

Uniwersytet im. Adama Mickiewicza w Poznaniu

Wydział Matematyki i Informatyki

**Wybrane wielowymiarowe
metody statystyczne dla
wielozmiennych danych funkcjonalnych**

Łukasz Waszak

Rozprawa doktorska w celu uzyskania stopnia
doktora nauk matematycznych w zakresie matematyki

promotor: prof. dr hab. Mirosław Krzyśko

promotor pomocniczy: dr Tomasz Górecki

Poznań 2015

Spis treści

Wstęp	2
1 Dane funkcjonalne	5
2 Składowe główne dla wielozmiennych danych funkcjonalnych	9
2.1 Klasyczne składowe główne	9
2.2 Składowe główne dla wielozmiennych danych funkcjonalnych	11
2.3 Symulacyjny wybór liczby elementów bazy dla wielozmiennych składowych głównych	18
3 Zmienne dyskryminacyjne dla wielozmiennych danych funkcjonalnych	25
3.1 Klasyczne zmienne dyskryminacyjne	25
3.2 Zmienne dyskryminacyjne dla wielozmiennych danych funkcjonalnych	27
4 Zmienne kanoniczne dla wielozmiennych danych funkcjonalnych	34
4.1 Klasyczne zmienne kanoniczne	34
4.2 Zmienne kanoniczne dla wielozmiennych danych funkcjonalnych	36
5 Zastosowanie wybranych metod statystycznych do wielozmiennych danych funkcjonalnych	45
5.1 Przykład 1.	45
5.1.1 Funkcjonalne składowe główne	47
5.1.2 Funkcjonalne zmienne dyskryminacyjne	53
5.1.3 Funkcjonalne zmienne kanoniczne	59
5.2 Przykład 2.	67
5.2.1 Funkcjonalne składowe główne	69
5.2.2 Funkcjonalne zmienne dyskryminacyjne	75
5.2.3 Funkcjonalne zmienne kanoniczne	81
Bibliografia	90

Wstęp

W klasycznych metodach statystycznych obiekty podlegające badaniu charakteryzowane są za pomocą cech obserwowanych w ustalonym momencie czasu. W prezentowanej rozprawie zakładamy, że obiekty charakteryzowane są za pomocą zmiennych funkcyjnych. Czym są zmienne funkcyjne? Zmienna funkcyjna X jest zmienną losową przyjmującą wartości w pewnej przestrzeni funkcyjnej E . Zbiór danych funkcyjnych jest próbą $\{X_1, \dots, X_n\}$ (oznaczaną również przez $\{X_1(t), \dots, X_n(t)\}$, jeśli jest to wygodne) pobraną z rozkładu zmiennej funkcyjnej X . Termin dane funkcyjne pojawił się po raz pierwszy w pracy Ramsaya i Dalzella (1991). W dalszym ciągu zakładamy, że E jest przestrzenią Hilberta wszystkich funkcji całkowalnych z kwadratem na pewnym przedziale $[a, b]$, tj. przestrzenią $L_2([a, b])$.

W tym przypadku dane funkcyjne mogą być reprezentowane w postaci

$$X(t) = \sum_{b=0}^{\infty} c_b \varphi_b(t),$$

gdzie $\varphi_b(t)$ są znanymi, ustalonymi funkcjami ortonormalnymi lub inaczej elementami ortonormalnej bazy $\{\varphi_0, \varphi_1, \dots\}$. Zauważmy, że reprezentacja funkcji za pomocą nieskończonego szeregu ortonormalnego wymaga znajomości nieskończonej liczby współczynników c_b . Niestety nikt z nas nie potrafi radzić sobie z nieskończoną liczbą współczynników. W związku z tym do aproksymacji funkcji $X(t)$ wykorzystuje się ucięty (skończony) szereg ortonormalny, zwany inaczej sumą częściową, postaci

$$X_B(t) := \sum_{b=0}^B c_b \varphi_b(t) = \mathbf{c}' \boldsymbol{\varphi}(t) = \boldsymbol{\varphi}'(t) \mathbf{c}.$$

Parametr B , będący liczbą naturalną, nazywa się punktem ucięcia. Zazwyczaj tylko niewielka liczba współczynników rozwinięcia jest istotna, a wszystkie pozostałe są mało znaczące. Prowadzi to do istotnej redukcji danych.

Z grubsza rzecz biorąc, główny problem statystyczny polega na optymalnym wyborze punktu ucięcia B oraz optymalnym oszacowaniu współczynników c_b . Problemom tym jest poświęcony Rozdział 1 tej rozprawy.

W tym miejscu można postawić naturalne pytanie: czy w rzeczywistości istnieją dane funkcyjne? Pytanie to ma istotne znaczenie, gdyż w praktyce wartości obserwowanego procesu

losowego $X(t)$ są zawsze rejestrowane w dyskretnych momentach czasu t_1, t_2, \dots, t_J , rzadziej lub gęściej rozmieszczonych w przedziale zmienności argumentu t . Tak więc ostatecznie mamy zawsze do czynienia z szeregiem czasowym $\{x(t_1), x(t_2), \dots, x(t_J)\}$ lub inaczej z wysokowymiarowym wektorem obserwacji. Istnieją jednakże liczne powody, by szeregi takie modelować jako elementy przestrzeni funkcjonalnej, ponieważ dane funkcjonalne mają wiele zalet w porównaniu z innymi sposobami reprezentowania szeregów czasowych.

Po pierwsze, łatwo radzą sobie z problemem brakujących obserwacji, nieuniknionym problemem w wielu dziedzinach badań. Niestety, większość metod analizy danych wymaga kompletnych szeregów czasowych. Jednym z rozwiązań jest po prostu usunięcie szeregu czasowego mającego brakujące wartości ze zbioru danych, ale działanie takie może prowadzić, i na ogół prowadzi, do utraty informacji. Inna możliwość, to posłużenie się jedną z wielu metod statystycznych predykcji brakujących danych, ale wówczas wyniki będą zależały od metody interpolacji. W przeciwieństwie do tego typu działań, w przypadku danych funkcjonalnych, problem brakujących obserwacji jest rozwiązany poprzez wyrażenie szeregów czasowych w postaci zbioru krzywych ciągłych.

Po drugie, dane funkcjonalne w sposób naturalny zachowują strukturę obserwacji, tj. zachowują zależność czasową obserwacji i biorą pod uwagę informację o każdym pomiarze.

Po trzecie, momenty obserwacji nie muszą być równomiernie rozmieszczone w poszczególnych szeregach czasowych.

Po czwarte, dane funkcjonalne unikają „przekleństwa” nadmiernej wymiarowości. Gdy całkowita liczba punktów czasowych, w których dokonuje się obserwacji przekracza liczbę rozpatrywanych szeregów czasowych, większość metod statystycznych nie daje zadowalających wyników ze względu na przeparametryzowanie. Aby uniknąć tego problemu, najczęściej stosuje się techniki redukcji wymiaru, takie jak analiza składowych głównych. Jednakże w tym przypadku pewne informacje o strukturze przestrzennej i czasowej danych mogą zostać utracone. W przypadku danych funkcjonalnych można uniknąć tego problemu, ponieważ szeregi czasowe zostają zastąpione zbiorem krzywych ciągłych niezależnych od całkowitej liczby punktów czasowych, w których dokonuje się obserwacji.

Chociaż prace dotyczące danych funkcjonalnych pojawiały się już wcześniej, to za symboliczny moment startu metod statystycznych dla danych funkcjonalnych należy przyjąć ukazanie się monografii Ramsaya i Silvermana (1997), monografii skierowanej do szerokiego kręgu odbiorców i z przewagą zagadnień praktycznych nad teorią. Uzupełnieniem tej pięknej monografii o dalsze aspekty praktyczne była książka Ramsaya i Silvermana (2002). W roku 2005 ukazało się wydanie drugie monografii z roku 1997. Wydanie to pociągnęło za sobą prawdziwy wysyp prac związanych z funkcjonalną analizą danych. Kolejnymi znaczącymi pozycjami w dziedzinie analizowania danych funkcjonalnych były książki: Clarkson i inni (2005), Ferraty i Vieu (2006) – książka o nastawieniu teoretycznym, Bosq i Blanke (2007), Dabo-Niang i Ferraty (2008),

Ramsay i inni (2009) oraz Ferraty i Romain (Eds) (2011). Najnowszą pozycją uzupełniającą literaturę dotyczącą danych funkcjonalnych jest książka Horvatha i Kokoszki (2012). Zawiera ona wyważoną mieszankę aspektów teoretycznych i praktycznych funkcjonalnej analizy danych.

Wśród prac przeglądowych dotyczących tej tematyki wymienić należy następujące prace: Rice (2004), Muller (2005), Gonzalez-Manteiga i Vieu (2011) – zawiera ona bogatą bibliografię, Delsol i inni (2011), Febrero-Bande i Oviedo de la Fuente (2012) oraz Cuevas (2014).

Wśród wielu metod statystycznych skonstruowanych dla danych funkcjonalnych poczesne miejsce zajmują trzy metody określane wspólnym mianem - metody redukcji wymiaru. Są to: analiza składowych głównych, analiza zmiennych dyskryminacyjnych oraz analiza korelacji i zmiennych kanonicznych. Klasyczne wersje tych metod zakładają, że rozpatrywane obiekty charakteryzowane są wieloma cechami. Tymczasem, w przypadku danych funkcjonalnych, dotychczas istniejące prace przyjmują, że obiekty charakteryzowane są za pomocą jednowymiarowych danych funkcjonalnych. Jest tu widoczna rozbieżność między założeniami w przypadku metod klasycznych i metod dla danych funkcjonalnych. W celu usunięcia tej rozbieżności, w prezentowanej rozprawie, skonstruowane zostały metody redukcji wymiaru dla wielowymiarowych danych funkcjonalnych. Metody te opisane są w Rozdziałach 2–4. Rozdział 5 zawiera konkretne przykłady stosowania tych metod.

Podziękowania

Składam serdeczne podziękowania Panu Profesorowi Mirosławowi Krzyśce oraz Panu Doktorowi Tomaszowi Góreckiemu za okazaną życzliwość, opiekę naukową oraz wskazówki udzielone mi podczas pisania pracy, które w istotny sposób przyczyniły się do ulepszenia niniejszej rozprawy.

Rozdział 1

Dane funkcjonalne

We wstępie powiedzieliśmy, że jednostki statystyczne podlegające badaniu będą charakteryzowane za pomocą wielu cech mierzonych w wielu momentach czasowych, tj. będą charakteryzowane ze pomocą szeregów czasowych. Powiedzieliśmy również, że z wielu względów dogodniej jest reprezentować szeregi czasowe za pomocą funkcji ciągłych określonych na pewnym przedziale. Pokażemy teraz sposób przejścia od szeregu czasowego do funkcji ciągłej.

Niech x_j oznacza zaobserwowaną wartość cechy X w j -tym momencie czasowym t_j , gdzie $j = 1, 2, \dots, J$. Wówczas dane te składają się z J par (t_j, x_j) . Takie dane dyskretne można wygładzić za pomocą pewnej funkcji ciągłej $x(t)$, gdzie $t \in I$ oraz I jest zbiorem spójnym takim, że $t_j \in I$, dla $j = 1, \dots, J$ (Ramsay i Silverman (2005)). Załóżmy, że funkcję $x(t)$ można przedstawić jako kombinację liniową skończonej liczby $B + 1$ ortonormalnych funkcji bazowych w następującej postaci:

$$x(t) = \sum_{b=0}^B c_b \varphi_b(t), \quad t \in I, \quad (1.1)$$

gdzie $\{\varphi_b\}$ jest układem ortonormalnych funkcji bazowych, a c_0, c_1, \dots, c_B są współczynnikami, które podlegają estymacji.

Przypomnijmy, że w przestrzeni Hilberta $L_2(I)$ funkcji całkownych z kwadratem zbiór funkcji $\{\varphi_b\}$ jest nazywany ortonormalnym, wtedy i tylko wtedy gdy iloczyn skalarny dowolnych dwóch funkcji z tej przestrzeni jest równy:

$$\langle \varphi_i(t), \varphi_j(t) \rangle = \int_I \varphi_i(t) \varphi_j(t) dt = \delta_{ij},$$

gdzie δ_{ij} jest deltą Kroneckera ($\delta_{ij} = 1$, gdy $i = j$ oraz 0 poza tym). Najczęściej wybieranymi układami funkcji bazowych są bazy Fouriera oraz Legendre'a. Baza Fouriera zdefiniowana jest w następujący sposób:

$$\begin{aligned} \varphi_0(t) &= \frac{1}{\sqrt{T}}, \\ \varphi_{2k-1}(t) &= \sqrt{\frac{2}{T}} \sin \frac{2\pi kt}{T}, \end{aligned}$$

$$\varphi_{2k}(t) = \sqrt{\frac{2}{T}} \cos \frac{2\pi kt}{T},$$

gdzie $t \in [0, T]$, $k = 1, 2, \dots$

Natomiast baza wielomianów Legendre'a na przestrzeni $L_2([-1, 1])$ jest zdefiniowana następująco:

$$\tilde{p}_k(x) = \sqrt{\frac{2k+1}{2}} p_k(x),$$

gdzie

$$p_{k+1}(x) = \frac{1}{k+1} [(2k+1)xp_k(x) - kp_{k-1}(x)], \quad k \geq 1,$$

$$p_0(x) = 1, \quad p_1(x) = x.$$

W przypadku bazy Legendre'a ograniczenie się do przestrzeni $L_2([-1, 1])$ nie powoduje utraty ogólności, gdyż każdy skończony przedział $[a, b]$ można przekształcić na przedział $[-1, 1]$ za pomocą przekształcenia:

$$x = \frac{2}{b-a}t - \frac{b+a}{b-a}, \quad t \in [a, b], \quad x \in [-1, 1].$$

W pracy Górecki i Krzyśko (2012a) pokazano, że baza Fouriera prowadzi do minimalnej liczby wyrazów rozwinięcia funkcji w szereg, co jest cechą nadzwyczaj pożądaną. Np. baza Legendre'a prowadzi częstokroć do liczby wyrazów dziesięciokrotnie większej do bazy Fouriera. Stąd rekomendowany jest wybór bazy Fouriera.

Niech $\mathbf{x} = (x_1, x_2, \dots, x_J)'$ będzie wektorem obserwacji, $\mathbf{c} = (c_0, c_1, \dots, c_B)'$ nieznanymi współczynnikami oraz $\Phi(t)$ macierzą wymiaru $J \times (B+1)$, której elementami są wartości orthonormalnych funkcji bazowych $\varphi_b(t_j)$ w kolejnych punktach czasowych t_j , $b = 0, 1, \dots, B$, $j = 1, 2, \dots, J$. Wektor $\mathbf{c} = (c_0, \dots, c_B)'$ w wyrażeniu (1.1) jest estymowany metodą najmniejszych kwadratów, w ten sposób aby minimalizować funkcję:

$$S(\mathbf{c}) = (\mathbf{x} - \Phi(t)\mathbf{c})' (\mathbf{x} - \Phi(t)\mathbf{c}).$$

Różniczkując funkcję $S(\mathbf{c})$ względem \mathbf{c} , otrzymujemy estymator najmniejszych kwadratów postaci:

$$\hat{\mathbf{c}} = (\Phi'(t)\Phi(t))^{-1} \Phi'(t)\mathbf{x}.$$

Estymacja wektora \mathbf{c} metodą najmniejszych kwadratów jest w tej tematyce powszechnie stosowana, poczynając od monografii Ramsaya i Silvermana (2005) i kończąc na pracy przeglądowej Cuevasa (2014). W pracy Górecki, Krzyśko, Wołyński (2015) porównane zostały trzy typy estymacji wektora \mathbf{c} : klasyczna metoda najmniejszych kwadratów, metoda oparta na funkcji logistycznej oraz metoda bazująca na tzw. lokalnej wygładzarce liniowej wykorzystującej funkcje jądrowe. Szczególnym przypadkiem ostatniej metody jest jądrowy estymator Nadaraya-Watsona. Okazało się, że klasyczna metoda najmniejszych kwadratów i funkcja logistyczna dają

porównywalne wyniki, nieco gorsze wyniki daje wygładzarka liniowa, a estymator Nadarayi-Watsona jest zdecydowanie najgorszy. W świetle tych badań użycie estymatora najmniejszych kwadratów jest uprawnione.

Założmy, że mamy n niezależnych zbiorów $\{(t_{i1}, x_{i1}), \dots, (t_{iJ}, x_{iJ})\}$, $i = 1, \dots, n$. Każdy z tych zbiorów z osobna można wygładzić za pomocą pewnej funkcji ciągłej postaci:

$$x_i(t) = \sum_{b=0}^{B_i} \hat{c}_{ib} \varphi_b(t), \quad i = 1, \dots, n, \quad t \in I.$$

Otrzymamy wówczas n wartości B_i , $i = 1, 2, \dots, n$. Dobór optymalnej wartości B_i i wspólnej wartości B omówiono w Podrozdziale 2.3. Dalej zakładając będziemy, że funkcja ciągła $x_i(t)$ wygładzająca zbiór $\{(t_{i1}, x_{i1}), \dots, (t_{iJ}, x_{iJ})\}$ ma następującą postać:

$$x_i(t) = \sum_{b=0}^B \hat{c}_{ib} \varphi_b(t), \quad i = 1, \dots, n, \quad t \in I. \quad (1.2)$$

W dotychczasowych pracach dotyczącej tematyki analizy danych, obiekty są scharakteryzowane za pomocą tylko i wyłącznie jednej cechy obserwowanej w wielu momentach czasowych (wielowymiarowość ze względu na czas). Nasze dalsze rozważania uogólnimy teraz na przypadek $p \geq 2$ cech. Wówczas dane składają się z n niezależnych funkcji wektorowych $\mathbf{x}_i(t) = (x_{i1}(t), x_{i2}(t), \dots, x_{ip}(t))'$, $i = 1, \dots, n$, przy czym składowe $x_{ib}(t)$ są klasycznymi danymi funkcjonalnymi postaci (1.2). Zbiór danych $\{\mathbf{x}_1(t), \dots, \mathbf{x}_n(t)\}$ nazywa się zbiorem wielozmiennych danych funkcjonalnych.

Założmy, że d -ta składowa funkcji wektorowej $\mathbf{x}(t)$ może być reprezentowana za pomocą skończonej liczby ortonormalnych funkcji bazowych φ_b :

$$x_d(t) = \sum_{b=0}^{B_d} c_{db} \varphi_b(t), \quad t \in I, \quad d = 1, 2, \dots, p,$$

gdzie c_{db} są zmiennymi losowymi takimi, że $E(c_{db}) = 0$, $\text{Var}(c_{db}) < \infty$, $d = 1, 2, \dots, p$, $b = 0, \dots, B_d$. Niech

$$\mathbf{c} = (c_{10}, \dots, c_{1B_1}, \dots, c_{p0}, \dots, c_{pB_p})'$$

będzie wektorem współczynników podlegających estymacji metodą najmniejszych kwadratów, natomiast

$$\Phi(t) = \begin{bmatrix} \varphi'_{B_1}(t) & \mathbf{0}' & \dots & \mathbf{0}' \\ \mathbf{0}' & \varphi'_{B_2}(t) & \dots & \mathbf{0}' \\ \dots & \dots & \dots & \dots \\ \mathbf{0}' & \mathbf{0}' & \dots & \varphi'_{B_p}(t) \end{bmatrix} \quad (1.3)$$

macierzą, której elementami są ortonormalne funkcje bazowe, przy czym $\varphi_{B_d}(t) = (\varphi_0(t), \dots, \varphi_{B_d}(t))'$ jest wektorem ortonormalnych funkcji bazowych odpowiadającym

d -tej składowej funkcji wektorowej $\mathbf{x}(t)$, $d = 1, \dots, p$, oraz $K = B_1 + B_2 + \dots + B_p$. Wówczas funkcję wektorową $\mathbf{x}(t)$ możemy zapisać równoważnie w następującej postaci:

$$\mathbf{x}(t) = \Phi(t)\mathbf{c}, \quad t \in I, \quad \mathbb{E}(\mathbf{c}) = \mathbf{0}. \quad (1.4)$$

Rozdział 2

Składowe główne dla wielozmiennych danych funkcjonalnych

2.1 Klasyczne składowe główne

Metoda składowych głównych jest jedną z najbardziej popularnych metod redukcji wymiaru danych, opisanych za pomocą dużej liczby skorelowanych zmiennych. W tym celu dane pierwotne (oryginalne) przekształca się w nowy zbiór zmiennych, tzw. składowych głównych, tak aby nowe zmienne miały maksymalną wariancję i były ze sobą nieskorelowane. Pozwala to m.in. na przedstawienie danych wielowymiarowych na płaszczyźnie (płaszczyzna dwóch pierwszych składowych głównych), zaobserwowanie obiektów podobnych czy też odstających. Składowe główne zostały zaproponowane przez Pearsona (1901), a następnie rozwinięte przez Hotellinga (1936). Poniżej omówimy ich konstrukcję.

Niech $\mathbf{X} = (X_1, X_2, \dots, X_p)' \in \mathbb{R}^p$ będzie p -wymiarowym wektorem losowym z zerowym wektorem wartości oczekiwanych $\boldsymbol{\mu} = \mathbf{0}$ i nieujemnie określoną macierzą kowariancji $\boldsymbol{\Sigma}$. W pierwszym kroku szukamy unormowanej kombinacji liniowej $U = u_1X_1 + u_2X_2 + \dots + u_pX_p = \mathbf{u}'\mathbf{X}$, $\mathbf{u}'\mathbf{u} = 1$, $\mathbf{u} = (u_1, \dots, u_p)' \in \mathbb{R}^p$, składowych wektora \mathbf{X} , maksymalizującej wariancję nowej zmiennej U , tzn. szukamy takiego wektora $\mathbf{u} \in \mathbb{R}^p$, aby:

$$\lambda = \max_{\mathbf{u} \in \mathbb{R}^p} \text{Var}(U) = \max_{\mathbf{u} \in \mathbb{R}^p} \text{Var}(\mathbf{u}'\mathbf{X}) = \max_{\mathbf{u} \in \mathbb{R}^p} \mathbf{u}' \text{Var}(\mathbf{X}) \mathbf{u} = \max_{\mathbf{u} \in \mathbb{R}^p} \mathbf{u}' \boldsymbol{\Sigma} \mathbf{u}, \quad (2.1)$$

przy dodatkowym warunku ograniczającym

$$\mathbf{u}'\mathbf{u} = 1, \quad \mathbf{u} = (u_1, \dots, u_p)' \in \mathbb{R}^p.$$

Stosując metodę mnożników Lagrange'a i rozwiązując równanie macierzowe

$$(\boldsymbol{\Sigma} - \lambda_1 \mathbf{I}_p) \mathbf{u}_1 = \mathbf{0}, \quad \mathbf{u}'_1 \mathbf{u}_1 = 1,$$

znajdujemy wektor \mathbf{u}_1 , który maksymalizuje wyrażenie (2.1), tzn.

$$\lambda_1 = \max_{\mathbf{u} \in \mathbb{R}^p} \text{Var}(U) = \text{Var}(U_1) = \mathbf{u}'_1 \boldsymbol{\Sigma} \mathbf{u}_1, \quad (2.2)$$

przy dodatkowym warunku ograniczającym

$$\mathbf{u}'_1 \mathbf{u}_1 = 1, \quad \mathbf{u}_1 = (u_{11}, \dots, u_{1p})' \in \mathbb{R}^p.$$

Unormowana kombinacja liniowa $U_1 = u_{11}X_1 + u_{12}X_2 + \dots + u_{1p}X_p = \mathbf{u}'_1 \mathbf{X}$, gdzie $\mathbf{u}_1 \in \mathbb{R}^p$ jest unormowanym wektorem własnym macierzy $\boldsymbol{\Sigma}$ odpowiadającym największej wartości własnej λ_1 , nazywa się pierwszą składową główną wektora \mathbf{X} .

Następnie tworzymy drugą unormowaną kombinację liniową $U_2 = \mathbf{u}'_2 \mathbf{X}$, mającą maksymalną wariancję (maksymalizacja wyrażenia (2.1)), która nie jest skorelowana z pierwszą składową główną U_1 , tzn. $\text{Cov}(U_1, U_2) = 0$. Wartość λ_2 oraz wektor $\mathbf{u}_2 \in \mathbb{R}^p$ znajdujemy analogicznie jak powyżej korzystając z metody mnożników Lagrange'a. Unormowana kombinacja liniowa $U_2 = u_{21}X_1 + u_{22}X_2 + \dots + u_{2p}X_p = \mathbf{u}'_2 \mathbf{X}$, $\mathbf{u}'_2 \mathbf{u}_2 = 1$, $\mathbf{u}_2 \in \mathbb{R}^p$, spełniająca warunek $\text{Cov}(U_1, U_2) = 0$, gdzie \mathbf{u}_2 jest unormowanym wektorem własnym macierzy $\boldsymbol{\Sigma}$ odpowiadającym drugiej (największej) wartości własnej λ_2 , $\lambda_1 \geq \lambda_2$, nazywa się drugą składową główną wektora \mathbf{X} .

Taki proces tworzenia kolejnych zmiennych możemy kontynuować k razy, ($k < p$). Dla $(k + 1)$ -szej składowej głównej ponownie tworzymy unormowaną kombinację liniową $U_{k+1} = \mathbf{u}'_{k+1} \mathbf{X}$, $\mathbf{u}_{k+1} \in \mathbb{R}^p$, mającą maksymalną wariancję (maksymalizacja wyrażenia (2.1)), która nie jest skorelowana z k pierwszymi składowymi głównymi U_κ , $\kappa = 1, \dots, k$, tzn. $\text{Cov}(U_{\kappa_1}, U_{\kappa_2}) = 0$ dla $\kappa_1 \neq \kappa_2$. Wartość λ_{k+1} oraz wektor $\mathbf{u}_{k+1} \in \mathbb{R}^p$ znajdujemy analogicznie jak poprzednio korzystając z metody mnożników Lagrange'a. Unormowana kombinacja liniowa $U_{k+1} = u_{k+1,1}X_1 + u_{k+1,2}X_2 + \dots + u_{k+1,p}X_p = \mathbf{u}'_{k+1} \mathbf{X}$, spełniająca warunek $\text{Cov}(U_{\kappa_1}, U_{\kappa_2}) = \delta_{\kappa_1 \kappa_2}$, gdzie \mathbf{u}_{k+1} jest unormowanym wektorem własnym macierzy $\boldsymbol{\Sigma}$ odpowiadającym $(k + 1)$ -szej wartości własnej λ_{k+1} , $\lambda_1 \geq \dots \geq \lambda_{k+1}$, nazywa się $(k + 1)$ -szą składową główną wektora \mathbf{X} .

W praktyce macierz $\boldsymbol{\Sigma}$ nie jest znana i estymujemy ją na podstawie n niezależnych realizacji $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n$ wektora losowego \mathbf{X} . Estymator macierzy $\boldsymbol{\Sigma}$ jest postaci:

$$\hat{\boldsymbol{\Sigma}} = \frac{1}{n} \sum_{i=1}^n (\mathbf{x}_i - \bar{\mathbf{x}})(\mathbf{x}_i - \bar{\mathbf{x}})'$$

W przypadku, gdy obserwacje te są scentrowane (tzn. $\bar{\mathbf{x}} = \mathbf{0}$), estymator ten możemy zapisać w postaci:

$$\hat{\boldsymbol{\Sigma}} = \frac{1}{n} \sum_{i=1}^n \mathbf{x}_i \mathbf{x}'_i = \frac{1}{n} (\mathbf{x}_1, \dots, \mathbf{x}_n) (\mathbf{x}_1, \dots, \mathbf{x}_n)'$$

Rozwiązując równanie $|\hat{\boldsymbol{\Sigma}} - \hat{\lambda} \mathbf{I}_p| = 0$ otrzymamy oszacowania p wartości własnych $\hat{\lambda}_1, \dots, \hat{\lambda}_p$. Odpowiadające im wektory własne $\hat{\mathbf{u}}_1, \dots, \hat{\mathbf{u}}_p$ otrzymujemy z układu równań $(\hat{\boldsymbol{\Sigma}} - \hat{\lambda}_k \mathbf{I}_p) \hat{\mathbf{u}}_k = \mathbf{0}$, $k = 1, \dots, p$. Składowe główne \hat{U}_{ik} otrzymane z próby są wówczas postaci $\hat{U}_{ik} = \hat{\mathbf{u}}'_k \mathbf{x}_i$, $i = 1, 2, \dots, n$, $k = 1, 2, \dots, p$.

2.2 Składowe główne dla wielozmiennych danych funkcjonalnych

Pionierską pracą dotyczącą analizy składowych głównych dla jednowymiarowych danych funkcjonalnych w przestrzeni L_2 jest praca Dauxois, Pousse i Romain (1982). Prace Besse i Ramsay (1986) oraz Besse (1988) dotyczą konstrukcji składowych głównych dla jednowymiarowych danych funkcjonalnych w przestrzeni funkcjonalnej H^m , w której zakłada się istnienie $m - 1$ absolutnie ciągłych pochodnych i całkowalność z kwadratem m -tej pochodnej, $m \geq 1$. Kolejną pracą związaną z funkcjonalną analizą składowych głównych, o której warto wspomnieć, jest praca Ramsaya i Dalzella (1991). W pracy tej po raz pierwszy pojawia się termin „dane funkcjonalne”. W już dojrzałej formie konstrukcja składowych głównych dla jednowymiarowych danych funkcjonalnych przedstawiona jest w monografiach Ramsaya i Silvermana (1997, 2002, 2005).

Wkładem ośrodka poznańskiego do tej tematyki są prace Góreckiego i Krzyński (2012a, 2012b) oraz Krzyński i innych (2013). W pierwszej z nich zbadano dobór bazy i liczby jej elementów w procesie aproksymacji szeregów czasowych funkcjami ciągłymi. W drugiej z nich przedstawiono wersję jądrową funkcjonalnych składowych głównych (składowe główne nieliniowe), a trzecia jest opisem wykorzystania funkcjonalnych składowych głównych w doświadczalnictwie rolniczym.

W dotychczasowych pracach, badane obiekty były charakteryzowane tylko i wyłącznie za pomocą jednej cechy obserwowanej dynamicznie (wielowymiarowość ze względu na czas). Opisem bardziej spójnym z klasyczną analizą składowych głównych jest opis badanych obiektów za pomocą wielu cech obserwowanych w wielu momentach czasowych (dane podwójnie wielowymiarowe).

Dane funkcjonalne wygodnie jest traktować jako realizacje $x_i(t)$ pewnego procesu stochastycznego $X(t)$ z ciągłym parametrem czasu $t \in I$, gdzie funkcje $x_i(t)$ są postaci (1.1). Uogólnimy teraz nasze rozważania na przypadek wielowymiarowych danych funkcjonalnych $\mathbf{x}_i(t)$ będących realizacjami wektorowego procesu stochastycznego $\mathbf{X}(t) \in L_2(I^p)$. Wykorzystując fakt, że funkcja wektorowa $\mathbf{x}(t)$ jest postaci (1.4), proces stochastyczny $\mathbf{X}(t)$ można zapisać w postaci $\mathbf{X}(t) = \Phi(t)\mathbf{c}$, gdzie $E(\mathbf{c}) = \mathbf{0}$. W analizie składowych głównych dla wielozmiennych danych funkcjonalnych interesuje nas znalezienie unormowanego iloczynu skłarnego

$$U = \langle \mathbf{u}(t), \mathbf{X}(t) \rangle = \int_I \mathbf{u}'(t)\mathbf{X}(t)dt,$$

mającego maksymalną wariancję dla wszystkich $\mathbf{u}(t)$ takich, że $\mathbf{u}(t) \in L_2(I^p)$, $\langle \mathbf{u}(t), \mathbf{u}(t) \rangle = 1$. Niech

$$\lambda_1 = \sup_{\mathbf{u}(t) \in L_2(I^p)} \text{Var}(\langle \mathbf{u}(t), \mathbf{X}(t) \rangle) = \text{Var}(\langle \mathbf{u}_1(t), \mathbf{X}(t) \rangle),$$

gdzie $\langle \mathbf{u}_1(t), \mathbf{u}_1(t) \rangle = 1$. Iloczyn skalarny $U_1 = \langle \mathbf{u}_1(t), \mathbf{X}(t) \rangle$ nazywa się pierwszą funkcjonalną składową główną procesu $\mathbf{X}(t)$, a funkcja wektorowa $\mathbf{u}_1(t)$ pierwszą wektorową funkcją

wagową. Wyrażenie $(\lambda_1, \mathbf{u}_1(t))$ nazywać będziemy pierwszym układem głównym procesu losowego $\mathbf{X}(t)$.

Analogicznie szukamy drugiej funkcjonalnej składowej głównej $U_2 = \langle \mathbf{u}_2(t), \mathbf{X}(t) \rangle$ maksymalizującej wyrażenie $\text{Var}(\langle \mathbf{u}(t), \mathbf{X}(t) \rangle)$ takiej, że $\langle \mathbf{u}_2(t), \mathbf{u}_2(t) \rangle = 1$ i która nie jest skorelowana z pierwszą funkcjonalną składową główną U_1 , tzn. spełnia warunek $\langle \mathbf{u}_1(t), \mathbf{u}_2(t) \rangle = 0$. Wyrażenie $(\lambda_2, \mathbf{u}_2(t))$ nazywać będziemy drugim układem głównym procesu $\mathbf{X}(t)$.

W dalszym ciągu zakładamy, że $\mathbf{X}(t)$ ma postać (1.4). Można także założyć (Ramsay i Silverman (2005)), że wektorowa funkcja wagowa $\mathbf{u}(t)$ i proces losowy $\mathbf{X}(t)$ leżą w tej samej przestrzeni, tzn. funkcja $\mathbf{u}(t)$ może zostać zapisana w następującej postaci:

$$\mathbf{u}(t) = \Phi(t)\mathbf{u},$$

gdzie $\mathbf{u} \in \mathbb{R}^{K+p}$. Wówczas

$$\langle \mathbf{u}(t), \mathbf{X}(t) \rangle = \langle \Phi(t)\mathbf{u}, \Phi(t)\mathbf{c} \rangle = \mathbf{u}' \langle \Phi(t), \Phi(t) \rangle \mathbf{c} = \mathbf{u}' I_{K+p} \mathbf{c} = \mathbf{u}' \mathbf{c},$$

przy czym

$$E(\langle \mathbf{u}(t), \mathbf{X}(t) \rangle) = \mathbf{u}' E(\mathbf{c}) = \mathbf{u}' \mathbf{0} = 0,$$

$$\text{Var}(\langle \mathbf{u}(t), \mathbf{X}(t) \rangle) = \mathbf{u}' E(\mathbf{c}\mathbf{c}') \mathbf{u} = \mathbf{u}' \Sigma \mathbf{u}.$$

Rozważmy pierwszą funkcjonalną składową główną procesu losowego $\mathbf{X}(t)$:

$$\lambda_1 = \sup_{\mathbf{u}(t) \in L_2(IP)} \text{Var}(\langle \mathbf{u}(t), \mathbf{X}(t) \rangle) = \text{Var}(\langle \mathbf{u}_1(t), \mathbf{X}(t) \rangle), \quad (2.3)$$

gdzie $\langle \mathbf{u}_1(t), \mathbf{u}_1(t) \rangle = 1$.

W ogólności k -ta funkcjonalna składowa główna $U_k = \langle \mathbf{u}_k(t), \mathbf{X}(t) \rangle$ spełnia warunki:

$$\lambda_k = \sup_{\mathbf{u}(t) \in L_2(IP)} \text{Var}(\langle \mathbf{u}(t), \mathbf{X}(t) \rangle) = \text{Var}(\langle \mathbf{u}_k(t), \mathbf{X}(t) \rangle),$$

gdzie

$$\langle \mathbf{u}_k(t), \mathbf{u}_l(t) \rangle = \begin{cases} 1, & k = l, \\ 0, & k \neq l. \end{cases}$$

Wyrażenie $(\lambda_k, \mathbf{u}_k(t))$ nazywać będziemy k -tym układem głównym procesu $\mathbf{X}(t)$.

Zdefiniujmy teraz składowe główne dla wektora losowego \mathbf{c} . Otóż k -ta składowa główna $U_k^* = \langle \mathbf{u}_k, \mathbf{c} \rangle$ wektora losowego \mathbf{c} spełnia następujący warunek:

$$\gamma_k = \sup_{\mathbf{u} \in \mathbb{R}^{K+p}} \text{Var}(\langle \mathbf{u}, \mathbf{c} \rangle) = \sup_{\mathbf{u} \in \mathbb{R}^{K+p}} \mathbf{u}' \text{Var}(\mathbf{c}) \mathbf{u} = \sup_{\mathbf{u} \in \mathbb{R}^{K+p}} \mathbf{u}' \Sigma \mathbf{u} = \mathbf{u}'_k \Sigma \mathbf{u}_k,$$

przy dodatkowym warunku ograniczającym

$$\mathbf{u}'_k \mathbf{u}_l = \begin{cases} 1, & k = l, \\ 0, & k \neq l, \end{cases}$$

gdzie $k, l = 1, \dots, K, K = B_1 + \dots + B_p$. Wyrażenie (γ_k, \mathbf{u}_k) nazywać będziemy k -tym układem głównym wektora \mathbf{c} .

Wyrażenie (2.3) jest równoważne wyrażeniu

$$\gamma_1 = \sup_{\mathbf{u} \in \mathbb{R}^{K+p}} \text{Var}(\langle \mathbf{u}, \mathbf{c} \rangle) = \sup_{\mathbf{u} \in \mathbb{R}^{K+p}} \mathbf{u}' \text{Var}(\mathbf{c}) \mathbf{u} = \mathbf{u}'_1 \Sigma \mathbf{u}_1 = \lambda_1,$$

gdzie $\mathbf{u}'_1 \mathbf{u}_1 = 1$. Oczywiście jest to definicja pierwszej składowej głównej wektora losowego \mathbf{c} . Z drugiej strony, jeżeli zaczniemy rozważania od pierwszego układu głównego wektora losowego \mathbf{c} zdefiniowanego przez (γ_1, \mathbf{u}_1) , otrzymamy pierwszy układ główny procesu losowego $\mathbf{X}(t)$ z równań:

$$\lambda_1 = \gamma_1, \quad \mathbf{u}_1(t) = \Phi(t) \mathbf{u}_1.$$

Możemy rozszerzyć nasze rozważania na drugi układ główny i kolejne.

Wyznaczanie klasycznych składowych głównych dla wektorów losowych zostało opisane w Podrozdziale 2.1. Udowodniliśmy zatem twierdzenie, które mówi, że analiza składowych głównych dla wektorów losowych jest równoważna analizie składowych głównych dla procesów losowych.

Twierdzenie 2.1. *k -ty układ główny $(\lambda_k, \mathbf{u}_k(t))$ procesu losowego $\mathbf{X}(t)$ jest zależny od k -tego układu głównego (γ_k, \mathbf{u}_k) wektora losowego \mathbf{c} poprzez następujące równości:*

$$\lambda_k = \gamma_k, \quad \mathbf{u}_k(t) = \Phi(t) \mathbf{u}_k, \quad t \in I,$$

gdzie $k = 1, \dots, K + p, K = B_1 + B_2 + \dots + B_p$.

Z Twierdzenia 2.1 wynika zatem, że wystarczy znaleźć k -ty układ główny wektora losowego \mathbf{c} , a następnie korzystając ze znanej postaci macierzy $\Phi(t)$ ortonormalnych funkcji bazowych, wyznaczyć k -ty układ główny procesu losowego $\mathbf{X}(t)$. Oznacza to, że konstrukcja składowych głównych dla procesu losowego $\mathbf{X}(t) = \Phi(t) \mathbf{c}$ sprowadza się w istocie do konstrukcji składowych głównych dla wektora losowego \mathbf{c} . Składowa główna U_k ma bowiem postać $U_k = \mathbf{u}'_k \mathbf{c}$, gdzie \mathbf{u}_k jest wektorem własnym macierzy kowariancji Σ wektora losowego \mathbf{c} , $k = 1, 2, \dots, K + p$.

Analiza składowych głównych dla wektora losowego \mathbf{c} bazuje na macierzy Σ . W praktyce macierz ta nie jest znana. Estymujemy ją na podstawie n niezależnych realizacji $\mathbf{x}_1(t), \mathbf{x}_2(t), \dots, \mathbf{x}_n(t)$ postaci $\mathbf{x}_i(t) = \Phi(t) \hat{\mathbf{c}}_i$ procesu losowego $\mathbf{X}(t)$, gdzie $\hat{\mathbf{c}}_i$ jest scentrowanym estymatorem wektora \mathbf{c}_i , $i = 1, 2, \dots, n$.

Niech $\hat{\mathbf{C}} = (\hat{\mathbf{c}}_1, \hat{\mathbf{c}}_2, \dots, \hat{\mathbf{c}}_n)'$. Wówczas estymator macierzy Σ jest następującej postaci:

$$\hat{\Sigma} = \frac{1}{n} \hat{\mathbf{C}}' \hat{\mathbf{C}}.$$

Jeżeli $n > K + p$, wówczas macierz $\hat{\Sigma}$ jest dodatnio określona z prawdopodobieństwem 1.

Niech ponadto $\hat{\gamma}_1 \geq \hat{\gamma}_2 \geq \dots \geq \hat{\gamma}_s$ będą niezerowymi wartościami własnymi macierzy $\hat{\Sigma}$, a $\hat{\mathbf{u}}_1, \hat{\mathbf{u}}_2, \dots, \hat{\mathbf{u}}_s$ odpowiadającymi im wektorami własnymi, gdzie $s = \text{rzęd}(\hat{\Sigma})$.

Wówczas k -ty układ główny procesu losowego $\mathbf{X}(t)$ wyznaczony z próby ma następującą postać:

$$(\hat{\lambda}_k = \hat{\gamma}_k, \hat{\mathbf{u}}_k(t) = \Phi(t)\hat{\mathbf{u}}_k), \quad k = 1, \dots, s.$$

Stąd współczynniki rzutu i -tej realizacji $\mathbf{x}_i(t)$ procesu losowego $\mathbf{X}(t)$ na k -tą funkcjonalną składową główną są równe:

$$\hat{U}_{ik} = \langle \hat{\mathbf{u}}_k(t), \mathbf{x}_i(t) \rangle = \langle \Phi(t)\hat{\mathbf{u}}_k, \Phi(t)\hat{\mathbf{c}}_i \rangle = \hat{\mathbf{u}}_k' \langle \Phi(t), \Phi(t) \rangle \hat{\mathbf{c}}_i = \hat{\mathbf{u}}_k' \hat{\mathbf{c}}_i,$$

dla $i = 1, 2, \dots, n, k = 1, 2, \dots, s$.

Ostatecznie współczynniki rzutu i -tej realizacji $\mathbf{x}_i(t)$ procesu losowego $\mathbf{X}(t)$ na płaszczyznę dwóch pierwszych funkcjonalnych składowych głównych z próby są równe $(\hat{\mathbf{u}}_1' \hat{\mathbf{c}}_i, \hat{\mathbf{u}}_2' \hat{\mathbf{c}}_i)$, $i = 1, 2, \dots, n$.

Prawdziwe są następujące własności funkcjonalnych składowych głównych:

Własność 2.1. *Wariancje składowych*

Niech \mathbf{u}_j będzie wektorem własnym macierzy kowariancji Σ wektora losowego \mathbf{c} w reprezentacji $\mathbf{X}(t) = \Phi(t)\mathbf{c}$ odpowiadającym wartości własnej λ_j , unormowanemu tak, że $\mathbf{u}_j' \mathbf{u}_j = 1$ i niech U_j oznacza j -tą składową główną, $j = 1, 2, \dots, K + p$. Wówczas

$$\text{Var}(U_j) = \lambda_j.$$

Wiemy, że U_j jest postaci $U_j = \mathbf{u}_j' \mathbf{c}$. Stąd

$$\text{Var}(U_j) = \text{Var}(\mathbf{u}_j' \mathbf{c}) = \mathbf{u}_j' \text{Var}(\mathbf{c}) \mathbf{u}_j = \mathbf{u}_j' \Sigma \mathbf{u}_j.$$

Ale \mathbf{u}_j jest wektorem własnym macierzy Σ . Stąd $(\Sigma - \lambda_j \mathbf{I}_p) \mathbf{u}_j = 0$ lub równoważnie $\Sigma \mathbf{u}_j = \lambda_j \mathbf{u}_j$ lub $\mathbf{u}_j' \Sigma \mathbf{u}_j = \lambda_j \mathbf{u}_j' \mathbf{u}_j = \lambda_j$, bo $\mathbf{u}_j' \mathbf{u}_j = 1$. Ponieważ $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_{K+p}$, to składowa główna U_1 ma maksymalną wariancję λ_1 , natomiast składowa główna U_{K+p} ma minimalną wariancję λ_{K+p} .

Własność 2.2. *Ortogonalność i nieskorelowanie składowych*

Składowe główne U_1, U_2, \dots, U_{K+p} są wzajemnie ortogonalne, tzn. dla $i \neq j$

$$\langle U_i, U_j \rangle = U_i' U_j = (\mathbf{u}_i' \mathbf{c})' \mathbf{u}_j' \mathbf{c} = \mathbf{c}' \mathbf{u}_i \mathbf{u}_j' \mathbf{c} = 0,$$

ponieważ $\mathbf{u}_i \mathbf{u}_j' = 0$ dla $i \neq j$.

Składowe główne U_1, U_2, \dots, U_{K+p} są wzajemnie nieskorelowane. Współczynnik korelacji między składowymi $U_i = \mathbf{u}_i' \mathbf{c}$ oraz $U_j = \mathbf{u}_j' \mathbf{c}$ jest równy:

$$\rho(U_i, U_j) = \frac{\text{Cov}(U_i, U_j)}{\sqrt{\text{Var}(U_i) \text{Var}(U_j)}} = \frac{\mathbf{u}_i' \Sigma \mathbf{u}_j}{\sqrt{\lambda_i \lambda_j}}, \quad i, j = 1, 2, \dots, K + p, i \neq j.$$

Ponieważ \mathbf{u}_i oraz \mathbf{u}_j są wektorami własnymi macierzy kowariancji Σ , to $\Sigma \mathbf{u}_j = \lambda_j \mathbf{u}_j$ lub równoważnie $\mathbf{u}_i' \Sigma \mathbf{u}_j = \lambda_j \mathbf{u}_i' \mathbf{u}_j = 0$, bo $\mathbf{u}_i' \mathbf{u}_j = 0$ dla wszystkich $i \neq j$. Zatem $\rho(U_i, U_j) = 0$ dla wszystkich $i \neq j$.

Własność 2.3. *Niezmienniczość składowych*

Składowe główne nie są niezmiennicze ze względu na zmianę skali. Wiemy, że k -ta składowa główna jest postaci $U_k = \mathbf{u}'_k \mathbf{c}$, gdzie \mathbf{c} jest wektorem losowym w reprezentacji $\mathbf{X}(t) = \Phi(t)\mathbf{c}$, a \mathbf{u}_k jest wektorem własnym macierzy kowariancji Σ wektora losowego \mathbf{c} . Niech $\mathbf{d} = \mathbf{D}\mathbf{c}$, gdzie \mathbf{D} jest macierzą nieosobliwą. Wówczas

$$\text{Var}(\mathbf{d}) = \text{Var}(\mathbf{D}\mathbf{c}) = \mathbf{D} \text{Var}(\mathbf{c})\mathbf{D}' = \mathbf{D}\Sigma\mathbf{D}'$$

oraz k -ta składowa główna odpowiadająca macierzy $\mathbf{D}\Sigma\mathbf{D}'$ jest postaci $V_k = \mathbf{v}'_k \mathbf{d}$, gdzie \mathbf{v}_k jest k -tym wektorem własnym macierzy $\mathbf{D}\Sigma\mathbf{D}'$. Aby $U_k = V_k$ musi zajść równość $\mathbf{u}'_k \mathbf{c} = \mathbf{v}'_k \mathbf{D}\mathbf{c}$ lub równoważnie $\mathbf{u}_k = \mathbf{D}'\mathbf{v}_k$ lub $\mathbf{v}_k = (\mathbf{D}')^{-1}\mathbf{u}_k$. Jednakże $\mathbf{v}_k = (\mathbf{D}')^{-1}\mathbf{u}_k$ nie jest wektorem własnym macierzy $\mathbf{D}\Sigma\mathbf{D}'$.

Własność 2.4. *Odległości*

Odległości w przestrzeni wektorowych procesów losowych $\mathbf{X}(t)$ są równe odległościom w przestrzeni wektorowej wszystkich składowych głównych, tj.

$$\langle \mathbf{X}_i(t) - \mathbf{X}_j(t), \mathbf{X}_i(t) - \mathbf{X}_j(t) \rangle = (\mathbf{U}_i - \mathbf{U}_j)'(\mathbf{U}_i - \mathbf{U}_j),$$

gdzie $\mathbf{X}_i(t) = \Phi(t)\mathbf{c}_i$, $\mathbf{U}_i = (U_{i1}, U_{i2}, \dots, U_{i,K+p})'$, $U_{ik} = \mathbf{u}'_k \mathbf{c}_i$, $i = 1, 2, \dots, n$, $k = 1, 2, \dots, K + p$. Mamy

$$\begin{aligned} \langle \mathbf{X}_i(t) - \mathbf{X}_j(t), \mathbf{X}_i(t) - \mathbf{X}_j(t) \rangle &= \int_I (\mathbf{X}_i(t) - \mathbf{X}_j(t))'(\mathbf{X}_i(t) - \mathbf{X}_j(t)) dt = \\ &= \int_I (\mathbf{c}_i - \mathbf{c}_j)' \Phi'(t) \Phi(t) (\mathbf{c}_i - \mathbf{c}_j) dt = (\mathbf{c}_i - \mathbf{c}_j)' (\mathbf{c}_i - \mathbf{c}_j) \end{aligned}$$

oraz

$$\begin{aligned} (\mathbf{U}_i - \mathbf{U}_j)'(\mathbf{U}_i - \mathbf{U}_j) &= (\mathbf{c}_i - \mathbf{c}_j)' (\mathbf{u}_1 \mathbf{u}'_1 + \mathbf{u}_2 \mathbf{u}'_2 + \dots + \mathbf{u}_{K+p} \mathbf{u}'_{K+p}) (\mathbf{c}_i - \mathbf{c}_j) = \\ &= (\mathbf{c}_i - \mathbf{c}_j)' (\mathbf{c}_i - \mathbf{c}_j), \end{aligned}$$

bo $\mathbf{u}_1 \mathbf{u}'_1 + \mathbf{u}_2 \mathbf{u}'_2 + \dots + \mathbf{u}_{K+p} \mathbf{u}'_{K+p} = \mathbf{I}_{K+p}$.

Własność 2.5. *Miara zmienności*

Posługując się składowymi głównymi, badacz musi zdecydować, ile z nich jest istotnych do opisu konkretnych danych. Składowe nieistotne należy pominąć. Będą to składowe związane z najmniejszymi wartościami własnymi macierzy kowariancji Σ wektora losowego \mathbf{c} w reprezentacji $\mathbf{X}(t) = \Phi(t)\mathbf{c}$. Możemy posłużyć się następującymi kryteriami:

1. Procent wariancji

Wiemy, że k -ta składowa główna jest postaci $U_k = \mathbf{u}'_k \mathbf{c}$, gdzie \mathbf{c} jest wektorem losowym w reprezentacji $\mathbf{X}(t) = \mathbf{\Phi}(t)\mathbf{c}$, a \mathbf{u}_k jest k -tym wektorem własnym macierzy kowariancji $\mathbf{\Sigma}$ wektora losowego \mathbf{c} . Z własności (2.1) wynika, że $\text{Var}(U_k) = \lambda_k$, $k = 1, 2, \dots, K + p$. Stąd

$$\sum_{k=1}^{K+p} \text{Var}(U_k) = \sum_{k=1}^{K+p} \lambda_k = \text{tr}(\mathbf{\Sigma}).$$

W analizie składowych głównych oczekujemy, że dla pewnego małego j suma $\lambda_1 + \lambda_2 + \dots + \lambda_j$ będzie bliska wyrażeniu $\text{tr}(\mathbf{\Sigma}) = \lambda_1 + \lambda_2 + \dots + \lambda_{K+p}$. Jeśli tak jest, to j pierwszych składowych głównych wyjaśnia dobrze zmienność procesu $\mathbf{X}(t)$, a pozostałe składowe wnoszą niewiele, ponieważ ich wariancje są małe. Wskaźnik

$$\frac{\lambda_1 + \dots + \lambda_j}{\lambda_1 + \dots + \lambda_{K+p}} \cdot 100\% \quad (2.4)$$

jest procentową miarą wyjaśnienia zmienności wektorowego procesu $\mathbf{X}(t)$ przez pierwszych j składowych głównych. Jeżeli wskaźnik (2.3) jest równy lub większy od ustalonej przez nas wartości, np. 80%, to pozostałe składowe główne pomijamy.

2. Wartość średnia z wartości własnych

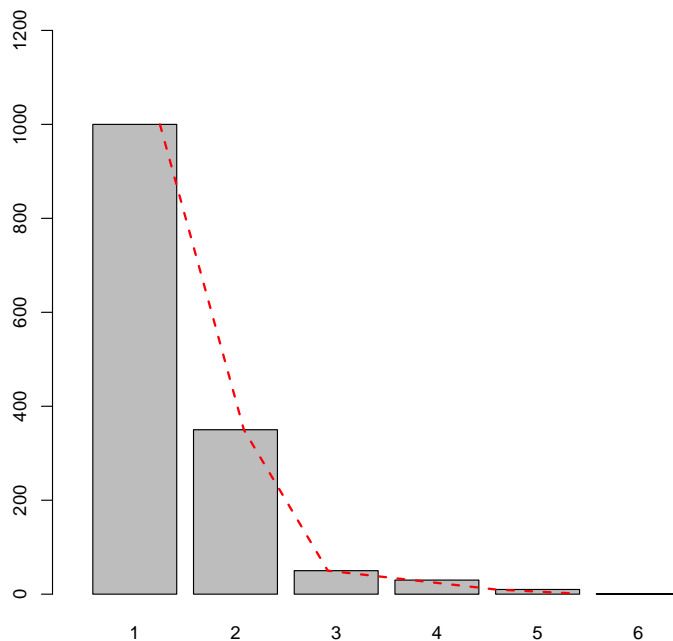
Pomijamy te składowe główne, którym odpowiadające wartości własne są mniejsze od średniej

$$\bar{\lambda} = \frac{1}{K+p} \sum_{k=1}^{K+p} \lambda_k.$$

Jest to zarazem średnia wariancja składowych wektora losowego \mathbf{c} , ponieważ $\text{tr}(\mathbf{\Sigma}) = \lambda_1 + \lambda_2 + \dots + \lambda_{K+p}$.

3. Wykres piargowy (osypisko)

Wartości własne numerujemy w porządku malejącym. Na osi odciętych zaznaczamy numery wartości własnych, na osi rzędnych zaznaczamy wielkości wartości własnych i wielkości te łączymy. Uzyskany w ten sposób wykres nazywany jest wykresem piargowym lub osypiskiem. Termin ten wprowadził Cattell (1966). Przykładowy wykres piargowy pokazany jest na Rysunku 2.1. Pierwsze dwie wartości własne istotnie różnią się od pozostałych wykazujących trend liniowy o małym spadku. Składowe główne poza dwiema pierwszymi, pomijamy.



Rysunek 2.1: Wykres piargowy.

Na zakończenie tych rozważań zajmijmy się wkładem procesu $X_j(t)$ w budowę składowej głównej U_k . Składowa główna U_k wyznaczona jest przez wektorową funkcję wagową $\mathbf{u}_k(t) = (u_{k1}(t), u_{k2}(t), \dots, u_{kp}(t))'$. Proponujemy następujące rozwiązania. Składowa $X_j(t)$ procesu wektorowego $\mathbf{X}(t) = (X_1(t), X_2(t), \dots, X_p(t))'$ ma największy wkład w budowę składowej głównej U_k w chwili t , jeżeli

$$|u_{kj}(t)| = \max_{1 \leq i \leq p} |u_{ki}(t)|.$$

Niech P_j będzie polem pod modułem składowej $u_{kj}(t)$ wektorowej funkcji wagowej, dla $t \in I$ i niech

$$P_j^* = \frac{P_j}{\sum_{i=1}^p P_i}, \quad j = 1, 2, \dots, p. \quad (2.5)$$

Składowa $X_j(t)$ procesu wektorowego $\mathbf{X}(t) = (X_1(t), X_2(t), \dots, X_p(t))'$ ma największy wkład w budowę składowej głównej U_k , dla wszystkich $t \in I$, jeżeli

$$P_j^* = \max_{1 \leq i \leq p} |P_i^*|.$$

2.3 Symulacyjny wybór liczby elementów bazy dla wielozmiennych składowych głównych

Stopień gładkości każdej funkcji $x_i(t)$ określonej wzorem (1.2) zależy od wielkości B (im mniejsza wartość B tym większy stopień wygładzenia). W rozdziale tym porównamy symulacyjnie różne sposoby wyboru liczby elementów bazy B przy konstrukcji wielozmiennych składowych głównych. W tym celu wygenerujemy dziesięć procesów losowych. Sześć procesów $X_i, i = 1, 2, 3, 4, 5, 6$, w oparciu o pracę Berrendero i in. (2011), postaci:

$$X_1(t) = -5 + t^{\frac{1}{5}}(1-t)^{6-\frac{1}{5}} + B_1 + 0,4B_2,$$

$$X_4(t) = -5 + t^{\frac{1}{5}}(1-t)^{6-\frac{4}{5}} + B_1 + 0,4B_2,$$

$$X_5(t) = -5 + t^{\frac{5}{5}}(1-t)^{6-\frac{5}{5}} + B_1 + 0,4B_2,$$

$$X_2(t) = (\sin(2\pi t^2))^5 + 0,4B_1 + B_2,$$

$$X_3(t) = (\sin(2\pi t^2))^7 + 0,4B_1 + B_2,$$

$$X_6(t) = (\sin(2\pi t^2))^9 + 0,4B_1 + B_2,$$

gdzie $B_i(t)$ są procesami Browna danymi wzorami:

$$B_i(t) = \frac{1}{\sqrt{p}} \sum_{k=1}^{Jt} X_k, \quad i = 1, 2, X_k \sim N(0, 1/25),$$

przy czym każdy proces Browna generowany jest osobno.

Ponadto wygenerujemy cztery procesy $X_i, i = 7, 8, 9, 10$ w oparciu o pracę Ferraty i Vieu (2003) oraz Preda i in. (2007), postaci:

$$X_7(t) = -5 + t^{\frac{1}{5}}(1-t)^{6-\frac{1}{5}} + U_2 h_3(t) + U_3 h_2(t) + \sqrt{0,1} \varepsilon(t),$$

$$X_8(t) = -5 + t^{\frac{1}{5}}(1-t)^{6-\frac{1}{5}} + U_3 h_1(t) + U_1 h_3(t) + \sqrt{0,1} \varepsilon(t),$$

$$X_9(t) = -5 + t^{\frac{1}{5}}(1-t)^{6-\frac{1}{5}} + U_2 h_1(t) + U_1 h_2(t) + \sqrt{0,1} \varepsilon(t),$$

$$X_{10}(t) = -5 + t^{\frac{1}{5}}(1-t)^{6-\frac{1}{5}} + U_1 h_1(t) + U_2 h_2(t) + U_3 h_3(t) + \sqrt{0,1} \varepsilon(t),$$

gdzie procesy $U_i, i = 1, 2, 3$, są generowane niezależnie z rozkładów normalnych postaci:

$$U_1 \sim N(1/2, 1/12), \quad U_2 \sim N(0, 1/12), \quad U_3 \sim N(0, 2/3),$$

natomiast $\varepsilon(t) \sim N(0, 1/5)$ jest białym szumem. Funkcje $h_i(t), i = 1, 2, 3$ są natomiast dane wzorami:

$$h_1(t) = (6 - |t - 11|)_+, \quad h_2(t) = (6 - |t - 7|)_+, \quad h_3(t) = (6 - |t - 15|)_+.$$

Przypomnijmy, że baza Fouriera prowadzi do minimalnej liczby wyrazów rozwinięcia funkcji $x(t)$ w szereg (Górecki i Krzyśko (2012a)) i taka baza będzie wykorzystywana przy estymacji opartej na metodzie najmniejszych kwadratów.

W eksperymencie symulacyjnym porównamy jak wybór kryterium informacyjnego wpływa na wyjaśnienia zmienności przez dwie pierwsze składowe główne, tj. zweryfikujemy wartość poniższego wskaźnika (im większa wartość tym większy stopień wyjaśnienia wariancji zmiennych pierwotnych):

$$\frac{\lambda_1 + \lambda_2}{\lambda_1 + \dots + \lambda_{K+p}} \cdot 100\%. \quad (2.6)$$

Do dalszej analizy wybrano 3 kryteria informacyjne: BIC, eBIC, AIC.

W pracy Shmueli (2010) przy zagadnieniach dotyczących dopasowania modelu rekomendowane jest użyciu Bayesowskiego Kryterium Informacyjnego BIC (Schwarz (1978)), które jest zdefiniowane w następujący sposób:

$$\text{BIC}(x(t)) = J \ln \left(\frac{\mathbf{e}'\mathbf{e}}{J} \right) + (B + 1) \left(\frac{\ln J}{J} \right),$$

gdzie $\mathbf{e} = (e_1, \dots, e_J)'$ jest wektorem błędów takim, że $e_j = x_j - \sum_{b=0}^B \hat{c}_b \varphi_b(t_j)$, $j = 1, 2, \dots, J$, a B wyznacza liczbę elementów bazy.

W przypadku gdy przestrzeń parametrów jest duża, w literaturze proponowane jest wykorzystanie zmodyfikowanych wersji powyższego kryterium BIC. Jedno z takich kryteriów zostało opisane w pracy Chen i Chen (2008), funkcjonującego w literaturze jak rozszerzone kryterium informacyjne eBIC. Kryterium to dane jest wzorem:

$$\text{eBIC}(x(t)) = J \ln \left(\frac{\mathbf{e}'\mathbf{e}}{J} \right) + (B + 1) (\ln J + 2\gamma \ln(B_{max} + 1)),$$

gdzie B_{max} oznacza maksymalną liczbę elementów bazy. Do dalszej analizy można przyjąć, że $\gamma = 0,5$. Wówczas otrzymamy:

$$\text{eBIC}(x(t)) = J \ln \left(\frac{\mathbf{e}'\mathbf{e}}{J} \right) + (B + 1) (\ln J + \ln(B_{max} + 1)).$$

Ostatnim wykorzystywanym kryterium będzie kryterium informacyjne AIC zaproponowane przez Akaike (1973), które jest zazwyczaj stosowane przy wyborze najlepszego modelu prognostycznego. Kryterium to jest wyrażone wzorem:

$$\text{AIC}(x(t)) = J \ln \left(\frac{\mathbf{e}'\mathbf{e}}{J} \right) + 2(B + 1).$$

Trzy wybrane kryteria różnią się w zasadzie definicją funkcji "kary", która zależy od liczby punktów pomiaru danego procesu i wielkości bazy. Zauważmy, że używając kryterium BIC można wyrazić dwa pozostałe kryteria (eBIC oraz AIC) w języku BIC:

$$\text{eBIC}(x(t)) = \text{BIC}(x(t)) + (B + 1) \ln(B_{max} + 1),$$

$$\text{AIC}(x(t)) = \text{BIC}(x(t)) - (B + 1)(\ln J - 2).$$

Niezależnie od kryterium informacyjnego wykonując n powtórzeń każdego z procesów $X_i, i = 1, 2, \dots, 10$, otrzymamy dla każdego z procesów n wartości $B_i, i = 1, 2, \dots, n$. Wśród wszystkich wartości B_1, B_2, \dots, B_n jedną wartość wspólną B będziemy wybierać na cztery sposoby, a mianowicie jako wartość modalną (D), maksimum (Max), minimum (Min) oraz średnią (\bar{X}) z liczb B_1, B_2, \dots, B_n . Zatem łącznie będziemy wykorzystywać 12 różnych sposobów wyboru parametru B (3 kryteria, dla każdego 4 metody wyboru B).

Symulacje wykonano dla różnej długości przedziałów czasowych $[0, T]$ (pomiaru dokonywane są w J ustalonych punktach czasowych) oraz różnej liczby p opisującej rozmiar wielozmiennego procesu wektorowego. Wybraliśmy $p = 2, 3, 4, 5, 10$, oraz $T = 10, 50, 100$, i dla każdej pary (p, T) wykonamy 1000 symulacji, łącznie (3 kryteria informacyjna \times 4 sposoby wyboru liczby elementów bazy \times 3 rodzaje przedziałów czasowych \times 5 różnych wymiarów danych funkcjonalnych \times 1000 powtórzeń) 180000 wielozmiennych analiz składowych głównych, badając ich wpływ na wyjaśnienie zmienności przez dwie pierwsze funkcjonalne składowe główne, który jest mierzony za pomocą wskaźnika opisanego wzorem (2.6).

Poniżej przedstawiono wektorowe procesy $\mathbf{X}(t)$ użyte do symulacji, w zależności od ich wymiaru p :

1. $p = 2$: $\mathbf{X}(t) = (X_1(t), X_2(t))'$
2. $p = 3$: $\mathbf{X}(t) = (X_1(t), X_2(t), X_3(t))'$
3. $p = 4$: $\mathbf{X}(t) = (X_1(t), X_2(t), X_3(t), X_4(t))'$
4. $p = 5$: $\mathbf{X}(t) = (X_1(t), X_2(t), X_3(t), X_4(t), X_5(t))'$
5. $p = 10$: $\mathbf{X}(t) = (X_1(t), X_2(t), X_3(t), X_4(t), X_5(t), X_6(t), X_7(t), X_8(t), X_9(t), X_{10}(t))'$

W Tab. 2.1 przedstawiono uśrednione wyniki symulacji dla miary danej wzorem (2.6), tj. $\frac{\lambda_1 + \lambda_2}{\lambda_1 + \dots + \lambda_{K+p}} \cdot 100\%$. Dla każdej pary (p, T) ustalono ranking (uwzględniający kryterium i metodę wyboru B) przyznając rangi od 1, dla najmniejszej wartości miary (2.6), do 12 dla wartości największej.

W Tab. 2.2 przedstawiono ranking przy ustalonym p dla każdego T z osobna. Dla $p = 2$ i $p = 4$ najlepsze okazało się kryterium eBIC, zaś dla $p = 3$ kryterium BIC, najczęściej z minimalną liczbą elementów bazy. Dla $p = 5$ najwięcej razy pierwsze miejsce zajęło również kryterium BIC, tym razem ze średnią z liczby elementów bazy, przy czym sumarycznie najwięcej punktów przypada na kryterium AIC z maksymalną liczbą elementów bazy. Dla $p = 10$ najczęściej na pierwszym miejscu znalazły się bazy wybierane przy zastosowaniu kryterium AIC, natomiast tym razem sumarycznie zarówno kryterium BIC z maksymalną liczbą elementów bazy, eBIC z

minimalną liczbą elementów bazy jak i AIC z dominantą z liczby elementów bazy zajęły ex aequo pierwsze miejsca.

W Tab. 2.3 T było ustalone, a rangowanie odbywało się dla każdego p . Najlepsze rezultaty dla $T = 10$ i $T = 100$ uzyskano przy zastosowaniu kryterium BIC oraz eBIC. Natomiast dla $T = 50$ zbliżone w rankingu są kryteria eBIC oraz AIC.

Przyjrzyjmy się dalej tylko trzem pierwszym miejscom ("podium") dla każdego p . W Tab. 2.4 (góra) przedstawiono liczbę wystąpień trzech pierwszych miejsc z podziałem na poszczególne kryteria informacyjne. Dalej zaprezentowano (w kolumnie "Suma") ile razy dane kryterium wystąpiło na "podium" (tj. zajęło pierwsze, drugie lub trzecie miejsce). Aby uwzględnić wagę poszczególnych miejsc, za 1. miejsce przyznaliśmy wagę 3, za drugie wagę 2, a za 3. wagę 1. W ostatnim wierszu ("Ważona suma") przedstawiono ocenę każdego kryterium uwzględniając wagi miejsc. Wówczas dla małych p ($p = 2, 3, 4$) najlepsze okazuje się kryterium BIC oraz eBIC. Dla $p = 5$ oraz $p = 10$ lepsze okazuje się kryterium AIC. W Tab. 2.4 (dół) mamy analogiczne wyniki, tym razem dla ustalonego T . Dla $T = 10$ najlepsze okazuje się eBIC, dla $T = 50$ AIC. Natomiast dla $T = 100$ równie wysokie miejsca zajęły kryteria BIC oraz eBIC.

Tablica 2.1: Uśrednione wyniki wskaźnika (2.6) na podstawie symulacji

Kryterium	BIC				eBIC				AIC			
	D	Max	Min	\bar{X}	D	Max	Min	\bar{X}	D	Max	Min	\bar{X}
p=2												
T=10	82,26	81,80	83,19	84,35	83,79	82,07	88,37	84,16	81,33	85,76	86,23	85,30
T=50	80,73	81,94	77,90	81,42	83,96	82,76	90,87	82,31	80,20	81,35	80,86	83,58
T=100	79,01	80,39	82,95	83,45	84,33	82,39	83,20	82,23	82,98	79,98	78,94	80,78
p=3												
T=10	84,04	85,34	81,04	73,14	82,85	78,30	77,91	78,51	77,89	82,24	66,96	79,37
T=50	75,80	74,68	76,43	72,33	69,55	75,11	72,24	68,82	73,55	72,19	75,34	69,65
T=100	73,27	72,75	76,77	75,80	72,89	74,34	75,80	74,14	73,22	73,52	75,32	70,84
p=4												
T=10	67,36	64,24	65,38	67,35	68,88	67,67	61,13	66,42	61,28	59,75	62,62	59,73
T=50	58,49	59,93	58,23	53,37	55,90	58,68	58,97	55,55	58,53	62,82	59,12	59,31
T=100	55,53	55,85	56,84	58,02	60,19	57,93	56,98	58,02	57,44	55,53	59,42	57,89
p=5												
T=10	64,33	56,57	62,63	68,52	61,81	61,56	64,99	53,60	68,46	62,43	54,41	64,00
T=50	53,49	55,21	57,69	54,97	59,08	57,76	57,73	55,47	54,32	59,11	57,07	60,97
T=100	53,03	57,50	57,36	58,97	57,49	54,01	56,31	57,31	57,74	58,22	57,29	56,24
p=10												
T=10	33,08	34,05	33,17	32,91	28,09	33,29	34,63	35,28	26,80	24,61	36,97	31,43
T=50	20,27	22,03	21,59	23,66	22,31	22,01	23,67	20,40	24,97	21,30	19,63	22,45
T=100	21,10	21,35	20,50	19,67	20,93	19,57	19,97	21,89	22,89	19,55	20,59	19,75

Tablica 2.2: Ranking uśrednionych wyników wskaźnika (2.6) przy ustalonym p , 12 - najlepsze, 1 - najgorsze

Kryterium	BIC				eBIC				AIC			
Metoda	D	Max	Min	\bar{X}	D	Max	Min	\bar{X}	D	Max	Min	\bar{X}
p=2												
T=10	4	2	5	8	6	3	12	7	1	10	11	9
T=50	3	7	1	6	11	9	12	8	2	5	4	10
T=100	2	4	8	11	12	7	10	6	9	3	1	5
Suma	9	13	14	25	29	19	34	21	12	18	16	24
Suma	61				103				70			
p=3												
T=10	11	12	8	2	10	5	4	6	3	9	1	7
T=50	11	8	12	6	2	9	5	1	7	4	10	3
T=100	5	2	12	10	3	8	11	7	4	6	9	1
Suma	27	22	32	18	15	22	20	14	14	19	20	11
Suma	99				71				64			
p=4												
T=10	10	6	7	9	12	11	3	8	4	2	5	1
T=50	5	11	4	1	3	7	8	2	6	12	9	10
T=100	2	3	4	9,5	12	8	5	9,5	6	1	11	7
Suma	17	20	15	19,5	27	26	16	19,5	16	15	25	18
Suma	71,5				88,5				74			
p=5												
T=10	9	3	7	12	5	4	10	1	11	6	2	8
T=50	1	4	7	3	10	9	8	5	2	11	6	12
T=100	1	9	7	12	8	2	4	6	10	11	5	3
Suma	11	16	21	27	23	15	22	12	23	28	13	23
Suma	75				72				87			
p=10												
T=10	6	9	7	5	3	8	10	11	2	1	12	4
T=50	2	7	5	10	8	6	11	3	12	4	1	9
T=100	9	10	6	3	8	2	5	11	12	1	7	4
Suma	17	26	18	18	19	16	26	25	26	6	20	17
Suma	79				86				69			

Tablica 2.3: Ranking uśrednionych wyników wskaźnika (2.6) przy ustalonym T , 12 - najlepsze, 1 - najgorsze

Kryterium	BIC				eBIC				AIC			
Metoda	D	Max	Min	\bar{X}	D	Max	Min	\bar{X}	D	Max	Min	\bar{X}
T=10												
p=2	4	2	5	8	6	3	12	7	1	10	11	9
p=3	11	12	8	2	10	5	4	6	3	9	1	7
p=4	10	6	7	9	12	11	3	8	4	2	5	1
p=5	9	3	7	12	5	4	10	1	11	6	2	8
p=10	6	9	7	5	3	8	10	11	2	1	12	4
Suma	40	32	34	36	36	31	39	33	21	28	31	29
Suma	142				139				109			
T=50												
p=2	3	7	1	6	11	9	12	8	2	5	4	10
p=3	11	8	12	6	2	9	5	1	7	4	10	3
p=4	5	11	4	1	3	7	8	2	6	12	9	10
p=5	1	4	7	3	10	9	8	5	2	11	6	12
p=10	2	7	5	10	8	6	11	3	12	4	1	9
Suma	22	37	29	26	34	40	44	19	29	36	30	44
Suma	114				137				139			
T=100												
p=2	2	4	8	11	12	7	10	6	9	3	1	5
p=3	5	2	12	10	3	8	11	7	4	6	9	1
p=4	2	3	4	9,5	12	8	5	9,5	6	1	11	7
p=5	1	9	7	12	8	2	4	6	10	11	5	3
p=10	9	10	6	3	8	2	5	11	12	1	7	4
Suma	19	28	37	45,5	43	27	35	39,5	41	22	33	20
Suma	129,5				144,5				116			

Tablica 2.4: "Podium" uśrednionych wyników wskaźnika (2.6) przy ustalonym p (górze) oraz T (dół)

Kryterium	p=2			p=3			p=4			p=5			p=10		
	BIC	eBIC	AIC	BIC	eBIC	AIC	BIC	eBIC	AIC	BIC	eBIC	AIC	BIC	eBIC	AIC
1. miejsce	0	3	0	3	0	0	0	2	1	2	0	1	0	0	3
2. miejsce	1	1	1	1	1	1	1	1	1	0	0	3	0	3	0
3. miejsce	0	1	2	1	1	1	2	1	1	0	2	1	2	1	0
Suma	1	5	3	5	2	2	3	4	3	2	2	5	2	4	3
Ważona suma	2	12	4	12	3	3	4	9	6	6	2	10	2	7	9

Kryterium	T=10			T=50			T=100		
	BIC	eBIC	AIC	BIC	eBIC	AIC	BIC	eBIC	AIC
1. miejsce	2	2	1	1	1	3	2	2	1
2. miejsce	1	2	2	2	2	1	1	2	2
3. miejsce	1	3	1	1	3	1	3	2	1
Suma	4	7	4	4	6	5	6	6	4
Ważona suma	9	13	8	8	10	12	11	12	8

Rozdział 3

Zmienne dyskryminacyjne dla wielozmiennych danych funkcjonalnych

3.1 Klasyczne zmienne dyskryminacyjne

W poprzednim rozdziale omówiliśmy redukcję wymiarowości za pomocą składowych głównych. W przypadku, gdy posiadamy dodatkową informację o obserwacjach w postaci ich przynależności do jednej z klas (grup), możemy tę informację wykorzystać do konstrukcji nowych jednowymiarowych zmiennych, tzw. zmiennych dyskryminacyjnych. Metoda zmiennych dyskryminacyjnych w przypadku dwóch klas została wprowadzona przez Fishera (1936), a następnie uogólniona na przypadek $L > 2$ klas. W stosunku do składowych głównych, metoda ta różni się tym, że obserwujemy dwa rodzaje zmienności, zmienność między- i wewnątrzklasową.

Weźmy pod uwagę L klas obiektów. Niech $Y \in \{1, 2, \dots, L\}$ będzie etykietą, a $\mathbf{X} \in \mathbb{R}^p$ wektorem cech, którymi scharakteryzowane są obiekty należące do tych L klas. Będziemy w dalszym ciągu zakładać, że $P(Y = l) = q_l > 0$, gdzie $q_1 + q_2 + \dots + q_L = 1$, $E(\mathbf{X}|Y = l) = \boldsymbol{\mu}_l$ oraz $\text{Var}(\mathbf{X}|Y = l) = \boldsymbol{\Sigma} > 0$, dla $l = 1, \dots, L$. Niech ponadto

$$\begin{aligned}\boldsymbol{\mu} &= \sum_{l=1}^L q_l \boldsymbol{\mu}_l, \\ \boldsymbol{\Delta} &= \sum_{l=1}^L q_l (\boldsymbol{\mu}_l - \boldsymbol{\mu})(\boldsymbol{\mu}_l - \boldsymbol{\mu})', \\ \boldsymbol{\Omega} &= \boldsymbol{\Delta} + \boldsymbol{\Sigma}.\end{aligned}$$

Nieujemnie określona macierz $\boldsymbol{\Delta}$ nosi nazwę macierzy zmienności międzygrupowej (patrz np. Krzyśko (1990), str. 77), dodatnio określona macierz $\boldsymbol{\Sigma}$ jest macierzą zmienności wewnątrzgrupowej, a nieujemnie określona macierz $\boldsymbol{\Omega}$ jest macierzą zmienności całkowitej. Niech $U = \mathbf{u}'\mathbf{X}$, gdzie $\mathbf{u}, \mathbf{X} \in \mathbb{R}^p$ oraz niech

$$J(\mathbf{u}) = \frac{\text{Var}_B(U)}{\text{Var}_T(U)} = \frac{\mathbf{u}'\boldsymbol{\Delta}\mathbf{u}}{\mathbf{u}'\boldsymbol{\Omega}\mathbf{u}}, \quad (3.1)$$

gdzie $\text{Var}_B(U)$ jest wariancją międzygrupową zmiennej U a $\text{Var}_T(U)$ jest wariancją całkowitą zmiennej U , będzie miarą, która wskazuje, w jakim stopniu zmienne X_1, \dots, X_p różnicują (dyskryminują) L klas obiektów. Pierwszą zmienną dyskryminacyjną nazywać będziemy zmienną $U_1 = \mathbf{u}'_1 \mathbf{X}$ spełniającą warunek

$$\lambda_1 = \max_{\mathbf{u} \in \mathbb{R}^p} J(\mathbf{u}) = \frac{\mathbf{u}'_1 \Delta \mathbf{u}_1}{\mathbf{u}'_1 \Omega \mathbf{u}_1},$$

przy dodatkowym warunku ograniczającym

$$\mathbf{u}'_1 \Omega \mathbf{u}_1 = 1,$$

zapewniającym jednoznaczność pierwszej zmiennej dyskryminacyjnej U_1 . W celu znalezienia wektora $\mathbf{u}_1 \in \mathbb{R}^p$ maksymalizującego miarę $J(\mathbf{u})$, stosujemy metodę mnożników Lagrange'a, dochodząc do uogólnionego zagadnienia własnego

$$\Delta \mathbf{u}_1 = \lambda_1 \Omega \mathbf{u}_1$$

lub równoważnie

$$(\Delta - \lambda_1 \Omega) \mathbf{u}_1 = \mathbf{0}. \quad (3.2)$$

Jeżeli macierz Ω jest nieosobliwa, wówczas uogólnione zagadnienie własne możemy sprowadzić do klasycznego zagadnienia własnego postaci

$$(\Omega^{-1} \Delta - \lambda_1 I) \mathbf{u}_1 = \mathbf{0}.$$

W przypadku, gdy macierz Ω jest osobliwa, możemy dokonać regularyzacji tej macierzy lub użyć uogólnionej odwrotności lub też skorzystać z rozkładu tej macierzy według jej wartości osobliwych (Krzyśko i in. (2014)). Zatem pierwszą zmienną dyskryminacyjną wektora \mathbf{X} jest kombinacja liniowa

$$U_1 = u_{11}X_1 + u_{12}X_2 + \dots + u_{1p}X_p = \mathbf{u}'_1 \mathbf{X},$$

gdzie \mathbf{u}_1 jest wektorem własnym równania (3.2) odpowiadającym największej wartości własnej λ_1 . Następnie tworzymy drugą zmienną dyskryminacyjną $U_2 = \mathbf{u}'_2 \mathbf{X}$ maksymalizującą miarę (3.1) oraz spełniającą dodatkowe warunki $\mathbf{u}'_i \Omega \mathbf{u}_i = 1, i = 1, 2$ oraz $\mathbf{u}'_1 \Omega \mathbf{u}_2 = 0$. Kontynuując ten proces możemy utworzyć k -tą zmienną dyskryminacyjną $U_k = \mathbf{u}'_k \mathbf{X}$ maksymalizującą miarę (3.1) przy dodatkowych warunkach

$$\mathbf{u}'_k \Omega \mathbf{u}_l = \begin{cases} 1, & k = l, \\ 0, & k \neq l, \end{cases}$$

$k, l = 1, \dots, s = \text{rzęd} \Delta$.

Uwaga 3.1. Standardowo miara $J(\mathbf{u})$ ma postać

$$J(\mathbf{u}) = \frac{\text{Var}_B(U)}{\text{Var}_T(U)} = \frac{\mathbf{u}'\Delta\mathbf{u}}{\mathbf{u}'\Sigma\mathbf{u}}.$$

Alte

$$\max_{\mathbf{u} \in \mathbb{R}^p} \frac{\mathbf{u}'\Delta\mathbf{u}}{\mathbf{u}'\Omega\mathbf{u}} = \min_{\mathbf{u} \in \mathbb{R}^p} \frac{\mathbf{u}'\Omega\mathbf{u}}{\mathbf{u}'\Delta\mathbf{u}} = \min_{\mathbf{u} \in \mathbb{R}^p} \left(1 + \frac{1}{\frac{\mathbf{u}'\Delta\mathbf{u}}{\mathbf{u}'\Sigma\mathbf{u}}} \right) = \max_{\mathbf{u} \in \mathbb{R}^p} \frac{\mathbf{u}'\Delta\mathbf{u}}{\mathbf{u}'\Sigma\mathbf{u}}.$$

Stąd miary te prowadzą do identycznych rozwiązań, a tym samym identycznych zmiennych dyskryminacyjnych.

Zauważmy, że konstrukcja zmiennych dyskryminacyjnych wymaga znajomości wektorów $\boldsymbol{\mu}_1, \dots, \boldsymbol{\mu}_L$ oraz macierzy Σ, Δ i Ω . W praktyce parametry te nie są nam znane i zachodzi potrzeba posłużenia się ich ocenami z próby. Niech x_{l1}, \dots, x_{ln_l} będzie próbą pochodzącą z grupy l , gdzie $l = 1, \dots, L$ oraz niech $n = n_1 + \dots, n_L$. Wówczas:

$$\hat{q}_l = \frac{n_l}{n}, \quad \hat{\boldsymbol{\mu}}_l = \bar{\mathbf{x}}_l = \frac{1}{n_l} \sum_{j=1}^{n_l} \mathbf{x}_{lj}, \quad l = 1, \dots, L,$$

$$\hat{\boldsymbol{\mu}} = \bar{\mathbf{x}} = \frac{1}{n} \sum_{l=1}^L \sum_{j=1}^{n_l} \mathbf{x}_{lj},$$

$$\hat{\Sigma} = \frac{1}{n} \sum_{l=1}^L \sum_{j=1}^{n_l} (\mathbf{x}_{lj} - \bar{\mathbf{x}}_l)(\mathbf{x}_{lj} - \bar{\mathbf{x}}_l)',$$

$$\hat{\Delta} = \frac{1}{n} \sum_{l=1}^L n_l (\mathbf{x}_l - \bar{\mathbf{x}})(\mathbf{x}_l - \bar{\mathbf{x}})',$$

$$\hat{\Omega} = \hat{\Delta} + \hat{\Sigma} = \frac{1}{n} \sum_{l=1}^L \sum_{j=1}^{n_l} (\mathbf{x}_{lj} - \bar{\mathbf{x}})(\mathbf{x}_{lj} - \bar{\mathbf{x}})'$$

3.2 Zmienne dyskryminacyjne dla wielozmiennych danych funkcjonalnych

Powyżej została omówiona konstrukcja zmiennych dyskryminacyjnych dla wektorów losowych \mathbf{X} . Oczywiście chcielibyśmy nasze rozważania rozszerzyć teraz na dane funkcjonalne. Funkcjonalna analiza dyskryminacyjna została wprowadzona przez Jamesa i Hastiego (2001), i była jedną z pierwszych prac dotyczących problemu klasyfikacji dla danych funkcjonalnych. Taka sama metoda jest zastosowana w monografii Ramsaya i Silvermana (2005). Natomiast Rossi i Villa (2006) zastosowali metodę wektorów nośnych do klasyfikacji takich danych. Częstkowa metoda najmniejszych kwadratów w analizie dyskryminacyjnej dla danych funkcjonalnych została wprowadzona przez Predę i in. (2007). Araki i in. (2009) zastosowali natomiast funkcjonalną dyskryminację logistyczną opartą na technice regularyzacji. Pierwszą pracą, w której podano

konstrukcję zmiennych dyskryminacyjnych dla jednowymiarowych danych funkcjonalnych jest praca Shina (2008). Wariant jądrowy tej metody opisany jest w pracy Góreckiego i innych (2014). Niestety w dotychczasowych publikacjach zmienne dyskryminacyjne były zdefiniowane tylko i wyłącznie dla klasycznych danych funkcjonalnych, tj. wielowymiarowość była rozpatrywana tylko ze względu na czas. Naturalnym uogólnieniem wydaje się jednak uogólnienie na wielozmienne dane funkcjonalne. Podobnie jak w przypadku składowych głównych możemy uogólnić nasze rozważania na wielozmienne dane funkcjonalne $\mathbf{X}(t)$, wykorzystując znaną już reprezentację wielowymiarowego procesu losowego $\mathbf{X}(t) = \Phi(t)\mathbf{c}$, $E(\mathbf{c}) = \mathbf{0}$.

Naszym celem jest skonstruowanie iloczynu skalarnego postaci:

$$U = \langle \mathbf{u}(t), \mathbf{X}(t) \rangle = \int_I \mathbf{u}'(t)\mathbf{X}(t)dt,$$

którego wariancja międzyklasowa jest maksymalna w porównaniu z wariancją całkowitą, gdzie $t \in I$, $\mathbf{u}(t) \in L_2(I^p)$. Przypomnijmy, że funkcja wektorowa postaci $\mathbf{u}(t) = (u_1(t), u_2(t), \dots, u_p(t))'$ nazywana jest wektorową funkcją wagową. Dokładniej, pierwsza funkcjonalna zmienna dyskryminacyjna $U_1 = \langle \mathbf{u}_1(t), \mathbf{X}(t) \rangle$ zdefiniowana jest jako:

$$\lambda_1 = \sup_{\mathbf{u}(t) \in L_2(I^p)} \frac{\text{Var}_B(\langle \mathbf{u}(t), \mathbf{X}(t) \rangle)}{\text{Var}_T(\langle \mathbf{u}(t), \mathbf{X}(t) \rangle)} = \frac{\text{Var}_B(\langle \mathbf{u}_1(t), \mathbf{X}(t) \rangle)}{\text{Var}_T(\langle \mathbf{u}_1(t), \mathbf{X}(t) \rangle)},$$

przy dodatkowym warunku ograniczającym

$$\text{Var}_T(\langle \mathbf{u}_1(t), \mathbf{X}(t) \rangle) = 1,$$

gdzie $\text{Var}_B(\langle \mathbf{u}(t), \mathbf{X}(t) \rangle)$ i $\text{Var}_T(\langle \mathbf{u}(t), \mathbf{X}(t) \rangle)$ są odpowiednio wariancjami zmienności międzyklasowej i całkowitej zmiennej losowej U . Powyższy warunek zapewnia jednoznaczność pierwszej zmiennej dyskryminacyjnej U_1 .

W dalszym ciągu zakładając będziemy, że $\mathbf{X}(t)$ ma postać (1.4). Można także założyć (Ramsay i Silverman (2005)), że wektorowa funkcja wagowa $\mathbf{u}(t)$ i proces losowy $\mathbf{X}(t)$ leżą w tej samej przestrzeni, tzn. funkcja $\mathbf{u}(t)$ może zostać zapisana w następującej postaci:

$$\mathbf{u}(t) = \Phi(t)\mathbf{u},$$

gdzie $\mathbf{u} \in \mathbb{R}^{K+p}$. Wówczas

$$\langle \mathbf{u}(t), \mathbf{X}(t) \rangle = \langle \Phi(t)\mathbf{u}, \Phi(t)\mathbf{c} \rangle = \mathbf{u}' \langle \Phi(t), \Phi(t) \rangle \mathbf{c} = \mathbf{u}' I_{K+p} \mathbf{c} = \mathbf{u}' \mathbf{c},$$

przy czym

$$E(\langle \mathbf{u}(t), \mathbf{X}(t) \rangle) = \mathbf{u}' E(\mathbf{c}) = \mathbf{u}' \mathbf{0} = 0.$$

Stąd zmienność międzyklasowa iloczynu skalarnego $\langle \mathbf{u}(t), \mathbf{X}(t) \rangle$ wynosi:

$$\text{Var}_B(\langle \mathbf{u}(t), \mathbf{X}(t) \rangle) = \mathbf{u}' \text{Var}_B(\mathbf{c})\mathbf{u} = \mathbf{u}' \mathbf{B}\mathbf{u},$$

a zmienność całkowita wyraża się wzorem:

$$\text{Var}_T(\langle \mathbf{u}(t), \mathbf{X}(t) \rangle) = \mathbf{u}' \text{Var}_T(\mathbf{c})\mathbf{u} = \mathbf{u}'\mathbf{T}\mathbf{u},$$

gdzie \mathbf{B} oraz \mathbf{T} są odpowiednio macierzami kwadratów i iloczynów zmienności wewnątrzklasowej i całkowitej wektora losowego \mathbf{c} .

Dla pierwszej funkcjonalnej zmiennej dyskryminacyjnej procesu $\mathbf{X}(t)$ mamy:

$$\lambda_1 = \sup_{\mathbf{u}(t) \in L_2(I^p)} \frac{\text{Var}_B(\langle \mathbf{u}(t), \mathbf{X}(t) \rangle)}{\text{Var}_T(\langle \mathbf{u}(t), \mathbf{X}(t) \rangle)} = \sup_{\mathbf{u} \in \mathbb{R}^{K+p}} \frac{\mathbf{u}' \text{Var}_B(\mathbf{c})\mathbf{u}}{\mathbf{u}' \text{Var}_T(\mathbf{c})\mathbf{u}} = \frac{\mathbf{u}'_1 \mathbf{B}\mathbf{u}_1}{\mathbf{u}'_1 \mathbf{T}\mathbf{u}_1}, \quad (3.3)$$

pod warunkiem, że $\mathbf{u}'_1 \mathbf{T}\mathbf{u}_1 = 1$.

Analogicznie możemy skonstruować k -tą funkcjonalną zmienną dyskryminacyjną

$$U_k = \langle \mathbf{u}_k(t), \mathbf{X}(t) \rangle,$$

gdzie wektorowa funkcja wagowa $\mathbf{u}_k(t)$ jest zdefiniowana w następujący sposób:

$$\lambda_k = \sup_{\mathbf{u}(t) \in L_2(I^p)} \frac{\text{Var}_B(\langle \mathbf{u}(t), \mathbf{X}(t) \rangle)}{\text{Var}_T(\langle \mathbf{u}(t), \mathbf{X}(t) \rangle)} = \sup_{\mathbf{u} \in \mathbb{R}^{K+p}} \frac{\mathbf{u}' \text{Var}_B(\mathbf{c})\mathbf{u}}{\mathbf{u}' \text{Var}_T(\mathbf{c})\mathbf{u}} = \frac{\mathbf{u}'_k \mathbf{B}\mathbf{u}_k}{\mathbf{u}'_k \mathbf{T}\mathbf{u}_k},$$

przy dodatkowym warunku ograniczającym

$$\text{Cov}_T(\mathbf{u}'_k \mathbf{T}\mathbf{u}_k, \mathbf{u}'_l \mathbf{B}\mathbf{u}_l) = \begin{cases} 1, & k = l, \\ 0, & k \neq l, \end{cases}$$

$k, l = 1, \dots, s$, $s = \min(K + p, L - 1)$, $K = B_1 + \dots + B_p$.

Wyrażenie $(\lambda_k, \mathbf{u}_k(t))$ nazywać będziemy k -tym układem dyskryminacyjnym procesu $\mathbf{X}(t)$.

Zdefiniujmy teraz klasyczne zmienne dyskryminacyjne dla wektora losowego \mathbf{c} (patrz Sekcja 3.1). k -ta zmienna dyskryminacyjna $U_k^* = \langle \mathbf{u}_k, \mathbf{c} \rangle$ tego wektora spełnia warunek:

$$\gamma_k = \sup_{\mathbf{u} \in \mathbb{R}^{K+p}} \text{Var}_B(\langle \mathbf{u}, \mathbf{c} \rangle) = \text{Var}_B(\langle \mathbf{u}_k, \mathbf{c} \rangle) = \mathbf{u}'_k \text{Var}_B(\mathbf{c})\mathbf{u}_k = \mathbf{u}'_k \mathbf{B}\mathbf{u}_k,$$

przy dodatkowym warunku ograniczającym

$$\text{Cov}_T(\mathbf{u}'_k \mathbf{T}\mathbf{u}_k, \mathbf{u}'_l \mathbf{B}\mathbf{u}_l) = \begin{cases} 1, & k = l, \\ 0, & k \neq l. \end{cases}$$

Wyrażenie (γ_k, \mathbf{u}_k) nazywać będziemy k -tym układem dyskryminacyjnym wektora losowego \mathbf{c} .

Zauważmy, że wyrażenie (3.3) jest równoważne wyrażeniu

$$\gamma_1 = \sup_{\mathbf{u} \in \mathbb{R}^{K+p}} \text{Var}_B(\langle \mathbf{u}, \mathbf{c} \rangle) = \sup_{\mathbf{u} \in \mathbb{R}^{K+p}} \mathbf{u}' \text{Var}_B(\mathbf{c})\mathbf{u} = \mathbf{u}'_1 \mathbf{B}\mathbf{u}_1,$$

gdzie $\mathbf{u}'_1 \mathbf{T}\mathbf{u}_1 = 1$.

Jest to definicja pierwszej zmiennej dyskryminacyjnej wektora losowego \mathbf{c} . Z drugiej strony, jeżeli pierwszy układ dyskryminacyjny (γ_1, \mathbf{u}_1) definiuje pierwszą zmienną dyskryminacyjną

wektora losowego \mathbf{c} , to otrzymujemy pierwszy układ dyskryminacyjny procesu losowego $\mathbf{X}(t)$ z poniższych równości:

$$\lambda_1 = \gamma_1, \quad \mathbf{u}_1(t) = \Phi(t)\mathbf{u}_1.$$

Możemy rozszerzyć nasze rozważania na drugi układ dyskryminacyjny i kolejne.

Prawdziwe jest zatem następujące twierdzenie:

Twierdzenie 3.1. *k -ty układ dyskryminacyjny $(\lambda_k, \mathbf{u}_k(t))$ procesu losowego $\mathbf{X}(t)$ jest zależny od k -tego układu dyskryminacyjnego (γ_k, \mathbf{u}_k) wektora losowego \mathbf{c} poprzez następujące równości:*

$$\lambda_k = \gamma_k, \quad \mathbf{u}_k(t) = \Phi(t)\mathbf{u}_k, \quad t \in I,$$

gdzie $k = 1, \dots, s$, $s = \min(K + p, L - 1) = \text{rzqd}(\mathbf{B})$, $K = B_1 + B_2 + \dots + B_p$.

Z Twierdzenia 3.1 wynika, że konstrukcja zmiennych dyskryminacyjnych dla procesu losowego $\mathbf{X}(t) = \Phi(t)\mathbf{c}$ sprowadza się w istocie do konstrukcji klasycznych zmiennych dyskryminacyjnych dla wektora losowego \mathbf{c} . Zmienna dyskryminacyjna U_k ma bowiem postać $U_k = \mathbf{u}'_k \mathbf{c}$, gdzie \mathbf{c} jest wektorem losowym w reprezentacji wektorowej procesu losowego $\mathbf{X}(t)$ postaci $\mathbf{X}(t) = \Phi(t)\mathbf{c}$, natomiast \mathbf{u}_k jest wektorem własnym w uogólnionym zagadnieniu własnym $(\mathbf{B} - \lambda_k \mathbf{T})\mathbf{u}_k = \mathbf{0}$, gdzie \mathbf{B} oraz \mathbf{T} są macierzami zmienności międzyklasowej i całkowitej wektora losowego \mathbf{c} , $k = 1, 2, \dots, s$.

W praktyce jednak macierze zmienności $\text{Var}_B(\mathbf{c})$ oraz $\text{Var}_T(\mathbf{c})$ nie są znane i muszą być estymowane na podstawie próby. Niech $\mathbf{x}_{l1}(t), \mathbf{x}_{l2}(t), \dots, \mathbf{x}_{ln_l}(t)$ będzie próbą pochodzącą z l -tej klasy, gdzie $l = 1, 2, \dots, L$. Każda z funkcji $\mathbf{x}_{li}(t)$ ma następującą postać:

$$\mathbf{x}_{li}(t) = \Phi(t)\hat{\mathbf{c}}_{li},$$

gdzie $\hat{\mathbf{c}}_{lj} = (\hat{c}_{10}^{(lj)}, \dots, \hat{c}_{1K_1}^{(lj)}, \dots, \hat{c}_{p0}^{(lj)}, \dots, \hat{c}_{pK_p}^{(lj)})'$, $i = 1, 2, \dots, n_l$, $l = 1, 2, \dots, L$. Niech ponadto

$$\bar{\mathbf{c}}_l = \frac{1}{n_l} \sum_{i=1}^{n_l} \hat{\mathbf{c}}_{li}$$

będzie wektorem średnich w l -tej klasie, $l = 1, \dots, L$. Wówczas

$$\hat{\text{Var}}_B(\mathbf{c}) = \hat{\mathbf{B}} = \frac{1}{L} \sum_{l=1}^L n_l \bar{\mathbf{c}}_l \bar{\mathbf{c}}_l',$$

$$\hat{\text{Var}}_T(\mathbf{c}) = \hat{\mathbf{T}} = \frac{1}{n} \sum_{l=1}^L \sum_{i=1}^{n_l} n_l \hat{\mathbf{c}}_{li} \hat{\mathbf{c}}_{li}'.$$

Następnie znajdujemy niezerowe wartości własne $\hat{\gamma}_1 \geq \hat{\gamma}_2 \geq \dots \geq \hat{\gamma}_s$ i odpowiadające im wektory własne $\hat{\mathbf{u}}_1, \hat{\mathbf{u}}_2, \dots, \hat{\mathbf{u}}_s$ równania $(\hat{\mathbf{B}} - \hat{\gamma} \hat{\mathbf{T}})\hat{\mathbf{u}} = \mathbf{0}$, gdzie $s = \min(K + p, L - 1)$, $K = B_1 + B_2 + \dots + B_p$. Wówczas k -ty układ dyskryminacyjny procesu losowego $\mathbf{X}(t)$ wyznaczony z próby ma następującą postać:

$$(\hat{\lambda}_k = \hat{\gamma}_k, \hat{\mathbf{u}}_k(t) = \Phi(t)\hat{\mathbf{u}}_k), \quad k = 1, \dots, s = \min(K + p, L - 1).$$

Stąd współczynniki rzutu i -tej realizacji $\mathbf{x}_{li}(t)$ procesu losowego $\mathbf{X}(t)$ należącego do l -tej klasy na k -tą funkcjonalną zmienną dyskryminacyjną są równe:

$$\hat{U}_{lik} = \langle \hat{\mathbf{u}}_k(t), \mathbf{x}_{li}(t) \rangle = \hat{\mathbf{u}}_k' \hat{\mathbf{c}}_{li},$$

dla $i = 1, 2, \dots, n_l, k = 1, 2, \dots, s, l = 1, 2, \dots, L$.

Ostatecznie współczynniki rzutu i -tej realizacji $\mathbf{x}_{li}(t)$ procesu losowego $\mathbf{X}(t)$ należącego do l -tej klasy na płaszczyznę dwóch pierwszych funkcjonalnych zmiennych dyskryminacyjnych z próby są równe $(\hat{\mathbf{u}}_1' \hat{\mathbf{c}}_{li}, \hat{\mathbf{u}}_2' \hat{\mathbf{c}}_{li}), i = 1, 2, \dots, n, l = 1, \dots, L$.

Prawdziwe są następujące własności funkcjonalnych zmiennych dyskryminacyjnych:

Własność 3.1. *Wariancje zmiennych*

Niech \mathbf{u}_j będzie wektorem własnym w następującym zagadnieniu własnym $(\mathbf{B} - \lambda_j \mathbf{T})\mathbf{u}_j = \mathbf{0}$ wektora losowego \mathbf{c} w reprezentacji $\mathbf{X}(t) = \Phi(t)\mathbf{c}$ odpowiadającym wartości własnej λ_j unormowanym tak, że $\mathbf{u}_j' \mathbf{T} \mathbf{u}_j = 1$ i niech U_j oznacza j -tą składową główną, $j = 1, 2, \dots, s$. Wówczas

$$\text{Var}_B(U_j) = \lambda_j.$$

Wiemy, że U_j jest postaci $U_j = \mathbf{u}_j' \mathbf{c}$. Stąd

$$\text{Var}_B(U_j) = \text{Var}_B(\mathbf{u}_j' \mathbf{c}) = \mathbf{u}_j' \text{Var}_B(\mathbf{c}) \mathbf{u}_j = \mathbf{u}_j' \mathbf{B} \mathbf{u}_j.$$

Ale \mathbf{u}_j spełnia warunek $\mathbf{B} \mathbf{u}_j = \lambda_j \mathbf{T} \mathbf{u}_j$ lub równoważnie $\mathbf{u}_j' \mathbf{B} \mathbf{u}_j = \lambda_j \mathbf{u}_j' \mathbf{T} \mathbf{u}_j = \lambda_j$, bo z założenia wiemy, że $\mathbf{u}_j' \mathbf{T} \mathbf{u}_j = 1$. Ponieważ $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_s$, to zmienna dyskryminacyjna U_1 ma maksymalną wariancję λ_1 , natomiast zmienna dyskryminacyjna U_s ma minimalną wariancję λ_s .

Własność 3.2. *Ortogonalność i nieskorelowanie zmiennych*

Zmienne dyskryminacyjne U_1, U_2, \dots, U_s nie są wzajemnie ortogonalne, tzn.

$$\langle U_i, U_j \rangle = U_i' U_j = (\mathbf{u}_i' \mathbf{c})' \mathbf{u}_j' \mathbf{c} = \mathbf{c}' \mathbf{u}_i \mathbf{u}_j' \mathbf{c} \neq 0.$$

Zmienne dyskryminacyjne U_1, U_2, \dots, U_s są natomiast wzajemnie nieskorelowane. Współczynnik korelacji między zmiennymi $U_i = \mathbf{u}_i' \mathbf{c}$ oraz $U_j = \mathbf{u}_j' \mathbf{c}$ jest równy:

$$\rho(U_i, U_j) = \frac{\text{Cov}_B(U_i, U_j)}{\sqrt{\text{Var}_B(U_i) \text{Var}_B(U_j)}} = \frac{\mathbf{u}_i' \mathbf{B} \mathbf{u}_j}{\sqrt{\lambda_i \lambda_j}}, \quad i, j = 1, 2, \dots, s, i \neq j.$$

Ponieważ $\mathbf{B} \mathbf{u}_j = \lambda_j \mathbf{T} \mathbf{u}_j$ lub równoważnie $\mathbf{u}_i' \mathbf{B} \mathbf{u}_j = \lambda_j \mathbf{u}_i' \mathbf{T} \mathbf{u}_j$ lub $\mathbf{u}_i' \mathbf{B} \mathbf{u}_j = 0$, bo $\mathbf{u}_i' \mathbf{T} \mathbf{u}_j = 0$, to $\rho(U_i, U_j) = 0$ dla $i \neq j$.

Własność 3.3. *Niezmienniczość zmiennych*

Zmienne dyskryminacyjne są niezmiennicze ze względu na zmianę skali. Wiemy, że k -ta zmienna dyskryminacyjna jest postaci $U_k = \mathbf{u}'_k \mathbf{c}$, gdzie \mathbf{c} jest wektorem losowym w reprezentacji $\mathbf{X}(t) = \Phi(t)\mathbf{c}$, a \mathbf{u}_k jest wektorem własnym w następującym uogólnionym zagadnieniu własnym $(\mathbf{B} - \lambda_k \mathbf{T})\mathbf{u}_k = \mathbf{0}$. Niech $\mathbf{d} = \mathbf{D}\mathbf{c}$, gdzie \mathbf{D} jest macierzą nieosobliwą. Ponadto zachodzi $|\mathbf{B} - \lambda_k \mathbf{T}| = 0$ lub równoważnie $0 = |\mathbf{D}||\mathbf{B} - \lambda_k \mathbf{T}||\mathbf{D}'| = |\mathbf{D}\mathbf{B}\mathbf{D}' - \lambda_k \mathbf{D}\mathbf{T}\mathbf{D}'|$. Oznacza to, że pierwiastki równania $(\mathbf{B} - \lambda_k \mathbf{T})\mathbf{u}_k = \mathbf{0}$ są niezmiennicze ze względu na przekształcenia $\mathbf{d} = \mathbf{D}\mathbf{c}$. Załóżmy dalej, że $(\mathbf{D}\mathbf{B}\mathbf{D}' - \lambda_k \mathbf{D}\mathbf{T}\mathbf{D}')\mathbf{v}_k = \mathbf{0}$ co jest równoważne temu, że $(\mathbf{B} - \lambda_k \mathbf{T})\mathbf{D}'\mathbf{v}_k = \mathbf{0}$. Z drugiej strony pierwotny problem $(\mathbf{B} - \lambda_k \mathbf{T})\mathbf{u}_k = \mathbf{0}$ możemy zapisać równoważnie jako $(\mathbf{B} - \lambda_k \mathbf{T})\mathbf{D}'(\mathbf{D}')^{-1}\mathbf{u}_k = \mathbf{0}$. Aby $U_k = V_k$ musiałby zachodzić warunek $\mathbf{v}_k = (\mathbf{D}')^{-1}\mathbf{u}_k$. Wówczas $V_k = \mathbf{v}'_k \mathbf{d} = \mathbf{u}'_k (\mathbf{D})^{-1} \mathbf{D}\mathbf{c} = \mathbf{u}'_k \mathbf{c} = U_k$.

Własność 3.4. *Odległości*

Odległości w przestrzeni wektorowych procesów losowych $\mathbf{X}(t)$ nie są równe odległościom w przestrzeni wektorowej wszystkich zmiennych dyskryminacyjnych, tj.

$$\langle \mathbf{X}_i(t) - \mathbf{X}_j(t), \mathbf{X}_i(t) - \mathbf{X}_j(t) \rangle \neq (\mathbf{U}_i - \mathbf{U}_j)'(\mathbf{U}_i - \mathbf{U}_j),$$

gdzie $\mathbf{X}_i(t) = \Phi(t)\mathbf{c}_i$, $\mathbf{U}_i = (U_{i1}, U_{i2}, \dots, U_{is})'$, $U_{ik} = \mathbf{u}'_k \mathbf{c}_i$, $i = 1, 2, \dots, n$, $k = 1, 2, \dots, s$. Mamy

$$\begin{aligned} \langle \mathbf{X}_i(t) - \mathbf{X}_j(t), \mathbf{X}_i(t) - \mathbf{X}_j(t) \rangle &= \int_I (\mathbf{X}_i(t) - \mathbf{X}_j(t))' (\mathbf{X}_i(t) - \mathbf{X}_j(t)) dt = \\ &= \int_I (\mathbf{c}_i - \mathbf{c}_j)' \Phi'(t) \Phi(t) (\mathbf{c}_i - \mathbf{c}_j) dt = (\mathbf{c}_i - \mathbf{c}_j)' (\mathbf{c}_i - \mathbf{c}_j) \end{aligned}$$

oraz

$$\begin{aligned} (\mathbf{U}_i - \mathbf{U}_j)'(\mathbf{U}_i - \mathbf{U}_j) &= (\mathbf{c}_i - \mathbf{c}_j)' (\mathbf{u}_1 \mathbf{u}'_1 + \mathbf{u}_2 \mathbf{u}'_2 + \dots + \mathbf{u}_s \mathbf{u}'_s) (\mathbf{c}_i - \mathbf{c}_j) \neq \\ &\neq (\mathbf{c}_i - \mathbf{c}_j)' (\mathbf{c}_i - \mathbf{c}_j), \end{aligned}$$

bo $\mathbf{u}_1 \mathbf{u}'_1 + \mathbf{u}_2 \mathbf{u}'_2 + \dots + \mathbf{u}_s \mathbf{u}'_s \neq \mathbf{I}_s$.

Własność 3.5. *Miara zmienności*

Posługując się zmiennymi dyskryminacyjnymi, badacz musi zdecydować, ile z nich jest istotnych do opisu konkretnych danych. Zmienne nieistotne należy pominąć. Będą to zmienne związane z najmniejszymi wartościami własnymi macierzy $\mathbf{T}^{-1}\mathbf{B}$ wektora losowego \mathbf{c} w reprezentacji $\mathbf{X}(t) = \Phi(t)\mathbf{c}$. Możemy posłużyć się następującymi kryteriami:

1. Procent wariancji

Wiemy, że k -ta zmienna dyskryminacyjna jest postaci $U_k = \mathbf{u}'_k \mathbf{c}$, gdzie \mathbf{c} jest wektorem losowym w reprezentacji $\mathbf{X}(t) = \Phi(t)\mathbf{c}$, a \mathbf{u}_k jest k -tym wektorem własnym macierzy $\mathbf{T}^{-1}\mathbf{B}$ wektora losowego \mathbf{c} . Z własności (3.1) wynika, że $\text{Var}_B(U_k) = \lambda_k$, $k = 1, 2, \dots, s$. Stąd

$$\sum_{k=1}^s \text{Var}_B(U_k) = \sum_{k=1}^s \lambda_k = \text{tr}(\mathbf{T}^{-1}\mathbf{B}).$$

W analizie zmiennych dyskryminacyjnych oczekujemy, że dla pewnego małego j suma $\lambda_1 + \lambda_2 + \dots + \lambda_j$ będzie bliska wyrażeniu $\text{tr}(\mathbf{T}^{-1}\mathbf{B}) = \lambda_1 + \lambda_2 + \dots + \lambda_s$. Jeśli tak jest, to j pierwszych zmiennych dyskryminacyjnych wyjaśnia dobrze zmienność procesu $\mathbf{X}(t)$, a pozostałe zmienne wnoszą niewiele, ponieważ ich wariancje są małe. Wskaźnik

$$\frac{\lambda_1 + \dots + \lambda_j}{\lambda_1 + \dots + \lambda_s} \cdot 100\% \quad (3.4)$$

jest procentową miarą wyjaśnienia zmienności wektorowego procesu $\mathbf{X}(t)$ przez pierwszych j zmiennych dyskryminacyjnych. Jeżeli wskaźnik (3.4) jest równy lub większy od ustalonej przez nas wartości, np. 80%, to pozostałe zmienne dyskryminacyjne pomijamy. Dwa dalsze kryteria są identyczne z wymienionymi we Własności 2.5.

Na koniec zauważmy, że wkład procesu $X_j(t)$ w budowę zmiennej dyskryminacyjnej U_k można zdefiniować tak, jak to uczyniono w przypadku składowych głównych.

Rozdział 4

Zmienne kanoniczne dla wielozmiennych danych funkcjonalnych

4.1 Klasyczne zmienne kanoniczne

Poprzednie dwa rozdziały dotyczyły redukcji wymiaru pewnego wektora losowego \mathbf{X} . Często interesuje nas jednak badanie zależności pomiędzy dwoma zbiorami cech. Klasycznym narzędziem do badania zależności pomiędzy pewną zmienną objaśnianą Y a zespołem q zmiennych objaśniających X_1, X_2, \dots, X_q jest regresja wielokrotna. Uogólnieniem tego przypadku jest taka sytuacja, gdy interesuje nas zależność pomiędzy zespołem p cech Y_1, Y_2, \dots, Y_p a zespołem q cech X_1, X_2, \dots, X_q . De facto jest to zależność pomiędzy dwoma wektorami losowymi $\mathbf{Y} = (Y_1, Y_2, \dots, Y_p)' \in \mathbb{R}^p$ oraz $\mathbf{X} = (X_1, X_2, \dots, X_q)' \in \mathbb{R}^q$. Metoda ta została wprowadzona przez Hotellinga (1936) i funkcjonuje w literaturze pod nazwą analizy korelacji kanonicznych. Podobnie jak poprzednio redukcję wymiaru otrzymuje się dzięki konstrukcji nowych zmiennych jednowymiarowych, tzw. zmiennych kanonicznych. W tym przypadku obserwujemy zmienność poszczególnych wektorów oraz interakcję pomiędzy nimi.

Niech $\mathbf{Y} = (Y_1, Y_2, \dots, Y_p)' \in \mathbb{R}^p$ będzie p -wymiarowym, a $\mathbf{X} = (X_1, X_2, \dots, X_q)' \in \mathbb{R}^q$ q -wymiarowym wektorem losowym. Bez utraty ogólności możemy założyć, że $p \leq q$. Niech ponadto

$$\text{Var}(\mathbf{Y}) = \boldsymbol{\Sigma}_{11},$$

$$\text{Var}(\mathbf{X}) = \boldsymbol{\Sigma}_{22},$$

$$\text{Cov}(\mathbf{Y}, \mathbf{X}) = \boldsymbol{\Sigma}_{12} = \boldsymbol{\Sigma}'_{21}.$$

Na początek rozpatrzmy kombinacje liniowe $U = u_1 Y_1 + u_2 Y_2 + \dots + u_p Y_p = \mathbf{u}'\mathbf{Y}$ składowych wektora \mathbf{Y} oraz kombinacje liniowe $V = v_1 X_1 + v_2 X_2 + \dots + v_q X_q = \mathbf{v}'\mathbf{X}$ składowych wektora \mathbf{X} . Wektory \mathbf{u} i \mathbf{v} chcemy dobrać tak, aby współczynnik korelacji pomiędzy nowymi zmiennymi

U i V był maksymalny, tzn.

$$\rho = \rho(U, V) = \max_{\mathbf{u} \in \mathbb{R}^p, \mathbf{v} \in \mathbb{R}^q} \frac{\text{Cov}(U, V)}{\sqrt{\text{Var}(U) \text{Var} V}} = \max_{\mathbf{u} \in \mathbb{R}^p, \mathbf{v} \in \mathbb{R}^q} \frac{\mathbf{u}' \boldsymbol{\Sigma}_{12} \mathbf{v}}{\sqrt{\mathbf{u}' \boldsymbol{\Sigma}_{11} \mathbf{u} \mathbf{v}' \boldsymbol{\Sigma}_{22} \mathbf{v}}}. \quad (4.1)$$

Ponieważ współczynnik korelacji pomiędzy zmiennymi U i V jest równy współczynnikowi korelacji pomiędzy zmiennymi $\varsigma_1 U$ i $\varsigma_2 V$, gdzie ς_1 i ς_2 są stałymi, to wektory \mathbf{u} i \mathbf{v} możemy unormować tak, aby $\text{Var}(U) = 1 = \text{Var} V$. Wówczas zagadnienie (4.1) sprowadza się do maksymalizacji wyrażenia:

$$\rho = \rho(U, V) = \max_{\mathbf{u} \in \mathbb{R}^p, \mathbf{v} \in \mathbb{R}^q} \mathbf{u}' \boldsymbol{\Sigma}_{12} \mathbf{v},$$

przy dodatkowych warunkach ograniczających

$$\mathbf{u}' \boldsymbol{\Sigma}_{11} \mathbf{u} = 1 = \mathbf{v}' \boldsymbol{\Sigma}_{22} \mathbf{v}. \quad (4.2)$$

Niech $\mathbf{C} = \boldsymbol{\Sigma}_{11}^{-1} \boldsymbol{\Sigma}_{12}$ oraz $\mathbf{D} = \boldsymbol{\Sigma}_{22}^{-1} \boldsymbol{\Sigma}_{21}$. Stosując metodę mnożników Lagrange'a i rozwiązując równania macierzowe

$$(\mathbf{C}\mathbf{D} - \rho_1^2 \mathbf{I}_p) \mathbf{u}_1 = \mathbf{0},$$

$$(\mathbf{D}\mathbf{C} - \rho_1^2 \mathbf{I}_q) \mathbf{v}_1 = \mathbf{0},$$

znajdujemy wektory \mathbf{u}_1 oraz \mathbf{v}_1 , które maksymalizują wyrażenie (4.1), tzn.

$$\rho_1 = \max_{\mathbf{u} \in \mathbb{R}^p, \mathbf{v} \in \mathbb{R}^q} \text{Cov}(U, V) = \text{Cov}(U_1, V_1) = \mathbf{u}_1' \boldsymbol{\Sigma}_{12} \mathbf{v}_1, \quad (4.3)$$

przy dodatkowych warunkach ograniczających (4.2). Ponieważ interesują nas jedynie niezerowe rozwiązania powyższego zagadnienia, a macierze $\mathbf{C}\mathbf{D}$ oraz $\mathbf{D}\mathbf{C}$ mają te same niezerowe wartości własne ρ^2 , to możemy nasze zagadnienie zredukować do poniższego układu równań:

$$(\boldsymbol{\Sigma} - \rho_1^2 \mathbf{I}_p) \mathbf{u}_1 = \mathbf{0},$$

przy dodatkowych warunkach ograniczających (4.2), przy czym $\boldsymbol{\Sigma} = \mathbf{C}\mathbf{D}$. Wektor \mathbf{v}_1 otrzymujemy natomiast z równości $\mathbf{v}_1 = \rho_1^{-1} \mathbf{D}\mathbf{u}_1$, przy dodatkowym warunku ograniczającym (4.2).

Dodatni pierwiastek z ρ_1^2 nazywamy pierwszym współczynnikiem korelacji kanonicznej, a unormowane kombinacje liniowe $U_1 = u_{11}Y_1 + u_{12}Y_2 + \dots + u_{1p}Y_p = \mathbf{u}_1' \mathbf{Y}$ oraz $V_1 = v_{11}X_1 + v_{12}X_2 + \dots + v_{1q}X_q = \mathbf{v}_1' \mathbf{X}$ nazywamy zmiennymi kanonicznymi odpowiednio przestrzeni \mathbf{Y} oraz \mathbf{X} .

Analogicznie można skonstruować k -tą parę kombinacji liniowych $U_k = \mathbf{u}_k' \mathbf{Y}$ oraz $V_k = \mathbf{v}_k' \mathbf{X}$, $k = 1, 2, \dots, p$. Wartość ρ_k oraz wektory \mathbf{u}_k i \mathbf{v}_k , $k = 1, 2, \dots, p$, znajdujemy w opisany powyżej sposób. Kombinacje liniowe $U_k = u_{k1}Y_1 + u_{k2}Y_2 + \dots + u_{kp}Y_p = \mathbf{u}_k' \mathbf{Y}$ oraz $V_k = v_{k1}X_1 + v_{k2}X_2 + \dots + v_{kq}X_q = \mathbf{v}_k' \mathbf{X}$, spełniają warunki $\text{Var}(U_k) = 1 = \text{Var}(V_k)$ oraz dodatkowo k -ta para zmiennych kanonicznych (U_k, V_k) nie jest skorelowana z $k - 1$ pierwszymi parami zmiennych

kanonicznych. Parę (U_k, V_k) nazywa się k -tą parą zmiennych kanonicznych odpowiadających k -temu współczynnikowi korelacji kanonicznej ρ_k .

W praktyce macierze Σ_{11} , Σ_{22} oraz Σ_{12} nie są znane i estymujemy je na podstawie n niezależnych realizacji $\mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_n$ wektora losowego \mathbf{Y} oraz $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n$ wektora losowego \mathbf{X} . Estymatory tych macierzy są odpowiednio postaci:

$$\hat{\Sigma}_{11} = \frac{1}{n} \sum_{i=1}^n (\mathbf{y}_i - \bar{\mathbf{y}})(\mathbf{y}_i - \bar{\mathbf{y}})',$$

$$\hat{\Sigma}_{22} = \frac{1}{n} \sum_{i=1}^n (\mathbf{x}_i - \bar{\mathbf{x}})(\mathbf{x}_i - \bar{\mathbf{x}})',$$

$$\hat{\Sigma}_{12} = \frac{1}{n} \sum_{i=1}^n (\mathbf{y}_i - \bar{\mathbf{y}})(\mathbf{x}_i - \bar{\mathbf{x}})'$$

W przypadku, gdy obserwacje te są scentrowane (tzn. $\bar{\mathbf{y}} = \mathbf{0}$ oraz $\bar{\mathbf{x}} = \mathbf{0}$), estymatory te możemy zapisać w postaci:

$$\hat{\Sigma}_{11} = \frac{1}{n} \sum_{i=1}^n \mathbf{y}_i \mathbf{y}_i' = \frac{1}{n} (\mathbf{y}_1, \dots, \mathbf{y}_n) (\mathbf{y}_1, \dots, \mathbf{y}_n)',$$

$$\hat{\Sigma}_{22} = \frac{1}{n} \sum_{i=1}^n \mathbf{x}_i \mathbf{x}_i' = \frac{1}{n} (\mathbf{x}_1, \dots, \mathbf{x}_n) (\mathbf{x}_1, \dots, \mathbf{x}_n)',$$

$$\hat{\Sigma}_{12} = \frac{1}{n} \sum_{i=1}^n \mathbf{x}_i \mathbf{y}_i' = \frac{1}{n} (\mathbf{x}_1, \dots, \mathbf{x}_n) (\mathbf{y}_1, \dots, \mathbf{y}_n)'$$

Rozwiązując równanie $|\hat{\Sigma} - \hat{\rho}^2 \mathbf{I}_p| = 0$ otrzymamy oceny p wartości własnych $\hat{\rho}_1^2, \dots, \hat{\rho}_p^2$. Odpowiadające im wektory własne $\hat{\mathbf{u}}_1, \dots, \hat{\mathbf{u}}_p$ otrzymujemy z układu równań $(\hat{\Sigma} - \hat{\rho}_k^2 \mathbf{I}_p) \hat{\mathbf{u}}_k = \mathbf{0}$, $k = 1, \dots, p$, przy czym $\hat{\Sigma} = \hat{\Sigma}_{11}^{-1} \hat{\Sigma}_{12} \hat{\Sigma}_{22}^{-1} \hat{\Sigma}_{21}$. Wektory $\hat{\mathbf{v}}_k$ otrzymujemy natomiast z równości $\hat{\mathbf{v}}_k = \rho_k^{-1} \hat{\mathbf{D}} \hat{\mathbf{u}}_k$. Zmienne kanoniczne \hat{V}_{ik} oraz \hat{U}_{ik} otrzymane z próby są wówczas postaci $\hat{V}_{ik} = \hat{\mathbf{v}}_k' \mathbf{y}_i$ oraz $\hat{U}_{ik} = \hat{\mathbf{u}}_k' \mathbf{x}_i$, $i = 1, 2, \dots, n$, $k = 1, 2, \dots, p$.

4.2 Zmienne kanoniczne dla wielozmiennych danych funkcjonalnych

Za pierwszą pracę dotyczącą tematyki funkcjonalnej analizy kanonicznej można uznać pracę Hannana (1961), w której rozpatrywana jest funkcjonalna analiza kanoniczna dla stacjonarnych procesów gaussowskich. Natomiast Dauxois i Pousse (1976) opisują analizę kanoniczną w uogólnionych przestrzeniach Hilberta, a Brillinger (1975) charakteryzuje analizę kanoniczną dla wielowymiarowych stacjonarnych szeregów czasowych poprzez redukcję do klasycznej analizy kanonicznej dzięki wykorzystaniu stacjonarności. Z kolei Leurgans i in. (1993) wykorzystali wygładzanie danych funkcjonalnych w analizie kanonicznej używając funkcji sklepanych. Liczne zastosowania w biologii i medycynie zostały wskazane w pracy Leurgansa i in. (1993), a także

w ekologii i ochronie środowiska w pracy Service, Rice i Chavez (1998). Bogata literatura z tej tematyki została przytoczona w pracy He i in. (2002). Warto jeszcze wspomnieć o pracach He i inni (2000, 2004) oraz Eubank i Hsing (2008). Jądrowa wersja zmiennych kanonicznych dla klasycznych danych funkcjonalnych z wykorzystaniem rozwinięcia w bazach ortogonalnych została opisana w pracy Krzyżki i Waszaka (2013). Podobnie jak poprzednio możemy uogólnić nasze rozważania na obiekty, które są realizacjami wielozmiennych danych funkcjonalnych $\mathbf{y}(t)$ oraz $\mathbf{x}(t)$. Wykorzystując reprezentację danych funkcjonalnych opisaną na początku, możemy założyć, że składowe $Y_g(t)$ procesu losowego $\mathbf{Y}(t)$ oraz składowe $X_h(t)$ procesu $\mathbf{X}(t)$ mogą zostać przedstawione odpowiednio za pomocą skończonej liczby ortonormalnych funkcji bazowych $\{\varphi_e\}$ oraz $\{\varphi_f\}$:

$$Y_g(t) = \sum_{e=0}^{E_g} \alpha_{ge} \varphi_e(t), \quad g = 1, 2, \dots, p, \quad t \in I_1,$$

$$X_h(t) = \sum_{f=0}^{F_h} \beta_{hf} \varphi_f(t), \quad h = 1, 2, \dots, q, \quad t \in I_2.$$

Wprowadźmy dodatkowo następującą notację:

$$\boldsymbol{\alpha} = (\alpha_{10}, \dots, \alpha_{1E_1}, \dots, \alpha_{p0}, \dots, \alpha_{pE_p})',$$

$$\boldsymbol{\beta} = (\beta_{10}, \dots, \beta_{1F_1}, \dots, \beta_{q0}, \dots, \beta_{qF_q})',$$

$$\Phi_1(t) = \begin{bmatrix} \boldsymbol{\varphi}'_{E_1}(t) & \mathbf{0}' & \dots & \mathbf{0}' \\ \mathbf{0}' & \boldsymbol{\varphi}'_{E_2}(t) & \dots & \mathbf{0}' \\ \dots & \dots & \dots & \dots \\ \mathbf{0}' & \mathbf{0}' & \dots & \boldsymbol{\varphi}'_{E_p}(t) \end{bmatrix},$$

$$\Phi_2(t) = \begin{bmatrix} \boldsymbol{\varphi}'_{F_1}(t) & \mathbf{0}' & \dots & \mathbf{0}' \\ \mathbf{0}' & \boldsymbol{\varphi}'_{F_2}(t) & \dots & \mathbf{0}' \\ \dots & \dots & \dots & \dots \\ \mathbf{0}' & \mathbf{0}' & \dots & \boldsymbol{\varphi}'_{F_q}(t) \end{bmatrix},$$

gdzie $\boldsymbol{\varphi}_{E_1}, \dots, \boldsymbol{\varphi}_{E_p}$ oraz $\boldsymbol{\varphi}_{F_1}, \dots, \boldsymbol{\varphi}_{F_q}$ są wektorami, których składowymi są ortonormalne funkcje bazowe odpowiednio przestrzeni $L_2(I_1)$ oraz $L_2(I_2)$, a $K_1 = E_1 + E_2 + \dots + E_p$, $K_2 = F_1 + F_2 + \dots + F_q$. Używając powyższej notacji procesy $\mathbf{Y}(t)$ oraz $\mathbf{X}(t)$ można zapisać w następującej postaci:

$$\mathbf{Y}(t) = \Phi_1(t)\boldsymbol{\alpha}, \quad E(\boldsymbol{\alpha}) = \mathbf{0},$$

$$\mathbf{X}(t) = \Phi_2(t)\boldsymbol{\beta}, \quad E(\boldsymbol{\beta}) = \mathbf{0}.$$

Funkcjonalne zmienne kanoniczne U oraz V dla procesów losowych $\mathbf{Y}(t)$ oraz $\mathbf{X}(t)$ można zdefiniować w następujący sposób:

$$U = \langle \mathbf{u}(t), \mathbf{Y}(t) \rangle = \int_{I_1} \mathbf{u}'(t)\mathbf{Y}(t)dt,$$

$$V = \langle \mathbf{v}(t), \mathbf{X}(t) \rangle = \int_{I_2} \mathbf{v}'(t) \mathbf{X}(t) dt,$$

gdzie funkcje wektorowe $\mathbf{u}(t)$ oraz $\mathbf{v}(t)$ nazywane są wektorowymi funkcjami wagowymi. Funkcje wagowe $\mathbf{u}(t)$ oraz $\mathbf{v}(t)$ są dobierane w ten sposób, aby maksymalizować współczynnik

$$\rho = \frac{\text{Cov}(U, V)}{\sqrt{\text{Var}(U) \text{Var}(V)}},$$

przy dodatkowych warunkach ograniczających

$$\text{Var}(U) = \text{Var}(V) = 1. \quad (4.4)$$

Współczynnik ρ nazywany jest współczynnikiem korelacji kanonicznej. Jednakże maksymalizacja tego współczynnika nie daje zadawalających wyników, bowiem w przypadku funkcjonalnym możemy dowolnie wybrać funkcję wektorową $\mathbf{u}(t)$, skonstruować zmienną kanoniczną $U = \langle \mathbf{u}(t), \mathbf{Y}(t) \rangle$, a następnie znaleźć funkcję wektorową $\mathbf{v}(t)$ taką, aby współczynnik korelacji $\rho = \frac{\text{Cov}(U, V)}{\sqrt{\text{Var}(U) \text{Var}(V)}}$ był równy 1, gdzie $V = \langle \mathbf{v}(t), \mathbf{X}(t) \rangle$. Funkcje wagowe $\mathbf{u}(t)$ oraz $\mathbf{v}(t)$ nie dają zatem użytecznej informacji o sile powiązań między zbiorami cech, co wyraźnie wskazuje na potrzebę zastosowania technik wygładzania. Prosty sposób wygładzenia jest modyfikacja warunków ograniczających (4.4) poprzez dodanie pewnego współczynnika kary:

$$\text{Var}(U^{(N)}) = \text{Var}\left(\int_{I_1} \mathbf{u}'(t) \mathbf{Y}(t) dt\right) + \lambda \text{PEN}_2(\mathbf{u}(t)) = 1, \quad (4.5)$$

$$\text{Var}(V^{(N)}) = \text{Var}\left(\int_{I_2} \mathbf{v}'(t) \mathbf{X}(t) dt\right) + \lambda \text{PEN}_2(\mathbf{v}(t)) = 1, \quad (4.6)$$

gdzie współczynnik kary PEN_2 jest scałkowanym kwadratem drugich pochodnych, tj.:

$$\begin{aligned} \text{PEN}_2(\mathbf{u}(t)) &= \int_{I_1} \left(\frac{\partial^2 \mathbf{u}(t)}{\partial t^2}\right)' \frac{\partial^2 \mathbf{u}(t)}{\partial t^2} dt = \int_{I_1} \left(\frac{\partial^2 \Phi_1(t) \mathbf{u}}{\partial t^2}\right)' \frac{\partial^2 \Phi_1(t) \mathbf{u}}{\partial t^2} dt \\ &= \mathbf{u}' \int_{I_1} \left(\frac{\partial^2 \Phi_1(t)}{\partial t^2}\right)' \frac{\partial^2 \Phi_1(t)}{\partial t^2} dt \mathbf{u}, = \mathbf{u}' \mathbf{R}_1 \mathbf{u}, \\ \mathbf{R}_1 &= \int_{I_1} \left(\frac{\partial^2 \Phi_1(t)}{\partial t^2}\right)' \frac{\partial^2 \Phi_1(t)}{\partial t^2} dt, \end{aligned} \quad (4.7)$$

oraz

$$\begin{aligned} \text{PEN}_2(\mathbf{v}(t)) &= \int_{I_2} \left(\frac{\partial^2 \mathbf{v}(t)}{\partial t^2}\right)' \frac{\partial^2 \mathbf{v}(t)}{\partial t^2} dt = \int_{I_2} \left(\frac{\partial^2 \Phi_2(t) \mathbf{v}}{\partial t^2}\right)' \frac{\partial^2 \Phi_2(t) \mathbf{v}}{\partial t^2} dt \\ &= \mathbf{v}' \int_{I_2} \left(\frac{\partial^2 \Phi_2(t)}{\partial t^2}\right)' \frac{\partial^2 \Phi_2(t)}{\partial t^2} dt \mathbf{v}, = \mathbf{v}' \mathbf{R}_2 \mathbf{v}, \\ \mathbf{R}_2 &= \int_{I_2} \left(\frac{\partial^2 \Phi_2(t)}{\partial t^2}\right)' \frac{\partial^2 \Phi_2(t)}{\partial t^2} dt. \end{aligned} \quad (4.8)$$

Współczynnik kary PEN_2 jest uogólnieniem na przypadek wielozmienny współczynnika wprowadzanego przez Ramsaya i Silvermana (2005) na str. 84. Napisali oni: „Współczynnik

(PEN₂) jest popularnym sposobem na oszacowanie gładkości funkcji. Kwadrat drugiej pochodnej $[D^2x(t)]^2$ funkcji w momencie t jest często nazywany jej krzywizną w punkcie t , przy czym linia prosta, z czym się wszyscy zgadzamy, nie ma krzywizny i ma drugą pochodną równą zero. Zatem naturalną miarą krzywizny funkcji jest scałkowany kwadrat drugiej pochodnej

$$\text{PEN}_2 = \int [D^2x(s)]^2 ds.$$

Można oczekiwać, że wysokowymiarowe funkcje uzyskują wysokie wartości współczynnika PEN₂, ponieważ ich drugie pochodne są większe przynajmniej na interesującym nas przedziale.” Dalej na str. 205 napisali: „Efekt wprowadzenia współczynnika kary z ograniczeniem dla kandydatów na zmienne kanoniczne jest taki, że bierzemy pod uwagę nie tylko ich wariancje, ale także krzywiznę i porównujemy ważoną sumę tych dwóch wielkości w terminach kowariancji”. I dalej: „Nasza metoda wprowadzenia wygładzania lub regularyzacji jest podobna do techniki regresji grzbietowej, która jest często wykorzystywana w przetwarzaniu obrazów i źle postawionych problemów do poprawiania uwarunkowań rozważanych macierzy kowariancji.”

Teraz pierwszy współczynnik korelacji kanonicznej ρ_1 i odpowiadające mu wektorowe funkcje wagowe $\mathbf{u}_1(t)$ oraz $\mathbf{v}_1(t)$ można zdefiniować w następujący sposób:

$$\rho_1 = \sup_{\mathbf{u}(t) \in L_2(I_1^p), \mathbf{v}(t) \in L_2(I_2^q)} \frac{\text{Cov}(\langle \mathbf{u}(t), \mathbf{Y}(t) \rangle, \langle \mathbf{v}(t), \mathbf{X}(t) \rangle)}{\sqrt{\text{Var}(U^{(N)}) \text{Var}(V^{(N)})}}, \quad (4.9)$$

przy dodatkowych warunkach ograniczających

$$\text{Var}(U^{(N)}) = \text{Var}(V^{(N)}) = 1. \quad (4.10)$$

W ogólności k -ty współczynnik korelacji kanonicznej ρ_k i odpowiadające mu wektorowe funkcje wagowe $\mathbf{u}_k(t)$ oraz $\mathbf{v}_k(t)$ są zdefiniowane w następujący sposób:

$$\rho_k = \sup_{\mathbf{u}(t) \in L_2(I_1^p), \mathbf{v}(t) \in L_2(I_2^q)} \text{Cov}(\langle \mathbf{u}(t), \mathbf{Y}(t) \rangle, \langle \mathbf{v}(t), \mathbf{X}(t) \rangle) =$$

$$\text{Cov}(\langle \mathbf{u}_k(t), \mathbf{Y}(t) \rangle, \langle \mathbf{v}_k(t), \mathbf{X}(t) \rangle),$$

gdzie $\mathbf{u}_k(t)$ oraz $\mathbf{v}_k(t)$ spełniają warunki (4.5) i (4.6), oraz k -ta para zmiennych kanonicznych (U_k, V_k) nie jest skorelowana z $k - 1$ zmiennymi kanonicznymi, gdzie

$$U_k = \langle \mathbf{u}_k(t), \mathbf{Y}(t) \rangle,$$

$$V_k = \langle \mathbf{v}_k(t), \mathbf{X}(t) \rangle,$$

są funkcjonalnymi zmiennymi kanonicznymi. Taką procedurę nazywa się wygładzoną analizą korelacji kanonicznych.

Wyrażenie $(\rho_k, \mathbf{u}_k(t), \mathbf{v}_k(t))$ nazywać będziemy k -tym układem kanonicznym pary procesów losowych $\mathbf{Y}(t)$ oraz $\mathbf{X}(t)$.

Niech

$$\text{Var}(\boldsymbol{\alpha}) = \text{E}(\boldsymbol{\alpha}\boldsymbol{\alpha}') = \boldsymbol{\Sigma}_{11},$$

$$\text{Var}(\boldsymbol{\beta}) = \text{E}(\boldsymbol{\beta}\boldsymbol{\beta}') = \boldsymbol{\Sigma}_{22},$$

$$\text{Cov}(\boldsymbol{\alpha}, \boldsymbol{\beta}) = \text{E}(\boldsymbol{\alpha}\boldsymbol{\beta}') = \boldsymbol{\Sigma}_{12}.$$

Zdefiniujmy teraz zmienne kanoniczne $U^* = \langle \mathbf{u}, \boldsymbol{\alpha} \rangle$ oraz $V^* = \langle \mathbf{v}, \boldsymbol{\beta} \rangle$ odpowiednio dla wektorów losowych $\boldsymbol{\alpha}$ oraz $\boldsymbol{\beta}$. k -ta korelacja kanoniczna γ_k i związane z nią wektory wagowe \mathbf{u}_k oraz \mathbf{v}_k są zdefiniowane następująco:

$$\gamma_k = \sup_{\mathbf{u} \in \mathbb{R}^{E+p}, \mathbf{v} \in \mathbb{R}^{F+q}} \text{Cov}(\langle \mathbf{u}, \boldsymbol{\alpha} \rangle, \langle \mathbf{v}, \boldsymbol{\beta} \rangle) = \mathbf{u}'_k \boldsymbol{\Sigma}_{12} \mathbf{v}_k,$$

przy dodatkowych warunkach ograniczających

$$\mathbf{u}'_k (\boldsymbol{\Sigma}_{11} + \lambda \mathbf{R}_1) \mathbf{u}_k = 1,$$

$$\mathbf{v}'_k (\boldsymbol{\Sigma}_{22} + \lambda \mathbf{R}_2) \mathbf{v}_k = 1,$$

gdzie macierze \mathbf{R}_1 i \mathbf{R}_2 są zdefiniowane odpowiednio za pomocą wyrażeń (4.7) oraz (4.8), a k -ta para zmiennych kanonicznych (U_k^*, V_k^*) nie jest skorelowana z pierwszymi $k - 1$ parami zmiennych kanonicznych. Wyrażenie $(\gamma_k, \mathbf{u}_k, \mathbf{v}_k)$ nazywać będziemy k -tym układem kanonicznym wektorów losowych $\boldsymbol{\alpha}$ oraz $\boldsymbol{\beta}$.

Prawdziwe jest następujące twierdzenie:

Twierdzenie 4.1. k -ty układ kanoniczny $(\rho_k, \mathbf{u}_k(t), \mathbf{v}_k(t))$ pary procesów losowych $\mathbf{Y}(t)$ oraz $\mathbf{X}(t)$ jest zależny od k -tego układu kanonicznego $(\gamma_k, \mathbf{u}_k, \mathbf{v}_k)$ pary wektorów losowych $\boldsymbol{\alpha}$ oraz $\boldsymbol{\beta}$ poprzez następujące równości:

$$\rho_k = \gamma_k, \quad \mathbf{u}_k(t) = \boldsymbol{\Phi}_1(t) \mathbf{u}_k, t \in I_1, \quad \mathbf{v}_k(t) = \boldsymbol{\Phi}_2(t) \mathbf{v}_k, t \in I_2,$$

gdzie $k = 1, \dots, \min(K_1 + p, K_2 + q)$, $K_1 = E_1 + \dots + E_p$, $K_2 = F_1 + \dots + F_q$.

Dowód. Funkcja kowariancji procesu losowego $\mathbf{Y}(t)$ jest równa:

$$r_{\mathbf{Y}\mathbf{Y}}(s, t) = \text{E}[\mathbf{Y}(s)\mathbf{Y}(t)] = \boldsymbol{\Phi}_1(s) \text{E}(\boldsymbol{\alpha}\boldsymbol{\alpha}') \boldsymbol{\Phi}_1(t) = \boldsymbol{\Phi}'_1(s) \boldsymbol{\Sigma}_{11} \boldsymbol{\Phi}'_1(t).$$

Analogicznie otrzymujemy:

$$r_{\mathbf{X}\mathbf{X}}(s, t) = \boldsymbol{\Phi}_2(s) \boldsymbol{\Sigma}_{22} \boldsymbol{\Phi}'_2(t),$$

$$r_{\mathbf{Y}\mathbf{X}}(s, t) = \boldsymbol{\Phi}_1(s) \boldsymbol{\Sigma}_{12} \boldsymbol{\Phi}'_2(t).$$

Bez utraty ogólności możemy założyć, że macierze $\boldsymbol{\Sigma}_{11}$ oraz $\boldsymbol{\Sigma}_{22}$ są pełnego rzędu kolumnowego.

Można założyć (Ramsay i Silverman (2005)), że funkcja wagowa $\mathbf{u}(t)$ i proces losowy $\mathbf{Y}(t)$ leżą w tej samej przestrzeni, tzn. funkcja $\mathbf{u}(t)$ może zostać zapisana w następującej postaci:

$$\mathbf{u}(t) = \boldsymbol{\Phi}(t) \mathbf{u},$$

gdzie $\mathbf{u} \in \mathbb{R}^{K_1+p}$. Wówczas

$$\langle \mathbf{u}(t), \mathbf{Y}(t) \rangle = \langle \Phi_1(t)\mathbf{u}, \Phi_1(t)\boldsymbol{\alpha} \rangle = \mathbf{u}' \langle \Phi_1(t), \Phi_1(t) \rangle \boldsymbol{\alpha} = \mathbf{u}' \mathbf{I}_{K_1+p} \boldsymbol{\alpha} = \mathbf{u}' \boldsymbol{\alpha}.$$

Analogicznie dla $\mathbf{v}(t) \in L_2(I_2)^q$ mamy:

$$\langle \mathbf{v}(t), \mathbf{X}(t) \rangle = \mathbf{v}' \boldsymbol{\beta},$$

gdzie $\mathbf{v} \in \mathbb{R}^{K_2+q}$. Stąd otrzymujemy:

$$\mathbb{E}(\langle \mathbf{u}(t), \mathbf{Y}(t) \rangle) = \mathbf{u}' \mathbb{E}(\boldsymbol{\alpha}) = \boldsymbol{\alpha}' \mathbf{0} = 0,$$

$$\mathbb{E}(\langle \mathbf{v}(t), \mathbf{X}(t) \rangle) = \mathbf{v}' \mathbb{E}(\boldsymbol{\beta}) = \boldsymbol{\beta}' \mathbf{0} = 0,$$

$$\text{Var}(\langle \mathbf{u}(t), \mathbf{Y}(t) \rangle) = \mathbf{u}' \boldsymbol{\Sigma}_{11} \mathbf{u},$$

$$\text{Var}(\langle \mathbf{v}(t), \mathbf{X}(t) \rangle) = \mathbf{v}' \boldsymbol{\Sigma}_{22} \mathbf{v},$$

$$\text{Cov}(\langle \mathbf{u}(t), \mathbf{Y}(t) \rangle, \langle \mathbf{v}(t), \mathbf{X}(t) \rangle) = \mathbf{u}' \boldsymbol{\Sigma}_{12} \mathbf{v}.$$

Rozważmy pierwszą korelację kanoniczną pomiędzy procesami losowymi $\mathbf{Y}(t)$ oraz $\mathbf{X}(t)$:

$$\begin{aligned} \rho_1 = & \sup_{\mathbf{u}(t) \in L_2(I_1^p), \mathbf{v}(t) \in L_2(I_2^q)} \text{Cov}(\langle \mathbf{u}(t), \mathbf{Y}(t) \rangle, \langle \mathbf{v}(t), \mathbf{X}(t) \rangle) = \\ & \text{Cov}(\langle \mathbf{u}_1(t), \mathbf{Y}(t) \rangle, \langle \mathbf{v}_1(t), \mathbf{X}(t) \rangle), \end{aligned}$$

gdzie $\mathbf{u}(t)$ oraz $\mathbf{v}(t)$ spełniają odpowiednio ograniczenia (4.5) oraz (4.6). Jest to równoważne temu, że

$$\gamma_1 = \sup_{\mathbf{u} \in \mathbb{R}^{K_1+p}, \mathbf{v} \in \mathbb{R}^{K_2+q}} \mathbf{u}' \boldsymbol{\Sigma}_{12} \mathbf{v} = \mathbf{u}'_1 \boldsymbol{\Sigma}_{12} \mathbf{v}_1,$$

przy dodatkowych warunkach ograniczających

$$\mathbf{u}'_1 (\boldsymbol{\Sigma}_{11} + \lambda \mathbf{R}_1) \mathbf{u}_1 = 1,$$

$$\mathbf{v}'_1 (\boldsymbol{\Sigma}_{22} + \lambda \mathbf{R}_2) \mathbf{v}_1 = 1,$$

gdzie macierze \mathbf{R}_1 oraz \mathbf{R}_2 są wyrażone odpowiednio za pomocą równości (4.7) oraz (4.8). Jest to definicja pierwszej korelacji kanonicznej pomiędzy wektorami losowymi $\boldsymbol{\alpha}$ oraz $\boldsymbol{\beta}$.

Z drugiej strony, jeżeli zaczniemy od pierwszego układu kanonicznego $(\gamma_1, \mathbf{u}_1, \mathbf{v}_1)$ pary wektorów losowych $\boldsymbol{\alpha}$ oraz $\boldsymbol{\beta}$, to pierwszy układ kanoniczny pary procesów losowych $\mathbf{Y}(t)$ oraz $\mathbf{X}(t)$ otrzymamy z równości:

$$\rho_1 = \gamma_1, \quad \mathbf{u}_1(t) = \Phi_1(t) \mathbf{u}_1, \quad \mathbf{v}_1(t) = \Phi_2(t) \mathbf{v}_1.$$

Możemy rozszerzyć nasze rozważania na drugi układ kanoniczny i kolejne. □

Z Twierdzenie 4.1 wynika, że zmienne kanoniczne U_k i V_k dla pary procesów losowych $\mathbf{Y}(t) = \Phi_1(t)\boldsymbol{\alpha}$ oraz $\mathbf{X}(t) = \Phi_2(t)\boldsymbol{\beta}$ są postaci $U_k = \mathbf{u}'_k\boldsymbol{\alpha}$ oraz $V_k = \mathbf{v}'_k\boldsymbol{\beta}$, gdzie \mathbf{u}_k oraz \mathbf{v}_k są wektorami wagowymi w zmiennych kanonicznych $U_k^* = \mathbf{u}'_k\boldsymbol{\alpha}$ oraz $V_k^* = \mathbf{v}'_k\boldsymbol{\beta}$ dla pary wektorów losowych $\boldsymbol{\alpha}$ oraz $\boldsymbol{\beta}$, $k = 1, 2, \dots, \min(K_1 + p, K_2 + q)$.

Analiza korelacji kanonicznych dla wektorów losowych $\boldsymbol{\alpha}$ oraz $\boldsymbol{\beta}$ bazuje na macierzach $\boldsymbol{\Sigma}_{11}$, $\boldsymbol{\Sigma}_{22}$ i $\boldsymbol{\Sigma}_{12}$, które są nieznane. Estymujemy je na podstawie n niezależnych realizacji $\mathbf{y}_1(t), \mathbf{y}_2(t), \dots, \mathbf{y}_n(t)$ postaci $\mathbf{y}_i(t) = \Phi_1(t)\hat{\boldsymbol{\alpha}}_i$ procesu losowego $\mathbf{Y}(t)$, oraz $\mathbf{x}_1(t), \mathbf{x}_2(t), \dots, \mathbf{x}_n(t)$ postaci $\mathbf{x}_i(t) = \Phi_2(t)\hat{\boldsymbol{\beta}}_i$ procesu losowego $\mathbf{X}(t)$, $i = 1, 2, \dots, n$, gdzie

$$\hat{\boldsymbol{\alpha}}_i = (\hat{\alpha}_{10}^{(i)}, \dots, \hat{\alpha}_{1E_1}^{(i)}, \dots, \hat{\alpha}_{p0}^{(i)}, \dots, \hat{\alpha}_{pE_p}^{(i)})',$$

$$\hat{\boldsymbol{\beta}}_i = (\hat{\beta}_{10}^{(i)}, \dots, \hat{\beta}_{1F_1}^{(i)}, \dots, \hat{\beta}_{q0}^{(i)}, \dots, \hat{\beta}_{qF_q}^{(i)})'.$$

Niech

$$\hat{\mathbf{A}} = (\hat{\boldsymbol{\alpha}}_1, \dots, \hat{\boldsymbol{\alpha}}_n)'$$

oraz

$$\hat{\mathbf{B}} = (\hat{\boldsymbol{\beta}}_1, \dots, \hat{\boldsymbol{\beta}}_n)'.$$

Wówczas estymatory macierzy $\boldsymbol{\Sigma}_{11}$, $\boldsymbol{\Sigma}_{22}$ oraz $\boldsymbol{\Sigma}_{12}$ są odpowiednio postaci:

$$\hat{\boldsymbol{\Sigma}}_{11} = \frac{1}{n}\hat{\mathbf{A}}'\hat{\mathbf{A}},$$

$$\hat{\boldsymbol{\Sigma}}_{22} = \frac{1}{n}\hat{\mathbf{B}}'\hat{\mathbf{B}},$$

$$\hat{\boldsymbol{\Sigma}}_{12} = \frac{1}{n}\hat{\mathbf{A}}'\hat{\mathbf{B}}.$$

Niech ponadto $\hat{\mathbf{C}} = \hat{\boldsymbol{\Sigma}}_{11}^{-1}\hat{\boldsymbol{\Sigma}}_{12}$ oraz $\hat{\mathbf{D}} = \hat{\boldsymbol{\Sigma}}_{22}^{-1}\hat{\boldsymbol{\Sigma}}_{21}$, gdzie $\hat{\boldsymbol{\Sigma}}_{21} = \hat{\boldsymbol{\Sigma}}_{12}'$. Macierze $\hat{\mathbf{C}}\hat{\mathbf{D}}$ i $\hat{\mathbf{D}}\hat{\mathbf{C}}$ mają te same niezerowe wartości własne $\hat{\gamma}_k^2$, a odpowiadające im wektory własne $\hat{\mathbf{u}}_k$ oraz $\hat{\mathbf{v}}_k$ są wyznaczone z równości:

$$(\hat{\mathbf{C}}\hat{\mathbf{D}} - \hat{\gamma}_k^2\mathbf{I}_{K_1+p})\hat{\mathbf{u}}_k = \mathbf{0},$$

$$(\hat{\mathbf{D}}\hat{\mathbf{C}} - \hat{\gamma}_k^2\mathbf{I}_{K_2+q})\hat{\mathbf{v}}_k = \mathbf{0},$$

$k = 1, \dots, \min(K_1 + p, K_2 + q)$.

Wówczas k -ty układ kanoniczny pary procesów losowych $\mathbf{Y}(t)$ oraz $\mathbf{X}(t)$ wyznaczony z próby ma następującą postać:

$$(\hat{\lambda}_k = \hat{\gamma}_k, \hat{\mathbf{u}}_k(t) = \Phi_1(t)\hat{\mathbf{u}}_k, \hat{\mathbf{v}}_k(t) = \Phi_2(t)\hat{\mathbf{v}}_k), \quad k = 1, \dots, \min(K_1 + p, K_2 + q).$$

Stąd współczynniki rzutu i -tej realizacji $\mathbf{y}_i(t)$ procesu $\mathbf{Y}(t)$ na k -tą funkcjonalną zmienną kanoniczną są równe:

$$\hat{U}_{ik} = \langle \hat{\mathbf{u}}_k(t), \mathbf{y}_i(t) \rangle = \int_{I_1} \hat{\mathbf{u}}_k(t)\mathbf{y}_i(t)dt = \hat{\boldsymbol{\alpha}}_i'\hat{\mathbf{u}}_k.$$

Analogicznie współczynniki rzutu i -tej realizacji $\mathbf{x}_i(t)$ procesu $\mathbf{X}(t)$ na k -tą funkcjonalną zmienną kanoniczną są równe:

$$\hat{V}_{ik} = \hat{\beta}'_i \hat{v}_k,$$

gdzie $i = 1, 2, \dots, n$, $k = 1, \dots, \min(K_1 + p, K_2 + q)$.

Ostatecznie współczynniki rzutu i -tej realizacji $(\mathbf{y}_i(t), \mathbf{x}_i(t))$ procesów losowych $\mathbf{Y}(t)$ oraz $\mathbf{X}(t)$ na płaszczyznę dwóch pierwszych funkcjonalnych zmiennych kanonicznych z próby są równe $(\hat{v}'_1 \hat{\alpha}_i, \hat{u}'_1 \hat{\beta}_i)$, $i = 1, 2, \dots, n$. Ponadto rzut na płaszczyznę $(\hat{u}'_1 \hat{\alpha}_i, \hat{u}'_2 \hat{\alpha}_i)$, $i = 1, 2, \dots, n$, jest rzutem na płaszczyznę dwóch pierwszych składowych głównych procesu $\mathbf{Y}(t)$, a $(\hat{v}'_1 \hat{\beta}_i, \hat{v}'_2 \hat{\beta}_i)$, $i = 1, 2, \dots, n$, jest rzutem na płaszczyznę dwóch pierwszych składowych głównych procesu $\mathbf{X}(t)$.

Prawdziwe są następujące własności zmiennych kanonicznych: zmienne kanoniczne nie są wzajemnie ortogonalne, tzn. dla $i \neq j$:

$$\langle U_i, U_j \rangle = U'_i U_j = (\mathbf{u}'_i \boldsymbol{\alpha})' \mathbf{u}'_j \boldsymbol{\alpha} = \boldsymbol{\alpha}' \mathbf{u}_i \mathbf{u}'_j \boldsymbol{\alpha} \neq 0.$$

$$\langle V_i, V_j \rangle = V'_i V_j = (\mathbf{v}'_i \boldsymbol{\beta})' \mathbf{v}'_j \boldsymbol{\beta} = \boldsymbol{\beta}' \mathbf{v}_i \mathbf{v}'_j \boldsymbol{\beta} \neq 0.$$

$$\langle U_i, V_j \rangle = U'_i V_j = (\mathbf{u}'_i \boldsymbol{\alpha})' \mathbf{v}'_j \boldsymbol{\beta} = \boldsymbol{\alpha}' \mathbf{u}_i \mathbf{v}'_j \boldsymbol{\beta} \neq 0.$$

Ponadto zmienne kanoniczne są skorelowane, a współczynnik korelacji (przy założeniu, że $\text{Var}(U_i) = 1 = \text{Var}(V_j)$) między zmiennymi kanonicznymi $U_i = \mathbf{u}'_i \boldsymbol{\alpha}$ oraz $V_j = \mathbf{v}'_j \boldsymbol{\beta}$ jest równy:

$$\begin{aligned} \rho(U_i, V_j) &= \text{Cov}(\mathbf{u}'_i \boldsymbol{\alpha}, \mathbf{v}'_j \boldsymbol{\beta}) = \mathbf{u}'_i \text{Cov}(\boldsymbol{\alpha}, \boldsymbol{\beta}) \mathbf{v}_j = \\ &= \mathbf{u}'_i \boldsymbol{\Sigma}_{12} \mathbf{v}_j = \begin{cases} \gamma_i = \rho_i, & \text{gdzie } i = j, \\ 0, & \text{gdzie } i \neq j. \end{cases} \end{aligned}$$

Dodatkowo zmienne kanoniczne $U_i = \mathbf{u}'_i \boldsymbol{\alpha}$ oraz $U_j = \mathbf{u}'_j \boldsymbol{\alpha}$ nie są wzajemnie skorelowane, tzn.

$$\begin{aligned} \rho(U_i, U_j) &= \text{Cov}(U_i, U_j) = \text{Cov}(\langle \mathbf{u}_i(t), \mathbf{X}(t) \rangle, \langle \mathbf{u}_j(t), \mathbf{X}(t) \rangle) = \\ &= \text{Cov}(\mathbf{u}'_i \boldsymbol{\alpha}, \mathbf{u}'_j \boldsymbol{\alpha}) = \mathbf{u}'_i \text{Cov}(\boldsymbol{\alpha}, \boldsymbol{\alpha}) \mathbf{u}_j = \mathbf{u}'_i \boldsymbol{\Sigma}_{11} \mathbf{u}_j = \begin{cases} 1, & \text{gdzie } i = j, \\ 0, & \text{gdzie } i \neq j. \end{cases} \end{aligned}$$

Analogicznie zmienne kanoniczne $V_i = \mathbf{v}'_i \boldsymbol{\beta}$ oraz $V_j = \mathbf{v}'_j \boldsymbol{\beta}$ nie są wzajemnie skorelowane, tzn.

$$\begin{aligned} \rho(V_i, V_j) &= \text{Cov}(V_i, V_j) = \text{Cov}(\langle \mathbf{v}_i(t), \mathbf{Y}(t) \rangle, \langle \mathbf{v}_j(t), \mathbf{Y}(t) \rangle) = \\ &= \text{Cov}(\mathbf{v}'_i \boldsymbol{\beta}, \mathbf{v}'_j \boldsymbol{\beta}) = \mathbf{v}'_i \text{Cov}(\boldsymbol{\beta}, \boldsymbol{\beta}) \mathbf{v}_j = \mathbf{v}'_i \boldsymbol{\Sigma}_{22} \mathbf{v}_j = \begin{cases} 1, & \text{gdzie } i = j, \\ 0, & \text{gdzie } i \neq j. \end{cases} \end{aligned}$$

Ostatecznie możemy zapisać:

$$\rho(U_i, V_j) = \begin{cases} \rho_i, & \text{gdzie } i = j, \\ 0, & \text{gdzie } i \neq j. \end{cases}$$

oraz $\rho(U_i, U_j) = 0 = \rho(V_i, V_j)$, gdy $i \neq j$. Własności te są treścią Twierdzenia 8.1 ze str. 252 podręcznika Krzyśko (2009).

O wkładzie poszczególnych składowych wektorowych procesów losowych $\mathbf{Y}(t)$ oraz $\mathbf{X}(t)$ w budowę zmiennych kanonicznych można wnioskować na podstawie wektorowych funkcji wagowych $\mathbf{u}(t)$ oraz $\mathbf{v}(t)$ analogicznie jak w przypadku funkcjonalnych składowych głównych.

Rozdział 5

Zastosowanie wybranych metod statystycznych do wielozmiennych danych funkcjonalnych

5.1 Przykład 1.

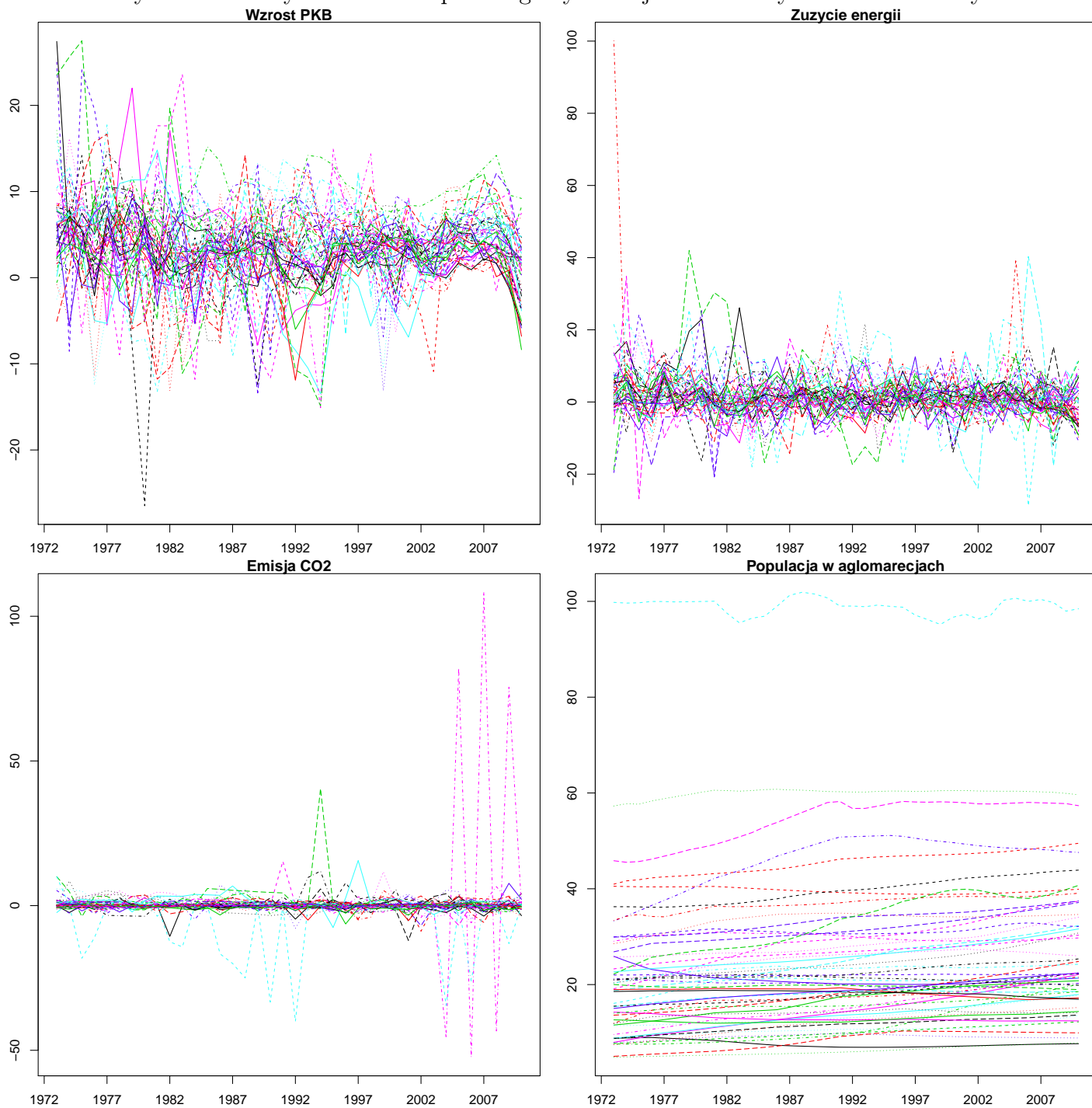
Dane pochodzą ze strony internetowej Banku Światowego (<http://data.worldbank.org/>). Do analizy wybrano 54 kraje świata ($n = 54$). Nazwy tych krajów zostały zaprezentowane w Tab. 5.1, a wykresy wartości poszczególnych cech na Rys. 5.5. Obserwacje były prowadzone w latach 1972–2009 ($J = 38$). Każdy kraj należy do jednej z 4 klas (wg stanu z 2009):

1. Niski dochód (Low-income) (PKB \$1.025 lub mniej), $n_1 = 3$
2. Nisko-średni dochód (Lower-middle-income) (PKB \$1.026 do \$4.035), $n_2 = 19$
3. Średni dochód (Upper-middle-income) (PKB \$4.036 do \$12.475), $n_3 = 14$
4. Wysoki dochód (High-income) (PKB \$12.476 lub więcej), $n_4 = 18$

i scharakteryzowany jest za pomocą 4 procesów:

- X_1 : Wzrost PKB (roczny wzrost w %)
- X_2 : Zużycie energii (stopa wzrostu zużycia oleju w kg na osobę)
- X_3 : Emisja CO₂ (stopa wzrostu emisji w kt)
- X_4 : Populacja w aglomeracjach pow. 1 miliona ludności (% całej populacji kraju)

Rysunek 5.1: Wykres cech dla poszczególnych krajów w badanym okresie czasowym



Dane zostały przekształcone do danych funkcjonalnych za pomocą metody opisanej w Rozdziale 1., a do obliczeń użyto bazę Fouriera. Momentom czasowym $t_i \in [0, T] = [0, 38]$ przyporządkowane zostały poszczególne lata w następujący sposób: $t_1 = 0,5(1972), t_2 = 1,5(1973), \dots, t_{38} = 37,5(2009)$. Optymalna wielkość bazy B dla każdego procesu $X_i, i = 1, 2, 3, 4$ jest omówiona poniżej, oddzielnie dla funkcjonalnych składowych głównych, zmiennych dyskryminacyjnych oraz zmiennych kanonicznych.

Tablica 5.1: Kod i nazwa kraju

Klasa	Kod kraju	Nazwa kraju	Klasa	Kod kraju	Nazwa kraju
1	COD	Congo, Dem. Rep.		CRI	Costa Rica
	KEN	Kenya		CUB	Cuba
	TGO	Togo		DOM	Dominican Republic
2	BOL	Bolivia	3	ECU	Ecuador
	CMR	Cameroon		IRN	Iran, Islamic Rep.
	COG	Congo, Rep.		PAN	Panama
	CIV	Cote d'Ivoire		PER	Peru
	EGY	Egypt, Arab Rep.		ZAF	South Africa
	SLV	El Salvador		THA	Thailand
	GHA	Ghana	4	AUS	Australia
	HND	Honduras		AUT	Austria
	IND	India		CAN	Canada
	IDN	Indonesia		DNK	Denmark
	MAR	Morocco		FIN	Finland
	NIC	Nicaragua		FRA	France
	NGA	Nigeria		GRC	Greece
	PAK	Pakistan		HUN	Hungary
	PRY	Paraguay		ISR	Israel
PHL	Philippines	ITA	Italy		
SEN	Senegal	JPN	Japan		
SDN	Sudan	KOR	Korea, Rep.		
SYR	Syrian Arab Republic	NLD	Netherlands		
3	DZA	Algeria	PRT	Portugal	
	ARG	Argentina	SAU	Saudi Arabia	
	CHL	Chile	SGP	Singapore	
	CHN	China	ESP	Spain	
	COL	Colombia	SWE	Sweden	

5.1.1 Funkcjonalne składowe główne

W Tab. 5.2(góra) zaprezentowano wariacje wyjaśnione przez dwie pierwsze funkcjonalne składowe główne w zależności od kryterium informacyjnego i metody wyboru liczby elementów bazy,

a w Tab. 5.2(dół) odpowiadającą im liczbę elementów bazy B dla poszczególnych procesów. Można zauważyć, że w prezentowanym przykładzie w przypadku funkcjonalnych składowych głównych, wpływ na wyjaśnienie wariancji ma jedynie metoda wyboru liczby elementów bazy bez względu na wybrane kryterium informacyjne. W związku z tym możemy ograniczyć się jedynie do kryterium BIC. Optymalna liczba elementów bazy B została wybrana przy użyciu metody minimum, gdyż taki wybór pozwala na najlepsze wyjaśnienie wariancji przez dwie pierwsze funkcjonalne składowe główne i wynosi 2 dla wszystkich procesów X_1, X_2, X_3 oraz X_4 .

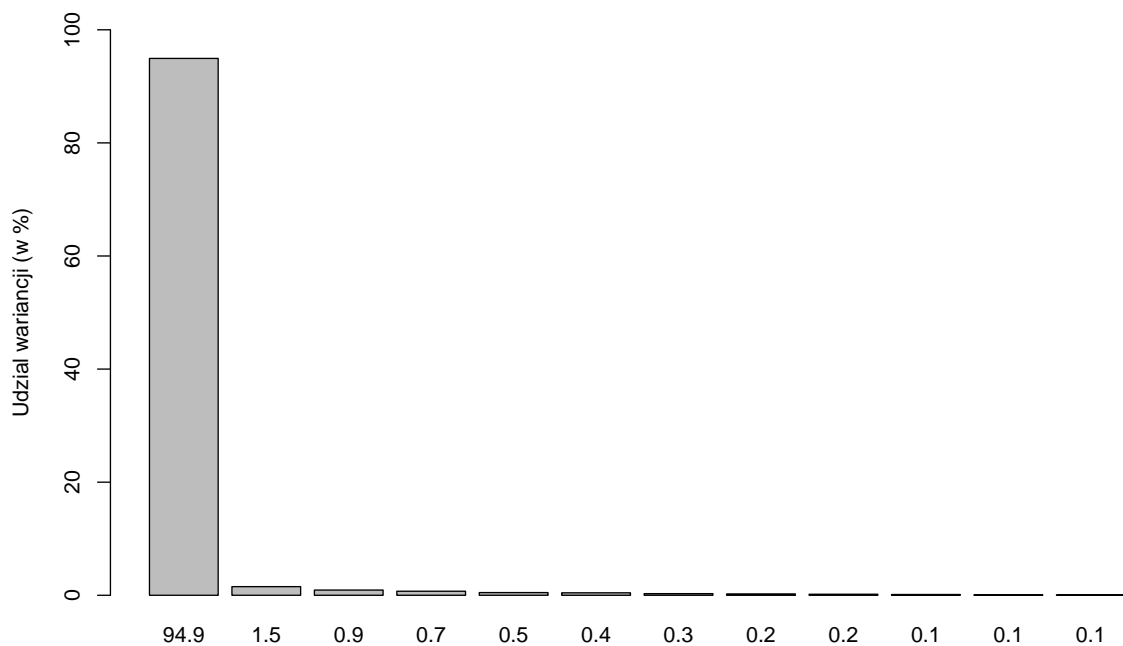
Obiektami statystycznymi w funkcjonalnej analizie składowych głównych są 54 państwa ($n = 54$) scharakteryzowane czterema ($p = 4$) danymi funkcjonalnymi $\mathbf{x}_i(t) = (x_{i1}(t), x_{i2}(t), x_{i3}(t), x_{i4}(t))'$, $t \in [0, 38]$, $i = 1, 2, \dots, 54$. Nie bierze się pod uwagę przynależności obiektu do jednej z czterech wyróżnionych grup państw. Funkcje wektorowe $\mathbf{x}_1(t), \mathbf{x}_2(t), \dots, \mathbf{x}_{54}(t)$ mają postać

$$\mathbf{x}_i(t) = \Phi(t)\hat{\mathbf{c}}_i,$$

gdzie $\Phi(t)$ jest macierzą postaci (1.3), natomiast wektor $\hat{\mathbf{c}}_i$ ma postać

$$\hat{\mathbf{c}}_i = (\hat{c}_{i10}(t), \dots, \hat{c}_{i1B_1}, \dots, \hat{c}_{i40}(t), \dots, \hat{c}_{i4B_4})',$$

gdzie $B_1 = B_2 = B_3 = B_4 = 2, i = 1, 2, \dots, 54$. W kroku pierwszym z wektorów $\hat{\mathbf{c}}_1, \hat{\mathbf{c}}_2, \dots, \hat{\mathbf{c}}_{54}$ budujemy macierz $\hat{\Sigma}$, a następnie znajdujemy jej wartości własne $\hat{\gamma}_k$ i odpowiadające im wektory $\hat{\mathbf{u}}_k$. Stosunki danej wartości własnej do sumy wszystkich wartości własnych wyrażone w procentach pokazane są na Rys. 5.2. Z Rys. 5.2 wynika, że 94,9% zmienności całkowitej przypada na pierwszą funkcjonalną składową główną.



Rysunek 5.2: Udział wariancji dla funkcjonalnych składowych głównych (MFPCA).

Tablica 5.2: Wariancje wyjaśnione przez dwie pierwsze funkcjonalne składowe główne (górze) oraz liczba elementów bazy dla poszczególnych procesów X_i PCA (dół)

Wariancje wyjaśnione przez dwie pierwsze składowe główne						
Kryterium	Metoda	λ_1	$\lambda_1/\Sigma\lambda_i(\%)$	λ_2	$\lambda_2/\Sigma\lambda_i(\%)$	$(\lambda_1 + \lambda_2)/\Sigma\lambda_i (\%)$
BIC	min	$9,07 \times 10^3$	94,94	$1,46 \times 10^2$	1,53	96,47
	max	$9,10 \times 10^3$	88,95	$1,92 \times 10^2$	1,87	90,83
	dominanta	$9,07 \times 10^3$	94,75	$1,46 \times 10^2$	1,53	96,28
	średnia	$9,07 \times 10^3$	94,18	$1,50 \times 10^2$	1,56	95,74
eBIC	min	$9,07 \times 10^3$	94,94	$1,46 \times 10^2$	1,53	96,47
	max	$9,10 \times 10^3$	88,95	$1,92 \times 10^2$	1,87	90,83
	dominanta	$9,07 \times 10^3$	94,75	$1,46 \times 10^2$	1,53	96,28
	średnia	$9,07 \times 10^3$	94,18	$1,50 \times 10^2$	1,56	95,74
AIC	min	$9,07 \times 10^3$	94,94	$1,46 \times 10^2$	1,53	96,47
	max	$9,10 \times 10^3$	88,95	$1,92 \times 10^2$	1,87	90,83
	dominanta	$9,07 \times 10^3$	94,75	$1,46 \times 10^2$	1,53	96,28
	średnia	$9,07 \times 10^3$	94,18	$1,50 \times 10^2$	1,56	95,74

Baza PCA					
Kryterium	Metoda	X_1	X_2	X_3	X_4
BIC	min	3	3	3	3
	max	19	9	17	13
	dominanta	3	3	3	7
	średnia	5	3	3	5
eBIC	min	3	3	3	3
	max	19	9	17	13
	dominanta	3	3	3	7
	średnia	5	3	3	5
AIC	min	3	3	3	3
	max	19	9	17	13
	dominanta	3	3	3	7
	średnia	5	3	3	5

W kroku drugim tworzymy wektorowe funkcje wagowe

$$\hat{\mathbf{u}}_k(t) = \Phi(t)\hat{\mathbf{u}}_k,$$

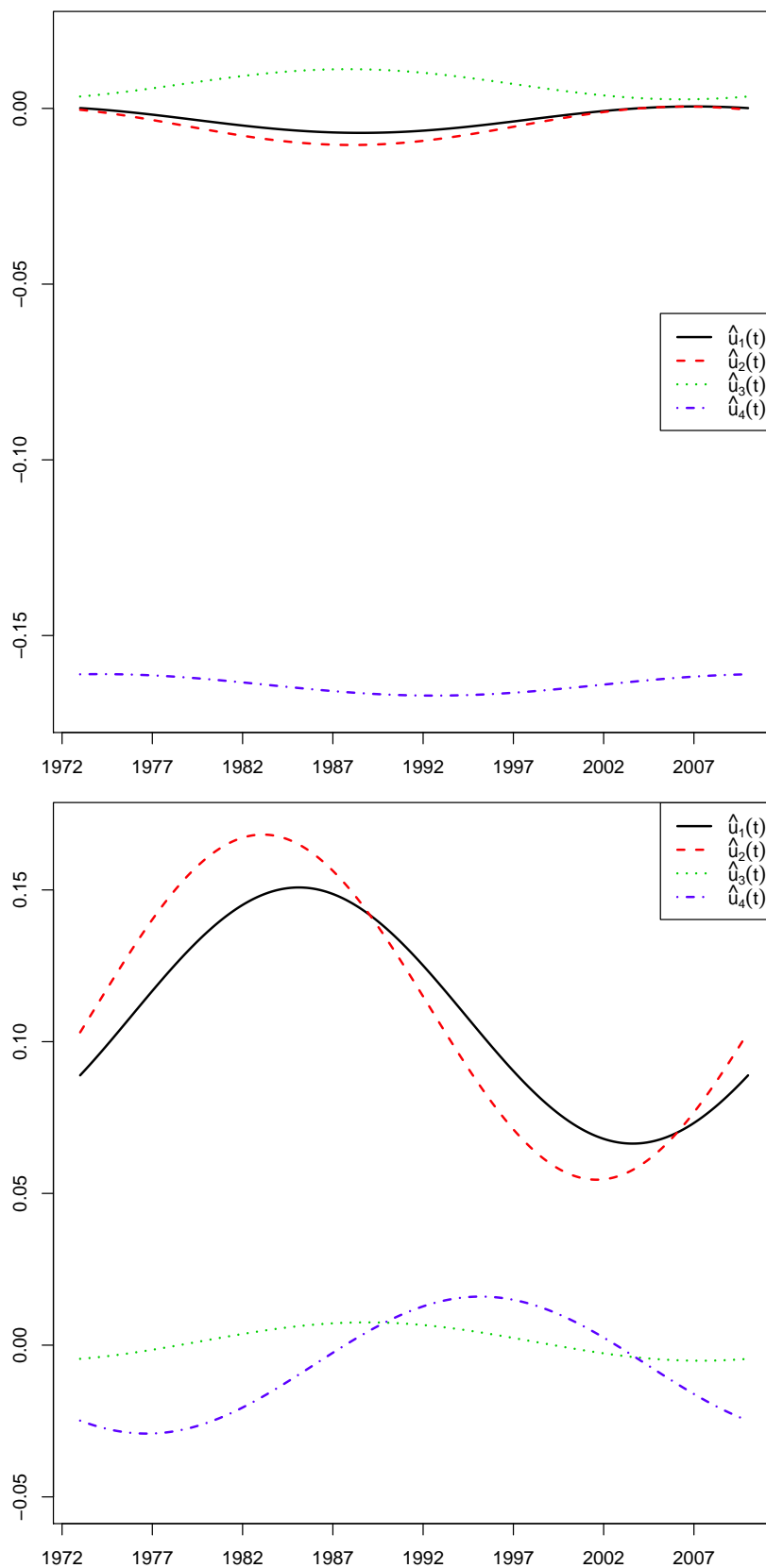
gdzie $k = 1, \dots, 16$, oraz odpowiadające im funkcjonalne składowe główne postaci

$$\hat{U}_k = \langle \hat{\mathbf{u}}_k(t), \mathbf{X}(t) \rangle = \hat{\mathbf{u}}_k' \hat{\mathbf{c}}.$$

Wykresy czterech składowych wektorowej funkcji wagowej dla pierwszej i drugiej funkcjonalnej składowej głównej pokazane są odpowiednio na Rys. 5.3.

W określonym punkcie czasowym t , im większa jest wartość bezwzględna składowej wektorowej funkcji wagowej, tym większy jest udział tej składowej w budowie danej funkcjonalnej składowej głównej. Z Rys. 5.3 (górze) widzimy, że największy udział w budowie pierwszej funkcjonalnej składowej głównej przez wszystkie brane pod uwagę lata obserwacji ma proces $X_4(t)$ (populacja w aglomeracjach), a jego udział w budowie tej zmiennej wynosi 92%. Oznacza to, że konstrukcję pierwszej składowej głównej moglibyśmy właściwie ograniczyć do obserwacji liczby ludności żyjących w aglomeracjach poszczególnych państw. Z Rys. 5.3 (dół) widzimy natomiast, że na określonych przedziałach czasowych naprzemiennie największy udział w budowie drugiej funkcjonalnej składowej głównej mają procesy $X_2(t)$ (zużycie energii) i $X_1(t)$ (wzrost PKB). Widzimy, że dwa procesy łącznie, zużycie energii i wzrost PKB, mają 92% udział w budowie drugiej składowej głównej. Jednocześnie możemy zauważyć, że emisja dwutlenku węgla ma bardzo mały wpływ (poniżej 5% zarówno na budowę pierwszej jak i drugiej składowej głównej). Udział danego procesu pierwotnego $X_i(t)$ w budowie pierwszej i drugiej funkcjonalnej składowej głównej oszacować można za pomocą pola pod modułem funkcji wagowych odpowiadających temu procesowi i opisane to zostało w Rozdziale 2., a wyniki te zawiera Tab. 5.3, przy czym P_i oznacza pole pod krzywą $|u_i(t)|$, P_i^* dane jest wzorem (2.5).

Wzajemne położenie 54 państw w układzie dwóch pierwszych funkcjonalnych składowych

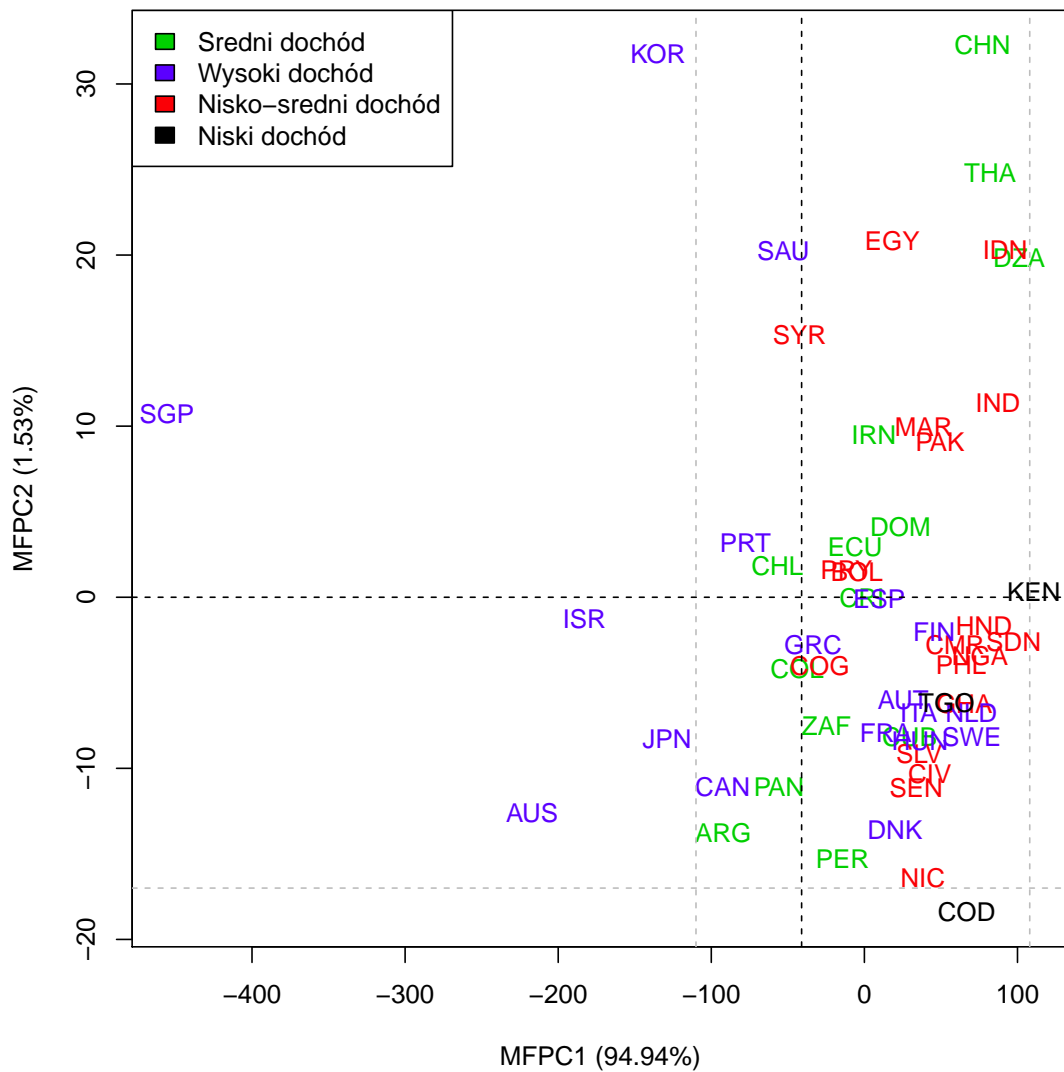


Rysunek 5.3: Funkcje wagowe dla pierwszej funkcjonalnej składowej głównej \hat{U}_1 (góra). Funkcje wagowe dla drugiej funkcjonalnej składowej głównej \hat{U}_2 (dół).

głównych (\hat{U}_1, \hat{U}_2) pokazane jest na Rys. 5.4. Układ dwóch pierwszych funkcjonalnych składowych głównych zachowuje 96,3% zmienności całkowitej. Z rysunku tego można zaobserwować kraje odstające takie jak Singapur (SGP), Korea Południowa (KOR), Chiny (CHN) czy Demokratyczna Republika Konga (COD). Kraje dla których pierwsza funkcjonalna składowa $U_1 < -41$ należą do krajów o średnim lub wysokim PKB, a w dodatku gdy składowa ta $U_1 < -110$ to zawsze mamy do czynienia z krajem o wysokim PKD. Wszystkie kraje o najniższym PKB mają zawsze drugą funkcjonalną składową główną $U_2 < 0$. Zauważmy, że jeżeli jednocześnie $U_1 > 0$ oraz $U_2 > -41$ to z pewnością mamy do czynienia z krajem o średnim lub nisko-średnim dochodzie. Najtrudniej identyfikowalne są państwa leżące w prostokącie $-17 < U_1 < 0$ oraz $-41 < U_2 < 108$, który skupia kraje należące do wszystkich wybranych przez nas klas.

Tablica 5.3: Pola pod funkcjami wagowymi $u_i(t)$ dla pierwszej i drugiej funkcjonalnej składowej głównej.

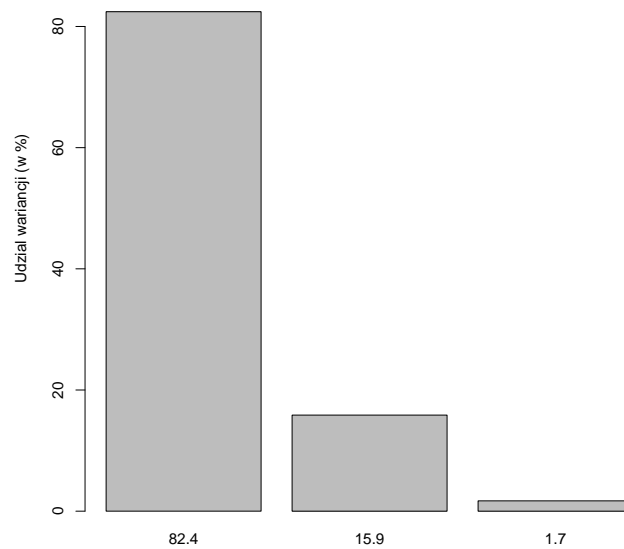
Funkcje wagowe $u_i(t)$ dla pierwszej funkcjonalnej składowej głównej U_1			
i	P_i	P_i^*	ΣP_i^*
4	6,07	92%	92%
3	0,25	4%	95%
2	0,19	3%	98%
1	0,12	2%	100%
Funkcje wagowe $u_i(t)$ dla drugiej funkcjonalnej składowej głównej U_2			
i	P_i	P_i^*	ΣP_i^*
2	4,12	47%	47%
1	4,02	45%	92%
4	0,55	6%	98%
3	0,15	2%	100%



Rysunek 5.4: Rzut 54 krajów na płaszczyznę dwóch pierwszych funkcjonalnych składowych głównych (\hat{U}_1, \hat{U}_2) (MFDC).

5.1.2 Funkcjonalne zmienne dyskryminacyjne

W Tab. 5.4(góra) zaprezentowano wariacje wyjaśnione przez dwie pierwsze funkcjonalne zmienne dyskryminacyjne w zależności od kryterium informacyjnego i metody wyboru liczby elementów bazy, a w Tab. 5.4(dół) odpowiadającą im liczbę elementów bazy B dla poszczególnych procesów. Można zauważyć, że analogicznie jak przy funkcjonalnych składowych głównych w prezentowanym przykładzie dla funkcjonalnych zmiennych dyskryminacyjnych, wpływ na wyjaśnienie wariacji ma metoda wyboru liczby elementów bazy bez względu na wybrane kryterium informacyjne. Tak jak poprzednio ograniczymy się jedynie do kryterium BIC. W tym przypadku dla zmiennych dyskryminacyjnych optymalna liczba elementów bazy B została wybrana przy



Rysunek 5.5: Udział wariancji dla funkcjonalnych zmiennych dyskryminacyjnych (MFDC).

użyciu metody maksimum, która pozwala na najlepsze wyjaśnienie wariancji przez dwie pierwsze zmienne i wynosi ona odpowiednio 18, 8, 16 oraz 12 odpowiednio dla procesów X_1, X_2, X_3 i X_4 .

W konstrukcji funkcjonalnych zmiennych dyskryminacyjnych, w porównaniu z konstrukcją funkcjonalnych składowych głównych, uwzględniamy dodatkowo informację o podziale 54 państw na 4 rozłączne grupy ($L = 4$). Z wektorów \hat{c}_i budujemy estymator \hat{B} macierzy zmienności między klasami oraz estymator \hat{T} macierzy zmienności całkowitej, a następnie znajdujemy niezerowe wartości własne $\hat{\gamma}_k$ macierzy $\hat{T}^{-1}\hat{B}$ oraz odpowiadające im wektory $\hat{u}_k, k = 1, 2, 3$. Stosunki danej wartości własnej do sumy wszystkich wartości własnych wyrażone są w procentach na Rys. 5.5. Z Rys. 5.5 wynika, że 82,4% zmienności całkowitej przypada na pierwszą funkcjonalną zmienną dyskryminacyjną.

W kroku drugim tworzymy wektorowe funkcje wagowe

$$\hat{u}_k(t) = \Phi(t)\hat{u}_k,$$

gdzie $k = 1, 2, 3$, oraz odpowiadające im funkcjonalne zmienne dyskryminacyjne postaci

$$\hat{U}_k = \langle \hat{u}_k(t), \mathbf{X}(t) \rangle = \hat{u}_k' \hat{c}.$$

Wykresy czterech składowych wektorowej funkcji wagowej dla pierwszej i drugiej funkcjonalnej zmiennej dyskryminacyjnej pokazane są odpowiednio na Rys. 5.6. Z Rys. 5.6 widzimy, że w przypadku funkcjonalnych zmiennych dyskryminacyjnych brak jest funkcji wagowej która by dominowała na całym przedziale czasowym tak jak to miało miejsce przy konstrukcji funk-

Tablica 5.4: Wariancje wyjaśnione przez dwie pierwsze funkcjonalne zmienne dyskryminacyjne (górze) oraz liczba elementów bazy dla poszczególnych procesów X_i (dół).

Wariancje wyjaśnione przez dwie pierwsze funkcjonalne zmienne dyskryminacyjne						
Kryterium	Metoda	λ_1	$\lambda_1/\Sigma\lambda_i(\%)$	λ_2	$\lambda_2/\Sigma\lambda_i(\%)$	$(\lambda_1 + \lambda_2)/\Sigma\lambda_i(\%)$
BIC	min	1,21	56,14	0,73	33,95	90,09
	max	$9,33 \times 10^7$	82,44	$1,79 \times 10^7$	15,86	98,30
	dominanta	3,11	74,75	0,80	19,24	93,99
	średnia	2,52	65,67	0,91	23,82	89,49
eBIC	min	1,21	56,14	0,73	33,95	90,09
	max	$9,33 \times 10^7$	82,44	$1,79 \times 10^7$	15,86	98,30
	dominanta	3,11	74,75	0,80	19,24	93,99
	średnia	2,52	65,67	0,91	23,82	89,49
AIC	min	1,21	56,14	0,73	33,95	90,09
	max	$9,33 \times 10^7$	82,44	$1,79 \times 10^7$	15,86	98,30
	dominanta	3,11	74,75	0,80	19,24	93,99
	średnia	2,52	65,67	0,91	23,82	89,49

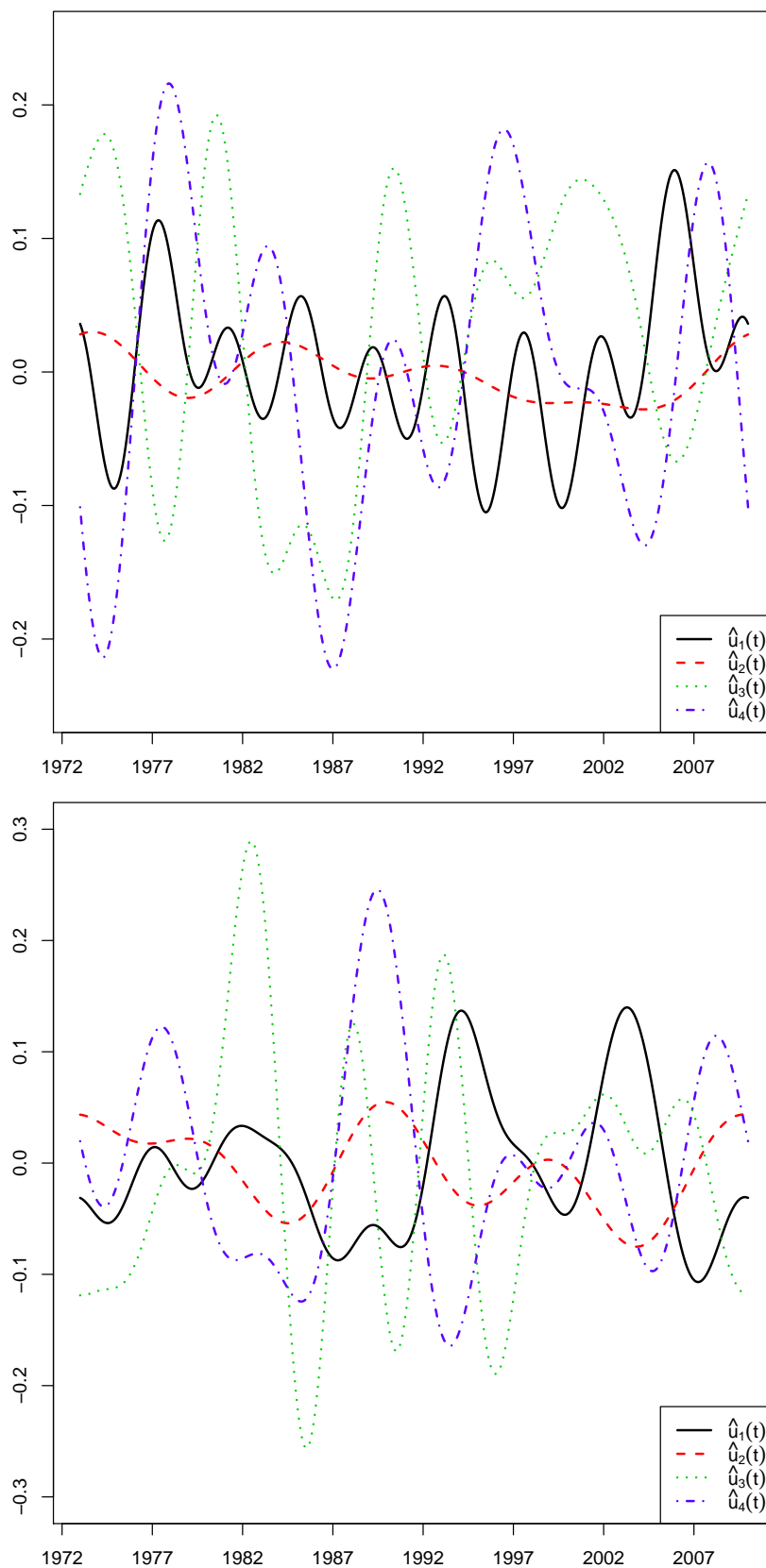
Baza DC					
Kryterium	Metoda	X_1	X_2	X_3	X_4
BIC	min	3	3	3	3
	max	19	9	17	13
	dominanta	3	3	3	7
	średnia	5	3	3	5
eBIC	min	3	3	3	3
	max	19	9	17	13
	dominanta	3	3	3	7
	średnia	5	3	3	5
AIC	min	3	3	3	3
	max	19	9	17	13
	dominanta	3	3	3	7
	średnia	5	3	3	5

Tablica 5.5: Pola pod funkcjami wagowymi $u_i(t)$ dla pierwszej i drugiej funkcjonalnej zmiennej dyskryminacyjnej.

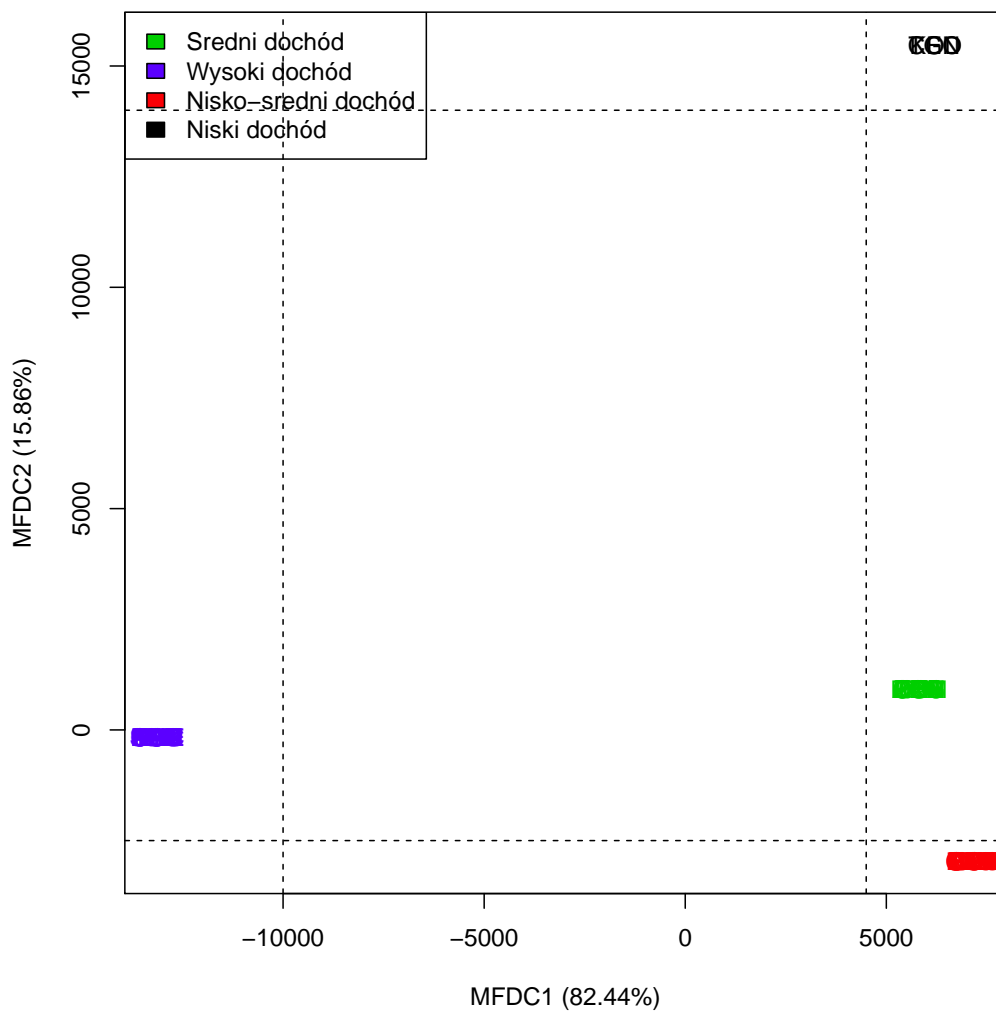
Funkcje wagowe $u_i(t)$ dla pierwszej funkcjonalnej zmiennej dyskryminacyjnej U_1			
i	P_i	P_i^*	ΣP_i^*
3	3,45	39%	39%
4	3,35	38%	77%
1	1,56	18%	95%
2	0,54	6%	100%
Funkcje wagowe $u_i(t)$ dla drugiej funkcjonalnej zmiennej dyskryminacyjnej U_2			
i	P_i	P_i^*	ΣP_i^*
3	3,30	37%	37%
4	2,73	30%	67%
1	1,87	21%	88%
2	1,12	12%	100%

cjonalnych składowych głównych. W budowie pierwszej i drugiej funkcjonalnej zmiennej dyskryminacyjnej największy udział na proces $X_3(t)$ (emisja CO_2) oraz $X_4(t)$ (populacja w aglomeracjach), natomiast na niewielkich przedziałach czasowych dominuje proces $X_1(t)$ (wzrost PKB). Podczas gdy emisja CO_2 na największy udział (powyżej 36%) w budowie obu funkcjonalnych zmiennych dyskryminacyjnych, to w przypadku funkcjonalnych składowych głównych miała wpływ na poziomie poniżej 5%. Udziały poszczególnych procesów w budowie pierwszej i drugiej funkcjonalnej zmiennej dyskryminacyjnej procesu wektorowego $\mathbf{X}(t)$ zawiera Tab. 5.5.

Wzajemne położenie 54 państw w układzie dwóch pierwszych funkcjonalnych zmiennych dyskryminacyjnych pokazane jest na Rys. 5.7. Układ dwóch pierwszych funkcjonalnych zmiennych dyskryminacyjnych zachowuje 98,30% zmienności całkowitej. Wszystkie kraje, dla których pierwsza funkcjonalna zmienna dyskryminacyjna $U_1 < -10000$ należą do krajów wysokim PKB. Kraje o PKB nie większym niż \$12.475 mają zawsze pierwszą zmienną $U_1 > 4500$. Na podstawie drugiej funkcjonalnej zmiennej dyskryminacyjnej możemy rozróżniać kraje o niskim i nisko-średnim dochodzie. Otóż, gdy druga funkcjonalna zmienna dyskryminacyjna $U_2 < -2500$, to wówczas są to zawsze kraje o nisko-średnim dochodzie, a gdy $U_2 > 14000$ to o niskim dochodzie. Kraje o PKB większym niż \$4.036 mają drugą zmienną $-2500 < U_2 < 14000$ (kraj o co najmniej średnim dochodzie). Używając jednocześnie dwóch funkcjonalnych zmiennych dyskryminacyjnych możemy jednoznacznie rozróżniać kraje o średnim PKB, tzn. dla wszystkich krajów z tej grupy zachodzą warunki $-2500 < U_2 < 14000$ oraz $U_1 > 4500$. W porównaniu z rzutem na dwie pierwsze funkcjonalne składowe główne, podział na 4 jednorodne grupy jest tutaj wyraźnie widoczny.



Rysunek 5.6: Funkcje wagowe dla pierwszej funkcjonalnej zmiennej dyskryminacyjnej \hat{U}_1 (góra). Funkcje wagowe dla drugiej funkcjonalnej zmiennej dyskryminacyjnej \hat{U}_2 (dół).



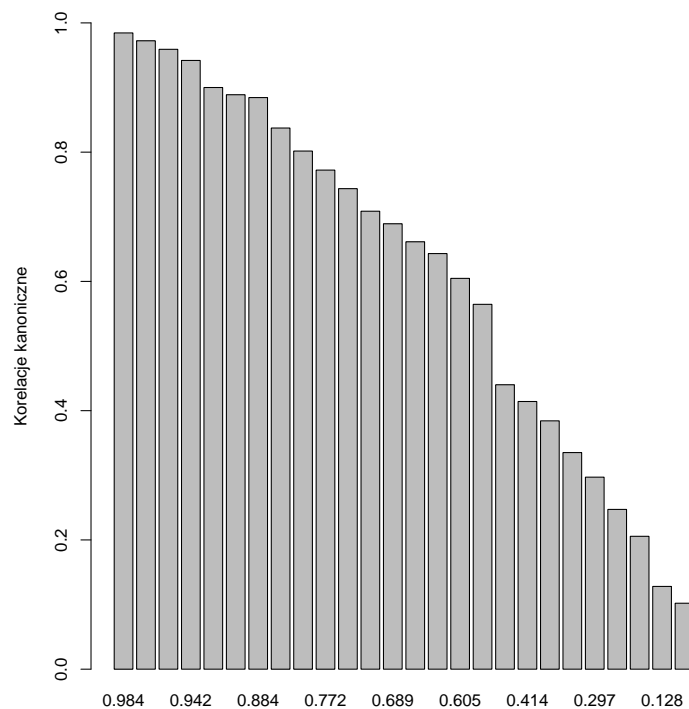
Rysunek 5.7: Rzut 54 krajów na płaszczyznę dwóch pierwszych funkcjonalnych zmiennych dyskryminacyjnych (\hat{U}_1, \hat{U}_2) (MFDC).

5.1.3 Funkcjonalne zmienne kanoniczne

W konstrukcji funkcjonalnych zmiennych kanonicznych nie uwzględniamy podziału 54 państw na 4 grupy, a czterowymiarowy proces stochastyczny dzielimy na dwie części: $\mathbf{Y}(t) = (X_2(t), X_3(t))'$ oraz $\mathbf{X}(t) = (X_1(t), X_4(t))'$. W naszym przypadku $p = q = 2$. Interesuje nas zależność między procesami $\mathbf{Y}(t)$ oraz $\mathbf{X}(t)$. Budujemy estymatory macierzy $\hat{\Sigma}_{11}$, $\hat{\Sigma}_{22}$ i $\hat{\Sigma}_{12}$, a następnie znajdujemy niezerowe wartości własne $\hat{\gamma}_k^2$ i odpowiadające im wektory $\hat{\mathbf{u}}_k$ macierzy $\hat{\mathbf{C}}\hat{\mathbf{D}}$ oraz wartości własne $\hat{\gamma}_k^2$ i odpowiadające im wektory $\hat{\mathbf{v}}_k$ macierzy $\hat{\mathbf{D}}\hat{\mathbf{C}}$, gdzie $\hat{\mathbf{C}} = \hat{\Sigma}_{11}^{-1}\hat{\Sigma}_{12}$ oraz $\hat{\mathbf{D}} = \hat{\Sigma}_{22}^{-1}\hat{\Sigma}_{21}$, $\hat{\Sigma}_{21} = \hat{\Sigma}'_{12}$, $k = 1, \dots, 6$.

W Tab. 5.6 zaprezentowano dwa pierwsze współczynniki korelacji oraz odpowiadającą im liczbę elementów bazy dla poszczególnych procesów X_i w zależności od kryterium informacyjnego i metody wyboru liczby elementów bazy. Podobnie jak poprzednio przy konstrukcji funkcjonaln składowych głównych i zmiennych dyskryminacyjnych, współczynniki korelacji zależą jedynie od metody wyboru liczby elementów bazy bez względu na wybrane kryterium informacyjne. Optymalna liczba elementów bazy B została wybrana przy użyciu kryterium maksimum analogicznie jak przy funkcjonalnych zmiennych dyskryminacyjnych, i wynosi ona odpowiednio 18, 8, 16 oraz 12 odpowiednio dla procesów X_1, X_2, X_3 i X_4 .

Wartości własne $\hat{\gamma}_k$, nazywane korelacjami kanonicznymi, pokazane są na Rys. 5.8.



Rysunek 5.8: Współczynniki korelacji kanonicznych (MFCCA).

Tablica 5.6: Dwa pierwsze współczynniki korelacji kanonicznych (górze) oraz liczba elementów bazy dla poszczególnych procesów X_i (dół)

Dwa pierwsze współczynniki korelacji			
Kryterium	Metoda	ρ_1	ρ_2
BIC	min	0,9078	0,8453
	max	0,9845	0,9723
	dominanta	0,9509	0,8810
	średnia	0,9317	0,8963
eBIC	min	0,9078	0,8453
	max	0,9845	0,9723
	dominanta	0,9509	0,8810
	średnia	0,9317	0,8963
AIC	min	0,9078	0,8453
	max	0,9845	0,9723
	dominanta	0,9509	0,8810
	średnia	0,9317	0,8963

Baza CCA					
Kryterium	Metoda	X_1	X_2	X_3	X_4
BIC	min	3	3	3	3
	max	19	9	17	13
	dominanta	3	3	3	7
	średnia	5	3	3	5
eBIC	min	3	3	3	3
	max	19	9	17	13
	dominanta	3	3	3	7
	średnia	5	3	3	5
AIC	min	3	3	3	3
	max	19	9	17	13
	dominanta	3	3	3	7
	średnia	5	3	3	5

W kroku drugim tworzymy wektorowe funkcje wagowe

$$\hat{\mathbf{u}}_k(t) = \Phi_1(t)\hat{\mathbf{u}}_k, \quad \hat{\mathbf{v}}_k(t) = \Phi_2(t)\hat{\mathbf{v}}_k,$$

odpowiadające procesom $\mathbf{Y}(t)$ oraz $\mathbf{X}(t)$, gdzie $k = 1, \dots, 6$. Funkcjom tym odpowiadają funkcjonalne zmienne kanoniczne postaci

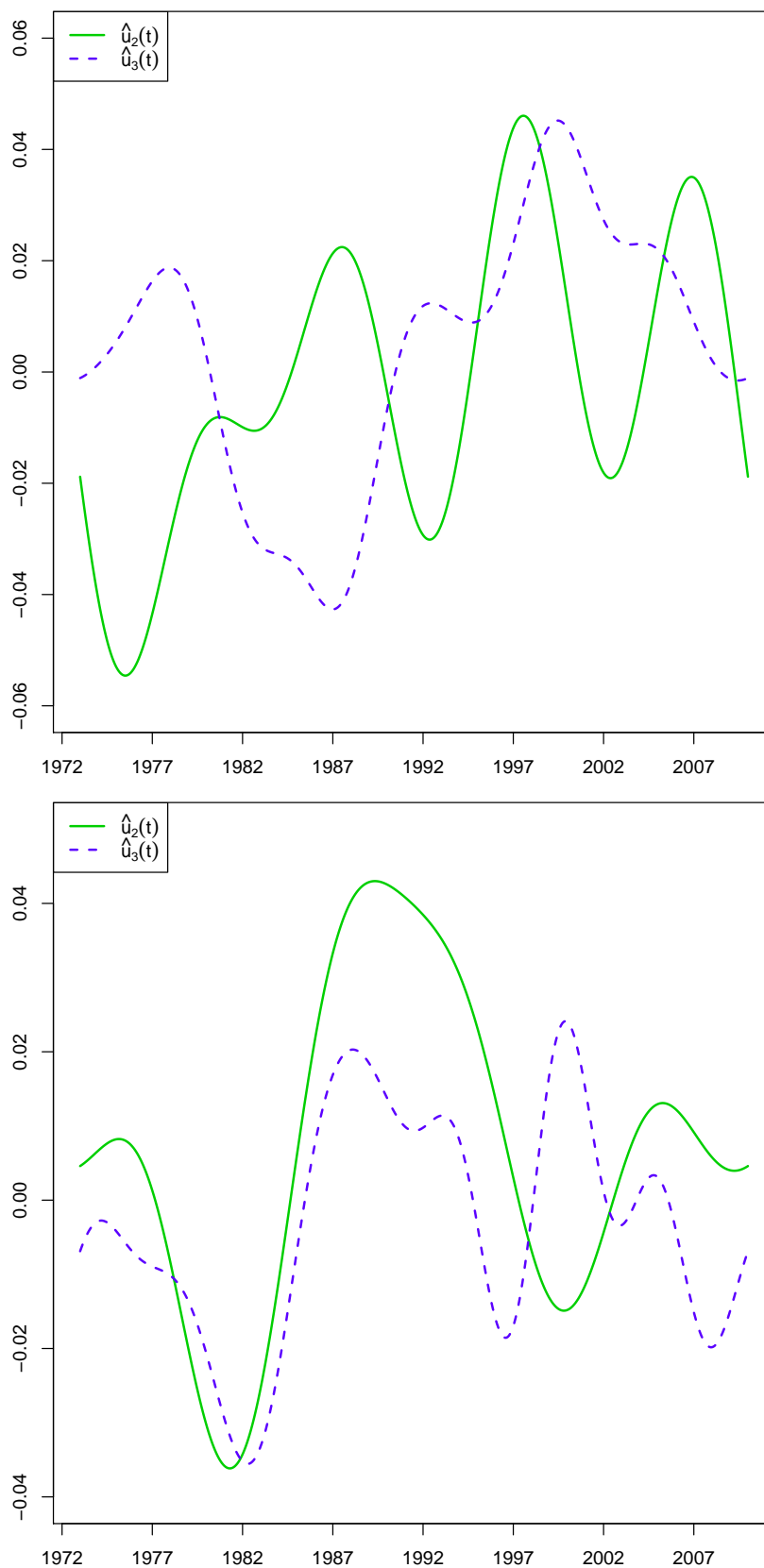
$$\hat{U}_k = \langle \hat{\mathbf{u}}_k(t), \mathbf{Y}(t) \rangle, \quad \hat{V}_k = \langle \hat{\mathbf{v}}_k(t), \mathbf{X}(t) \rangle,$$

odpowiadające procesom $\mathbf{Y}(t)$ oraz $\mathbf{X}(t)$. Wykresy dwóch składowych wektorowej funkcji wagowej dla pierwszej i drugiej funkcjonalnej zmiennej kanonicznej procesów $\mathbf{Y}(t)$ oraz $\mathbf{X}(t)$ pokazane są na Rys. 5.9 oraz 5.10.

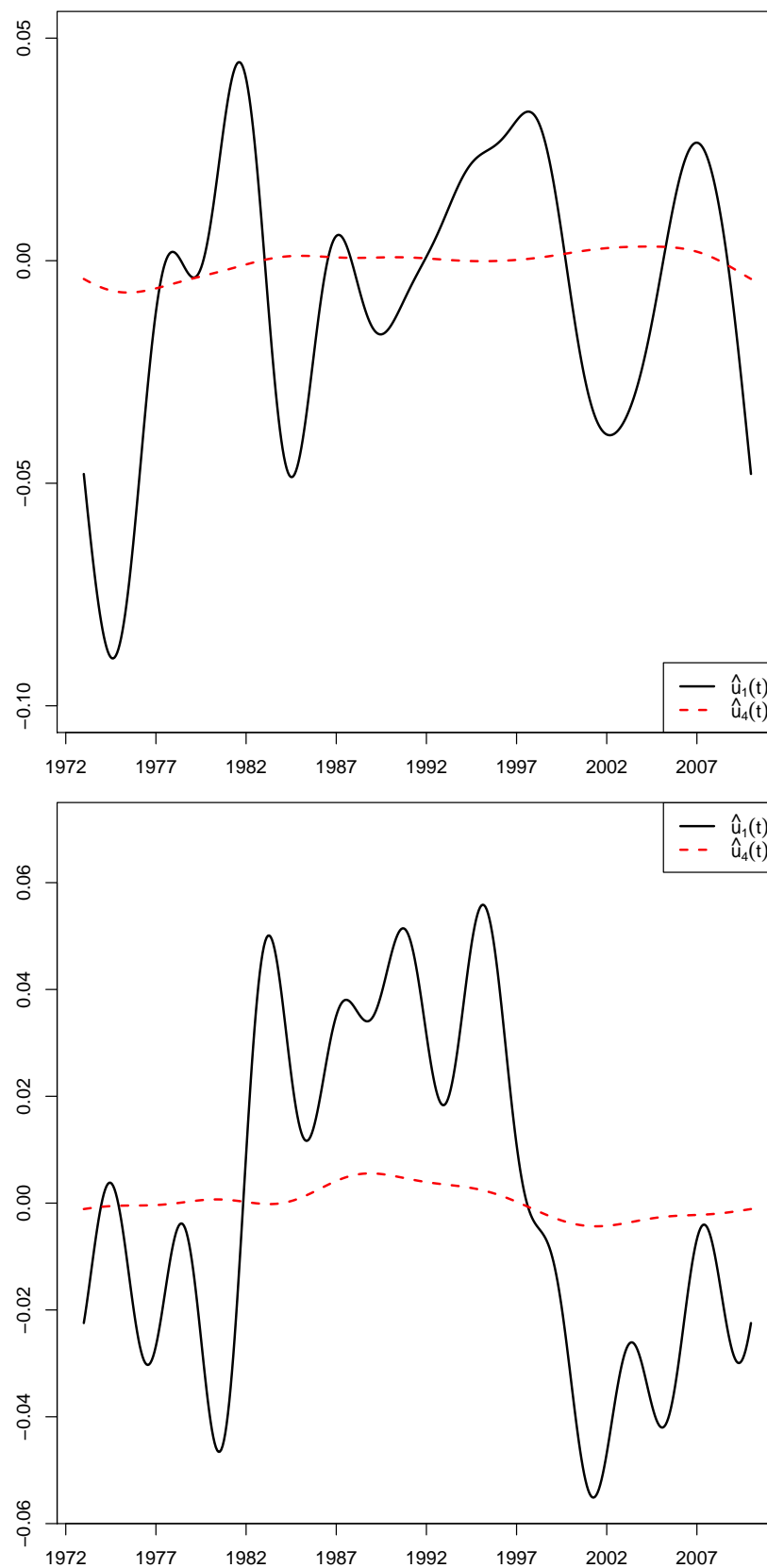
Z Rys. 5.9 widzimy, że naprzemiennie dominuje $X_2(t)$ (zużycie energii) oraz $X_3(t)$ (emisja CO_2), przy czym, zarówno w przypadku pierwszej i drugiej funkcjonalnej zmiennej kanonicznej procesu $\mathbf{Y}(t)$, wkład procesu opisującego zużycie energii w budowę tych zmiennych jest nieco większy (ponad 50%) niż procesu opisującego emisję CO_2 . Z kolei Rys. 5.10 ilustruje wyraźnie dominację procesu $X_1(t)$ (wzrost PKB), którego udział przy konstrukcji dwóch pierwszych funkcjonalnych zmiennych kanonicznych procesu $\mathbf{X}(t)$ jest na poziomie powyżej 90%. Przypomnijmy, że udział populacji w aglomeracjach w budowie pierwszej funkcjonalnej składowej głównej i zmiennej dyskryminacyjnej był również na poziomie poniżej 20%. Udziały poszczególnych procesów w budowie pierwszej i drugiej funkcjonalnej zmiennej kanonicznej procesów wektorowych $\mathbf{X}(t)$ oraz $\mathbf{Y}(t)$ zawiera Tab. 5.7.

Wzajemne położenie 54 państw w układach (\hat{U}_1, \hat{U}_2) , (\hat{V}_1, \hat{V}_2) oraz (\hat{U}_1, \hat{V}_1) funkcjonalnych zmiennych kanonicznych pokazane jest na Rys. 5.11 oraz 5.12.

Przypomnijmy, że $\mathbf{Y}(t) = (X_2(t), X_3(t))'$ oraz $\mathbf{X}(t) = (X_1(t), X_4(t))'$. Proces wektorowy $\mathbf{Y}(t)$ możemy nazwać grupą procesów ekologicznych, a proces $\mathbf{X}(t)$ grupą procesów ekonomicznych. Interesowała nas siła zależności pomiędzy tymi dwoma grupami procesów. Rys. 5.12 pokazuje rzut na płaszczyznę pierwszej pary funkcjonalnych zmiennych kanonicznych (\hat{U}_1, \hat{V}_1) . Współczynnik korelacji ρ_1 związany z tą parą jest na poziomie 0,98, co oznacza, że obserwacje leżą blisko prostej ($y = x$). Można to interpretować jako zastępowalność grupy procesów ekologicznych poprzez ekonomiczne i na odwrót. Jest to szczególnie ważne w niektórych praktycznych zagadnieniach, gdy np. pomiar jakiś procesów jest bardzo drogi lub wręcz niemożliwy. Ponadto układy (\hat{U}_1, \hat{U}_2) , (\hat{V}_1, \hat{V}_2) traktować można jako układy dwóch pierwszych funkcjonalnych składowych głównych dla wektorowych procesów $\mathbf{Y}(t)$ oraz $\mathbf{X}(t)$ odpowiednio.



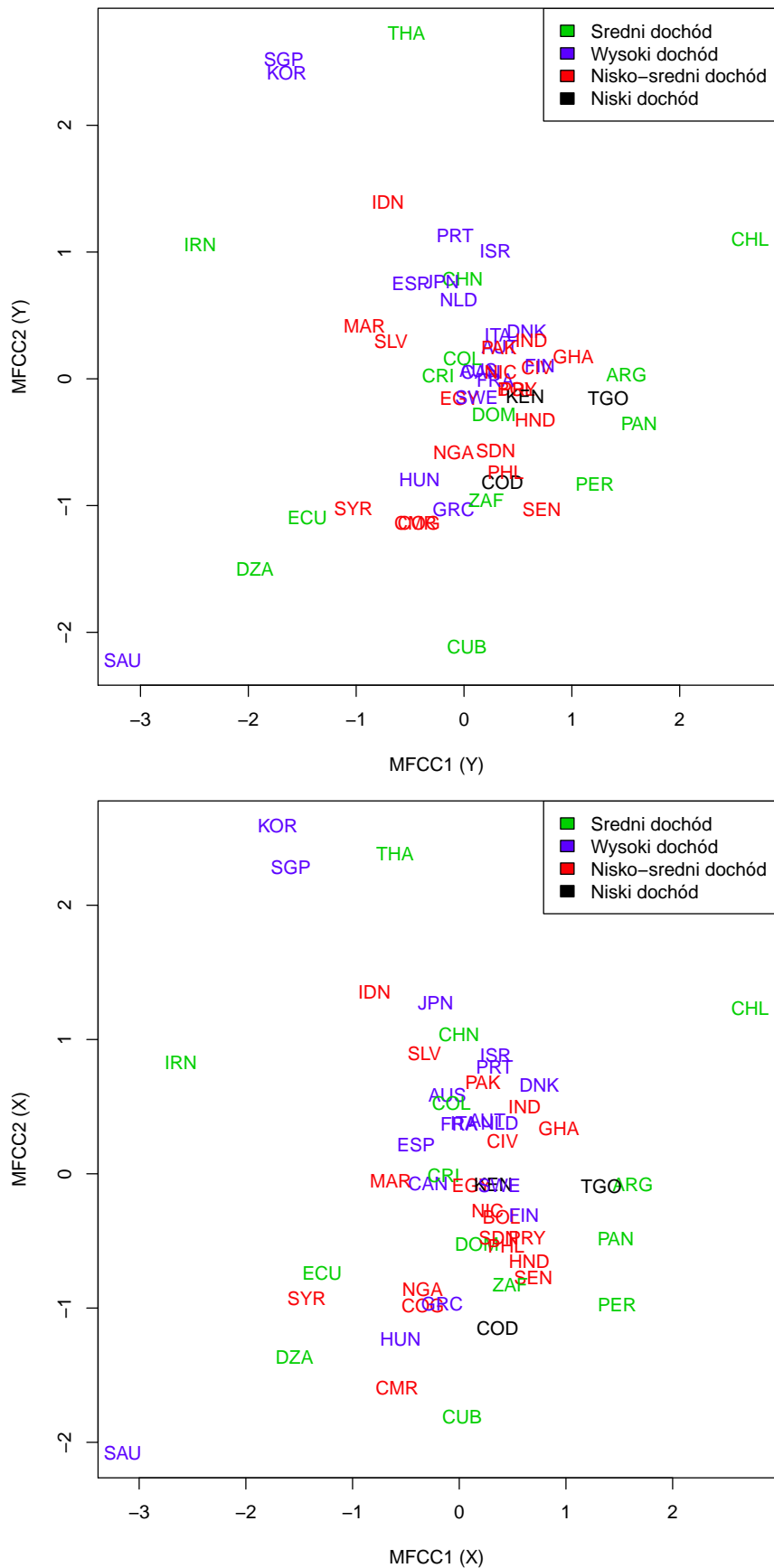
Rysunek 5.9: Funkcje wagowe dla pierwszej funkcjonalnej zmiennej kanonicznej \hat{U}_1 (górze). Funkcje wagowe dla drugiej funkcjonalnej zmiennej kanonicznej \hat{U}_2 (dół).



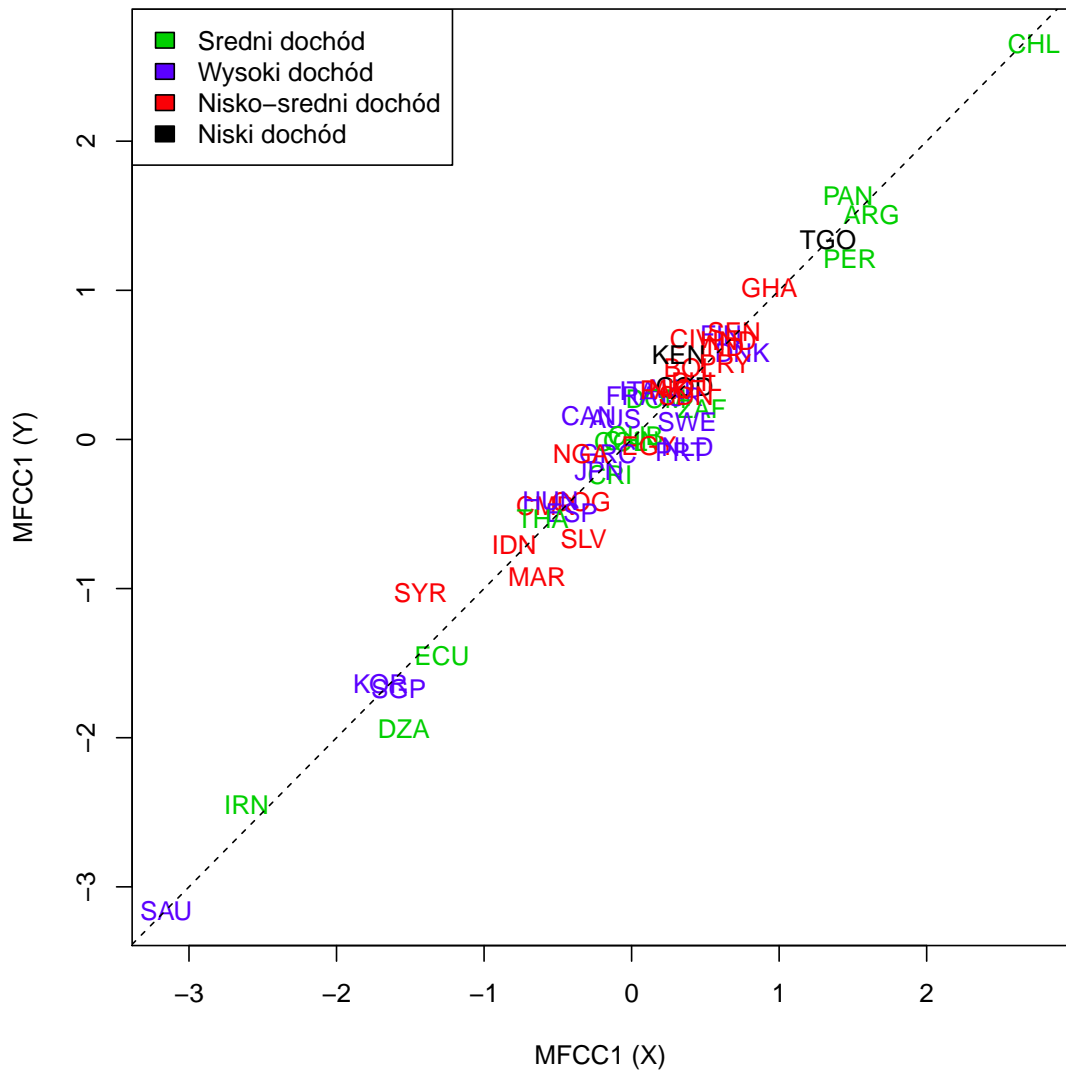
Rysunek 5.10: Funkcje wagowe dla pierwszej funkcjonalnej zmiennej kanonicznej \hat{V}_1 (z lewej). Funkcje wagowe dla drugiej funkcjonalnej zmiennej kanonicznej \hat{V}_2 (z prawej).

Tablica 5.7: Pola pod funkcjami wagowymi $u_i(t)$ dla pierwszej i drugiej funkcjonalnej zmiennej kanonicznej.

Funkcje wagowe dla pierwsza funkcjonalnej zmiennej kanonicznej V_1			
i	P_i	P_i^*	ΣP_i^*
1	0,89	92%	92%
4	0,08	8%	100%
Funkcje wagowe dla drugiej funkcjonalnej zmiennej kanonicznej V_2			
i	P_i	P_i^*	ΣP_i^*
1	1,02	93%	93%
4	0,08	7%	100%
Funkcje wagowe dla pierwsza funkcjonalnej zmiennej kanonicznej U_1			
i	P_i	P_i^*	ΣP_i^*
2	0,79	52%	52%
3	0,73	48%	100%
Funkcje wagowe dla drugiej funkcjonalnej zmiennej kanonicznej U_2			
i	P_i	P_i^*	ΣP_i^*
2	0,67	58%	58%
3	0,48	42%	100%



Rysunek 5.11: Rzut 54 krajów na płaszczyznę dwóch pierwszych funkcjonalnych zmiennych kanonicznych (\hat{U}_1, \hat{U}_2) przestrzeni $\mathbf{Y}(t)$ (góra) oraz (\hat{V}_1, \hat{V}_2) przestrzeni $\mathbf{X}(t)$ (dół) (MFCCA).



Rysunek 5.12: Rzut 54 krajów na płaszczyznę pierwszych funkcjonalnych zmiennych kanonicznych (\hat{U}_1, \hat{V}_1) (MFCCA).

5.2 Przykład 2.

Dane zostały udostępnione przez dra hab. inż. Rafała Nadulskiego z Uniwersytetu Przyrodniczego w Lublinie, a szczegółowy opis znajduje się w pracy Nadulskiego (2005). Badania prowadzono na jabłkach 6 odmian:

1. Arlet
2. Gala
3. Gloster
4. Golden (Golden Delicious)
5. Jonagold (Jonagold Decosta)
6. Szampion

Owoce pochodziły ze specjalistycznego gospodarstwa sadowniczego mieszczącego się na terenie gminy Markuszów. Owoce zrywano w pierwszej połowie października. Badania wykonano w sezonie 2009/2010. Dojrzałość jabłek określano na podstawie testu skrobiowego. Do badań pobrano owoce z szypułkami o zbliżonej wielkości. Owoce przechowywano w chłodni z kontrolowaną atmosferą (ULO) i w chłodni zwykłej (CH):

- ULO - przy obniżonej zawartości tlenu
- CH - atmosfera niezmienniona (taka jaką oddychamy)

Owoce przeznaczone do badań przechowywano w siatkach umieszczonych w ażurowych, plastikowych skrzynkach. Skrzynki ustawiano w sposób zapewniający swobodną cyrkulację powietrza. Testy wytrzymałościowe przeprowadzono w Katedrze Inżynierii i Maszyn Spożywczych Uniwersytetu Przyrodniczego w Lublinie. Owoce po wyjęciu z chłodni przetrzymywano w pomieszczeniu laboratoryjnym do osiągnięcia równowagi z warunkami otoczenia. Pomiary prowadzono w stałych warunkach w temperaturze $21^{\circ} \pm 1^{\circ}$ oraz przy wilgotności $65\% \pm 5\%$. Owoce badano bezpośrednio po zbiorze a następnie w odstępach co 15 dni. Badania mechanicznych właściwości tekstury prowadzono przy pomocy pięciu metod (z użyciem standardowej aparatury wytrzymałościowej Instron 4302 wyposażonej w głowicę pomiarową o zakresie do 10 kN), tj.:

- Metoda 1 (M1) - test penetrometryczny polegający na wbijaniu penetrometru o średnicy 11 mm na głębokość 8 mm
- Metoda 2 (M2) - test przebijania skórki penetrometrem o średnicy 8 mm

- Metoda 3 (M3) - test podwójnego ściskania próbki wyciętej z mięszu (wys. 10 mm, śr. 15 mm) (test TPA)
- Metoda 4 (M4) - test ściskania próbki j/w ale jednokrotny w zakresie sprężystym.
- Metoda 5 (M5) - test rozciągania próbki wyciętej z mięszu, kształt pierścienia szer. 10 mm, śr. wew. 11 mm, śr. zew 17 mm

W badaniach wykorzystano oprogramowanie firmy Instron serii IX. W przypadku każdego testu stosowano prędkość obciążania próbki wynosząca $20 \text{ mm} \cdot \text{min}^{-1}$. Każdy pomiar wykonano w dziesięciu powtórzeniach. W każdej z metod badano różne zastawy cech. Cechy badane w metodzie M1:

- X_1 : Fm (Fsf11) - siła penetracji (jędrność), [N]
- X_2 : SLm (SLsf11) - współczynnik sprężystości w teście penetracji, [$N \cdot \text{mm}^{-1}$]
- X_3 : Pm (Psf11) - praca penetracji, [$J \cdot 100^{-1}$]

Cechy badane w metodzie M2:

- X_4 : Fp (Fp8) - siła potrzebna do przebicia skórki owocu, [N]
- X_5 : Lp (Lp8) – deformacja odpowiadająca sile Fp (Fp8), [mm]
- X_6 : SLp (SLPp8) - współczynnik sprężystości w teście przebicia skórki, [$N \cdot \text{mm}^{-1}$]
- X_7 : Pp (Pp8) - praca potrzebna do przebicia skórki, [$J \cdot 100^{-1}$]

Cechy badane w metodzie M3:

- X_8 : FI (Fs1) - twardość I, [J]
- X_9 : FII (Fs2) - twardość II, [J]
- X_{10} : PI (Ps1) - praca deformacji I, [$J \cdot 100^{-1}$]
- X_{11} : PII (Ps2) - praca deformacji II, [$J \cdot 100^{-1}$]
- X_{12} : Cs – spoistość, [$\cdot 100^{-1}$]
- X_{13} : SLI (SL1s)- współczynnik sprężystości w teście podwójnego ściskania, [$N \cdot \text{mm}^{-1}$]

Cechy badane w metodzie M4:

- X_{14} : Fz (Fkr) - siła potrzebna do trwałej deformacji mięszu, [N]

- X_{15} : Lz (Lkr) - Lz (Lkr) - deformację względną odpowiadającą sile Fz (Fkr), [mm]
- X_{16} : Lz (Lkr) - deformację względną odpowiadającą sile Fz (Fkr), [%]
- X_{17} : Ss (S) - sprężystość, [mm · 10⁻¹]
- X_{18} : Ez - moduł sprężystości mięszu w teście ściskania, [MPa · 10⁻¹]

Cechy badane w metodzie M5:

- X_{19} : Fr – siła zrywająca próbkę, [N · 10⁻¹]
- X_{20} : Lr – deformacja odpowiadająca sile Fr, [N · 10⁻¹]
- X_{21} : SLr - współczynnik sprężystości, [N · mm⁻¹]
- X_{22} : Er - moduł sprężystości w teście zrywania, [MPa · 10⁻¹]

Reasumując przez 180 dni, co 15 dni badano 6 odmian jabłek (liczba klas $L = 6$) za pomocą 5 metod badawczych w 10 powtórzeniach. Pomiaru były dokonywane w dniach 0, 15, 30, ..., 165 i 180, łącznie w 13 punktach czasowych ($J = 13$). Momentom czasowym $t_i \in [0, T] = [0, 13]$, przyporządkowane zostały dni pomiarów w następujący sposób: $t_1 = 0,5$ (0 dzień), $t_2 = 1,5$ (15 dzień), $t_3 = 2,5$ (30 dzień), ..., $t_{13} = 12,5$ (180 dzień). Każda odmiana była badana w 10 powtórzeniach w każdym momencie czasowym i scharakteryzowana była za pomocą 22 cech, które uzyskano za pomocą 5 metod badawczych. Dane zostały przekształcone do danych funkcjonalnych za pomocą metody opisanej w Rozdziale 1. Do obliczeń użyto bazy Fouriera. Optymalny wybór liczby elementów bazy B dla poszczególnych procesów $X_i, i = 1, 2, \dots, 22$, zostanie omówiony poniżej, z osobna dla każdej z metod prezentowanych w Rodziałach 2.-4.

5.2.1 Funkcjonalne składowe główne

W Tab. 5.8(góra) zaprezentowano wariacje wyjaśnione przez dwie pierwsze funkcjonalne składowe główne w zależności od kryterium informacyjnego i metody wyboru liczby elementów bazy, a w Tab. 5.8(dół) odpowiadającą im liczbę elementów bazy dla poszczególnych procesów. Tak jak w poprzednim przykładzie i tym razem w przypadku funkcjonalnych składowych głównych, wpływ na wyjaśnienie wariacji ma jedynie metoda wyboru liczby elementów bazy bez względu na wybrane kryterium informacyjne. Możemy ograniczyć się znów do kryterium BIC. Optymalna liczba elementów bazy B została wybrana przy użyciu metody minimum, gdyż taki wybór pozwala na najlepsze wyjaśnienie wariacji przez dwie pierwsze funkcjonalne składowe główne i wynosi 2 dla wszystkich procesów $X_i, i = 1, \dots, 22$.

Obiektami statystycznymi w funkcjonalnej analizie składowych głównych jest 120 jabłek (10 powtórzeń każdej z 6 odmian w 2 warunkach przechowywania) ($n = 120$) scharakteryzowanych

Tablica 5.8: Wariancje wyjaśnione przez dwie pierwsze funkcjonalne składowe główne (góra) oraz liczba elementów bazy dla poszczególnych procesów X_i (dół).

Wariancje wyjaśnione przez dwie pierwsze składowe główne						
Kryterium	Metoda	λ_1	$\lambda_1/\Sigma\lambda_i(\%)$	λ_2	$\lambda_2/\Sigma\lambda_i(\%)$	$(\lambda_1 + \lambda_2)/\Sigma\lambda_i (\%)$
BIC	min	$5,24 \times 10^4$	75,41	$6,48 \times 10^3$	9,34	84,75
	max	$5,28 \times 10^4$	69,55	$8,45 \times 10^3$	11,12	80,67
	dominanta	$5,24 \times 10^4$	75,11	$6,54 \times 10^3$	9,37	84,48
	średnia	$5,24 \times 10^4$	74,90	$6,59 \times 10^3$	9,42	84,32
eBIC	min	$5,24 \times 10^4$	75,41	$6,48 \times 10^3$	9,34	84,75
	max	$5,28 \times 10^4$	69,55	$8,45 \times 10^3$	11,12	80,67
	dominanta	$5,24 \times 10^4$	75,11	$6,54 \times 10^3$	9,37	84,48
	średnia	$5,24 \times 10^4$	74,90	$6,59 \times 10^3$	9,42	84,32
AIC	min	$5,24 \times 10^4$	75,41	$6,48 \times 10^3$	9,34	84,75
	max	$5,28 \times 10^4$	69,55	$8,45 \times 10^3$	11,12	80,67
	dominanta	$5,24 \times 10^4$	75,11	$6,54 \times 10^3$	9,37	84,48
	średnia	$5,24 \times 10^4$	74,90	$6,59 \times 10^3$	9,42	84,32

PCA											
Kryterium	Metoda	Nr procesu X_i									
		1	2	3,4	5-7	8-10	11	12	13,14	15,16	17-22
BIC	min	3	3	3	3	3	3	3	3	3	3
	max	11	11	11	11	11	11	11	11	9	11
	dominanta	3	3	3	3	3	3	5	3	3	3
	średnia	3	5	3	5	3	5	5	3	3	3
eBIC	min	3	3	3	3	3	3	3	3	3	3
	max	11	11	11	11	11	11	11	11	9	11
	dominanta	3	3	3	3	3	3	5	3	3	3
	średnia	3	5	3	5	3	5	5	3	3	3
AIC	min	3	3	3	3	3	3	3	3	3	3
	max	11	11	11	11	11	11	11	11	9	11
	dominanta	3	3	3	3	3	3	5	3	3	3
	średnia	3	5	3	5	3	5	5	3	3	3

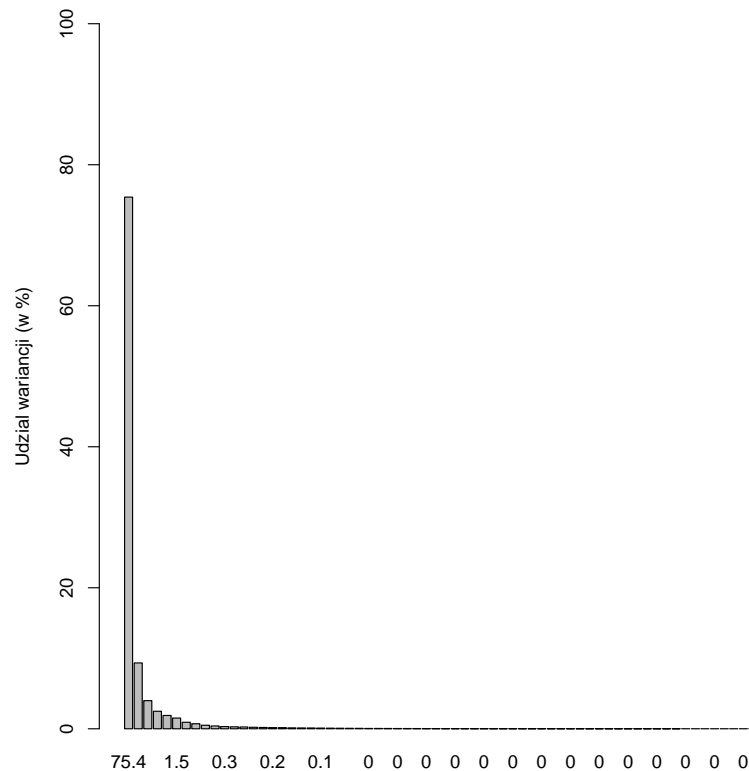
22 ($p = 22$) danymi funkcjonalnymi $\mathbf{x}_i(t) = (x_{i1}(t), x_{i2}(t), \dots, x_{i22}(t))'$, $t \in [0, 13]$, $i = 1, 2, \dots, 120$. Nie bierze się pod uwagę przynależności obiektu do jednej z sześciu wyróżnionych odmian ani warunków przechowywania. Funkcje wektorowe $\mathbf{x}_1(t), \mathbf{x}_2(t), \dots, \mathbf{x}_{120}(t)$ mają postać

$$\mathbf{x}_i(t) = \Phi(t)\hat{\mathbf{c}}_i,$$

gdzie $\Phi(t)$ jest macierzą postaci (1.3), natomiast wektor $\hat{\mathbf{c}}_i$ ma postać

$$\hat{\mathbf{c}}_i = (\hat{c}_{i10}(t), \dots, \hat{c}_{i1B_1}, \dots, \hat{c}_{i22,0}(t), \dots, \hat{c}_{i22B_{22}})'$$

gdzie $B_1 = B_2 = \dots = B_{22} = 2$, $i = 1, 2, \dots, 120$. W kroku pierwszym z wektorów $\hat{\mathbf{c}}_1, \hat{\mathbf{c}}_2, \dots, \hat{\mathbf{c}}_{120}$ budujemy macierz $\hat{\Sigma}$, a następnie znajdujemy jej wartości własne $\hat{\gamma}_k$ i odpowiadające im wektory $\hat{\mathbf{u}}_k$. Stosunki danej wartości własnej do sumy wszystkich wartości własnych wyrażone w procentach pokazane są na Rys. 5.13. Z Rys. 5.13 wynika, że 75,4% zmienności całkowitej



Rysunek 5.13: Udział wariancji dla poszczególnych funkcjonalnych składowych głównych (MFPCA).

przypada na pierwszą funkcjonalną składową główną.

W kroku drugim tworzymy wektorowe funkcje wagowe

$$\hat{\mathbf{u}}_k(t) = \Phi(t)\hat{\mathbf{u}}_k,$$

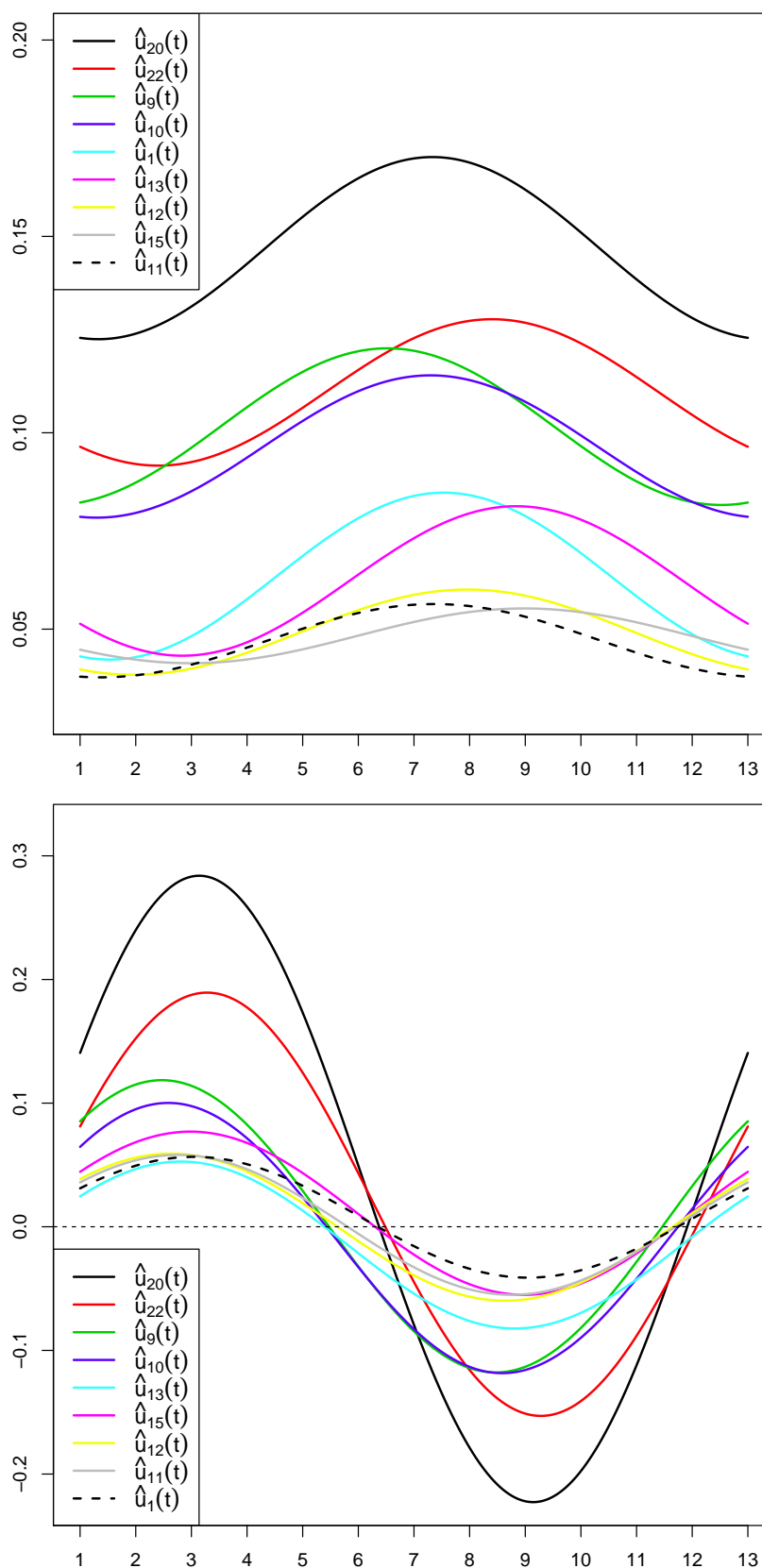
gdzie $k = 1, \dots, 46$, oraz odpowiadające im funkcjonalne składowe główne postaci

$$\hat{U}_k = \langle \hat{\mathbf{u}}_k(t), \mathbf{X}(t) \rangle = \hat{\mathbf{u}}_k' \hat{\mathbf{c}}.$$

Wykresy dziewięciu największych (w sensie pola pod modułem funkcji wagowej) składowych wektorowej funkcji wagowej dla pierwszej i drugiej funkcjonalnej składowej głównej pokazane są odpowiednio na Rys. 5.14.

W określonym punkcie czasowym t , im większa jest wartość bezwzględna składowej wektorowej funkcji wagowej, tym większy jest udział tej składowej w budowie danej funkcjonalnej składowej głównej. Z Rys. 5.14 (góra) widzimy, że największy udział w budowie pierwszej funkcjonalnej składowej głównej przez wszystkie brane pod uwagę momenty obserwacji ma proces X_{20} (deformacja odpowiadająca sile Fr), a jego udział w budowie tej zmiennej wynosi 15%. Proces ten ma również największy wpływ na budowę drugiej funkcjonalnej składowej głównej przez większość badanych momentów czasowych, z udziałem na poziomie 22%. Zauważmy, że dziewięć z dwudziestu dwóch wybranych procesów, tj. siła penetracji (X_1), twardość II (X_9), praca deformacji I (X_{10}), praca deformacji II (X_{11}), spoistość (X_{12}), współczynnik sprężystości w teście podwójnego ściskania (X_{13}), deformację względną odpowiadającą sile Fz (X_{15}), deformacja odpowiadająca sile Fr (X_{20}), moduł sprężystości w teście zrywania (X_{22}), ma łącznie 76% udział w budowie pierwszej funkcjonalnej składowej głównej oraz 81% udział w budowie drugiej funkcjonalnej składowej głównej. Żaden z tych dziewięciu procesów nie jest uzyskany przy pomocy testu przebijania skórki penetrometrem o średnicy 8 mm (Metodą M2). Ponadto udział każdego z procesów, opisujących współczynnik sprężystości w teście penetracji (X_2), współczynnik sprężystości w teście przebicia skórki (X_6), pracę potrzebną do przebicia skórki (X_7), twardość I (X_8), deformację względną odpowiadającą sile Fz (X_{16}), sprężystość (X_{17}), moduł sprężystości mięszu w teście ściskania (X_{18}), siłę zrywającą próbkę (X_{19}) oraz współczynnik sprężystości (X_{21}), jest na poziomie poniżej 3% i to zarówno w przypadku pierwszej jak i drugiej funkcjonalnej składowej głównej. Udział danego procesu pierwotnego $X_i(t)$ w budowie pierwszej i drugiej funkcjonalnej składowej głównej oszacować można za pomocą pola pod modułem funkcji wagowych odpowiadających temu procesowi i opisane to zostało w Rozdziale 2., a wyniki te zawiera Tab. 5.9, przy czym P_i oznacza pole pod krzywą $|u_i(t)|$, a P_i^* dane jest wzorem (2.5).

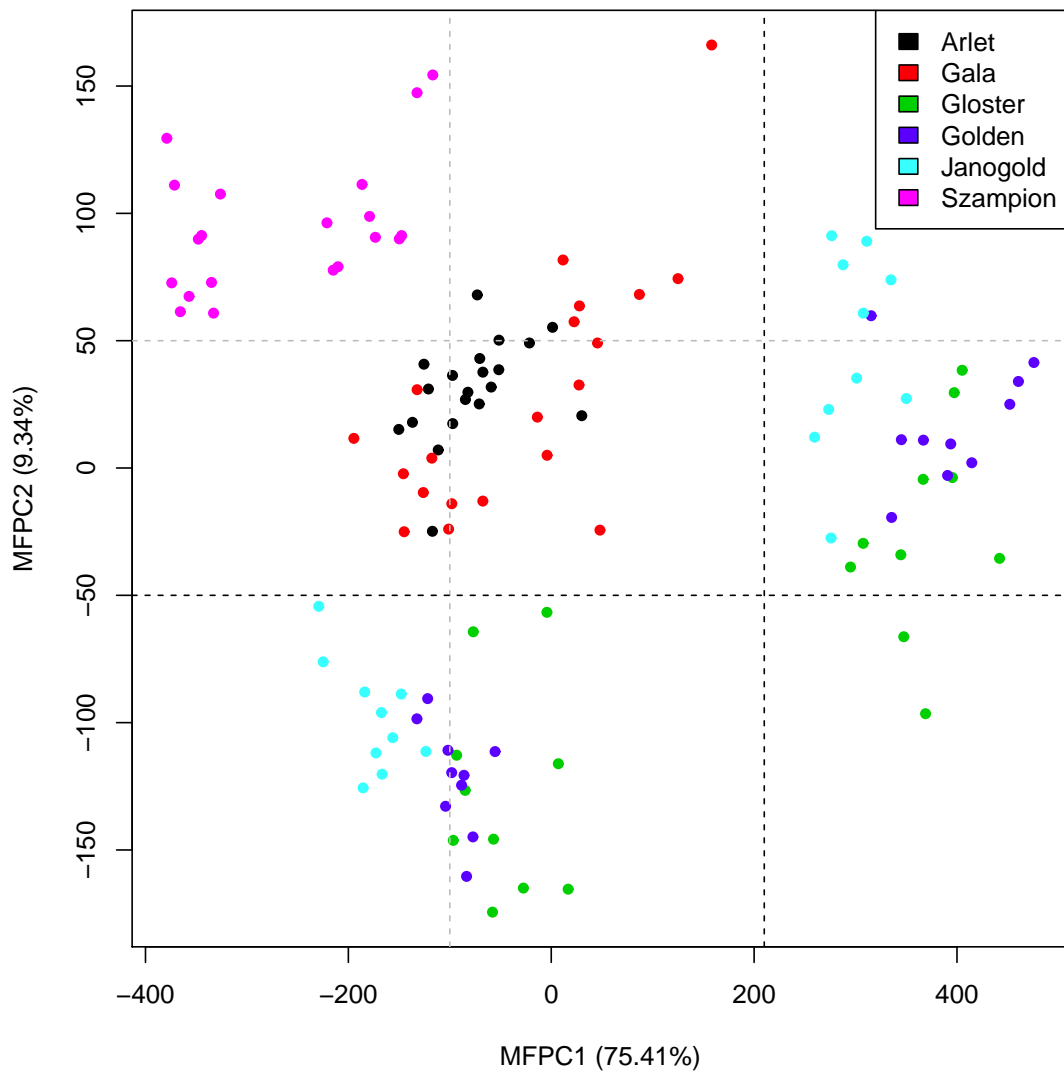
Wzajemne położenie 120 jabłek układzie dwóch pierwszych funkcjonalnych składowych głównych (\hat{U}_1, \hat{U}_2) pokazane jest na Rys. 5.15. Układ dwóch pierwszych funkcjonalnych składowych głównych zachowuje 84,8% zmienności całkowitej. W układzie tym wszystkie jabłka typu Arlet, Gala i Szampion leżą w prostokącie $-400 < U_1 < 210$ oraz $U_2 > -50$. Pozostałe odmiany leżące poza tym prostokątem, tj. gdy $U_1 > 210$ lub $U_2 < -50$, są zwsze odmianami Gloster, Golden lub Janogold. Najtrudniej rozróżnialne są odmiany Gloster i Golden, które osiągają dojrzałość konsumpcyjną na przełomie listopada i grudnia. Natomiast najlepiej odseparowana jest odmiana Szampion, która osiąga dojrzałość konsumpcyjną w połowie października i dla której zawsze zachodzi $U_1 < -100$ oraz $U_2 > 50$.



Rysunek 5.14: Funkcje wagowe dla pierwszej funkcjonalnej składowej głównej \hat{U}_1 (góra). Funkcje wagowe dla drugiej funkcjonalnej składowej głównej \hat{U}_2 (dół).

Tablica 5.9: Pola pod funkcjami wagowymi $u_i(t)$ dla pierwszej i drugiej funkcjonalnej składowej głównej.

Funkcje wagowe $u_i(t)$ dla pierwszej składowej głównej U_1			
i	P_i	P_i^*	ΣP_i^*
20	1,76	15%	15%
22	1,32	12%	27%
9	1,22	11%	38%
10	1,16	10%	48%
1	0,76	7%	55%
13	0,75	7%	61%
12	0,59	5%	66%
15	0,58	5%	71%
11	0,56	5%	76%
14	0,45	4%	80%
3	0,42	4%	84%
4	0,42	4%	88%
5	0,39	3%	91%
19	0,28	2%	94%
18	0,21	2%	95%
2	0,19	2%	97%
7	0,14	1%	98%
21	0,09	1%	99%
8	0,08	1%	100%
6	0,01	0%	100%
16	0,01	0%	100%
17	0,01	0%	100%
Funkcje wagowe $u_i(t)$ dla drugiej funkcjonalnej składowej głównej U_2			
i	P_i	P_i^*	ΣP_i^*
20	1,95	22%	22%
22	1,31	15%	36%
9	0,9	10%	46%
10	0,84	9%	55%
13	0,53	6%	61%
15	0,51	6%	67%
12	0,45	5%	72%
11	0,43	5%	77%
1	0,38	4%	81%
14	0,36	4%	85%
3	0,22	2%	87%
4	0,22	2%	90%
19	0,22	2%	92%
5	0,18	2%	94%
18	0,11	1%	95%
7	0,1	1%	96%
21	0,1	1%	98%
2	0,09	1%	99%
8	0,04	0%	99%
16	0,04	0%	99%
17	0,04	0%	100%
6	0,01	0%	100%



Rysunek 5.15: Rzut 6 odmian jabłek na płaszczyznę dwóch pierwszych składowych głównych (\hat{U}_1, \hat{U}_2) (MFPCA).

5.2.2 Funkcjonalne zmienne dyskryminacyjne

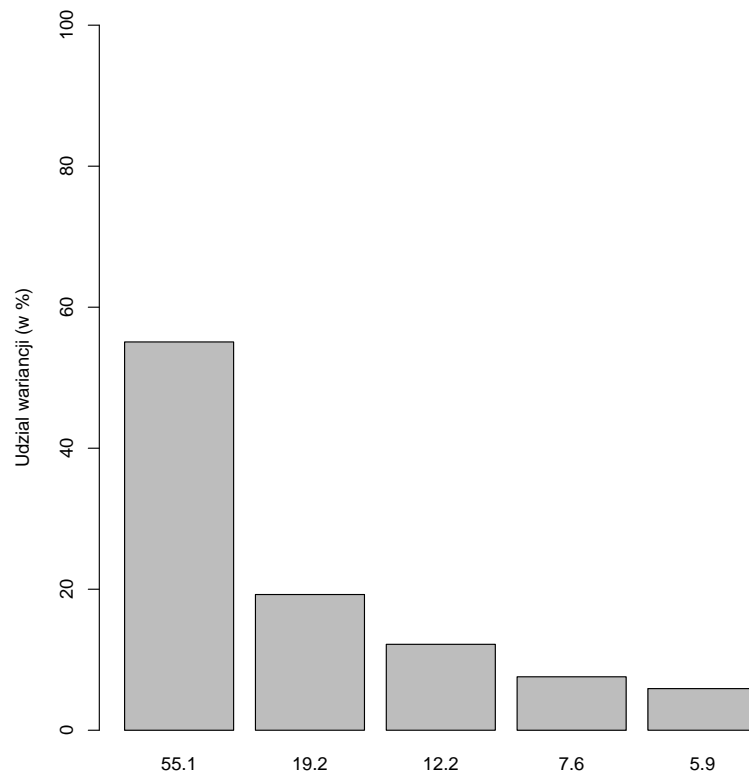
W Tab. 5.10(góra) zaprezentowano wariacje wyjaśnione przez dwie pierwsze zmienne dyskryminacyjne w zależności od kryterium informacyjnego i metody wyboru liczby elementów bazy, a w Tab. 5.10(dół) odpowiadającą im liczbę elementów bazy dla poszczególnych procesów. Analogicznie jak przy funkcjonalnych składowych głównych w prezentowanym przykładzie w zmiennych dyskryminacyjnych, wpływ na wyjaśnienie wariacji ma metoda wyboru liczby elementów bazy bez względu na wybrane kryterium informacyjne. Ponownie ograniczymy się jedynie do kryterium BIC. W tym przypadku dla zmiennych dyskryminacyjnych optymalna liczba elementów bazy B została wybrana przy użyciu metody związanej z dominantą, która pozwala na najlepsze wyjaśnienie wariacji przez dwie pierwsze zmienne i wynosi ona 4 dla procesu X_{12} i X_{20} oraz 2 dla pozostałych procesów $X_i, i = 1, \dots, 11, 13, \dots, 19, 21, 22$.

Tablica 5.10: Wariancje wyjaśnione przez dwie pierwsze funkcjonalne zmienne dyskryminacyjne (górną) oraz liczba elementów bazy dla poszczególnych procesów X_i (dół).

Wariancje wyjaśnione przez dwie pierwsze zmienne dyskryminacyjne						
Kryterium	Metoda	λ_1	$\lambda_1/\Sigma\lambda_i(\%)$	λ_2	$\lambda_2/\Sigma\lambda_i(\%)$	$(\lambda_1 + \lambda_2)/\Sigma\lambda_i (\%)$
BIC	min	$6,32 \times 10$	54,21	$2,19 \times 10$	18,79	72,30
	max	$4,80 \times 10^9$	45,28	$2,19 \times 10^9$	20,70	65,98
	dominanta	$7,33 \times 10$	55,08	$2,56 \times 10$	19,25	74,32
	średnia	$9,56 \times 10$	54,38	$3,52 \times 10$	19,43	73,81
eBIC	min	$6,32 \times 10$	54,21	$2,19 \times 10$	18,79	72,30
	max	$4,80 \times 10^9$	45,28	$2,19 \times 10^9$	20,70	65,98
	dominanta	$7,33 \times 10$	55,08	$2,56 \times 10$	19,25	74,32
	średnia	$9,56 \times 10$	54,38	$3,52 \times 10$	19,43	73,81
AIC	min	$6,32 \times 10$	54,21	$2,19 \times 10$	18,79	72,30
	max	$4,80 \times 10^9$	45,28	$2,19 \times 10^9$	20,70	65,98
	dominanta	$7,33 \times 10$	55,08	$2,56 \times 10$	19,25	74,32
	średnia	$9,56 \times 10$	54,38	$3,52 \times 10$	19,43	73,81

DC														
Kryterium	Metoda	Nr procesu X_i												
		1	2	3,4	5	6,7	8-10	11	12	13,14	15,16	17-19	20	21,22
BIC	min	3	3	3	3	3	3	3	3	3	3	3	3	3
	max	11	11	11	11	11	11	11	11	11	11	9	11	11
	dominanta	3	3	3	3	3	3	3	3	5	3	3	3	5
	średnia	3	5	3	5	5	3	5	5	3	3	3	3	5
eBIC	min	3	3	3	3	3	3	3	3	3	3	3	3	3
	max	11	11	11	11	11	11	11	11	11	11	9	11	11
	dominanta	3	3	3	3	3	3	3	3	5	3	3	3	5
	średnia	3	5	3	5	5	3	5	5	3	3	3	3	5
AIC	min	3	3	3	3	3	3	3	3	3	3	3	3	3
	max	11	11	11	11	11	11	11	11	11	11	9	11	11
	dominanta	3	3	3	3	3	3	3	3	5	3	3	3	5
	średnia	3	5	3	5	5	3	5	5	3	3	3	3	5

W konstrukcji funkcjonalnych zmiennych dyskryminacyjnych, w porównaniu z konstrukcją funkcjonalnych składowych głównych uwzględniamy dodatkowo informację o podziale 120 jabłek na 6 rozłącznych odmian ($L = 6$). Z wektorów $\hat{\mathbf{c}}_i$ budujemy estymator $\hat{\mathbf{B}}$ macierzy zmienności między klasami oraz estymator $\hat{\mathbf{T}}$ macierzy zmienności całkowitej, a następnie znajdujemy niezerowe wartości własne $\hat{\gamma}_k$ macierzy $\hat{\mathbf{T}}^{-1}\hat{\mathbf{B}}$ oraz odpowiadające im wektory $\hat{\mathbf{u}}_k, k = 1, 2, 3, 4, 5$. Stosunki danej wartości własnej do sumy wszystkich wartości własnych wyrażone są w procentach na Rys. 5.16. Z Rys. 5.16 wynika, że 54,2% zmienności całkowitej przypada na pierwszą funkcjonalną zmienną dyskryminacyjną.



Rysunek 5.16: Udział wariancji dla poszczególnych funkcjonalnych zmiennych dyskryminacyjnych (MFDC).

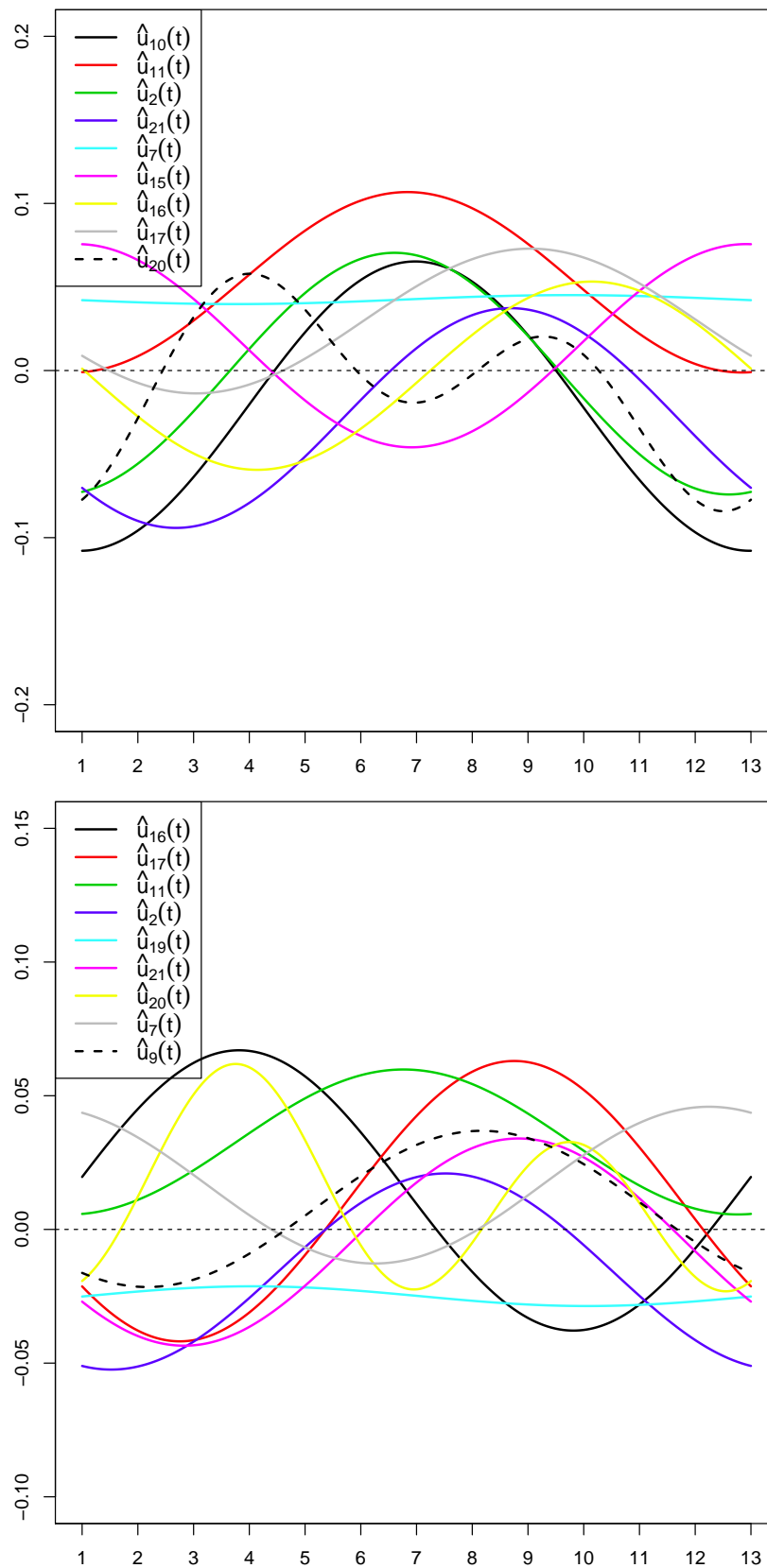
W kroku drugim tworzymy wektorowe funkcje wagowe

$$\hat{\mathbf{u}}_k(t) = \mathbf{\Phi}(t)\hat{\mathbf{u}}_k,$$

gdzie $k = 1, 2, 3, 4, 5$, oraz odpowiadające im funkcjonalne zmienne dyskryminacyjne postaci

$$\hat{U}_k = \langle \hat{\mathbf{u}}_k(t), \mathbf{X}(t) \rangle = \hat{\mathbf{u}}_k' \hat{\mathbf{c}}.$$

Wykresy dziewięciu największych (w sensie pola pod modułem funkcji wagowej) składowych wektorowej funkcji wagowej dla pierwszej i drugiej funkcjonalnej zmiennej dyskryminacyjnej pokazane są na Rys. 5.17.



Rysunek 5.17: Funkcje wagowe dla pierwszej funkcjonalnej zmiennej dyskryminacyjnej \hat{U}_1 (góra). Funkcje wagowe dla drugiej funkcjonalnej zmiennej dyskryminacyjnej \hat{U}_2 (dół).

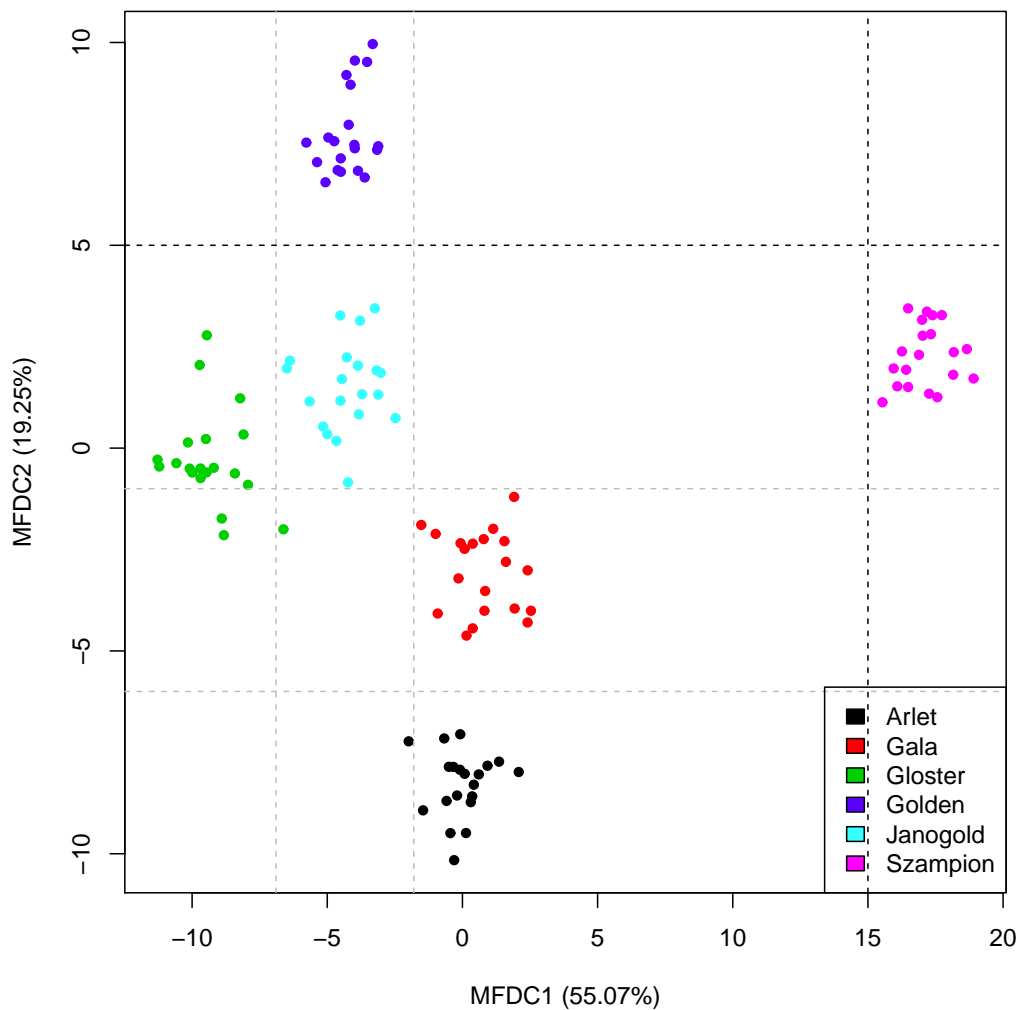
W określonym punkcie czasowym t , im większa jest wartość bezwzględna składowej wektorowej funkcji wagowej, tym większy jest udział tej składowej w budowie danej funkcjonalnej zmiennej dyskryminacyjnej. Z Rys. 5.17 (góra) widzimy, że w przypadku funkcjonalnych zmiennych dyskryminacyjnych żadna funkcja wagowa nie dominuje na całym przedziale czasowym tak jak to miało miejsce w przypadku funkcjonalnych składowych głównych. W budowie pierwszej funkcjonalnej zmiennej dyskryminacyjnej 63% udziału mają łącznie następujące procesy: współczynnik sprężystości w teście penetracji (X_2), praca potrzebna do przebicia skórki (X_7), praca deformacji I (X_{10}), praca deformacji II (X_{11}), deformację względną odpowiadającą sile Fz (X_{15}), deformację względną odpowiadającą sile Fz (X_{16}), sprężystość (X_{17}), deformacja odpowiadająca sile Fr (X_{20}), współczynnik sprężystości (X_{21}). Natomiast w budowie drugiej funkcjonalnej zmiennej dyskryminacyjnej 62% udział mają łącznie następujące procesy: współczynnik sprężystości w teście penetracji (X_2), praca potrzebna do przebicia skórki (X_7), twardość II (X_9), praca deformacji II (X_{11}), deformację względną odpowiadającą sile Fz (X_{16}), sprężystość (X_{17}), siła zrywająca próbkę (X_{19}), deformacja odpowiadająca sile Fr (X_{20}), współczynnik sprężystości (X_{21}). Warto zauważyć, że udział żadnego procesu w budowie dwóch pierwszych funkcjonalnych zmiennych dyskryminacyjnych nie przekracza 10%, co oznacza że większość procesów ma istotny udział w wyjaśnieniu zmienności danych pierwotnych przez te zmienne. Udział danego procesu pierwotnego $X_i(t)$ w budowie pierwszej i drugiej funkcjonalnej zmiennej dyskryminacyjnej ponownie oszacować można za pomocą pola pod modułem funkcji wagowych odpowiadających temu procesowi i opisane to zostało w Rozdziale 2., a wyniki te zawiera Tab. 5.11, przy czym P_i oznacza pole pod krzywą $|u_i(t)|$, a P_i^* dane jest wzorem (2.5).

Wzajemne położenie 120 jabłek układzie dwóch pierwszych funkcjonalnych zmiennych dyskryminacyjnych (\hat{U}_1, \hat{U}_2) przedstawione jest na Rys. 5.18. Układ dwóch pierwszych funkcjonalnych zmiennych dyskryminacyjnych zachowuje 74,3% zmienności całkowitej. W porównaniu z rzutem na dwie pierwsze funkcjonalne składowe główne, podział na 6 jednorodnych grupy jest tutaj wyraźnie widoczny. Wszystkie jabłka dla których pierwsza funkcjonalna zmienna dyskryminacyjna $U_1 > 15$ należą do gatunku Szampion. Gdy druga funkcjonalna zmienna dyskryminacyjna $U_2 > 5$, to należą do gatunku Golden, natomiast jeżeli $U_2 < -6$ to mamy do czynienia z gatunkiem Arlet. Używając dwóch zmiennych dyskryminacyjnych możemy również wydzielić trzy pozostałe gatunki. A mianowicie, gdy $-6,9 < U_1 < -1,8$ oraz $-6 < U_2 < -1$ to obserwacje znajdują w klasie Janogold, zaś gdy $-1,8 < U_1 < 15$ oraz $-1 < U_2 < 5$ to w klasie Gala. Dla gatunku Gloster zachodzą warunki $U_1 < -6$ oraz $-6,9 < U_2 < 5$, przy czym jedna obserwacja jako jedyna z wszystkich leży w sąsiednim prostokącie ($-6,9 < U_1 < -1,8$ oraz $-6 < U_2 < -1$).

Tablica 5.11: Pola pod funkcjami wagowymi $u_i(t)$ dla pierwszej i drugiej funkcjonalnej zmiennej dyskryminacyjnej.

Funkcje wagowe $u_i(t)$ dla pierwszej funkcjonalnej zmiennej dyskryminacyjnej U_1			
i	P_i	P_i^*	ΣP_i^*
10	0,68	9%	9%
11	0,64	9%	18%
2	0,55	7%	25%
21	0,55	7%	33%
7	0,51	7%	40%
15	0,48	6%	46%
16	0,43	6%	52%
17	0,41	6%	58%
20	0,4	5%	63%
22	0,37	5%	68%
6	0,34	5%	73%
19	0,34	5%	77%
12	0,29	4%	81%
14	0,29	4%	85%
3	0,19	3%	88%
4	0,19	3%	90%
8	0,17	2%	92%
18	0,17	2%	95%
13	0,15	2%	97%
5	0,12	2%	98%
1	0,06	1%	99%
9	0,06	1%	100%

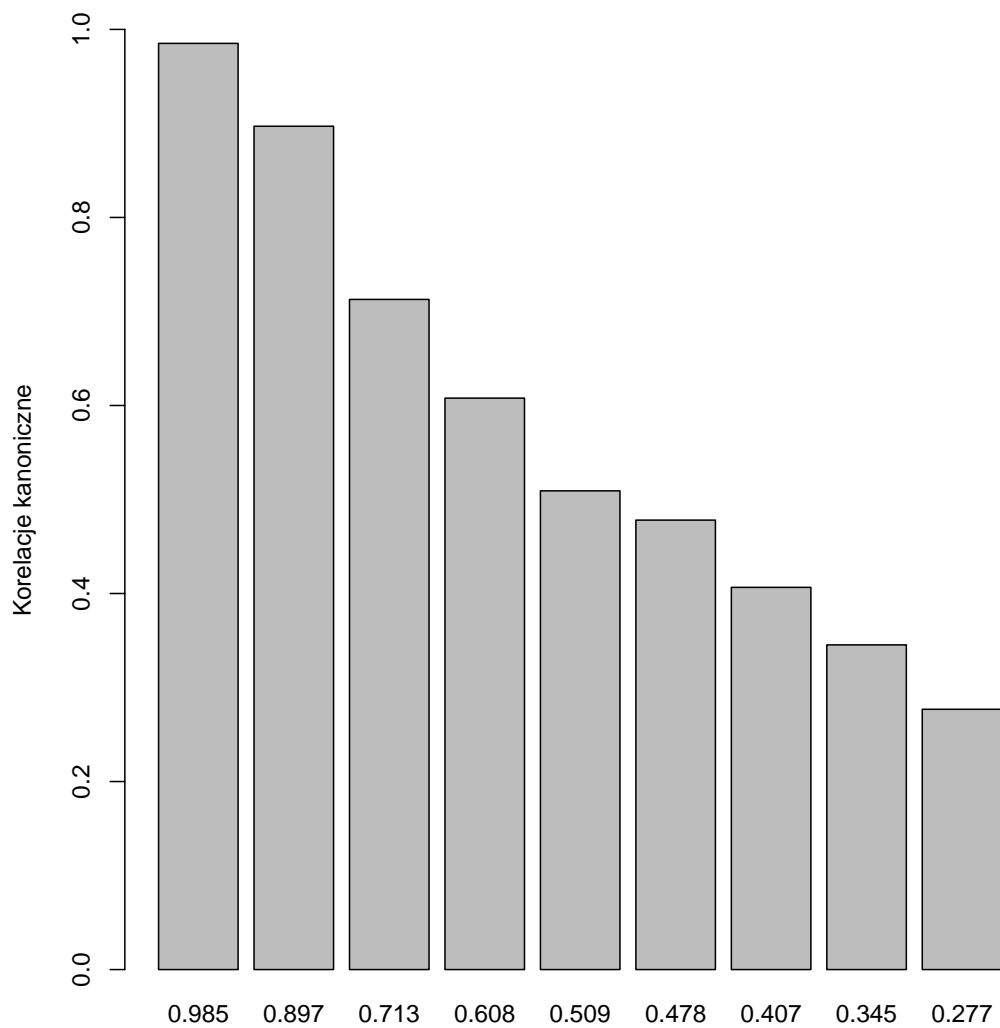
Funkcje wagowe $u_i(t)$ dla drugiej funkcjonalnej zmiennej dyskryminacyjnej U_2			
i	P_i	P_i^*	ΣP_i^*
16	0,42	9%	9%
17	0,41	9%	18%
11	0,39	8%	26%
2	0,31	7%	32%
19	0,3	6%	39%
21	0,3	6%	45%
20	0,29	6%	51%
7	0,26	6%	57%
9	0,23	5%	62%
15	0,22	5%	66%
8	0,21	4%	71%
10	0,18	4%	75%
22	0,18	4%	79%
12	0,17	4%	82%
1	0,14	3%	85%
3	0,14	3%	88%
4	0,14	3%	91%
14	0,13	3%	94%
5	0,08	2%	96%
13	0,08	2%	97%
6	0,07	1%	99%
18	0,06	1%	100%



Rysunek 5.18: Rzut 6 odmian jabłek na płaszczyznę dwóch pierwszych zmiennych dyskryminacyjnych (\hat{U}_1, \hat{U}_2) (MFDC).

5.2.3 Funkcjonalne zmienne kanoniczne

W przypadku funkcjonalnych zmiennych kanonicznych zaprezentujemy konstrukcje dla dwóch wielowymiarowych procesów stochastycznych. Przeprowadzono obliczenia w zakresie wszystkich par zestawów cech uzyskanych różnymi metodami (łącznie $\binom{5}{2} = 10$ par). W dalszej części zaprezentujemy tylko zależności między cechami mierzonymi w metodzie 3 i 1 z uwagi na to, że pierwszy współczynnik korelacji kanonicznej dla tej pary był największy spośród wszystkich uzyskanych. Mianowicie będziemy badać zależność między procesami $\mathbf{Y}(t) = (X_8(t), X_9(t), X_{10}(t), X_{11}(t), X_{12}(t), X_{13}(t))'$ oraz $\mathbf{X}(t) = (X_1(t), X_2(t), X_3(t))'$. W naszym przypadku $p = 5$, a $q = 3$. Budujemy estymatory macierzy $\hat{\Sigma}_{11}, \hat{\Sigma}_{22}$ i $\hat{\Sigma}_{12}$, a następnie znajdujemy niezerowe wartości własne $\hat{\gamma}_k^2$ i odpowiadające im wektory $\hat{\mathbf{u}}_k$ macierzy $\hat{\mathbf{C}}\hat{\mathbf{D}}$ oraz wartości własne $\hat{\gamma}_k^2$ i odpowiadające im wektory $\hat{\mathbf{v}}_k$ macierzy $\hat{\mathbf{D}}\hat{\mathbf{C}}$, gdzie $\hat{\mathbf{C}} = \hat{\Sigma}_{11}^{-1}\hat{\Sigma}_{12}$ oraz $\hat{\mathbf{D}} = \hat{\Sigma}_{22}^{-1}\hat{\Sigma}_{21}, \hat{\Sigma}_{21} = \hat{\Sigma}_{12}'$, $k = 1, \dots, 9$.



Rysunek 5.19: Współczynniki korelacji kanonicznych (MFCCA).

W Tab. 5.12 zaprezentowano dwa pierwsze współczynniki korelacji oraz odpowiadającą im liczbę elementów bazy B dla poszczególnych procesów X_i w zależności od kryterium informacyjnego i metody wyboru liczby elementów bazy. Podobnie jak poprzednio, współczynniki korelacji zależą jedynie od metody wyboru liczby elementów bazy bez względu na wybrane kryterium informacyjne. Optymalna liczba elementów bazy B została wybrana przy użyciu kryterium BIC z metodą związaną z dominantą, analogicznie jak przy funkcjonalnych zmiennych dyskryminacyjnych, i wynosi ona 4 dla procesu X_{12} oraz 2 dla pozostałych procesów $X_i, i = 1, 2, 3, 8, 9, 10, 11, 13$.

Wartości własne $\hat{\gamma}_k$, nazywane korelacjami kanonicznymi, pokazane są na Rys. 5.12. W kroku drugim tworzymy wektorowe funkcje wagowe

$$\hat{\mathbf{u}}_k(t) = \Phi_1(t)\hat{\mathbf{u}}_k, \quad \hat{\mathbf{v}}_k(t) = \Phi_2(t)\hat{\mathbf{v}}_k,$$

odpowiadające procesom $\mathbf{Y}(t)$ oraz $\mathbf{X}(t)$, gdzie $k = 1, \dots, 9$. Funkcjom tym odpowiadają funk-

Tablica 5.12: Dwa pierwsze współczynniki korelacji oraz liczba elementów bazy dla poszczególnych procesów X_i CCA

Dwa pierwsze współczynniki korelacji			
Kryterium	Metoda	ρ_1	ρ_2
BIC	min	0,9850	0,8983
	max	0,9781	0,9143
	dominanta	0,9851	0,8970
	średnia	0,9850	0,8989
eBIC	min	0,9850	0,8983
	max	0,9781	0,9143
	dominanta	0,9851	0,8970
	średnia	0,9850	0,8989
AIC	min	0,9850	0,8983
	max	0,9781	0,9143
	dominanta	0,9851	0,8970
	średnia	0,9850	0,8989

CCA										
Kryterium	Metoda	Nr procesu X_i								
		1	2	3	8	9	10	11	12	13
BIC	min	3	3	3	3	3	3	3	3	3
	max	11	11	11	11	11	11	11	11	11
	dominanta	3	3	3	3	3	3	3	5	3
	średnia	3	5	3	3	3	3	5	5	3
eBIC	min	3	3	3	3	3	3	3	3	3
	max	11	11	11	11	11	11	11	11	11
	dominanta	3	3	3	3	3	3	3	5	3
	średnia	3	5	3	3	3	3	5	5	3
AIC	min	3	3	3	3	3	3	3	3	3
	max	11	11	11	11	11	11	11	11	11
	dominanta	3	3	3	3	3	3	3	5	3
	średnia	3	5	3	3	3	3	5	5	3

cjonalne zmienne kanoniczne postaci

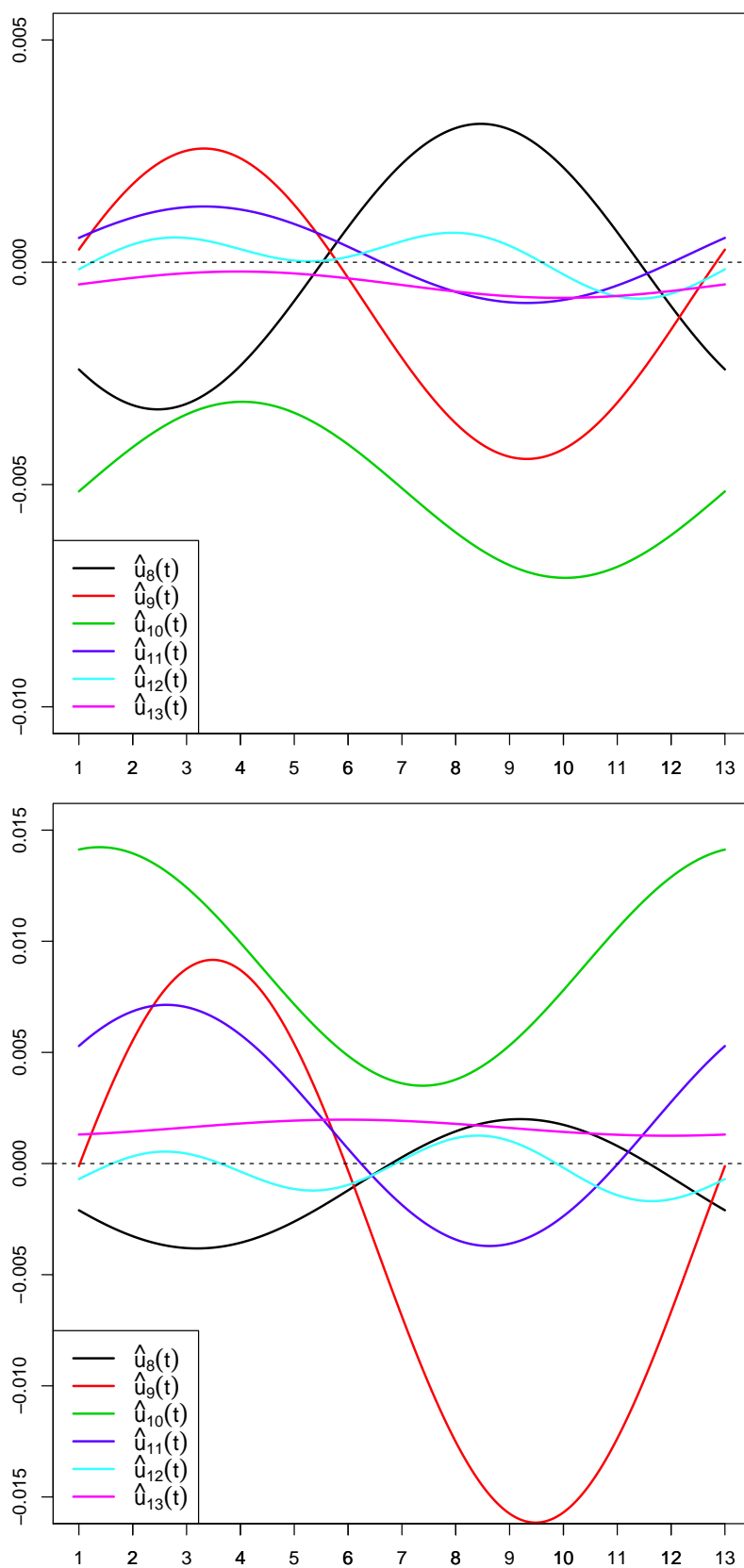
$$\hat{U}_k = \langle \hat{\mathbf{u}}_k(t), \mathbf{Y}(t) \rangle, \quad \hat{V}_k = \langle \hat{\mathbf{v}}_k(t), \mathbf{X}(t) \rangle,$$

odpowiadające procesom $\mathbf{Y}(t)$ oraz $\mathbf{X}(t)$. Wykresy dwóch składowych wektorowej funkcji wagowej dla pierwszej i drugiej funkcjonalnej zmiennej kanonicznej procesów $\mathbf{Y}(t)$ oraz $\mathbf{X}(t)$ pokazane są na Rys. 5.20 oraz 5.21.

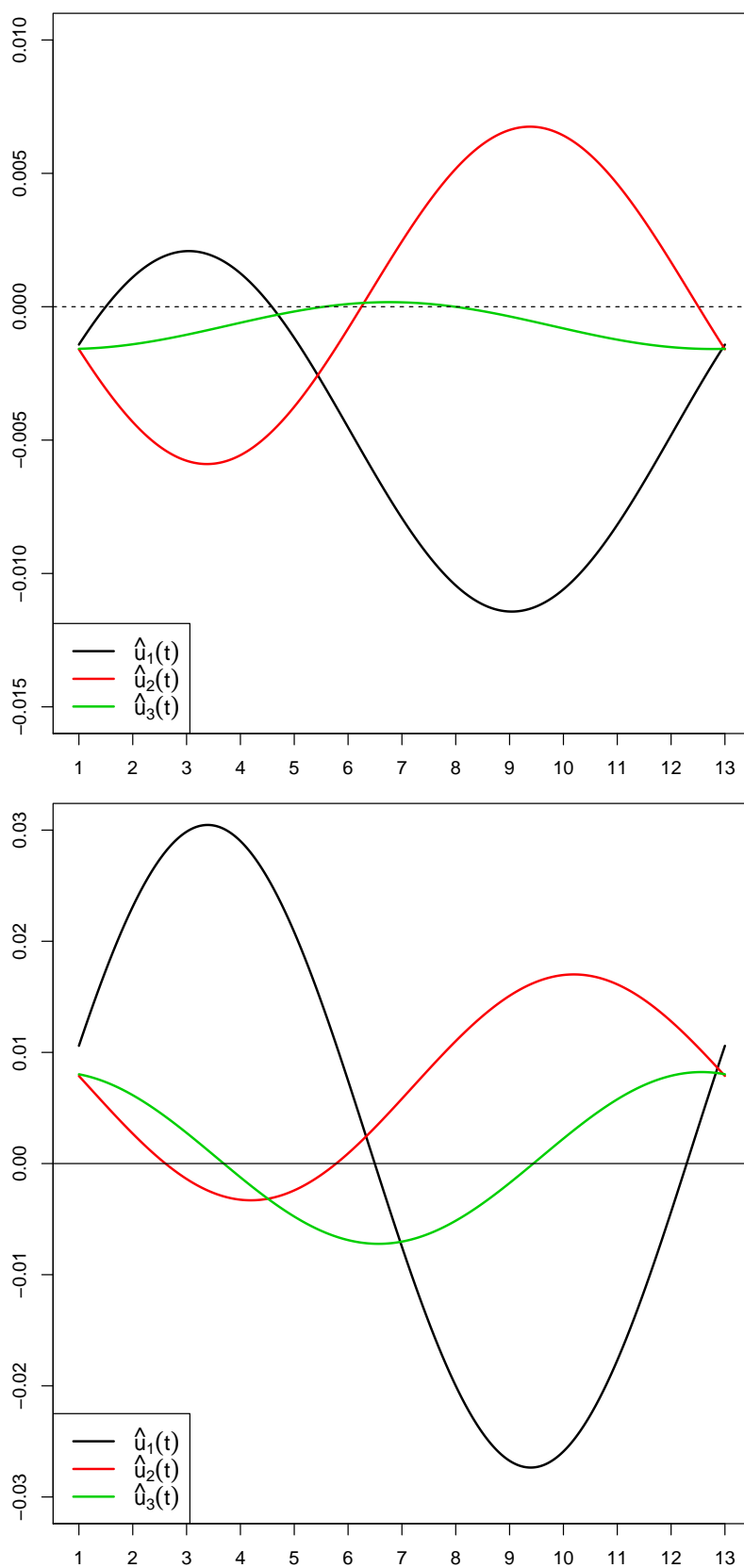
W określonym punkcie czasowym t , im większa jest wartość modułu składowej wektorowej funkcji wagowej, tym większy jest wkład tej składowej w konstrukcję danej funkcjonalnej zmiennej kanonicznej. Z Rys. 5.21 widzimy, że na określonych przedziałach czasowych największy (ponad 50%) udział w budowie zarówno pierwszej V_1 jak i drugiej V_2 funkcjonalnej zmiennej kanonicznej procesu $\mathbf{X}(t)$ ma siła penetracji (X_1). Ponadto siła penetracji (X_1) oraz współczynnik sprężystości w teście penetracji (X_2) łącznie mają powyżej 80% wkład w budowę dwóch pierwszych funkcjonalnych zmiennych kanonicznych procesu $\mathbf{X}(t)$. Natomiast w konstrukcji pierwszej jak i drugiej funkcjonalnej zmiennej kanonicznej procesu $\mathbf{Y}(t)$ największe udziały mają twardość II (X_9) oraz praca deformacji I (X_{10}), które mają łącznie 69% udział w budowie zmiennej U_1 oraz 72% udział w budowie zmiennej U_2 . Należy jeszcze zauważyć, że spoistość (X_{12}) nie wnosi nic do konstrukcji dwóch pierwszych funkcjonalnych zmiennych kanonicznych obydwu przestrzeni wektorowych. Podobnie jak poprzednio, całkowity udział poszczególnych procesów $X_i(t)$ w strukturze poszczególnych zmiennych kanonicznych może być oszacowany z użyciem pola pod modułami funkcji wagowych odpowiadających temu procesowi. Udziały te dla pierwszej i drugiej funkcjonalnej zmiennej kanonicznej procesów wektorowych $\mathbf{Y}(t)$ oraz $\mathbf{X}(t)$ zawiera Tab. 5.13.

Wzajemne położenie 120 jabłek w układach (\hat{U}_1, \hat{U}_2) , (\hat{V}_1, \hat{V}_2) oraz (\hat{U}_1, \hat{V}_1) funkcjonalnych zmiennych kanonicznych pokazane jest na Rys. 5.22 oraz Rys. 5.23.

Proces wektorowy $\mathbf{Y}(t)$ są to cechy badane w metodzie M3, a proces $\mathbf{X}(t)$ to cechy badane w metodzie M1. Interesowała nas siła zależności pomiędzy tymi dwoma grupami procesów. Rys. 5.23 pokazuje rzut na płaszczyznę pierwszej pary funkcjonalnych zmiennych kanonicznych (\hat{U}_1, \hat{V}_1) . Współczynnik korelacji ρ_1 związany z tą parą jest na poziomie 0,98, a więc obserwacje leżą blisko prostej ($y = x$). Można to interpretować jako zastępowalność grupy procesów badaną metodą M1 i M3. Oznacza to w praktyce, że pomiar sześciu procesów mierzonych metodą M3 może zostać zastąpiony trzema procesami mierzonymi metodą M1. Ponadto układy (\hat{U}_1, \hat{U}_2) , (\hat{V}_1, \hat{V}_2) traktować można jako układy dwóch pierwszych funkcjonalnych składowych głównych dla wektorowych procesów $\mathbf{Y}(t)$ oraz $\mathbf{X}(t)$ odpowiednio.



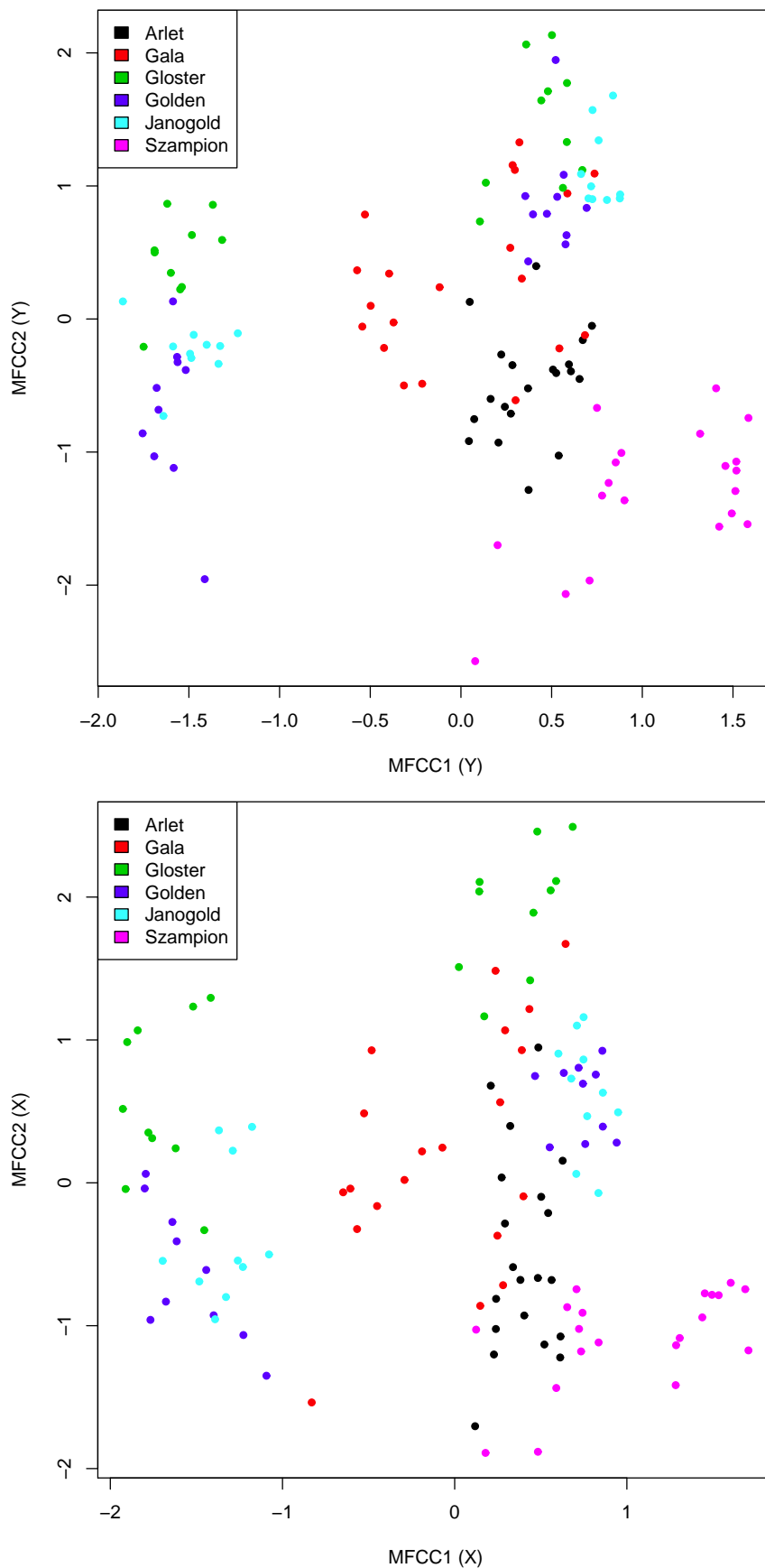
Rysunek 5.20: Funkcje wagowe dla pierwszej funkcjonalnej zmiennej kanonicznej \hat{U}_1 (góra). Funkcje wagowe dla drugiej funkcjonalnej zmiennej kanonicznej \hat{U}_2 (dół).



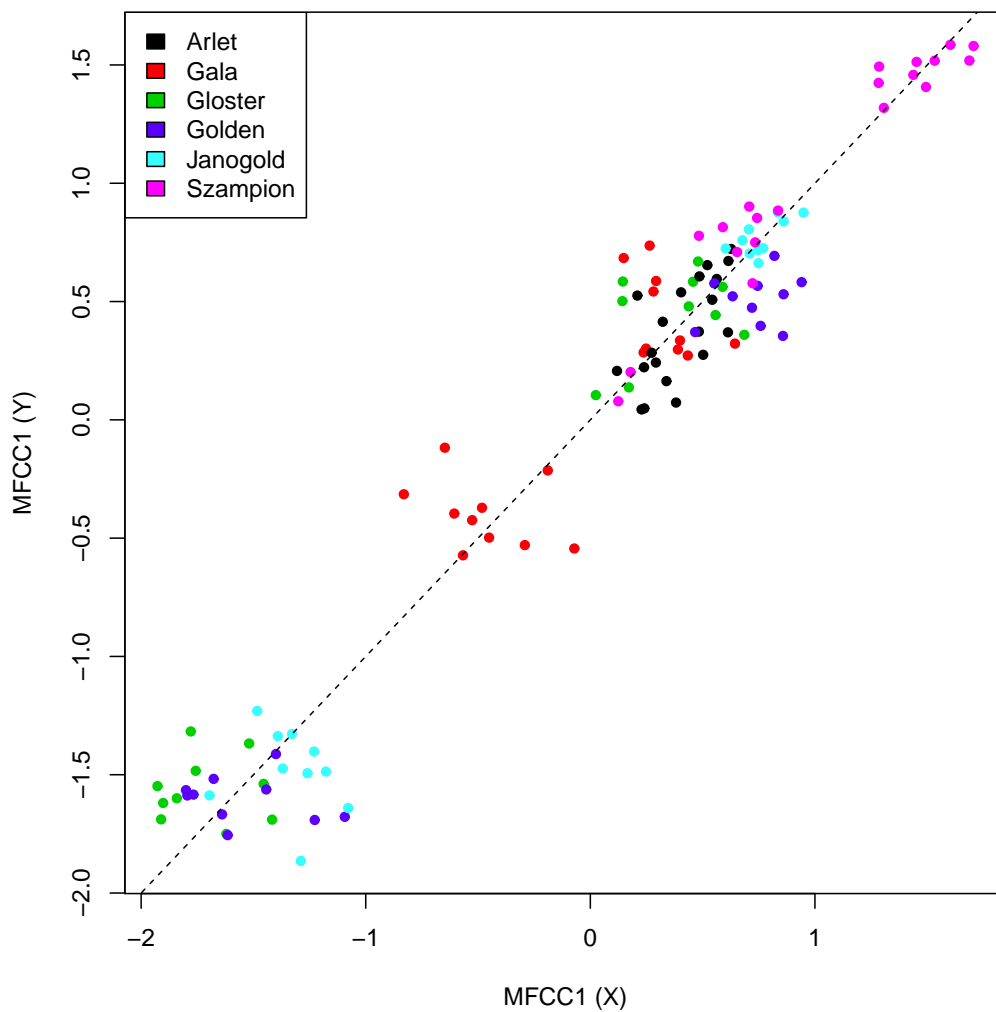
Rysunek 5.21: Funkcje wagowe dla pierwszej funkcjonalnej zmiennej kanonicznej \hat{V}_1 (góra). Funkcje wagowe dla drugiej funkcjonalnej zmiennej kanonicznej \hat{V}_2 (dół).

Tablica 5.13: Pola pod funkcjami wagowymi $u_i(t)$ dla pierwszej i drugiej funkcjonalnej zmiennej kanonicznej.

Funkcje wagowe dla pierwsza funkcjonalnej zmiennej kanonicznej V_1			
i	P_i	P_i^*	ΣP_i^*
1	0,06	51%	51%
2	0,048	41%	92%
3	0,009	8%	100%
Funkcje wagowe dla drugiej funkcjonalnej zmiennej kanonicznej V_2			
i	P_i	P_i^*	ΣP_i^*
1	0,22	59%	59%
2	0,096	26%	84%
3	0,059	16%	100%
Funkcje wagowe dla pierwsza funkcjonalnej zmiennej kanonicznej U_1			
i	P_i	P_i^*	ΣP_i^*
10	0,06	46%	46%
9	0,03	23%	69%
8	0,02	15%	85%
11	0,01	8%	92%
13	0,01	8%	100%
12	0	0%	100%
Funkcje wagowe dla drugiej funkcjonalnej zmiennej kanonicznej U_2			
i	P_i	P_i^*	ΣP_i^*
10	0,11	38%	38%
9	0,1	34%	72%
11	0,04	14%	86%
13	0,02	7%	93%
8	0,02	7%	100%
12	0	0%	100%



Rysunek 5.22: Rzut 6 odmian jabłek na płaszczyznę dwóch pierwszych funkcjonalnych zmiennych kononicznych (\hat{U}_1, \hat{U}_2) przestrzeni $Y(t)$ (górze) oraz (\hat{V}_1, \hat{V}_2) przestrzeni $X(t)$ (dół) (MFCCA).



Rysunek 5.23: Rzut 6 odmian jabłek na płaszczyznę dwóch pierwszych funkcjonalnych zmiennych kononicznych (\hat{U}_1, \hat{V}_1) (MFCCA).

Bibliografia

- [1] Akaike, H., *Information theory and an extension of the maximum likelihood principle*. B. N. Petrov i F. Csaki (red.), Second International Symposium on Information Theory. Akademiai Kiado. Budapest (1973).
- [2] Araki Y., Konishi S., Kawano S., Matsui H., *Functional logistic discrimination via regularized basis expansions*. Communications in Statistics - Theory and Methods 38, 16-17 (2009).
- [3] Berrendero J.R., Justel A., Svarc M., *Principal components for multivariate functional data*. Computational Statistics and Data Analysis 55, 2619 – 2634 (2011).
- [4] Besse P., *Spline functions and optimal metric in linear principal component analysis*. In Component and Correspondence Analysis: Dimensional Reduction by Functional Approximation (eds J.L.A. van Rijkevorseel and J. de Leeuw). Wiley, New York (1988).
- [5] Besse P., Ramsay J.O, *Principal components analysis of sampled functions*. Psychometrika 51, 285-311 (1986).
- [6] Bosq D., Blanke D., *Inference and Prediction in Large Dimensions*. Wiley, Chichester (2007).
- [7] Brillinger D. R., *Time Series: Data Analysis and Theory, Second Edition*. Holden Day, San Francisco (1975).
- [8] Cattell R.B., *The scree test for the number of factors*. Multivariate Behavior Research 1, 245-276 (1966).
- [9] Clarkson D.B., Fraley C., Gu C.C, Ramsay J.O, *S+ Functional Data Analysis User's Manual for Windows*. Springer Verlag, New York (2005).
- [10] Cuevas A., *A partial overview of the theory of statistics with functional data*. Journal of Statistical Planning and Inference 147, 1-23 (2014).

-
- [11] Chen J., Chen. Zehua, , *Extended Bayesian Information Criteria for Model Selection with Large Model Spaces*. Biometrika Vol. 9, 759-771 (2008).
- [12] Dabo-Niang S, Ferraty F., *Functional and Operational Statistics*. Physica-Verlag, Heidelberg (2005).
- [13] Dauxois J., Pousse A., *Les analyses factorielles en calcul des probabilités et en statistique: essai d'étude synthétique. DSc Thesis*. University of Toulouse, Toulouse (1976).
- [14] Dauxois J., Pousse A., Romain Y., *Asymptotic theory for the principal component analysis of a vector random function:some applications to statistical inference*. Journal of Multivariate Analysis 12, 136–154 (1982).
- [15] Delsol L.,Ferraty F.,Martínez-Calvo A., *Functional data analysis: an interdisciplinary statistical topic*. In: Gettler-Summa M., Bottou L., Goldfarb B., Murtagh F., Pardoux C., Touati M.(Eds.). Chapmanand Hall, CRC, Boca Raton. Statistical Learning and Data Science, 189–195 (2011).
- [16] Eubank R.L, Hsing T., *Canonical correlation for stochastic processes*. Stochastic Processes and their Applications 118 (9), 1634-1661 (2008).
- [17] Febrero-Bande M., Oviedo de la Fuente M.,*Statistical computing in functional data analysis: the R package fda.usc*. Journal of Statistical Software 51, 1–28 (2012).
- [18] Ferraty F., Romain Y, (Eds.), *The Oxford Handbook of Functional Data Analysis*. Oxford University Press, Oxford (2011).
- [19] Ferraty F., Vieu P., *Curves discrimination: a nonparametric functional approach*. Computational Statistics and Data Analysis 44, 161 – 173 (2003).
- [20] Ferraty F., Vieu P., *Nonparametric Functional Data Analysis, Theory and Practice*. Springer, New York (2006).
- [21] Fisher R.A., *The use of multiple measurements in taxonomic problem.*, *Annals of Eugenics* 7, 179-188 (1936).
- [22] González-Manteiga W., Vieu P., *Methodological richness of functional data analysis*. In: Gettler-Summa M., Bottou L.,Goldfarb B., Murtagh F., Pardoux C., Touati M.(Eds.). Chapmanand Hall, CRC. Statistical Learning and Data Science, 197–203 (2011).
- [23] Górecki T., Krzyśko M., *Functional principal components analysis*. In: Data Analysis Methods and Its Applications (J. Pocięcha and R. Decker, eds), Wydawnictwo C.H. Beck, Warszawa, 71-87 (2012a).

- [24] Górecki T., Krzyśko M., *A kernel version of functional principal components*. Statistics in Transition-new series 13(3), 559-568 (2012b).
- [25] Górecki T., Krzyśko M., Waszak., Ł., *Functional discriminant coordinates*. Communications in Statistics - Theory and Methods 43(5), 1013-1025 (2014).
- [26] Górecki T., Krzyśko M., Wołyński. W., *Classification problem based in regression models for multidimensional functional data.*, Statistics in Transition new series Vol. 16 No. 1, 97-110 (2015).
- [27] Hannan E.J., *The general theory of canonical correlation and its relation to functional analysis*. Journal of the Australian Mathematical Society 21, 229-242 (1961).
- [28] He G., Muller H.G, Wang J.L , *Extending correlation and regression from multivariate to functional data*. In: Asymptotics in Statistics and Probability (M.L. Puri (Ed.)), 2000 VSP, 1-14 (2000).
- [29] He G., Muller H.G, Wang J.L , *Functional canonical analysis for square integrable stochastic processes*. Journal of the Multivariate Analysis (2002).
- [30] He G., Muller H.G, Wang J.L , *Methods of canonical analysis for functional data*. Journal of Statistical Planning and Inference 122, 141-159 (2004).
- [31] Horváth L.,Kokoszka P., *Inference for Functional Data with Applications*. Springer, New York (2012).
- [32] Hotelling H., *Relations between two sets of variates*. Biometrika 28, 321-377 (1936).
- [33] James G.M., Hastie T.J., *Functional linear discriminant analysis for irregularly sampled curves*. Journal of the Royal Statistical Society B 63, 533–550 (2001).
- [34] Krzyśko M., *Analiza dyskryminacyjna*. Wydawnictwo Naukowo-Techniczne, Warszawa (1990).
- [35] Krzyśko M., *Podstawy wielowymiarowego wnioskowania statystycznego*. Wydawnictwo Naukowe UAM, Poznań (2009).
- [36] Krzyśko M., Derejko A., Górecki T., Gacek E., *Principal component analysis for functional data on grain yield of winter wheat cultivars*. Biometrical Letters 50(2), 81-94 (2013).
- [37] Krzyśko M., Waszak. Ł., *Functional canonical correlation analysis for functional data*. Biometrical Letters 50(2), 95-105 (2013).

- [38] Krzyśko M., Waszak Ł., Wołyński W., *Comparison of kernel discriminant coordinates algorithms*. Communications in Statistics - Simulation and Computation 43(9), 2138-2148 (2014).
- [39] Leurgans S.E., Moyeed R.A., Silverman B.W., *Canonical correlation analysis when the data are curves*. Journal of the Royal Statistical Society Series B 55, 725-740 (1993).
- [40] Muller H.-G., *Functional modelling and classification of longitudinal data*. Scandinavian Journal of Statistics 32, 223-240 (2005).
- [41] Nadulski R., *Metodyczne aspekty badań mechanicznych właściwości tekstury jabłek*. Wydawnictwo Akademii Rolniczej w Lublinie, Lublin (2005).
- [42] Pearson K., *On lines and planes of closest fit to systems of points in space*. Philosophical Magazine 2, 559-572 (1901).
- [43] Preda C., Saporta G., Leveder C., *PLS classification of functional data*. Computational Statistics 22, 223-235 (2007).
- [44] Ramsay J.O., *Functional data analysis*. In: S. Kotz, C.B. Read and D.L. Danks (Eds.), Encyclopedia of Statistical Sciences. Wiley, New York (1997).
- [45] Ramsay J. O., Dalzell C. J., *Some tools for functional data analysis*. Journal of the Royal Statistical Society Series B (53), 539-572 (1991).
- [46] Ramsay J.O, Hooker G, Graves S., *Functional Data Analysis with R and MATLAB, Use R*. Springer Science + Business Media (2009).
- [47] Ramsay J.O., Silverman B.W, *Functional Data Analysis*. Springer, New York (1997).
- [48] Ramsay J.O., Silverman B.W, *Functional Data Analysis*, Second Edition. Springer, New York (2005).
- [49] Ramsay J.O., Silverman B.W, *Applied Functional Data Analysis*. Springer, New York (2002).
- [50] Rice J. A., *Functional and longitudinal data analysis: perspectives on smoothing*. Statistica Sinica 14, 631-647 (2004).
- [51] Rossi F. Villa N., *Support vector machine for functional data classification*. Neurocomputing 69, 730-742 (2006).
- [52] Service S.K., Rice J.A., Chavez F.P., *Relationship between physical and biological variables during the upwelling period in Monterey Bay, CA*. Deep-Sea Research II, 1669-1685 (1998).

-
- [53] Shin H., *An extension of Fisher's discriminant analysis for stochastic processes*. Journal of Multivariate Analysis 99, 1191-1216 (2008).
- [54] Shmueli G., *To explain or to predict?*. Statistical Science 25(3), 289-310 (2010).
- [55] Schwarz G., *Estimating the dimension of a model*. Annals of Statistics 6 (2), 461-464 (1978).