeśli istnieje jądro reprodukujące dla przestrzeni Hilberta $(H, \langle\cdot,\cdot\rangle)$, to jest ono jedyne, to naturalnym nasuwającym się od razu pytaniem jest:

Kiedy istnieje jądro reprodukujące dla przestrzeni Hilberta $(H, \langle\cdot,\cdot\rangle)$?

Twierdzenie 3.2. *Niech H oznacza przestrzeń Hilberta składającą się z funkcji f określonych w zbiorze E. Jądro reprodukujące K istnieje dla H wtedy i tylko wtedy, gdy dla każdego $w \in E$, odwzorowanie*

$$\phi_w : f \to f(w), \quad f \in H$$

jest liniowym i ograniczonym funkcjonałem na H.

Dowód. Załóżmy najpierw, że istnieje jądro reprodukujące K dla przestrzeni H. Korzystając z własności jądra reprodukującego, otrzymujemy

$$|f(w)| = |\langle f, K_w\rangle| \leqslant \|f\|\|K_w\| = \|f\|\sqrt{\langle K_w, K_w\rangle} = \sqrt{K_w(w)}\|f\|.$$

Zatem ϕ_w jest liniowym i ograniczonym funkcjonałem na H.

Z drugiej strony, jeśli ϕ_w jest liniowym i ograniczonym funkcjonałem liniowym, to z twierdzenia Riesza o reprezentacji funkcjonału w przestrzeni Hilberta [7, str. 13] wynika istnienie dokładnie jednej funkcji g_w należącej do H i takiej, że

$$f(w) = \langle f, g_w\rangle.$$

Zatem funkcja $K : E \times E \to \mathbb{C}$ określona przez $K(z, w) \stackrel{\text{def}}{=} g_w(z)$ jest jądrem reprodukującym dla H. □

Z pierwszej części dowodu twierdzenia 3.2 wynika następujący wniosek.

Wniosek 3.1. Zbieżność ciągu funkcyjnego $\{f_n\}$ w normie w przestrzeni Hilberta z jądrem reprodukującym implikuje jego punktową zbieżność, czyli dla każdego punktu $z \in E$ mamy

$$\lim_{n\to\infty} f_n(z) = f(z).$$

Ponadto, jeśli $A \subset E$ oraz $\sup_{w\in A}\|K_w\| < \infty$, to zbieżność ciągu $\{f_n\}$ jest jednostajna na A.

Dowód. Dla każdego $w \in E$ i każdego $n \in \mathbb{N}$

$$|f_n(w) - f(w)| = |\langle f_n - f, K_w\rangle| \leqslant \|f_n - f\|\|K_w\|.$$

Stąd

$$\sup_{w\in A}|f_n(w) - f(w)| \leqslant \sup_{w\in A}\|K_w\|\|f_n - f\|.$$

□

Ta miła własność pokazuje dlaczego przestrzenie Hilberta z jądrem reprodukującym są używane w zastosowaniach matematyki.

Naszym następnym celem będzie udzielenie odpowiedzi na pytanie:

Kiedy funkcja $K : E \times E \to \mathbb{C}$ jest jądrem reprodukującym?

Zanim wypowiemy stosowne twierdzenie wprowadzimy najpierw definicję dodatnio określonych jąder.

Definicja 3.2. *Jądro $K : E \times E \to \mathbb{C}$ jest nazywane dodatnio określonym, jeśli dla każdej dodatniej liczby całkowitej n oraz dla dowolnych skończonych ciągów punktów $z_1, \cdots, z_n \in E$ oraz $c_1, \cdots, c_n \in \mathbb{C}$ spełniony jest warunek*

$$\sum_{k=1}^{n}\sum_{j=1}^{n} c_k \overline{c_j} K(z_k, z_j) \geqslant 0.$$

Warto zauważyć, że dodatnia określoność jądra jest równoważna z dodatnią określonością macierzy

$$[K(z_k, z_j)]_{n\times n} = \begin{bmatrix} K(z_1,z_1) & K(z_1,z_2) & \cdots & K(z_1,z_n) \\ K(z_2,z_1) & K(z_2,z_2) & \cdots & K(z_2,z_n) \\ \cdots & \cdots & \cdots & \cdots \\ K(z_n,z_1) & K(z_n,z_2) & \cdots & K(z_n,z_n) \end{bmatrix},$$

dla dowolnej dodatniej liczby całkowitej n i dowolnych punktów $z_1, \cdots, z_n$ ze zbioru E [20].

Podamy teraz twierdzenie, które podaje charakteryzację jąder reprodukujących dla przestrzeni Hilberta H w terminach dodatnio określonych macierzy [21].

Twierdzenie 3.3. *Niech H oznacza przestrzeń Hilberta składającą się z funkcji f określonych w zbiorze E. Funkcja $K : E \times E \to \mathbb{C}$ jest jądrem reprodukującym dla przestrzeni Hilberta H wtedy i tylko wtedy, gdy dla dowolnej dodatniej liczby całkowitej n i dowolnych punktów $z_1, \cdots, z_n$ ze zbioru E macierz $[K(z_k, z_j)]_{n\times n}$ jest dodatnio określona.*

Dowód. Załóżmy, że $K : E \times E \to \mathbb{C}$ jest jądrem reprodukującym dla przestrzeni Hilberta H. Niech n będzie dowolną dodatnią liczbą całkowitą. Dla dowolnego skończonego ciągu punktów $z_1, \cdots, z_n \in E$ oraz $c_1, \cdots, c_n \in \mathbb{C}$ mamy

$$\begin{aligned} 0 &\leqslant \left\| \sum_{j=1}^{n} c_j K(\cdot, z_j) \right\|_H^2 \\ &= \sum_{j=1}^{n} \sum_{k=1}^{n} c_j \overline{c_k} \langle K(\cdot, z_j), K(\cdot, z_k) \rangle \\ &= \sum_{j=1}^{n} \sum_{k=1}^{n} c_j \overline{c_k} K(z_k, z_j) \end{aligned}$$

Zatem macierz $[K(z_k, z_j)]_{n \times n}$ jest dodatnio określona.

Dowód warunku dostatecznego na to, aby funkcja K była jądrem reprodukującym dla przestrzeni Hilberta H można znaleźć w [4] lub [21]. □

Przykładem jądra dodatnio określonego jest funkcja

$$K : E \times E \ni (z, w) \to f(z)\overline{f(w)} \in \mathbb{C},$$

gdzie f jest dowolną funkcją na E [24].

Poniższe twierdzenie podaje prosty sposób znalezienia jawnego wzoru na jądro reprodukujące w ośrodkowej przestrzeni Hilberta z jądrem reprodukującym.

Twierdzenie 3.4. *Niech H będzie ośrodkową przestrzenią Hilberta składającą się z funkcji o wartościach zespolonych określonych w zbiorze E z jądrem reprodukującym K. Dla dowolnej bazy ortonormalnej $\{e_n\}_{n=0}^{\infty}$ dla przestrzeni H mamy*

$$\bigwedge_{w \in E} K_w(\cdot) = K(\cdot, w) = \sum_{n=0}^{\infty} e_n(\cdot)\overline{e_n(w)}, \tag{3.1}$$

gdzie powyższy szereg jest zbieżny w H.

Odwrotnie, jeśli warunek (3.1) jest spełniony dla zbioru ortonormalnego $\{e_n\}_{n=0}^{\infty}$, to zbiór ten jest zupełny oraz przestrzeń Hilberta H jest ośrodkowa. Po-

nadto warunek (3.1) implikuje zbieżność bezwzględną szeregu

$$K(z,w) = \sum_{n=0}^{\infty} e_n(z)\overline{e_n(w)}$$

w $E \times E$.

Dowód. Ustalmy $w \in E$. Funkcja K_w jako element przestrzeni H ma rozwinięcie w szereg Fouriera

$$\sum_{n=0}^{\infty} c_n e_n(\cdot),$$

gdzie

$$c_n = \langle K_w, e_n\rangle = \overline{\langle e_n, K_w\rangle} = \overline{e_n(w)}.$$

Odwrotnie, jeśli warunek (3.1) jest spełniony dla układu ortonormalnego $\{e_n\}_{n=0}^{\infty}$, to dla dowolnej funkcji $\varphi \in H$ oraz dla każdego $w \in E$ mamy

$$\varphi(w) = \langle \varphi, K_w\rangle = \sum_{n=0}^{\infty} \langle \varphi, e_n\rangle e_n(w).$$

Stąd wynika, że układ $\{e_n\}_{n=0}^{\infty}$, jest zupełny oraz na mocy twierdzenia 2.3 przestrzeń H jest ośrodkowa.

Ustalmy $z, w \in E$. Korzystając z warunku (3.1), otrzymujemy

$$K(z,w) = \langle K_w, K_z\rangle = \left\langle \sum_{n=0}^{\infty} e_n(\cdot)\overline{e_n(w)}, \sum_{m=0}^{\infty} e_m(\cdot)\overline{e_m(z)} \right\rangle = \sum_{n=0}^{\infty} e_n(z)\overline{e_n(w)}.$$

Stąd oraz z nierówności Höldera dla sum (zob. np. [16]) otrzymujemy

$$\sum_{n=0}^{\infty} |e_n(z)\overline{e_n(w)}| \leqslant \left(\sum_{n=0}^{\infty} |e_n(z)|^2\right)^{\frac{1}{2}} \left(\sum_{n=0}^{\infty} |e_n(w)|^2\right)^{\frac{1}{2}} = K(z,z)K(w,w) < \infty.$$

□

4 Przestrzenie funkcji analitycznych z jądrem reprodukującym

Zanim podamy przykłady przestrzeni Hilberta z jądrem reprodukującym, wprowadzimy pojęcie izometrycznego izomorfizmu przestrzeni unitarnych $(H_1, \langle\cdot,\cdot\rangle_{H_1})$ i $(H_2, \langle\cdot,\cdot\rangle_{H_2})$.

Definicja 4.1. *Niech $(H_1, \langle\cdot,\cdot\rangle_{H_1})$ i $(H_2, \langle\cdot,\cdot\rangle_{H_2})$ będą przestrzeniami unitarnymi. Przestrzenie H_1 i H_2 nazywamy izometrycznie izomorficznymi, jeśli istnieje liniowe i różnowartościowe odworowanie L przestrzeni H_1 na przestrzeń H_2 takie, że*

$$\langle Lf, Lg\rangle_{H_2} = \langle f, g\rangle_{H_1},$$

dla każdego $f, g \in H_1$. Odwzorowanie L nazywamy izometryczym izomorfizmem H_1 na H_2.

Zauważmy, że izometryczny izomorfizm L pomiędzy dwoma przestrzeniami unitarnymi H_1 i H_2 przekształca ciągi Cauchy'ego $\{f_n\}_{n=0}^{\infty}$ w H_1 na ciągi Cauchy'ego $\{Lf_n\}_{n=0}^{\infty}$ w H_2. Stąd izometryczny izomorfizm zachowuje zupełność.

4.1 Przestrzeń Hardy'ego H^2 w kole jednostkowym

Niech $\mathbb{D}$ oznacza koło jednostkowe na płaszczyźnie zespolonej $\mathbb{C}$, czyli $\mathbb{D} = \{z \in \mathbb{C} : |z| < 1\}$. Przestrzeń Hardy'ego H^2, definiujemy jako przestrzeń funkcji $f = \sum_{n=0}^{\infty} a_n z^n,\ z \in \mathbb{D}$, takich, że

$$\|f\|_{H^2}^2 = \sum_{n=0}^{\infty} |a_n|^2 < \infty \tag{4.1}$$

z iloczynem skalarnym

$$\langle f, g\rangle_{H^2} = \sum_{n=0}^{\infty} a_n \overline{b_n}, \quad f = \sum_{n=0}^{\infty} a_n z^n, \quad g = \sum_{n=0}^{\infty} b_n z^n.$$

Zauważmy, że warunek (4.1) zapewnia zbieżność szeregu potęgowego $\sum_{n=0}^{\infty} a_n z^n$ na $\mathbb{D}$. Istotnie, korzystając z nierówności Höldera dla sum, otrzymujemy

$$\left|\sum_{n=0}^{\infty} a_n z^n\right| \leqslant \sum_{n=0}^{\infty} |a_n||z^n| \leqslant \left(\sum_{n=0}^{\infty} |a_n|^2\right)^{\frac{1}{2}} \left(\sum_{n=0}^{\infty} |z^{2n}|\right)^{\frac{1}{2}}$$
$$= \|f\|_{H^2} \frac{1}{\sqrt{1-|z|^2}}.$$

W celu uzasadnienia, że przestrzeń Hardy'ego H^2 jest przestrzenią Hilberta, rozpatrzmy liniowe odwzorowanie $L : H^2 \to l^2(\mathbb{Z}^+)$ określone wzorem $L(f) = (a_0, a_1, \ldots)$, gdzie $l^2(\mathbb{Z}^+)$, oznacza przestrzeń składającą się z nieskończonych ciągów $\{a_n\}_{n=0}^{\infty}$ liczb rzeczywistych lub zespolonych, dla których $\sum_{n=0}^{\infty} |a_n|^2 < \infty$, z iloczynem skalarnym

$$\langle a, b\rangle = \sum_{n=0}^{\infty} a_n \overline{b_n}, \text{ gdzie } a = \{a_n\}_{n=0}^{\infty}, \quad b = \{b_n\}_{n=0}^{\infty}.$$

Ponieważ $l^2(\mathbb{Z}^+)$ jest przestrzenią Hilberta [16, 6.7], a $L : H^2 \to l^2(\mathbb{Z}^+)$ jest izometrycznym izomorfizmem, to przestrzeń Hardy'ego H^2 jest przestrzenią Hilberta.

Ponieważ

$$|f(z)| \leqslant \|f\|_{H^2} \frac{1}{\sqrt{1-|z|^2}},$$

to z twierdzenia 3.2 wynika, że H^2 jest przestrzenią Hilberta z jądrem reprodukującym.

Z definicji iloczynu skalarnego w przestrzeni H^2 wynika, że funkcje $z^n, n = 0, 1, 2\ldots$ stanowią układ ortonormalny w tej przetrzeni. Ponieważ

$$\|f\|_{H^2}^2 = \sum_{n=0}^{\infty} |\langle f, e_n\rangle|^2, \quad f \in H^2,$$

to na mocy twierdzenia 2.2 $\{z^n\}_{n=0}^{\infty}$ jest bazą ortonormalną dla H^2. Stąd oraz z twierdzenia 2.3 wynika, że przestrzeń H^2 jest ośrodkowa. Zatem jądro reprodukujące dla H^2 jest dane wzorem

$$K^{H^2}(z, w) = \sum_{n=0}^{\infty} z^n \overline{w^n} = \frac{1}{1 - z\overline{w}}, \quad z \in \mathbb{D}.$$

4.2 Przestrzeń Bergmana A^2 w kole jednostkowym $\mathbb{D}$

Oznaczmy przez $L^2(\mathbb{D})$ przestrzeń wszystkich funkcji f określonych w kole jednostkowym $\mathbb{D}$ o wartościach w $\mathbb{C}$ takich, że

$$\frac{1}{\pi} \iint\limits_{\{x+iy \in \mathbb{D}\}} |f(x+iy)|^2 dxdy < \infty$$

z iloczynem skalarnym

$$\langle f, g \rangle = \frac{1}{\pi} \iint\limits_{\{x+iy \in \mathbb{D}\}} f(x+iy)\overline{g(x+iy)}dxdy, \quad f, g \in L^2(\mathbb{D}).$$

Przestrzeń Bergmana A^2 w kole jednostkowym $\mathbb{D}$ definiujemy jako podprzestrzeń przestrzeni $L^2(\mathbb{D})$ składającą się z funkcji analitycznych w kole jednostkowym $\mathbb{D}$. Dla $f, g \in A^2$ przyjmujemy

$$\langle f, g \rangle_{A^2} = \frac{1}{\pi} \iint\limits_{\{x+iy \in \mathbb{D}\}} f(x+iy)\overline{g(x+iy)}dxdy.$$

Korzystając ze wzoru na zamianę zmiennych w całce podwójnej na współrzędne biegunowe (zob. np. [13, str. 117-118]), otrzymujemy

$$\langle f, g \rangle_{A^2} = \frac{1}{\pi} \int_0^1 \int_0^{2\pi} f(re^{i\theta})\overline{g(re^{i\theta})} r d\theta dr.$$

Twierdzenie 4.1. *Dla każdego $z \in \mathbb{D}$ odwzorowanie*

$$\phi_z : f \mapsto f(z), \quad f \in A^2$$

jest liniowym i ograniczonym funkcjonałem w A^2.

Dowód. Niech $z \in \mathbb{D}$ będzie dowolnie ustalonym punktem oraz niech $\delta(z) = 1 - |z|$. Wtedy koło $B(z, \delta(z)) = \{\xi \in \mathbb{C} : |\xi - z| < \delta(z)\}$ zawiera się w $\mathbb{D}$. Zatem z twierdzenia o wartości średniej [14, str. 106] zastosowanego do funkcji f^2 wynika, że dla $0 \leqslant r < \delta(z)$ mamy

$$|f(z)|^2 \leqslant \frac{1}{2\pi} \int_0^{2\pi} |f(z + re^{i\theta})|^2 d\theta.$$

Stąd

$$\begin{aligned} 2|f(z)|^2 \int_0^{\delta(z)} r dr &\leqslant \frac{1}{\pi} \int_0^{\delta(z)} \int_0^{2\pi} |f(z+re^{i\theta})|^2 d\theta r dr \\ &= \frac{1}{\pi} \iint\limits_{\{x+iy \in B(z,\delta(z))\}} |f(x+iy)|^2 dx dy \\ &\leqslant \|f\|_{A^2}^2, \end{aligned}$$

co daje

$$|f(z)| \leqslant \frac{\|f\|_{A^2}}{\delta(z)} = \frac{1}{1-|z|} \|f\|_{A^2}.$$

□

Z dowodu twierdzenia 4.1 wynika następujący wniosek.

Wniosek 4.1. Niech $A \subset \mathbb{D}$ będzie zwarty. Istnieje wtedy taka stała $C(A)$, że dla $f \in A^2$

$$\sup_{z \in A} |f(z)| \leqslant C(A) \|f\|_{A^2}.$$

Dowód. Niech $A \subset \mathbb{D}$ będzie zwarty. Istnieje wtedy takie $0 < R < 1$, że $A \subset \overline{B(0,R)} \subset \mathbb{D}$, gdzie $\overline{B(0,R)} = \{z \in \mathbb{C} : |z| \leqslant R\}$. Ponieważ dla każdego $z \in \mathbb{D}$

$$|f(z)| \leqslant \frac{1}{1-|z|} \|f\|_{A^2},$$

to

$$\sup_{z \in A} |f(z)| \leqslant \sup_{z \in \overline{B(0,R)}} |f(z)| \leqslant \sup_{z \in \overline{B(0,R)}} \frac{1}{1-|z|} \|f\|_{A^2} = \frac{1}{1-R} \|f\|_{A^2}.$$

□

Twierdzenie 4.2. *Przestrzeń Bergmana A^2 jest domkniętą podprzestrzenią przestrzeni $L^2(\mathbb{D})$.*

Dowód. Niech $\{f_n\}$ będzie ciągiem elementów przestrzeni A^2 zbieżnym w normie do $f \in L^2(\mathbb{D})$. W szczególności $\{f_n\}$ jest ciągiem Cauchy'ego w A^2. Korzystając z wniosku 4.1 otrzymujemy

$$\sup_{z\in A}|f_n(z)-f_m(z)| \leqslant \frac{1}{1-R}\|f_n-f_m\|_{A^2},$$

gdzie A jest dowolnym zwartym podzbiorem koła jednostkowego $\mathbb{D}$. Stąd oraz z kryterium Cauchy'ego (zob. np. [14, str.46]) wynika, że $\{f_n\}$ jest jednostajnie zbieżny na każdym zwartym pozdbiorze A koła jednostkowego $\mathbb{D}$. Z twierdzenia Weierstrassa (zob. np. [14, str.109]) otrzymujemy, że f jest funkcją analityczną w $\mathbb{D}$. Zatem $f \in A^2$. □

Ponieważ $L^2(\mathbb{D})$ jest przestrzenią Hilberta (zob. np. [18, twierdzenie 3.11]) oraz A^2 jest domkniętą podprzestrzenią przestrzeni $L^2(\mathbb{D})$, to przestrzeń A^2 jest również przestrzenią Hilberta.

Łatwo można sprawdzić, że funkcje $e_n(z) = \sqrt{n+1}z^n, n = 0,1,2,\ldots$ stanowią układ ortonormalny w przestrzeni A^2. Istotnie, jeśli $n = m$, to

$$\begin{aligned}\langle e_n, e_m\rangle_{A^2} &= \sqrt{(n+1)(m+1)}\frac{1}{\pi}\int_0^1\int_0^{2\pi} r^{n+m+1}e^{i\theta(n-m)}drd\theta\\ &= 2(n+1)\int_0^1 r^{2n+1}dr = 1.\end{aligned}$$

Jeżeli $n \neq m$, to oczywiście $\langle e_n, e_m\rangle = 0$. Ponadto fakt, że stanowią one również bazę w tej przestrzeni jest równoważny tożsamości Parsevala

$$\|f\|^2_{A^2} = \sum_{n=0}^{\infty}|\langle f, e_n\rangle|^2, \quad f \in A^2.$$

Załóżmy, że

$$f(z) = \sum_{n=0}^{\infty} a_n z^n \in A^2.$$

Wykażemy, że

$$\|f\|^2_{A^2} = \sum_{n=0}^{\infty}\frac{|a_n|^2}{n+1}.$$

Niech $S_n(z) = \sum_{k=0}^{n} a_k z^k$. Wtedy dla $\delta < 1$ mamy

$$\begin{aligned}\frac{1}{\pi}\int_0^{\delta}\int_0^{2\pi} |S_n(re^{i\theta})|^2 r dr d\theta &= \frac{1}{\pi}\int_0^{\delta}\int_0^{2\pi} \left(\sum_{k=0}^{n} a_k r^k e^{ik\theta}\right)\left(\sum_{l=0}^{n} \overline{a_l r^l e^{il\theta}}\right) r dr d\theta \\ &= 2\sum_{k=0}^{n} |a_k|^2 \int_0^{\delta} r^{2k+1} dr \\ &= \sum_{k=0}^{n} \frac{|a_k|^2 \delta^{2k+2}}{k+1}.\end{aligned}$$

Ponieważ $S_n(z) \to f(z)$ jednostajnie w kole $|z| \leqslant \delta < 1$, to

$$\frac{1}{\pi}\int_0^{\delta}\int_0^{2\pi} |f(re^{i\theta})|^2 r dr d\theta = \sum_{k=0}^{\infty} \frac{|a_k|^2 \delta^{2k+2}}{k+1}.$$

Łatwo można wykazać, że granica $\lim_{\delta\to 1^-} \sum_{k=0}^{\infty} \frac{|a_k|^2\delta^{2k+2}}{k+1}$ istnieje (zob. np. rozwiązanie zadania 3.3.18 [12]), Przechodząc do granicy z $\delta \to 1^-$ dostajemy żądaną nierówność. Ponieważ $\langle f, e_n\rangle_{A^2} = \frac{a_n}{\sqrt{n+1}}$, to tożsamość Parsevala została wykazana.

Na mocy twierdzenia 3.4 jądro reprodukujące w przestrzeni A^2 dane jest wzorem

$$K^{A^2}(z,w) = \sum_{n=0}^{\infty} (n+1) z^n \overline{w^n} = \frac{1}{(1-z\overline{w})^2}.$$

4.3 Ważona przestrzeń Bergmana $A^2_\alpha, \alpha > -1$

Dla $-1 < \alpha < \infty$ ważona przestrzeń Bergmana A^2_α jest przestrzenią funkcji f holomorficznych w kole jednostkowym $\mathbb{D}$ takich, że

$$\frac{1}{\pi}\int_0^{1}\int_0^{2\pi} |f(re^{i\theta})|^2 (1-r^2)^{\alpha} r d\theta dr < \infty,$$

Przestrzeń A^2_α jest przestrzenią Hilberta z iloczynem skalarnym

$$\langle f, g\rangle_{A^2_\alpha} = \frac{1}{\pi}\int_0^{1}\int_0^{2\pi} f(re^{i\theta})\overline{g(re^{i\theta})}(1-r^2)^{\alpha} r d\theta dr.$$

Ponadto, jeśli

$$f(z) = \sum_{n=0}^{\infty} a_n z^n \qquad \text{oraz} \qquad g(z) = \sum_{n=0}^{\infty} b_n z^n,$$

to

$$\langle f, g \rangle_{A^2_\alpha} = \sum_{n=0}^{\infty} a_n \overline{b_n} \frac{n!\Gamma(\alpha+2)}{\Gamma(n+\alpha+2)},$$

gdzie Γ oznacza funkcję gamma Eulera. W przypadku gdy $\alpha = 0$, otrzymujemy klasyczną przestrzeń Bergmana ($A^2_0 = A^2$.)

W podobny sposób, jak w przypadku klasycznej przestrzeni Bergmana można wykazać, że ważona przestrzeń Bergmana jest przestrzenią z jądrem reprodukującym [8]. Ponadto mamy

$$K^{A^2_\alpha}(z, w) = \frac{1}{(1 - z\overline{w})^{2+\alpha}}.$$

5 Zastosowania jąder reprodukujących do rozwiązywania pewnych problemów ekstremalnych

5.1 Podprzestrzenie niezmiennicze oraz funkcje ekstremalne dla A^2_α

Niech H oznacza przestrzeń Hilberta funkcji holomorficznych w kole jednostkowym $\mathbb{D}$. Domkniętą podprzestrzeń I przestrzeni H nazywamy niezmienniczą, jeśli jest niezmiennicza ze względu na operator mnożenia przez z, to znaczy prawdziwa jest implikacja

$$f(z) \in I \implies zf(z) \in I. \tag{5.1}$$

Przyjmując oznaczenie $zI = \{zf : f \in I\}$ warunek (5.1) możemy zapisać

$$zI \subset I.$$

Najmniejszą niezmienniczą podprzestrzeń zawierającą zbiór $E \subset H$ będziemy oznaczać symbolem $[E]$ i nazywać *niezmienniczą podprzestrzenią generowaną przez* E.

Niezmienniczą podprzestrzeń generowaną przez funkcję $f \in H$, będziemy oznaczać przez $[f]$. Jest to domknięcie przestrzeni linowej, której elementami są iloczyny wielomianu przez funkcję f.

Dla domkniętych podprzestrzeni $M, N \subset H$, takich że $M \subset N$ przyjmujemy

$$N \ominus M = N \cap M^{\perp},$$

gdzie

$$M^{\perp} = \{g \in H : \bigwedge_{f \in M} \langle f, g \rangle = 0\}.$$

W 1949 roku A. Beurling wykazał, że każda niezmiennicza podprzestrzeń $I \subset H^2, I \neq \{0\}$ jest postaci

$$I = \varphi H^2 = [\varphi],$$

gdzie φ jest funkcją wewnętrzną dla H^2, tzn. $\varphi \in H^\infty$ (φ jest funkcją analityczną w kole jednostkowym $\mathbb{D}$ oraz $\sup_{z \in \mathbb{D}} |\varphi(z)|) < \infty$) i jej wartości brzegowe spełniają warunek $|\varphi(e^{i\theta})| = 1$ prawie wszędzie [5].

Ponieważ $I \ominus zI = \mathbb{C}\varphi$, to twierdzenie Beurlinga oznacza, że $I = [I \ominus zI]$.

Podprzestrzenie niezmiennicze ważonej przestrzeni Bergmana mają dużo bardziej złożoną strukturę niż w przypadku przestrzeni Hardy'ego. Dla przestrzeni Bergmana można skonstruować podprzestrzenie niezmiennicze I takie, że wymiar $\dim(I \ominus zI) = n$, gdzie n jest dowolną liczbą naturalną lub $\dim(I \ominus zI) = \infty$ ([2, 10] oraz [11, str. 176–179]). Wymiar przestrzeni $I \ominus zI$ nazywamy *indeksem* I.

Przykładem niezmienniczej podprzestrzeni I_C, gdzie $C = \{c_n\} \subset \mathbb{D}$, jest przestrzeń wszystkich funkcji z A^2_α zerujących się na danym ciągu C. Indeks takiej przestrzeni wynosi 1.

Będziemy dodatkowo zakładać, że $c_n \neq 0$. Wtedy istnieje dokładnie jedno rozwiązanie następującego problemu ekstremalnego

$$\sup\{|f(0)| : f \in I_C, \|f\|_\alpha \leqslant 1\}. \tag{5.2}$$

Rozwiązanie problemu ekstremalnego (5.2) nazywamy *funkcją ekstremalną* oraz będziemy oznaczać je przez $G^{\alpha}_{I_C}$. Funkcja ekstremalna $G^{\alpha}_{I_C}$ jest określona wzorem

$$G^{\alpha}_{I_C}(z) = \frac{K^{\alpha}_{I_C}(z,0)}{\sqrt{K^{\alpha}_{I_C}(0,0)}},$$

gdzie $K^{\alpha}_{I_C}$ jest jądrem reprodukującym dla przestrzeni I_C, tzn. $K^{\alpha}_{I_C}(\in I_C$ oraz dla każdego $f \in I_C$

$$f(z) = \langle f, K^{\alpha}_{I_C}(\cdot, z)\rangle_{\alpha}, \qquad z \in \mathbb{D}.$$

Istotnie, ponieważ

$$K^{\alpha}_{I_C}(0,0) = \langle K^{\alpha}_{I_C}(\cdot,0), K^{\alpha}_{I_C}(\cdot,0)\rangle_{\alpha} = \|K^{\alpha}_{I_C}(\cdot,0)\|^2_{\alpha},$$

to dla dowolnego $f \in I_C$ mamy

$$|f(0)| = |\langle f, K^{\alpha}_{I_C}(\cdot,0)\rangle_{\alpha}| \leqslant \|f\|_{\alpha}\|K^{\alpha}_{I_C}(\cdot,0)\|_{\alpha} = \|f\|_{\alpha}\sqrt{K^{\alpha}_{I_C}(0,0)}.$$

Łatwo teraz sprawdzić, że dla $f = G^{\alpha}_{I_C}$ w powyższej nierówności zachodzi równość.

Niech $C = \{c\}$, gdzie $c \neq 0$ jest dowolnie ustalonym punktem $\mathbb{D}$, i niech G^{α}_{c} oznacza rozwiązanie powyższego problemu ekstremalnego w przypadku gdy $C = \{c\}$. H. Hedenmalm, K. Zhu w 1992 roku otrzymali następujący wzór na funkcję ekstremalną G^{α}_{c}

$$G^{\alpha}_{c}(z) = \frac{1 - \left(\frac{1-|c|^2}{1-\bar{c}z}\right)^{\alpha+2}}{\sqrt{1-(1-|c|^2)^{\alpha+2}}}.$$

W roku 2017 wzór ten został uogólniony dla podprzestrzeni niezmienniczych $I_{k\times c}$ ważonej przestrzeni Bergmana mających zero w punkcie $z = c, c \neq 0$ o wielokrotności co najmniej k [17].

5.2 Twierdzenie typu Beurlinga

Przypomnijmy, że twierdzenie Beurlinga dla H^2 oznacza w szczególności, że $I = [I \ominus zI]$.

Mówimy, że **twierdzenie typu Beurlinga** jest spełnione dla przestrzeni Hilberta H, jeśli

$$I = [I \ominus zI],$$

dla dowolnej niezmienniczej podprzestrzeni $I \subset H$.

W 1996 roku A. Aleman, S. Richter, C. Sundberg wykazali, że twierdzenie typu Beurlinga jest spełnione dla przestrzeni $A^2 = A^2_0$ [1].

Następnie w latach 2001-2003 ukazały się prace S. Shimorina, w których autor wykazał równość $I = [I \ominus zI]$ dla dowolnej niezmienniczej podprzestrzeni $I \subset A^2_\alpha$ dla $\alpha \in (-1, 1]$ [22, 23].

W 1992 roku H. Hedenmalm i K. Zhu wykazali, że dla każdego $c \in \mathbb{D}\backslash\{0\}$ funkcja ekstremalna G^α_c ma dokładnie jedno zero w $\mathbb{D}$ ($z = c$) wtedy i tylko wtedy, gdy $-1 < \alpha \leqslant 4$ [9]. Ponieważ indeks $I_{\{c\}}$ wynosi 1, więc gdyby twierdzenie typu Beurlinga było prawdziwe dla $\alpha > 4$, to mielibyśmy

$$I_{\{c\}} = [I_{\{c\}} \ominus zI_{\{c\}}] = [G^\alpha_c].$$

Jednak ostatnia równość jest nieprawdziwa, gdyż $I_{\{c\}}$ zawiera również funkcje, które zerują się tylko w jednym punkcie ($z = c$), natomiast wszystkie funkcje z przestrzeni $[G^\alpha_c]$ zerują się w co najmniej w dwóch punktach.

W pracy [17] została zbadana liczba zer funkcji ekstremalnej $G^\alpha_{I_{k\times c}}$ dla $\alpha = 2, 3, 4$. Funkcja ta nie ma dodatkowych zer w kole jednostkowym w przypadku gdy $\alpha = 2$ oraz gdy $\alpha = 3$. Natomiast, jeśli $\alpha = 4$ oraz c jest wystarczająco blisko okręgu jednostkowego, to funkcja $G^\alpha_{I_{k\times c}}$ oprócz zera rzędu k w punkcie c ma dwa dodatkowe zera w kole jednostkowym. Stąd, stosując podobne rozumowanie jak dla $\alpha > 4$, otrzymujemy, że twierdzenie typu Beurlinga nie jest prawdziwe dla A^2_4.

5.3 Problem otwarty

Do tej pory nie zostało rozstrzygnięte czy twierdzenie typu Beurlinga jest spełnione dla ważonej przestrzeni Bergmana A^2_α w przypadku gdy $\alpha \in (1, 4)$. Ze względu na pewne własności ważonych przestrzeni Bergmana A^2_α dla $\alpha > 1$ można spodziewać się, że twierdzenie to jest fałszywe dla takich przestrzeni [23].

Podziękowania

Serdeczne podziękowania składam Panu prof. dr hab. Witoldowi Rzymowskiemu za wnikliwe zapoznanie się z pracą oraz za wiele cennych i trafnych uwag dotyczących strony redakcyjnej, dzięki którym praca stała się bardziej precyzyjna i dokładna.

Literatura

[1] A. Aleman, S. Richter, C. Sundberg, Beurling's Theorem for the Bergman space, *Acta Math.* 177, 1996, 275–310.

[2] C. Apostol, H. Bercovici, C. Foiaş, C. Pearcy, Invariant subspaces, dilation theory, and the structure of the predual of a dual algebra, I, *J. Funct. Anal.* 63, 1985, 369–404.

[3] N. Aronszajn, Theory of reproducing kernels, *Trans. Amer. Math. Soc.* 68, 1950, 337–404.

[4] A. Berlinet, C. Thomas-Agnan, *Reproducing kernel Hilbert spaces in probability and statistic,* Kluwer Academic Publishers, Boston, Dordrecht, London, 2004.

[5] A. Beurling, On two problems concerning linear transformations in Hilbert space, *Acta Math.* 81, 1949, 239–255.

[6] A. Bobrowski, *Analiza funkcjonalna jeden i pół. Szkic o zupełności,* Politechnika Lubelska, Lublin, 2015.

[7] J. B. Conway, *A Course in Functional Analysis*, Springer-Verlag, New York, 1985.

[8] P. Duren, A. Schuster, *Bergman spaces, American Mathematical Society*, Providence, RI, 2004.

[9] H. Hedenmalm, K. Zhu, On the failure of optimal factorization for certain weigted Bergman spaces, *Complex Variables Theory Appl.* 19, 1992, 165–176.

[10] H. Hedenmalm, An invariant subspace of the Bergman space having the codimension two property, *J. Reine Angew. Math.* 443, 1993, 1–9.

[11] H. Hedenmalm, B. Korenblum, K. Zhu, *Theory of Bergman spaces*, Springer-Verlag, New York, 2000.

[12] W. J. Kaczor, M. T. Nowak, *Zadania z analizy matematycznej, tom 2*, Wydawnictwo Naukowe PWN, Warszawa, 2005.

[13] W. Krysicki, L. Włodarski, *Analiza matematyczna w zadaniach, część II*, Wydawnictwo Naukowe PWN, Warszawa, 1974.

[14] F. Leja, *Funkcje zespolone,* Wydawnictwo Naukowe PWN, Warszawa, 2006.

[15] A, Makarewicz, Application of the Maxima in the teaching of science subjects at different levels of knowledge, *Adv. Sci. Technol. Res. J.* 8(24), 2014, 44–50.

[16] J. Musielak, *Wstęp do analizy funkcjonalnej*, Państwowe Wydawnictwo Naukowe, Warszawa 1989.

[17] M. Nowak, R. Rososzczuk, M. Wołoszkiewicz-Cyll, Extremal functions in the weighted Bergman spaces, *Complex Variables and Elliptic Equations*, 62 (1), 2017, 98–109.

[18] W. Rudin, *Analiza rzeczywista i zespolona,* Wydawnictwo Naukowe PWN, Warszawa, 1998.

[19] W. Rzymowski, *Przestrzenie metryczne w analizie,* Wydawnictwo UMCS, Lublin, 2000.

[20] W. Rzymowski, *Macierze i operatory,* Wydawnictwo UMCS, Lublin, 2004.

[21] S. Saitoh, *Theory of reproducing kernels and its applications*, Pitman Research Notes in Mathematics 189, New York, 1988.

[22] S. M. Shimorin, Wold-type decompositions and wandering subspaces for operators close to isometries, *J. Reine Angew. Math.* 531, 2001, 147–189.

[23] S. Shimorin, On Beurling-type theorems in weighted l^2 and Bergman spaces, *Proc. Amer. Math. Soc.* 131 (6), 2003, 1777–1787.

[24] F. H. Szafraniec, *Przestrzenie Hilberta z jądrem reprodukującym,* Wydawnictwo Uniwerytetu Jagielońskiego, Kraków, 2004.

[25] F. H. Szafraniec, Przypadek Stanisława Zaremby - oportunizm czy nonszalancja, *Kwartalnik Historii Nauki i Techniki R.* 61(1), 2016, 117–127.

[26] S. Zaremba, L'équation biharmonique et une class remarquable de functions fondamentales harmoniques, *Bulletin International de l'Académie des Sciences de Cracovie, Classe des Sciences Mathématiques et Naturelles*, 1907, 147–196.

[27] S. Zaremba, Sur le calcul numerérique des foctions demandées dans le probléme de Dirichlet et le probéme hydrodynamique, *Bulletin International de l'Académie des Sciences de Cracovie, Classe des Sciences Mathématiques et Naturelles*, 1909, 125–195.

[28] K. Zhu, *Operator Theory in Function Spaces*, Marcel Dekker, New York, 1990.

Jak unikać błędów w opracowaniach statystycznych?

Dariusz Majerek[1]

Streszczenie

Jedno z najpowszechniej stosowanych narzędzi w badaniach naukowych, czyli statystyka jest równocześnie często bardzo źle wykorzystywane. Błędy jakie można napotkać w publikacjach naukowych, związane z niewłaściwym użyciem statystyki, są często zatrważające. Niniejsza praca przybliża podstawowe błędy, opisuje ich źródła oraz wskazuje sposoby ich unikania.

Abstract

One of the most commonly used tools for deepening knowledge in various fields, that is, the statistics is unfortunately very badly used. Errors that can be encountered in scientific publications related to the misuse of statistics are often alarming. This work introduces basic errors, describes their sources and indicates how to avoid them.

Słowa kluczowe: błąd wnioskowania, dobór próby, populacja, hipoteza

[1]Katedra Matematyki Stosowanej, Wydział Podstaw Techniki, Politechnika Lubelska
e-mail: d.majerek@pollub.pl

1 Wstęp

Statystyka, a dokładniej wnioskowanie statystyczne, jest dziś powszechnie stosowane w poszerzaniu i upowszechnianiu wiedzy z różnych dziedzin, począwszy od nauk ścisłych, a skończywszy na naukach społecznych. Badania naukowe prowadzone na całym świecie, od wielu lat korzystają z analizy ilościowej, w celu potwierdzenia lub odrzucenia pewnych hipotez, modelowania badanych zjawisk czy odnalezienia struktury zależności pomiędzy zmiennymi. Wszystkie te działania powinny się odbywać zgodnie z zasadą ujętą w metodologii badań, ściśle opisującą przebieg przygotowania eksperymentu, jego właściwe przeprowadzenie oraz sposób analizowania otrzymanych wyników. W zależności od obszaru wiedzy, w którym stosowana jest analiza statystyczna, przygotowywane są publikacje szczegółowo opisujące specyfikę badań danej dziedziny, możliwości i zakres wykorzystania w nich metod ilościowych. Można powiedzieć o pewnego rodzaju „kanonach" przeprowadzania badań, formułowania hipotez, prezentacji wyników oraz publikowania wniosków. Z owymi arkanami wiedzy na temat zasad obowiązujących w metodologii badań, w tym również metod statystycznych potrzebnych do przeprowadzenia badań, badacz powinien zapoznać się podczas studiów. Niestety poziom wiedzy akademickiej z zakresu wnioskowania statystycznego jest często niewystarczający. Wynika to zarówno z niedostosowania programu studiów do gwałtownego rozwoju technik statystycznych oraz metod ich implementacji, jak i z braku praktycznych umiejętności ich stosowania. Na wielu, nawet szanowanych uczelniach, brakuje zajęć laboratoryjnych ze statystyki, które mogłyby pomóc opanować umiejętności praktycznego zastosowania, i tak wąskiego zakresu, metod statystycznych.

Efektem tego stanu rzeczy jest to, że naukowcy prowadzący badania w swoich obszarach, nie posiadają wystarczającej wiedzy do przeprowadzenia badań we właściwy sposób. Wówczas często zgłaszają się z tym problemem do osób mniej lub bardziej kompetentnych w analizie danych. W ostatnich latach powstaje coraz więcej publikacji w szanowanych czasopismach, zarówno popularno-naukowych, jak i naukowych, na temat niewłaściwego wykorzystania wnioskowania statystycz-

nego przejawiającego się w błędnym doborze metody statystycznej, próby badanej populacji, na którą będą uogólniane wnioski płynące z badań, czy w końcu niewłaściwym prezentowaniu wyników. Powstają artykuły i książki poświęcone tylko i wyłącznie błędom, jakie popełniamy korzystając z metod statystycznych [10]. Zainteresowanych czytelników odsyłamy do publikacji umieszczonych w bibliografii wskazujących błędy metodologiczne [2–6, 8, 9]. Do najczęściej pojawiających się błędów związanych z wnioskowaniem statystycznym zaliczamy: brak przygotowania eksperymentu, brak reprezentatywności próby, niewłaściwy pomiar zmiennych, użycie niewłaściwych metod statystycznych, brak walidacji modeli, niewłaściwie formułowane wnioski czy nieczytelne lub wręcz umyślnie wprowadzające w błąd wizualizacje wyników. Niestety tylko część z tych błędów jest wynikiem braku wiedzy, czy nieumiejętności przeprowadzania badań zgodnie z powszechnie obowiązującą procedurą. Istnieją udowodnione przypadki, w których badania naukowe były celowo fałszowane, bądź podczas ich opracowania decydowano się na pominięcie ważnych elementów na etapie planowania i przeprowadzania eksperymentu, jak również wyciągania wniosków na podstawie otrzymanych wyników [11].

Kolejną, być może najpoważniejszą konsekwencją niewłaściwego stosowania statystyki jest poddawanie w wątpliwość skuteczności technik analizy danych w procesie badawczym. Od wielu lat upowszechnia się pogląd, czasami przyjmujący nawet formę żartów, że statystyka to jedno wielkie oszustwo. Najbardziej znane, autorstwa Benjamina Disraeliego - dziewiętnastowiecznego Premiera Wielkiej Brytanii, brzmi: „There are three kinds of lies: lies, damned lies, and statistics", spopularyzowane przez Marka Twaina w jego autobiografii. Sądy te biorą się głównie z niezrozumienia istoty losowości próby, czy błędów weryfikacyjnych w testowaniu hipotez, ale również z niedostatecznej dbałości o zachowanie wspomnianych wcześniej reguł metodologii badań.

Wskazane wyżej problemy w zastosowaniu statystyki, obiekcje co do jej użyteczności oraz osobiste doświadczenia autora związane z pomocą w opracowaniu statystycznym materiału w różnych dziedzinach nauki, były przyczynkiem do pochylenia się nad tym problemem. Oprócz wskazywania potencjalnych i realnych

przyczyn popełniania błędów we wnioskowaniu statystycznym, zostanie również podanych kilka wskazówek mogących pomóc w unikaniu „pułapek statystycznych".

Typowe błędy wnioskowania statystycznego oraz sposoby ich unikania podzielono na kilka części. I tak, podrozdział 2 odnosi się do problemów związanych ze zbieraniem danych, doborem próby, jej reprezentatywnością. W podrozdziale 3 opisano typowe błędy w doborze metod badawczych, ze szczególnym uwzględnieniem testowania hipotez. Podrozdział 4 całkowicie będzie poświęcony ukazaniu konieczności walidacji otrzymanych modeli statystycznych. Choć autor nie ma wątpliwości, że niniejsza publikacja nie będzie remedium na wszelkie błędy popełniane we wnioskowaniu statystycznym, to ma nadzieję, że przybliży ona nieco problem uczciwości w badaniach naukowych i uzmysłowi konsekwencje pewnych luk w postępowaniu badacza.

2 Problem doboru właściwego materiału badań

W podręcznikach do metodologii nauk znajdziemy niejednokrotnie zalecenie, aby bardzo wyraźnie określić, już na etapie planowania badania, co chcemy odkryć, jakie zmienne będą nam do tego potrzebne. Wiąże się to z postawieniem jasnych, zrozumiałych i przede wszystkim weryfikowalnych hipotez. To właśnie dokładne określenie zmiennych użytych w badaniu, sposobu ich mierzenia (z tym wiąże się dokładność pomiaru), liczby pomiarów oraz wyraźnego podziału na zmienne zależne i niezależne, leżą u podstaw poprawnie przeprowadzonego badania. Niestety codzienna praktyka pokazuje, iż badacze pomijają jeden lub kilka elementów z tej wyliczanki, która skutkuje mniejszymi lub większymi problemami na etapie konstruowania modelu oraz późniejszego wnioskowania. Zdarza się, nierzadko, że owe pominięcia ważnych elementów badania nie pozwalają ostatecznie rozszerzyć wyników z poziomu próby na populację. Samo badanie nie jest wówczas całkowicie bezużyteczne, bo jest swego rodzaju studium przypadku, które może posłużyć przyszłym badaczom do oceny wielkości próby badanej, czy w ogóle do zmiany kierunku badań, ale nie można z niego stworzyć pewnego prawa nauki. Od tej reguły istnieje jedno odstępstwo, mianowicie w przypadku, gdy w obszarze badanych

obiektów znajdują się takie, które przeczą obecnemu stanowi wiedzy, to należy zweryfikować sąd dotyczący tego prawa nauki. Rozwój wiedzy ma charakter kumulatywny, czyli odkrycia badaczy „sumują" się w pewien coraz bardziej spójny sposób. Przykładowo powszechnie znane prawo dotyczące grawitacji, sformułowane przez Newtona, było przyjmowane jako ogólnie obowiązujące we wszechświecie, aż do czasów Einsteina, który zauważył „nieścisłości" w tej teorii. Uzupełnił on nasze postrzeganie grawitacji o nowy relatywistyczny element.

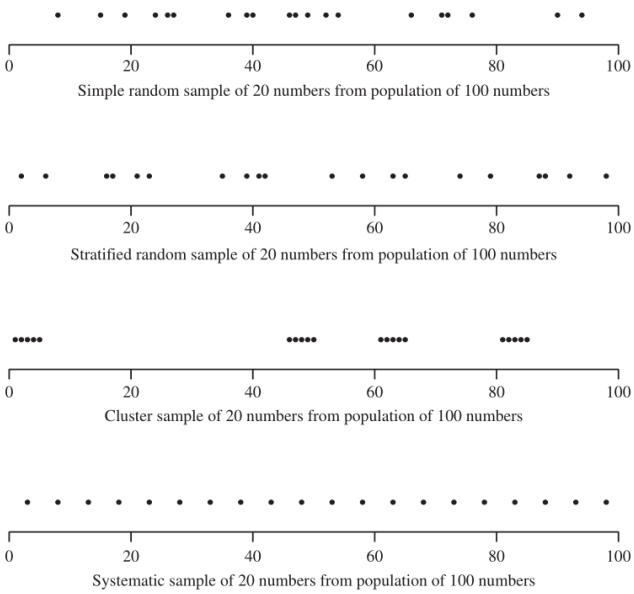
Zalecenia dotyczące poprawnego przeprowadzania badań każą przed ich podjęciem przeprowadzić dogłębną analizę literatury tematu z obszaru planowanych badań. Może się bowiem okazać, że dokonane przez nas odkrycie jest już powszechnie znaną wiedzą. Z punktu widzenia wnioskowania statystycznego, ma to jeszcze inną wartość, mianowicie, pozwala to na uwzględnienie w badaniu wszystkich zmiennych zakłócających wartość zmiennej wyjściowej (wynikowej), co w konsekwencji pozwala na dokładniejsze oszacowanie efektu oddziaływania kluczowej dla badania zmiennej niezależnej. Jeśli, przykładowo, celem badacza jest ustalenie wielkości wpływu interwencji w postaci szkoleń zawodowych, na aktywność zawodową badanych, to konieczne jest włączenie do modelu statystycznego opisującego tą zależność zmiennych takich jak fakt uczestniczenia w innych programach aktywizujących zawodowo, czy liczbę ofert pracy, które przyjął badany od Urzędu Pracy. Łatwo można sobie wyobrazić, że na rynku znajdują się osoby z różnym poziomem aktywności w poszukiwaniu nowej pracy, co może być czynnikiem decydującym o ewentualnym zatrudnieniu. Wówczas efekt oddziaływania w postaci np. wojewódzkiego programu rozwoju aktywności zawodowej może nie przynosić pożądanych efektów, a szacowanie efektu netto tego programu może być obciążone błędem, jeśli nie uwzględnimy jakiejś zmiennej towarzyszącej mającej wpływ na aktywność w poszukiwaniu pracy. W kontekście planowania doświadczeń ma to wielkie znaczenie, ponieważ na etapie zbierania danych wiemy, że powinniśmy uwzględnić pewne zmienne a innych nie. To znów wpływa na konstrukcję narzędzia do zbierania danych, ustalenie odpowiedniego operatu losowania, czy zawężenia lub rozszerzenia populacji generalnej, której dotyczyć będzie badanie. Dobór

odpowiedniej próby i populacji jest krytycznym punktem, niemal wszystkich badań opartych o analizę statystyczną. Wybór sposobu pozyskania próby i jej wielkość, są kluczowe dla zachowania reprezentatywności próby oraz do uzyskania trafności wewnętrznej i zewnętrznej badania, czyli oceny, czy oszacowany efekt nie posiada obciążenia selekcyjnego oraz czy wyniki badań można uogólniać na szerszą grupę obiektów. Określenie populacji generalnej, której dotyczyć będzie badanie ma na celu uściślenie do jakiej zbiorowości można będzie uogólniać wyniki badań, choć o tym mogą decydować również same wyniki.

Zasady randomizacji mówiące, że dobór obiektów do próby oraz przyporządkowanie ich do grup porównywanych ma być losowy, powoduje, że wykonanie badań z zachowaniem tych reguł jest kosztowne, logistycznie uciążliwe oraz czasem z powodów etycznych wręcz niemożliwe do wykonania. Wyobraźmy sobie przykładowo, że mamy dokonać oceny poziomu wydatków na żywność przez gospodarstwa domowe w Polsce. Jako operatu losowania możemy użyć adresów zamieszkania, ale w ten sposób gospodarstwa domowe wylosowane do próby pochodziłyby z całej Polski, a to generowałoby duże koszty takiego badania i powodowało trudności logistyczne. Przykłady badań klinicznych pokazują, że wykonie badań z zachowaniem obu zasad randomizacji jest czasami niemożliwe do wykonania z powodów etycznych. Nie można bowiem zmuszać kogoś do przyjęcia leku eksperymentalnego, lub narażania na szwank swojego zdrowia, ze względu na dobro badań.

Zważywszy na powyższe problemy w doborze próby, stosowane są różne metody zbierania materiału badawczego. Dzielimy je na probabilistyczne i nieprobabilistyczne. Te pierwsze dobierane są w oparciu o pewien schemat losowania. Nie zawsze jest to losowanie proste, czyli z zachowaniem zasady, że każdy obiekt może być wylosowany z tym samym prawdopodobieństwem. Do grupy tej należą losowanie warstwowe, grupowe, okresowe. Cechą charakterystyczną tych prób jest to, iż można na podstawie praw rachunku prawdopodobieństwa oszacować dokładność wnioskowania na niej opartego. Próby nieprobabilistyczne (celowe) to próby kwotowe, ochotnicze, eksperckie, charakteryzujące się tym, że nie zapewniają re-

prezentatywności próby i nie można na ich podstawie decydować o precyzji szacunków. Dobór do prób celowych i eksperckich odbywa się w sposób arbitralny co często jest ich największą wadą.



Rysunek 1: Przykłady prób 20 elementowych z liczb od 1 do 100 losowanych czterema rożnymi schematami: losowanie proste, warstwowe (gdzie warstwami były kolejne dziesiątki liczb a z każdej warstwy losowano 2 liczby), grupowe (gdzie było 20 grup składających się z 5 kolejnych liczb i wylosowano 4 grupy) i systematyczne (startujące od 3 z krokiem 5)

Z powodu ograniczeń budżetowych, czasowych i logistycznych grupa nieprobalistycznych metod doboru próby dominuje, szczególnie w naukach społecznych. Konsekwencją tego są wyniki badań, których zasięg jest ograniczony. Najmniej polecana spośród wszystkich metod pozyskiwania danych jest próba ochotnicza. Istnieją prace, które wskazują pewien profil psychologiczny osób zgłaszających się do takich badań, a to może mieć swoje konsekwencje w postaci obciążenia wyników [14, 15]. Zdarzają się również takie sytuacje, iż zachowaliśmy wszelkie starania dotyczące sposobu zbierania materiału badawczego, ale część osób objętych na-

szym badaniem umarło, wycofało się lub z jakiegoś innego powodu stali się dla nas niedostępni. Braki danych mogą być także powodowane innymi przyczynami, np. respondenci nie chcieli wypełniać rubryki dotyczącej poziomu zarobków. W przypadku niewielu braków, które dodatkowo mają charakter losowy, można użyć technik statystycznych w celu ich uzupełnienia. Natomiast, jeśli braki nie są losowe lub jest ich znacząca liczba, należy przemyśleć ewentualne badanie uzupełniające. Oczywiście nie zawsze da się takie uzupełnienie przeprowadzić i to nie tylko ze względów finansowych, czy logistycznych, ale np. dlatego, że pomiędzy poprzednim pomiarem (tym właściwym) a uzupełniającym wystąpił dodatkowy czynnik, który w istotny sposób może wpłynąć na wartości w próbie uzupełniającej.

Ostatni aspekt doboru próby, czyli jej liczebność, w sposób istotny zależy od założonej precyzji z jaką chcemy mierzyć zmienną zależną (patrz nierówność (2.1)), ale czasami również od innych czynników takich jak moc testu statystycznego, czy pewna minimalna liczebność w badanych grupach. Wśród domorosłych statystyków panuje przekonanie, że zwiększenie liczebności próby spowoduje, iż ta stanie się bardziej reprezentatywna. Niestety jest to pogląd błędny, ponieważ, jeśli sposób jej doboru jest niewłaściwy, to jej wielkość nie zmniejszy obciążenia wyniku badań (obciążenie to nazywamy selekcyjnym). Dobrą praktyką w kontekście odpowiedniego przygotowania docelowego zbioru danych jest przeprowadzenie badania pilotażowego na dużo mniejszej próbie. Pozwala ona na uchwycenie głównych problemów pojawiających się na etapie zbierania danych, pomaga określić dokładności szacunków poszczególnych zmiennych, ich rozkłady, zależności między nimi oraz na wyliczenie minimalnej liczebności próby docelowej.

Lemat 2.1. *Wzór na minimalną liczebność próby, przy zadanej precyzji e przyjmuje postać*

$$n_0 \geqslant \left(\frac{z_{\alpha/2}s}{e}\right)^2, \tag{2.1}$$

gdzie n_0 oznacza minimalną liczebność próby potrzebną do osiągnięcia oszacowania mierzonej wielkości z dokładnością e, $z_{\alpha/2}$ jest kwantylem rozkładu normalnego standaryzowanego, a s odchyleniem standardowym mierzonej wielkości.

Bardzo często dobór elementów do próby odbywa się przy użyciu schematu losowania warstwowego, uwzględniającego dodatkowo koszt przeprowadzania badań w poszczególnych grupach, który nie musi być identyczny dla wszystkich warstw. Powodem konieczności stosowania losowania warstwowego jest heterogeniczność populacji ze względu na badaną cechę. Próba prosta w tym przypadku mogłaby się okazać nieefektywna w kontekście zmienności estymatorów wyznaczonych na jej podstawie. Odpowiednio dobrana próba warstwowa, może znacznie obniżyć wariancję estymatora. Pokazuje to następujący przykład.

Przykład 2.1. Dane pochodzą ze spisu rolniczego przeprowadzanego regularnie co 5 lat w Stanach Zjednoczonych [13]. Próba zawiera informacje na temat liczby akrów zajętych przez farmy w poszczególnych latach objętych badaniem w podziale na regiony. Na podstawie próby prostej 300 elementowej można policzyć, że w roku 1992 liczba akrów zajętych przez farmy wynosiła 916927110, przy błędzie standardowym estymacji na poziomie 58169381. Ponieważ badana populacja nie jest jednorodna ze względu na badaną cechę (regiony zachodnie charakteryzują się większymi farmami), to losowanie warstwowe z uwzględnieniem regionów może poprawić precyzję oszacowania liczby akrów zajętych przez farmy. Istotnie przy doborze warstwowym proporcjonalnym próby o liczebności 300 elementów, otrzymujemy oszacowanie szukanej wielkości na poziomie 909736035, przy błędzie standardowym 50417248. Poprawa precyzji oszacowania wyrażona wariancją wynosi $50417248^2/58169381^2 \approx 0.75$. Oznacza to, że w przypadku losowania warstwowego potrzebujemy $225 = 300 \cdot 0.75$ elementów w próbie aby otrzymać precyzję zbliżoną do tej z losowania prostego.

W powyższym przykładzie można również dodać jeszcze jeden aspekt towarzyszący pozyskiwaniu danych, a mianowicie koszt. Jeśli nakłady pieniężne na pozyskiwanie informacji o farmach w różnych regionach byłyby inne (np. z powodu odległości), to losowanie optymalne (uwzględniające koszty lub zmienności w warstwach) wydawałoby się bardziej racjonalne.

Z pewnością należy pamiętać, iż jakość zebranego materiału przełoży się w sposób znaczący na jakość otrzymanych wyników. Wybitny amerykański statystyk

Raymond J. Carroll zwykł na temat złej jakości danych mawiać „Garbage in, gargabe out". Dlatego na etapie planowania badań warto „zatrudnić" statystyka, który może wskazać istotne elementy na tym etapie procesu badawczego. Pamiętajmy bowiem, że nawet najbardziej wyrafinowane techniki statystyczne nie sprawią, że ze „słabych" danych da się uzyskać wartościowe wyniki.

3 Dobór metod statystycznych

Niewłaściwy dobór materiału badawczego nie jest jedynym powodem błędów we wnioskowaniu statystycznym. W równym stopniu decyzja o wyborze metod statystycznych użytych w badaniu może skutkować fałszywymi wnioskami. W niniejszym opracowaniu będą wzięte pod uwagę techniki z zakresu: estymacji punktowej i przedziałowej, weryfikacji hipotez, modelowania zależności za pomocą regresji liniowej i nieliniowej.

Na wstępie należy nadmienić, że wszystkie metody statystyczne mają pewne ograniczenia, zarówno w obszarze ich stosowalności, jak i potencjalnych wniosków, które można na ich podstawie wyciągać o populacji. Tylko dobre poznanie i spełnienie wszelkich założeń danej metody oraz świadomość jej ograniczeń daje nam gwarancję, że otrzymane wyniki będą poprawne. Przez właściwy dobór metody statystycznej rozumiemy, właśnie poznanie warunków jej stosowalności oraz ich spełnienie.

Niestety szereg publikacji naukowych oraz własne doświadczenia autora nabyte podczas konsultacji dotyczących doboru metod statystycznych pokazują, że czasami ograniczony zasób technik analizy danych oraz przyzwyczajenia badaczy do pewnych metod, a czasami także brak wiedzy na temat możliwości ich zastosowania w konkretnym problemie powodują, że stosowane są niewłaściwie. Sytuacje, w których badacz wykorzystuje niewłaściwą technikę statystyczną, sugerując się sugestią promotora, autora podobnej pracy czy własnym błędnym doświadczeniem z poprzednich badań, są na porządku dziennym. Nie jest to uwaga, która dotyczy wyłącznie niestatystyków, ponieważ również konsultanci statystyczni

pracujący w zespołach badawczych, a także ci, którzy czuwają nad poprawnością metodologiczną publikacji nie zawsze dobrze wywiązują się ze swoich zadań, ulegając często presji środowisk, w których funkcjonują. Panuje bowiem zwyczaj, że wybrana technika statystyczna, mimo że nie jest adekwatna w danej sytuacji, zostaje zastosowana, ponieważ zakłada się, że recenzenci nie są gotowi na pewne „nowości" i mogliby odrzucić pracę z powodu braku wiedzy na temat właściwych metod. Inną wielką bolączką dzisiejszych prac naukowych wykorzystujących techniki statystyczne jest to, że nie są w nich publikowane dane dotyczące weryfikacji założeń poszczególnych metod stosowanych w pracy, wielkości prób, informacje o rozkładach poszczególnych zmiennych, czasie kiedy były mierzone czy warunkach w jakich został przeprowadzony eksperyment. Wielu badaczy usprawiedliwia się faktem, iż ograniczenia edytorskie dotyczące liczby stron, nie pozwalają na pełne zaprezentowanie wyników wraz z wszelkimi szczegółami zastosowanych metod. Niestety taki stan rzeczy powoduje pewne komplikacje w sytuacji gdy inni naukowcy chcieliby wykorzystać otrzymane wyniki badań. Czasami zwykła replikacja przeprowadzonych eksperymentów nie jest możliwa do wykonania w laboratorium w innej części świata, ponieważ pewne szczegóły pominięte w opublikowanych pracach okazują się niezbędne. Z drugiej strony należy wyróżnić te prace naukowe, w których wszelkie informacje o przeprowadzonym badaniu, wykorzystanych technikach, spełnieniu założeń oraz zestawienia tabelaryczne zostały dołączone w postaci dodatków.

3.1 Estymacja punktowa

Estymacja punktowa jest chyba najczęściej stosowanym narzędziem statystycznym w ramach prac naukowych, raportów, spisów powszechnych czy badań opinii publicznej. Polega ona na wyznaczeniu wartości estymatora na podstawie próby $\hat{\theta}$, który przybliża nieznaną wartość parametru rozkładu θ badanej cechy. Podstawowy kurs statystyki informuje nas, że dobry estymator, to taki, który jest:

- nieobciążony – czyli przeciętnie wartość estymatora „trafia" w parametr estymowany ($E(\hat{\theta}) = \theta$);

- zgodny – czyli wraz ze wzrostem liczebności próby, rośnie prawdopodobieństwo, że oszacowanie będzie przyjmować wartości coraz bliższe szacowanego parametru ($\wedge_{\varepsilon>0} \lim_{n\to\infty} P(|\hat{\theta}-\theta|<\varepsilon)=1$);
- najefektywniejszy – czyli taki, którego błąd średniokwadratowy[2] jest najmniejszy; w klasie estymatorów nieobciążonych jest to estymator o najmniejszej wariancji.

Wspomniane charakterystyki estymatorów wprowadza się dlatego, iż nie istnieją estymatory najlepsze, czyli takie, które w klasie wszystkich estymatorów jednostajnie minimalizują błąd średniokwadratowy. Przy okazji unika się tak bezsensownych estymatorów jak np. stała. Niestety nie zawsze estymatory nieobciążone istnieją, a estymatory nieobciążone o minimalnej wariancji mogą okazać się gorsze od estymatorów obciążonych. Czasami statystycy dorzucają kolejne wymagania dotyczące estymatorów, takie jak np. odporność[3]. Własność ta oznacza, że odstępstwa od normalności rozkładu badanej cechy, w postaci braku symetrii, ciężkoogonowości[4] lub występowaniu wartości ekstremalnych, nie wpływają znacząco na wartość estymatora. W badaniu tendencji centralnej zarobków Polaków posłużymy się raczej medianą, czyli statystyką bardziej odporną na prawostronną asymetrię badanego rozkładu oraz występowanie ewentualnych wartości odstających, niż średnią, która jest mniej odporna.

W estymacji punktowej należy pamiętać, że w przypadku cech ciągłych prawdopodobieństwo, że estymator przyjmuje wartość równą parametrowi estymowanemu jest zerowe. Estymacja punktowa zawsze jest obarczona pewnym błędem, ponieważ wyznaczana jest na podstawie próby, czyli pewnej realizacji zmiennej losowej. Każda inna realizacja może skutkować inną wartością estymatora. Dodatkowo dla prób probabilistycznych możemy szacować tą niepewność. Brak zrozumienia tego faktu skutkuje nieuzasadnioną krytyką statystyki jako wiarygodnej metody naukowej. Rachunek prawdopodobieństwa, który jest głównym narzędziem do budo-

[2] $MSE(\hat{\theta}) = E(\hat{\theta}-\theta)^2$

[3] ang. *robust*

[4] ang. *heavy-tailed*

wy modeli statystycznych oraz ich oceny (w szczególności estymatorów), określa z jaką niepewnością będziemy mieli do czynienia stosując to czy inne narzędzie statystyczne. Można zatem powiedzieć, że jedynie uwiarygadnia on wyniki, jeśli użyjemy prawidłowego narzędzia do analizy danych.

Bardzo częstym błędem w estymacji punktowej jest pomijanie kontekstu wielkości populacji, z której została pobrana próba, przy szacowaniu dokładności estymatora. Statystycy bowiem zalecają stosowanie tzw. poprawki dla skończonej populacji[5], w przypadku populacji mało licznych. Przykładowo jeśli wielkość populacji to 150 obiektów, a w próbie znajduje się ich 30, to każdy element jest reprezentantem siebie i 4 innych obserwacji nie ujętych w próbie. Wówczas szacowanie wartości oczekiwanej na podstawie średniej z próby $\bar{x}$ odbywa się z pewnym błędem wyrażonym np. wariancją $Var(\bar{x})$. Dla odmiany, gdy próba stanowi całą populację, zmienność związana z szacowaną wartością jest równa zero. Aby uwzględnić proporcję jaką próba stanowi w stosunku do populacji zastosowano właśnie poprawkę dla skończonej populacji

$$Var(\bar{x}) = \left(1 - \frac{n}{N}\right)\frac{s^2}{n}, \tag{3.1}$$

gdzie n i N oznaczają odpowiednio wielkość próby i populacji, a s^2 jest wariancją próbkową. Oznacza to, że dla próby 30 elementowej z populacji 150 elementowej, jeśli wariancja średniej bez korekty wynosi np. 50, to z korektą wyniesie $40 = 0.8 \cdot 50$, co stanowi znaczną poprawę precyzji oszacowania. W przypadku dużych populacji poprawka dla skończonej populacji nie ma takiego znaczenia. Rozważmy dwie populacje wielkości 1000000 i 100000000 elementów. Z obu populacji wylosujmy po 1000 elementów. Wówczas korekta wariancji w pierwszym przypadku wynosi 0.999, a w drugim 0.99999.

Błąd średniokwadratowy, który jest często używaną miarą do oceny jakości estymacji, wyznacza swego rodzaju kompromis pomiędzy obciążeniem[6] a wariancją estymatora.

[5] ang. *finite population correction*

[6] dokładniej jego kwadratem

Można to wyrazić równaniem

$$MSE(\hat{\theta}) = Var(\hat{\theta}) + (b(\hat{\theta}))^2, \tag{3.2}$$

gdzie $b(\hat{\theta}) = E(\hat{\theta}) - \theta$ jest obciążeniem estymatora. Najlepszy estymator charakteryzowałby się oczywiście minimalnym błędem średniokwadratowym, ale nie da się jednocześnie minimalizować obu jego składowych. Dlatego w kontekście analizowanego problemu możemy minimalizować wariancję, godząc się na pewne niewielkie obciążenie lub żądać nieobciążoności licząc się z większa wariancją estymatora. Nigdy jednak nie poszukujemy estymatorów o skrajnych wartościach obu składowych.

Ciekawym, niestety rzadko stosowanym w badaniach naukowych sposobem na obniżenie wariancji estymatora jest użycie estymacji ilorazowej[7]. Pierwszym udokumentowanym przykładem użycia tej metody jest oszacowanie z dużą dokładnością populacji Francuzów przez Pierre Simon de Laplace'a w 1802 roku. Wówczas nikt nie myślał jeszcze o spisach powszechnych. Zastosował on metodę estymacji ilorazowej, polegającej na wykorzystaniu ilorazu cechy badanej i silnie skorelowanej z nią zmiennej towarzyszącej[8]. Laplace wykorzystał jako zmienną towarzyszącą liczbę zarejestrowanych narodzin w analizowanych gminach, która jest silnie skorelowana z populacją. Estymator ilorazowy przyjmuje postać

$$B = \frac{\bar{y}}{\bar{x}} = \frac{\hat{t}_y}{\hat{t}_x}, \tag{3.3}$$

gdzie y to cecha interesująca badacza, x to cecha pomocnicza, $\hat{t}_y = \sum_{i=1}^{n} y_i$ a $\hat{t}_x = \sum_{i=1}^{n} x_i$. W estymacji sumy wartości badanej cechy nie możemy wziąć po prostu $\hat{t}_y = N \cdot \bar{y}$, ponieważ nie znamy wielkości populacji N. Wiemy natomiast jak ją oszacować, bo $N \approx t_x/\bar{x}$, gdzie t_x jest sumą wartości zmiennej pomocniczej w populacji. Wówczas nieznana suma wartości cechy y jest równa

$$\hat{t}_{yr} = \bar{y}\frac{t_x}{\bar{x}}. \tag{3.4}$$

[7] ang. *ratio estimation*

[8] tzw. cecha pomocnicza - ang. *auxiliary variable*

Przykład 3.1. [12] Załóżmy, że celem naszego badania jest oszacowanie ile ryb o długości przekraczającej 12 cm jest w wyłowionej sieci. Na wstępie zważymy cały połów, wiedząc, że cecha pomocnicza x, czyli waga jest silnie skorelowana z liczbą ryb dłuższych niż 12 cm, czyli y. Następnie losujemy próbę prostą ryb z sieci i zliczamy, ile z nich spełnia warunki zadania. Odnosząc tą liczbę do liczebności próby otrzymamy proporcję $\bar{y}$. Wówczas wszystkie parametry do wyliczenia $\hat{t}_{yr}$ są już znane. Jak widać użycie zmiennej, której statystyki $\hat{t}_x$ i $\bar{x}$ można łatwo obliczyć, uprościło wyznaczenie statystyki zmiennej badanej.

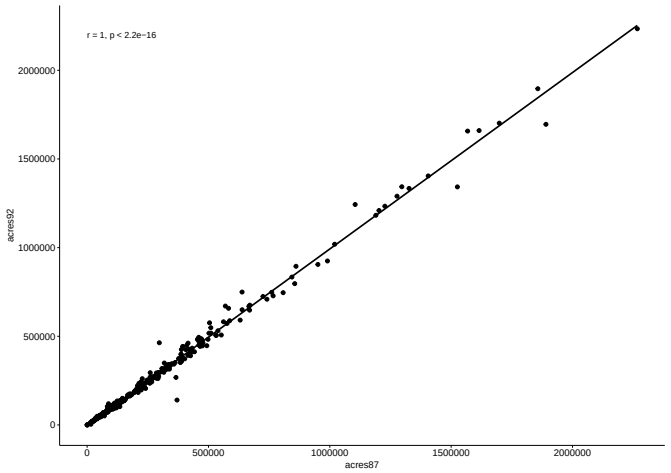
Przykład 3.2. Wracając do danych z przykładu 2.1 możemy jeszcze raz oszacować liczbę akrów zajętych przez farmy, tylko tym razem metodą ilorazową. Szacunki na podstawie próby prostej przyniosły wariancję estymatora na poziomie $58169381^2 = 3.383677e+15$, przy estymacie 916927110. Zatem wariancja naszego oszacowania jest duża. Wprowadzając zmienną pomocniczą – liczbę akrów zajętych przez farmy w 1987 roku – otrzymamy iloraz $\hat{B} = 0.9865652$, ponieważ zmienne są silnie ze sobą skorelowane (patrz rysunek 2). Dodatkowo zostały oszacowane: liczba akrów zajętych przez farmy w 1987 roku oraz wariancja tej wielkości. W przeciwieństwie do wariancji obu składowych, iloraz charakteryzuje się bardzo niewielką zmiennością $3.306794e-05$. Ostatecznie możemy wyliczyć, że

$$\hat{t}_{yr} = \hat{B} \cdot t_x \approx \hat{B} \cdot \hat{t}_x = 0.9865652 \cdot 929413560 = 916927110. \tag{3.5}$$

Ponieważ użyliśmy $\hat{t}_x$ w miejsce t_x, to liczba zajętych akrów przez farmy jest identyczna jak w przykładzie 2.1. Korzyść odnosimy w postaci zwiększonej precyzji oszacowania. Wariancja estymatora wyznaczonego metodą ilorazową wynosi $2.8564e+13$ i stanowi jedynie około 0.8% wariancji estymatora klasycznego[9].

Jeszcze jednym błędem pojawiającym się niektórych publikacjach naukowych jest użycie niewłaściwego estymatora danej wielkości w stosunku do sposobu pozyskiwania danych. Zdarza się mianowicie, że nieświadomie niektórzy badacze używają estymatorów przeznaczonych dla losowania prostego, które są najbardziej

[9] $2.8564e+13/3.383677e+15 = 0.008441704$



Rysunek 2: Wykres zależności pomiędzy liczbą akrów zajętych przez farmy w 1987 i 1992 roku

popularne, do prób warstwowych, grupowych lub do losowań wielostopniowych. Uwaga ta dotyczy szczególnie badań z wykorzystaniem ankiet.

Podsumowując ten podrozdział należy jeszcze raz wyraźnie podkreślić, że nie ma estymatorów najlepszych. Każdej estymacji punktowej powinna towarzyszyć krótka refleksja nad tym, jaki parametr szacujemy, jaki estymator będzie optymalny z punktu widzenia błędu średniokwadratowego, precyzji szacunku oraz rodzaju próby.

3.2 Estymacja przedziałowa

Rzadko estymacja punktowa jest jedynym i wystarczającym narzędziem analizy ilościowej stosowanym w pracach naukowych. Oszacowania punktowe stosowane w statystyce opisowej są często uzupełniane przez estymację przedziałową, polegającą na znalezieniu takiego obszaru przestrzeni parametrów, aby z zadaną pewnością[10] nieznany parametr lub parametry należały do tego obszaru.

[10] nazywaną poziomem ufności równym najczęściej $1 - \alpha = 0.95$

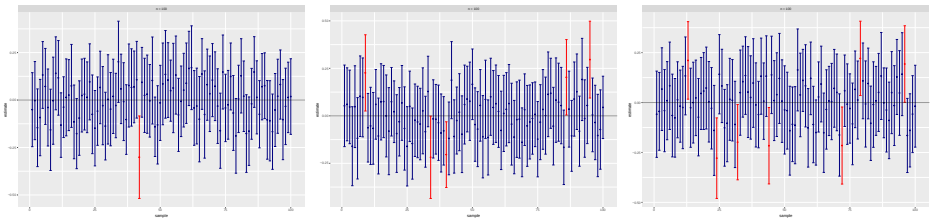
$(1-\alpha)\%$-ową realizacją przedziału ufności dla nieznanego parametru θ jest przedział $(\underline{\theta}, \overline{\theta})$ spełniający warunek

$$P\left(\underline{\theta} \leqslant \theta \leqslant \overline{\theta}\right) = 1 - \alpha \tag{3.6}$$

gdzie $\underline{\theta}$ nazywamy dolnym, a $\overline{\theta}$ górnym końcem przedziału ufności.

Niestety mimo, że estymacja przedziałowa łączy w sobie zalety estymacji punktowej oraz precyzję oszacowania, to rzadko jest wykorzystywana w pracach naukowych. Często estymacja punktowa jest połączona z weryfikacją hipotez, ale rzadko uzupełniona o przedziały ufności.

Intuicja leżąca u podstaw konstrukcji przedziałów ufności jest taka, aby wyznaczyć w jakim przedziale leżeć będzie wartość szacowanego parametru, ze z góry zadaną pewnością. Przykładowo, jeśli 95%-ową realizacją przedziału ufności dla nieznanej wartości oczekiwanej wzrostu Polaków jest przedział (157.4, 184.7)[11], to nieznany parametr μ należy do tego przedziału z prawdopodobieństwem 0.95. Nie oznacza to jednak, że na 100 wylosowanych prób z tego samego rozkładu, 95 będzie miało przedział ufności pokrywający nieznany parametr (patrz rysunek 3).



Rysunek 3: 95%-owe realizacje przedziałów ufności dla trzech 100-elementowych próby wylosowanych z rozkładu $N(0,1)$

Jak widać na rysunku 3, dla każdego losowania 100-elementowych prób realizacja 95%-owego przedziału ufności zawiera inną liczbę przedziałów pokrywających nieznany parametr. W pierwszym przypadku jest ich 99, w drugim 95,

[11] wartości czysto hipotetyczne

a w trzecim 93. Każda kolejna realizacja może nam dać inny wynik, ponieważ zależą one od prób, które są wyznaczane losowo. Można natomiast zauważyć pewną prawidłowość, że liczba przedziałów w kolejnych powtórzeniach tego doświadczenia będzie oscylowała wokół 95. Zwiększając liczbę prób losowanych z tego samego rozkładu, na podstawie których wyznaczane są końce przedziałów, odsetek tych, które pokrywają nieznany parametr, będzie się zbliżał do 0.95. Ponieważ każda 100-elementowa próba w naszym przykładzie była losowana niezależnie z tego samego rozkładu, to możemy je równie dobrze potraktować jako 300 prób 100-elementowych z tego rozkładu. W tym przypadku 287 przedziałów pokrywa nieznany parametr i stanowi to 0.9567 wszystkich przedziałów ufności. Zrozumienie istoty pewności pokrycia nieznanego parametru przez przedział ufności jest konieczne, tym bardziej, że zazwyczaj wyznaczamy tylko jeden przedział ufności i nie mamy pewności, że żądany parametr leży w tym przedziale.

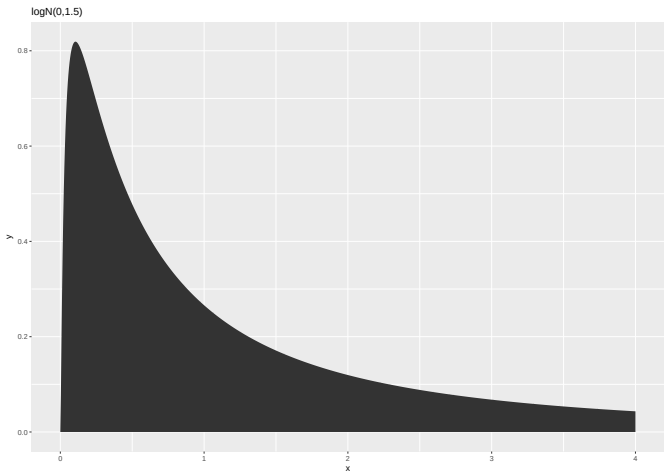
O szerokości przedziału ufności decydują przede wszystkim trzy czynniki:

- liczebność próby na podstawie, której szacujemy parametr – im jest ona większa, tym przedział węższy;
- poziomu ufności z jakim chcemy szacować parametr – im większy, tym szerszy przedział;
- zmienność szacowanego parametru – im większa, tym szerszy jest przedział ufności.

Najczęściej stosowane przedziały ufności dla nieznanej wartości oczekiwanej, zakładają normalność rozkładu, z którego losowana była próba lub żądają aby wielkość tej próby była stosunkowo duża. Oba te warunki nastręczają w praktyce badaczy wielu problemów, oczywiście wówczas, gdy ci mają tego świadomość. Zdarzają się bowiem przypadki, że postać rozkładu z którego losowano próbę nie jest znana i, co gorsza, nie jest badana. Z drugiej strony ograniczenia finansowe, logistyczne i etyczne, o których wspomniano w rozdziale 2 powodują, że próby są często mało liczne. Istotna odchyłka od normalności rozkładu badanej cechy, w postaci

asymetrii rozkładu lub ciężkoogonowości powoduje, że zakładana pewność pokrycia nieznanego parametru przez przedział ufności może być nieprawdziwa. Aby zilustrować jaki wpływ na pokrycie przedziałem ufności ma asymetria rozkładu przytoczymy następujący przykład.

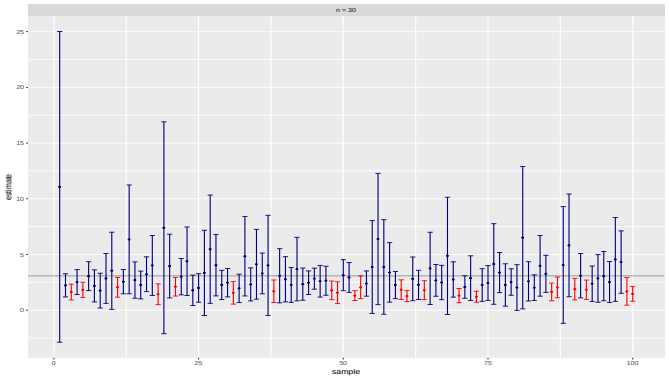
Przykład 3.3. Z rozkładu $logN(0, 1.5)$ wylosujemy 100 prób po 30 elementów. Następnie wyznaczymy dla każdej próby przedział ufności dla wartości oczekiwanej i sprawdzimy ile z tych przedziałów zawierało prawdziwą wartość. Rozkład logarytmiczno-normalny charakteryzuje się silną asymetrią prawostronną (patrz rysunek 4), która będzie wpływać znacząco na pokrycie przez przedziały ufności wartości oczekiwanej, czyli $\exp(\mu + \sigma^2/2) \approx 3.08$.



Rysunek 4: Funkcja gęstości rozkładu $logN(0, 1.5)$ z wyraźną prawostronną asymetrią rozkładu.

Istotnie, wiele z przedziałów (dokładnie 22) nie zawiera wartości oczekiwanej. Na rysunku 5 zaznaczono je kolorem czerwonym.

Naruszenie założenia o normalności rozkładu, z którego pobrano próbę, może mieć swoje konsekwencje w braku pokrycia nieznanego parametru przez przedział ufności. Są jednak takie, które nie wymagają normalności rozkładu. Centralne



Rysunek 5: 95% realizacje przedziałów ufności dla nieznanej średniej z rozkładu $logN(0, 1.5)$

twierdzenie graniczne mówi, że gdy próba $x_1, \ldots, x_n$ jest dostatecznie duża i pobrana niezależnie z rozkładu o średniej μ i wariancji σ^2, to rozkład odpowiednio standaryzowanej zmiennej losowej jest zbliżony do normalnego, czyli

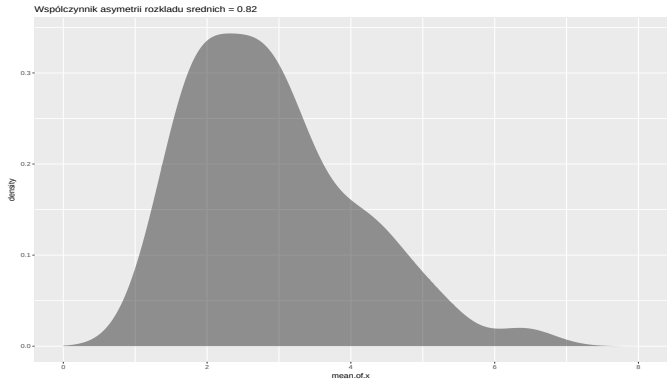
$$\frac{\bar{x} - \mu}{\sigma/\sqrt{n}} \xrightarrow{D} N(0,1). \tag{3.7}$$

Dla danych z powyższego przykładu, jeśli wyliczymy średnie z poszczególnych prób, to otrzymamy rozkład następującej postaci

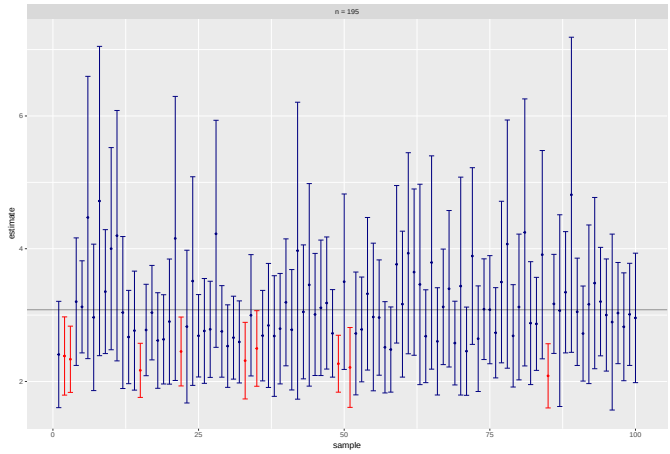
Szybkość zbieżności do rozkładu normalnego zależy jednak od asymetrii rozkładu i dla rozkładów skośnych wymagana jest większa liczba elementów w próbie, aby osiągnąć zbieżność. W roku 1977 Cochran [7] zaproponował formułę, która uwzględniać będzie asymetrię rozkładu w szybkości zbieżności. Następnie Sugden [16] dopracował regułę Cochrana i dziś funkcjonuje ona w postaci

$$n_{min} = 28 + 25 \left(\frac{\sum_{i=1}^{n} (y_i - \bar{y})^3}{n s^3} \right)^2. \tag{3.8}$$

Według tej reguły dla danych z poprzedniego przykładu liczebność minimalna powinna wynosić 195. Istotnie przy tej liczebności prób, znacznie poprawia się pokrycie nieznanej wielkości przez przedziały ufności (patrz rysunek 7).



Rysunek 6: Rozkład gęstości jądrowej średnich ze 100 prób z $logN(0, 1.5)$



Rysunek 7: 95% realizacje przedziałów ufności dla prób o liczebności 195 pobranych z rozkładu $logN(0, 1.5)$

Jeszcze innym sposobem radzenia sobie z asymetrią rozkładu lub małą liczebnością próby są metody bootstrapowe wyznaczania przedziałów ufności. Polegają one na wielokrotnym losowaniu z powtórzeniami elementów z próby, tak aby uzyskać wiele prób (najczęściej 1000) tej samej liczebności o zbliżonym rozkładzie. Na podstawie tych prób wyznacza się estymatę badanego parametru, która charak-

teryzuje się pewną zmiennością związaną z różnicami w wylosowanych próbach. W jednej z metod bootstrapowych wyznacza się kwantyle rzędu 0.025 i 0.975 z rozkładu estymat, które stanowią końce przedziału ufności. Istnieją również inne metody bootstrapowe, zarówno nieparametryczne (nie zakładające postaci rozkładu, z którego losowano próby), jak i parametryczne, w których losuje się próby z rozkładu znanej postaci. Niestety w tym przypadku również oszacowania na końce przedziału ufności są obarczone pewnym obciążeniem. Wynika ono z rozkładu populacji, z której losowano próby. Nie mniej jednak metody bootstrapowe mogą być alternatywą dla klasycznych przedziałów ufności, w sytuacji nie spełnienia założeń ich dotyczących, a szczególnie stają się przydatne w sytuacji szacowania parametrów, dla których klasycznych przedziałów ufności nie ma.

W niewielu publikacjach naukowych można spotkać prawidłowo przeprowadzone wnioskowanie oparte o estymację przedziałową. Po pierwsze rzadko jest stosowana, a jeśli już, to bez żadnej refleksji nad rozkładem populacji czy liczebnością próby na podstawie, której są szacowane końce przedziałów.

3.3 Weryfikacja hipotez

Badania naukowe najczęściej polegają na stawianiu hipotez, a następnie na podstawie dostępnego materiału badawczego ich weryfikowaniu. Nie trzeba chyba nikogo przekonywać, jak użytecznym w tym kontekście narzędziem są testy statystyczne. Niemal każda praca naukowa lub popularno-naukowa, korzystająca z metod ilościowych, stosuje testy. Biorąc pod uwagę jak szerokie spektrum testów statystycznych jest dostępnych w niemal każdym programie statystycznym, nie dziwi fakt, iż tak często naukowcy po nie sięgają. Czy zawsze słusznie? A jeśli tak, to czy we właściwy sposób z nich korzystają?

Dogłębna analiza wielu publikacji naukowych pod kątem stosowania testów oraz wieloletnie doświadczenia autora w opracowaniu statystycznym materiałów w ramach zespołów badawczych, pozwala stwierdzić, że istnieje coś takiego jak „kult" istotności statystycznej [17]. Przejawia się on przede wszystkim w bezrefleksyjnym stosowaniu testów na każdym etapie badań i do każdego zagadnienia. Po-

nadto stosuje się je w sposób niewłaściwy, używając ich jak procedur decyzyjnych a nie pomocy w ocenie wiarygodności hipotezy postawionej w badaniu. Czasami można odnieść wrażenie, że nie ważna jest wielkość efektu, a to czy jest on istotny statystycznie. Należy tu przypomnieć, że testy statystyczne powstały na użytek naukowców po to, aby pomóc im w ocenie wiarygodności stawianych hipotez, ale nie miały ich zwolnić z myślenia, czy ów wynik testu oprócz istotności statystycznej ma istotne znaczenie dla rozwijanej dziedziny. Autor najpowszechniej stosowanego testu do porównywania średnich dwóch grup niezależnych, czyli William Sealy Gosset (pseudonim Student)[12], całe życie zajmował się produkcją piwa w browarach Guinnessa w Dublinie. Tworząc podwaliny pod dzisiejszą teorię weryfikacji hipotez, jako praktyk w którego żyłach płynęła krew raczej ekonomisty niż matematyka, poszukiwał metod sprawdzenia jakie konfiguracje składników i metod produkcji dadzą lepsze piwo. Zupełnie inne podejście do testowania hipotez miał uczeń Gosseta, czyli Ronald Aylmer Fisher. Równie genialny, zajmujący się genetyką, rolnictwem i eugeniką, skłaniał się raczej ku temu, aby stworzyć pewien zbiór reguł powalający przyjąć lub odrzucić poddaną ocenie hipotezę. Różnica w podejściu do stosowania testów polegała na tym, że Gosset oprócz wyniku testu w swoich decyzjach brał pod uwagę również czynnik ekonomiczny wiążący się z przeprowadzenie eksperymentu i ewentualnymi jego konsekwencjami, a Fisher raczej podążał w stronę matematycznej formuły, dobrze sformalizowanej, prowadzącej do podjęcia decyzji.

Przeglądając wiele publikacji naukowych, można odnieść wrażenie, że podejście przyświecające Gossetowi nie znalazło wielu naśladowców, również wśród statystyków. Była jednak grupa naukowców, która podobnie jak Student widziała w stosowaniu testów statystycznych jako procedury decyzyjnej duże zagrożenie. Byli wśród nich: Edgeworth, Gosset, Egon Pearson, Jeffreys, Borel, Neyman, Wald, Wolfowitz, Yule, Deming, Yates, L. J. Savage, de Finetti, Good, Lindley, Feynman, Lehmann, DeGroot, Bernardo, Chernoff, Raiffa, Arrow, Blackwell, Friedman, Mosteller, Kruskal, Mandelbrot, Wallis, Roberts, Granger, Leamer, Press, Moore, Ber-

[12] stąd nazwa test t-Studenta

ger, Gigerenzer, Freedman, Rothman, Zellner [17]. Nie oznacza to oczywiście, że sformalizowany opis sposobów weryfikacji hipotez nie ma sensu. To raczej bezrefleksyjne stosowanie testów statystycznych, bez poznania istoty niepewności towarzyszącej każdej procedurze testowej jest bolączką dzisiejszych błędnych jego zastosowań.

Każdy test statystyczny dopuszcza wystąpienie dwóch rodzajów błędów weryfikacyjnych:

- błąd I rodzaju – polega na odrzuceniu prawdziwej hipotezy;
- błąd II rodzaju – polega na przyjęciu fałszywej hipotezy.

Nie jest możliwe jednoczesne minimalizowanie obu błędów, dlatego przyjęto kontrolować prawdopodobieństwo błędu I rodzaju na poziomie[13] $\alpha = 0.05$ i tak konstruować testy, aby błąd II rodzaju był jak najmniejszy.

Prawidłowo przeprowadzona procedura testowa polega na tym, że ustala się hipotezę zerową oraz hipotezę alternatywną. Ta druga jest na ewentualność gdybyśmy byli „zmuszeni" do odrzucenia hipotezy zerowej. Następnie ustala się poziom istotności, czyli dopuszczalnego poziomu błędu I rodzaju dla naszego twierdzenia. Kolejnym krokiem jest wybór testu statystycznego oraz sprawdzenie założeń wymaganych do jego zastosowania. Jeśli założenia są spełnione, to możemy wyznaczyć statystykę testową i sprawdzić czy należy ona do obszaru krytycznego danego testu. W przypadku, gdy statystyka testowa należy do zbioru, zaleca się odrzucenie hipotezy zerowej na korzyść alternatywnej pamiętając, że jest 5% szans, iż decyzja ta jest błędna. Jeśli natomiast statystyka testowa nie leży w obszarze krytycznym, to mówimy, że nie ma podstaw do odrzucenia hipotezy zerowej, co najczęściej jest błędnie utożsamiane z potwierdzeniem hipotezy zerowej. Warto w tym momencie wyznaczyć moc użytego testu, aby się zorientować jakie jest prawdopodobieństwo popełnienia błędu II rodzaju. Na podstawie wielkości badanego efektu, wyniku testu oraz mocy testu możemy podjąć decyzję o przyjęciu lub odrzuceniu hipotezy zerowej. W przypadku niespełnienia założeń testu, nie możemy oprzeć naszego

[13] zwanym poziomem istotności

wnioskowania na poziomie istotności przyjętym w procedurze. Najczęściej zaleca się wówczas poszukać alternatywnego testu. Oczywiście wszystkie założenia dotyczące doboru próby oraz podziału na grupy wymieniane w podrozdziale 2 muszą być również spełnione.

Odnosząc się do przytoczonej powyżej procedury testowania hipotez, należy wspomnieć o błędach, które są najczęściej popełniane. Bardzo częstym błędem w prowadzonych badaniach jest stawianie hipotez po przeprowadzeniu testów ją weryfikujących. Prowadzi to czasem do modyfikacji hipotez na potrzeby wyniku, ponieważ od hipotezy alternatywnej zależy obszar krytyczny testu. Ustalanie poziomu istotności przed rozpoczęciem testowania jest zasadne, ponieważ jeśli wynik testu ujawni, że również przy niższym poziomie istotności uda się odrzucić hipotezę, to niewłaściwą praktyką jest zmniejszanie poziomu istotności przez niektórych badaczy. Inna realizacja tej samej zmiennej losowej w postaci nowej próby, może doprowadzić do zmiany decyzji przy zaostrzonym poziomie istotności.

Wybór testu w procedurze weryfikacyjnej jest często najtrudniejszym zadaniem. Niewłaściwy może skutkować wyciągnięciem złych wniosków. Przy doborze testu należy się kierować kilkoma kryteriami:

1. badanym zagadnieniem (np. czy jest to porównanie średnich, czy badanie niezależności cech jakościowych);

2. skalą pomiarową w której mierzone są interesujące nas wielkości;

3. niezależnością/zależnością pomiarów (należy ustalić czy pomiary odbywały się na tych samych obiektach rozłożone w czasie czy na niezależnych podmiotach/przedmiotach);

4. rozkładem badanych cech;

5. liczebnością prób (choć ten aspekt nie zawsze występuje);

6. mocą testu.

Swego rodzaju standardem testowania hipotez stało się podawanie statystyki testowej oraz krytycznego poziomu istotności[14] zamiast obszaru krytycznego. W zależności od tego czy $p < \alpha$ czy $p > \alpha$, test „każe" odrzucić hipotezę H_0 na korzyść H_1 lub nie daje podstaw do odrzucenia hipotezy H_0. Jednak przed podjęciem decyzji weryfikacyjnej warto sprawdzić jak duża jest moc przeprowadzonego testu oraz oszacować wielkość badanego efektu.

Badając wielkość efektu oddziaływania pewnej zmiennej niezależnej, w trzech sytuacjach test będzie nas wprowadzał w błąd:

1. Jeśli nie ma żadnego faktycznego efektu, ale niedoskonałości techniczne przeprowadzonej procedury weryfikacyjnej doprowadzą do wykazania istotności statystycznej.

2. Istnieje faktyczny efekt oddziaływania, ale ze względu na dużą zmienność badanej cechy nie osiągnęliśmy poziomu istotności statystycznej.

3. Nie istnieje żaden znaczący efekt oddziaływania, ale precyzja oszacowania badanej wartości jest tak duża, że powoduje istotność statystyczną efektu.

Wspominając niedoskonałości techniczne testów, miano na myśli zarówno te wynikające z niepewności towarzyszącej testom statystycznym, jak i braku wiedzy badacza, jaki test będzie optymalny. O ile pierwszej z wymienionych przyczyn nie da się wyeliminować, to upowszechnianie wiedzy z zakresu weryfikacji hipotez oraz poszerzaniu umiejętności jej zastosowania w praktyce, może znacząco obniżyć prawdopodobieństwo wystąpienia pomyłki wynikającej z doboru testu. Sytuacja przedstawiona w drugim punkcie jest problemem, z którym się spotyka wielu badaczy. Dotyczy on zarówno wątpliwości czy publikować wyniki, które nie są „istotne statystycznie" , jak i czy ów brak istotności jest wystarczającym powodem, aby przestać się zajmować danym zagadnieniem. Istnieją udokumentowane przypadki gdy brak istotności statystycznej efektu stanowił zagrożenie dla życia ludzi. Firma farmaceutyczna Merck chcąc wprowadzić na rynek nowy lek (Vioxx) na uśmierzenie

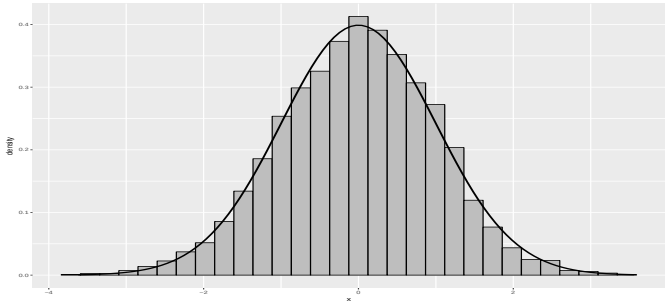
[14] ang. *p-value*

bólu, przeprowadziła testy kliniczne, pod kątem występowania efektów niepożądanych po zażyciu tego leku. Porównywali oni swój produkt z Naproxenem, czyli powszechnie znanym i stosowanym lekiem przeciwbólowym. Opublikowano wyniki testu i okazało się, że brak jest podstaw do orzeczenia, że Vioxx powoduje więcej efektów niepożądanych niż Naproxen. W roku 2003 w Stanach Zjednoczonych na atak serca po zażyciu Vioxxu zmarła kobieta. Po tym incydencie, specjalna komisja oraz niezależni naukowcy, wrócili do wyników badań klinicznych. Okazało się, że po pierwsze, faktycznie nie było istotności statystycznej różnic w frakcji wystąpienia efektów niepożądanych obu leków, ale wyniki jednostkowe pokazały, że Naproxen wywołał 1 atak serca, a Vioxx aż 5. Po drugie, dogłębna analiza dokumentów oraz rozmowy ze świadkami badań wykazały, że producent Vioxxu zataił 3 przypadki śmiertelne po zażyciu tego medykamentu. W tym samym roku lek został wycofany z rynku.

Przykład ten pokazuje, że „kult" istotności statystycznej, narzucany w dobrej wierze, również przez instytucje czuwające na straży naszego zdrowia, doprowadził do tragedii. Prawdopodobnie zbyt mała próba, była powodem tego, że różnica w odsetkach efektów niepożądanych obu leków nie była istotna.

Wspomniana wielkość próby może się znacząco przyczynić do odrzucenia, bądź nieodrzucenia hipotezy zerowej. Istnieje bowiem bardzo silna zależność między liczebnością próby a mocą testu. Niektórzy statystycy mawiają, że nie powinniśmy pytać o istotność statystyczną efektu, bo tę się zawsze da osiągnąć zwiększając wielkość próby. Rzeczywiście mamy czasem do czynienia z sytuacjami co najmniej dziwnymi.

Przykład 3.4. Losując 5000-elementową próbę z rozkładu $N(0,1)$ jesteśmy przekonani, że badana cecha ma rozkład normalny. Potwierdza to kształt histogramu (patrz rysunek 8)
Natomiast testy statystyczne Shapiro-Wilka ($p = 0.01$) i Lillieforsa-Kołmogorowa ($p = 0.04$) tego nie potwierdzają. Oba każą odrzucić hipotezę, że próba pochodzi z rozkładu normalnego. Faktycznie test Shapiro-Wilka jest raczej skierowany do prób o mniejszej liczebności (choć ta z przykładu nie jest niedopuszczalna), ale



Rysunek 8: Histogram próby z rozkładu $N(0,1)$ o liczebności 5000 wraz z funkcją gęstości tego rozkładu

test Lillieforsa powinien sobie poradzić z tym problemem.

Podobnie zaskakujące wyniki możemy uzyskać przy weryfikacji hipotez parametrycznych.

Przykład 3.5. Tym razem wylosujemy dwie próby, jedną z rozkładu $N(0,3)$, a drugą z rozkładu $N(2,3)$. Obie próby będą o liczebności 10. Weryfikując hipotezę $H_0 : \mu_1 = \mu_2$ o tym, że średnie obu populacji są równe otrzymujemy brak istotności statystycznej ($p = 0.606$) stosując odpowiedni w takiej sytuacji test, czyli test t-Studenta dla dwóch prób niezależnych. Średnie obu prób wynoszą odpowiednio $\bar{x}_1 = 0.65$ i $\bar{x}_2 = 1.45$, a odchylenia standardowe $s_1 = 3.95$ i $s_2 = 2.75$. Stosunkowo duża zmienność obu prób oraz niewielka liczebność przekładają się na dużą zmienność różnicy średnich i stąd brak istotności. Natomiast, jeśli tylko zwiększymy próby do liczebności np. 30, to otrzymamy istotność statystyczną testu różnicy. Wówczas ($p = 0.0008$), a estymaty średniej i odchylenia standardowego z prób wynoszą odpowiednio $\bar{x}_1 = -0.89$ i $\bar{x}_2 = 1.83$ oraz $s_1 = 3.15$ i $s_2 = 2.85$. Tym razem odchylenie standardowe w pierwszej próbie jest nieco mniejsze, ale to zwiększona liczebność próby oraz większa różnica miedzy średnimi obu prób, powodują istotność statystyczną różnicy. Oczywiście powiększając sukcesywnie próbę spodziewamy się, że różnica średnich będzie oscylowała wokół 2, a jej odchylenie standardowe będzie malało z powodu powiększania się liczebności, a to będzie skutkowało

jeszcze większa mocą testu. W pierwszym porównaniu moc testu wyniosła 0.08, a w drugim 0.93.

Powyższe przykłady pokazują jak bardzo liczebność próby może wpłynąć na decyzję weryfikacyjną. Dlatego warto przed wykonaniem testu określić wymaganą liczebność próby na podstawie mocy testu (minimum 0.8).

Ostatni rodzaj błędu popełnianego podczas testowania hipotez, który zostanie przytoczony w tym opracowaniu, to brak kontroli błędu I rodzaju. Polega on na błędnym przeświadczeniu, że jeśli się wykona powiedzmy 5 testów różnic między średnimi i każdy z nich jest przeprowadzony na poziomie istotności 0.05, to błąd I rodzaju dla decyzji wynikającej z wszystkich pięciu testów jest też kontrolowany na poziomie 0.05. Niestety, ponieważ testy te były wykonywane niezależnie, to nie można powiedzieć, że prawdopodobieństwo błędu I rodzaju będzie nie większe niż 0.05. W przypadku pięciu porównań, kontrola odbywa się na poziomie 0.23, a np. przy dziesięciu porównaniach 0.4. Dlatego zaleca się stosowanie porównań zaplanowanych, czy zwykłej analizy wariancji zanim przystąpi się do porównywania grup między sobą. Test ANOVA[15] służy do porównania wielu średnich, przy jednoczesnej kontroli błędu I rodzaju na poziomie 0.05.

Podsumowując rozważania na temat zagrożeń czyhających na „użytkownikówű testów statystycznych, należy podkreślić, że konieczna jest stosowna wiedza na temat rodzajów testów i ich założeń. Świadomość niepewności towarzyszącej każdej procedurze weryfikacji hipotez, również pomaga w wyciąganiu wniosków z przeprowadzonych testów. Na koniec jeszcze raz podkreślamy, że testy statystyczne są techniką uwiarygadniania sądów na temat badanych hipotez, nie mogą jednak za nas podejmować decyzji.

3.4 Modelowanie zależności pomiędzy zmiennymi

Codziennie jesteśmy zasypywani danymi o zależności między zmiennymi, czy wpływie czasu na wzrost lub spadek pewnych indeksów. W wielu badaniach na-

[15] ang. *Analysis of Variance*

ukowych rozważa się wpływ zmiennych niezależnych na zmienną wynikową. Siłę związku oraz jego postać wyraża się przy pomocy analizy korelacji i regresji. Pamiętajmy jednak, że faktycznym obiektem zainteresowań wielu naukowców jest związek przyczynowo-skutkowy między zmiennymi, a to nie jest to samo co zależność statystyczna. Istnieje wiele przykładów pozornych związków, które pokazują, że wystąpienie istotnej korelacji między zmiennymi jest jedynie warunkiem koniecznym wystąpienia związku przyczynowo-skutkowego, ale nie dostatecznym. Przykładowo poziom wydatków na naukę w Stanach Zjednoczonych w kolejnych latach jest silnie skorelowany z liczbą samobójstw przez powieszenie ($r = 0.997$). Również liczba ofiar uduszenia przez zaplątanie we własne prześcieradło jest silnie skorelowana z poziomem spożycia sera na przestrzeni lat ($r = 0.94$). Śmiało możemy przypuszczać, że te związki statystyczne nie znajdują potwierdzenia w zależności przyczynowo-skutkowej. Dlatego z tym większą uwagą powinniśmy stosować narzędzia statystyczne do modelowania zależności.

O ile współczynnik korelacji informuje nas o statystycznym związku liniowym pomiędzy cechami, to regresja liniowa pozwala go opisać w postaci równania

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_p x_p + \varepsilon, \tag{3.9}$$

gdzie $x_1, \ldots, x_p$ są zmiennymi niezależnymi modelu, y jest cechą zależną, $\beta_0, \ldots, \beta_p$ są parametrami strukturalnymi modelu, a ε jest niezależnym błędem losowym o rozkładzie $N(0, \sigma)$. Parametr β_i informuje nas o tym o ile wzrośnie wartość y jeśli zmienna x_i zmieni swoją wartość o jeden, przy jednoczesnym zachowaniu pozostałych zmiennych na tym samym poziomie.

Największe zagrożenia, które czyhają na naukowców stosujących analizę regresji czy korelację, to:

1. Ograniczony zakres w przestrzeni cech, na którym model jest prawdziwy.

2. Niewłaściwa specyfikacja modelu (np. gdy szacujemy model liniowy, podczas gdy zależność jest krzywoliniowa).

3. Zmienne zakłócające (brak ich uwzględnienia w modelu może pozorować wpływ pewniej cechy na zmienną wynikową).

4. Brak spełnienia założeń dotyczących modelu.

5. Brak zdolności uogólniania wyników z próby na szerszą populację.

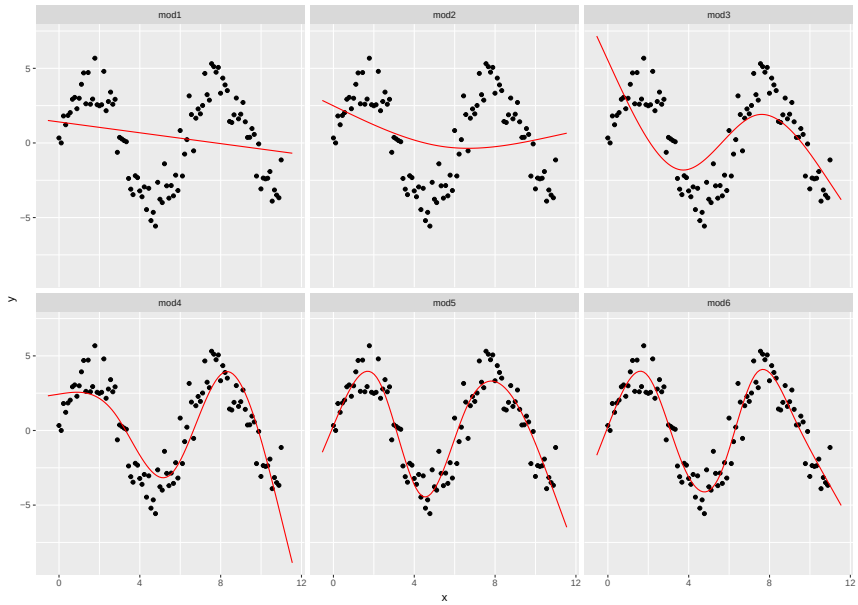
Niebezpieczeństwo opisane w punkcie pierwszym jest związane z faktem, że prawie każde prawo nauki ma pewne ograniczone spektrum swojego działania. O ile dla typowych wartości cech model dobrze opisuje związek między nimi, to dla wartości ekstremalnych nie zawsze. Należy wówczas wyraźnie określić zakres stosowalności modelu, tak, aby nie dokonać predykcji zmiennej wynikowej poza zakresem, gdzie rzeczywista relacja zachodzi.

Niewłaściwa specyfikacja modelu oznacza, że prawdziwy wpływ cech niezależnych na wartość wynikową jest niewłaściwie opisany równaniem regresji. Jeśli liczba cech umożliwia wykreślenie zależności pomiędzy nimi w układzie współrzędnych, to łatwo można dostrzec niedoskonałość własnego modelu, ponieważ wówczas linia regresji jest niedostatecznie dopasowana do danych. Nieco trudniej jest, gdy zbudowaliśmy model regresji wielorakiej, w której rysunek nie jest już pomocny. Wówczas musimy się posłużyć wykresem reszt do wykrycia ewentualnej złej specyfikacji modelu. Jeśli zależność jest dobrze opisana za pomocą równania, to twierdzenie Gaussa-Markowa mówi, że reszty z modelu mają być nieskorelowane, o średniej równej zero i stałej wariancji σ^2. Dodatkowo, aby móc testować istotność statystyczną poszczególnych parametrów, wymagana jest także normalność rozkładu reszt. Jak widzimy, niepoprawna postać zależności wyrażona równaniem regresji jest ściśle związana z założeniami, które muszą być spełnione, aby estymatory parametrów modelu były BLUE[16].

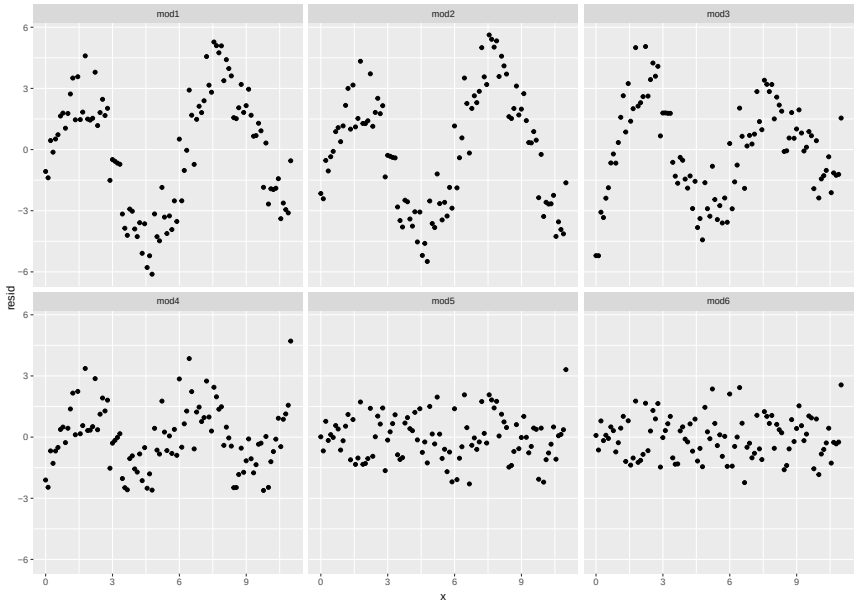
Przykład 3.6. Dla danych sztucznie utworzonych na podstawie równania $y = 4\sin(x)$ zniekształconych niewielkim szumem o rozkładzie normalnym, zbudujemy kilka

[16] ang. *Best Linear Unbiased Estimator*

modeli regresji i ocenimy ich dopasowanie. Ponieważ funkcja sinus da się przybliżyć za pomocą wielomianu odpowiedniego stopnia, to właśnie w tej rodzinie modeli będziemy poszukiwać rozwiązania. Jak widać na rysunku 9 dopasowanie wielomianów stopni od 1 do 3 nie jest wystarczająco dobrym opisem zależności. Dopiero wielomiany stopni od 4 do 6 dobrze opisują ukryty związek. Moglibyśmy zatem zadać pytanie: wielomian, którego stopnia będzie najlepszy? Jest to poniekąd pytanie o właściwą specyfikację modelu. Aby odpowiedzieć na to pytanie, należy uważnie przejrzeć wykresy reszt (patrz rysunek 10). Reszty modeli wielomianowych stopni 5 i 6 nie wykazują autokorelacji reszt, heterogeniczności wariancji, oscylują wokół 0, czyli istotnie mogą spełniać założenia twierdzenia Gaussa-Markowa. Rozkłady reszt przypominają rozkład normalny już od wielomianu 4 stopnia (patrzy rysunek 11). Ostatecznie uwzględniając zasadę, iż model o prostszej postaci jest lepszy, wybralibyśmy dopasowanie za pomocą wielomianu 5 stopnia.



Rysunek 9: Wykresy rozrzutu wraz z dopasowanymi liniami regresji będącymi wielomianami stopni od 1 do 6



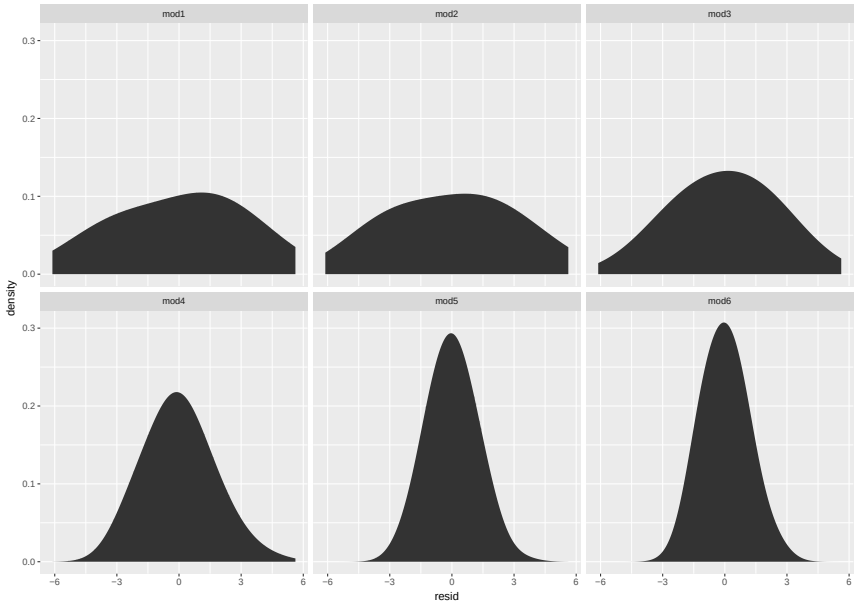
Rysunek 10: Wykresy reszt wszystkich sześciu modeli

Przy okazji tego przykładu pojawiło się jeszcze jedno zjawisko, które w istotny sposób decyduje o możliwościach predykcyjnych modeli, a mianowicie przeuczenie[17] lub niedouczenie[18] modelu. Przeuczenie polega na dopasowaniu modelu zbyt dokładnie do uczącego zbioru danych (czyli zbioru, na podstawie którego szacowane są parametry). Wówczas model oprócz ukrytej zależności pomiędzy zmiennymi, opisuje również część szumu towarzyszącemu niepewności pomiarowej i różnicom osobniczym badanych obiektów. Przeuczenie przejawia się również w włączeniu do modelu zbyt wielu zmiennych niezależnych. Niedouczenie przejawia się w niedostatecznym dopasowaniu modelu do danych zbioru uczącego, wynikającym z niewłaściwej postaci lub braku w modelu kluczowych dla zależności zmiennych. W przykładzie poprzednim wielomiany stopnia 1,2 i 3 wykazywały niedouczenie.

Poszukując optymalnego modelu jesteśmy zawsze zmuszeni do wyboru mię-

[17] ang. *overfitting*

[18] ang. *underfitting*



Rysunek 11: Rozkłady reszt wszystkich sześciu modeli

dzy mniejszym obciążeniem poszczególnych parametrów modelu, a ich wariancją. Stosunkowo niewielkie obciążenie oznacza, że funkcyjna postać zależności opisana równaniem regresji jest zbliżona do prawdziwej. Drugi z wymienionych czynników określa z jaką niepewnością został wyznaczony model. Jeśli wariancja modelu jest niska, to oznacza to, że kolejne replikacje cech ujętych w modelu nie będą znacząco wpływały na jego postać. Przy dużej wariancji parametry modelu potrafią się diametralnie zmieniać od próby do próby. Nie ma sztywnych reguł dotyczących wyboru poziomów obciążenia i wariancji. Decyzję o wielkości obu czynników podejmuje się na podstawie celu badań. Jeśli jest nim wskazanie zmiennych, które mają istotny wpływ na zmienną wynikową, to model z niższą wariancją będzie odpowiedni. Pozostawia się wówczas w modelu jedynie istotne zmienne. Jeśli celem jest sterowanie pewnym procesem opisanym równaniem lub predykcja przyszłych wartości, to często badacze decydują się na model z mniejszym obciążeniem.

Ostatnim problemem, który zostanie podjęty w tym podrozdziale są zmienne zakłócające[19]. W sytuacjach wspomnianych przy okazji pozornych korelacji, mamy często do czynienia ze zmiennymi zakłócającymi. Przykładowo jeśli badamy korelację pomiędzy liczbą oddziałów straży pożarnych, które przyjechały do gaszenia pożaru, a wielkością strat z nim związanych, to może się okazać, że jest ona dodatnia i bardzo wysoka. Wpływ liczby oddziałów na straty jest zaniedbywalny, jeśli dodamy do modelu zmienną wyrażającą wielkość pożaru (zmienna zakłócająca). Pomocą w odnalezieniu takich zmiennych może być dogłębna analiza literatury dotycząca związku analizowanych cech.

Podobny efekt „znikania" istotności pewnych czynników uzyskamy, jeśli włączymy do modelu zmienne nadmiarowe (redundantne), czyli takie, które przekazują do modelu podobne informacje. Szczególnie często zdarza się to w sytuacji, gdy model zawiera wiele zmiennych. Trudno wówczas uniknąć nadmiarowości, ponieważ przy dużej liczbie zmiennych, co najmniej kilka będzie ze sobą silnie skorelowanych, będąc atrybutami badanych obiektów. Zjawisko redundancji jest o tyle niebezpieczne, że oszacowania wariancji parametrów stojących przy zmiennych nadmiarowych są zazwyczaj duże, co nie pozwala na ocenę istotności statystycznej wpływu tych zmiennych. Z pomocą w takich sytuacjach przychodzą techniki redukcji wymiarów, jak PCR[20], czy PLSR[21] oraz metody z grupy regresji regularyzowanej, jak regresja grzbietowa[22], LASSO[23], ElasticNet.

Podsumowując, badając związki pomiędzy cechami należy się wystrzegać wyciągania wniosków o zależności przyczynowo-skutkowej wyłącznie na podstawie wysokiej korelacji. Nawet dobrze wyglądający model regresji, który wydaje się właściwe opisywać wpływ zmiennych niezależnych na zmienną wynikową, może okazać się błędny, ponieważ brak w nim zmiennej zakłócającej, która jest prawdziwą przyczyną ukrytej zależności. Konieczne jest również sprawdzanie założeń, na-

[19] ang. *confounding variables*
[20] ang. *Principal Component Regression*
[21] ang. *Partial Least Squares Regression*
[22] ang. *ridge regression*
[23] ang. *Least Absolute Shrinkage and Selection Operator*

wet, gdy w publikacji się ich nie umieszcza, ponieważ informacja o tym, że warunki użycia metody najmniejszych kwadratów zostały spełnione, może przyspieszyć prace nad rozwojem danej teorii przez innych naukowców. Warto aby najczęściej przytaczana statystyka w kontekście modeli liniowych, czyli R^2 [24], była uzupełniana przez takie parametry jak: statystyka F, błąd standardowy estymacji oraz przedziały ufności dla parametrów β_i. W kontekście modeli ważna jest również ich walidacja, o której będzie traktował kolejny podrozdział.

4 Walidacja wyników badań

Celem stosowania wszelakich technik statystycznych jest osiągnięcie wiarygodnego opisu badanej rzeczywistości w oparciu o wyniki analiz. Dotyczy to estymacji punktowej, przedziałowej, weryfikacji hipotez czy też modeli regresyjnych i klasyfikacyjnych. We wszystkich wspomnianych obszarach konieczne jest weryfikowanie wyników analiz pod kątem ich wrażliwości na zmieniające się dane. Wartość poznawcza płynąca z estymowanego modelu jest istotna, jeśli przy zastosowaniu go na nowym zbiorze danych otrzymamy dopasowanie na poziomie zbliżonym do tego, które otrzymaliśmy w procesie budowania modelu.

Do oceny wrażliwości otrzymanego wyniku lub modelu możemy zastosować wiele technik próbkowania[25] oraz równie wiele miar dopasowania. W zależności od zastosowanej techniki statystycznej do opisu rzeczywistości czy weryfikacji hipotez, stosuje się różne metody ich weryfikacji. Najczęściej stosowaną jest podział próby na część uczącą, na której buduje się modele statystyczne, oraz część testową lub walidacyjną, na której szacuje się dopasowanie modeli.

Walidację modeli statystycznych stosuje się po to, aby sprawdzić czy otrzymane wyniki nie są dziełem przypadku. Może się bowiem zdarzyć, że istotność statystyczna różnic lub efektu oddziaływania pewnego czynnika wynika tylko i wyłącznie ze zbiegu okoliczności. Próba ucząca otrzymana w sposób losowy może

[24] współczynnik determinacji

[25] ang. *resampling*

zawierać dane sprzyjające odrzuceniu lub potwierdzeniu hipotezy, mimo, że trend ten nie utrzymuje się w całej populacji. Jeszcze innym powodem konieczności stosowania walidacji modeli jest fakt, iż część modeli ma tendencje do przeuczania. Polega ono na tym, że poziom dopasowania modelu na zbiorze danych uczących jest wysoki, natomiast dla nowego zestawu danych z tej samej populacji poziom dopasowania spada drastycznie. Oznacza to, że stosowanie tego modelu jako opisu zależności pomiędzy zmiennymi jest mocno wątpliwe.

Niestety wiele badań naukowych przeprowadzanych jest w oparciu o metody ilościowe nie poparte żadną weryfikacją otrzymanych modeli. Nierzadko zdarza się, że ze względu na koszty związane z zebraniem odpowiednio obszernego materiału badawczego, pomija się badanie wrażliwości wyników na różnice osobnicze, czy homogeniczność poszczególnych grup badawczych. Dobrą praktyką w tej materii jest replikacja badań na zupełnie nowej próbie badawczej. Przykładowo, jeśli oszacowany model wykazuje dobre dopasowanie na próbie osób z jednego szpitala, to warto przeprowadzić jego weryfikację na pacjentach innego szpitala.

Bardzo często w procesie budowy i oceny dopasowania modelu stosuje się metody takie jak: k-krotny sprawdzian krzyżowy, wielokrotny podział na próbę uczącą i testową[26] oraz bootstrap. Na etapie budowy modelu metody te pozwalają na korektę parametrów, natomiast w procesie walidacji umożliwiają ocenę dopasowania.

W estymacji punktowej i przedziałowej można stosować metody próbkowania w celu zmniejszenia obciążenia selekcyjnego. Techniki jackknife[27] oraz bootstrap pozwalają oszacować parametry rozkładu zmniejszając obciążenie ale również wyznaczyć przedziały ufności, dla których nie ma klasycznych wzorów.

W weryfikacji hipotez wykorzystuje się za to testy permutacyjne, których zasada działania polega na wyznaczeniu rozkładu statystyki testowej na podstawie prób powstałych przez permutację obiektów. Przykładowo jeśli weryfikujemy hipotezę o równości średnich dwóch populacji, to należy najpierw obliczyć statystykę testową dla oryginalnych danych. Następnie dokonać permutacji zmiennej grupującej

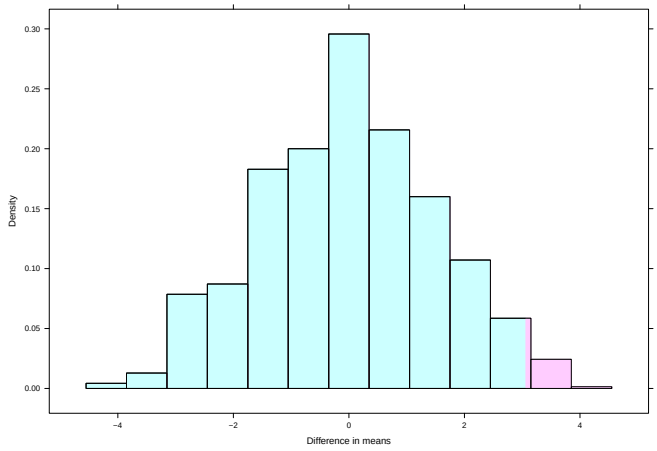
[26] ang. *leave-group-out cross-validation*

[27] polega na uśrednieniu estymatorów uzyskanych z prób, w których usuwany jest kolejno jeden element

(albo zależnej – nie ma to znaczenia) i obliczyć statystykę testową. Powtórzyć tę czynność wielokrotnie (zaleca się 1000 razy), a ponieważ hipoteza zerowa mówi, że średnie w obu grupach są identyczne, to permutacje nie powinny istotnie wpłynąć na statystykę testową, która w przypadku równych wariancji obliczana jest ze wzoru

$$T = \bar{x}_A - \bar{x}_B, \tag{4.1}$$

gdzie $\bar{x}_A, \bar{x}_B$ oznaczają średnie prób A i B. Na koniec wyliczamy p-value (dla hipotezy alternatywnej $H_1 : \mu_A > \mu_B$) jako odsetek tych permutacji, których $T_i^* > T_0$, gdzie T_i^* oznacza statystykę testową i-tej permutacji, a T_0 próby oryginalnej. Metoda ta może być stosowana do statystyk, których rozkładu nie znamy. Warunkiem jej stosowalności jest wymienność obserwacji przy założeniu prawdziwości hipotezy H_0. Dla testu różnic oznacza to równość wariancji.



Rysunek 12: Przykład rozkładu różnic średnich. Różowym kolorem zaznaczone wartości $T_i^* > T_0$.

Nie istnieje najlepsza metoda próbkowania, ponieważ każda z nich ma pewne zalety i ograniczenia. W próbach o małej liczności zaleca się stosowanie 10-krotnego sprawdzianu krzyżowego, ponieważ kontrola obciążenia selekcyjnego oraz wariancji estymatorów jest stosunkowo niewielka, a dodatkowo metoda ta nie jest

wymagająca obliczeniowo. W porównywaniu modeli proponuje się stosowanie bootstrapu zważywszy na fakt, iż losujemy wówczas ze zwracaniem elementy próby, a to obniża wariancje estymatorów. Dla dużych prób różnice pomiędzy technikami próbkowania się zacierają.

Podsumowując, stosowanie weryfikacji uzyskanych wyników jest obowiązkiem każdego badacza. Tylko modele poddane analizie wrażliwości na zmienność, która charakteryzuje opisywane zmienne może posłużyć jako narzędzie predykcyjne lub klasyfikacyjne.

5 Podsumowanie

Mając na uwadze liczbę błędów jakie można znaleźć w publikacjach w zakresie wykorzystania metod ilościowych, należy stwierdzić, że stosowanie statystyki w badaniach naukowych nie jest rzeczą prostą. Niewłaściwy dobór materiału badawczego, metod, czy sposobu wnioskowania należą do najczęściej powtarzających się błędów. Dla poprawienia tego stanu rzeczy, niewątpliwie należy zacieśnić współpracę między naukowcami wielu dziedzin a statystykami. Taka współpraca wydaje się być nieodzowna, ponieważ pozwoli ona na kontrolowanie poprawności przeprowadzanego badania na każdym jego etapie. Niestatystycy tylko w ograniczonym zakresie poznają reguły wnioskowania statystycznego, a paleta metod statystycznych, którymi dysponują jest często bardzo wąska. W połączeniu z brakiem wystarczającej wiedzy na temat planowania i przygotowania eksperymentu, prowadzi to do niewłaściwych wniosków wyciąganych na podstawie badań ilościowych. Jeszcze innym problemem sygnalizowanym w tej pracy jest nieufność środowisk naukowych w stosunku do stosowania technik statystycznych, szczególnie tych bardziej zaawansowanych, znanych mniejszemu gronu naukowców. Jedynym remedium na taki stan rzeczy wydaje się być upowszechnianie wiedzy z zakresu metod ilościowych. Niestety również statystycy będący konsultantami projektów badawczych czy recenzentami wydawniczymi nie zawsze przykładają się należycie do swoich obowiązków. Ulegają oni czasami presji autorów prac, ażeby przymknąć

oko na pewne nieścisłości statystyczne pojawiające się w pracy. Nie ma tu jednak prostych rozwiązań, jedyne co można w tej sytuacji zrobić, to zaapelować o uczciwość merytoryczną przeprowadzanych badań oraz ich weryfikacji. Pozostaje mieć nadzieję, że problemy zasygnalizowane w niniejszej publikacji pobudzą do właściwego podejścia do wnioskowania statystycznego i przełożą się na podniesienie jakości prac naukowych w zakresie wykorzystania technik ilościowych.

Literatura

[1] D. G. Altman. The scandal of poor medical research, *BMJ: British Medical Journal*, 308(6928), 1994, 591–593.

[2] D. G. Altman. Poor-quality medical research: what can journals do? *Jama*, 287(21), 2002, 2765–2767.

[3] T. C. Badrick and R. J. Flatman. The inappropriate use of statistics. *New Zealand Journal of Medical Laboratory Science*, 53(3), 1999, 95–103.

[4] J. O. Berger and T. Sellke. Testing a point null hypothesis: The irreconcilability of p values and evidence. *Journal of the American statistical Association*, 82(397), 1987, 112–122.

[5] J. M. Bland and D. G. Altman. Comparing methods of measurement: why plotting difference against standard method is misleading. *The Lancet*, 346(8982), 1995, 1085–1087.

[6] G. E. P. Box and G. C. Tiao. A note on criterion robustness and inference robustness. *Biometrika*, 51(1/2), 1964, 169–173.

[7] W. G. Cochran. *Sampling Techniques: 3d Ed.* J. Wiley, 1977.

[8] D. R. Cox. Some problems connected with statistical inference. *The Annals of Mathematical Statistics*, 29(2), 1958, 357–372.

[9] T. J. Duggan and C. W. Dean. Common misinterpretations of significance levels in sociological journals. *The American Sociologist*, 1968, 45–46.

[10] P. I. Good and J. W. Hardin. *Common errors in statistics (and how to avoid them)*. John Wiley & Sons, 2012.

[11] J. R. Lisse, M. Perlman, G. Johansson, J. R. Shoemaker, J. Schechtman, C. S. Skalky, M. E. Dixon, A. B. Polis, A. J. Mollen, and G. P. Geba. Gastrointestinal tolerability and effectiveness of rofecoxib versus naproxen in the treatment of osteoarthritis: a randomized, controlled trial. *Annals of internal medicine*, 139(7), 2003, 539–546.

[12] S. Lohr. *Sampling: design and analysis*. Nelson Education, 2009.

[13] United States Department of Agriculture. Census of agriculture. https://www.agcensus.usda.gov/. Dostęp: 1 marca 2018.

[14] R. Rosenthal, R. L. Rosnow, R. Rosenthal, and R. L. Rosnow. The volunteer subject. *Artifacts in behavioral research*, 2009, 48–92.

[15] R. Rosethal and R. L. Rosnow. Essentials of behavioral research: Methods and data analysis, 1991.

[16] R. A. Sugden, T. M. F. Smith, and R. P. Jones. Cochran's rule for simple random sampling. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 62(4), 2000, 787–793.

[17] S. Ziliak and D. Nansen McCloskey. *The cult of statistical significance: How the standard error costs us jobs, justice, and lives*. University of Michigan Press, 2008.

Aktualne problemy w badaniach nad prawami wielkich liczb

Anna Kuczmaszewska[1]

Streszczenie

Prawa wielkich liczb to jeden z obszarów rachunku prawdopodobieństwa bogato reprezentowany w badaniach. Początki tego działu sięgają XVIII wieku, kiedy to Jakob Bernoulli sformułował pierwsze prawo wielkich liczb. Początkowe rozważania dotyczące praw wielkich liczb odnosiły się tylko do niezależnych zmiennych losowych, zbieżności według prawdopodobieństwa i zbieżności z prawdopodobieństwem 1. Statystyczne zastosowania praw wielkich liczb, zastosowania w teorii niezawodności i analizie ryzyka spowodowały zainteresowanie prawami wielkich liczb w odniesieniu do ciągów zmiennych losowych o różnych strukturach zależności. W rozdziale tym przedstawione zostaną aktualne tendencje w obszarze badań nad prawami wielkich liczb z uwzględnieniem różnych struktur zależności oraz różnych, nowych, rodzajów zbieżności takich jak zbieżność kompletna (ang. *complete convergence*) i zbieżność kompletna momentowa (ang. *complete moment convergence*).

Abstract

The law of large numbers is a group of theorems in probability theory which are richly represented in literature, The first law of large numbers was formulated by Jakob Bernoulli in the eighteenth century. The fundamental

[1]Katedra Matematyki Stosowanej; Politechnika Lubelska
e-mail: a.kuczmaszewska@pollub.pl

considerations regarding the laws of large numbers relate to independent random variables, convergence in probability and convergence with probability 1. Statistical applications of laws of large numbers, applications in the theory of reliability and risk analysis have resulted interest in the laws of large numbers for sequences of random variables with different dependence structure In this chapter there will be presented the current trends in the field of research on the laws of large numbers taking into account the various structures of dependence and the various new types of convergence such as complete convergence and complete moment convergence.

Słowa kluczowe: prawo wielkich liczb, zbieżność kompletna, zbieżność kompletna momentowa, zależne zmienne losowe

1 Wstęp

Prawa wielkich liczb to grupa twierdzeń w teorii prawdopodobieństwa, które opisują związki podstawowych pojęć (wielkości) probabilistycznych z ich intuicyjnymi oszacowaniami występującymi w zastosowaniach praktycznych. Do takich pojęć należą np. prawdopodobieństwo i jego intuicyjny estymator – częstość względna oraz wartość oczekiwana i jej estymator – wartość przeciętna z próby.

Pierwsze prawo wielkich liczb sformułował Jakob Bernoulli w roku 1713: *Z prawdopodobieństwem dowolnie bliskim 1 można się spodziewać, iż przy dostatecznie wielkiej liczbie prób częstość danego zdarzenia losowego będzie się dowolnie mało różniła od jego prawdopodobieństwa*.

Twierdzenie to, Bernoulli nazwał *Złotym Twierdzeniem*, a w literaturze pozostało pod nazwą *Twierdzenie Bernoulliego*. Nazwa *Prawo wielkich liczb* pojawiła się dopiero w roku 1835 za sprawą Simeona Denisa Poissona.

Zdefiniowanie dokładnego obszaru zainteresowania twierdzeń typu prawa wielkich liczb nie jest proste. Najogólniej, mówimy, że prawo wielkich liczb zapewnia zbieżność, w pewnym określonym sensie, średniej arytmetycznej

$$\xi_n = \frac{X_1 + X_2 + \cdots + X_n}{b_n}$$

zmiennych losowych $X_1, X_2, \ldots, X_n$ do zmiennej losowej X, gdy $n \to \infty$ zaś $\{b_n, n \geqslant 1\}$ jest rosnącym ciągiem liczb dodatnich rozbieżnym do nieskończoności.

Klasa twierdzeń nazywanych prawami wielkich liczb zawiera obecnie wiele twierdzeń, które nie odnoszą się już bezpośrednio do samego pojęcia prawdopodobieństwa. Zaliczyć do nich możemy np. twierdzenia o szeregach Fouriera sumowalnych w sensie Cesaro. Dlatego mówimy, że prawa wielkich liczb to twierdzenia zapewniające zbieżność ciągu $\{\xi_n, n \geqslant 1\}$ uwzględniające różne aspekty prawdopodobieństwa.

Prawa wielkich liczb możemy rozważać biorąc pod uwagę różne rodzaje zbieżności i otrzymywać różne typy praw (mocne prawa wielkich liczb lub słabe prawa wielkich liczb).

Grupa twierdzeń określanych mianem słabego prawa wielkich liczb, to twierdzenia, w których uwzględniana jest tzw. słaba zbieżność, tzn. zbieżność według prawdopodobieństwa.

Pierwsze prawo wielkich liczb sformułowane przez Bernoulliego wyrażone we współczesnej terminologii ma następująca postać.

Twierdzenie 1.1. *Jeśli S_n jest liczbą sukcesów w schemacie n niezależnych prób Bernoulliego z prawdopodobieństwem sukcesu w pojedynczej próbie równym p, to dla każdego $\varepsilon > 0$*

$$\lim_{n\to\infty} P\big[|\frac{S_n}{n} - p| \leqslant \varepsilon\big] = 1. \tag{1.1}$$

Oznacza to, że niezależnie od wyboru szerokości ($\varepsilon > 0$) przedziału wokół wartości oczekiwanej p, prawdopodobieństwo dla dostatecznie dużych n będzie dowolnie bliskie 1.

Liczbę sukcesów S_n możemy wyrazić jako sumę n niezależnych zmiennych losowych X_i, $1 \leqslant i \leqslant n$ o rozkładzie zero-jedynkowym

$$P\big[X_i = 1\big] = p = 1 - P\big[X_i = 0\big], \quad 1 \leqslant i \leqslant n.$$

Oznacza to, że twierdzenie 1.1 możemy traktować jako szczególny przypadek słabego prawa wielkich liczb Markowa.

Twierdzenie 1.2. *Jeśli* $\{X_n, n \geqslant 1\}$ *jest ciągiem niezależnych zmiennych losowych spełniających warunek Markowa*

$$\lim_{n\to\infty} \frac{\sum_{i=1}^{n} Var(X_i)}{n^2} = 0,$$

to spełniony jest warunek (1.1).

Twierdzenie Markowa rozszerza klasę ciągów spełniających słabe prawo wielkich liczb na przypadek ciągów o różnych rozkładach.

Można je również rozszerzyć na klasę ciągów $\{X_n, n \geqslant 1\}$ parami niezależnych zmiennych losowych.

Drugą ważną grupę twierdzeń typu prawa wielkich liczb stanowią twierdzenia dotyczące zbieżności prawie pewnej. Do klasycznych twierdzeń z tej grupy zaliczamy mocne prawo wielkich liczb Kołmogorowa i twierdzenia Kołmogorowa.

Twierdzenie 1.3. *Niech* $\{X_n, n \geqslant 1\}$ *będzie ciągiem niezależnych zmiennych losowych z wartością oczekiwaną* $EX_n = 0$ *i skończoną wariancją* $Var(X_n) < \infty$, $n \geqslant 1$, *spełniającym warunek*

$$\sum_{n=1}^{\infty} \frac{Var(X_n)}{n^2} < \infty. \tag{1.2}$$

Wtedy

$$\frac{\sum_{i=1}^{n} X_i}{n} \longrightarrow 0 \ \ p.p., \ n \longrightarrow \infty. \tag{1.3}$$

Twierdzenie Kołmogorowa jest uogólnieniem mocnego prawa wielkich liczb Kołmogorowa (twierdzenie 1.3) polegającym na zastąpieniu warunku (1.2) ogólniejszym

$$\sum_{n=1}^{\infty} \frac{Var(X_n)}{b_n^2} < \infty \tag{1.4}$$

z ogólniejszą tezą

$$\frac{\sum_{i=1}^{n} X_i}{b_n} \longrightarrow 0 \ \ p.p., \ n \longrightarrow \infty, \tag{1.5}$$

gdzie $\{b_n, n \geqslant 1\}$ jest rosnącym do nieskończoności ciągiem liczb dodatnich.

Warunek Kołmogorowa (1.2) jest jednym z bardziej znanych warunków dostatecznych mocnego prawa wielkich liczb, może być jednak stosowany wyłącznie do ciągów niezależnych zmiennych losowych ze skończonymi wariancjami.

Chung [9] w pracy z 1947 roku osłabił założenie o skończonych wariancjach i zaproponował zastąpienie go warunkiem $E\varphi(|X_n|) < \infty$, $n \geqslant 1$, gdzie φ jest pewną funkcją ciągłą, określoną w przedziale $\langle 0, \infty)$, nieujemną i taką, że $\lim_{n\to\infty} \varphi(x) = \infty$ oraz spełniającą jeden z warunków

$$\Big(\varphi(x)/x\Big) \searrow \quad \text{lub} \quad \Big(\varphi(x)/x\Big) \nearrow \text{ i } \Big(\varphi(x)/x^2\Big) \searrow, \quad \text{gdy} \quad x \longrightarrow \infty.$$

Przy tych założeniach uogólnił twierdzenie Kołmogorowa pokazując, że jeśli ciąg $\{X_n, n \geqslant 1\}$ niezależnych zmiennych losowych z zerową wartością oczekiwaną spełnia warunek

$$\sum_{n=1}^{\infty} \frac{E\varphi\Big(|X_n|\Big)}{\varphi(n)} < \infty,$$

to $\{X_n, n \geqslant 1\}$ spełnia mocne prawo wielkich liczb, gdy $b_n = n$, $n \geqslant 1$.

Twierdzenie Chunga pozwala rozważać mocne prawo wielkich liczb dla szerszej klasy ciągów niezależnych zmiennych losowych niż twierdzenie Kołmogorowa. Jednak w przypadku ciągów zmiennych losowych o wariancjach $Var(X_n) = \sigma_n^2 \sim n/(\log n)^\delta$, $0 < \delta \leqslant 1$ oba wspomniane wyżej kryteria nie mogą być stosowane z uwagi na rozbieżność szeregu Kołmogorowa z warunku (1.2). Przykład takiego ciągu rozważał Teicher [45] i zaproponował zastąpienie warunku Kołmogorowa warunkiem ogólniejszym

$$\sum_k := \sum_{j_k=k}^{\infty} j_k^{-2k} \sigma_{j_k}^2 \sum_{j_{k-1}=k-1}^{j_k-1} \sigma_{j_{k-1}}^2 \cdots \sum_{j_1=1}^{j_2-1} \sigma_{j_1}^2 < \infty, \quad k \geqslant 2,$$

który, przy pewnych dodatkowych założeniach, zapewnia prawdziwość mocnego prawa wielkich liczb dla ciągu $\{X_n, n \geqslant 1\}$ gdy $\sigma^2(X_n)$ jest rzędu $n/(\log n)^\delta$.

Warunek Kołmogorowa (1.2) może być rozpatrywany w takim ujęciu jako pierwszy wyraz ciągu warunków $\{\sum_k, k \geqslant 1\}$. Twierdzenie Teichera rozszerza klasę ciągów, o których możemy twierdzić, że spełniają mocne prawo wielkich liczb,

jednak odnosi się wyłącznie do **niezależnych** zmiennych losowych o **skończonych drugich momentach**.

Kolejną grupą twierdzeń typu prawa wielkich liczb są tzw. prawa wielkich liczb typu Hsu-Robbinsa. Są to twierdzenia, w których badana jest kompletna zbieżność ciągu $\{\xi_n, n \geqslant 1\}$.

Pojęcie kompletnej zbieżności wprowadzili Hsu i Robbins [21] w roku 1947. Oni też zapoczątkowali badanie tej zbieżności w kontekście prawa wielkich liczb.

Definicja 1.1 ([21]). *Mówimy, że ciag $\{X_n, n \geqslant 1\}$ jest zbieżny kompletnie do zmiennej X jeśli*

$$\sum_{n=1}^{\infty} P(|X_n - X| > \varepsilon) < \infty,$$

dla dowolnego $\varepsilon > 0$.

Na mocy lematu Borela-Cantelliego zbieżność kompletna implikuje zbieżność prawie pewną, tym samym twierdzenia typu prawa wielkich liczb otrzymane dla zbieżności kompletnej są mocniejsze i implikują odpowiednie wyniki dla zbieżności prawie pewnej i zbieżności według prawdopodobieństwa.

Hsu i Robbins dowiedli następującego twierdzenia.

Twierdzenie 1.4 ([21]). *Niech $\{X, X_n, n \geqslant 1\}$ będzie ciągiem niezależnych zmiennych losowych o jednakowym rozkładzie takich, że $EX = 0$ i $EX^2 < \infty$. Wtedy*

$$\forall \varepsilon > 0 \quad \sum_{n=1}^{\infty} P\Big[|S_n| > \varepsilon n\Big] < \infty.$$

Erdős [13] udowodnił ponadto, że warunek $EX = 0$ i $EX^2 < \infty$ jest również warunkiem koniecznym dla prawa wielkich liczb typu Hsu-Robbinsa dla ciągu niezależnych zmiennych losowych o jednakowym rozkładzie.

Wynik Hsu-Robbinsa-Erdősa jest jednym z podstawowych twierdzeń w obszarze prawa wielkich liczb. Był on wielokrotnie uogólniany w różnych aspektach przez wielu autorów. Na przykład Katz [24] dowiódł, że jeśli $\{X, X_n, n \geqslant 1\}$ jest

ciągiem niezależnych zmiennych losowych o jednakowym rozkładzie z wartością oczekiwaną $EX = 0$ i $E|X|^{2p} < \infty$ dla pewnego $1 \leqslant p < 2$, to

$$\sum_{n=1}^{\infty} P\Big[|S_n| > \varepsilon n^{1/p}\Big] < \infty \quad \forall \varepsilon > 0.$$

Innym aspektem uogólniania wyniku Hsu i Robbinsa jest badanie tzw. szybkości zbieżności w prawie wielkich liczb tzn. badanie, jak szybko wielkości takie jak $P\big[|S_n| \geqslant n^{\alpha} \cdot \varepsilon\big]$ dążą do zera, jeśli zmienne losowe $X_n, n \geqslant 1$ mają momenty rzędu większego od zera. Można to robić poprzez bezpośrednie badanie tych wielkości lub przez badanie zbieżności sum typu $\sum_{n=1}^{\infty} c_n P\big[|S_n| \geqslant n\varepsilon\big]$, gdzie $\{c_n, n \geqslant 1\}$ jest ciągiem stałych dodatnich. Problem ten był początkowo rozważany głównie w przypadku zmiennych losowych o jednakowym rozkładzie. Wynik Hsu i Robbinsa [21] oraz Erdősa [13] jest rozwiązaniem tak postawionego problemu, gdy $c_n = n^t$ i $t = 0$. W przypadku $t = -1$ problem rozwiązał Spitzer [44]. Szybkością zbieżności w prawie wielkich liczb dla ciągu niezależnych zmiennych losowych o jednakowym rozkładzie zajmował się również Gut [14–17], który w swoich pracach dowiódł kilku twierdzeń dotyczących warunków koniecznych i dostatecznych dla podciągów oraz dla ciągów niezależnych zmiennych losowych o jednakowym rozkładzie, również dla zmiennych losowych o wskaźnikach wielowymiarowych.

Jednym z bardziej znanych uogólnień wyniku Hsu-Robbinsa-Erdősa w kontekście problemu szybkości zbieżności jest twierdzenie z pracy Guta [18] będące połączeniem wyników Katza [24], Bauma i Katza [2] oraz Chowa [7] z normalizacją typu Marcinkiewicza-Zygmunda. Pokazuje ono równoważność pomiędzy założeniem o istnieniu p-tego momentu wyrazów ciągu $\{X, X_n, n \geqslant 1\}$ niezależnych zmiennych losowych o jednakowym rozkładzie, a zbieżnością szeregów:

$$\sum_{n=1}^{\infty} n^{\alpha p-2} P[|\sum_{i=1}^{n} X_i| > n^{\alpha}\varepsilon],$$

$$\sum_{n=1}^{\infty} n^{\alpha p-2} P[\max_{k \leqslant n} |\sum_{i=1}^{k} X_i| > n^{\alpha}\varepsilon]$$

i

$$\sum_{n=1}^{\infty} n^{\alpha p-2} P[\sup_{k \geqslant n} k^{-\alpha} |\sum_{i=1}^{k} X_i| > \varepsilon],$$

dla dowolnego $\varepsilon > 0$ i $\alpha p \geqslant 1$, $\alpha > \frac{1}{2}$.

Problem kompletnej zbieżności w prawie wielkich liczb rozważany jest również w odniesieniu do tablic $\{X_{ij}, i \geqslant 1, j \geqslant 1\}$ zmiennych losowych. Np. Hu, Móricz i Taylor [22] pokazali, że warunkiem koniecznym i dostatecznym na to, by ciąg średnich

$$n^{-\frac{1}{p}} \sum_{k=1}^{n} X_{nk}$$

tablicy zmiennych losowych $\{X_{nk}, n \geqslant 1, 1 \leqslant k \leqslant n\}$ dążył kompletnie do zera, gdy n dąży do nieskończoności i $1 \leqslant p < 2$, jest by $E|X_{11}|^{2p} < \infty$.

Wynik ten jest nie tylko bezpośrednim uogólnieniem twierdzenia Hsu-Robbinsa-Erdősa na przypadek dwuwymiarowych tablic zmiennych losowych, ale dodatkowo wprowadza do rozważań normalizację typu Marcinkiewicza-Zygmunda.

We wszystkich cytowanych wyżej wynikach występowało założenie o niezależności zmiennych losowych, ich jednakowym rozkładzie oraz założenie o istnieniu momentu rzędu $r > 1$.

Statystyczne zastosowania praw wielkich liczb wymagają odpowiedzi na dwa ważne pytania: o szybkość zbieżności częstości względnych lub średnich z próby do wartości prawdopodobieństwa lub odpowiednio wartości oczekiwanej oraz o zachowanie się średnich z próby w przypadku, gdy elementy w tej próbie nie mają jednakowego rozkładu lub są zależne.

Rozważania poświęcone zależnym zmiennym losowym są szczególnie istotne ze względu na fakt, że nie zawsze elementy w badanej próbie są niezależne, a ponadto, nawet jeśli takie są, to sprawdzenie tego faktu bywa czasami niemożliwe lub bardzo skomplikowane. Stąd w bardzo wielu statystycznych modelach założenie niezależności zmiennych losowych jest zastępowane założeniem o występowaniu określonego typu zależności.

Problem ten znajduje odzwierciedlenie w aktualnych rozważaniach dotyczących prawa wielkich liczb. Rozważane są różne struktury zależności, spośród których szczególne zainteresowanie koncentruje się na zależnościach typu ujemnego i zależnościach typu mieszającego.

Jeszcze inną tendencją, jaką można zaobserwować w badaniach nad prawami wielkich liczb, jest poszukiwanie nowych rodzajów zbieżności zmiennych losowych i próba odpowiedzi na pytanie, jak dla tych nowych, często mocniejszych rodzajów zbieżności, zachowują się znane klasyczne wyniki dotyczące praw wielkich liczb.

Celem niniejszej pracy jest przedstawienie aktualnych kierunków badań nad prawami wielkich liczb, które uwzględniają różne struktury zależności oraz różne rodzaje zbieżności zmiennych losowych.

2 Zmienne losowe zależne

Rezygnacja z założenia o niezależności zmiennych losowych nasuwa oczywiste pytanie, czy rozważane zmienne losowe (cechy statystyczne) tworzą jakąś strukturę zależności, czy występują pomiędzy nimi jakieś związki, które można określić za pomocą pewnej zależności funkcyjnej. W literaturze znane są różne typy takich struktur.

Jedną z nich jest zależność typu ujemnego. W zastosowaniach praktycznych bardzo często przyrosty jednych zmiennych losowych wywołują spadek wartości innych zmiennych losowych, stąd w wielu przypadkach założenie o niezależności zastępowane jest bardziej odpowiednim założeniem o występowaniu ujemnej zależności. Do takich przypadków zaliczyć możemy losowanie bez zwracania ze skończonej populacji, jak również procedury rangowania w przypadku statystyk nieparametrycznych. Pojęcie ujemnej zależności wprowadził Lehmann [27] w roku 1966 dla przypadku dwóch zmiennych losowych, na przypadek większej liczby zmiennych uogólnili je Ebrahimi i Ghosh [12].

Definicja 2.1. *Zmienne losowe $X_1, X_2, \ldots, X_n$ są ujemnie zależne (ND), jeśli*

$$P\Big[X_1 \leqslant x_1, \ldots, X_n \leqslant x_n\Big] \leqslant \prod_{i=1}^{n} P\Big[X_i \leqslant x_i\Big]$$

oraz

$$P\Big[X_1 > x_1, \ldots, X_n > x_n\Big] \leqslant \prod_{i=1}^{n} P\Big[X_i > x_i\Big]$$

dla dowolnych $x_1, \ldots, x_n \in \mathbb{R}$. Nieskończona rodzina zmiennych losowych jest ujemnie zależna, jeśli każda skończona podrodzina jest ujemnie zależna.

Szczególnym przypadkiem zależności ujemnej jest ujemne stowarzyszenie. Pojęcie to zostało wprowadzone w pracach: Alam i Saxena [1] oraz Joag-Dev i Proschan [23]. Ważną cechą tej struktury jest to, że niemalejące funkcje na rozłącznych podzbiorach ujemnie stowarzyszonych zmiennych losowych zachowują ujemne stowarzyszenie. Takiej własności nie mają inne typy zależności ujemnej jak np. zależność ujemna dolna, zależność ujemna górna itp. Wiele spośród znanych rozkładów wielowymiarowych ma własność ujemnego stowarzyszenia, np.: wielowymiarowy rozkład hipergeometryczny, ujemnie skorelowany rozkład normalny, rozkład wielomianowy, wielowymiarowy rozkład Dirichleta.

Definicja 2.2. *Skończona rodzina zmiennych losowych $\{X_i, 1 \leqslant i \leqslant n\}$ jest ujemnie stowarzyszona, jeśli dla każdej pary A, B rozłącznych i niepustych podzbiorów zbioru $\{1, 2, \ldots, n\}$ oraz dowolnych rzeczywistych i niemalejących współrzędnościowo funkcji f_1 i f_2*

$$Cov\big(f_1(X_i, i \in A), f_2(X_j, j \in B)\big) \leqslant 0,$$

gdzie f_1 i f_2 są takie, że powyższa kowariancja istnieje.

Nieskończona rodzina zmiennych losowych jest ujemnie stowarzyszona, jeśli każda jej skończona podrodzina jest ujemnie stowarzyszona.

Ze względu na szerokie zastosowanie w wielowymiarowej analizie statystycznej, teorii niezawodności i analizie ryzyka, ujemne stowarzyszenie jest ważnym rodzajem zależności.

Struktura ujemnego stowarzyszenia jest dość dobrze zbadana i opisana. Najważniejsze własności tej struktury zostały opisane w pracy Joag-Dev and Proschan [23], Newman [33] dowiódł centralnego twierdzenia granicznego dla ciągu ujemnie stowarzyszonych zmiennych losowych, Matuła [31] twierdzenia o trzech szeregach, a Shao [39] nierówności typu Rosenthala oraz twierdzenia o kompletnej zbieżności w prawie wielkich liczb, zaś Liang [28] wykazał prawdziwość mocnego prawa wielkich liczb w wersji dla zbieżności kompletnej średnich ważonych.

Wynik otrzymany w 2000 roku przez Shao [39] możemy dziś traktować jako klasyczny, będący zarazem uogólnieniem wyniku Hsu-Robbinsa [21] na przypadek ujemnie stowarzyszonych zmiennych losowych.

Twierdzenie 2.1 ([39]). *Niech $\frac{1}{2} < \alpha \leqslant 1$ i $\alpha p \leqslant 1$. Jeśli $\{X_n, n \geqslant 1\}$ jest ciągiem ujemnie stowarzyszonych zmiennych losowych o jednakowym rozkładzie i $EX_1 = 0$, to następujące warunki sa równoważne:*

(i) $\quad E|X_1|^p < \infty,$

(ii) $\quad \sum_{n=1}^{\infty} n^{\alpha p-2} P\big[\max_{1\leqslant j\leqslant n} |S_j| > \varepsilon n^{\alpha}\big] < \infty$ *dla dowolnego* $\varepsilon > 0,$

gdzie $S_n = \sum_{i=1}^{n} X_i.$

W roku 2011 Gut i Stadtmüller [19] dowiedli twierdzenia dotyczącego kompletnej zbieżności dla niezależnych zmiennych losowych o jednakowym rozkładzie dla ciągów $\{X, X_n, n \geqslant 1\}$ spełniających warunki:

$$EX = 0 \quad \text{i} \quad E\exp\big(\ln^{\alpha}|X|\big) < \infty, \quad \text{dla} \quad \alpha > 1. \tag{2.1}$$

Wynik ten został uogólniony na przypadek zmiennych losowych ujemnie stowarzyszonych przez Wu i Jianga [49].

Twierdzenie 2.2. *Niech $\{X, X_n, n \geqslant 1\}$ będzie ciągiem ujemnie stowarzyszonych zmiennych losowych o jednakowym rozkładzie spełniajacych warunki (2.1) dla $\alpha > 1$. Wtedy*

$$\sum_{n=1}^{\infty} \exp\big(\ln^{\alpha} n\big) \frac{\ln^{\alpha-1} n}{n^2} P\big[\max_{1\leqslant k\leqslant n} |S_k| > n\beta\big] < \infty \quad \textit{dla} \quad \beta > 1. \tag{2.2}$$

Jeśli warunek (2.2) jest spełniony dla $\beta > 0$, to

$$E\exp\bigl(\ln^{\alpha}|X/(2\beta)|\bigr) < \infty.$$

Ponadto jeśli $\beta \leqslant 1/2$, to

$$E\exp\bigl(\ln^{\alpha}|X|\bigr) < \infty$$

i jeśli $\beta > 1/2$, to

$$E\exp\bigl((1-\lambda)\ln^{\alpha}|X|\bigr) < \infty$$

dla dowolnego $\lambda > 0$.

Inną, niedawno zdefiniowaną, strukturą zależności, zaliczaną do grupy struktur ujemnie zależnych, jest struktura rozszerzonych ujemnie zależnych zmiennych losowych (ang. *extended negatively dependent*) (END). Została ona po raz pierwszy wprowadzona w 2009 roku przez Liu [29] w następujący sposób.

Definicja 2.3. *Mówimy, że zmienne losowe $X_k, k = 1, \cdots, n$ są:*

(i) oddolnie ujemnie zależne w sposób rozszerzony (lower extended negatively dependent) (LEND), jeśli istnieje taka stała $M > 0$, że dla wszystkich liczb rzeczywistych $x_k, k = 1, \cdots, n$,

$$P\left\{\bigcap_{k=1}^{n}(X_k \leqslant x_k)\right\} \leqslant M\prod_{k=1}^{n}P\{X_k \leqslant x_k\}; \tag{2.3}$$

(ii) odgórnie ujemnie zależne w sposób rozszerzony (upper extended negatively dependent) (UEND), jeśli istnieje taka stała $M > 0$, że dla wszystkich liczb rzeczywistych $x_k, k = 1, \cdots, n$,

$$P\left\{\bigcap_{k=1}^{n}(X_k > x_k)\right\} \leqslant M\prod_{k=1}^{n}P\{X_k > x_k\}; \tag{2.4}$$

(iii) ujemnie zależne w sposób rozrzerzony (extended negatively dependent) (END), jeśli są jednocześnie LEND i UEND.

Ciąg $\{X_k, k = 1, 2, \cdots\}$ jest LEND/UEND/END jeśli istnieje taka stała $M > 0$, że dla dowolnej liczby naturalnej n, zmienne losowe $X_1, X_2, \ldots, X_n$ są odpowiednio LEND/UEND/END.

W przypadku gdy $M = 1$, warunek dla struktury ujemnie zależnych zmiennych losowych w sposób rozszerzony jest warunkiem określającego strukturę zmiennych losowych ujemnie zależnych zdefiniowaną w pracy Joag-Dev i Proschan [23]. Wykazali oni, że struktura ujemnego stowarzyszenia jest strukturą z zależnością typu ujemnego, a tym samym, że ujemnie stowarzyszone zmienne losowe tworzą strukturę zawierającą się w strukturze zmiennych losowych ujemnie zależnych w sposób rozszerzony (END).

Struktura END zawiera nie tylko wiele znanych struktur typu ujemnej zależności, ale co ciekawe, również niektóre struktury typu zależności dodatniej. Stąd badanie granicznego zachowania się zmiennych losowych o ujemnie zależnej w sposób rozszerzony strukturze ma istotne znaczenie. Literatura dotycząca tej struktury jest bardzo bogata i stale pojawiają się nowe pozycje. Na przykład Liu [29] przedstwiał wyniki dotyczace problemu wielkich odchyleń dla zmiennych losowych z ciężkimi ogonami, Shen [40] pewne nierówności probabilistyczne dostarczające narzędzi do badania tej struktury, Qiu i inn. [37], Wang i inn. [47] i Shen i inn. [41] rozważali kompletną zbieżność dla zmiennych losowych END, a Chen i inn. [6] oraz Yan [50, 51] zajmowali się mocnym prawem wielkich liczb dla tych zmiennych losowych.

Do grupy struktur typu ujemnego zaliczana jest również struktura typu NSD (ujemnej superaddytywnej zależności). Istota tej zależności wiąże się z pojęciem funkcji superaddytywnej.

Definicja 2.4. *Funkcję* $\phi : \mathbb{R}^n \longrightarrow \mathbb{R}$ *nazywamy superaddytywną, jeśli*

$$\phi(\mathbf{x} \vee \mathbf{y}) + \phi(\mathbf{x} \wedge \mathbf{y}) \geqslant \phi(\mathbf{x}) + \phi(\mathbf{y}),$$

dla dowolnych $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$*, gdzie symbole* $\vee$ *i* $\wedge$ *oznaczają odpowiednio maksimum i minimum po współrzędnych dla* $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$*.*

Definicję ujemnej superaddytywnej zależności wprowadził Hu [20] w roku 2000.

Definicja 2.5. *Zmienne losowe $X_1, X_2, \ldots, X_n$ są ujemnie superaddytywnie zależne (NSD), jeśli istnieje taka superddytywna funkcja $\phi : \mathbb{R}^n \longrightarrow \mathbb{R}$, że*

$$E\phi(X_1, X_2, \ldots, X_n) \leqslant E\phi(X_1^*, X_2^*, \ldots, X_n^*), \tag{2.5}$$

gdzie zmienne losowe $X_1^*, X_2^*, \ldots, X_n^*$ są niezależne i takie, że X_i^* i X_i mają taki sam rozkład dla każdego $1 \leqslant i \leqslant n$ oraz istnieją wszystkie wartości oczekiwane w (2.5).

Ciąg zmiennych losowych $\{X_n, n \geqslant 1\}$ jest ujemnie superaddytywnie zależny (NSD), jeśli dla dowolnego $n \geqslant 1$ zmienne losowe $X_1, X_2, \ldots, X_n$ są ujemnie superaddytywnie zależne .

Definicja kolejnej nowej struktury pojawiła się w wyniku rozważań nowego niestandardowego modelu ryzyka, w którym zmienne losowe opisujące wielkości roszczeń i zmienne losowe wyrażające czasy pomiędzy kolejnymi zgłoszeniami roszczenia tworzą ciągi o zależnych wewnętrznie strukturach, lecz wzajemnie niezależne jeden od drugiego. Definicję tej struktury podali w 2013 roku Wang i inn. [46]. Nazwali ją strukturą szeroko zależną typu orthant (widely orthant dependent) (WOD). Struktura ta jest rozszerzeniem struktury END. Uogólnienie polega na zastąpieniu stałych M w warunkach (2.3) i (2.4) ciągami, odpowiednio $\{g_L(n), n \geqslant 1\}$ i $\{g_U(n), n \geqslant 1\}$. W szczególnych przypadkach ciągi te mogą być definiowane następująco:

$$g_L(n) = \sup_{x_i \in (-\infty,\infty), 1 \leqslant i \leqslant n} \frac{P\{\bigcap_{k=1}^n (X_k \leqslant x_k)\}}{\prod_{k=1}^n P\{X_k \leqslant x_k\}}, \quad n \geqslant 1$$

i

$$g_U(n) = \sup_{x_i \in (-\infty,\infty), 1 \leqslant i \leqslant n} \frac{P\{\bigcap_{k=1}^n (X_k > x_k)\}}{\prod_{k=1}^n P\{X_k > x_k\}}, \quad n \geqslant 1.$$

Bliższa analiza tej struktury pozwala zauważyć, że pokrywa ona nie tylko wiele struktur typu ujemnego, ale również wybrane struktury dodatnio zależnych zmiennych losowych. Poprzez nałożenie różnych warunków na ciągi $\{g_L(n), n \geqslant 1\}$ i $\{g_U(n), n \geqslant 1\}$ możemy rozważać ciągi $\{X_n, n \geqslant 1\}$ ze strukturą zależności WOD w kontekście prawa wielkich liczb.

Innym sposobem osłabiania założenia o niezależności zmiennych losowych jest założenie zależności typu „mieszania". To pojęcie opisuje asymptotyczną niezależność zmiennych losowych, gdy różnica ich indeksów dąży do nieskończoności. Rozważane są wtedy ciągi zmiennych losowych „oddzielonych jeden od drugiego" i „prawie niezależnych". Rodzaj tej „prawie niezależności" opisują odpowiednio zdefiniowane współczynniki, których wartość zależy od σ-ciał generowanych przez te zmienne losowe. W pracach rozważane są różne typy ciągów mieszających np. ciągi ϕ-mieszające, ρ-mieszające oraz ρ^*-mieszające.

Definicja 2.6. *Ciąg zmiennych losowych $\{X_n, n \geqslant 1\}$ nazywamy ϕ-mieszającym (lub jednostajnie silnie mieszającym), jeśli*

$$\phi(n) = \sup_{k\geqslant 1, A\in\mathcal{F}_1^k, P(A)>0, B\in\mathcal{F}_{k+n}^\infty} \big|P(B|A) - P(B)\big| \longrightarrow 0, \quad gdy \quad n \longrightarrow \infty,$$

gdzie $\mathcal{F}_m^n$ jest σ-ciałem generowanym przez zmienne losowe $X_m, X_{m+1}, \ldots, X_n$.

Definicja 2.7. *Ciąg zmiennych losowych $\{X_n, n \geqslant 1\}$ nazywamy ciągiem ρ-mieszającym, jeśli*

$$\rho(n) = \sup_{k\in N} \rho\big(\mathcal{F}_1^k, \mathcal{F}_{k+n}^\infty\big) \longrightarrow 0, \quad gdy \quad n \longrightarrow \infty.$$

Dla zdefiniowania kolejnej zależności ρ^*-mieszania, nazywanej też zależnością $\tilde{\rho}$-mieszania, wykorzystujemy ciąg współczynników $\rho^*(n)$, $n \geqslant 0$ (oznaczane też symbolem $\tilde{\rho}(n)$):

$$\rho^*(n) = \sup_{S,T}\Big(\sup_{X\in L^2(\mathcal{F}_S), Y\in L^2(\mathcal{F}_T)} \frac{cov(X,Y)}{\sqrt{\mathbf{Var}X \cdot \mathbf{Var}Y}}\Big), \tag{2.6}$$

gdzie S, T są skończonymi podzbiorami zbioru liczb naturalnych takimi, że $dist(S,T) = \inf_{x\in S, y\in T} |x-y| \geqslant n$.

Różne graniczne własności tego współczynnika były badane przez Bradleya [3, 4] i Millera [32].

Definicja 2.8. *Ciąg* $\{X_n, n \geqslant 1\}$ *jest ciągiem* ρ^**-mieszającym (*$\tilde{\rho}$*-mieszającym), jeśli*

$$\lim_{n\to\infty} \rho^*(n) < 1. \tag{2.7}$$

Bezpośrednio z wzoru (2.6) wynika następująca własność współczynników $\rho^*(n)$:

$$0 \leqslant \rho^*(n+1) \leqslant \rho^*(n) \leqslant \rho^*(0) = 1.$$

Oznacza to, że warunek (2.7) jest równoważny istnieniu stałej $k \in \mathbb{N}$ takiej, że

$$\rho^*(k) < 1. \tag{2.8}$$

W pracach dotyczących zależności ρ^*-mieszania występuje na ogół założenie, że $\rho^*(n) < 1$ dla n większych od pewnego n_0.

Zależność ρ^*-mieszania wydaje się być podobna do zależności ρ-mieszania, jednak są to dwa zupełnie różne typy zależności.

Rozważania dotyczące twierdzenia typu Bauma-Katza dla ciągów ρ^*-mieszających rozpoczęli w roku 1999 Peligrad i Gut [34]. Przedstawili oni dowód mocnego prawa wielkich liczb typu Bauma-Katza dla ciągów ρ^*-mieszających o jednakowym rozkładzie

Twierdzenie 2.3. *[34] Jeżeli* $\{X, X_n, n \geqslant 1\}$ *jest ciągiem* ρ^**-mieszającym o jednakowym rozkładzie,* $\alpha p > 1$*,* $\alpha > \frac{1}{2}$ *i* $EX = 0$ *dla* $p > 1$*, to warunek istnienia skończonego momentu rzędu* p*,* $E|X|^p < \infty$ *jest równoważny warunkowi*

$$\sum_{n=1}^{\infty} n^{\alpha p-2} P\big[\max_{1\leqslant j\leqslant n} |S_j| > \varepsilon n^{\alpha}\big] < \infty.$$

Cai [5] uzupełnił ten wynik o przypadek $\alpha p = 1$.

W pracy z 2004 roku Shixin [42] pokazał, dla ciągów ρ^*-mieszających o jednakowym rozkładzie, że założenia $EX = 0$ i $E|X|^p < \infty$, $1 < p \leqslant 2$ są wystarczające na to, by dla dowolnego $\varepsilon > 0$, dowolnego $\delta > 0$ i pewnego $\alpha \geqslant \max\{(1+\delta)/p, 1\}$ spełniony był warunek

$$\sum_{n=1}^{\infty} n^{\alpha p-2-\delta} P\Big[|S_n| > \varepsilon n^{\alpha}\Big] < \infty.$$

W szczególnym przypadku, gdy $\delta = 1$, $p = 2$ i $\alpha = 1$, Shixin [42] otrzymał następujący rezultat: *jeżeli ciąg* $\{X, X_n, n \geqslant 1\}$ *jest ciągiem* ρ^**-mieszającym o jednakowym rozkładzie z zerową wartością oczekiwaną i skończonym drugim momentem, to*

$$\sum_{n=1}^{\infty} n^{-1} P\Big[|S_n| > \varepsilon n\Big] < \infty.$$

3 Warunek konieczny i dostateczny kompletnej zbieżności w prawie wielkich liczb

W badaniach dowolnej populacji ze względu na określoną cechę X, najbardziej pożądanym jest taki wynik, który podaje warunek konieczny i dostateczny na to, by elementy tej populacji posiadały cechę X. W przypadku praw wielkich liczb, chcielibyśmy znaleźć warunek konieczny i dostateczny na to, by ciąg zmiennych losowych, spełniający być może jakieś dodatkowe warunki, spełniał prawo wielkich liczb. Przykładem takiego rezultatu jest twierdzenie Hsu-Robbinsa-Erdősa; *warunkiem koniecznym i dostatecznym na to, by ciąg niezależnych zmiennych losowych* $\{X_n, n \geqslant 1\}$ *o jednakowym rozkładzie spełniał warunek*

$$\forall \varepsilon > 0 \quad \sum_{n=1}^{\infty} P\Big[|S_n| > \varepsilon n\Big] < \infty \tag{3.1}$$

jest zerowa wartość oczekiwana i skończony drugi moment elementów tego ciągu.

Warunek konieczny i dostateczny (WKD) został sformułowany dla kompletnej zbieżności w prawie wielkich liczb dla ciągu niezależnych zmiennych losowych o jednakowym rozkładzie. Dla ciągów o różnych rozkładach w przypadku ogólnym, tzn. bez dodatkowych założeń, takiego warunku jak dotąd, nie udało się otrzymać. W różnych pracach podejmowane były próby podania warunku WKD;

wszystkie jednak zakładały dodatkowe ograniczenia dotyczące rozkładów zmiennych $X_1, X_2, \ldots$.

Spătaru [43] założył, że problem warunku koniecznego i dostatecznego dla kompletnej zbieżności w prawie wielkich liczb dla ciągu niezależnych zmiennych losowych można sprowadzić do poszukiwania związków pomiędzy warunkiem (3.1) a warunkami

$$\sum_{n=1}^{\infty}\sum_{k=1}^{n} P\Big[|X_k| \geqslant n\Big] < \infty \tag{3.2}$$

i

$$\lim_{n\to\infty} \frac{1}{n}\sum_{k=1}^{n} E\Big(X_k I[|X_k| < n]\Big) = 0. \tag{3.3}$$

Udowodnił on, że (3.1) w ogólnym przypadku implikuje (3.2) i (3.3). Dla implikacji w drugą stronę wprowadził pewne warunki pomocnicze. Rozważał taki ciąg niezależnych zmiennych losowych $\{X_n, n \geqslant 1\}$, że zbiór różnych rozkładów zmiennych $X_1, X_2, \ldots$ jest skończony, tzn., że istnieje skończony zbiór zmiennych losowych $Y_1, Y_2, \ldots Y_k$ o tej własności, że dla każdego $n \in N$ istnieje takie $i = i(n)$, że zmienne X_n i $Y_{i(n)}$ mają takie same rozkłady. Dla takiego ciągu wprowadził podział zbioru N na rozłączne podzbiory $N_1, N_2, \ldots, N_k$ o tej własności, że jeśli $n \in N_i$, to zmienna X_n ma taki sam rozkład jak zmienna Y_i. Dodatkowe warunki, przy których udowodnił, że (3.2) i (3.3) są wystarczające dla (3.1), mają postać warunków opisujących graniczne zachowanie się ciągów $\alpha_i(n)$, $i = 1, \ldots, k$, gdzie $\alpha_i(n) = Card\{j \in N_i : j \leqslant n\}$.

Problem warunku WKD podejmował też w swoich pracach Pruss [35,36]. Rozważał on ciągi niezależnych zmiennych losowych oraz tablice niezależnych w wierszu zmiennych losowych. W pracy [36] podał cztery rodzaje słabszych warunków dla ciągów $\alpha_i(n)$, przy których warunki (3.2) i (3.3) implikują (3.1), a w pracy [35], dla tablic $\{X_{nk}, 1 \leqslant k \leqslant n, n \geqslant 1\}$, założył, że zmienne w każdym wierszu tworzą regularne niezależne pokrycie pewnej innej zmiennej losowej ξ.

Definicja 3.1 ([35]). *Niech $\xi_1, \xi_2, \ldots, \xi_n$ będzie ciągiem zmiennych losowych i niech ξ będzie zmienną losową, która może być zdefiniowana nawet na innej przestrzeni*

probabilistycznej. Mówimy, że $\xi_1, \xi_2, \ldots, \xi_n$ tworzą regularne pokrycie zmiennej ξ, gdy

$$E\Big[G(\xi)\Big] = \frac{1}{n}\sum_{k=1}^{n} E\Big[G(\xi_k)\Big]$$

dla dowolnej mierzalnej funkcji G, dla której obie strony równości są dobrze określone.

Udowodnił, że w takim przypadku warunkiem koniecznym i wystarczającym kompletnej zbieżności ciągu $\frac{1}{n}\sum_{i=1}^{n} X_{ni}$ do pewnej stałej rzeczywistej L jest istnienie drugiego momentu zmiennej ξ ($E\xi^2 < \infty$) oraz równość $E\xi = L$. Tym samym podał uogólnienie wyniku Hsu-Robbinsa-Erdősa na przypadek tablic zmiennych losowych tworzących w każdym wierszu regularne niezależne pokrycie zmiennej ξ.

Problem warunku koniecznego i dostatecznego kompletnej zbieżności w prawie wielkich liczb dla ciągów zmiennych losowych ujemnie stowarzyszonych był rozważany również w pracach [25] i [26]. Podstawowym celem przyjętym w tych pracach była rezygnacja z założenia o jednakowym rozkładzie i zastąpienie go nowym, słabszym warunkiem tzw. słabego ograniczenia.

Definicja 3.2 ([25]). *Zmienne losowe $\{X_n, n \geqslant 1\}$ są słabo ograniczone przez zmienną losową ξ (która może być zdefiniowana na innej przestrzeni probabilistycznej), jeśli istnieją takie stałe $c_1, c_2 > 0$, że dla dowolnego $n \in \mathbb{N}$ i $x \in \mathbb{R}$*

$$c_1 \cdot P\big[|\xi| > x\big] \leqslant \frac{1}{n}\sum_{k=1}^{n} P\big[|X_k| > x\big] \leqslant c_2 \cdot P\big[|\xi| > x\big]. \tag{3.4}$$

Warunek ten pozwala, np. zamiast zmiennych losowych o jednakowym rozkładzie, rozważać ciągi zmiennych losowych o różnych rozkładach, zbieżnych według rozkładu do pewnej zmiennej X.

Twierdzenie 3.1 ([25]). *Niech $\alpha > \frac{1}{2}$ i $\alpha p > 1$. Niech $\{X_n, n \geqslant 1\}$ będzie ciągiem ujemnie stowarzyszonych zmiennych losowych słabo ograniczonych przez zmienną losową X. Ponadto zakładamy, że dla $p \geqslant 1$ $EX_n = 0$, dla wszystkich $n \geqslant 1$.*

Wtedy następujące warunki są równoważne:
(i) $E|X|^p < \infty$*;*
(ii) $\sum_{n=1}^{\infty} n^{\alpha p-2} P\left(\max_{1\leqslant i\leqslant n} |S_i| > \varepsilon n^{\alpha}\right) < \infty$ *dla dowolnego* $\varepsilon > 0$.

Wnioski otrzymane z tego twierdzenia są uogólnionymi wersjami znanych wyników dla niezależnych zmiennych losowych o jednakowych rozkładach. Np. kładąc $\alpha = \frac{1}{t}$ i $p = 2t$ dla $0 < t < 2$ dostajemy uogólnienie rezultatu Hu, Móricza i Taylora [22] na przypadek ciągu ujemnie stowarzyszonych zmiennych losowych, bez założenia o jednakowym rozkładzie.

Natomiast, gdy rozważamy ciągi o jednakowych rozkładach otrzymujemy wersję twiedzenia typu Bauma-Katza dla zmiennych losowych ujemnie stowarzyszonych (Wniosek 2.2 [25]). Ponadto, w przypadku, gdy $\{X_n, n \geqslant 1\}$ jest ciągiem słabo zdominowanym przez pewną zmienną losową X o skończonym p-tym momencie, to z twierdzenia 3.1 dostajemy główny wynik Hu, Móricza i Taylora [22] w wersji dla zmiennych losowych ujemnie stowarzyszonych.

4 Zbieżność kompletna momentowa

Pierwsze twierdzenia typu prawa wielkich liczb to tzw. *słabe prawa wielkich liczb*, formułowane dla zbieżności według prawdopodobieństwa. Kolejnym krokiem w rozwoju tych twierdzeń są rozważania dla zbieżności z prawdopodobieństwem jeden, zbieżności mocniejszej, dające w efekcie twierdzenia typu *mocnego prawa wielkich liczb*. Zdefiniowanie w 1947 roku nowego rodzaju zbieżności, zbieżności kompletnej, zapoczątkowało nowy typ mocnego prawa wielkich liczb określany mianem *prawa wielkich liczb typu Hsu-Robbinsa*, *kompletną zbieżnością w prawie wielkich liczb* lub *szybkością zbieżności w prawie wielkich liczb*. W roku 1988 Chow [8] zdefiniował jeszcze mocniejszy rodzaj zbieżności i określił go mianem kompletnej momentowej zbieżności (ang. *complete moment convergence*).

Definicja 4.1 ([8]). *Niech* $\{Z_n, n \geqslant 1\}$ *będzie ciągiem zmiennych losowych,* $q > 0$*,* $a_n > 0$ *i* $b_n > 0$ *dla* $n \geqslant 1$*. Jeśli*

$$\sum_{n=1}^{\infty} a_n E\{b_N^{-1}|Z_n| - \varepsilon\}_+^q < \infty, \tag{4.1}$$

to mówimy, że ciąg $\{Z_n, n \geqslant 1\}$ *spełnia warunek kompletnej momentowej zbieżności dla ciągów* $\{a_n, n \geqslant 1\}$ *i* $\{b_n, n \geqslant 1\}$*.*

Dla niezależnych zmiennych losowych Chow [8] dowiódł następującego twierdzenia.

Twierdzenie 4.1 ([8]). *Niech* $\{X, X_n, n \geqslant 1\}$ *będzie ciągiem niezależnych zmiennych losowych o jednakowym rozkładzie z wartością oczekiwaną* $EX = 0$ *i niech* $1 \leqslant p < 2$ *oraz* $\gamma > p$*. Jeśli*

$$E[|X|^\gamma + |X|\log(1 + |X|)] < \infty,$$

to

$$\sum_{n=1}^{\infty} n^{\frac{\gamma}{p}-2-\frac{1}{p}} E\big(|S_n| - \varepsilon n^{\frac{1}{p}}\big)_+ < \infty, \quad \textit{dla dowolnego} \quad \varepsilon > 0.$$

Aktualnie, problematyka zbieżności kompletnej i zbieżności kompletnej momentowej należą do najważniejszych w teorii pawdopodobieństwa. Ostatnie wyniki dotyczące kompletnej momentowej zbieżności można znaleźć między innymi w pracach Qiu i Chena [38], Wu i Jianga [49], Dinga i inn. [11] oraz Liu i inn. [30].

Qui i Chen [38], zainspirowani wynikiem Guta i Stadtmüllera [19], dowiedli twierdzenia dotyczącego zbieżności kompletnej momentowej dla ciągu niezależnych zmiennych losowych $\{X, X_n, n \geqslant 1\}$ o jednakowym rozkładzie z wartością oczekiwaną $EX = 0$ spełniającego warunek $E\exp(\ln^\alpha |X|) < \infty$ dla $\alpha > 0$. Wu i Jiang [49] rozszerzyli ten wynik na zmienne losowe ujemnie stowarzyszone bez konieczności nakładania jakichkolwiek dodatkowych warunków.

Twierdzenie 4.2 ([49]). *Niech* $\{X, X_n, n \geqslant 1\}$ *będzie ciagiem ujemnie stowarzyszonych zmiennych losowych o jednakowym rozkładzie spełniających warunki*

$EX = 0$ *i* $E\exp(\ln^{\alpha}|X|) < \infty$ *dla* $\alpha > 0$. *Wtedy*

$$\sum_{n=1}^{\infty} \exp\left(\ln^{\alpha} n\right) \frac{\ln^{\alpha-1} n}{n^{2+q}} E\{\max_{1\leqslant k\leqslant n} |S_k| - n\beta\}_+^q < \infty \quad \textit{dla} \quad \beta > 1 \quad \textit{i} \quad q > 0. \tag{4.2}$$

Do interesujących wyników zaliczyć należy również twierdzenie dotyczące kompletnej momentowej zbieżności dla procesu średniej ruchomej generowanej przez klasę ciągów spełniajacych nierówność typu Rosenthala i spełniających warunek słabego zdominowania (Ding i inn. [11]). Wynik ten zaliczyć możemy do grupy twierdzeń rozważających ciągi o różnych rozkładach.

W pracy Liu i inn. [30] rozważana jest kompletna i kompletna momentowa zbieżność dla najogólniejszej z omawianych tu struktur zależności, dla zależności typu WOD (widely orthant dependent). Struktura ta zawiera w sobie takie struktury jak ujemnie stowarzyszone (NA) zmienne losowe, ujemnie superaddytywnie zależne (NSD) zmienne losowe, ujemnie zależne typu orthant (NOD) zmienne losowe oraz szeroko zależne typu orthant (WOD) zmienne losowe. Wynik przedstawiony w pracy [30] jest uogólnieniem klasycznego wyniku Chowa [8] na przypadek zmiennych losowych o strukturze zależności typu WOD przy dokładnie takich samych założeniach momentowych jak w twierdzeniu 4.1 dla niezależnych zmiennych losowych.

5 Zastosowania

Do klasycznych zastosowań praw wielkich liczb należą zastosowania w statystyce np. do badania zgodności estymatorów parametrów regresji liniowej dla modelu z błędami losowymi. Klasyczna definicja estymatora zgodnego mówi, że *estymator* $\hat{\theta} = \hat{\theta}_n$ *jest zgodny, jeśli jest stochastycznie zbieżny do szacowanego parametru* θ*:*

$$\forall \varepsilon > 0 \lim_{n\to\infty} P\left[|\hat{\theta}_n - \theta| < \varepsilon\right] = 1. \tag{5.1}$$

Oznacza to, że wraz ze wzrostem liczebności próby, prawdopodobieństwo, że oszacowanie parametru θ przy pomocy estymatora $\hat{\theta}$ będzie różnić się dowolnie mało od samego parametru dąży do 1. Zbieżność we wzorze (5.1) jest zbieżnością według prawdopodobieństwa, a więc rozważając mocniejszy rodzaj zbieżności estymatora $\hat{\theta}_n$ do parametru θ , np. zbieżność kompletną lub zbieżność kompletną momentową, możemy dostać mocniejszy rodzaj zgodności tego estymatora.

W pracach dotyczących zbieżności kompletnej lub kompletnej momentowej, bardzo często, prezentowane są zastosowania otrzymanych wyników do badania zgodności (kompletnej lub kompletnej momentowej) MNK-estymatorów (estymatorów otrzymanych metodą najmniejszych kwadratów) parametrów regresji liniowej w EV-modelu (modelu z błędami losowymi). Na ogół, gdy w modelach najprostszych, błędy losowe są pomijane, MNK-estymatory nieznanych parametrów są estymatorami obciążonymi, dlatego model uwzględniający błędy jest modelem dokładniejszym.

Model rozważany w pracy Wanga i inn. [48] jest modelem zaproponowanym przez Deatona [10] w celu skorygowania błędów próbkowania. W modelu tym obserwujwmy dwie cechy w konkretnie ustalonych punktach, a obserwacje obarczone są błędami losowymi.

Przyjmujemy następującą postać modelu

$$\begin{cases} \eta_i = \theta + \beta x_i + \varepsilon_i, \quad \xi_i = x_i + \delta_i, \quad 1 \leqslant i \leqslant n; \\ E\varepsilon_i = E\delta_i, \quad 1 \leqslant i \leqslant n \end{cases}, \tag{5.2}$$

gdzie θ i β są nieznanymi parametrami, x_i, $1 \leqslant i \leqslant n$ są nielosowymi ustalonymi punktami, ε_i, δ_i, $i = 1, 2, ...$ błędami losowymi a ξ_i, η_i, $i = 1, 2, ...$ zmiennymi losowymi opisującymi obserwacje. Z warunku (5.2) mamy

$$\eta_i = \theta + \beta\xi_i + \nu_i, \quad \nu_i = \varepsilon_i - \beta\delta_i, \quad 1 \leqslant i \leqslant n.$$

Dla tak zbudowanego modelu estymator $\hat{\beta}_n$ parametru β i $\hat{\theta}_n$ parametru θ, otrzymane metodą najmniejszych kwadratów, dane są wzorami:

$$\hat{\beta}_n = \frac{\sum_{i=1}^{n}(\xi_i - \bar{\xi}_n)(\eta_i - \bar{\eta}_n)}{\sum_{i=1}^{n}(\xi_i - \bar{\xi}_n)^2}, \tag{5.3}$$

$$\hat{\theta}_n = \bar{\eta}_n - \hat{\beta}_n \bar{\xi}_n, \tag{5.4}$$

gdzie: $\bar{\xi}_n = n^{-1}\sum_{i=1}^n \xi_i$ oraz $\bar{\eta}_n = n^{-1}\sum_{i=1}^n \eta_i$.

Dla powyższego modelu otrzymujemy następujące oszacowania

$$\hat{\beta}_n - \beta = \frac{\sum_{i=1}^n (\delta_i - \bar{\delta}_n)\varepsilon_i - \sum_{i=1}^n (x_i - \bar{x}_n)(\varepsilon_i - \beta\delta_i) - \beta\sum_{i=1}^n (\delta_i - \bar{\delta}_n)^2}{\sum_{i=1}^n (\xi_i - \bar{\xi}_n)^2} \tag{5.5}$$

i

$$\hat{\theta}_n - \theta = (\beta - \hat{\beta}_n)\bar{x}_n + (\beta - \hat{\beta}_n)\bar{\delta}_n + \bar{\varepsilon}_n - \beta\bar{\delta}_n, \tag{5.6}$$

$\bar{x}_n = n^{-1}\sum_{i=1}^{\infty} x_i$.

W pracy [48] rozważany jest model (5.2) z błędami losowymi $\{\varepsilon_i, i \geqslant 1\}$ oraz $\{\delta_i, i \geqslant 1\}$, które tworzą ściśle stacjonarne ciągi zmiennych losowych z wewnętrzną strukturą zależności typu NSD (ujemnie superaddytywnie zależne) wzajemnie niezależne jeden od drugiego. Twierdzenia dotyczące kompletnej zbieżności dla ważonych sum zmiennych losowych ujemnie superaddytywnie zależnych przedstawione w pracy [48] zostały wykorzystane do badania kompletnej zgodności estymatorów (5.3) i (5.4). Dla pierwszego z tych estymatorów sformułowane zostało następujące twierdzenie.

Twierdzenie 5.1 ([48]). *Niech $\{\varepsilon_i, i \geqslant 1\}$ i $\{\delta_i, i \geqslant 1\}$ będą ciągami ściśle stacjonarnych zmiennych losowych z wewnętrzną strukturą zależności typu NSD, wzajemnie niezależne jeden od drugiego i opisujące błędy losowe w modelu (5.2). Załóżmy, że $E|\varepsilon_1|^{4p} < \infty$ i $E|\delta_1|^{4p} < \infty$ dla pewnego $p > 0$. Jeśli istnieje taka stała $\tau > 0$, że*

$$\max_{1\leqslant i\leqslant n} \frac{|x_i - \bar{x}_n|}{\sqrt{S_n} n^{\tau - 1/p}} = O(1),$$

$$\frac{n^r}{\sqrt{S_n}} = O(1)$$

i

$$\frac{n^{1-r}}{\sqrt{S_n}} \longrightarrow 0, \quad \textit{gdy} \quad n \longrightarrow \infty,$$

gdzie $S_n = \sum_{i=1}^{n}(x_i - \bar{x}_n)^2$ *i* $\bar{x}_n = n^{-1}\sum_{i=1}^{\infty} x_i$ *dla dowolnego* $n \geqslant 1$, *to*

$$\frac{\sqrt{S_n}}{n^{\tau}}(\hat{\beta}_n - \beta) \longrightarrow 0 \quad \textit{kompletnie,} \quad \textit{gdy} \quad n \longrightarrow \infty \tag{5.7}$$

Kompletną zgodność estymatora $\hat{\theta}_n$ zapewnia twierdzenie.

Twierdzenie 5.2 ([48]). *Jeśli, przy założeniach twierdzenia 5.1, istnieje stała* ν *spełniająca warunki:*

$$0 < \nu < \min(1 - 1/p, 1/2) \quad \textit{oraz} \quad \frac{n^{\tau+\nu}}{\sqrt{S_n}}|\bar{x}_n| = O(1),$$

to

$$n^{\nu}(\hat{\theta}_n - \theta) \longrightarrow 0 \quad \textit{kompletnie,} \quad \textit{gdy} \quad n \longrightarrow \infty. \tag{5.8}$$

Twierdzenie 5.1 i odpowiednio twierdzenie 5.2 podają warunki dostateczne dla zgodości kompletnej estymatora $\hat{\beta}$ i estymatora $\hat{\theta}$ w modelu (5.2). Kluczowym punktem rozważań dowodowych są zależności (5.5) i (5.6). Różnica $\hat{\beta}_n - \beta$, we wzorze (5.4) jest rozłożona na cztery części $\hat{\beta}_n(\delta_i - \bar{\delta}_n)\varepsilon_i$, $\sum_{i=1}^{n}(x_i - \bar{x}_n)(\varepsilon_i - \beta\delta_i)$, $\beta\sum_{i=1}^{n}(\delta_i - \bar{\delta}_n)^2$ i $\sum_{i=1}^{n}(\xi_i - \bar{\xi}_n)^2$. Dowód tezy (5.7) polega na badaniu zbieżności poszczególnych wyrażeń, z odpowiednim ciągiem normalizującym, i wykorzystaniu odpowiednich twierdzeń ([48]) dla kompletnej zbieżności sum ważonych zmiennych losowych o strukturze ujemnej superaddytywnej zależności. Analogicznie przebiega dowód tezy (5.8) w twierdzeniu 5.2.

Twierdzenia typu mocnego prawa wielkich liczb formułowane dla zmiennych losowych z różnymi coraz bardziej ogólnymi strukturami zależności i coraz mocniejszymi rodzajami zbieżności pozwalają określać warunki dostateczne dla mocniejszej zgodności estymatorów dla prób losowych z różnymi strukturami zależności. Dlatego modele otrzymywane z wykorzystaniem tych wyników są bliższe rzeczywistości.

Literatura

[1] K. Alam, K.M.L. Saxena, Positive dependence in multivariate distrbutions, *Comm. Statist. A* 10, 1981, 1183–1196.

[2] I.E. Baum, M. Katz, Convergence rates in the law of large numbers, *Trans. Amer. Math. Soc.*, 120, 1965, 108–123.

[3] R.C. Bradley, On the spectral density and asymptotic normality of weakly dependent random fields, *J. Theor. Probab.*, 5, 1992, 355–373.

[4] R.C. Bradley, Equivalent mixing condition for random fields, *Ann. Probab.*, 21, 1993, 1921–1926.

[5] G. H. Cai, Strong law of large numbers for ρ^*-mixing sequences with different distributions, *Discrete Dynamics in Nature and Society*, 2007, Acticle ID 27648, 1–7.

[6] Y. Chen, A. Chen, K. Ng , The strong law of large numbers for extended negatively dependent random variables, *J. Appl. Probab.* 47, 2010, 908–922.

[7] Y.S. Chow, Delayed sums and Borel summability of independent, identically distributed random variables, *Bull. Inst. Math. Acad. Sinica*, 1973, 207–2020.

[8] Y. S. Chow, On the rate of moment convergence of sample sums and extremes, *Bull. Inst. Math. Acad.Sin.*, 16, 1988, 177–201.

[9] K.L. Chung, Note on some strong laws of large numbers, *Amer. J. Math.*, 69, 1947, 189–192.

[10] A. Deaton, Panel data from a time series of cross-section, *Journal of Econometrics*, 30, 1985, 109–126.

[11] Y.Ding, Y. Wu, S. Ma, X. Tao, X. Wang, Complete convergence and complete moment convergence for widely orthant-dependent random variables, *Commun. Stat. Theory Methods*, 46, 2017, 10903–10913.

[12] N. Ebrahimi, M. Ghosh, Multivariate negative dependence, *Commun. Stat. Theory Methods*, 10, 1981, 307–337.

[13] P. Erdős, On a theorem of Hsu and Robbins, *Ann. Math. Statistics*, 20, 1949, 286–291.

[14] A. Gut, Marcinkiewicz laws and complete convergence rates in the law of large numbers for random variables with multidimentional indices, *Ann. Probab.*, 6, 1978, 469–482.

[15] A. Gut, Convergence rates for probabilities of moderate deviations for sums of random variables with multidimentional indices, *Ann. Probab.*, 8, 1978, 298–313.

[16] A. Gut, Complete convergence rates for randomly indexed partial sums with an application to some first passage times, *Acta Math. Hung.*, 42, 1983, 225–232.

[17] A. Gut, Correction to "Complete convergence rates for randomly indexed partial sums with an application to some first passage times", *Acta Math. Hung.*, 45, 1985, 235–236.

[18] A. Gut, Complete convergence for arrays, *Periodica Math. Hung.*, 25, 1992, 51–75.

[19] A. Gut,U. Stadtmüller, An intermediate Baum-Katz theorem, *Stat. Probab. Lett.*, 81, 2011, 1486–1492.

[20] T.Z. Hu, Negatively superadditive dependence of random variablrs with applications, *Chinese Journal of Applied Probability and Statistics*, 16, 2000, 133–1444.

[21] P.L. Hsu, H. Robbins, Complete convergence and the law of large numbers, *Proc. Nat. Acad. Sci. USA*, 33, 1947, 25–31.

[22] T.C. Hu, F. Móricz, R.L. Taylor, Strong laws of large numbers for arrays of rowwise independent random variables, *Acta Math. Hung.*, 54, 1989, 153–162.

[23] K. Joag-Dev, F. Proschan, Negative association of random variables with applications, *Ann. Statist.*, 17, 1999, 963–991.

[24] M.L. Katz, The probability in the tail of a distribution, *Ann. Math. Statist.*, 34, 1963, 312–318.

[25] A. Kuczmaszewska, On complete convergence in Marcinkiewicz-Zygmund type SLLN for negatively associated random variables, *Acta Math. Hungar.*, 2010, 116–130.

[26] A.Kuczmaszewska, Z.A. Lagodowski, Convergence rates in the SLLN for some classes of dependent fields, *J. Math. Anal. Appl.*, 380, 2011, 571–584.

[27] E.L. Lehmann, Some concepts of dependence, *Ann. Math. Statist.*, 37, 1966, 1137–1153.

[28] H. Y. Liang, C. Su, Complete convergence for weighted sums of NA sequences, *Statist. Probab. Lett.*, 45, 1999, 85–95.

[29] L. Liu, Precise large deviations for dependent random variables with heavy tails. *Stat. Probab. Lett.*, 79, 2009, 1290–1298.

[30] X. Liu, Y. Shen, J. Yang, Y. Lu, Complete moment convergence of widely orthant dependent random variables, *Commun. Stat. Theory Methods*, 46, 2017, 7256–7265.

[31] P. Matula, A note on the almost sure convergence of sums of negatively dependent random variables, *Statist. Probab. Lett.*, 15 1992, 201–213.

[32] C. Miller, Three theorems on ρ^*-mixing random fields, *J. Theor. Probab.*, 7, 1994, 867–882.

[33] C. M. Newman, Asymptotic independence and limit theorems for positively and negatively dependent random variables, Inequalities in Statistics and Probability (Hayward 1984);. Tong, YL (ed), Institute of Mathematical Statistics, 127–140.

[34] M. Peligrad, A. Gut, Almost-sure results for class of dependent random variables, *J. Theor, Probab.*, 12, 1999, 87–104.

[35] A.R. Pruss, Randomly sampled Riemann sums and complete convergence in the law of large numbers for the case without identical distributions, *Proc. Amer. Math. Soc.*, 124, 1996, 919–929.

[36] A.R. Pruss, On Spãtaru's extension of the Hsu-Robbins-Erdős law of large numbers, *J. Math. Anal. Appl.*, 199, 1996, 558–576.

[37] D.H. Qiu, P.Y. Chen, R.G. Antonini, A. Volodin, On the complete convergence for arrays of rowwise extended negatively dependent random variables. *J. Korean Math. Soc.*, 50, 2013, 379–392.

[38] D.H. Qiu, P.Y. Chen, Complete moment convergence for i.i.d. random variables, *Stat. Probab. Lett.*, 91, 2014, 76–82.

[39] Q.M. Shao, A comparison theorem on moment inequalities between negatively associated and independent random variables, *J. Theor. Probab.*, 13, 2000, 343–356.

[40] A. Shen, Probability inequalities for END sequence and their applications, *J. Inequal. Appl.*, 98, 2011, 1–12.

[41] A. Shen, M. Xue, W. Wang, Complete convergence for weighted sums of extended negatively dependent random variables. *Commun. Stat. Theory Methods*, 46, 2017, 1433–1444.

[42] G. Shixin, Almost sure convergence for $\tilde{\rho}$-mixing random variable sequences, *Statist. Probab. Lett.*, 67, 2004, 289–298.

[43] A. Spătaru, A generalization of the Hsu-Robbins-Erdős law of large numbers, *J. Math. Anal. Appl.*, 187, 1994, 371–383.

[44] F.L. Spitzer, A combinatorial lemma and its application to probability theory, *Trans. Amer. Math. Soc.*, 82, 1956, 323–339.

[45] H. Teicher, Some new conditions for the strong law, *Proc. Nat. Acad. Sci. U.S.A.*, 59, 1968,705–707.

[46] K. Wang, Y. Wang, Q. Gao, Uniform asymptotics for finite-time ruin probability of dependent risk model with constant interest rate, *Methodol. Comput. Appl.Probab.*, 15, 2013, 111–124.

[47] X.J.Wang, X.Q. Li, S.H. Hu, X.H. Wang, On complete convergence for an extended negatively dependent sequence, *Commun. Stat. Theory Methods*, 43, 2014, 2923–2937.

[48] X. Wang, A. Shen, Z. Chen, S. Hu, Complete convergence for weighted sums of NSD random variables and its application in the EV regression model, *Test*. 24, 2015, 166–184.

[49] Q. Wu, Y. Jiang, Complete convergence and complete moment convergence for negatively associated sequences of random variables, *Journal of Inequality and Applications*, 2016, DOI10.1186/s13660-016-1107-z92016)

[50] J.G. Yan, Strong Stability of a type of Jamison weighted sums for END random variables, *J. Korean Math. Soc.*, 54, 2017, 897–907.

[51] J.G. Yan, Almost sure convergence for weighted sums of WNOD random variables and its applications in nonparametric regression models, *Commun. Stat. Theory Methods*, 2017, doi.org/10.1080/03610926.2017.1364390

Biblioteka Cyfrowa Politechniki Lubelskiej

Działalność wydawnicza Politechniki Lubelskiej

65 lat POLITECHNIKI LUBELSKIEJ

POLITECHNIKA LUBELSKA
WYDZIAŁ PODSTAW TECHNIKI

ISBN: 978-83-7947-311-3

Ministerstwo Nauki i Szkolnictwa Wyższego

PATRONAT HONOROWY

PREZYDENT MIASTA LUBLIN
KRZYSZTOF ŻUK

Lubelska Wyżyna IT