Ryszard Jasiński

METODY PRZESTRZENI LINIOWEJ W EKONOMETRII

praca doktorska

Promotor doc. dr Władysław Bukietyński

W Y Ż S Z A   S Z K O Ł A   E K O N O M I C Z N A

W r o c ł a w   1973

S P I S   T R E Ś C I

str.

str.

W S T Ę P

J. G. Kemeny w swojej pracy [21] napisał: "Nauka rozwiązała problem odpowiedniego języka. Znalazła ona język doskonale precyzyjny i dość bogaty aby sprostać wszystkim jej problemom. Językiem tym jest matematyka".

Formułowanie problemów ekonomii w matematycznej postaci odbywa się najczęściej za pośrednictwem ekonometrii, która wykorzystuje w swoich badaniach głównie statystykę matematyczną i algebrę liniową.

Następujące względy pozwalają sprowadzić zasadnicze problemy ekonometrii do przestrzeni liniowych. Cechy obiektów ekonomicznych z racji wielu wpływających na nie czynników niezdeterminowanych mają charakter losowy, traktuje się je przeto jako zmienne losowe, a jak wiadomo - dostatecznie obszerna klasa zmiennych losowych tworzy pewien typ przestrzeni liniowej. Inną przyczyną pozwalającą ograniczyć metody matematyczne stosowane w ekonometrii do przestrzeni liniowych jest fakt, że zmienne losowe - cechy obiektów ekonomicznych są najczęściej dane jako ciągi liczb będących ich realizacjami. Ciągi te można również traktować jako elementy pewnej przestrzeni liniowej.

Wybór metod przestrzeni liniowych nie przesądza indeterministycznego charakteru zjawisk ekonomicznych. Zawsze bowiem można uznać, że przestrzeń, w której cechy są reprezentowane przez wektory, jest po prostu skończenie wymiarową przestrzenią liniową, nie zakładając probabilistycznego "rodowodu" tych wektorów.

Ograniczenie się do metod przestrzeni liniowych implikuje skoncentrowanie uwagi na prostych zależnościach liniowych. Postępowanie to daje się uzasadnić pragmatycznie. Takie modelowanie zagadnień ekonomicznych pozwala na łatwe wyciąganie wniosków ich dotyczących. Powstaje jednak wtedy problem adekwatności tego prostego aparatu formalnego do rzeczywistości ekonomicznej. Jako argumentu za nim przemawiającego można by użyć zasady W. Ockhama[1/], głoszącej unikanie niepotrzebnego komplikowania i mnożenia bytów wyjaśniających fakty empiryczne. Ponadto znany jest pogląd A. Einsteina[2/], który twierdził, że Stwórca nie mógł pominąć okazji uczynienia prostymi związków występujących w rzeczywistości.

Tematyka pracy wiąże się ściśle z badaniami prowadzonymi w Instytucie Metod Rachunku Ekonomicznego WSE we Wrocławiu. Realizacja podjętego tematu ma następujący przebieg. W pierwszym etapie /rozdziały I, II, III/ została przedstawiona teoria przestrzeni liniowych ze szczególnym uwzględnieniem tak zwanej przestrzeni Hilberta. W przestrzeni tej zdefiniowane zostały zasadnicze pojęcia statystyki matematycznej używane w ekonometrii i zbadane główne własności ich charakterystyk. Tym samym, pojęciom owym została dana interpretacja geometryczna. Umożliwiło to uproszczenie dowodów znanych twierdzeń i osiągnięcie wielu nowych

1/ Porównaj: W. Tatarkiewicz [36] .
2/ Porównaj: J.G. Kemeny [21]

wyników dotyczących ważnych problemów ekonometrii, przy czym niektóre z nich są nowymi spostrzeżeniami w teorii przestrzeni Hilberta, ale formułuje się je często w języku pojęć statystyki matematycznej aby ułatwić wyciągnięcie wniosków istotnych w ekonometrii. Przedstawione w trzech pierwszych rozdziałach wyniki zostały wykorzystane w rozdziale V, w którym omówiony został problem zależności liniowej między cechami obiektów ekonomicznych i zaproponowany sposób wyboru cech do modelu liniowego, uwzględniający między innymi nierozważany dotąd w literaturze problem ich niezmienniczości.

Inne zastosowanie przestrzeni liniowych obejmuje prezentowaną w rozdziale IV metodę wyboru cech najlepiej charakteryzujących obiekty ekonomiczne. W metodzie tej eliminowane są cechy powtarzające informacje i cechy, które słabo separują obiekty, co również nie było dotychczas uwzględniane.

Kontynuacją rozważań rozdziału IV jest rozdział VI. Wybierając bowiem cechy najlepiej charakteryzujące obiekty ustalamy w pewnym sensie "optymalną" przestrzeń liniową, w której z kolei obiekty są reprezentowane jako wektory. Takie odwzorowanie zbioru obiektów pozwala na badanie relacji między nimi, np.: wybór obiektu najlepszego - wzorca, jak również wybór podzbiorów obiektów do siebie podobnych, co ma zastosowanie między innymi w programowaniu matematycznym. Rodział VI poza omówieniem znanych i nowych rezultatów dotyczących wspomnianych relacji, zawiera nowe propozycje ich zastosowań do rozwiązywania zagadnień pojawiających się w aktualnie prowadzonej polityce gospodarczej.

Ponadto w rozdziale tym jest przedstawiony problem wzajemnych związków przestrzeni cech i przestrzeni obiektów, a także związek taksonomii ze statystyką, umożliwiający wykorzystanie metod pierwszej w zagadnieniach dotyczących cech - zmiennych losowych.

Niektóre z poruszonych w pracy problemów zostały zbyt lapidarnie potraktowane. Wymagają one odrębnego opracowania i są przedmiotem aktualnie prowadzonych przez autora uporczywych badań. Innego rodzaju trudności, które napotkał autor są natury redakcyjnej: w oznaczeniach nie udało się uniknąć niekonsekwencji spowodowanych najczęściej rozbieżnościami tradycji ponujących w teorii przestrzeni liniowych i w ekonometrii - dotyczy to zwłaszcza oznaczania zmiennych losowych i ich realizacji.

R o z d z i a ł I

DEFINICJE I TWIERDZENIA PODSTAWOWE

W rozdziale tym przedstawimy definicje pojęć i twierdzenia dotyczące rachunku prawdopodobieństwa, algebry liniowej i przestrzeni Hilberta, z których będziemy korzystali w pracy.

Tylko nieliczne z przytoczonych faktów są mniej znane. Znalazły się jednak tutaj aby ułatwić czytanie pracy; ponadto zdarza się, że w sformułowaniu nawet popularnych pojęć występują w literaturze pewne rozbieżności co stwarza konieczność podania ich w takiej postaci w jakiej będą używane w ciągu dalszym niniejszego tekstu. Liczne powołania i odnośniki pozwalają znaleźć sformułowania i dowody cytowanych twierdzeń.

§ 1. Rachunek prawdopodobieństwa i statystyka.

Niech $\langle \Omega ; \mathcal{L} ; P \rangle$ oznacza przestrzeń probabilistyczną, gdzie

Ω - zbiór zdarzeń elementarnych,
$\mathcal{L}$ - zbiór zdarzeń losowych,
P - prawdopodobieństwo.

/1/ Definicja

Zmienną losową o wartościach rzeczywistych nazywamy funkcję x rzeczywistą taką, że dla każdego rzeczywistego t zbiory postaci

$$\{u : x(u) \leqslant t\}$$

należą do $\mathcal{B}$

/2/ <u>Definicja</u>

Mówimy, że pewna własność zachodzi prawie wszędzie, gdy prawdopodobieństwo zdarzenia losowego, dla którego własność ta nie zachodzi, jest równe zero.
Na przykład, jeśli dwie zmienne losowe są sobie równe, znaczy to tyle, że prawdopodobieństwo zdarzenia losowego, na którym funkcje te przyjmują różne wartości jest równe zero. Wszystkie relacje między zmiennymi losowymi i własności zmiennych losowych należy rozumieć jako zachodzące prawie wszędzie. W ciągu dalszym pracy nie będziemy tego faktu odnotowywali.

/3/ <u>Twierdzenie</u>

Jeśli funkcje x i y są zmiennymi losowymi, to zmiennymi losowymi są również funkcje:

$$x + a,\ ax + by,\ |x| \ \text{i}\ xy,$$

gdzie a i b są liczbami rzeczywistymi.

Jeśli ponadto $y \neq 0$ /prawie wszędzie/, to $\frac{x}{y}$ jest również zmienną losową.

Dowód /3/ można znaleźć np. w pracy S. Zubrzycki [38], str. 104 i dalsze.

/4/ Definicja

Statystyką nazywamy zmienną losową będącą funkcją łącznej zmiennej losowej /wektora losowego/

$$\varphi(x_1, \dots, x_n),$$

/5/ Definicja

Ciąg zmiennych losowych $\{x_n\}$ jest zbieżny według prawdopodobieństwa /inaczej stochastycznie zbieżny/ do zmiennej losowej x, jeśli dla każdego $\varepsilon > 0$

$$P\left(\left\{|x_n - x| \geqslant \varepsilon\right\}\right) \longrightarrow 0 \qquad \text{przy } n \longrightarrow \infty.$$

Pisze się wtedy

$$x = P - \lim x_n. \tag{6}$$

/7/ Definicja

Rozkładem prawdopodobieństwa /dystrybuantą/ zmiennej losowej x nazywamy funkcję

$$F(t) = P(x \leqslant t).$$

/8/ Definicja

Wartością oczekiwaną dowolnej funkcji f(x) zmiennej losowej x nazywamy

$$E\big(f(x)\big) = \int_{-\infty}^{\infty} f(t)\, dF(t).$$

/9/ Definicja

Łącznym rozkładem prawdopodobieństwa /dystrybuantą łączną/ skończonego zbioru zmiennych losowych $x_1,\dots,x_n$ nazywamy funkcję

$$F(t_1,\dots,t_n) = P(x_1 \leqslant t_1,\dots,x_n \leqslant t_n)$$

/10/ Definicja

Wartością oczekiwaną dowolnej funkcji $f(x_1,\dots,x_n)$ zmiennych losowych $x_1,\dots,x_n$ nazywamy

$$E\big(f(x_1,\dots,x_n)\big) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(t_1,\dots,t_n)\, dF(t_1,\dots,t_n)$$

/11/ Definicja

Warunkową wartością oczekiwaną zmiennej losowej x pod warunkiem, że zmienne losowe $x_1,\dots,x_n$ przyjmą odpowiednio wartości $t_1,\dots,t_n$ nazywamy zmienną losową

$$\varphi(t_1,\dots,t_n) = E\left(x/\left\{x_1,\dots,x_n\right\}\right) = E\left(x/x_1 = t_1,\dots,x_n = t_n\right) .$$

/12/ Definicja

Macierzą kowariancji zmiennych losowych $x_1,\dots,x_n$ nazywamy macierz

$$Q\left(x_1,\dots,x_n\right) = \left(c_{ij}\right) ,$$

gdzie

$$c_{ij} = E\left[\left(x_i - E(x_i)\right)\left(x_j - E(x_j)\right)\right].$$

/13/ Definicja

Zmienne losowe x_i, x_j nazywają się nieskorelowane, jeśli

$$E\left(x_i x_j\right) = E(x_i)\; E(x_j) .$$

/14/ Definicja

Zmienne losowe x_i, x_j są niezależne stochastycznie, jeśli

$$F(x_i, x_j) = F(x_i)\ F(x_j),$$

tzn. łączny rozkład prawdopodobieństwa zmiennych x_i, x_j jest równy iloczynowi rozkładów zmiennych x_i i x_j.

/15/ Twierdzenie

Zmienne x_i, x_j są nieskorelowane wtedy i tylko wtedy, gdy

$$Cov(x_i, x_j) = 0.$$

/16/ Twierdzenie

Jeśli zmienne losowe są niezależne stochastycznie, to są nieskorelowane.

Twierdzenie odwrotne zachodzi tylko w przypadku, gdy zmienne x_i, x_j mają rozkłady normalne.

/17/ Definicja

Wariancją zmiennej losowej x nazywamy wyrażenie:

$$\sigma(x) = Var(x) = E\left(x - E(x)\right)^2.$$

§ 2. Przestrzeń liniowa.

/18/ Definicja

Przestrzenią liniową L nad ciałem C nazywa się zbiór, którego elementy nazywane są wektorami, dla których określone są następujące działania: dodawanie i mnożenie przez elementy ciała C, spełniające warunki

dla x,y,z należących do przestrzeni liniowej L i a,b,c należących do ciała C

$$x + y = y + x \qquad (x + y) + z = x + (y + z)$$
$$a(x + y) = ax + ay \qquad (a + b)x = ax + bx$$
$$a(bx) = (ab)x \qquad 1 \cdot x = x$$

Jeśli $x + y = x + z$ to $y = z$.

Iloczyn $0 \cdot x$ jest równy dla każdego x temu samemu elementowi Θ, który nazywa się wektorem zerowym.
Symbolem $-x$ oznacza się iloczyn -1 x. Różnicę $x - y$ definiuje się następująco:

$$x - y = x + (-y).$$

/19/ Definicja

Elementy $x_1, \ldots, x_n$ przestrzeni liniowej nazywają się liniowo niezależne, jeśli z równości

$$a_1x_1 + a_2x_2 + \dots + a_nx_n = \Theta ,$$

gdzie $a_1, \dots, a_n$ należą do ciała C, wynika że

$$a_1^2 + \dots + a_n^2 = 0.$$

Można łatwo dowieść, że zbiór wszystkich kombinacji liniowych dowolnego zbioru wektorów jest przestrzenią liniową, a stąd

/20/ Definicja

Przestrzenią liniową M generowaną przez układ wektorów $x_1, \dots, x_n$ nazywa się zbiór wszystkich kombinacji liniowych tych wektorów i oznacza się

$$M = [x_1, \dots, x_n].$$

/21/ Definicja

Bazą przestrzeni liniowej L nazywa się taki układ wektorów liniowo niezależnych, który generuje całą przestrzeń L.

/22/ Definicja

Wymiarem przestrzeni liniowej nazywa się moc zbioru wektorów tworzących bazę.

Jeśli baza składa się ze skończonej liczby wektorów, to wymiar przestrzeni jest równy po prostu ich ilości. Wymiar przestrzeni L oznacza się przez dim L.

/23/ Definicja

Przekształceniem liniowym nazywa się taką funkcję f, która odwzorowuje przestrzeń liniową L w siebie $(f : L \longrightarrow L)$ i ponadto

$$f(xa + by) = af(x) + bf(y) ,$$

gdzie $x, y \in L$, $a,b \in C$.

Można wykazać, że dla każdego przekształcenia liniowego f istnieje macierz A taka, że

$$f(x) = Ax.$$

Łatwo sprawdzić, że każda macierz A określa przekształcenie liniowe f przez powyższą tożsamość.

/24/ Definicja

Funkcjonałem liniowym nazywamy funkcję $f : L \longrightarrow C$ taką, że

$$f(ax + by) = af(x) + bf(y) .$$

/25/ Definicja

Wartością własną przekształcenia liniowego f /macierzy odpowiadającej temu przekształceniu/ nazywamy taki skalar a, dla którego

$$f(x) = ax \qquad (x \in L),$$

lub co na jedno wychodzi

$$Ax = ax,$$

przy czym wektor x nazywa się wektorem własnym.

Zachodzi następujące twierdzenie

/26/ Twierdzenie

Skalar a jest wartością własną macierzy A wtedy i tylko wtedy, gdy a jest pierwiastkiem równania

$$\det (A - aI) = 0,$$

gdzie I oznacza macierz jednostkową.

§ 3. Przestrzeń Hilberta

W niniejszej pracy będziemy rozpatrywali tylko przestrzenie nad ciałem liczb rzeczywistych R.

/27/ Definicja

Iloczynem skalarnym $(x|y)$ wektorów x, y, należących do przestrzeni liniowej L nad ciałem R, nazywamy funkcję o wartościach w R określoną na iloczynie kartezjańskim $L \times L$, która ma następujące własności:

1. $(x|y) = (y|x)$,
2. $(ax + by|z) = a(x|z) + b(y|z)$,
   gdzie $a, b \in R$.
3. $(x|x) \geqslant 0$,
   przy czym równość ma miejsce wtedy i tylko wtedy, gdy $x = \Theta$.

/28/ Definicja

Przestrzenią unitarną nazywamy przestrzeń liniową z iloczynem skalarnym.

/29/ Definicja

Wektory x, y nazywają się ortogonalne, jeśli

$$(x|y) = 0,$$

co zapisuje się:

$$x \perp y.$$

/30/ Definicja

Zbiór wektorów $\{x_1,\dots,x_n\}$ nazywa się ortogonalny, jeśli $x_i \perp x_j$ dla $i,j = 1,\dots,n$, $i \neq j$.

Uwaga

Jeżeli zbiór $\{x_1,\dots,x_n\}$ jest ortogonalny i nie zawiera wektora zerowego to jest on liniowo niezależny.
Twierdzenie odwrotne nie jest prawdziwe.

/31/ Definicja

Wektor y jest ortogonalny do zbioru $\{x_1,\dots,x_n\}$, jeśli $y \perp x_i$ dla $i = 1,\dots,n$.

Uwaga

Jeśli $(x|y) = 0$ dla każdego y, to $x = \Theta$, ponieważ wtędy jest również $(x|x) = 0$, co wobec własności iloczynu skalarnego implikuje $x = \Theta$.

Za pomocą iloczynu skalarnego można wprowadzić tzw. normę wektora. Pojęcie to jest uogólnieniem długości wektora.

/32/ Definicja

Normę - $\|x\|$ wektora x określa równość

$$\|x\| = \sqrt{(x|x)}$$

Z własności iloczynu skalarnego wynikają następujące własności normy:

$$\|x\| \geqslant 0; \qquad \|ax\| = |a| \; \|x\| ;$$
$$\|x + y\| \leqslant \|x\| + \|y\|;$$
$$|(x|y)| \leqslant \|x\| \; \|y\|,$$

przy czym równość zachodzi wtedy i tylko wtedy, gdy x = ay, lub x = Θ.

/33/ Lemat

Jeśli $x \perp y$, to

$$\|x + y\|^2 = \|x\|^2 + \|y\|^2$$

Treścią tego lematu jest fakt, że w przestrzeni unitarnej jest prawdziwe twierdzenie Pitagorasa. To między innymi pozwala na wykorzystywanie w rozważaniach dotyczących przestrzeni unitarnych analogii geometrycznych.

W przestrzeni unitarnej można, za pomocą dotąd wprowadzonych pojęć, zdefiniować kąt i odległość między wektorami.

/34/ Definicja

Cosinus kąta między wektorami $x \neq \theta$ i $y \neq \theta$ określa się za pomocą równości:

$$\cos\alpha = \frac{(x|y)}{\|x\|\,\|y\|},$$

gdzie α nazywa się kątem między wektorami x i y $\left(\alpha = \sphericalangle(x,y)\right)$

/35/ Definicja

Odległością między wektorami x i y nazywa się funkcję $d : L \times L \to R$ określoną równością

$$d(x,y) = \|x - y\|.$$

Z własności normy wynikają następujące własności odległości:

1. $d(x,y) \geqslant 0$,

   przy czym równość ma miejsce wtedy i tylko wtedy, gdy $x = y$.
2. $d(x,y) = d(y,x)$,
3. $d(x,y) \leqslant d(y,z) + d(x,z)$.

Funkcję d spełniającą warunki 1, 2 i 3 nazywa się też metryką.

/36/ Lemat

$$\|x - y\|^2 = \|x\|^2 + \|y\|^2 - 2\|x\|\,\|y\| \cos \sphericalangle (x,y).$$

Za pomocą pojęcia metryki można zdefiniować zbieżność ciągu elementów przestrzeni unitarnej.

/37/ Definicja

Ciąg $\{x_n\}$ jest zbieżny do pewnego elementu x, jeśli dla każdego $\varepsilon > 0$ istnieje takie n_0, że dla $n > n_0$

$$d(x_n,x) < \varepsilon .$$

/38/ Definicja

Podzbiór A przestrzeni unitarnej jest domknięty, jeśli każdy ciąg zbieżny elementów A ma granicę należącą do zbioru A.

/39/ Definicja

Funkcja $f \; L \to R$ jest ciągła w punkcie x, jeśli dla każdego $\varepsilon > 0$ istnieje takie $\eta > 0$, że z nierówności

$$d(x_n,x) < \eta$$

wynika

$$|f(x_n) - f(x)| < \varepsilon$$

/39'/ Lemat

Funkcja f jest ciągła wtedy i tylko wtedy, gdy z $x_n \longrightarrow x$ wynika, że

$$f(x_n) \longrightarrow f(x).$$

/40/ Lemat

Iloczyn skalarny jest funkcją ciągłą, tzn. jeśli

$$x_n \longrightarrow x \text{ i } y_n \longrightarrow y, \quad \text{to} \quad (x_n | y_n) \longrightarrow (x|y).$$

/41/ Definicja

Ciąg (x_n) spełnia warunek Cauchy'ego, jeśli dla każdego $\varepsilon > 0$ istnieje takie k, że dla $n_o, m_o > k$,

$$d(x_{n_o}, x_{m_o}) < \varepsilon.$$

/42/ Definicja

Przestrzeń nazywa się zupełna, jeśli z tego, że ciąg (x_n) spełnia warunek Cauchy'ego wynika, że ciąg (x_n) jest zbieżny.

/43/ Definicja

Przestrzenią Hilberta nazywa się przestrzeń unitarną zupełną w sensie metryki

$$d(x,y) = \|x - y\| ,$$

gdzie norma jest określona za pomocą iloczynu skalarnego, jak w definicji /32/.

Przytoczymy teraz pewne twierdzenia i definicje dotyczące przestrzeni Hilberta, na które będziemy się powoływali w ciągu dalszym pracy. Dowody twierdzeń można znaleźć np. w pracy D. Luenbergera [26].

/44/ Definicja

Dopełnieniem ortogonalnym podzbioru S przestrzeni unitarnej nazywa się zbiór wszystkich wektorów ortogonalnych do S. Dopełnienie ortogonalne zbioru S oznacza się przez $S^{\perp}$.

Uwaga

Zbiór $S^{\perp}$ jest podprzestrzenią liniową. Wynika to stąd, że suma wektorów prostopadłych do S jest prostopadła do S i iloczyn ax, gdzie $x \perp S$, $a \in R$ jest wektorem prostopadłym do S.

/45/ Definicja

Przestrzeń liniowa L jest sumą prostą dwu podprzestrzeni

K i M, jeśli każdy wektor $x \in L$ ma jednoznaczne przedstawienie postaci

$$x = k + m,$$

gdzie $k \in K$, $m \in M$. Pisze się wtedy

$$L = K \oplus M.$$

/46/ Twierdzenie

Jeśli M jest domkniętą liniową podprzestrzenią przestrzeni Hilberta H, to

$$H = M \oplus M^{\perp} \quad \text{i} \quad M = \left(M^{\perp}\right)^{\perp}.$$

/47/ Twierdzenie /klasyczne twierdzenie o rzutowaniu/

Niech H będzie przestrzenią Hilberta i M domkniętą podprzestrzenią H. Dla każdego wektora $x \in H$ istnieje jedyny wektor $m_0 \in H$ taki, że

$$\| x - m_0 \| \leqslant \| x - m \| \qquad \text{dla każdego} \quad m \in M.$$

Ponadto: m_0 jest jedynym minimalizującym wektorem wtedy i tylko wtedy, gdy $x - m_0$ jest ortogonalny do M.

Jeśli podprzestrzeń $M \subset H$ jest generowana przez zbiór wektorów $\{x_1,\dots,x_n\}$ liniowo niezależny, to

$$(y - m_o | x_j) = 0 \qquad \text{dla } j = 1,\dots,n$$

i

$$m_o = \sum_{i=1}^{n} a_i x_i \,, \tag{/47'/}$$

stąd

$$\left(y - \sum_{i=1}^{n} a_i x_i | x_j\right) = 0,$$

co można zapisać

$$\begin{bmatrix} (x_1|x_1) & & (x_n|x_1) \\ & & \\ (x_1|x_n) & & (x_n|x_n) \end{bmatrix} \begin{bmatrix} a_1 \\ \cdot \\ \cdot \\ \cdot \\ a_n \end{bmatrix} = \begin{bmatrix} (y|x_1) \\ \\ \cdot \\ (y|x_n) \end{bmatrix} \,, \tag{/48/}$$

stąd a_i wyznacza się z wzoru

$$\begin{bmatrix} a_1 \\ \cdot \\ \cdot \\ \cdot \\ a_n \end{bmatrix} = \left((x_i|x_j)\right)^{-1} \begin{bmatrix} (y|x_1) \\ \\ \cdot \\ (y|x_n) \end{bmatrix} \tag{/48'/}$$

Jeśli $\{x_i\}$ są zortonormalizowane, tzn $\|x_i\| = 1$ i $\{x_i\}$ jest zbiorem ortogonalnym, to $a_i = (y|x_i)$

Twierdzenie /47/ daje jednoznaczne rozwiązanie problemu aproksymacji liniowej, ma przeto zastosowanie np. w metodzie najmniejszych kwadratów. W naszych rozważaniach będziemy się na nie często powoływali, między innymi przy wyborze zmiennych objaśniających.

Twierdzenie /47/ nie jest prawdziwe w dowolnej przestrzeni unitarnej. Tak więc przestrzeń Hilberta jest minimalnym aparatem formalnym koniecznym do formułowania i rozwiązywania wielu problemów ekonometrii.

§ 4. Przykłady przestrzeni Hilberta.

/49/ Przykład

Najbardziej znanym przykładem przestrzeni Hilberta jest przestrzeń euklidesowa E^n. Wektorami tej przestrzeni są n-elementowe ciągi liczb rzeczywistych. Działania na nich określa się następująco:

$$(a_1,a_2,\dots,a_n) + (b_1,b_2,\dots,b_n) = (a_1+b_1,a_2+b_2,\dots,a_n+b_n) ;$$

$$\alpha(a_1,a_2,\dots,a_n) = (\alpha a_1 , \alpha a_2,\dots, \alpha a_n) , \qquad \alpha\in R,$$

iloczyn skalarny wektorów

$$x = (a_1,\dots,a_n), \quad y = (b_1,\dots,b_n)$$

$$(x|y) = a_1b_1 + a_2b_2 + \dots + a_nb_n.$$

Stąd norma wektora i odległość miedzy wektorami, określone przez iloczyn skalarny, są postaci:

$$\|x\|= \sum_{i=1}^{n} a_i^2 ,$$

$$d(x,y) = \sum_{i=1}^{n} (a_i - b_i)^2.$$

Łatwo sprawdzić, że tak określone działania, norma i odległość spełniają wcześniej wymienione własności tych pojęć, ponadto z analizy matematycznej wiadomo, że E^n jest w sensie przyjętej metryki zupełna. Tak więc przestrzeń euklidesowa E^n jest modelem przestrzeni Hilberta.

/50/ <u>Przykład</u>

Sformułujmy teraz ważny dla dalszego ciągu pracy przykład przestrzeni Hilberta zmiennych losowych L_2^*.

Niech x,y,... oznaczają zmienne losowe o wartościach rzeczywistych takie, że

$$E|x|^2<\infty \quad , \quad E|y|^2<\infty,\ldots$$

gdzie E(x) oznacza wartość oczekiwaną zmiennej x.

Zbiór takich zmiennych losowych z dodawaniem i mnożeniem przez liczby rzeczywiste określonymi w zwykły sposób, iloczynem skalarnym danym równością

$$(x|y) = Exy$$

jest przestrzenią Hilberta /por. np. M. Loeve [25] ./.

Normę w tej przestrzeni określa się następująco:

$$\|x\| = (E|x|^2)^{\frac{1}{2}}$$

/51/ Definicja

Odległość d zmiennych losowych x,y określamy jako

$$d(x,y) = \left(E|x - y|^2\right)^{\frac{1}{2}} = \|x - y\|.$$

Z powyższych definicji i związku niezależności liniowej z ortogonalnością wektorów wynika

/52/ Lemat

Jeśli zmienne losowe x,y są nieskorelowane to są niezależne liniowo.

/53/ Lemat

Jeśli zmienne losowe x,y mają rozkłady normalne to z równości

$$(x|y) = 0$$

wynika, że x i y są nieskorelowane.

/54/ Przykład

O zmiennych losowych rozpatrywanych w przykładzie /50/ można wykazać, korzystając z nierówności Cauchy'ego, że mają one również skończone wartości oczekiwane.

Rozważmy zbiór zmiennych losowych scentrowanych ich wartościami oczekiwanymi, tzn. zmiennych postaci

$$x = x^* - E(x^*).$$

Dla prostoty zapisu, zmienne scentrowane będą oznaczone przez $x, y, \ldots$. Dla takich zmiennych

$$(x|y) = E(xy),$$

a więc iloczyn skalarny jest kowariancją. Norma natomiast jest odchyleniem standardowym wobec równości

$$\|x\| := \left(E|x|^2\right)^{\frac{1}{2}} = \sqrt{\mathrm{Var}(x)} = \sigma(x).$$

Odeległość między zmiennymi x, y określa formuła:

$$d(x,y) = \left(E|x-y|^2\right)^{\frac{1}{2}}.$$

/55/ Przykład

Model rozpatrywany w przykładzie /50/ można uogólnić rozpatrując przestrzeń wektorów, których składowe są zmiennymi losowymi.

/56/ Definicja

n-wymiarowym wektorem losowym nazywa się n-elementowy ciąg zmiennych losowych

$$x = (x_1, \dots, x_n)$$

przy czym

$$E|x_i|^2 < \infty \qquad \text{dla } i = 1, \dots, n.$$

Przestrzeń Hilberta H_2 wektorów losowych można otrzymać z danego zbioru wektorów losowych w pewnym stopniu analogicznie jak w przykładzie /50/. Wektory losowe dodaje się i mnoży przez skalar tak jak wektory przestrzeni E^n. Iloczyn skalarny definiuje się

$$(x|y) = E\left(\sum_{i=1}^{n} x_i y_i\right) = \sum_{i=1}^{n} E(x_i y_i),$$

gdzie x jest wektorem losowym o składowych x_i, a $y = (y_1, \dots, y_n)$ / z nierówności Cauchy'ego wynika, że również $E(x_i y_i) < \infty$ /.

Wygodną postacią definicji iloczynu skalarnego jest zapis:

$$(x|y) = E(xy'),$$

przy czym y' oznacza wektor y transponowany.

Wprowadzając pojęcie śladu macierzy

/57/ $$Tr(A) = \sum_{i=1}^{n} a_{ii},$$

gdzie A jest macierzą kwadratową o elementach a_{ij},

iloczyn skalarny wektorów x,y można zapisać:

$$(x|y) = Tr(E(x'y)).$$

Norma wektora x ma wtedy postać:

$$\|x\| = (Tr(E(x'x))^{\frac{1}{2}},$$

a odległość d między x i y wyraża się wzorem

$$d(x,y) = \|x - y\|.$$

§ 5. Niektóre własności macierzy, macierze i wyznaczniki Grama, formy kwadratowe.

W paragrafie tym zostaną sformułowane pewne specjalne twierdzenia i definicje dotyczące macierzy, wyznaczników i form kwadratowych, które zostaną wykorzystane w dalszych rozważaniach.

/58/ Definicje

Rzędem macierzy A nazywamy maksymalną ilość liniowo niezależnych kolumn /wierszy/ tej macierzy.

/59/ Twierdzenie

Rząd macierzy A' jest równy rzędowi macierzy A.

/60/ Definicja

Minorem głównym macierzy kwadratowej A nazywamy minor powstały przez skreślenie w macierzy A wierszy i kolumn o tych samych wskaźnikach.

/61/ Twierdzenie

Rząd macierzy symetrycznej jest równy najwyższemu ze stopni

jej minorów głównych różnych od zera.
Dowód tego twierdzenia jest np. w: A. Mostowski i M. Stark [29].

/62/ Twierdzenie

Niech M różne od zera będzie minorem głównym stopnia r macierzy symatrycznej A. Niech wszystkie minory główne powstające z M przez dodanie jednego wiersza i kolumny, jak również minory główne powstające z M przez dodanie dwóch wierszy i kolumn będą równe zero. Wówczas zachodzi równość

$$\text{rząd } A = r.$$

Twierdzenie to jest udowodnione np. w: A. Mostowski i M. Stark [29]

/63/ Definicja

Macierzą Grama $G(x_1,\dots,x_n)$ nazywamy macierz $G = (a_{ij})$, przy czym

$$a_{ij} = (x_i | x_j),$$

gdzie x_i, x_j $(i, j = 1,\dots,n)$ są wektorami przestrzeni unitarnej.

Wobec symetrii iloczymu skalarnego macierz Grama jest macierzą symetryczną.

Macierz Grama wykorzystamy do wyznaczenia współczynników rzutu prostopadłego m_0 wektora y. Wzór /48'/ można mianowicie zapisać w postaci:

/64/
$$G \begin{bmatrix} a_1 \\ \\ a_n \end{bmatrix} = \begin{bmatrix} (y|x_1) \\ \\ \cdot \\ (y|x_n) \end{bmatrix}$$

stąd

$$\begin{bmatrix} a_1 \\ \cdot \\ \cdot \\ \cdot \\ a_n \end{bmatrix} = (G^T)^{-1} \begin{bmatrix} (y|x_1) \\ \\ \cdot \\ \cdot \\ (y|x_n) \end{bmatrix} \quad .$$

/65/ Definicja

Wyznacznikiem Grama $g(x_1,\ldots,x_n)$ nazywamy wyznacznik macierzy Grama

$$g(x_1,\ldots,x_n) = \det G(x_1,\ldots,x_n) .$$

/66/ Twierdzenie

Wyznacznik Grama $g(x_1,\ldots,x_n)$ jest różny od zera wtedy i tylko wtedy, gdy wektory $x_1,\ldots,x_n$ są liniowo niezależne.

Dowód tego twierdzenia jest np. w pracy D. Luenbergera [26].

/67/ <u>Twierdzenie</u>

Wymiar macierzy Grama $G(x_1,\ldots,x_n)$ jest równy liczbie liniowo niezależnych wektorów układu $x_1,\ldots,x_n$.

Za pomocą wyznaczników Grama można sformułować twierdzenie dotyczące odległości wektora należącego do przestrzeni Hilberta H od podprzestrzeni $K \subset H$. Twierdzenie to jest uzupełnieniem twierdzenia /47/ o rzutowaniu.

/68/ <u>Twierdzenie</u>

Niech $y \in H$ i dany jest układ liniowo niezależnych wektorów $x_1,\ldots,x_n$ generujący podprzestrzeń K. Oznaczmy przez d odległość wektora y od podprzestrzeni K.
Wtedy

/69/ $$d^2 = \frac{g(x_1,\ldots,x_n,y)}{g(x_1,\ldots,x_n)}$$

Dowód tego twierdzenia można znaleźć np. w pracy D. Luenbergera [26].

Z równości /69/ wynikają nierówności

/70/ $$\frac{g(x_1,\ldots,x_{n+1})}{g(x_1,\ldots,x_n)} = d_n^2 > 0 \qquad (\text{dla } n = 1,2,\ldots,m-1)$$

Ponadto wiadomo, że

$$/71/ \qquad g(x_1) = (x_1|x_1) > 0.$$

Z nierówności /70/ i /71/ wynika następująca interpretacja wyznacznika Grama.

/72/ Twierdzenie

Wyznacznik Grama $g(x_1,\dots,x_n)$ jest kwadratem objętości równoległościanu rozpiętego na wektorach $x_1,\dots,x_n$.

Dowód

Oznaczmy

$$g_n = g(x_1,\dots,x_n) , \qquad (n = 1,\dots,m-1).$$

Z /70/ i /71/ otrzymujemy

$$/73/ \qquad \begin{aligned} \sqrt{g_1} &= \|x_1\| = V_1 \\ \sqrt{g_2} &= V_1 \; h_1 = V_2 \\ &\;\;\vdots \\ \sqrt{g_m} &= V_{m-1} \; h_{m-1} = V_m , \end{aligned}$$

przy czym tę ostatnią formułę wraz z równością /73/ można traktować jako definicję indukcyjną objętości n-wymiarowego równoległościanu rozpiętego na wektorach $x_1,\dots,x_n$.

/74/ Uwaga

Wnioskiem z interpretacji geometrycznej wyznacznika Grama jest fakt, że jego wartość nie zależy od kolejności wektorów, tzn.:

$$g(x_1,\dots,x_n) = g(x_{\alpha_1},\dots,x_{\alpha_n}),$$

gdzie $\alpha_1,\dots,\alpha_n$ jest dowolną permutacją liczb $1,\dots,n$.

Przytoczymy teraz pewne nierówności między wyznacznikami Grama, które zostaną wykorzystane w ciągu dalszym.

Z faktu, że dowolny wektor y można rozłożyć na składowe: m_o i prostopadły doń wektor m_1, wynika że

$$(y|y) = (m_o + m_1|m_o + m_1) = (m_o|m_o) + (m_1|m_1) \geqslant (m_1|m_1) = d^2 ,$$

co daje

$$g(x_1,\dots,x_n,y) \leqslant g(x_1,\dots,x_n)\ g(y),$$

a stąd

/75/ $$g(x_1,\dots,x_n) \leqslant g(x_1)\ g(x_2) \dots g(x_n).$$

Nierówność /75/ nazywa się nierównością Hadamarda. Jej sens geometryczny jest taki, że objętość dowolnego równoległościanu jest nie większa niż objętość prostopadłościanu rozpiętego na wektorach o tej samej długości. Ogólniejszą postacią nierówności /75/ jest tzw. uogólniona nierówność Hadamarda:

/76/ $$g(x_1,\dots,x_n) \leqslant g(x_1,\dots,x_p)\; g(x_{p+1},\dots,x_n) \;,$$

przy czym równość ma miejsce wtedy i tylko wtedy, gdy każdy z wektorów $x_1,\dots,x_p$ jest ortogonalny do każdego z wektorów $x_{p+1},\dots,x_n$.

Dowód /76/ można znaleźć np. w: F.R. Gantmacher [9].

Z uwagi /74/ i nierówności /76/ wynika, że

/77/ $$g(x_1,\dots,x_n) \leqslant g(x_{\alpha_1},\dots,x_{\alpha_p})\; g(x_{\alpha_{p+1}},\dots,x_{\alpha_n}).$$

/78/ Definicja

Formą kwadratową nazywamy wyrażenie

/79/ $$xAx' = \sum_{i,j}^{n} a_{ij}x_i x_j \;,$$

gdzie A jest symetryczną macierzą stopnia n, która nazywa się macierzą formy kwadratowej a x' jest wektorem transponowanym wektora x o n składowych

$$x = (x_1, \ldots, x_n) .$$

/80/ Definicja

Forma kwadratowa xAx' nazywa się dodatnio określona, jeśli dla $x \neq 0$

$$xAx' > 0;$$

ujemnie określona, jeśli

$$xAx' < 0;$$

nieujemnie określona, jeśli

$$xAx' \geqslant 0;$$

i niedodatnio określona, jeśli

$$xAx' \leqslant 0.$$

Poniższe twierdzenia są kryteriami pozwalającymi ustalić określoność formy.

/81/ Twierdzenie

Forma kwadratowa /79/ jest dodatnio określona wtedy i tylko wtedy, gdy wszystkie minory kątowe jej macierzy dą dodatnie. Stąd można udowodnić, że jeśli $xAx' > 0$ to i wszystkie minory główne macierzy A są dodatnie /por. np.: A.P.Miszina, I.W.Proskuriakow [27] /.

/83/ Twierdzenie

Forma /79/ jest nieujemnie określona wtedy i tylko wtedy, gdy wszystkie jej minory główne są nieujemne /por. np.: A.P.Miszina i I.W.Proskuriakow [27] /.

/84/ Twierdzenie

Forma /79/ jest dodatnio określona wtedy i tylko wtedy, gdy jej macierz A można przedstawić w postaci

/85/ $$A = C'C,$$

gdzie C jest macierzą nieosobliwą stopnia n.
Dowód /84/ jest np.: w F.R. Gantmacher [9] .

/86/ Twierdzenie

Forma /79/ rzędu r /tzn. rząd A = r/ jest nieujemna wtedy i tylko wtedy, gdy można ją przedstawić w postaci /85/, gdzie C jest macierzą stopnia n rzędu r. Można przy tym przyjąć, że pierwsze r wierszy macierzy C są liniowo niezależne, a pozostałe wiersze są zerowe.

Dowód tego twierdzenia można znaleźć w cytowanej już pracy [9].

R o z d z i a ł II

INTERPRETACJA POJĘĆ STATYSTYKI MATEMATYCZNEJ W PRZESTRZENI HILBERTA

W rozdziale wykorzystamy fakt, że zmienne losowe o skończonych wariancjach z odpowiednimi operacjami są modelem przestrzeni Hilberta. Spostrzeżenie to pozwoli nam dać interpretację geometryczną pewnych ważnych w ekonometrii pojęć statystyki matematycznej. Staną się one przez to bliższe intuicji, co pozwoli łatwiej nimi operować: uprościć dowody niektórych twierdzeń, wykorzystać geometryczne analogie w konstrukcjach znajdujących zastosowanie w ekonometrii, dokonać pewnych spostrzeżeń dotyczących problemów ekonometrii a łatwo dających się zauważyć wtedy, gdy sformułuje się je w języku geometrii.

§ 1. Współczynnik korelacji.

Załóżmy, że x,y,z,... oznaczają zmienne losowe będące elementami omawianej w rozdziale I przestrzeni L_2, a więc o zmiennych losowych tu rozpatrywanych zakładamy, że mają wartości oczekiwane równe zero i skończone wariancje.

W L_2 iloczyn skalarny zmiennych x,y określa się następująco:

$$(x|y) = E(xy),$$

stąd norma dowolnej zmiennej x jest określona wzorem

/1/ $$\|x\| = \left(E|x|^2\right)^{\frac{1}{2}},$$

a odległość między zmiennymi x,y

$$d(x,y) = \|x - y\|.$$

W ciągu dalszym elementy L_2 będą nazywane wektorami, bądź zmiennymi.

Ponieważ zmienne losowe mają wartości oczekiwane równe zero, kowariancja jest iloczynem skalarnym:

/2/ $$\mathrm{Cov}(x,y) = E\left[\left(x - E(x)\right)\left(y - E(y)\right)\right] = E(xy) = (x|y),$$

natomiast wariancja dowolnej zmiennej x jest równa kwadratowi jego długości:

$$/3/ \qquad \mathrm{Var}(x) = E(x - E(x))^2 = E(x)^2 = \|x\|^2.$$

Tak więc w przestrzeni L_2 wariancja zmiennej jest kwadratem jej normy, a stąd odchylenie standardowe zmiennej jest jej długością, można mówić wobec tego, że zmienne losowe zestandaryzowane mają długość równą jeden.

Z powyższych równości współczynnik korelacji $r(x,y)$ między zmiennymi losowymi $x,y \in L_2$ wyraża się wzorem:

$$/4/ \qquad r(x,y) = \frac{E(xy)}{\sqrt{E|x|^2}\sqrt{E|y|^2}} = \frac{(x|y)}{\|x\|\,\|y\|}$$

Ostatnie wyrażenie /por. rozdział I, definicja /34// jest równe cosinusowi[1] kąta α między zmiennymi x i y, a więc

$$/5/ \qquad r(x,y) = \cos\alpha .$$

Znajdziemy teraz związek współczynnika korelacji $r(x,y)$ z odległością $d(x,y)$ między zmiennymi losowymi x i y. Odległość spełnia następujące równości:

$$/6/ \qquad d^2(x,y) = \|x - y\|^2 = \|x\|^2 + \|y\|^2 - 2\|x\|\,\|y\|\cos\alpha ,$$

gdzie $\alpha = \measuredangle(x,y)$.

1/ Zauważmy, że cosinus może być użyty jako miara zależności dowolnych zmiennych losowych o skończonych, różnych od zera wariancjach.

Podstawiając do /6/ równości /4/ i /5/ otrzymujemy twierdzenie wyrażające zależność między współczynnikiem korelacji a odległością.

/7/ <u>Twierdzenie</u>

$$d(x,y) = \sqrt{Var(x) + Var(y) - 2\sqrt{Var(x)}\sqrt{Var(y)}\; r(x,y)}$$

Dla zmiennych x,y zestandaryzowanych zależność ta ma postać

$$d(x,y) = \sqrt{2 - 2r(x,y)}$$

<u>Wnioski</u>

1. Zmienne losowe x,y zestandaryzowane są zależne liniowo, gdy odległość między nimi jest równa zero. Wynika to z równości /5/ i /6/ oraz faktu, że jeśli współczynnik korelacji między zmiennymi losowymi jest równy jedności, to zmienne te są zależne.

2. Jeśli $r(x,y) = 0$, to zmienne x,y są ortogonalne, co implikuje ich liniową niezależność. Wynika to z równości /5/ i /6/ oraz twierdzenia Pitagorasa /por. rozdział I, lemat /33//.

§ 2. Macierz współczynników korelacji.

Oznaczmy przez R macierz, której elementami są współczynniki korelacji między zmiennymi losowymi $x_1,\dots,x_n$

/8/
$$R = \begin{bmatrix} 1 & r(x_1,x_2) & & r(x_1,x_n) \\ \dots & \dots & \dots & \dots \\ r(x_n,x_1) & r(x_n,x_2) & \dots & 1 \end{bmatrix}$$

Zakładając, że zmienne $x_1,\dots,x_n$ są zestandaryzowane, co oznacza, że przyjmujemy dodatkowo

/9/
$$\sigma(x_i) = \sqrt{\mathrm{Var}(x_i)} = \| x_i \| = 1,$$

wobec /2/ i /4/ otrzymujemy

/10/
$$R = \begin{bmatrix} 1 & (x_1|x_2) & & (x_1|x_n) \\ \dots & \dots & \dots & \dots \\ (x_n|x_1) & (x_n|x_2) & & 1 \end{bmatrix},$$

stąd macierz współczynników korelacji jest macierzą Grama wektorów $x_1,\dots,x_n \in L_2$, a więc ma własności tej macierzy: jest między innymi nieujemnie określona [1/]. Własność ta charakteryzuje dokładnie macierz współczynników korelacji. Zachodzi mianowicie następujące

/11/ Twierdzenie

Symetryczna macierz $R = (r_{ij})$ taka, że $r_{ii} = 1$, $|r_{ij}| \leqslant 1$, $(i,j = 1,\dots,n)$ jest macierzą współczynników korelacji wtedy i tylko wtedy, gdy jest nieujemnie określona.

Dowód

Jeśli macierz R jest macierzą współczynników korelacji, to ponieważ może być interpretowana jako macierz Grama, jest nieujemnie określona.

Jeśli natomiast macierz R ma własności wymienione w twierdzeniu to wobec wzoru /10/ może być rozpatrywana jako macierz kowariancji zmiennych losowych zestandaryzowanych /standaryzacja nie zmienia współczynników korelacji/, a symetryczna macierz jest macierzą kowariancji wtedy i tylko wtedy, gdy jest nieujemnie określona /por. np.: S. Zubrzycki [38] s.235/. Kończy to dowód .

1/ Porównaj np.: F.R. Gantmacher [9] s. 228 i dalsze.

W niektórych przypadkach między elementami macierzy R współczynników korelacji zachodzą pewne związki, co jest treścią następującego twierdzenia /12/.

Niech $Rz\left(y/[x_1,\dots,x_n]\right) = Rz(y/M)$ oznacza rzut prostopadły[1/] wektora y na podprzestrzeń $M = [x_1,\dots,x_n]$ generowaną przez wektory $x_1,\dots,x_n$.

/12/ Twierdzenie

Między współczynnikami korelacji zachodzi równość

$$r_{ik} = r_{ij}\, r_{jk} \qquad i < j < k$$

wtedy i tylko wtedy, gdy

$$Rz\left(x_k/[x_1,\dots,x_j]\right) = Rz\left(x_k/[x_j]\right) \qquad \left(i < j,\ j=2,\dots\right)$$

Dowód

Załóżmy, że wektory $x_1,\dots,x_j$ są zestandaryzowane.
Niech

$$Rz\left(x_k/[x_1,\dots,x_j]\right) = Rz\left(x_k/[x_j]\right) = r_{jk}x_j,$$

1/ Porównaj rozdział I, twierdzenie /47/.

przy czym ostatnia równość ma miejsce wobec /5/.

Stąd

$$x_k - r_{jk}x_j \perp x_i \qquad \text{dla } i < j \quad \left(\text{z twierdzenia /47/}\right)$$

a więc

$$\left(x_k - r_{jk}x_j \middle| x_i\right) = 0$$

i z własności iloczynu skalarnego otrzymujemy

$$\left(x_k \middle| x_i\right) - r_{jk}\left(x_j \middle| x_i\right) = 0,$$

co wobec przyjętych oznaczeń można zapisać

$$r_{ik} = r_{ij}\, r_{jk}.$$

Na odwrót, jeśli ta ostatnia równość jest spełniona, to dla każdego $i < j$

$$r_{jk}\left(x_j \middle| x_i\right) = \left(x_k \middle| x_i\right),$$

a stąd wynika, że

$$x_k - r_{jk}\, x_j \perp x_i$$

co implikuje /wobec jednoznaczności rzutu prostopadłego/

$$r_{jk}\, x_j = Rz\left(x_k / [x_j]\right) = Rz\left(x_k / [x_1, \ldots, x_j]\right)$$

i tym samym twierdzenie zostało udowodnione.

Ważne z punktu widzenia konstrukcji przykładów i numerycznych eksperymentów dotyczących np. wyboru zmiennych do modelu, jest następujące zagadnienie: jak mając zadaną macierz współczynników korelacji wyznaczyć wektory realizacji zmiennych losowych realizujące tę macierz ?. Odpowiedź na to pytanie daje

/13/ Twierdzenie

Jeśli macierz R stopnia p jest nieujemnie określona to można ją przedstawić w postaci iloczynu

$$R = C'C\ ,$$

gdzie C jest macierzą trójkątną górną stopnia p.

Dowód twierdzenia i wzory na elementy macierzy C można znaleźć np. w pracy: F.R. Gantmacher[9] s. 228 i dalsze. Wobec twierdzenia /11/ niniejszego rozdziału szukanymi wektorami realizacji zmiennych losowych są kolumny macierzy C.

Również, ze względu na eksperymenty związane z wyborem zmiennych do modelu, ważna jest możliwość realizacji macierzy współczynników korelacji o samych równych elementach, tzn. macierzy, w której $r_{ij} = a$ dla $i \neq j$. Macierz taka jest nieujemnie określona dla dowolnego $0 \leqslant a \leqslant 1$. Wynika to z równości[1]

/14/ $$\det R = \det \begin{bmatrix} 1 & a & & a \\ a & 1 & & a \\ \cdots & \cdots & \cdots & \cdots \\ a & a & & 1 \end{bmatrix} = (1-a)^{n-1}\left[(n-1)\,a + 1\right]$$

gdzie n oznacza stopień macierzy R.

/15/ <u>Lemat</u>

Jeśli $-1 \leqslant a < 0$, to macierz R jest nieujemnie określona tylko dla

$$n \leqslant 1 - \frac{1}{a}\,.$$

<u>Dowód</u>

Aby macierz R była nieujemnie określona musi być spełniona nierówność

$$(n-1)\,a + 1 \geqslant 0,$$

co kończy dowód lematu.

1/ Porównaj: Z. Hellwig [14] s. 234.

2/ Porównaj rozdział I, twierdzenie /83/.

W powyższych rozważaniach abstrahowaliśmy od wymiaru wektorów realizacji zmiennych losowych. Przechodząc do tej kwestii sformułujemy następujący, łatwy do udowodnienia

/16/ Lemat

W przestrzeni k-wymiarowej istnieje co najwyżej k+1 wektorów takich, że kąt między dowolnymi dwoma jest taki sam.

Problem istnienia skończenie wymiarowych wektorów realizujących daną macierz R stopnia n rozstrzyga

/17/ Twierdzenie

Jeśli macierz R stopnia n o samych równych elementach jest macierzą współczynników korelacji, przy czym $0 \leqslant a \leqslant 1$ lub dla $0 < a \leqslant -1$ $n \leqslant 1 - \frac{1}{a}$, to wektory realizacji $x_1, \ldots, x_n$, których współczynniki korelacji są elementami macierzy R są co najmniej n-1 wymiarowe.

Dowód

Elementy macierzy R są cosinusami kątów między wektorami $x_1, \ldots, x_n$ /por. wzór /5//. Stąd i z lematu /16/ wynika teza twierdzenia.

§ 3. Warunkowa wartość oczekiwana.

Z twierdzenia /47/ rozdziału I wynika, że każdą zmienną losową y należącą do przestrzeni Hilberta L_2 można przedstawić w postaci

$$y = y' + y'',$$

przy czym y'' jest rzutem prostopadłym wektora y na pewną podprzestrzeń M, a y' jest wektorem ortogonalnym do M. Jeśli podprzestrzeń M jest generowana przez układ $\{x_1, \ldots, x_n\}$, to

$$y'' = Rz(y/M) = Rz\left(y/\left[x_1, \ldots, x_n\right]\right).$$

/18/ Twierdzenie

Jeśli zmienne losowe $y, x_1, \ldots, x_n$ mają rozkłady normalne, to rzut prostopadły y na $\left[x_1, \ldots, x_n\right]$ jest równy warunkowej wartości oczekiwanej $E\left(y/x_1, \ldots, x_n\right)$ [1].

Dowód

Zmienna

$$y' = y - Rz\left(y/\left[x_1, \ldots, x_n\right]\right) = y - \sum_{j=1}^{n} a_j x_j$$

---

[1] Ukośna kreska w $E(\ /\ )$ oznacza warunkową wartość oczekiwaną natomiast w $Rz(x/[\ ])$ rzut x na $[\ ]$.

/por. rozdział /, wzór /47'// jest zmienną losową o rozkładzie normalnym ortogonalną do każdego x_i, stąd /wobec lematu /53/ z rozdziału I/ y' nie zależy od żadnego x_i, a więc

$$E\left(y'/\ x_1,\dots,x_n\right) = E\left(y'\right) = 0 =$$
$$= E\left(y - \sum_{j=1}^{n} a_j x_j\ /\ x_1,\dots,x_n\right) =$$
$$= E\left(y\ /\ x_1,\dots,x_n\right) - E\left(\sum_{j=1}^{n} a_j x_j\ /\ x_1,\dots,x_n\right).$$

Stąd wynika

$$E\left(y\ /\ x_1,\dots,x_n\right) = E\left(\sum_{j=1}^{n} a_j x_j\ /\ x_1,\dots,x_n\right) =$$
$$= \sum_{j=1}^{n} a_j x_j = Rz\left(y/\left[x_1,\dots,x_n\right]\right),$$

co też i należało udowodnić.

Można wykazać /por. M. Loeve [25], s. 485 i dalsze/, że dowolnej rodzinie zmiennych losowych o skończonych wariancjach odpowiada rodzina zmiennych losowych normalnych o tych samych wariancjach. Stąd rzut prostopadły można zawsze rozpatrywać jako warunkową wartość oczekiwaną w odpowiednio dobranych rodzinach zmiennych losowych o rozkładach normalnych. Rekapitulując, rzut zmiennej y na podprzestrzeń $\left[x_1,\dots,x_n\right]$ można zawsze uważać za regresję zmiennej y względem zmiennych $x_1,\dots,x_n$.

§ 4. Współczynnik rozrzutu.

Zdefiniujmy za M. Fiszem [6] wariancję uogólnioną ϑ zmiennych losowych $x_1, \dots, x_n$, jako

$$\vartheta = \det Q(x_1, \dots, x_n),$$

gdzie

$$Q(x_1, \dots, x_n) = \begin{bmatrix} Cov(x_1, x_1) & & Cov(x_1, x_n) \\ \dots & \dots & \dots \\ Cov(x_n, x_1) & & Cov(x_n, x_n) \end{bmatrix} \qquad /19/$$

jest macierzą kowariancji.

Jak wykazaliśmy, $Cov(x_i, x_j) = (x_i | x_j)$, a więc wariancja uogólniona, jest równa wyznacznikowi macierzy Grama:

$$\vartheta = \det G(x_1, \dots, x_n).$$

Można łatwo udowodnić[1/], że

$$\vartheta = Var(x_1) \qquad Var(x_n) \det R, \qquad /20/$$

1/ Porównaj np.: H. Cramer [3] s. 325.

gdzie R jest macierzą współczynników korelacji, tzn.

$$R = \left(r(x_i, x_j)\right).$$

/21/ <u>Definicja</u>

Współczynnikiem rozrzutu /rozsiewu [1]/ nazywamy wielkość określoną wzorem

$$\mathfrak{R} = \sqrt{\det R}$$

Do wykazania własności współczynnika rozrzutu wykorzystamy własności macierzy i wyznacznika Grama.

Standaryzacja nie zmienia współczynników korelacji, a więc wobec /19/ macierz współczynników korelacji jest macierzą Grama:

/22/ $$R = G(x_1, \ldots, x_n),$$

wynikają stąd i z twierdzenia /72/ rozdziału I następujące własności współczynnika $\mathfrak{R}$.

1/ Porównaj np.: H. Cramer[3]s. 372 i M. Fisz[6]s. 131.

/23/ Twierdzenie

Wartości współczynnika $\mathfrak{R}$ należą do przedziału domkniętego $[0,1]$.

Dowód

Macierz R jest nieujemnie określona. Stąd i z geometrycznej interpretacji wyznacznika Grama /por. twierdzenie /72/ rozdziału I / i założenia, że wektory $\{x_1,\ldots,x_n\}$ są zestandaryzowane, wynika że $\mathfrak{R}$ może być traktowany jako objętość "równoległościanu" o krawędziach, których długości są równe jeden, co kończy dowód twierdzenia.

/24/ Twierdzenie

Jeśli dla dowolnej pary x_i, x_j $(i \neq j)$ z układu wektorów $U = \{x_1,\ldots,x_n\}$ zachodzi równość

$$|r(x_i,x_j)| = 1$$

to

$$\mathfrak{R} = 0.$$

Dowód

Jeśli $|r(x_i,x_j)| = 1$, to wektory x_i,x_j są zależne liniowo; w takim razie $\mathfrak{R}$ jest objętością co najwyżej $n-1$ wymiarowego równoległościanu w przestrzeni n-wymiarowej, a więc $\mathfrak{R} = 0$.

/25/ Twierdzenie

Zmienne losowe $x_1,\ldots,x_n$ są parami nieskorelowane wtedy i tylko wtedy, gdy

$$\mathfrak{R}\ (x_1,\ldots,x_n) = 1.$$

Dowód

Uwzględniając fakt, że zmienne $x_1,\ldots,x_n$ są zestandaryzowane, $\mathfrak{R}(x_1,\ldots,x_n)$ jest objętością n-wymiarowego prostopadłościanu o wszystkich krawędziach równych jedności, co kończy dowód.

/26/ Twierdzenie

Jeśli

$$|r(x_i,x_j)| \longrightarrow 1 \qquad \text{dla dowolnych } i \neq j,$$

to

$$\mathfrak{R}\,(x_1,\ldots,x_n) \longrightarrow 0.$$

Dowód

Z uogólnionej nierówności Hadamarda i nierówności /77/ rozdziału I

$$0 \leqslant \mathfrak{R} = \sqrt{g(x_1,\ldots,x_n)} \leqslant \sqrt{g(x_1,\ldots,x_{i-1},x_{i+1},\ldots,x_{j-1},x_{j+1},\ldots,x_n)}\,\sqrt{g(x_i,x_j)}$$

stąd teza twierdzenia wynika z nierówności

$$g(x_i,x_j) = 1 - |r(x_i,x_j)|^2$$

/27/ Twierdzenie

Prawdopodobieństwo tego, że pomiędzy zmiennymi losowymi $x_1,\ldots,x_n$ zachodzi przynajmniej jeden związek liniowy jest równe jedności wtedy i tylko wtedy, gdy

$$\mathfrak{R}\,(x_1,\ldots,x_n) = 0.$$

Dowód tego twierdzenia, wobec założenia, że zmienne $x_1,\dots,x_n$ są zestandaryzowane, jest analogiczny do dowodu twierdzenia w pracy: M. Fisz [6] s, 72, sformułowanego i udowodnionego dla wariancji uogólnionej.

§ 5. Macierz informacyjna i współczynnik korelacji wielowymiarowej.

W teorii dotyczącej planowania eksperymentów macierzą informacyjną zmiennych losowych $x_1, \ldots, x_n$ nazywa się macierz $\mathcal{Y}$ zdefiniowaną[1/] jak następuje

$$\mathcal{Y}(x_1, \ldots, x_p) = \begin{bmatrix} E(x_1 x_1) & & E(x_1 x_p) \\ & & \\ E(x_p x_1) & \ldots & E(x_p x_p) \end{bmatrix}$$

Jeżeli założymy, że zmienne $x_1, \ldots, x_p$ należą do L_2, to macierz informacyjna jest macierzą Grama:

$$\mathcal{Y}(x_1, \ldots, x_p) = G(x_1, \ldots, x_p).$$

Stąd normę wektora losowego x można przy użyciu macierzy określić następująco:

$$\|x\| = (\mathrm{Tr}\, \mathcal{Y})^{\frac{1}{2}} ,$$

1/ Porównaj np.: [30]

gdzie

$$\operatorname{Tr} \mathcal{Y} = \sum_{i=1}^{p} E|x_i|^2$$

Wyznacznik macierzy informacyjnej ma pewną wadę jako miara informacji niesionej przez układ wektorów, mianowicie jego wartość zależy monotonicznie od ilości wektorów. Wykazuje to następujące

/28/ Twierdzenie

Między wyznacznikami macierzy informacyjnej zachodzi nierówność

$$f_1 = \det \mathcal{Y}(x_1,\dots,x_p) \geqslant f_2 = \det \mathcal{Y}(x_1,\dots,x_p,x_{p+1}) \,.$$

Dowód

Wyznacznik macierzy $\mathcal{Y}$ może być interpretowany jako kwadrat objętości równoległościanu rozpiętego na odpowiednich wektorach, stąd dla wektorów $x_1,\dots,x_p,x_{p+1}$ zestandaryzowanych : $0 \leqslant f_1 \leqslant 1$

$$0 \leqslant \| \operatorname{Rz} (x_{p+1} / [x_1,\dots,x_p])\| \leqslant 1$$

i

$$\| x_{p+1} - \operatorname{Rz} (x_{p+1} / [x_1,\dots,x_p])\| \leqslant 1 \,,$$

a stąd

$$f_1 \geqslant f_1 \left\| x_{p+1} - Rz\left(x_{p+1} / [x_1,\dots,x_p]\right)\right\| ,$$

co kończy dowód.

Interpretację współczynnika korelacji wielowymiarowej ϱ określa

/29/ Twierdzenie

Współczynnik ϱ jest cosinusem kąta między zmienną losową y a podprzestrzenią liniową generowaną przez zmienne $x_1,\dots,x_p$, tzn. kąta jaki tworzy y z jej rzutem prostopadłym na podprzestrzeń $[x_1,\dots,x_p]$.

Dowód

Jak wykazał Z. Pawłowski w [32] zachodzi następująca równość

$$\varrho = \sqrt{1 - \frac{\det R(x_1,\dots,x_p,y)}{\det R(x_1,\dots,x_p)}}$$

przy czym ułamek pod pierwiastkiem jest kwadratem odległości zmiennej y od podprzestrzeni $[x_1,\ldots,x_p]$ /por. twierdzenie /68/, rozdział I/. Stąd ϱ jest długością rzutu wektora o długości jeden na podprzestrzeń $[x_1,\ldots,x_p]$, a więc twierdzenie zostało dowiedzione.

§ 6. Metoda najmniejszych kwadratów i estymator Gaussa-Markowa.

Niech

$$\left\{y_1,\dots,y_m\right\}$$

będzie zbiorem n- wymiarowych wektorów losowych. Każdy element y_i ma n składowych y_{ij} $(j = 1,\dots,n)$, z których każda jest zmienną losową o skończonej wariancji.

/30/ Definicja[1/]

Przestrzenią Hilberta H n-wymiarowych wektorów losowych nazywamy zbiór wektorów, których składowe są kombinacjami liniowymi składowych wektorów y_i.

Każdy wektor tej przestrzeni ma postać

$$y = A_1y_1 + A_2y_2 + \dots + A_my_m,$$

gdzie A_i jest macierzą stopnia n.

1/ Konstrukcja tej przestrzeni Hilberta jest bardziej złożona od poprzednio omawianych. Kompensuje to jednak proste sformułowanie problemów estymacji.

Iloczyn skalarny wektorów przestrzeni H jest postaci:

$$(x|z) = E\left(\sum_{i=1}^{n} x_i z_i\right), \qquad (x,z \in H).$$

Stąd norma elementu $x \in H$

$$\|x\| = \left(\mathrm{Tr}\ E(xx')\right)^{\frac{1}{2}},$$

gdzie

$$E(xx') = \begin{bmatrix} E(x_1x_1) & E(x_1x_n) \\ \dots\dots\dots & \dots\dots\dots \\ E(x_nx_1) & E(x_nx_n) \end{bmatrix}$$

Jeżeli składowe wektorów losowych mają wartości oczekiwane równe zero, to macierz $E(xx')$ jest po prostu macierzą kowariancji, a z definicji iloczynu skalarnego wynika, że jest ona macierzą Grama zmiennych losowych $x_1,\dots,x_n$.

Załóżmy, że dysponujemy realizacją wektora losowego y, która jest m-elementowym ciągiem liczb rzeczywistych i realizacjami wektorów losowych $x_1,\dots,x_n$. Załóżmy ponadto, że zmienna y jest liniową kombinacją zmiennych $x_1,\dots,x_n$, wtedy oznaczając przez A n-wymiarowy $(n < m)$ wektor współczynników tej kombinacji otrzymamy

/31/ $$y = WA,$$

gdzie W jest macierzą $(m \times n)$, której kolumnami są realizacje wektorów $x_1, \dots, x_n$. Wartość wyrazu wolnego kombinacji liniowej wyrażającej y poprzez zmienne x_i otrzymamy jeśli do macierzy W dopiszemy kolumnę złożoną z samych jedynek[1/].

Ponieważ $n < m$, nie można na ogół wyznaczyć wektora A spełniającego dokładnie równość /31/. O szukanym wektorze A założymy więc, że minimalizuje normę

$$\| y - WA \| ,$$

tzn. minimalizuje odległość wektora y od podprzestrzeni generowanej przez kolumny macierzy W.

/32/ <u>Twierdzenie</u>

Jeśli macierz W jest macierzą o niezależnych liniowo kolumnach, a wektor y jest m-wymiarowy, to istnieje jedyny n-wymiarowy wektor $\hat{A}$ minimalizujący

$$\| y - WA \| ,$$

przy czym

1/ Porównaj np.: Z. Hellwig [13] s. 190.

/33/ $$\hat{A} = (W'W)^{-1} W'y.$$

Dowód

Istnienie wektora $\hat{A}$ i jego jednoznaczność wynikają z twierdzenia o rzutowaniu /twierdzenie /47/ rozdział I/. Ponadto macierz Grama odpowiadająca kolumnom macierzy W ma postać $W'W$.

Stąd

$$W'W\hat{A} = W'y,$$

a ponieważ kolumny macierzy W są liniowo niezależne, macierz $W'W$ ma odwrotną, stąd wynika równość /33/.

Załóżmy, że w wyniku doświadczeń otrzymaliśmy m-wymiarowy wektor realizacji y i że

$$y = WA + \varepsilon ,$$

gdzie: W jest znaną z eksperymentu macierzą;

A jest nieznanym n-wymiarowym wektorem;

ε jest m-wymiarowym wektorem losowym takim, że

$$E(\varepsilon) = \Theta$$

i znana jest macierz kowariancji

$$Q = E\left(\varepsilon\,\varepsilon'\right),$$

o której zakładamy, że jest dodatnio określona.
Wektor ε interpretujemy jako błędy pomiarów. Naszym zadaniem jest znalezienie estymatora dla A.

/34/ <u>Twierdzenie /Gauss-Markow/</u>

Liniowym nieobciążonym najefektywniejszym estymatorem dla A jest

$$\hat{A} = \left(W'Q^{-1}W\right)^{-1} W'Q^{-1}y,$$

a odpowiadająca mu kowariancja błędu ma postać

$$E\left[\left(\hat{A} - A\right)\left(\hat{A} - A\right)'\right] = \left(W'Q^{-1}W\right)^{-1}$$

Dowód tego twierdzenia można znaleźć np. w: D. Luenberger[26],

§ 7. Zależność współczynników regresji od współczynnika korelacji.

Przedstawione rezultaty wykorzystamy w rozważaniach dotyczących zależności współczynników korelacji między zmiennymi a współczynnikami regresji

$$Y = a_1X_1 + \dots + a_nX_n.$$

Problematyka ta jest szczególnie istotna, gdy jedną ze zmiennych jest czas. Ograniczmy nasze rozważania narazie do przypadku, w którym mamy do czynienia z płaszczyzną regresji

$$y = a_1X_1 + a_2X_2.$$

I tak załóżmy, że o współczynniku korelacji $r(Y,X_1)$ wiadomo, że jest większy od zera. Co można powiedzieć wtedy o współczynniku a_1 ? Odpowiedź na to pytanie daje następujące

/35/ Twierdzenie

Dla zestandaryzowanych wektorów realizacji y,x_1x_2

jeśli $r(y,x_1) > 0$ i $|r(x_1,x_2)|$ jest różne od jedności, to $a_1 > 0$ wtedy i tylko wtedy, gdy

$$r(y,x_1) > r(y,x_2)\, r(x_1,x_2).$$

<u>Dowód</u>

Współczynniki a_1 i a_2 wyznaczamy z układu równań

$$(y - a_1x_1 - a_2x_2 \mid x_1) = 0$$
$$(y - a_1x_1 - a_2x_2 \mid x_2) = 0 \quad ,$$

który można zapisać w postaci

$$a_1(x_1|x_1) + a_2(x_1|x_2) = (y|x_1)$$
$$a_1(x_1|x_2) + a_2(x_2|x_2) = (y|x_2)\ .$$

Stąd

$$a_1 = \frac{\begin{vmatrix} (y|x_1) & (x_1|x_2) \\ (y|x_2) & (y_2|x_2) \end{vmatrix}}{\begin{vmatrix} (x_1|x_1) & (x_1|x_2) \\ (x_2|x_1) & (x_2|x_2) \end{vmatrix}}$$

Wyznacznik w mianowniku jest zawsze dodatni, ponieważ założyliśmy, że $|r(x_1,x_2)| \neq 1$. Stąd też $a_1 > 0$ wtedy i tylko wtedy, gdy wyrażenie w liczniku jest dodatnie, tzn. gdy

/36/ $$(y|x_1)\ (x_2|x_2) > (x_1|x_2)\ (y|x_2).$$

Wobec założenia, że wektory y, x_1, x_2 są zestandaryzowane,

$$\begin{aligned}(x_2|x_2) &= 1\\ (y|x_1) &= r(y,x_1)\\ (x_1|x_2) &= r(x_1,x_2)\\ (y|x_2) &= r(y,x_2)\ ;\end{aligned}$$

podstawiając te równości do /36/ otrzymujemy

$$r(y,x_1) > r(y,x_2)\ r(x_1,x_2),$$

co kończy dowód.

Uogólnieniem twierdzenia /35/ na przypadek, w którym mamy do czynienia z większą liczbą zmiennych jest następujące

/37/ Twierdzenie

Niech $y, x_1, \dots, x_n$ oznaczają zestandaryzowane wektory realizacji

$$z = Rz\left(y \,/\, [x_2, \dots, x_n]\right).$$

Nie zmniejszając ogólności rozważań założymy ponadto, że

$$r(y, x_1) > 0 \qquad i \qquad |r(x_1, z)| \neq 1.$$

Przy tych założeniach współczynnik a_1 jest większy od zera wtedy i tylko wtedy, gdy

$$r(y, x_1) > r(y, z) \;\; r(x_1, z).$$

Dowód tego twierdzenia jest natychmiastowy jeśli wykorzystamy twierdzenie /7/ z rozdziału V i wektor y zrzutujemy na $[x_1, \dots, x_n]$ rzutując go na x_1 i na $[x_2, \dots, x_n]$ a następnie na podprzestrzeń generowaną przez x_1 i z. Rolę analogiczną do x_2 występującego w sformułowaniu twierdzenia /35/ spełnia teraz z.

R o z d z i a ł III

ESTYMACJA CHARAKTERYSTYK ROZKŁADÓW W PRZESTRZENI HILBERTA L_2

W rozdziale niniejszym zajmiemy się estymacją pewnych charaketrystyk rozkładów w przestrzeni Hilberta L_2 zmiennych losowych, takich jak iloczyn skalarny, norma, odległość, macierz Grama i charakterystyk za ich pomocą definiowanych. Zostaną przedstawione niektóre własności tych estymatorów. Czynimy to z tego względu, że na ogół w praktyce mamy do czynienia z realizacjami zmiennych losowych a nie z samymi zmiennymi, stąd też wynika konieczność operowania iloczynem skalarnym, normą, itp. jako wartościami próbkowymi odpowiednich statystyk.

Przedstawione rezultaty nie wyczerpują wszystkich zagadnień związanych z omawianą tematyką. Bogactwo i różnorodność problemów, które zostały tylko naszkicowane zasługuje na osobne opracowanie. Dotyczy to zagadnień związanych z rozkładami estymatorów.

§ 1. Przestrzeń realizacji zmiennych losowych.

W rozdziale I była rozpatrywana przestrzeń Hilberta zmiennych losowych o wartościach oczekiwanych równych zero i skończonych wariancjach. Przestrzeń ta była oznaczona przez L_2.

Niech wektory

$$x_1 = (x_{11}, x_{12}, \ldots, x_{1n})$$

$$x_2 = (x_{21}, x_{22}, \ldots, x_{2n})$$

$$\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots$$

$$x_p = (x_{p1}, x_{p2}, \ldots, x_{pn})$$

oznaczają realizacje zmiennych losowych $X_1, \ldots, X_p \in L_2$. Składowe wektora x_i są niezależnymi realizacjami zmiennej X_i.

Wprowadźmy iloczyn skalarny takich wektorów równością

$$/1/ \quad (x_i | x_j)_n = \frac{1}{n-1} \sum_{k=1}^{n} \left(x_{ik} - E(X_i)\right) \left(x_{jk} - E(X_j)\right) = \frac{1}{n-1} \sum_{k=1}^{n} x_{ik} x_{jk}$$

Dla ustalonego n zbiór wektorów $\{x_1, \ldots, x_p, \ldots\}$ z dodawaniem i iloczynem przez skalar po składowych, i iloczynem skalarnym określonym przez /1/ jest przestrzenią Hilberta, którą oznaczymy przez R^n.

Normę w R^n określa się

/2/ $$\|x_i\|_n = \sqrt{\frac{1}{n-1} \sum_{k=1}^{n} x_{ik}^2} = s(x_i) = s_i,$$

a stąd wzór na odległość elementów z R^n jest postaci

/3/ $$d_n(x_i, x_j) = \|x_i - x_j\|_n = \sqrt{\frac{1}{n-1} \sum_{k=1}^{n} (x_{ik} - x_{jk})^2} =$$

$$= \sqrt{s_i^2 + s_j^2 - 2(x_i | x_j)_n}$$

/4/ Uwaga

Iloczyn, norma i odległość zadane odpowiednio wzorami /1/, /2/ i /3/ jest iloczynem skalarnym, normą i odległością z próby.

Macierz Grama G_n wektorów $x_1, \ldots, x_p \in R^n$ ma postać

/5/ $$G_n(x_1,\dots,x_p) = \frac{1}{n-1}\begin{bmatrix} (x_1|x_1) & \dots & (x_1|x_p) \\ \dots & \dots & \dots \\ (x_p|x_1) & & (x_p|x_p) \end{bmatrix}$$

przy czym $(x_i|x_j)$ $(i,j = 1,\dots,p)$ jest iloczynem skalarnym wektorów x_i, x_j rozpatrywanych jako elementy przestrzeni E^n. Wobec /1/ macierz Grama G_n jest macierzą kowariancji z próby a det G_n wariancją uogólnioną z próby.

Macierz

/6/ $$R = (r_{ij}),$$

gdzie

$$r_{ij} = \frac{(x_i|x_j)\, n}{s_i\, s_j}$$

jest macierzą współczynników korelacji w R^n.
Wyrażenie $\mathcal{R} = \sqrt{\det R}$ jest współczynnikiem rozrzutu w R^n.

§ 2. Własności estymatorów.

W dalszych rozważaniach wykorzystamy twierdzenie dotyczące zmiennych losowych, które można znaleźć wraz z dowodem w pracy I.I. Gichman i A. W. Skorochod [10] s. 67, gdzie jest ono sformułowane dla funkcji mierzalnych; oczywiście zmienne losowe są funkcjami mierzalnymi.

/7/ Twierdzenie

Jeśli ciągi zmiennych losowych $\left(X_n^{(1)}\right), \ldots, \left(X_n^{(s)}\right)$ są zbieżne stochastycznie do zmiennych losowych odpowiednio $g_1, \ldots, g_s$ a $\varphi : R^s \to R$ jest ciągła, to

$$P - \lim \varphi\left(X_n^{(1)}, \ldots, X_n^{(s)}\right) = \varphi\left(g_1, \ldots, g_s\right)$$

Rozważmy następujące statystyki

/8/ $$c_{ij} = \frac{1}{n} \sum_{k=1}^{n} X_{ik} X_{jk} = \frac{1}{n} \Sigma_{ij} ;$$

/9/ $$s_i^2 = \frac{1}{n} \sum_{k=1}^{n} X_{ik}^2 ,$$

gdzie $X_{ik} \in L_2$ ma taki rozkład jak X_i.

/10/ $$S_i = \sqrt{S_i^2} \;;$$

/11/ $$r_{ij} = \frac{C_{ij}}{S_i S_j}$$

/12/ Uwaga

Statystyka /8/ jest zgodnym estymtorem kowariancji /por. A.S. Goldberger [11] s. 165/ $Cov(X_i, X_j)$, a wobec zależności / por. H. Cramer [3] s. 344/

/13/ $$E(C_{ij}) = \frac{n-1}{n} \, Cov(X_i, X_j)$$

statystyka

$$\frac{n}{n-1} \, C_{ij}$$

jest zgodnym i nieobciążonym estymatorem $Cov(X_i, X_j)$.

/14/ Uwaga

Statystyka /9/ jest zgodnym /por. A.S. Goldberger [11] s. 165/

i nieobciążonym estymatorem wariancji $Var(X_i)$.

Nieobciążoność wynika z faktu, że wartości oczekiwane zmiennych losowych X_{ik} są znane i równe wartości oczekiwanej $E(X_i)$. Ponadto jeśli X_i ma rozkład normalny to S^2 jest estymatorem najefektywniejszym[1/] $Var(X_i)$.

/15/ Uwaga

Wnioskiem z twierdzenia /7/ jest fakt, że S_i jest estymatorem zgodnym odchylenia standardowego $\sigma(X_i)$. Ponadto S_i jest estymatorem obciążonym. Natomiast w przypadku, gdy zmienna losowa $X_i \in L_2$ ma rozkład normalny i $E(X_i)$ jest znana /co w naszym przypadku ma miejsce/, to estymatorem nieobciążonym i zgodnym parametru $\sigma(X_i)$ jest statystyka

$$\hat{S}_i = \frac{\sqrt{n}\,\Gamma\left(\frac{1}{2}\,n\right)}{\sqrt{2}\,\Gamma\left(\frac{1}{2}\,(n+1)\right)}\,S_i\,,$$

gdzie Γ oznacza funkcję gamma.

Wynika to z twierdzenia /7/ i faktu, że przy wymienionych założeniach dotyczących zmiennych losowych X_i zachodzi równość[2/]

$$E(S_i) = \frac{\sqrt{2}\,\Gamma\left(\frac{1}{2}\,(n+1)\right)}{\sqrt{n}\,\Gamma\left(\frac{1}{2}\,n\right)}\,\sigma(X_i)$$

1/ Porównaj: Korn-Korn [23] s. 545.

2/ Porównaj np.: M. Fisz [6] s. 482.

/16/ Uwaga

Statystyka /11/ jest wobec twierdzenia /7/ estymatorem zgodnym współczynnika korelacji $r(X_i, X_j)$ określonego w L_2 i jak wykazał H. Cramer [3] s. 427, jest nieobciążona.

/17/ Twierdzenie

Iloczyn skalarny określony równością

$$(x_i | x_j)_n = \frac{1}{n} \sum_{k=1}^{n} x_{ik} x_{jk}$$

jest estymatorem zgodnym iloczynu skalarnego w L_2. Wynika to z własności estymatora /8/. Z wzoru /13/ wynika natomiast, że statystyka

$$(x_i | x_j)_n^* = \frac{1}{n-1} \sum_{k=1}^{n} x_{ik} x_{jk} = \frac{n}{n-1} (x_i | x_j)_n$$

jest zgodnym i nieobciążonym estymatorem iloczynu skalarnego w L_2, tzn.

$$P - \lim (x_i | x_j)_n = (x_i | x_j)$$

i

$$E\left((x_i|x_j)\right)_n = (x_i|x_j).$$

/18/ Twierdzenie

Norma określona wzorem

$$\|x_i\|_n = \sqrt{\frac{1}{n}\sum_{k=1}^{n} x_{ik}^2}$$

jest estymatorem zgodnym normy w L_2.

Wynika to z własności estymatora /10/ i z twierdzenia /7/. Z uwagi /15/ wynika ponadto, że istnieje zgodny i nieobciążony estymator normy w L_2. Natomiast kwadrat normy jest estymatorem zgodnym i nieobciążonym wobec założenia, że zmienne należące do L_2 mają znane wartości oczekiwane. Jeśli X_i ma rozkład normalny, to estymator kwadratu normy tej zmiennej jest ponadto najefektywniejszy. /porównaj z uwagą /14//.

/19/ Twierdzenie

Odległość zadana wzorem

$$d_n(x_i,x_j) = \frac{1}{n-1}\sum_{k=1}^{n}\left(x_{ik} - x_{jk}\right)^2$$

jest estymatorem zgodnym odległości w L_2.

Jest to wniosek z twierdzenia /18/ i twierdzenia /7/.

Macierz Grama $G_n(X_1,\dots,X_p)$ wektorów $X_1,\dots,X_p \in L_2$ /por. /5// jest macierzą o elementach c_{ij}. Ponieważ zgodnym i nieobciążonym estymatorem kowariancji w L_2 jest

$$\frac{n}{n-1}\, c_{ij},$$

to macierz

$$C_n = \left(\frac{n}{n-1}\, c_{ij}\right) = \frac{1}{n-1} \begin{bmatrix} \sum_{11} & \dots & \sum_{1p} \\ \dots & \dots & \dots \\ \sum_{p1} & \dots & \sum_{pp} \end{bmatrix}$$

jest zgodnym[1/] i nieobciążonym estymatorem macierzy Grama w L_2. Stąd i z twierdzenia /7/ otrzymujemy

/20/ <u>Twierdzenie</u>

Estymatorem zgodnym wyznacznika Grama $g(X_1,\dots,X_p)$ jest

---

1/ Zbieżność stochastyczną ciągu macierzy definiuje się jako zbieżność stochastyczną odpowiednich elementów macierzy; a wartością oczekiwaną macierzy, której elementami są zmienne losowe jest macierz wartości oczekiwanych.

wyznacznik

/21/ $$V_n = \det C_n .$$

Ponadto mają miejsce równości

/22/ $$V_n = \left(\frac{n}{n-1}\right)^p \det\left(c_{ij}\right) = \frac{1}{(n-1)^p} \det \begin{bmatrix} \Sigma_{11} & & \Sigma_{1p} \\ \cdots & \cdots & \cdots \\ \Sigma_{p1} & \cdots & \Sigma_{pp} \end{bmatrix}$$

Jak wykazaliśmy w § 4 rozdziału II, wyznacznik Grama w L_2 jest wariancją uogólnioną ϑ. Sformułujemy twierdzenie, które daje geometryczną interpretację estymatora wariancji uogólnionej ϑ będącej wobec powyższych rozważań estymatorem zgodnym i nieobciążonym ϑ.

/23/ Twierdzenie[1/]

Niech $x_1,\ldots,x_p$ oznaczają elementy przestrzeni R^n. Kwadrat objętości v równoległościanu rozpiętego na wektorach $x_1,\ldots,x_p$ jest równy

---

1/ Twierdzenie to udowodnił T. Anderson w pracy [1] s. 93 za pomocą innych środków.

$$v^2 = \det\left((x_i|x_j)_n\right) = \frac{1}{(n-1)^p} \det\left((x_i|x_j)\right),$$

gdzie $(x_i|x_j)$ oznacza jak poprzednio iloczyn skalarny w E^n.

Dowód

Wobec równości /1/ i /5/ macierz $\left((x_i|x_j)_n\right)$ jest macierzą Grama w R^n, a więc teza twierdzenia wynika natychmiast z geometrycznej interpretacji macierzy Grama podanej w twierdzeniu /72/ rozdziału I.

/24/ Uwaga

Wariancja uogólniona $\mathscr{V}$ może być interpretowana jako kwadrat objętości "równoległościanu" rozpiętego na zmiennych $X_1,\dots,X_p$, traktowanych jako wektory przestrzeni L_2. Stąd i z /7/ statystykę $\sqrt{V_n}$ można uważać za estymator zgodny objętości.

Twierdzenie /23/ wykorzystamy między innymi w wyborze zmiennych do modelu liniowego, co będzie przedmiotem rozważań przedstawionych w rozdziale V.

R o z d z i a ł IV

WYBÓR CECH CHARAKTERYSTYCZNYCH

§ 1. Uwagi wstępne.

Zastosowanie matematyki w ekonomii wymaga swoistego przekładu zjawisk i obiektów ekonomicznych na pojęcia używane w matematyce. I tak, załóżmy że każdemu z obiektów ekonomicznych rozważanego zbioru przyporządkujemy jest n-elementowy ciąg liczb będących wartościami wybranych a priori cech tych obiektów. Takie postawienie zagadnienia wymaga przeprowadzenia rozważań dotyczących właściwego wyboru cech nie poprzestając na ustaleniach, które pozwala nam dokonać intuicja i doświadczenie. Naszym zdaniem, w wyborze cech powinny decydować następujące względy.

/1/ Ilość wybranych cech powinna być możliwie mała. Wymaga tego ekonomiczne podejście do zagadnienia: badania, których celem jest ustalenie wartości cech są często bardzo kosztowne.

/2/ Nie należy wybierać cech silnie ze sobą związanych. Taki wybór powodowałby zbędne powielanie informacji i zwiększanie błędów /zanieczyszczanie informacji/.

/3/ Każda z wybranych cech powinna charaketryzować się dużą zmiennością przy przejściu od obiektu do obiektu. Nic istotnego do opisu obiektów nie wnoszą cechy, które nie zmieniają się lub zmieniają się bardzo nieznacznie przy zmianie obiektu. Nie separują one obiektów, a zatem nie są przydatne jako cechy je charakteryzujące.

Wybierając optymalny układ cech spowodujemy, że badanym obiektom będą przyporządkowane w sposób wzajemnie jednoznaczny punkty w przestrzeni o ustalonym wymiarze, a więc twory stricte matematyczne. Taka transformacja obiektów w zbiór punktów pozwala w dalszych badanich np. na rozdyskryminowanie ich, wybranie spośród nich obiektu "najlepszego" w sensie ustalonego kryterium, itp.
Krótko mówiąc - umożliwia opis i analizę zbioru obiektów za pomocą metod matematycznych.

Zanim sformułujemy własne propozycje wyboru cech, przedstawimy. posługując się przestrzenią Hilberta, metodę, która dotąd służyła temu celowi: tzw. metodę czynników głównych, której twórcą jest H. Hotteling.

W ciągu dalszym cechy charakteryzujące obiekty będziemy traktowali jako zmienne losowe. Przyczyną do tego skłaniającą jest sama istota cechy, na której wartości ma wpływ duża ilość niezdeterminowanych czynników, a także zakłócenia powodujące nierzetelność pomiarów.

§ 2. Metoda czynników głównych.

Załóżmy, że każdy z badanych obiektów jest opisany za pomocą p cech, których wartości będziemy uważali za realizacje p zmiennych losowych.

Zadaniem metody czynników głównych jest znalezienie takich kombinacji liniowych tych zmiennych, których wariancje są największe przy spełnieniu pewnych dodatkowych warunków nałożonych na współczynniki kombinacji. Postępowanie to jest uzasadnione faktem, że owe kombinacje o największych wariancjach najlepiej separują obiekty opisane zmiennymi losowymi, tzn. największa wariancja świadczy o tym, że wybrana kombinacja liniowa najsilniej wpływa na zróżnicowanie obiektów.

Jednym z możliwych sposobów[1/] wykładu analizy czynnikowej jest przedstawinie jej w terminach wartości własnych i wektorów własnych[2/].

Niech X będzie wektorem losowym należącym do przestrzeni Hilberta H_2 wektorów, których składowe są zmiennymi o skończonych wariancjach.

$$X = (X_1, \dots, X_p) \in H_2$$

1/ Porównaj np.: T. Anderson [1] s. 369 i dalsze.

2/ Porównaj rozdział I, definicja /25/.

oznacza p-wymiarowy wektor losowy, a

$$Q = \left(Cov\left(X_i, X_j\right)\right)$$

jego macierz kowariancji. Niech ponadto

$$B_1 = \left(b_1^1, \ldots, b_p^1\right)^{T}$$

oznacza p wymiarowy wektor kolumnowy taki, że

$$B_1' B_1 = 1.$$

Celem naszych poszukiwań są takie składowe wektora B_1 aby wariancja

$$Var\left(B_1' X\right) = E\left(B_1' X\right)^2 = B_1' CB_1 = \lambda_1$$

była największa. Jeśli wektor B_1 zostanie wyznaczony, następny krok polega na znalezieniu wektora B_2 ortogonalnego do B_1 spełniającego

$$B_2'B_2 = 1 ,$$

takiego, żeby z kolei wariancja kombinacji liniowej

$$Var(B_2' X) = \lambda_2$$

była największa.

Postępowanie jest zakończone jeśli znajdziemy p wektorów $B_1,\dots,B_p$ i p liczb $\lambda_1,\dots,\lambda_p$.

Można wykazać[1/], że wektory $B_1,\dots,B_p$ są wektorami własnymi a $\lambda_1,\dots,\lambda_p$ wartościami własnymi macierzy Q, przy czym składowe wektorów $B_1,\dots,B_p$ w analizie czynnikowej noszą nazwę ładunków czynnikowych. W ten sposób powstał układ p liniowo niezależnych takich wektorów, że wariancja kombinacji liniowej, której współczynnikami są składowe tych wektorów, jest maksymalna, a stąd owe p kombinacji ma własność najlepszego w sensie

---

1/ Porównaj np.: T. Anderson [4] s. 375.

wariancji separowania obiektów. Z dalszych rozważań wyklucza się te kombinacje liniowe, których wariancja jest mała.

Przedstawione postępowanie ma pewne mankamenty natury technicznej: błąd[1/] przy obliczaniu kolejnych wartości własnych szybko rośnie, ponadto wartości własne i wektory własne zależą od kowariancji a te z kolei od mian jednostek , w których są mierzone wartości zmiennych. Innego rodzaju wadą występująca w praktycznym zastosowaniu analizy czynnikowej polega na tym, że interpretacja czynników będących kombinacjami liniowymi zadanych zmiennych losowych jest bardzo kłopotliwa.

Stosowanie analizy czynnikowej jako sposobu eliminacji zmiennych ma tę niedogodność, że kwalifikowanie kombinacji liniowych jako mających małe wariancje jest arbitralna. Ponadto metoda ta stosowana w takim aspekcie nie uwzględnia stopnia skorelowania zmiennych.

Wobec powyższych uwag i zastrzeżeń dotyczących metody Hottelinga, proponujemy inne podejście do zagadnienia wyboru cech charakterystycznych realizujące postulaty omówione w paragrafie pierwszym.

1/ Porównaj: W.N. Faddiejewa [5] s. 124 i dalsze.

§ 3. Funkcja Ω wyboru układu charakterystycznego zmiennych losowych i jej własności.

Niech $\mathcal{U} = \{X_1,\ldots,X_p\}$ oznacza podzbiór przestrzeni L_2, w której ciąg (B_j) jest bazą.

Oznaczmy przez $A \in L_2$ zmienną postaci

$$A = \sum_i \frac{B_i}{\|B_i\|}$$

Na podzbiorach zbioru $\mathcal{U}$ określamy funkcję

$$\Omega: 2^{\mathcal{U}} \setminus \{\emptyset\} \longrightarrow R$$

/$2^{\mathcal{U}}$ oznacza rodzinę podzbiorów zbioru $\mathcal{U}$/ taką, że

/4/ $$\Omega(X_1,\ldots,X_m) = \frac{1}{m} \sum_{i=1}^{m} \omega_i ,$$

gdzie

$$/5/ \qquad \omega_i = \frac{\sqrt{1 - \left(\frac{(X_i | A)}{\|X_i\| \, \|A\|}\right)^2}}{\sum_{j=1}^{m} \frac{|(X_i | X_j)|}{\|X_i\| \, \|X_j\|}}$$

/6/ Definicja

Układem charakterystycznym zmiennych losowych nazywamy taki podzbiór zbioru $\mathcal{U}$, dla którego funkcja Ω przyjmuje wartość największą.

Poniższe własności funkcji Ω są uzasadnieniem roli jaką ona spełnia w wyborze układu charakterystycznego.

/7/ Własność

Funkcja Ω przyjmuje wartości z przedziału $[0,1]$.

Dowód

Z /2/ wynika, że $0 \leqslant \omega_i \leqslant 1$, ponieważ prawdziwa jest nierówność

$$(X|Y) \leqslant \|X\| \, \|Y\| ,$$

a więc największa wartość jaką może przyjąć ω_i jest jeden, stąd największa wartość sumy w /1/ jest równa m. Ponadto Ω jako suma nieujemnych składników jest nieujemna. Kończy to dowód.

/8/ Własność

Funkcja Ω przyjmuje wartość zero wtedy i tylko wtedy, gdy

$$\left|(X_i|A)\right| = \|X_i\|\,\|A\| \qquad \text{dla } i=1,\dots,m \quad ,$$

tzn. gdy $X_i = \alpha A$ /por. rozdział I/.

Dowód

Ω jako suma nieujemnych składników jest równa zero wtedy i tylko wtedy, gdy każdy z nich jest równy zero. To z kolei ma miejsce wtedy i tylko wtedy, gdy

$$\frac{\left|(X_i|A)\right|}{\|X_i\|\|A\|} = 1$$

dla każdego i, co kończy dowód.

/9/ Własność

Funkcja Ω przyjmuje wartość 1 wtedy i tylko wtedy, gdy

$$(X_i|X_j) = 0 \qquad \text{dla } i\neq j$$

i

$$(X_i|A) = 0 \qquad \text{dla } i = 1,..,m.$$

Dowód

$\Omega = 1$ wtedy i tylko wtedy, gdy $\omega_i = 1$ dla każdego i, co z kolei jest równoważne

$$(X_i|A) = 0 \quad \text{dla } i = 1,..,m$$

i

$$(X_i|X_j) = 0 \qquad \text{dla } i \neq j.$$

Kończy to dowód.

/10/ Własność

Jeśli zbiory $K, L \in 2^{\mathcal{U}}$ są takie, że

$$K \subset L,$$

to nie wynika stąd np., że

$$\Omega(K) \leq \Omega(L).$$

Własność tę wykażemy w § 5 omawiając estymację Ω.

Z powyższych własności funkcji Ω wynika, że może być ona użyta do wyboru podzbioru zbioru $\mathcal{U}$, najlepiej charakteryzującego dany zbiór obiektów, których cechy określamy za pomocą wartości jakie przyjmują zmienne losowe $X_1, \ldots, X_p$. Warto przy tym zauważyć, że wprowadzenie zmiennej A realizuje postulat separowania /porównaj /3/ /.

§ 4. Estymacja funkcji Ω, wybór układu cech charakterystycznych.

Za estymator funkcji Ω przyjmiemy statystykę, w której występują estymatory odpowiednich iloczynów skalarnych i norm.

Z twierdzenia /7/ rozdziału III wynika wtedy, że taki estymator funkcji Ω jest zgodny.

Przejdźmy teraz do omówienia realizacji Ω.

Załóżmy, że ustalony został a priori zbiór $U = \{x_1, \ldots, x_p\} \subset R^n$ realizacji zmiennych losowych /cech/, za pomocą których zamierza charakteryzować dane obiekty, np. kraje[1/], gospodarstwa rolne, przedsiębiorstwa, jedno przedsiębiorstwo w różnych momentach czasu, itp.

Ideę proponowanej przez nas metody wyboru cech realizującej postulaty wymienione w uwagach wstępnych, można streścić w następujący sposób.

Ze zbioru U wybieramy taki podzbiór elementów /cech/, że współczynniki korelacji między nimi są możliwie małe, a same cechy wykazują możliwie największą zmienność przy przejściu od obiektu do obiektu. Miarą tej zmienności mogłaby być wariancja, nie użyjemy jej jednak, ponieważ zależy ona od mian cech.

1/ W tym przypadku wybrany układ cech można by nazwać cywilizacją ze względu na ustalony w ten sposób system wartości, według których porównujemy i oceniamy stopień rozwoju krajów.

Do mierzenia zmienności użyjemy sinusa kąta między wektorem realizacji zmiennej /cechy/, a n-wymiarowym wektorem, którego wszystkie składowe są jednakowe. W ten sposób zostanie przez nas zrealizowany postulat /3/.

/11/ Definicja

Współczynnikiem separowalności indywidualnej cechy x_i /wektora realizacji zmiennej X_i/ nazywa się wielkość

$$\omega_{i.} = \frac{\sqrt{1 - \cos^2 \alpha_i}}{\sum_{j=1}^{m} r_{ij}} ,$$

gdzie

$$\alpha_i = \sphericalangle (x_j, a) ; \quad r_{ij} = r(x_i, x_j) ;$$

$a = (1, 1, \ldots, 1)$ ma tyle składowych ile jest realizacji zmiennej X_i.

/12/ Definicja

Współczynnikiem separowalności globalnej cech $x_1,\dots,x_m$ nazywa się wielkość

$$\Omega = \frac{1}{m} \sum_{j=1}^{m} \omega_j .$$

/13/ Definicja

Układem cech charakterystycznych /układem charakterystycznym/ obiektów opisanych a priori za pomocą zbioru U nazywamy taki podzbiór tego zbioru, dla którego współczynnik Ω przyjmie wartość największą.

Następujące twierdzenia określają własności współczynnika Ω.

/14/ Twierdzenie

Współczynnik Ω przyjmuje wartości z przedziału $[0,1]$.

Dowód

Z definicji współczynnika ω_j wynika, że $0 \leqslant \omega_j \leqslant 1$, ponieważ największą wartością jaką może przyjąć nieujemny licznik ułamka w /7/ jest jeden, a najmniejszą wartością nieujemnego mianownika tego ułamka jest również jeden. Stąd suma w definicji /8/ może przyjąć wartość co najwyżej równą m. Kończy to dowód.

/15/ Twierdzenie

Współczynnik Ω przyjmuje wartość zero, wtedy i tylko wtedy, gdy

$$\measuredangle(x_j, a) = 0 \qquad \text{dla } j = 1, \ldots, p.$$

Dowód

Współczynnik Ω jako suma nieujemnych składników jest równy zero wtedy i tylko wtedy, gdy każdy z nich jest równy zero. Ma to miejsce wtedy i tylko wtedy, gdy $|\cos\alpha_j| = 1$ dla każdego j, co kończy dowód.

/16/ Twierdzenie

Współczynnik $\Omega = 1$ wtedy i tylko wtedy, gdy x_i są parami nieskorelowane i ortogonalne do wektora stałego a.

Dowód

$\Omega = 1$ wtedy i tylko wtedy, gdy dla każdego j, $\omega_j = 1$, co z kolei jest równoważne $(x_j | a) = 0$ dla każdego j i $r_{ij} = 0$, dla $i \neq j$.

/17/ Uwaga

Jak wykazują przykłady /porównaj następny paragraf/ współczynnik Ω nie jest monotoniczny w sensie realizacji inkluzji między podzbiorami $\mathcal{U}$.

/18/ Uwaga

Jeśli współczynnik Ω przyjmie tę samą maksymalną wartość dla kilku podzbiorów U, to jako układ charakterystyczny przyjmuje się ten z owych podzbiorów, który jest najmniej liczny, wprowadzając współczynnik

$$\Omega^*(U_i) = \frac{\Omega(U_i)}{q_i},$$

gdzie q_i jest liczebnością zbioru U_i.

Jeśli i Ω^* nie pozwoli wyznaczyć jednoznacznie układu charakterystycznego /np. ze względu na równe liczebności zbiorów U_i/, to o wyborze układu decyduje doświadczenie badacza.

/19/ Uwaga

Jeśli dla układu charaketystycznego współczynnik Ω przyjmi wartość mniejszą od jedności, to wielkość $1 - \Omega$ może być interpretowana jako ta ilość informacji, której nie można uzyskać z wybranego a priori zbioru cech.

§ 5. Przykłady numeryczne.

Następujące przykłady zostały wykonane przy wykorzystaniu programu [39] sporządzonego przez W. Ostasiewicza. Program ten został opracowany na maszynę ODRA 1204.

/20/ Przykład

Dla dwóch wektorów x_1, x_2 dziesięciowymiarowych, dla których macierz

$$\begin{bmatrix} 1{,}0000 & 1{,}8352 & 0{,}5599 \\ 0{,}8352 & 1{,}0000 & 0{,}8087 \\ 0{,}5599 & 0{,}8087 & 1{,}0000 \end{bmatrix}$$

w pierwszej kolumnie i wierszu zawiera cosinusy kątów między wektorami realizacji i w której pozostałe elementy są odpowiednimi współczynnikami korelacji, uzyskano następujące rezultaty:

$$\Omega(x_1) = 0{,}5499$$

$$\Omega(x_2) = 0{,}8286$$

$$\Omega(x_1, x_2) = 0{,}7622.$$

Stąd, układem charakterystycznym jest układ $\{x_2\}$. Warto przy tym zauważyć, że zmienne x_1, x_2 są silnie skorelowane, a zmienna x_2 charakteryzuje się największą zmiennością w sensie przyjętej miary.

/21/ <u>Przykład</u>

Macierz, której elementami są / jak w przykładzie /20// cosinusy i współczynniki korelacji dla wektorów x_1, x_2, x_3, x_4, x_5 dziesięciowymiarowych, ma postać:

$$\begin{bmatrix} 1,0000 & 0,5761 & 0,9137 & 0,9325 & 0,8898 & 0,7935 \\ 0,5761 & 1,0000 & 0,6510 & 0,7755 & 0,7197 & 0,6103 \\ 0,9137 & 0,6510 & 1,0000 & 0,9159 & 0,8634 & 0,6686 \\ 0,9325 & 0,7755 & 0,9159 & 1,0000 & 0,9761 & 0,8243 \\ 0,8898 & 0,7197 & 0,8634 & 0,9761 & 1,0000 & 0,8381 \\ 0,7935 & 0,6103 & 0,6686 & 0,8243 & 0,8381 & 1,0000 \end{bmatrix}.$$

Wartości współczynników Ω są następujące:

$$\Omega(x_1) = 0{,}8174$$
$$\Omega(x_2) = 0{,}4064$$
$$\Omega(x_3) = 0{,}3612$$
$$\Omega(x_4) = 0{,}4564$$
$$\Omega(x_5) = 0{,}6086$$
$$\Omega(x_1,x_2) = 0{,}7412$$
$$\Omega(x_1,x_3) = 0{,}6638$$
$$\Omega(x_1,x_4) = 0{,}7407$$
$$\Omega(x_1,x_5) = 0{,}8855$$
$$\Omega(x_2,x_3) = 0{,}4006$$
$$\Omega(x_2,x_4) = 0{,}4630$$
$$\Omega(x_2,x_5) = 0{,}6083$$
$$\Omega(x_3,x_4) = 0{,}4137$$
$$\Omega(x_3,x_5) = 0{,}5316$$
$$\Omega(x_4,x_5) = 0{,}5794$$
$$\Omega(x_1,x_2,x_3) = 0{,}6294$$
$$\Omega(x_1,x_2,x_4) = 0{,}6831$$
$$\Omega(x_1,x_2,x_5) = 0{,}8037$$
$$\Omega(x_1,x_3,x_4) = 0{,}6281$$
$$\Omega(x_1,x_3,x_5) = 0{,}7315$$
$$\Omega(x_1,x_4,x_5) = 0{,}7778$$
$$\Omega(x_2,x_3,x_4) = 0{,}4318$$
$$\Omega(x_2,x_3,x_5) = 0{,}5332$$
$$\Omega(x_2,x_4,x_5) = 0{,}5722$$
$$\Omega(x_3,x_4,x_5) = 0{,}5197$$
$$\Omega(x_1,x_2,x_3,x_4) = 0{,}6050$$
$$\Omega(x_1,x_2,x_3,x_5) = 0{,}6936$$
$$\Omega(x_1,x_2,x_4,x_5) = 0{,}7305$$
$$\Omega(x_1,x_3,x_4,x_5) = 0{,}6793$$

$$\Omega(x_1,x_2,x_3,x_4) = 0{,}6050$$
$$\Omega(x_1,x_2,x_3,x_5) = 0{,}6936$$
$$\Omega(x_1,x_2,x_4,x_5) = 0{,}7305$$
$$\Omega(x_1,x_3,x_4,x_5) = 0{,}6793$$

$$\Omega(x_1,x_2,x_3,x_4,x_5) = 0{,}0553$$

Układem charakterystycznym jest układ $\{x_1, x_5\}$, przy czym współczynnik korelacji $r(x_1, x_5) = 0{,}6103$, $\cos \sphericalangle (x_1, a) = 0{,}5761$, $\cos \sphericalangle (x_5, a) = 0{,}7935$.

R o z d z i a ł V

OPTYMALNY WYBOR ZMIENNYCH DO MODELU LINIOWEGO

§ 1. Uwagi i twierdzenia wstępne.[1/]

Często postulowaną w praktyce zależnością między zmiennymi losowymi jest zależność typu

$$/1/ \qquad Y = a_1X_1 + \ldots + a_pX_p + a,$$

gdzie $X_1,\ldots,X_p$ zmienne losowe nazywane objaśniającymi /predyktantami/,

Y - zmienna losowa nazywana zmienną objaśniającą /predyktatą/.

Problem, któremu jest poświęcony niniejszy rozdział dotyczy optymalnego wyboru zmiennych do zależności postaci /1/. W ciągu dalszym wyjaśnimy co będziemy rozumieli przez optymalny wybór, a także powody, dla których nie można przyjąć, że najlepszy jest taki model, który zawiera jak największą ilość zmiennych.

1/ Ważniejsze wyniki tego i następnych paragrafów zostały opublikowane w pracy: R. Jasiński [18].

Wyborem zmiennych, oprócz przedstawionej w poprzednim rozdziale klasycznej metody czynników głównych zajmuje się Z. Hellwig w swojej pracy [14] dając jego rozwiązanie przy zastosowaniu tzw. współczynnika H pojemności nośników informacji. Proponowane przez nas rozwiązanie wykorzystuje metody przestrzeni Hilberta.

Zagadnienia związane z wyborem zmiennych mają aspekt A/ ilościowy i B/ jakościowy:

A_1/ Do modelu nie powinno wejść zbyt dużo zmiennych, ponieważ badania związane z uzyskiwaniem odpowiednich danych są często bardzo kosztowne. Ponadto

A_2/ wykażemy, że w pewnych przypadkach zwiększanie ilości zmiennych powoduje zmniejszanie współczynników przy zmiennych $X_1,\ldots,X_p$ w wyrażeniu /1/ co z kolei powoduje, że wpływ wartości zmiennych objaśniających na objaśnianą jest bardzo mały.

Jakościowy aspekt zagadnienia polega na wyborze właściwych zmiennych. Do modelu powinny wejść zmienne, które

B_1/ możliwie dużo różnią się od stałej co zostało częściowo omówione w poprzednim rozdziale;

B_2/ są między sobą słabo skorelowane;

B_3/ są silnie skorelowane ze zmienną objaśnianą.

Postulat B_2/ ma uzasadnienie w tym, że jeśli w modelu znajdą się zmienne silnie ze sobą skorelowane, spowoduje to niepotrzebne dublowanie informacji, a stąd niepotrzebne koszty, ponadto pomiary związane z uzyskiwaniem wartości realizacji zmiennych są obarczone błędami, a więc wzrosną zakłócenia zniekształcające ostateczną postać zależności /1/. Realizacja postulatu B_3/ spowoduje, że do modelu wejdą zmienne, których wartości silnie w wpływają na zmienną Y.

Rekapitulując, w modelu powinna znaleźć się możliwie mała ilość zmiennych różnych od stałej, słabo między sobą skorelowanych, silnie natomiast skorelowanych ze zmienną Y. Łatwo się jednak przekonać, że postulaty B_2 i B_3/ nie są niezależne. Mówi o tym następujące

/2/ Twierdzenie[1/]

Niech $Y, X_1, X_2 \in L_2$.

Jeśli

$$r(X_1, Y) = r_1 > 0,$$
$$r(X_2, Y) = r_2 > 0,$$

to

$$r_1 r_2 - \sqrt{(1 - r_1^2)(1 - r_2^2)} \leq r_{12} \leq r_1 r_2 + \sqrt{(1 - r_1^2)(1 - r_2^2)}$$

1/ Porównaj: R. Jasiński [16].

gdzie

$$r_{12} = r(X_1,X_2).$$

Dowód

Wobec założenia, że $Y, X_1, X_2 \in L_2$:

$$r_{12} = \cos\gamma \quad , \qquad r_1 = \cos\alpha \ , \ r_2 = \cos\beta \ ,$$

gdzie

/*/ $\alpha = \sphericalangle(X_1,Y), \ \beta = \sphericalangle(X_1,Y), \ \gamma = \sphericalangle(X_1,X_2)$

i $\qquad \alpha + \beta \leq \pi$

Przy ustalonych kątach α i β, zmienne /wektory/X_1X_2 można "obracać" tworząc koncentryczne "stożki" o osi symetrii Y. Stąd kąt γ między X_1,X_2 spełnia nierówność

$$|\alpha - \beta| \leq \gamma \leq \alpha + \beta \ ,$$

a zatem

$$\cos\alpha \cos\beta - \sqrt{(1-\cos^2\alpha \quad 1-\cos^2\beta)} \leqslant \cos\gamma \leqslant \cos\alpha + \cos\beta +$$

$$+ \sqrt{(1-\cos^2\alpha)(1-\cos^2\beta)} ,$$

co wobec równości /*/ kończy dowód.

/3/ Wniosek

Jeśli ustalimy, że do modelu wejdą zmienne, dla których współczynnik korelacji ze zmienną Y jest większy niż z góry zadana liczba to "wymuszona" zostanie wartość współczynnika korelacji między tymi zmiennymi.

/4/ Wniosek

Z nierówności, które są tezą twierdzenia /2/ wynika, że dla

$$r_1 \approx 1 \qquad \text{lub} \qquad r_2 \approx 1$$

zachodzi przybliżona równość[1/]

/5/ $$r_{12} \approx r_1 r_2$$

1/ Tę przybliżoną równość otrzymał za pomocą innych metod Z. Hellwig w swojej pracy [14].

O tym kiedy przybliżoną równość /5/ można zastąpić równością mówi następujące

/6/ Twierdzenie

$$r_{12} = r_1 r_2$$

wtedy i tylko wtedy, gdy dla dowolnych $i,j = 1,2,\dots,p$

$$X_1 - r_1 Y \perp X_2 .$$

Dowód

Dla $X_1,\dots,X_p$ zestandaryzowanych $r_{12} = (X_1|X_2)$, $r_1 = (Y|X_1)$ $r_2 = (Y|X_2)$.

Stąd równość

$$r_{12} = r_1 r_2$$

zajdzie wtedy i tylko wtedy, gdy

$$(X_1|X_2) - (Y|X_1)\ (Y|X_2) = 0,$$

co jest równoważne

$$\left(X_1|X_2\right) - r_1 \left(Y|X_2\right) = 0,$$

a ta z kolei zależność, wobec własności iloczynu skalarnego jest równoważna

$$\left(X_1 - r_1 Y|X_2\right) = 0,$$

c.b.d.o.

Przejdźmy do omówienia aspektu ilościowego wyboru zmiennych. Zanim wykażemy, że zwiększenie ilości zmiennych powoduje w pewnych przypadkach zmniejszenie współczynników w kombinacji liniowej /1/ udowodnimy dwa twierdzenia, z których będziemy w ciągu dalszym korzystali. Pierwsze z nich jest istotnym uogólnieniem twierdzenia o rzutowaniu na sumę prostą podprzestrzeni generowanych przez ortogonalne układy wektorów /porównaj np. W. Mlak [28] s. 82 i dalsze/.

/7/ <u>Twierdzenie</u>

Załóżmy, że wektory $X_1,\dots,X_p,\ X_{p+1},\dots,X_n \in L_2$ są liniowo niezależne / niekoniecznie ortogonalne/.

Niech Y' oznacza rzut Y na $[X_1,\dots,X_p]$, Y'' rzut Y na $[X_{p+1},\dots,X_n]$. Rzut wektora Y na podprzestrzeń

$$P = [X_1,\dots,X_p, X_{p+1},\dots,X_n]$$

jest równy rzutowi wektora Y na podprzestrzeń generowaną przez wektory Y' i Y'':

$$Rz\left(Y/ [X_1,\dots,X_p, X_{p+1},\dots,X_n]\right) = Rz\left(Y/ [Y', Y'']\right)$$

<u>Dowód</u>

Jeśli zortogonalizujemy /porównaj np. D. Luenberger [26] / zbiór wektorów $\{X_1,\dots,X_p, X_{p+1},\dots,X_n\}$, twierdzenie stanie się trywialne; teza wynika po prostu z łączności dodawania wektorów. Otrzymamy wtedy wektory $\tilde{y}'$ i $\tilde{y}''$ będące rzutami Y odpowiednio na zortogonalizowane podzbiory $\{\tilde{X}_1,\dots,\tilde{X}_p\}$ i $\{\tilde{X}_{p+1},\dots,\tilde{X}_n\}$, i wektor

$$\tilde{\tilde{y}} = \mathrm{Rz}\left(Y/\left[\tilde{X}_1,\dots,\tilde{X}_p,\ \tilde{X}_{p+1},\dots,\tilde{X}_n\right]\right) = \tilde{y}' + \tilde{y}''.$$

Dowód twierdzenia zostanie zakończony jeśli układ ortogonalny przekształcimy zpowrotem na układ $X_1,\dots,X_p,X_{p+1},\dots,X_n$.

/8/ Twierdzenie

Niech Y' oznacza rzut unormowanego wektora Y na podprzestrzeń

$$P = [X_1,\dots,X_p]\ ;$$

X - wektor unormowany i liniowo niezależny od Y';
Y''- rzut wektora Y na podprzestrzeń $[Y', X]$.

Wektor Y'' można wyrazić za pomocą kombinacji liniowej

$$Y'' = \alpha Y' + \beta X,$$

gdzie $|\alpha| < 1$ wtedy i tylko wtedy, gdy spełnione są nierówności

$$(Y'|X)\cdot(Y'-Y|X) < \delta$$

/9/

$$2\,\|Y'\|^2 - (Y'|X)(Y'+Y|X) > 0.$$

Dowód

Wyznaczmy współczynniki α i β . Wektor $Y - Y''$ jest ortogonalny do Y' i X, a więc

$$(Y - \alpha Y' - \beta X|Y') = 0$$

i

$$(Y - \alpha Y' - \beta X|X) = 0,$$

co jest równoważne

/10/
$$\begin{cases} \alpha(Y'|Y') + \beta(X|Y') = (Y|Y') \\ \alpha(Y'|X) + \beta = (Y|X). \end{cases}$$

Stąd

$$/11/ \qquad \alpha = \frac{(Y'|Y') - (Y'|X)(Y|X)}{(Y'|Y') - (Y'|X)^2} \; .$$

Wobec założonej niezależności wektorów Y' i X, mianownik ułamka /11/ jest większy od zera. Zatem po przekształceniu otrzymamy, że nierówność $|\alpha| < 1$ jest spełniona wtedy i tylko wtedy, gdy są spełnione nierówności /9/, co kończy dowód.

Sformułujmy warunki /9/ używając zamist iloczynów skalarnych współczynników korelacji.

Korzystając z zależności /4/ rozdziału II między współczynnikiem korelacji a iloczynem skalarnym i normą, i zakładając, że wektory Y i X są zestandaryzowane, otrzymamy równości

$$(Y|X) = \|Y\| \, \|X\| \; r(X,Y) = r(X,Y)$$
$$(Y'|X) = \|Y'\| \, \|X\| \; r(Y',X) = r(Y,Y') \; r(Y', X)$$
$$(Y'|Y') = \; r^2(Y', Y),$$

które po podstawieniu do układu /9/ dają

$$\begin{cases} [r(Y,Y') \; r(Y', X)]^2 - r(Y', Y) \; r(Y',X) \; r(X,Y) < 0 \\ 2r^2(Y',Y) - [r(Y,Y') \; r(Y',X)]^2 - r(Y,Y') \, r(Y',X) \; r(X,Y) > 0 \end{cases}$$

co jest z kolei równoważne

$$\begin{cases} r(Y,Y')\ r(Y',X)\ \left[r(Y,Y')\ \ r(Y',X)\ -r(X,Y)\right] < 0 \\ r(Y',Y)\ \left[2r(Y',Y)\ -r(Y',Y)\ r^2(Y',X)-r(Y',X)\ r(X,Y)\right] > 0 \end{cases}$$

Ponieważ Y′ jest rzutem Y, $r(Y,Y') > 0$, a zatem ostatni układ nierówności jest równoważny układowi

$$\begin{cases} r(Y',X)\left[r(Y,Y')\ \ r(Y',X)\ -r(X,Y)\right] < 0 \\ 2r(Y';Y)\ -r(Y',Y)\ \ r^2\ (Y',X)-r(Y',X)\ \ r(X,Y) > 0. \end{cases} \tag{12}$$

Tak więc układ nierówności /9/ jest równoważny układowi /12/ zapisanemu przy użyciu współczynników korelacji, a ten z kolei alternatywie poniższych układów

$$\begin{cases} r(Y',X) > 0 \\ r(Y,Y')\ r(Y',X)\ -r(X,Y) < 0 \\ 2r(Y',Y)\ -r(Y',Y)\ \ r^2\ (Y',X)-r(Y',X)\ \ r(X,Y) > 0\ ; \end{cases}$$

lub

$$\begin{cases} r(Y',X) < 0 \\ r(Y,Y')\ r(Y'X) - r(X,Y) > 0 \\ 2r(Y',Y) - r(Y',Y)\ r^2(Y',X) - r(Y',X)\ \ r(X,Y) > 0. \end{cases}$$

Sformułujemy i udowodnimy twierdzenie, które podaje warunki dostateczne na to, aby zwiększenie ilości zmiennych w modelu powodowało zmniejszenie bezwzględnych wartości współczynników przy zmiennych.

/13/ Twierdzenie

Niech Y' oznacza rzut wektora Y na podprzestrzeń

$$P = [X_1, \ldots, X_p],$$

a więc

$$Y' = a_1X_1 + \ldots + a_pX_p.$$

Jeśli X jest wektorem nie należącym do P i spełnia warunki

$$/14/ \quad \begin{cases} r(Y',X)\left[r(Y,Y')\; r(Y',X) - r(X,Y)\right] < 0 \\ 2r(Y,Y') - r(Y',Y)\; r^2(Y',X) - r(Y',X)\; r(X,Y) > 0 \end{cases}$$

to wektor

$$Y'' = Rz\left(Y/\left[X_1, \ldots, X_p, X\right]\right)$$

wyraża się wzorem

$$Y'' = b_1X_1 + b_2X_2 + \ldots + b_pX_p + bX,$$

gdzie

$$|b_i| < |a_i|$$

dla $i = 1, \ldots, p$.

Dowód

Wnioskiem z twierdzenia /7/ jest fakt, że Y'' można przedstawić w postaci kombinacji liniowej wektorów Y' i X:

$$Y'' = \alpha Y' + bX = \alpha a_1 X_1 + \ldots + \alpha a_p X_p + bX.$$

Jeśli spełnione są założenia /14/, to jak wykazaliśmy w twierdzeniu /8/

$$|\alpha| < 1,$$

a zatem mnożąc obustronnie tę nierówność przez $|a_i|$ i uwzględniając równość $\alpha\, a_i = b_i$ otrzymujemy

$$|b_i| = |a_i| = |\alpha|\,|a_i| < |a_i|,$$

co kończy dowód.

§ 2. Współczynnik W. wyboru zmiennych danych przez realizacje i jego własności.

Załóżmy, że elementy $U = \{x_1,\ldots,x_p\}$ i y oznaczają n-elementowe $(n > p+1)$ realizacje zmiennych $X_1,\ldots,X_p$, Y. Elementy $x_1,\ldots,x_p,y$ będziemy traktowali jako wektory przestrzeni R^n, ponadto założymy o nich, że są unormowane na jeden, to znaczy ich składowe są podzielone odpowiednio przez

$$\|x_i\|_n = \sqrt{\frac{1}{n} \sum_{j=1}^{n} x_{ij}^2}\,,$$

$$\|y\|_n = \sqrt{\frac{1}{n} \sum_{i=1}^{n} y_i^2}$$

W poprzednim paragrefie przedstawiliśmy postulaty $A_i(i=1,2)$ i $B_j(j=1,2,3)$, które powinny być spełnione przy wyborze zmiennych do modelu postaci /1/. Wprowadźmy funkcje, których wartości będą miarami stopnia realizacji tych postulatów.

Za miarę zależności zmiennych między sobą i niezależności od stałej, która w realizacji ma postać wektora

$$a = (1,\ldots,1),$$

przyjmujemy objętość V równoległościanu rozpiętego na wektorach $x_1,\dots,x_p,a$; przy czym zauważmy, że

$$\|a\|_n = 1.$$

Zależność między zmiennymi objaśniającymi, a zmienną y będziemy mierzyli odległością d wektora y od podprzestrzeni generowanej przez wektory $x_1,\dots,x_p$. Odległość ta, jak wykazaliśmy w rozdziale II § 5, ma ścisły związek ze współczynnikiem korelacji wielowymiarowej z próby między y a $x_1,\dots,x_p$:

$$d\left(y,\left[x_1,\dots,x_p\right]\right) = \sqrt{1-\varrho^2}\,.$$

Można ją traktować jako miarę aproksymacji liniowej zmiennej Y przez zmienne $X_1,\dots,X_p$.

Wprowadźmy teraz funkcję, której wartości posłużą nam jako kryterium wyboru zmiennych do modelu /1/.

/15/ <u>Definicja</u>

Współczynnikiem W nazywamy funkcję

$$W \quad 2^{U} \setminus \{\emptyset\} \to R$$

określoną dla $x_1,\ldots,x_p;y$ wzorem

$$W(x_1,\ldots,x_p;y) = 1 - \frac{d([x_1,\ldots,x_p],y)}{1 + V(x_1,\ldots,x_p,a)},$$

gdzie

$$V(x_1,\ldots,x_p,a) = \sqrt{g(x_1,\ldots,x_p,a)}.$$

Współczynnik W ma następujące własności uzasadniające jego przydatność do wyboru zmiennych objaśniających.

/16/ Własność

Współczynnik W przyjmuje wartości z odcinka domkniętego $[0,1]$.

/17/ Własność

Jeśli wektory $x_1,\ldots,x_p$ są liniowo niezależne, to

$$W = 1$$

wtedy i tylko wtedy, gdy istnieją takie skalary $a_1,\dots,a_r$ i wektory $x_{\alpha_1},\dots,x_{\alpha_r}$ należące do układu

$$\{x_1,\dots,x_p\},$$

że

$$y = a_1x_{\alpha_1} + \dots + a_rx_{\alpha_r} \qquad (r \leqslant p, \quad \alpha_i \leqslant p).$$

/18/ Własność

Jeśli $W = 0$, to wektor y jest ortogonalny do podprzestrzeni $[x_1,\dots,x_p]$, tzn.

$$r(y,x_i) = (y|x_i) = 0 \qquad \text{dla} \quad i = 1,\dots,p.$$

/19/ Własność

Jeśli $a \in [x_1,\dots,x_p]$

i wektor y jest ortogonalny do podprzestrzeni $[x_1,\ldots,x_p]$, to $W = 0$.

Oto dowody sformułowanych własności.

Dowód własności /16/

Wektor y ma długość równą jeden, a zatem

$$0 \leqslant d \leqslant 1.$$

Objętość V też spełnia nierówność

$$0 \leqslant V \leqslant 1,$$

wobec unormowania wektorów $x_1,\ldots,x_p,a$.

Stąd ułamek $\frac{d}{1+V}$ przyjmuje wartości z przedziału domkniętego $[0,1]$.

Dowód własności /17/

Jeśli

$$y = a_1x_1 + \ldots + a_px_p,$$

to z twierdzenia o rzutowaniu /rozdział I twierdzenie /47//
wynika, że

$$d = 0,$$

ale wtedy

$$W' = 1.$$

Jeśli natomiast

$$'W = 1,$$

to

$$\frac{d}{1 + V} = 0,$$

ale wtedy

$$d = 0,$$

tzn. y jest elementem podprzestrzeni $[x_1, \dots, x_p]$, a więc

$$y = a_1 x_1 + \dots + a_p x_p.$$

Dowód własności /18/

Jeśli

$$W = 0,$$

to

$$\frac{d}{1 + V} = 1,$$

tzn.

$$d = 1 + V. \tag{*}$$

Ponieważ $0 \leqslant d \leqslant 1$, $0 \leqslant V \leqslant 1$, to równość /*/ będzie spełniona wtedy i tylko wtedy, gdy

$$V = 0$$

i

$$d = 1,$$

co kończy dowód własności.

Dowód własności /19/

Jeśli

$$a \in [x_1, \ldots, x_p],$$

to

$$V(x_1, \ldots, x_p, a) = 0.$$

Jeśli ponadto założymy, że y jest ortogonalny do $[x_1, \ldots, x_p]$, to

$$d = 1,$$

stąd

$$\frac{d}{1 + V} = 1,$$

a więc

$$W = 0.$$

§ 3. Zastosowanie współczynnika W do wyboru zmiennych danych przez realizację.

Do modelu /1/ wybiera się te zmienne $x_{\alpha_1},\dots,x_{\alpha_r}$ spośród $x_1,\dots,x_p$, dla których

$$W\left(x_{\alpha_1},\dots,x_{\alpha_r}\right) = 1 - \frac{d\left(\left[x_{\alpha_1},\dots,x_{\alpha_r}\right], y\right)}{1 + V\left(x_{\alpha_1},\dots,x_{\alpha_r},a\right)}$$

przyjmie wartość największą, wprowadźmy przy tym określenie

/20/ Definicja

Układem minimalnym nazywamy taki podzbiór $\left\{x_{\alpha_1},\dots,x_{\alpha_r}\right\}$ zbioru $\left\{x_1,\dots,x_p\right\}$, dla którego współczynnik W przyjmuje wartość największą.

Przedstawimy teraz w punktach postępowanie związane z wyborem układu minimalnego przy wykorzystaniu współczynnika W.

1° Wybrać spośród wektorów $x_1,\dots,x_p$ maksymalny podukład liniowo niezależny. Oznaczymy go przez $\mathcal{X}$ i niech jego liczebność będzie $\hat{k}$.

2^o Unormować elementy ze zbioru $\mathcal{X}$.

3^o Dla każdego r-elementowego $(r < k)$ podzbioru zbioru $\mathcal{X}$ wyznaczyć wartość współczynnika W. Podzbiór, dla którego współczynnik ten przyjmie wartość największą jest układem minimalnym. Niech to będzie układ

$$x_{\alpha_1}, \ldots, x_{\alpha_r}.$$

4^o Wyznaczyć współczynnik $a_0, a_1, \ldots, a_r$ kombinacji

$$Y = a_1 X_{\alpha_1} + \ldots + a_r X_{\alpha_r} + a_0$$

za pomocą metody najmniejszych kwadratów /porównaj rozdział II paragraf 6/.

Uwagi

1. Wyznaczenie maksymalnego układu liniowo niezależnego spośród wektorów $x_1, \ldots, x_p$, wobec twierdzenia /66/ z rozdziału I,

jest w praktyce obliczeniowej proste: sprowadza się do badania rzędu macierzy Grama odpowiadającej tym wektorom. Oczywiście jeśli w macierzy Grama występują poza główną przekątną elementy równe jeden co do bezwzględnej wartości, to należy przede wszystkim wyeliminować wektor odpowiadający kolumnie lub wierszowi, w którym taki element się znajduje. Ponadto warto zauważyć, że wyznaczenie maksymalnego układu liniowo niezależnego pozwala wykluczyć ze zbioru rozpatrywanych zmiennych pewną ich ilość a ponieważ ze względu na zaokrąglenia czynione przez maszynę matematyczną stosowaną do obliczeń, wykluczone zostaną te zmienne, które są praktycznie zależne od pozostałych.

2. Jeśli dla $i,j = 1,\dots,p$

$$\left| r\left(x_i, x_j\right) \right| = 1,$$

to układ minimalny tworzy którakolwiek ze zmiennych $X_1,\dots,X_p$.

3. Jeśli dla kilku różnych podzbiorów zbioru $\mathcal{X}$ współczynnik W przyjmie tę samą największą wartość, to za układ minimalny będziemy uważać ten podzbiór, który jest najmniej liczny.

Jeśli liczebność układu minimalnego jest bardzo istotna

/np. ze względu na koszt lub pracochłonność uzyskiwania danych/, to można zastosować procedurę postoptymalizacyjną przyjmując za układ minimalny taki podzbiór $\mathcal{T} \subset \mathcal{K}$, dla którego wyrażenie

$$W^* = \frac{W(x_{\alpha_1}, \ldots, x_{\alpha_t})}{t}$$

przyjmie wartość największą, gdzie t oznacza liczebność zbioru $\mathcal{T}$.

4. Wartość 1 - W można interpretować jako tę ilość informacji, która nie została dostarczona przez a priori wybrane do modelu zmienne $X_1, \ldots, X_p$.

5. Współczynnik W jest związany z współczynnikiem ϱ korelacji wielowymiarowej zależnością

$$W = 1 - \frac{\sqrt{1 - \varrho^2}}{1 + V} .$$

Nie ma jednak tej wady co współczynnik ϱ zastosowany do wyboru zmiennych. W przeciwieństwie do ρ może być

$$\mathfrak{z}_s \subset \mathfrak{z}_t \subset \mathcal{K},$$

a

$$w(\mathfrak{z}_s) > w(\mathfrak{z}_t),$$

Wykażemy to na przykładach w paragrafie 5.

§.4. Postać współczynnika W w przestrzeni L_2.

Niech elementy zbioru $\mathfrak{U} = \{X_1, \ldots, X_p\}$ i Y oznaczają zmienne losowe należące do L_2, w której ciąg (B_i) jest bazą. Odległość d między zmienną losową Y, a podprzestrzenią generowaną przez zbiór $\{X_1, \ldots, X_p\}$ wyraża się wzorem /porównaj twierdzenie /68/ rozdział I/:

$$d\left(Y, [X_1, \ldots, X_p]\right) = \sqrt{\frac{g(X_1, \ldots, X_p, Y)}{g(X_1, \ldots, X_p)}}, \tag{21}$$

gdzie

$$g(X_1, \ldots, X_p) \qquad \text{i} \qquad g(X_1, \ldots, X_p, Y)$$

oznaczają wyznaczniki Grama w L_2.

Oznaczmy przez

$$\vartheta = \sqrt{g(X_1, \ldots, X_p, A)},$$

gdzie

$$A = \sum_i \frac{B_i}{\| B_i \|} ,$$

"objętość równoległościanu"rozpiętego na $X_1, \dots, X_p, A$ /porównaj z uwagą /24/,rozdział III/.

Przy tych oznaczeniach, współczynnik W ma w L_2 postać

$$W = 1 - \frac{d\left(Y, [X_1, \dots, X_p]\right)}{1 + \vartheta\left(X_1, \dots, X_p, A\right)} .$$

W paragrafie 2 rozdziału III wykazaliśmy, że wyznacznik V_n jest estymatorem zgodnym wyznacznika Grama w L_2. Stąd też podstawiając w /21/ odpowiednie estymatory, statystyka

$$d = \sqrt{\frac{V_n'}{V_n}}$$

jest wobec twierdzenia /7/ rozdziału III estymatorem zgodnym odległości.

Jeśli ponadto przez $\hat{\vartheta}$ oznaczymy estymator zgodny parametru ϑ, to również z twierdzenia /7/ rozdziału III wynika

/22/ <u>Twierdzenie</u>

Statystyka

$$\hat{\mathcal{W}} = 1 - \frac{\hat{d}}{1 + \hat{\vartheta}}$$

jest estymatorem zgodnym współczynnika $\mathcal{W}$.

/23/ <u>Twierdzenie</u>

Współczynnik $\mathcal{W}$ ma następujące własności:

1. $$\mathcal{W} \quad 2^{\mathfrak{U}} \setminus \{\emptyset\} \longrightarrow [0,1]$$

2. Jeśli zmienne $X_1, \ldots, X_p$ są liniowo niezależne, to

$$\mathcal{W} = 1$$

wtedy i tylko wtedy, gdy

$$Y \in \left[X_1,\dots,X_p\right]$$

3. Jeśli

$${}^{M}\mathcal{W} = 0,$$

to zmienna Y jest nieskorelowana z każdą ze zmiennych

$$X_1,\dots,X_p \ .$$

4. Jeśli

$$A \in \left[X_1,\dots,X_p\right]$$

i zmienna Y jest nieskorelowana z każdą ze zmiennych $X_1,\dots,X_p$, to

$${}^{M}\mathcal{W} = 0.$$

---

5. Współczynnik ${}^{M}\mathcal{W}$ nie jest monotoniczny[1/] w sensie inkluzji między podzbiorami $\mathcal{U}$. Wykazują to przykłady.

Dowód twierdzenia /23/ jest analogiczny do dowodów własności /16/, /17/,/18/ i /19/.

1/ Porównaj z uwagą 5.

§ 5. Przykłady numeryczne.

Autorem programu na EMC ODRA 1204, na podstawie którego zostały wykonane powyższe przykłady jest W. Ostasiewicz [40] .

/24/ Przykład

Do obliczeń zostały użyte trzy dziesięciowymiarowe wektory y - realizacje zmiennej objaśnianej, x_1, x_2 - realizacje zmiennych objaśniających o następującej macierzy współczynników korelacji

$$\begin{bmatrix} 1,0000 & 0,8352 & 0,5599 \\ 0,8352 & 1,0000 & 0,8087 \\ 0,5599 & 0,8087 & 1,0000 \end{bmatrix} .$$

Uzyskano następujące wartości współczynników W :

$$W(x_1) = 0,68$$

$$W(x_2) = 0,37$$

$$W(x_1, x_2) = 0,58 .$$

Układem minimalnym jest układ $\{x_1\}$, zauważmy przy tym, że

$$r(x_1, y) = 0{,}8352,$$

$$r(x_1, x_2) = 0{,}8087.$$

/25/ <u>Przykład</u>

Dla wektora y realizacji zmiennej objaśnianej i wektorów x_1, x_2, x_3, x_4, x_5 o macierzy symetrycznej

$$\begin{bmatrix}
1{,}0000 & & & & & \\
0{,}4048 & 1{,}0000 & & & & \\
0{,}2538 & -0{,}2482 & 1{,}0000 & & & \\
-0{,}6843 & -0{,}1888 & -0{,}3019 & 1{,}0000 & & \\
0{,}1032 & -0{,}5929 & 0{,}0[illegible]22 & -0{,}1949 & 1{,}0000 & \\
0{,}1951 & -0{,}1414 & -0{,}0111 & -0{,}4905 & 0{,}5904 & 1{,}0000
\end{bmatrix}$$

współczynników korelacji uzyskano następujące wartości współczynników W :

$$W(x_1) = 0{,}1639$$
$$W(x_2) = 0{,}0644$$
$$W(x_3) = 0{,}4683$$
$$W(x_4) = 0{,}0107$$
$$W(x_5) = 0{,}0381$$
$$W(x_1,x_2) = 0{,}1829$$
$$W(x_1,x_3) = 0{,}5318$$
$$W(x_1,x_4) = 0{,}1096$$
$$W(x_1,x_5) = 0{,}1770$$
$$W(x_2,x_3) = 0{,}4092$$
$$W(x_2,x_4) = 0{,}0734$$
$$W(x_2,x_5) = 0{,}1014$$
$$W(x_3,x_4) = 0{,}4008$$
$$W(x_3,x_5) = 0{,}3397$$
$$W(x_4,x_5) = 0{,}0306$$
$$W(x_1,x_2,x_3) = 0{,}4698$$
$$W(x_1,x_2,x_4) = 0{,}1463$$
$$W(x_1,x_2,x_5) = 0{,}2021$$
$$W(x_1,x_3,x_4) = 0{,}4364$$
$$W(x_1,x_3,x_5) = 0{,}4254$$
$$W(x_1,x_4,x_5) = 0{,}1214$$
$$W(x_2,x_3,x_4) = 0{,}3703$$
$$W(x_2,x_3,x_5) = 0{,}3357$$
$$W(x_2,x_4,x_5) = 0{,}0927$$
$$W(x_3,x_4,x_5) = 0{,}3021$$
$$W(x_1,x_2,x_3,x_4) = 0{,}4054$$
$$W(x_1,x_2,x_3,x_5) = 0{,}4046$$
$$W(x_1,x_2,x_4,x_5) = 0{,}1596$$
$$W(x_1,x_3,x_4,x_5) = 0{,}3567$$
$$W(x_2,x_3,x_4,x_5) = 0{,}3080$$
$$W(x_1,x_2,x_3,x_4,x_5) = 0{,}3529$$

Układem minimalnym jest układ $\{x_1, x_3\}$, przy czym

$$r(y, x_1) = 0{,}4048,$$

$$r(y, x_3) = -0{,}6843,$$

$$r(x_1, x_3) = -0{,}1888.$$

Rozdział VI

DYSKRYMINACJA I TAKSONOMIA W PRZESTRZENI LINIOWEJ

§ 1. Dyskryminacja

Załóżmy, że badając zbiór $\mathcal{Z}$ obiektów /mogą to być np. przedsiębiorstwa, gospodarstwa rolne, kraje, stany obiektu w wybranych momentach czasu itp./ wybraliśmy np. metodą przedstawioną w rozdziale IV p spośród rozważonych cech, za pomocą których obiekty te są opisane. Został ustalony więc układ charakterystycznych cech a co za tym idzie baza przestrzeni, w której obiekty będą reprezentowane jako jej elementy.

Niech $Z = \{A_1, \ldots, A_n\} \subset E^p$ oznacza zbiór, który będziemy utoższamiali ze zbiorem obiektów. Przez dyskryminację zbioru Z rozumiemy[1/] podział tego zbioru na podzbiory rozłączne niepuste i dające w sumie cały zbiór Z . Podzbiory takie nazywa się warstwami. Podział na warstwy powinien spełniać następujący warunek: w warstwie elementy powinny być bardziej skoncentrowane niż w całym zbiorze Z . Jest to o tyle trudne do zrealizowania, że łatwo doprowadza do sytuacji, w której zbiór Z rozbija się na podzbiory jednoelementowe. Innym ważnym problemem teorii dyskryminacji jest zagadnienie istnienia

---

1/ Porównaj: W. Bukietyński, Z. Hellwig, U. Królik, A.Smoluk [2] .

związku między zależnością cech jako zmiennych losowych, a konfiguracją obiektów za ich pomocą scharakteryzowanych. Problem ten jest o tyle istotny, że gdyby układ przestrzenny obiektów zależał od związków między cechami, to dyskryminacja nie odzwierciedlałaby własności zbioru obiektów. Traktuje o nim sformułowane dalej twierdzenie /2/.

Niech

$$A_1 = (x_{11}, x_{12}, \dots, x_{1p})$$

$$\dots\dots\dots\dots\dots\dots$$

$$A_n = (x_{n1}, x_{n2}, \dots, x_{np})$$

oznaczają wektory /obiekty/, których składowymi są wartości cech $c_1, c_2, \dots, c_p$. Przestrzeń euklidesową E^p, której elementami są obiekty, nazywamy przestrzenią obiektów. Cechy są wektorami o n składowych:

$$x_1 = (x_{11}, x_{21}, \dots, x_{n1})$$

$$\dots\dots\dots\dots\dots\dots$$

$$x_p = (x_{1p}, x_{2p}, \dots, x_{np})$$

Będziemy je traktowali jako elementy przestrzeni R^n realizacji zmiennych losowych, a przestrzeń R^n będzie nazywana przestrzenią cech i założymy o cechach, że są zestandaryzowane. Stąd współczynnik korelacji między cechami c_i, c_j

$$r_{ij} = r\left(x_i, x_j\right) = \sum_{k=1}^{n} x_{ki}\, x_{kj} \,.$$

Przyjmujemy ponadto, że odległość d_{ij} między obiektami jest określona równością

/1/ $$d_{ij} = d\left(A_i, A_j\right) = \sqrt{\sum_{k=1}^{p} \left(x_{ik} - x_{jk}\right)^2}$$

/2/ Twierdzenie[1/]

Odległości między elementami zbioru Z nie zależą od współczynników korelacji między cechami.

Dowód

Kwadrat odległości między obiektami A_i, A_j jest równy

1/ Porównaj: R. Jasiński [20].

$$d_{ij}^2 = \left(x_{i1}-x_{j1}\right)^2 + \ldots + \left(x_{is}-x_{js}\right)^2 + \ldots + \left(x_{iu}-x_{ju}\right)^2 + \ldots + \left(x_{ip}-x_{jp}\right)^2.$$

Jeśli zmienimy tak składowe s,u, że będą zachowane równości

/3/
$$x_{is} - x_{js} = a$$
$$x_{iu} - x_{ju} = b$$

/ a i b - pewne ustalone wielkości/, to odległość między A_i, A_j nie zmieni się.

Z zależności, /porównaj rozdziały II i III/

$$d\left(x_s, x_u\right) = \sqrt{\frac{1}{n}\sum_{k=1}^{n}\left(x_{ks} - x_{ku}\right)^2} = \sqrt{2 - 2\, r_{su}}$$

po przekształceniu otrzymujemy

$$r_{su} = \frac{2 - d^2\left(x_s, x_u\right)}{2} =$$

$$= 1 - \frac{1}{n}\left[\left(x_{1s}-x_{1u}\right)^2 + \ldots + \left(x_{is}-x_{iu}\right)^2 + \ldots + \left(x_{js}-x_{ju}\right)^2 + \ldots + \left(x_{ns}-x_{nu}\right)^2\right].$$

Jest zatem oczywiste, że można tak zmienić $x_{is}, x_{iu}, x_{js}, x_{ju}$ aby były spełnione zależności /3/ a zmieniły się przy tym $\left(x_{is} - x_{iu}\right)^2$ i $\left(x_{js} - x_{ju}\right)^2$, co kończy dowód twierdzenia.

/4/ Uwaga

Twierdzenie /2/ umożliwia przeprowadzenie dyskryminacji dowolnego zbioru obiektów bez analizowania siły związków między cechami. Innymi słowy sposób rozdyskryminowania zbioru obiektów zależy tylko od własności jego elementów. W szczególności może zostać rozdyskryminowany zbiór, którego elementami są realizacje pary zmiennych losowych zależnych.

§ 2. Taksonomia wrocławska.

W roku 1952 K. Florek, J. Łukaszewicz, J. Perkal, H. Steinhaus i S. Zubrzycki w [8] podali sposób nieliniowego uporządkowania zbioru punktów $Z = \{A_1, \dots, A_n\}$ będących elementami przestrzeni E^p, nazwany taksonomią wrocławską.

Przedstawimy uzyskane przez nich rezultaty zawarte w pracy cytowanej i w [7] tychże autorów.

Oznaczmy przez $d(A_i, A_j)$ odległość między elementami A_i, A_j zbioru Z. O zbiorze tym założymy, że odległość między jego elementami nie powtarzają się. Na elementach zbioru Z rozpina się graf spójny[1/] w następujący sposób:

1. Każdy element zbioru Z łączy się z najbliższym odcinkiem o długości równej odległości między tymi punktami. Tak otrzymany graf nazywa się skupieniem pierwszego rzędu.
2. Każde skupienie pierwszego rzędu łączy się z najbliższym, przy czym przez odległość dwu skupień S, P rozumie się najmniejszą odległość między punktami, z których jeden należy do skupienia S a drugi do skupienia P. Taki graf nazywa się skupieniem rzędu drugiego.

Postępowanie kontynuuje się dopóty, dopóki wszystkie elementy zbioru Z zostaną połączone. Tak uzyskany graf jest spój-

1/ Definicje dotyczące używanych tu pojęć teorii grafów można znaleźć np. w: R. Jasiński [19] lub w O. Ore [31].

ny i nie zawiera cykli, ponieważ odległości między punktami zbioru Z nie powtarzają się.

Ponadto w cytowanej już pracy [7] dowodzi się następujących twierdzeń.

Niech $D(Z)$ oznacza najkrótszy graf rozpięty na Z /tzn. taki, w którym suma długości łuków jest najmniejsza/, a $F(Z)$ graf rozpięty metodą taksonomii wrocławskiej.

/5/ Twierdzenie

Najkrótszym grafem rozpiętym na zbiorze Z jest graf uzyskany za pomocą taksonomii wrocławskiej, tzn.

$$D(Z) = F(Z).$$

/6/ Definicja

Długością podziału nazywamy sumę długości grafów $D(Z_k)$, gdzie $Z_1, \ldots, Z_m$ oznaczają rozłączne podzbiory zbioru Z takie, że

/7/ $$Z = Z_1 \cup \ldots \cup Z_m.$$

/8/ Definicja

Podział zbioru Z na m części nazywa się najlepszy jeśli jego długość jest minimalna.

/9/ <u>Twierdzenie</u>

Jeśli podział /7/ jest najlepszym podziałem zbioru Z na m części, to

$$\bigcup_{k=1}^{m} D(Z_k) \subset D(Z).$$

Z tego twierdzenia wynika natychmiast

/10/ <u>Twierdzenie</u>

Podział zbioru Z na m części jest najlepszy, jeśli w grafie D(Z) wyeliminuje się m - 1 najdłuższych łuków. Powstałe w wyniku tej operacji podgrafy spójne są najkrótszymi grafami $D(Z_k)$.

Numerycznym aspektem wyznaczania grafu F(Z) zajmował się G. Trybuś. Jest on autorem programu opublikowanego w [37], który pozwala wyznaczyć graf F(Z), przy czym liczebność zbioru Z jest ograniczona jedynie pojemnością pamięci maszyny cyfrowej.

§ 3. Związek taksonomii wrocławskiej ze statystyką.

Przestrzeń cech, która będzie przedmiotem naszych rozważań jest przestrzenią R^n. Wymiar tej przestrzeni jest określony przez liczbę obiektów, które są scharakteryzowane rozważanymi cechami. Każdy punkt przestrzeni jest wektorem, którego składowe są wartościami cechy dla poszczególnych obiektów.

W rozdziale II /twierdzenie /7// otrzymaliśmy zależność między współczynnikiem korelacji a odległością między wektorami realizacji zmiennych losowych.

/11/ $$d(x_i,x_j)=\sqrt{2-2r(x_i,x_j)}$$

Zależność ta umożliwi nam powiązanie znanej i rozbudowanej teorii dotyczącej współczynnika korelacji z taksonomią mającą wiele efektownych zastosowań[1/]. Powiązanie to umożliwia w pewnych przypadkach rozwiązywanie problemów ekonometrii dotyczących zmiennych losowych za pomocą metod taksonomii.

Z wzoru /11/ można wywnioskować, że zależność zmiennych losowych nie implikuje zerowania się odległości między nimi. Tak więc w ogólnym przypadku nie można zamiast macierzy współczynników korelacji badać macierzy odległości między cechami.

1/ Porównaj np.:Z. Hellwig [15].

Powstaje zatem problem w jakim "szczególnym" przypadku badanie odległości pozwala przenieść uzyskane wyniki na zagadnienie dotyczące zależności zmiennych, lub innymi słowy - kiedy między macierzami współczynników korelacji i odległości jest adekwatna odpowiedniość. Zanim na to pytanie odpowiemy, wprowadźmy definicję takiej odpowiedniości.

/12/ Definicja

Mówimy, że dla zbioru zmiennych danych przez realizacje /wektorów/ $\{x_i\}_{i \in N}$ istnieje adekwatna odpowiedniość między współczynnikami korelacji a odległościami tych zmiennych jeśli są spełnione następujące postulaty:

/I/ $\left|r(x_s,x_t)\right| = 1$ wtedy i tylko wtedy, gdy

$d(x_s,x_t) = 0 \qquad (s,t \in N)$;

/II/ $\left|r(x_s,x_t)\right| = \left|r(x_u,x_t)\right|$ wtedy i tylko wtedy, gdy

$d(x_s,x_t) = d(x_u,x_t) \qquad (s,t,u \in N)$.

Rozwiązaniem problemu dotyczącego adekwatnej odpowiedniości jest następujące

/13/ <u>Twierdzenie</u>

Dla zbioru zmiennych wektorów $\{x_i\}_{i \in N}$ zestandaryzowanych nie na adekwatnej odpowiedniości między macierzą $R = \left(r(x_i,x_j)\right)$ współczynników korelacji, a macierzą $D = \left(d(x_i,x_j)\right)$ odległości wtedy i tylko wtedy, gdy w zbiorze $\{x_i\}_{i \in N}$ istnieje taka trójka elementów x_i, x_j, x_k, że

$$r_{ij} = r(x_i,x_j) < 0,$$

/*/ $$r_{ik} = r(x_i,x_k) < 0,$$

$$r_{jk} = r(x_j,x_k) < 0.$$

Na pytanie czy taka macierz współczynników korelacji istnieje odpowiemy w ciągu dalszym pracy.

<u>Dowód</u>

Załóżmy, że istnieje trójka zmiennych x_i, x_j, x_k, spełniająca /*/. Wykażemy, że nie mogą być spełnione postulaty /I/ i /II/. Istotnie, jeśli pomnożymy wszystkie wektory trójki przez -1, to nie będzie spełniony postulat /II/, ponieważ z równości

/14/ $$r(x_i,x_j) = \cos \sphericalangle (x_i,x_j)$$

wyńika, że współczynniki korelacji r_{ij}, r_{jk}, r_{ik} będą nadal ujemne i otrzymamy

$$\left|r(-x_i,x_j)\right| = \left|r(x_i,x_j)\right|,$$

a wobec /14/ i twierdzenia /7/ rozdziału II

$$d(-x_i,x_j) < d(x_i,x_j).$$

Analogicznie można wykazać, że nie będą spełnione postulaty, gdy tylko jeden lub dowolne dwa wektory z rozpatrywanej trójki pomnożymy przez -1.

Załóżmy teraz, że nie istnieje w zbiorze $\{x_i\}$ trójka wektorów spełniająca /*/, tzn. może być co najwyżej

$$r(x_i,x_j) < 0 \qquad \text{dla pewnego } x_i \text{ i } j \in N \setminus \{i\}$$

i

$$r(x_k,x_l) \geqslant 0 \qquad \text{dla dowolnych } k,\ l \in N \setminus \{i\}.$$

Wtedy mnożąc wektor x_i przez -1 otrzymamy

/**/ $$r(x_s,x_t) \geqslant 0 \qquad \text{dla dowolnych } s,t \in N.$$

Stąd równość

$$r(x_s,x_t) = \left|r(x_s,x_t)\right| = 1,$$

wobec twierdzenia /7/ rozdziału II, jest równoważna

$$d(x_s,x_t) = 0.$$

Ponadto, z /**/ wynika, że równość

$$\left|r(x_i,x_j)\right| = \left|r(x_k,x_j)\right|$$

jest równoważna następującej

$$r(x_i,x_j) = r(x_k,x_j),$$

co z kolei wobec twierdzenia /7/ rozdziału II jest równoważne

$$d(x_i,x_j) = d(x_k,x_j).$$

Twierdzenie zostało więc udowodnione.

Z twierdzenia /13/ wynika następujący

/15/ Wniosek

Jeśli w macierzy współczynników korelacji nie ma elementów spełniających /ж/, to warunki /I/ i /II/ będą spełnione jeśli wektory, którym odpowiadają kolumny /wiersze/ zawierające ujemne współczynniki pomnożymy przez -1.

Zagadnienie adekwatności można uogólnić w następujący sposób: czy istnieje taka miara m zależności zmiennych losowych, że

$$m(x,y) = a$$

implikuje zależność zmiennych losowych x,y, która spełniałaby postulaty:

$m(x,y) = a$ wtedy i tylko wtedy, gdy $d(x,y) = 0$;

$m(x,y) = m(z,y)$ wtedy i tylko wtedy, gdy $d(x,y) = d(z,y)$.

W ogólnym przypadku odpowiedź na to pytanie jest negatywna, ponieważ z aksjomatów metryki wynika, że

$$d(x,y) = 0$$

wtedy i tylko wtedy, gdy x = y, a zmienne losowe są zależne np. wtedy, gdy x = -y.

Twierdzenie /13/ nasuwa pytanie: czy może istnieć macierz współczynników korelacji, której elementy spełniają warunek /12/?

Otóż wykażemy, że macierz współczynników korelacji może zawierać same ujemne elementy /oprócz elementów głównej przekątnej/.

W praktyce mamy do czynienia z realizacjami zmiennych losowych, które są n-elementowymi ciągami liczb, przeto badanie możliwości istnienia macierzy R o samych ujemnych współczynnikach przeprowadzimy dla n-wymiarowych wektorów.

/16/ Twierdzenie

Warunkiem koniecznym na to aby była możliwa konstrukcja macierzy stopnia n współczynników korelacji o samych elementach ujemnych /oprócz elementów głównej przekątnej/ jest aby ilość realizacji była równa co najmniej n - 1.
Dowód tego twierdzenia poprzedzimy lematem.

/17/ Lemat

W przestrzeni unitarnej n-wymiarowej R^n istnieje co najwyżej n + 1 wektorów, dla których kąt między dowolnymi dwoma różnymi jest większy niż $\frac{\pi}{2}$

Dowód

1° W przestrzeni R^1 są dwa takie wektory, np. p i $-p$.

2° Załóżmy, że lemat jest prawdziwy w przestrzeni R^n. Rozważmy przestrzeń R^{n+1} jako sumę prostą przestrzeni R^n i prostopadłej do niej przestrzeni R^1. Z założenia w R^n jest $n+1$ wektorów $p_1,\dots,p_{n+1}$, dla których

$$\sphericalangle (p_i,p_j) > \frac{\pi}{2} \quad ,$$

dla $i \neq j$ $\quad (i,j = 1,\dots,n+1)$.

W przestrzeni R^1 są dwa wektory p i $-p$ takie, że

$$\sphericalangle (p,p_i) = \frac{\pi}{2} \ , \qquad \sphericalangle (-p,p_i) = \frac{\pi}{2} \ ,$$

dla $i = 1,\dots,n+1$.

Ponieważ

$$\sphericalangle (p_i,p_j) > \frac{\pi}{2} \qquad (i,j = 1,2,\dots,n+1)$$

każdy kąt $\sphericalangle (p,p_i) = \frac{\pi}{2}$ można zwiększyć /wynika to z ciągłości iloczynu skalarnego/ zachowując nierówność

$$\measuredangle(p, p_i) > \frac{\pi}{2} .$$

Otrzymamy przy tym

$$\measuredangle(-p, p_i) < \frac{\pi}{2} ,$$

a więc w przestrzeni R^{n+1} jest nie więcej niż n+2 wektorów, dla których kąty między dowolnymi dwoma różnymi są większe niż $\frac{\pi}{2}$, co kończy dowód lematu.

Dowód twierdzenia /16/ jest teraz natychmiastowy, jeśli weźmiemy pod uwagę równość /14/.

Wprowadźmy używany w analizie zbioru cech metodą wskaźników J. Perkala[1/] i omawiany w pracy B. Kopocińskiego i L. Zubrzyckiej [22], tzw. układ cech słabo zgodny.

/18/ Definicja

Zmienne losowe $\{X_i\}_{i \in N}$ stanowią układ zgodny, jeśli

---

1/ Porównaj: J. Perkal [33].

wszystkie współczynniki korelacji między nimi są nieujemne.

/19/ Definicja

Zmienne losowe $\{X_i\}_{i \in N}$ stanowią układ słabo zgodny, jeśli istnieje ciąg liczb $\{e_i\}_{i \in N}$ takich, że $e_i = \pm 1$ i zmienne $\{e_i X_i\}_{i \in N}$ stanowią układ zgodny.

Układ słabo zgodny można scharakteryzować w poniższy sposób.

/20/ Twierdzenie

Układ $C = \{X_i\}_{i \in N}$ jest słabo zgodny wtedy i tylko wtedy, gdy w macierzy współczynników korelacji

$$R = (r_{ij}) = \left(r(X_i, X_j)\right)$$

nie występuje taka trójka elementów, że

$$r_{ij} < 0, \quad r_{jk} < 0, \quad r_{jk} < 0. \tag{21}$$

Dowód

Jeśli w macierzy R występuje trójka elementów spełniająca /21/, to w zbiorze C istnieją trzy zmienne X_i, X_j, X_k takie, że kąt między dowolnymi dwoma jest większy niż $\frac{\pi}{2}$
Stąd też zmienne te nie stanowią układu słabo zgodnego, a więc i C nie jest układem słabo zgodnym.

Załóżmy teraz, że nie istnieje w zbiorze C trójka zmiennych spełniająca /21/, tzn. może być co najwyżej

$$r\left(X_i, X_j\right) < 0 \qquad \text{dla pewnego } X_i \text{ i } j \in N \setminus \{i\}$$

i

$$r\left(X_k, X_l\right) \geqslant 0 \qquad \text{dla dowolnych } k, l \in N \setminus \{i\}.$$

Wtedy mnożąc X_i przez -1 otrzymamy $r\left(X_s, X_t\right) \geqslant 0$ dla dowolnych $s, t \in N$, co kończy dowód.

Z twierdzeń /13/ i /20/ otrzymujemy jako wniosek następujące

/22/ Twierdzenie

Adekwatna odpowiedniość między macierzą R a macierzą odległości $D = \left(d\left(X_i, X_j\right)\right)$ istnieje wtedy i tylko wtedy, gdy układ $\left\{X_i\right\}_{i \in N}$ jest słabo zgodny.

§ 4. Przegląd metod dyskryminacji i ich zastosowań.

Taksonomia wrocławska sugeruje dyskryminowanie zbioru przez rozcinanie połączeń między skupieniami począwszy od połączenia najdłuższego. Nie występuje jednak w niej explicite kryterium, według którego można by obiektywnie rozstrzygnąć ile połączeń należy wyeliminować, tzn. na ile pozbiorów ma rozpaść się zbiór. Stąd, proponowane postępowanie będzie miało zastosowanie tylko w przypadku, gdy liczba podzbiorów jest z góry ustalona, co również może mieć praktyczny sens np. przy podziale dychotomicznym.

Metodą dyskryminacji wykorzystującą taksonomię wrocławską jest następujące postępowanie[1/].

1. Metodą taksonomii wrocławskiej wyznacza się malejący ciąg $\{d_1, \ldots, d_{n-1}\}$ długości połączeń w grafie $F(Z)$.
2. Z tak otrzymanego ciągu tworzy się ciąg

$$w_2 = \frac{d_1}{d_2}, \qquad w_3 = \frac{d_2}{d_3}, \ldots .$$

1/ Porównaj: Z. Hellwig [15]

/23/ Definicja

Zbiór Z rozpada się w sposób naturalny na k części, jeśli

$$w_k < w_{k+1}$$

W pracy [8] jest podany przykład przyrodniczy naturalnego rozpadu zbioru.

Wprowadźmy pewną modyfikację rozpadu naturalnego, polegającą na tym, że w ciągu $w_2, w_3, \ldots$ wybiera się element największy i jeśli jest to

$$w_s = \frac{d_{s-1}}{d_s},$$

zbiór Z dzielimy na s części eliminując s - 1 najdłuższych połączeń. Taki podział wydaje się być subtelniejszy niż rozpad naturalny.

Wymienione w § 1 żądanie aby w warstwach elementy były bardziej skoncentrowane niż w całym Z, realizuje metoda dyskryminacji polegająca na tym, że grafie $F(Z)$ eliminuje się wszystkie połączenia dłuższe niż średnia arytmetyczna $\bar{d}$ długości połączeń w $F(Z)$ /porównaj: Z. Hellwig [15]/.

Za miarę koncentracji zbioru punktów przyjmujemy tu średnią arytmetyczną długości połączeń w najkrótszym grafie rozpiętym na tym zbiorze, ponieważ o zaproponowanym postępowaniu można udowodnić

/24/ Twierdzenie

Średnia arytmetyczna długości połączeń w każdym z podgrafów /warstw/ $Z_1, \ldots, Z_k$ grafu $F(Z)$ powstałych przez eliminację $k-1$ połączeń o długościach

$$d_1 > \ldots > d_{k-1} ,$$

takich, że

$$d_{k-1} < \bar{d} \quad i \quad d_k > \bar{d}$$

jest mniejsza niż $\bar{d}$.

Dowód

Oznaczmy przez $\bar{d}_i$ średnią arytmetyczną połączeń w $D(Z_i)$.

Z twierdzenia /10/ wynika, że podgrafy $Z_1, \ldots, Z_k$ powstałe w wyniku eliminacji $k - 1$ połączeń są najkrótsze, stąd w podgrafie Z_i $(i = 1, \ldots, k)$ wszystkie połączenia są krótsze niż $\bar{d}$, a więc

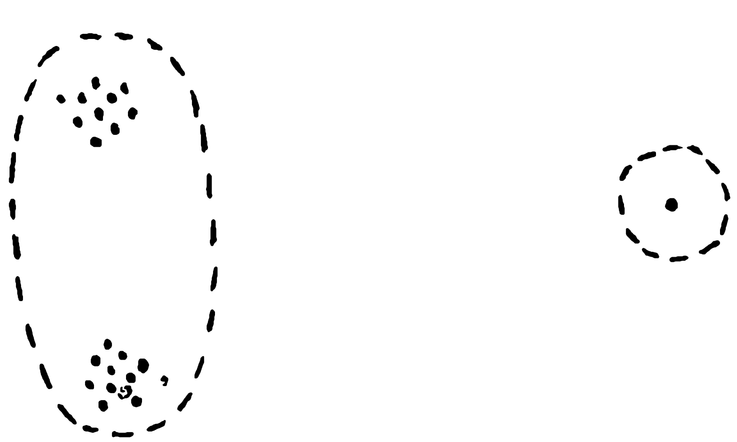
$$\bar{d}_i < \bar{d} \qquad \text{dla } i = 1, \ldots, k ,$$

co należało udowodnić.

/25/ Uwaga

Przedstawiona metoda dyskryminacji jest podziałem najlepszym na k części /porównaj: definicja /8//.

Jednakże wadą tej metody jest, że zbiór



zostanie rozdyskryminowany na dwa podzbiory, co może budzić kontrowersje. Tę niedogodność łagodzi eliminowanie wszystkich połączeń dłuższych niż $d - s_d$, gdzie s_d jest odchyleniem standardowym długości połączeń w $F(Z)$.

E. Stolarska, K. Zadora i J. Kolonko w [35] przedstawili metodę dyskryminacji polegającą na tym, że zbiór Z wpisuje się w siatkę utworzoną przez hipersześciany. Warstwy tworzą takie podzbiory zbioru Z, które są oddzielone od siebie hipersześcianami nie zawierającymi elementów zbioru Z. Za miarę koncentracji przejmiemy tu liczbę elementów przypadającą na jednostkę obszeru. Miarę tę można nazwać gęstością. Łatwo udowodnić, że gęstość ta jest większa w warstwach niż w całym zbiorze Z.

Inne podejście do problemu dyskryminacji prezentuje praca W. Bukietyńskiego i innych [2]. Wprowadzono tam tzw. kryterium dyskryminacji - tj. funkcję rzeczywistą F określoną na rodzinie możliwych klas podziałów na warstwy, przy czym przez dyskryminację optymalną rozumie się taki podział, dla którego funkcja F osiąga maksimum. W [2] zwrócono również uwagę na fakt, że każda dyskryminacja zbioru Z jest podziałem Z na klasy abstrakcji[1/], a jak wiadomo czyni to dowolna relacja równoważności /podobieństwa/ określona w zbiorze Z. Można np. przyjąć, że obiektami podobnymi są takie, które są zależne liniowo.
Jeśli jednak cechy, za pomocą których opisane obiekty są realizacjami zmiennych losowych ciągłych to prawdopodobieństwo, że dwa obiekty będą w tym sensie podobne, jest równe zero. Należy więc

1/ Porównaj np.: A. Grzegorczyk [12].

wprowadzić miarę podobieństwa i uważać za podobne takie obiekty, dla których owa miara przyjmie wartość mniejszą niż arbitralnie ustalona wielkość. Może to być np. $2s_m$, gdzie s_m jest odchyleniem standardowym miary. Za miarę podobieństwa w sensie liniowej zależności można przyjąć wartość bezwzględną cosinusa[1] kąta między wektorami reprezentującymi obiekty.

/26/ Uwagi o zastosowaniu dyskryminacji.

Znane są zastosowania dyskryminacji w zagadnieniach przyrodniczych /porównaj np.: K. Florek i inni [8] / i ekonomicznych np. do typologicznego podziału kraju /porównaj: Z. Hellwig [15] /, wyodrębniania regionów gospodarczych itp. Oprócz wymienionych aplikacji, dyskryminację można stosować w zagadnieniach programowania matematycznego dotyczących produkcji lub magazynowania, w których pojawia się konieczność agregowania wyrobów, a także w zagadnieniach klasyfikacji i oceny jakości wyrobów, których nie można uporządkować liniowo.

1/ Cosinus ten jest jak wiadomo równy pewnych przypadkach współczynnikowi korelacji, a więc można mówić o współczynniku korelacji między obiektami.

§ 5. Wybór obiektu optymalnego.

W paragrafie tym będziemy rozróżniali zbiór obiektów $\mathcal{Z}$ i utożsamiany z nim dotąd zbiór $Z \subset E^p$.

Z zagadnieniem dyskryminacji wiąże się problem obiektu "najlepszego" w zbiorze $\mathcal{Z}$ obiektów, lub w wybranej warstwie rozdyskryminowanego zbioru $\mathcal{Z}$. Zwykle za najlepszy uważa się taki element zbioru, dla którego pewna funkcja na nim określona osiąga ekstremum. Funkcja, która przyporządkowuje obiektom p - elementowe ciągi liczb będące realizacjami cech, jest funkcją wektorową

$$f : \mathcal{Z} \to E^p.$$

/27/ Definicja[1]

Mówimy, że funkcja wektorowa f osiąga w punkcie $L \in \mathcal{Z}$ maksimum, jeśli

$$f(L) \succ f(K) \qquad \text{dla każdego } K \in \mathcal{Z} \quad ,$$

przy czym relacja $\succ$ między wektorami oznacza, że między odpowiednimi składowymi wektorów zachodzi nierówność." $\geqslant$ " ; punkt L

---
[1] Porównaj: R. Kulikowski [24]

nazywa się często optymalny w sensie Pareto.

Tak zdefiniowane optimum wymaga założenia, że rozpatrywane cechy obiektów mają charakter progresywny. Spełnienie tego założenia w praktyce jest łatwe, uzyskuje się je przez odpowiednią zmianę miar cech.

Z optimum w sensie Pareto wiąże się w następujący sposób problem wyboru wag dla cech.

/28/ <u>Twierdzenie</u>[1/]

Jeśli element $\dot{X}_0$ należący do wypukłego zbioru $\mathcal{X}$ jest optymalny w sensie Pareto, to istnieje taki układ wag

$$e = \left(e_1, \dots, e_p\right),$$

gdzie

$$\sum_{i=1}^{p} e_i = 1, \quad e_i \geqslant 0,$$

1/ Porównaj: R. Kulikowski [24] .

że iloczyn skalarny

$$\left(e \mid f(X)\right)$$

przyjmie wartość największą dla elementu $X_0 \in \mathcal{Z}$.

Zbiór cech jest na ogół zbiorem skończonym, a więc nie zawsze istnieje w nim element, dla którego funkcja f osiąga maksimum, tzn. obiekt optymalny w sensie Pareto.

Jednym z możliwych sposobów sformułowania definicji obiektu optymalnego w taki sposób aby należał on do zbioru $\mathcal{Z}$, jest zaproponowana przez Z. Hellwiga wprowadzenie abstrakcyjnego obiektu - przeciwobrazu punktu

$$\mathfrak{a} = \left(x_{o1}, x_{o2}, \ldots, x_{op} \right) \in E^p,$$

gdzie

$$x_{oi} = \max_j x_{ji} , \quad \text{gdy i-ta cecha jest progresywna;}$$

$$x_{ok} = \min_j x_{jk} , \quad \text{gdy k-ta cecha jest regresywna}$$

/29/ Definicja

Obiekt B $\in \mathcal{Z}$ nazywa się $\mathfrak{z}$ - optymalny, jeśli odległość $\varphi(f(B), f(\mathfrak{a}))$ jego obrazu $f(B) \in E^p$ od $\mathfrak{a}$ jest najmniejsza, przy czym φ niekoniecznie musi być metryką euklidesową.

Jeśli każdemu obiektowi należącemu do $\mathcal{Z}$ przyporządkujemy liczbę będącą odległością jego obrazu od $\mathfrak{a}$, to obiekty można będzie między sobą porównywać. Zauważmy przy tym, że relacja $\prec$ między obiektami K, L $\in \mathcal{Z}$ zdefiniowana przez odległości ich obrazów od punktu $\mathfrak{a}$ w następujący sposób

$$K \prec L$$

wtedy i tylko wtedy, gdy

$$\varphi(f(K), \mathfrak{a}) \leq \varphi(f(L), \mathfrak{a})$$

nie jest relacją porządkującą[1/]. Jest ona wprawdzie zwrotna i przechodnia, ale nie jest antysymetryczna, tzn. nie spełnia warunku: jeśli $K \prec L$ i $L \prec K$, to $K = L$.

1/ Porównaj: H. Rasiowa [34]

Przedstawimy inną definicję obiektu optymalnego opartą na procedurze rozwiązywania zagadnień programowania wielocelowego, zaproponowaną przez S. Bartosiewiczową[1/].

Ponumerujmy liczbami naturalnymi składowe punktów zbioru $f(\mathcal{Z}) \subset E^p$ rosnąco dla cech progresywnych i malejąco dla cech regresywnych. Niech $b(K)$, $(K \in \mathcal{Z})$ oznacza sumę liczb naturalnych przyporządkowanych w wymieniony sposób poszczególnym składowym punktu $f(K)$.

/30/ <u>Definicja</u>

Obiekt $B \in \mathcal{Z}$ nazywa się γ - optymalny, jeśli $b(B)$ jest wartością największą funkcji b w zbiorze $\mathcal{Z}$

Wprowadzając funkcję b, można obiekty między sobą porównywać. Polega to na porównywaniu wartości funkcji b na obiektach.

Kryteria, według których wybieramy obiekt optymalny w sensie /29/ lub /30/ można nazwać, w odróżnieniu od występujących w programowaniu matematycznym, kryteriami wewnętrznymi. Wybór obiektu optymalnego w rozumieniu tych kryteriów zależy od postaci zbioru obiektów.

Metody wyboru obiektu optymalnego, w sensie którejkolwiek

1/ Porównaj: Elementy rachunku ekonomicznego [4] z powyższych definicji i wynikające z nich możliwości porównywania obiektów, można /oprócz zastosowań omówionych w [8] i [15] / naszym zdaniem wykorzystać do wyboru najlepszego produktu w danej klasie produktów lub najlepszej technologii a także do przewidywania kierunków rozwoju postępu technicznego. Problemy te są istotne w przypadku podejmowania decyzji dotyczących np. zakupu licencji.

O przydatności definicji /29/ lub /30/ do rozwiązania konkretnego problemu ekonomicznego decydują podstawowe założenia polityki ekonomicznej, jak również charakter problemu.

B I B L I O G R Á F I A

[1] T.Anderson: Wwiedenije w mnogomiernyj statisticzeskij analiz, Moskwa 1965

[2] W.Bukietyński, Z. Hellwig, U. Królik, A. Smoluk: Uwagi o dyskryminacji zbiorów skończonych, Zeszyty Naukowe WSE, 1971

[3] H. Cramer: Metody matematyczne w statystyce, Warszawa 1958

[4] Elementy rachunku ekonomicznego /praca zbiorowa/, Warszawa 1972

[5] W.N. Faddiejewa: Metody numeryczne algebry liniowej, Warszawa 1955

[6] M. Fisz: Rachunek prawdopodobieństwa i statystyka matematyczna, Warszawa 1965

[7] K. Florek, J, Łukaszewicz, J. Perkal, H, Steinhaus, S. Zubrzycki: Sur la liaison et la division des points d'un ensemble fini, Colloquium Matematice, 3 - 4, 1951

[8] Taksonomia wrocławska, Przegląd Antropologiczny, nr 27, 1951

[9] F.R. Gantmacher: Teoria matric, Moskwa 1966

[10] I.I. Gichman, A.W. Skorochod: Wstęp do teorii procesów stochastycznych, Warszawa 1968

[11] A.S. Golberger: Teoria ekonometrii, Warszawa 1972

[12] A. Grzegorczyk: Zarys logiki matematycznej, Warszawa 1961

[13] Z. Hellwig: Elementy rachunku prawdopodobieństwa i statystyki matematycznej, Warszawa 1970

[14] - : Problem optymalnego wyboru predyktant, Przegląd Statystyczny, T.XVI, z. 3 - 4, 1969

[15] - : Procedure of evaluating high level manpower data and topology of countries by mean of the taxonomic method, UNESCO, Paris 1967

[16] R. Jasiński: Interpretacja niektórych pojęć statystyki matematycznej w przestrzeni Hilberta, w druku w Zeszytach Naukowych WSE

[17] - : O wyborze cech charakterystycznych, w druku w Zeszytach Naukowych WSE

[18] - : Pewna metoda optymalnego wyboru zmiennych, w druku w Zeszytach Naukowych WSE

[19] - Wybrane zagadnienia zastosowań metod matematycznych w zarządzaniu, Wrocław 1968

[20] - : O pewnym twierdzeniu teorii dyskryminacji, Zeszyty Naukowe WSE nr 33, 1972

[21] J.G. Kemeny: Nauka w oczach filozofa, Warszawa 1967

[22] B. Kopociński, L. Zubrzycka: Uwaga o podziale zespołu cech na podzespoły zgodne, Zastosowania Matematyki nr. 7, 1964

[23] G. Korn. T.Korn: Sprawocznik po matiematikie, Moskwa 1970

[24] R. Kulikowski: Sterowanie w wielkich systemach, Warszawa 1970

[25] M. Loève: Teoria wierojatnostiej, Moskwa 1962

[26] D. Luenberger: Optimization by Vector Space Methods, New York 1969

[27] A.P. Miszina, I.W. Proskuriakow: Algebra wyższa, Warszawa 1966

[28] W. Mlak: Wstęp do teorii przestrzeni Hilberta, Warszawa 1970

[29] A. Mostowski, M. Stark: Elementy algebry wyższej, Warszawa 1968

[30] Nowyje idei w płanirowanii ekspierymienta, Moskwa 1969 /praca zbiorowa/

[31] O. Ore: Wstęp do teorii grafów, Warszawa 1966

[32] Z. Pawłowski: Wstęp do statystyki matematycznej, Warszawa 1966

[33] J. Perkal: On the Analysis of a set of characteristics, Zastosowania Matematyki nr V, 1960

[34] H. Rasiowa: Wstęp do matematyki współczesnej, Warszwa 1968

[35] J. Kolonko, E. Stolarska, K. Zadora: Prosta metoda dyskryminacji zbiorów skończonych, Przegląs Statystyczny 1970

[36] W. Tatarkiewicz: Historia filozofii, Warszawa 1970

[37]. G. Trybuś: Numeryczne aspekty wyznaczania dystansów /praca doktorska/, 1970

[38] S. Zubrzycki: Wykłady z rachunku prawdopodobieństwa i statystyki matematycznej, Warszawa 1970

[39] W. Ostasiewicz: Program WYBÓR CECH. Program ten znajduje się w Bibliotece Programów Ośrodka Obliczeniowego WSE we Wrocławiu

[40] - : Program JARY. Program ten znajduje się w Bibliotece Programów Ośrodka Obliczeniowego WSE we Wrocławiu.