POLSKA AKADEMIA NAUK
Instytut Badań Systemowych

BADANIA OPERACYJNE I SYSTEMOWE: ŚRODOWISKO NATURALNE, PRZESTRZEŃ, OPTYMALIZACJA

Olgierd Hryniewicz,
Andrzej Straszak,
Jan Studziński
red.

BADANIA OPERACYJNE I SYSTEMOWE: ŚRODOWISKO NATURALNE, PRZESTRZEŃ, OPTYMALIZACJA

INSTYTUT BADAŃ SYSTEMOWYCH • POLSKA AKADEMIA NAUK

Seria: BADANIA SYSTEMOWE
tom 63

Redaktor naukowy:

Prof. dr hab. inż. Jakub Gutenbaum

Warszawa 2008

Olgierd Hryniewicz, Andrzej Straszak, Jan Studziński

BADANIA OPERACYJNE I SYSTEMOWE: ŚRODOWISKO NATURALNE, PRZESTRZEŃ, OPTYMALIZACJA

Publikacja była opiniowana do druku przez zespół recenzentów, którego skład podano w treści tomu

ISBN 83-894-7519-7
EAN 9788389475190

Redakcja i opracowanie techniczne: Jan W. Owsiński, Aneta M. Pielak, Anna Gostyńska

Lista recenzentów artykułów, wchodzących w skład tomów serii „Badania Systemowe” związanych z konferencją BOS 2008

Dr Paweł Bartoszczuk
Dr inż. Lucyna Bogdan
Dr hab. inż. Zbigniew Buchalski
Mgr inż. Hanna Bury
Prof. dr hab. Marian Chudy
Dr Jan Gadomski
Mgr Grażyna Grabowska
Mgr inż. Andrzej Jakubowski
Dr hab. inż. Ignacy Kaliszewski
Dr Andrzej Kałuszko
Dr hab. Leszek Klukowski
Dr hab. inż. Wiesław Krajewski
Dr inż. Lech Kruś
Dr hab. inż. Marek Libura
Dr Barbara Mażbic-Kulma
Dr inż. Edward Michalewski
Dr inż. Jan W. Owsiński
Dr inż. Grażyna Petriczek
Dr inż. Henryk Potrzebowski
Dr Maciej Romaniuk
Prof. dr hab. Piotr Sienkiewicz
Dr hab. Henryk Spustek
Prof. dr hab. Andrzej Straszak
Dr hab. inż. Jan Studziński
Prof. dr hab. Tomasz Szapiro
Mgr Anna Szediw
Dr inż. Grażyna Szkatuła
Dr hab. inż. Tadeusz Witkowski
Dr Irena Woroniecka-Leciejewicz
Dr hab. Sławomir Zadrożny
Dr inż. Andrzej Ziółkowski

Komitety Konferencji
Badania Operacyjne i Systemowe 2008
Rembertów, Akademia Obrony Narodowej

Patronat honorowy

Bogdan Klich, Minister Obrony Narodowej
Maciej Nowicki, Minister Środowiska i Zasobów Naturalnych

Komitet Sterujący

Janusz Kacprzyk, Prezes Polskiego Towarzystwa Badań Operacyjnych i Systemowych
Olgierd Hryniewicz, Dyrektor Instytutu Badań Systemowych
Janusz Kręcikij, Komendant Akademii Obrony Narodowej

Komitet Programowy

Piotr Sienkiewicz, *Przewodniczący*
Jacek Mercik, *Wiceprzewodniczący*

Tomasz Ambroziak
Marian Chudy
Olgierd Hryniewicz
Ignacy Kaliszewski
Piotr Kulczycki
Barbara Mażbic-Kulma
Włodzimierz Ogryczak
Andrzej Piegat
Honorata Sosnowska
Andrzej Straszak
Ryszard Tadeusiewicz
Jan Węglarz
Ryszard Budziński
Ludosław Drelichowski
Adam A. Janiak
Józef Korbicz
Małgorzata Łatuszyńska
Zbigniew Nahorski
Wojciech Olejniczak
Krzysztof Santarek
Henryk Spustek
Tomasz Szapiro
Eugeniusz Toczyłowski
Tadeusz Witkowski
Bogdan Zdrodowski
Wojciech Cellary
Jerzy Hołubiec
Jerzy Józefczyk
Maciej Krawczak
Marek J. Malarski
Andrzej Najgebauer
Jan W. Owsiński
Roman Słowiński
Jan Stachowicz
Andrzej Szymonik
Tadeusz Trzaskalik
Stanisław Zajas

Komitet Organizacyjny

Jan W. Owsiński, Andrzej Kałuszko, Mieczysław Pelc, Zbigniew Piątek

Sekretariat

Krystyna Warzywoda, Monika Majkut, Aneta M. Pielak, Krzysztof Sęp, Anna Stachowiak, Halina Świeboda, Tadeusz Winiarski

Redakcja wydawnictw

Janusz Kacprzyk, Piotr Sienkiewicz, Andrzej Najgebauer, Olgierd Hryniewicz, Andrzej Straszak, Jan Studziński, Jan W. Owsiński, Zbigniew Nahorski, Tomasz Szapiro

Metody:
Optymalizacja, dane, analiza

PORZĄDKOWANIE I GRUPOWANIE OBIEKTÓW JAKO ZAGADNIENIE ESTYMACJI STATYSTYCZNEJ

Leszek Klukowski
Instytut Badań Systemowych PAN, Newelska 6, 01-447 Warszawa
Leszek.Klukowski@ibspan.waw.pl

W pracy przedstawiono metody estymacji relacji: porządku, tolerancji i równoważności, na podstawie porównań parami z błędami losowymi, przy wykorzystaniu koncepcji *nearest adjoining order* (NAO), sformułowanej przez Slatera (1961), zob. też David (1988). Przedstawione wyniki stanowią zwięzłą syntezę prac autora, Klukowski (1990, 1994, 2000, 2002, 2007a,b,c).

1. Wstęp

Porządkowanie i grupowanie obiektów na podstawie porównań parami z błędami losowymi ma na celu określenie relacji: ∗ porządku, ∗ tolerancji lub ∗ równoważności w zbiorze skończonym. Konieczność określenia tych relacji zachodzi w badaniach wykonywanych w wielu dziedzinach, m.in. w ekonomii, socjologii, medycynie. Do wyznaczania tych relacji stworzono wiele metod i algorytmów wychodząc z różnych koncepcji metodologicznych. Wiele spośród tych metod opiera się na koncepcjach heurystycznych, co sprawia, że niewiele wiadomo o własnościach wyników. Z kolei metody umożliwiające określenie własności wyników wymagają zwykle mocnych założeń o danych wejściowych (porównań parami); nie zawsze można zweryfikować słuszność tych założeń. Pożądaną cechą metod estymacji relacji, z punktu widzenia potrzeb praktyki, jest koniunkcja słabych założeń odnośnie danych wejściowych i korzystnych właściwości wyników.

W przypadku, gdy opis danych wejściowych, służących do wyznaczenia relacji, ma postać probabilistyczną, określenie relacji stanowi przedmiot badań statystyki matematycznej; może mieć postać zagadnienia estymacji statystycznej. Ważną klasą estymatorów dla takich problemów są estymatory określające „najbliższy sąsiedni porządek” (ang. *nearest adjoining order* - NAO), oparte na porównaniach par elementów rozważanego zbioru, zakłóconych błędami losowymi. Koncepcja ta została wprowadzona przez Slatera w 1961 r. i była rozwijana przez wielu autorów. Koncepcja estymatorów NAO jest zgodna z ogólną ideą estymacji statystycznej - polega na wyznaczeniu relacji o takiej postaci, która wykazuje najmniejsze różnice (lub największe podobieństwo), zgodnie z przyjętym kryterium, w stosunku do próby. Próbą są w tym przypadku wyniki porównań parami. Nie jest to zwykle próba prosta, ponieważ porównania poszczególnych par mogą nie być niezależne stochastycznie.

Celem niniejszej pracy jest przedstawienie konstrukcji estymatorów opartych na koncepcji NAO oraz ich podstawowych własności, w przypadku wymienionych powyżej rodzajów relacji.

2. Koncepcja i ogólna postać estymatorów

Idea konstrukcji estymatorów NAO jest następująca. Każda z rozważanych relacji ma postać rodziny podzbiorów $\chi_1^{(f)*}, ..., \chi_n^{(f)*}$, n>1 (indeks f – określa rodzaj relacji: p – odpowiada relacji porządku, t – tolerancji, e - równoważności) skończonego zbioru $\mathbf{X}=\{x_1, ..., x_m)$ ($m\geq 3$); w przypadku relacji porządku rodzina ta jest ciągiem (indeks podzbioru określa rangę zawartych w nim elementów). Dla każdej pary $(x_i, x_j)\in \mathbf{X}\times\mathbf{X}$ dane są wyniki porównań $g_{\ell k}(x_i,x_j)$ (k=1, ..., N; $N\geq 1$, indeks ℓ - oznacza rodzaj porównania: binarne lub wielowartościowe), określających relację w parze elementów (x_i, x_j), z możliwością błędu losowego. Porównanie może być dokonane przy użyciu testu statystycznego, sieci neuronowej, przez eksperta, itp. Przykładowo, w przypadku relacji równoważności porównanie binarne $g_{\ell k}(x_i,x_j)$ stwierdza przynależność pary (x_i, x_j) do wspólnego podzbioru ($g_{\ell k}(x_i,x_j)=0$) lub brak przynależności ($g_{\ell k}(x_i,x_j)=1$). Porównanie wielowartościowe może wyrażać, w przypadku relacji porządku, odległość między elementami x_i oraz x_j, np. różnicę rang (miejsc w uporządkowaniu). Relację $\chi_1^{(f)*}, ..., \chi_n^{(f)*}$ można scharakteryzować za pomocą pewnej funkcji $T_\ell(x_i,x_j)$, o postaci analogicznej do porównań $g_{\ell k}(x_i,x_j)$; różnica $T_\ell(x_i,x_j)$ - $g_{\ell k}(x_i,x_j)$ wyraża wówczas losowy błąd k-tego porównania pary (x_i, x_j) (zero oznacza porównanie bezbłędne).

Przyjmuje się nie ograniczające założenia nt. stochastycznych własności błędów porównań. W przypadku porównań binarnych zakłada się, że prawdopodobieństwo bezbłędnego wyniku jest większe niż błędnego. W przypadku porównań wielowartościowych zakłada się, że funkcja prawdopodobieństwa jest jednomodalna z medianą i modą w zerze oraz, że w przypadku porównań wielokrotnych (N>1) mają one własność stochastycznej niezależności:

$$P[(g_{\ell k}(x_i,x_j)=T_\ell(x_i,x_j))\cap(g_{\ell \nu}(x_r,x_s)=T_\ell(x_r,x_s))]=$$
$$P[(g_{\ell k}(x_i,x_j)=T_\ell(x_i,x_j)]\,P[(g_{\ell \nu}(x_i,x_j)=T_\ell(x_r,x_s)],\ (x_i, x_j, x_r, x_s\in \mathbf{X};\ k\neq\nu).(1)$$

Spełnienie powyższych założeń można weryfikować za pomocą testów statystycznych. Wyznaczenie wyniku estymacji, tj. oceny $\hat{\chi}_1^{(f)}, ..., \hat{\chi}_{\hat{n}}^{(f)}$, polega na określeniu takiej jej postaci, która wykazuje najniższy poziom niezgodności z porównaniami $g_{\ell k}(x_i,x_j)$. Poziom niezgodności można wyrazić, dla (dowolnej) relacji $\chi_1, ..., \chi_r$ oraz ustalonego k ($1\leq k\leq N$), w postaci:

$$\sum_{(x_i,x_j)\in\mathbf{X}\times\mathbf{X}}\left|t_\ell(x_i,x_j)-g_{\ell k}(x_i,x_j)\right|,$$

przy czym $t_\ell(x_i,x_j)$ – funkcja charakteryzująca postać relacji $\chi_1, ..., \chi_r$, analogiczna do funkcji $T_\ell(x_i,x_j)$ (charakteryzującej relację $\chi_1^{(f)*}, ..., \chi_n^{(f)*}$).

Wynik estymacji otrzymuje się, w przypadku N=1, jako rozwiązanie zadania programowania dyskretnego o postaci:

$$\min_{F_{\mathrm{X}}^{(\cdot)}} \sum_{(x_i,x_j)\in R_m} \left| t_\ell^{(\iota)}(x_i,x_j) - g_{\ell 1}(x_i,x_j) \right|, \tag{2}$$

gdzie: $F_{\mathrm{X}}^{(\cdot)}$ – zbiór rozwiązań dopuszczalnych, tj. wszystkich relacji danego rodzaju w zbiorze **X**,

R_m – zbiór par indeksów o postaci $R_m=\{<i,j> \mid 1\le i,j\le m; j>i\}$,

$t_\ell^{(\iota)}(x_i,x_j)$ – funkcja charakteryzująca ι-ty element zbioru $F_{\mathrm{X}}^{(\cdot)}$,

$g_{\ell 1}(x_i,x_j)$ – wynik porównania pary (x_i, x_j).

W przypadku porównań wielokrotnych (N>1), wyznaczenie wyniku estymacji jest dokonywane w podobny sposób, przy czym ulegają modyfikacji wyrażenia

$\mid t_\ell^{(\iota)}(x_i,x_j) - g_{\ell 1}(x_i,x_j) \mid$; są one zastępowane pewnymi ich funkcjami, np.:

$\frac{1}{N}\sum_{k=1}^{N} \mid t_\ell^{(\iota)}(x_i,x_j) - g_{\ell k}(x_i,x_j) \mid$ (uśrednienie różnic bezwzględnych),

$\mid t_\ell^{(\iota)}(x_i,x_j) - g_{\ell,me}(x_i,x_j) \mid$ (różnice bezwzględne z zastosowaniem mediany),

gdzie: $g_{\ell,me}(\cdot)$ - mediana z wyników porównań $g_{\ell 1}(\cdot)$, ..., $g_{\ell N}(\cdot)$, N – nieparzyste.

W niniejszej pracy przedstawiono estymatory oparte na uśrednianiu różnic bezwzględnych; zastosowanie mediany prowadzi do estymatora analogicznego do przypadku pojedynczego porównania każdej pary (N=1), przy czym rozkłady błędów $T_\ell(\cdot) - g_{\ell,me}(\cdot)$ są różne od rozkładów błędów $T_\ell(\cdot) - g_{\ell k}(\cdot)$ (zob. Klukowski, 2007a).

3. Estymacja relacji porządku

Relacja słabego porządku $\boldsymbol{R}^{(p)}$ w zbiorze **X** ma postać:

$$\boldsymbol{R}^{(p)} = \boldsymbol{I} \cup \boldsymbol{P} \tag{3}$$

gdzie: $\boldsymbol{I}$ - relacja równoważności,

$\boldsymbol{P}$ – relacja mocnej preferencji (przechodnia, antysymetryczna).

Relacja $\boldsymbol{R}^{(p)}$ generuje ciąg podzbiorów $\chi_1^{(p)*}, ..., \chi_n^{(p)*}$, spełniających warunki:

$$\bigcup_{i=1}^{n} \chi_i^{(p)*} = \mathbf{X}, \quad \chi_r^{(p)*} \cap \chi_s^{(p)*} = \{\mathbf{0}\} \qquad (r\neq s); \tag{4}$$

jeśli para $(x_i, x_j)\in \chi_q^{(p)*}$ $(1\le q\le n)$, to elementy x_i oraz x_j są równoważne, jeśli $x_i\in \chi_r^{(p)*}$ oraz $x_j\in \chi_s^{(p)*}$, przy czym $r<s$, to element x_i jest preferowany w stosunku do elementu x_j, a różnica rang wynosi: r-s.

3.1. Estymacja relacji porządku dla porównań wskazujących kierunek preferencji

Relację $\boldsymbol{R}^{(p)}$ można scharakteryzować jednoznacznie przy użyciu funkcji T_1: $\mathbf{X}\times\mathbf{X} \to D$, $D = \{-1, 0, 1\}$, przy czym:

$$T_1(x_i, x_j) = \begin{cases} 0 & \text{jeśli istnieje } q \text{ takie, że } (x_i, x_j) \in \chi_q^{(p)*}; \\ -1 & \text{jeśli } x_i \in \chi_r^{(p)*} \text{ oraz } x_j \in \chi_s^{(p)*},\ r<s; \\ 1 & \text{jeśli } x_i \in \chi_r^{(p)*} \text{ oraz } x_j \in \chi_s^{(p)*},\ r>s. \end{cases} \tag{5}$$

Zakłada się, że porównania mają postać $g_{1k}: \mathbf{X} \times \mathbf{X} \to D$, $D=\{-1, 0, 1\}$:

$$g_{1k}(x_i, x_j) = \begin{cases} 0 & \text{jeśli istnieje } q \text{ takie, że } (x_i, x_j) \in \chi_q^{(p)*}; \\ -1 & \text{jeśli } x_i \in \chi_r^{(p)*} \text{ oraz } x_j \in \chi_s^{(p)*},\ r<s; \\ 1 & \text{jeśli } x_i \in \chi_r^{(p)*} \text{ oraz } x_j \in \chi_s^{(p)*},\ r>s. \end{cases} \tag{6}$$

Zakłada się ponadto, że wyniki porównań są obarczone błędami losowymi, a funkcja prawdopodobieństwa błędów spełnia warunki:

$$P(g_{1k}(x_i, x_j) \neq T_1(x_i, x_j)) \leq \delta^{(k)}, \qquad (k=1, ..., N), \qquad \delta^{(k)} \leq \delta, \qquad \delta \in (0, ½) \tag{7}$$

oraz są stochastycznie niezależne, tzn. spełniają zależność (1).

W celu określenia postaci i własności estymatora relacji (w przypadku uśredniania porównań), niezbędne są definicje następujących zmiennych losowych:

$$U_{kij}^{(p)}(\chi_1^{(p)}, ..., \chi_r^{(p)}) = \begin{cases} 0 - \text{jesli } g_{1k}(x_i, x_j) = 0 \text{ oraz } <i, j> \in I(\chi_1^{(p)}, ..., \chi_r^{(p)}), \\ 1 \text{ jesli } |g_{1k}(x_i, x_j)| = 1 \text{ oraz } <i, j> \in I(\chi_1^{(p)}, ..., \chi_r^{(p)}), \end{cases} \tag{8}$$

$$V_{kij}^{(p)}(\chi_1^{(p)}, ..., \chi_r^{(p)}) = \begin{cases} 0 \text{ jesli } g_{1k}(x_i, x_j) = -1 \text{ oraz } <i, j> \in P_1(\chi_1^{(p)}, ..., \chi_r^{(p)}), \\ 1 \text{ jesli } g_{1k}(x_i, x_j) \geq 0 \text{ oraz } <i, j> \in P_1(\chi_1^{(p)}, ..., \chi_r^{(p)}), \end{cases}$$

(9)

$$Z_{kij}^{(p)}(\chi_1^{(p)}, ..., \chi_r^{(p)}) = \begin{cases} 0 \text{ jesli } g_{1k}(x_i, x_j) = 1 \text{ oraz } <i, j> \in P_2(\chi_1^{(p)}, ..., \chi_r^{(p)}), \\ 1 \text{ jesli } g_{1k}(x_i, x_j) \leq 0 \text{ oraz } <i, j> \in P_2(\chi_1^{(p)}, ..., \chi_r^{(p)}), \end{cases}$$

(10)

przy czym:

$$I(\chi_1^{(p)}, ..., \chi_r^{(p)}) = \{<i, j> \mid (x_i, x_j) \in \chi_q^{(p)}, 1 \leq q \leq r, j > i\}, \tag{11}$$

$$P_1(\chi_1^{(p)}, ..., \chi_r^{(p)}) = \{<i, j> \mid x_i \in \chi_r^{(p)}, x_j \in \chi_s^{(p)}, r < s, j > i\}, \tag{12}$$

$$P_2(\chi_1^{(p)}, ..., \chi_r^{(p)}) = \{<i, j> \mid x_i \in \chi_r^{(p)}, x_j \in \chi_s^{(p)}, r > s\ j > i\}, \tag{13}$$

oraz:

$$U_{ij,N}^{(p)}(\chi_1^{(p)}, ..., \chi_r^{(p)}) = \tfrac{1}{N} \sum_{k=1}^{N} U_{kij}^{(p)}(\chi_1^{(p)}, ..., \chi_r^{(p)}), \tag{14}$$

$$V_{ij,N}^{(p)}(\chi_1^{(p)}, ..., \chi_r^{(p)}) = \tfrac{1}{N} \sum_{k=1}^{N} V_{kij}^{(p)}(\chi_1^{(p)}, ..., \chi_r^{(p)}), \tag{15}$$

$$Z_{ij,N}^{(p)}(\chi_1^{(p)}, ..., \chi_r^{(p)}) = \tfrac{1}{N} \sum_{k=1}^{N} Z_{kij}^{(p)}(\chi_1^{(p)}, ..., \chi_r^{(p)}), \tag{16}$$

$$\overline{W}_N^{(p)}(\chi_1^{(p)}, ..., \chi_r^{(p)}) = \sum_{I(\chi_1^{(p)}, ..., \chi_r^{(p)})} \overline{U}_{ij,N}^{(p)}(\chi_1^{(p)}, ..., \chi_r^{(p)}) + \sum_{P_1(\chi_1^{(p)}, ..., \chi_r^{(p)})} \overline{V}_{ij,N}^{(p)}(\chi_1^{(p)}, ..., \chi_r^{(p)}) + \sum_{P_2(\chi_1^{(p)}, ..., \chi_r^{(p)})} \overline{Z}_{ij,N}^{(p)}(\chi_1^{(p)}, ..., \chi_r^{(p)}). \quad (17)$$

Symbole odpowiadające relacji $\chi_1^{(p)*}, ..., \chi_n^{(p)*}$ będą oznaczane: I^*, P_1^*, P_2^*, $\overline{U}_{ij,N}^{(p)*}$, $\overline{V}_{ij,N}^{(p)*}$, $\overline{Z}_{ij,N}^{(p)*}$, $\overline{W}_N^{(p)*}$, natomiast odpowiadające dowolnej innej relacji $\tilde{\chi}_1^{(p)}, ..., \tilde{\chi}_{\tilde{n}}^{(p)}$ będą oznaczane: $\tilde{I}$, $\tilde{P}_1$, $\tilde{P}_2$, $\tilde{\overline{U}}_{ij,N}^{(p)}$, $\tilde{\overline{V}}_{ij,N}^{(p)}$, $\tilde{\overline{Z}}_{ij,N}^{(p)}$, $\tilde{\overline{W}}_N^{(p)}$.

Podstawę do konstrukcji estymatora relacji preferencji stanowi następujące

Twierdzenie 1.

Zmienne losowe $\overline{W}_N^{(p)*}$ oraz $\tilde{\overline{W}}_N^{(p)}$ spełniają nierówności:

$$E(W_N^{(p)*} - \tilde{W}_N^{(p)}) < 0, \quad (18)$$

$$P(W_N^{(p)*} < \tilde{W}_N^{(p)}) \geq 1 - \exp\{-2N(\tfrac{1}{2} - \delta)^2\}. \quad (19)$$

Dowód zob. Klukowski (1994, pkt 5).

Nierówność (19) otrzymuje się na podstawie nierówności Hoeffdinga (zob. Hoeffding 1963). Twierdzenie 1 orzeka, że: wartość oczekiwana zmiennej losowej $\overline{W}_N^{(p)*}$, odpowiadającej relacji $\chi_1^{(p)*}, ..., \chi_n^{(p)*}$, jest mniejsza niż wartość oczekiwana zmiennej $\tilde{\overline{W}}_N^{(p)}$, odpowiadającej dowolnej innej relacji $\tilde{\chi}_1^{(p)}, ..., \tilde{\chi}_{\tilde{n}}^{(p)}$, a ponadto prawdopodobieństwo zdarzenia $\{\overline{W}_N^{(p)*} < \tilde{\overline{W}}_N^{(p)}\}$ dąży wykładniczo do jedności, gdy $N \to \infty$. Zasadne jest zatem przyjęcie estymatora $\hat{\chi}_1^{(p)}, ..., \hat{\chi}_{\hat{n}}^{(p)}$ w postaci minimalizującej wartość zmiennej losowej $\overline{W}_N^{(p)}(\chi_1^{(p)}, ..., \chi_r^{(p)})$, dla danych porównań $g_{1k}(x_i, x_j)$ (k=1, ..., N; $<i, j> \in R_m$). Ocenę relacji otrzymuje się w wyniku rozwiązania zadania optymalizacji dyskretnej o postaci:

$$\min_{F_X^{(p)}} \left\{ \sum_{<i,j>\in R_m} \sum_{k=1}^{N} \left| t_1^{(\iota)}(x_i, x_j) - g_{1k}(x_i, x_j) \right| \right\}, \quad (20)$$

gdzie: $F_X^{(p)}$ - zbiór rozwiązań dopuszczalnych (wszystkich relacji porządku w zbiorze **X**),

$t_1^{(\iota)}(x_i, x_j)$ – funkcja określająca ι-ty element zbioru $F_X^{(p)}$.

Liczba rozwiązań zadania (20) może przekraczać jeden; postać jednoznaczną można wówczas wybrać losowo lub wprowadzając dodatkowe kryterium, np. wybierając relację minimalizującą wyrażenie

$$\sum_{<i,j>\in P_1(\hat{\chi}_1^{(p)}, ..., \hat{\chi}_{\hat{n}}^{(p)}) \cup P_2(\hat{\chi}_1^{(p)}, ..., \hat{\chi}_{\hat{n}}^{(p)})} \sum_{k=1}^{N} \left| t_1(x_i, x_j) - g_{1k}(x_i, x_j) \right|.$$

3.2. Estymacja relacji porządku dla porównań oceniających różnice rang

W przypadku porównań wskazujących różnicę rang (miejsc w uporządkowaniu) funkcja określająca postać relacji wyraża się zależnością $T_2 : \mathbf{X} \times \mathbf{X} \rightarrow D$, $D=\{-(n-1), ..., 0, ..., n-1\}$, przy czym:

$$T_2(x_i, x_j) = d_{ij} \iff x_i \in \chi_r^{(p)*}, \quad x_j \in \chi_s^{(p)*}, \quad d_{ij} = r - s. \tag{21}$$

Wyniki porównań $g_{2k}(x_i, x_j)$ mają postać analogiczną, tj. $g_{2k}: \mathbf{X} \times \mathbf{X} \rightarrow D$, $D=\{-(m-1), ..., m-1\}$; wynik $g_{2k}(x_i, x_j)=c_{ijk}$ stanowi ocenę różnicy rang $T_2(x_i, x_j)$, w k-tym porównaniu, zakłóconą błędem losowym. Zbiór D zawiera liczby całkowite z zakresu: $-(m-1), ..., m-1$, ponieważ nie zakłada się znajomości liczby podzbiorów n.

Przyjmuje się następujące założenia nt. rozkładów błędów porównań:

- niezależność w k-tym i l-tym porównaniu, $k \neq l$ (spełnienie warunku (1));
- jednomodalność funkcji prawdopodobieństwa błędów porównań $T_2(x_i, x_j)$-$g_{2k}(x_i, x_j)$, z modą i medianą w zerze, tzn.:

$$\sum_{l \le 0} P(T_2(x_i, x_j) - g_{2k}(x_i, x_j) = l\,) > 1/2, \tag{22}$$

$$\sum_{l \ge 0} P(T_2(x_i, x_j) - g_{2k}(x_i, x_j) = l\,) > 1/2, \tag{23}$$

$$P(T_2(x_i, x_j) - g_{2k}(x_i, x_j) = l) \ge P(T_2(x_i, x_j) - g_{2k}(x_i, x_j) = l+1); \qquad l \ge 0, \tag{24}$$

$$P(T_2(x_i, x_j) - g_{2k}(x_i, x_j) = l) \ge P(T(x_i, x_j) - g_{2k}(x_i, x_j) = l-1); \qquad l \le 0, \tag{25}$$

przy czym funkcje prawdopodobieństwa błędów porównań spełniają warunki:

$$\alpha_{ijk}^{(p)}(l) = P(T_2(x_i, x_j) - g_{2k}(x_i, x_j) = l;\ T_2(x_i, x_j) = 0) \quad (-(m-1) \le l \le (m-1)), \tag{26}$$

$$\beta_{ijk}^{(p)}(l) = P(T_2(x_i, x_j) - g_{2k}(x_i, x_j) = l;\ T_2(x_i, x_j) < 0) \quad (-2(m-1) \le l \le 2(m-1)), \tag{27}$$

$$\gamma_{ijk}^{(p)}(l) = P(T_2(x_i, x_j) - g_{2k}(x_i, x_j) = l;\ T_2(x_i, x_j) > 0) \quad (-2(m-1) \le l \le 2(m-1)), \tag{28}$$

$$\sum_{l=-(m-1)}^{m-1} \alpha_{ijk}^{(p)}(l) = 1, \qquad \sum_{l=-2(m-1)}^{2(m-1)} \beta_{ijk}^{(p)}(l) = 1, \qquad \sum_{l=-2(m-1)}^{2(m-1)} \gamma_{ijk}^{(p)}(l) = 1.$$

Nie postulujemy zerowej wartości oczekiwanej błędów porównań. Dla uproszczenia zakładamy, że rozkłady błędów $T_2(x_i, x_j)$-$g_{2k}(x_i, x_j)$ ($(x_i, x_j) \in \mathbf{X} \times \mathbf{X}$) są identyczne dla wszystkich k, $1 \le k \le N$ (porównania $g_{21}(x_i, x_j)$, …, $g_{2N}(x_i, x_j)$ są niezależnymi zmiennymi losowymi o identycznych rozkładach).

W celu określenia postaci i własności proponowanego estymatora, niezbędne są definicje następujących zmiennych losowych:

$$U_{kij}^{(p)}(\chi_1^{(p)}, ..., \chi_r^{(p)}) = |\, g_{2k}(x_i, x_j) \,|; \qquad t_2(x_i, x_j) = 0, \tag{29}$$

$$V_{kij}^{(p)}(\chi_1^{(p)}, ..., \chi_r^{(p)}) = |\, t_2(x_i, x_j)\text{-}g_{2k}(x_i, x_j)\,| ; \qquad t_2(x_i, x_j)<0, \tag{30}$$

$$Z_{kij}^{(p)}(\chi_1^{(p)}, ..., \chi_r^{(p)}) = |\, t_2(x_i, x_j)\text{-}g_{2k}(x_i, x_j)\,| ; \qquad t_2(x_i, x_j)>0, \tag{31}$$

$$W_k^{(p)}(\cdot) = \sum_{<i,j>\in I(\cdot)} U_{kij}^{(p)}(\cdot) + \sum_{<i,j>\in P_1(\cdot)} V_{kij}^{(p)}(\cdot) + \sum_{<i,j>\in P_2(\cdot)} Z_{kij}^{(p)}(\cdot)\,, \tag{32}$$

$$\overline{U}_{ij,N}^{(p)}(\cdot) = \tfrac{1}{N}\sum_{k=1}^{N} U_{kij}^{(p)}(\cdot)\,, \tag{33}$$

$$\overline{V}_{ij,N}^{(p)}(\cdot) = \tfrac{1}{N}\sum_{k=1}^{N} V_{kij}^{(p)}(\cdot)\,, \tag{34}$$

$$\overline{Z}_{ij,N}^{(p)}(\cdot) = \tfrac{1}{N}\sum_{k=1}^{N} Z_{kij}^{(p)}(\cdot)\,, \tag{35}$$

$$\overline{W}_N^{(p)}(\cdot) = \sum_{<i,j>\in I(\cdot)} \overline{U}_{ij,N}^{(p)}(\cdot) + \sum_{<i,j>\in P_1(\cdot)} \overline{V}_{ij,N}^{(p)}(\cdot) + \sum_{<i,j>\in P_2(\cdot)} \overline{Z}_{ij,N}^{(p)}(\cdot) \tag{36}$$

(I, P_1, P_2 - zbiory indeksów zdefiniowane zależnościami (10) – (12)).

W rozważanym w tym punkcie problemie również zastosowana będzie konwencja oznaczeń odnosząca się do relacji $\chi_1^{(p)*}, ..., \chi_n^{(p)*}$ oraz $\tilde{\chi}_1^{(p)}, ..., \tilde{\chi}_{\tilde{n}}^{(p)}$.

Podstawę do konstrukcji estymatora relacji preferencji stanowi następujące

<u>Twierdzenie 2</u>.

Zmienne losowe $\overline{W}_N^{(p)}$ oraz $\tilde{\overline{W}}_N^{(p)}$ spełniają nierówności:

$$E(\overline{W}_N^{(p)}) < E(\tilde{\overline{W}}_N^{(p)})\,, \tag{37}$$

$$P(\overline{W}_N^{(p)} < \tilde{\overline{W}}_N^{(p)}) \geq 1-$$

$$\exp\left\{ -2N \frac{\left(\sum_{T_2(x_i,x_j)\neq t(x_i,x_j)} \sum_{k=1}^{N} E\left(\left| T_2(x_i,x_j) - g_{2k}(x_i,x_j) \right| - \left| \tilde{t}_2(x_i,x_j) - g_{2k}(x_i,x_j) \right| \right) \right)^2}{(2\vartheta(m-1))^2} \right\}$$

$$. \tag{38}$$

gdzie: ϑ -liczebność $\{T_2(x_i,x_j) \neq \tilde{t}_2(x_i,x_j)\}$.

Dowód: Klukowski (2007a, pkt 4).

Interpretacja nierówności (37)–(38) jest analogiczna do interpretacji nierówności z Twierdzenia 1. Zadanie optymalizacyjne, pozwalające uzyskać ocenę $\hat{\chi}_1^{(p)}, ..., \hat{\chi}_{\hat{n}}^{(p)}$, ma postać:

$$\min_{F_X^{(p)}} \left\{ \sum_{<i,j>\in R_m(\cdot)} \sum_{k=1}^{N} \left| t_2(x_i,x_j) - g_{2k}(x_i,x_j) \right| \right\}. \tag{39}$$

Rozwiązanie zadania (39) również może być wielokrotne.

Prawa strona nierówności (38) odpowiada określonej postaci relacji $\tilde{\chi}_1^{(p)}, ..., \tilde{\chi}_{\tilde{n}}^{(p)}$. Wyznaczenie jej wartości wymaga znajomości rozkładów błędów porównań $T_2(\cdot)$-$g_{2k}(\cdot)$; w przypadku nieznajomości - można je estymować lub oszacować. Estymacja rozkładów wymaga odpowiedniej liczności porównań N. Oszacowania prawej strony (38) można otrzymać przyjmując dodatkowe założenia, np.:

$$\sum_{l<0} P(T_2(x_i,x_j)-g_2(x_i,x_j)=l)=\sum_{l>0} P(T_2(x_i,x_j)-g_2(x_i,x_j)=l),$$

$$T_2(x_i,x_j)-g_2(x_i,x_j)=l)=T_2(x_i,x_j)-g_2(x_i,x_j)=l+1) \qquad (l\geq 1),$$

$$T_2(x_i,x_j)-g_2(x_i,x_j)=l)=T_2(x_i,x_j)-g_2(x_i,x_j)=l-1) \qquad (l\leq -1).$$

4. Estymacja relacji tolerancji

Relacja tolerancji $\boldsymbol{R}^{(t)}$ w zbiorze $\mathbf{X}=\{x_1, ..., x_m\}$ ($m\geq 3$) ma postać rodziny podzbiorów $\chi_1^{(t)*},...,\chi_n^{(t)*}$ ($1<n<m$), spełniającej warunki:

$$\bigcup_{q=1}^{n}\chi_q^{(t)*}=\mathbf{X}, \quad \chi_q^{(t)*}\neq\{\mathbf{0}\}, \quad \exists\, q, s\ (q\neq s):\ \chi_q^{(t)*}\cap\chi_s^{(t)*}\neq\{\mathbf{0}\}. \tag{40}$$

Rodzina $\chi_1^{(t)*},...,\chi_n^{(t)*}$ zawiera co najmniej dwa podzbiory o niepustej koniunkcji.

4.1. Binarne porównania parami

W celu zapewnienia jednoznaczności opisu relacji tolerancji przy użyciu funkcji binarnej, zakłada się, że każdy podzbiór $\chi_q^{(t)*}\subset\mathbf{X}$ zawiera element x_i, należący wyłącznie do niego, tj.: $x_i\in\chi_q^{(t)*}$ oraz $x_i\notin\chi_s^{(t)*}$, $s\neq q$. W przypadku porównań binarnych funkcja charakteryzująca relację tolerancji ma postać $T_3:\mathbf{X}\times\mathbf{X}\to D$, $D=\{0, 1\}$, przy czym:

$$T_3(x_i,x_j)=\begin{cases}0 & \text{jesli istnieją } q, s \text{ (nie wyklucza się } q=s\text{) takie,}\\ & \text{że } (x_i,x_j)\in\chi_q^{(t)*}\cap\chi_s^{(t)*};\\ 1 & \text{w przeciwnym przypadku.}\end{cases} \tag{41}$$

Porównania parami mają postać $g_{3k}:\mathbf{X}\times\mathbf{X}\to D$, $D=\{0, 1\}$, przy czym:

$$g_{3k}(x_i,x_j)=\begin{cases}0 & \text{jesli wynik } k-\text{tego porównania wskazuje, że para } (x_i,x_j)\\ & \text{znajduje się w koniunkcji } \chi_q^{(t)}\cap\chi_s^{(t)} \text{ (nie wyklucza się } q=s\text{);}\\ 1 & \text{w przeciwnym przypadku.}\end{cases} \tag{42}$$

Zakłada się, że błędy $T_3(x_i,x_j)-g_{3k}(x_i,x_j)$ są stochastycznie niezależne (warunek (1)), a ich funkcje prawdopodobieństwa spełniają warunki:

$$P(T_3(x_i,x_j)\neq g_{3k}(x_i,x_j))\leq\delta^{(k)}, \quad \delta^{(k)}\leq\delta, \qquad \delta\in(0, ½). \tag{43}$$

Funkcja $t_3(x_i, x_j)$, charakteryzująca dowolną postać relacji $\chi_1^{(t)}, ..., \chi_r^{(t)}$, wyraża się zależnością $t_3 : \mathbf{X} \times \mathbf{X} \to D$, D={0, 1}, przy czym:

$$t_3(x_i, x_j) = \begin{cases} 0 & \text{jesli istnieją } q, s \text{ (nie wyklucza się } q = s\text{) spelniajace warunek } (x_i, x_j) \in \chi_q^{(t)} \cap \chi_s^{(t)}; \\ 1 & \text{w przeciwnym przypadku.} \end{cases} \quad (44)$$

Podstawą do konstrukcji i określenia własności estymatora rozważanej relacji są zmienne losowe $U_{ijk}^{(t)}(\chi_1^{(t)}, ..., \chi_r^{(t)})$ oraz $V_{ijk}^{(t)}(\chi_1^{(t)}, ..., \chi_r^{(t)})$ o postaci:

$$U_{ijk}^{(t)}(\chi_1^{(t)}, ..., \chi_r^{(t)}) = \left| t_3(x_i, x_j) - g_{3k}(x_i, x_j) \right|; \quad <i, j> \in I^{(t)}(\chi_1^{(t)}, ..., \chi_r^{(t)}), \quad (45)$$

$$V_{ijk}^{(t)}(\chi_1^{(t)}, ..., \chi_r^{(t)}) = \left| t_3(x_i, x_j) - g_{3k}(x_i, x_j) \right|; \quad <i, j> \in J^{(t)}(\chi_1^{(t)}, ..., \chi_r^{(t)}), \quad (46)$$

gdzie:

$$I^{(t)}(\chi_1^{(t)}, ..., \chi_r^{(t)}) = \{<i, j> \mid \exists q, s \ \ (x_i, x_j) \in \chi_q^{(t)} \cap \chi_s^{(t)} \ , j>i \text{ nie wyklucza się } q=s\}, \quad (47)$$

$$J^{(t)}(\chi_1^{(t)}, ..., \chi_r^{(t)}) = R_m - I^{(t)}(\chi_1^{(t)}, ..., \chi_r^{(t)}). \quad (48)$$

Symbole odpowiadające relacji $\chi_1^{(t)*}, ..., \chi_n^{(t)*}$ będą oznaczane: $U_{ijk}^{(t)*}$, $V_{ijk}^{(t)*}$, $I^{(t)*}$, $J^{(t)*}$, a odpowiadające dowolnej innej relacji $\tilde{\chi}_1^{(t)}, ..., \tilde{\chi}_r^{(t)}$ – będą oznaczane $\tilde{U}_{ijk}^{(t)}$, $\tilde{V}_{ijk}^{(t)}$, $\tilde{I}^{(t)}$, $\tilde{J}^{(t)}$.

Podobnie jak w przypadku relacji porządku definiuje się zmienne: $W_k^{(t)}(\chi_1^{(t)}, ..., \chi_r^{(t)})$, $W_k^{(t)*}$, $\tilde{W}_k^{(t)}$ i $\overline{W}_N^{(t)}(\chi_1^{(t)}, ..., \chi_r^{(t)})$, $\overline{W}_N^{(t)*}$, $\tilde{\overline{W}}_N^{(t)}$ o postaci:

$$W_k^{(t)}(\chi_1^{(t)}, ..., \chi_r^{(t)}) = \sum_{<i,j> \in I^{(t)}(\chi_1^{(t)}, ..., \chi_r^{(t)})} U_{ijk}^{(t)}(\chi_1^{(t)}, ..., \chi_r^{(t)}) + \sum_{<i,j> \in J^{(t)}(\chi_1^{(t)}, ..., \chi_r^{(t)})} V_{ijk}^{(t)}(\chi_1^{(t)}, ..., \chi_r^{(t)}), \quad (49)$$

$$W_k^{(t)*} = \sum_{<i,j> \in I^{(t)*}} U_{ijk}^{(t)*} + \sum_{<i,j> \in J^{(t)*}} V_{ijk}^{(t)*}, \quad (50)$$

$$\tilde{W}_k^{(t)} = \sum_{<i,j> \in \tilde{I}^{(t)}} \tilde{U}_{ijk}^{(t)} + \sum_{<i,j> \in \tilde{J}^{(t)}} \tilde{V}_{ijk}^{(t)}, \quad (51)$$

$$\overline{W}_N^{(t)}(\chi_1^{(t)}, ..., \chi_r^{(t)}) = \frac{1}{N} \sum_{k=1}^{N} W_k^{(t)}(\chi_1^{(t)}, ..., \chi_r^{(t)}), \quad (52)$$

$$\overline{W}_N^{(t)*} = \frac{1}{N} \sum_{k=1}^{N} W_k^{(t)*}, \quad (53)$$

$$\tilde{\overline{W}}_N^{(t)} = \frac{1}{N} \sum_{k=1}^{N} \tilde{W}_k^{(t)}. \quad (54)$$

Własności zmiennych losowych $\overline{W}_N^{(t)*}$ oraz $\tilde{\overline{W}}_N^{(t)}$ określa poniższe

Twierdzenie 3.

Zmienne losowe $\overline{W}_N^{(t)*}$ oraz $\widetilde{\overline{W}}_N^{(t)}$ spełniają nierówności:

$$E(\overline{W}_N^{(t)*} - \widetilde{\overline{W}}_N^{(t)}) < 0, \tag{55}$$

$$P(\overline{W}_N^{(t)*} < \widetilde{\overline{W}}_N^{(t)}) \geq 1\text{-}\exp\{\text{-}2N(\tfrac{1}{2}-\delta)^2\}. \tag{56}$$

Dowód analogiczny do dowodu Twierdzenia 1, z punktu 3.

Interpretacja nierówności (55), (56) jest analogiczna do przypadku relacji porządku. Zadanie optymalizacyjne będące podstawą do wyznaczenia estymatora ma postać:

$$\min_{F_X^{(t)}} \sum_{<i,j>\in R_m} \sum_{k=1}^{N} \left| t_3^{(t)}(x_i, x_j) - g_{3k}(x_i, x_j) \right|, \tag{57}$$

gdzie: $F_X^{(t)}$ - zbiór rozwiązań dopuszczalnych (relacji tolerancji w zbiorze **X**),

$t_3^{(t)}(x_i, x_j)$ - funkcja charakteryzująca ι-ty element zbioru $F_X^{(t)}$.

Rozwiązanie zadania (57) również może być wielokrotne.

4.2. Porównania określające liczbę wspólnych cech pary elementów

Podany poniżej estymator jest przeznaczony dla przypadku, gdy każde z porównań określa liczebność koniunkcji podzbiorów, do której należy para elementów. Przynależność elementu do podzbioru można traktować jako posiadanie określonej cechy – przynależność do koniunkcji podzbiorów jest równoznaczna z posiadaniem określonego zestawu wspólnych cech przez parę elementów.

Relację $\boldsymbol{R}^{(t)}$ można scharakteryzować, w obecnym przypadku, przy użyciu funkcji $T_4 : \mathbf{X} \times \mathbf{X} \rightarrow D$, $D=\{0, 1, ..., n\}$, zdefiniowanej w następujący sposób:

$$T_4(x_i, x_j) = \#(\Omega_i^* \cap \Omega_j^*), \tag{58}$$

gdzie: Ω_i^* - zbiór o postaci $\Omega_i^* = \{s \mid x_i \in \chi_s^*\}$,

$\#(\Xi)$ – liczba elementów zbioru Ξ.

Alternatywnie relację $\boldsymbol{R}^{(t)}$ można scharakteryzować, przy wykorzystaniu funkcji $T_5 : \mathbf{X} \times \mathbf{X} \rightarrow D$, o postaci:

$$T_5(x_i, x_j) = \#(\Psi_i^* \cap \Psi_j^*), \tag{59}$$

gdzie: Ψ_i^* - zbiór zdefiniowany zależnością $\Psi_i^* = \{1, ..., n\}\text{-}\,\Omega_i^*$.

Funkcja $T_4(x_i, x_j)$ określa liczbę wspólnych cech (podobieństwo) pary elementów, a funkcja $T_5(x_i, x_j)$ liczbę cech nie posiadanych przez parę (x_i, x_j), wśród cech występujących w zbiorze **X** (obie funkcje łącznie pozwalają określić niepodobieństwo). Każda z tych funkcji może być podstawą do estymacji relacji tolerancji, jak również obie łącznie.

Porównania parami odpowiadające funkcjom $T_4(\cdot)$ oraz $T_5(\cdot)$ mają postać:

$$g_{4k}(x_i, x_j) = d_{4ijk}, \qquad d_{4ijk} \in D, \tag{60}$$

$$g_{5k}(x_i, x_j) = d_{5ijk}, \qquad d_{5ijk} \in D, \tag{61}$$

gdzie: $d_{\ell ijk}$ ($\ell=4, 5$) jest oceną wartości funkcji $T_\ell(x_i, x_j)$, zakłóconą błędem losowym, otrzymaną w k-tym porównaniu.

Funkcje prawdopodobieństwa błędów porównań można zapisać w postaci:

$$P(T_\ell(x_i, x_j) - g_{\ell k}(x_i, x_j) = l) = \alpha_{\ell ijk}(l) \quad ((x_i, x_i) \in \mathbf{X} \times \mathbf{X};\ \ \ell=4, 5; -n \leq l \leq n). \tag{62}$$

Przyjmuje się następujące założenia o błędach porównań:
- niezależność (tj. spełnienie warunku (1));
- jednomodalność funkcji prawdopodobieństwa (moda i mediana w zerze):

$$\sum_{l \leq 0} \alpha_{\ell ijk}(l) > \tfrac{1}{2}, \qquad \sum_{l \geq 0} \alpha_{\ell ijk}(l) > \tfrac{1}{2}, \tag{63}$$

$$\left.\begin{array}{ll} \alpha_{\ell ijk}(l) \geq \alpha_{\ell ijk}(l+1), & l \geq 0, \\ \alpha_{\ell ijk}(l) \geq \alpha_{\ell ijk}(l-1), & l < 0; \end{array}\right\} \tag{64}$$

- nieskorelowanie obu rodzajów porównań, tj. $g_{4k}(x_i, x_j)$ oraz $g_{5v}(x_i, x_j)$, tzn. zerowa kowariancja $Cov(g_{4k}(x_i, x_j), g_{5v}(x_r, x_s)) = 0$ $((x_i, x_j), (x_r, x_s) \in \mathbf{X} \times \mathbf{X})$ (istnienie korelacji oznacza, że zawartość informacyjna obu porównań jest podobna).

W celu uproszczenia rozważań przyjmuje się, że rozkłady błędów porównań danej pary (x_i, x_j) są identyczne dla wszystkich k ($1 \leq k \leq N$).

Zdefiniujmy dla dowolnej relacji tolerancji $\chi_1^{(t)}, ..., \chi_r^{(t)}$ w zbiorze $\mathbf{X}$, funkcje: $t_4(\cdot)$, $t_5(\cdot)$ – analogiczne do funkcji $T_4(\cdot)$, $T_5(\cdot)$:

$$t_4(x_i, x_j) = \#(\Omega_i \cap \Omega_j), \tag{65}$$

$$t_5(x_i, x_j) = \#(\Psi_i \cap \Psi_j), \tag{66}$$

gdzie: $\Omega_i = \{s \mid x_i \in \chi_s^{(t)}\}$ oraz $\Psi_i = \{1, ..., r\} - \Omega_i$. (67)

Własności proponowanych estymatorów opierają się na własnościach zmiennych losowych: $U_{\ell ijk}^{(t)}(\chi_1^{(t)}, ..., \chi_r^{(t)})$, $W_{\ell k}^{(t)}(\chi_1^{(t)}, ..., \chi_r^{(t)})$, $\overline{U}_{\ell ij,N}^{(t)}(\chi_1^{(t)}, ..., \chi_r^{(t)})$, $\overline{W}_{\ell,N}^{(t)}(\chi_1^{(t)}, ..., \chi_r^{(t)})$ ($\ell=4, 5$) zdefiniowanych w następujący sposób:

$$U_{\ell ijk}^{(t)}(\chi_1^{(t)}, ..., \chi_r^{(t)}) = \left| t_\ell(x_i, x_j) - g_{\ell k}(x_i, x_j) \right|, \tag{68}$$

$$W_{\ell k}^{(t)}(\chi_1^{(t)}, ..., \chi_r^{(t)}) = \sum_{<i,j> \in R_m} U_{\ell ijk}^{(t)}(\chi_1^{(t)}, ..., \chi_r^{(t)}), \tag{69}$$

$$\overline{U}_{\ell ij,N}^{(t)}(\chi_1^{(t)}, ..., \chi_r^{(t)}) = \frac{1}{N} \sum_{k=1}^{N} U_{\ell ijk}^{(t)}(\chi_1^{(t)}, ..., \chi_r^{(t)}), \tag{70}$$

$$\overline{W}^{(t)}_{\ell,N}(\chi_1^{(t)},...,\chi_r^{(t)})=\frac{1}{N}\sum_{k=1}^{N} W^{(t)}_{\ell k}(\chi_1^{(t)},...,\chi_r^{(t)}). \tag{71}$$

W celu uproszczenia zapisu będzie stosowana konwencja z poprzedniego punktu, tj. symbole odpowiadające relacji $\chi_1^{(t)*},...,\chi_n^{(t)*}$ będą oznaczane: $U^{(t)*}_{\ell ijk}$, $\overline{U}^{(t)*}_{\ell ij,N}$, $\overline{W}^{(t)*}_{\ell,N}$, a odpowiadające dowolnej innej relacji $\tilde{\chi}_1^{(t)},...,\tilde{\chi}_r^{(t)}$ - będą oznaczane: $\tilde{U}^{(t)}_{\ell ijk}$, $\tilde{\overline{U}}^{(t)}_{\ell ij,N}$, $\tilde{\overline{W}}^{(t)}_{\ell,N}$.

Własności zmiennych losowych $\overline{W}^{(t)*}_{\ell,N}$ oraz $\tilde{\overline{W}}^{(t)}_{\ell,N}$ są analogiczne do przypadku relacji porządku; określa je

Twierdzenie 4.

Zmienne losowe $\overline{W}^{(t)*}_{\ell,N}$ oraz $\tilde{\overline{W}}^{(t)}_{\ell,N}$ spełniają nierówności:

$$E(\overline{W}^{(t)*}_{\ell,N}) < E(\tilde{\overline{W}}^{(t)}_{\ell,N}), \tag{72}$$

$$P(\overline{W}^{*}_{\ell,N} < \tilde{\overline{W}}_{\ell,N}) \geq 1 -$$

$$\exp\left\{\frac{-2N(\sum_{T_\ell(x_i,x_j)\neq\tilde{t}_\ell(x_i,x_j)}\sum_{k=1}^{N}E(\left|T_\ell(x_i,x_j)-g_{\ell k}(x_i,x_j)\right|-\left|\tilde{t}_\ell(x_i,x_j)-g_{\ell k}(x_i,x_j)\right|))^2}{(2\vartheta(m-1))^2}\right\}, \tag{73}$$

gdzie: ϑ - liczność zbioru $\{(x_i,x_j)\ \left|\ T_\ell(x_i,x_j)\neq\tilde{t}_\ell(x_i,x_j)\right.\}$.

Dowód zob. Klukowski (2007b, pkt 3).

Interpretacja nierówności (72), (73) jest analogiczna do nierówności z twierdzeń 1 – 3. Zadania optymalizacyjne do wyznaczenia oceny relacji mają postać:

$$\min_{F_X^{(t)}}[\sum_{k=1}^{N}\sum_{<i,j>\in R_m}\left|t_\ell^{(t)}(x_i,x_j)-g_{\ell k}^{(t)}(x_i,x_j)\right|] \qquad (\ell=4 \text{ lub } 5), \tag{74}$$

$$\min_{F_X^{(t)}}[\sum_{k=1}^{N}\sum_{<i,j>\in R_m}(\left|t_4^{(t)}(x_i,x_j)-g_{4k}^{(t)}(x_i,x_j)\right|+\left|t_5^{(t)}(x_i,x_j)-g_{5k}^{(t)}(x_i,x_j)\right|)], \tag{75}$$

gdzie:

$F_X^{(t)}$ – zbiór rozwiązań dopuszczalnych (wszystkich relacji tolerancji w zbiorze **X**);

$t_\ell^{(t)}(\cdot)$ - funkcja charakteryzująca ι -ty element zbioru $F_X^{(t)}$.

Oba zadania mogą mieć rozwiązania wielokrotne. Zadanie (75) znajduje zastosowanie w przypadku dysponowania obydwoma rodzajami porównań. Zapewnia ono lepszą precyzję estymacji, w stosunku do zadania (74), ponieważ opiera się na próbie o podwojonej wielkości, tj. 2*N*.

5. Estymacja relacji równoważności

Relacja równoważności $\boldsymbol{R}^{(e)}$ rozbija zbiór $\mathbf{X} = \{x_1, ..., x_m\}$ ($3 \le m < \infty$) na podzbiory $\chi_1^{(e)*}, ..., \chi_n^{(e)*}$ $(n \ge 2)$ z koniunkcjami będącymi zbiorami pustymi, tj.:

$$\mathbf{X} = \bigcup_{i=1}^{n} \chi_i^{(e)*}, \ \chi_r^{(e)*} \cap \chi_s^{(e)*} = \{\mathbf{0}\}; \qquad r \neq s, \tag{76}$$

gdzie: $\{\mathbf{0}\}$ - zbiór pusty.

Relację równoważności można scharakteryzować za pomocą funkcji $T_6 : \mathbf{X} \times \mathbf{X} \to D$, $D=\{0, 1\}$, o postaci:

$$T_6(x_i, x_j) = \begin{cases} 0 & \text{jesli istnieje } q \text{ takie, że } (x_i, x_j) \in \chi_q^{(e)*}; \\ 1 & \text{w przeciwnym przypadku.} \end{cases} \tag{77}$$

Porównania parami mają, w rozważanym obecnie przypadku, postać funkcji $g_{6k} : \mathbf{X} \times \mathbf{X} \to D$, $D=\{0, 1\}$, zdefiniowanej w następujący sposób:

$$g_{6k}(x_i, x_j) = \begin{cases} 0 & \text{jesli wynik } k-\text{tego porownania wskazuje, że para } (x_i, x_j) \\ & \text{należa do wspólnego podzbioru, tj. } \exists q \text{ takie, że } (x_i, x_j) \in \chi_q^{(e)*}; \\ 1 & \text{w przeciwnym przypadku.} \end{cases} \tag{78}$$

Funkcje prawdopodobieństwa błędów porównań spełniają warunki:

$$P(g_{6k}(x_i, x_j) \neq T_6(x_i, x_j)) \le \delta^{(k)}; \quad \delta^{(k)} \le \delta; \qquad 0 < \delta < ½; \tag{79}$$

zakłada się, że porównania $g_{6k}(\cdot)$ oraz $g_{6l}(\cdot)$ są, w przypadku $k \neq l$, stochastycznie niezależne (warunek (1)).

Dla dowolnej relacji równoważności $\chi_1^{(e)}, ..., \chi_r^{(e)}$ w zbiorze **X**, zdefiniujmy funkcję $t_6(x_i, x_j)$ o postaci analogicznej do funkcji $T_6(x_i, x_j)$:

$$t_6(x_i, x_j) = \begin{cases} 0 & \text{jesli istnieje } q \text{ spelniajace warunek } (x_i, x_j) \in \chi_q^{(e)}; \\ 1 & \text{w przeciwnym przypadku;} \end{cases} \tag{80}$$

oraz zmienne losowe:

$$U_{ijk}^{(e)}(\chi_1, ..., \chi_r) = \left| t_6(x_i, x_j) - g_{6k}(x_i, x_j) \right|; \quad \langle i, j \rangle \in I^{(e)}(\chi_1^{(e)}, ..., \chi_r^{(e)}), \tag{81}$$

$$V_{ijk}^{(e)}(\chi_1, ..., \chi_r) = \left| t_6(x_i, x_j) - g_{6k}(x_i, x_j) \right|; \quad \langle i, j \rangle \in J^{(e)}(\chi_1^{(e)}, ..., \chi_r^{(e)}), \tag{82}$$

gdzie: $I^{(e)}(\chi_1^{(e)}, ..., \chi_r^{(e)})$ oraz $J^{(e)}(\chi_1^{(e)}, ..., \chi_r^{(e)})$ zbiory indeksów o postaci:

$$I^{(e)}(\chi_1^{(e)}, ..., \chi_r^{(e)}) = \{\langle i, j \rangle \mid \exists q \text{ taki, ze } (x_i, x_j) \in \chi_q; j > i\}, \tag{83}$$

$$J^{(e)}(\chi_1^{(e)}, ..., \chi_r^{(e)}) = R_m - I^{(e)}(\chi_1^{(e)}, ..., \chi_r^{(e)}). \tag{84}$$

Zbiór $I^{(e)}(\cdot)$ zawiera indeksy par takich, że oba elementy należą do wspólnego podzbioru, a zbiór $J^{(e)}(\cdot)$ – par należących do różnych podzbiorów.

Symbole odpowiadające relacji $\chi_1^{(e)*}, ..., \chi_n^{(e)*}$ będą oznaczane: $U_{kij}^{(e)*}$, $V_{kij}^{(e)*}$, $I^{(e)*}$, $J^{(e)*}$, $W_k^{(e)*}$ natomiast odpowiadające dowolnej innej relacji $\tilde{\chi}_1^{(e)}, ..., \tilde{\chi}_r^{(e)}$ - będą oznaczane: $\tilde{U}_{kij}^{(e)}$, $\tilde{V}_{kij}^{(e)}$, $W_k^{(e)*}$, $\tilde{I}^{(e)}$, $\tilde{J}^{(e)}$.

Podstawą do konstrukcji estymatora i określenia jego własności są zmienne losowe $W_k^{(e)}(\chi_1^{(e)}, ..., \chi_r^{(e)})$, $W_k^{(e)*}$, $\tilde{W}_k^{(e)}$ oraz $\overline{W}_N^{(e)}(\chi_1^{(e)}, ..., \chi_r^{(e)})$, o postaci:

$$W_k^{(e)}(\chi_1^{(e)}, ..., \chi_r^{(e)}) = \sum_{<i,j> \in I^{(e)}(\chi_1, ..., \chi_r)} U_{kij}^{(e)}(\chi_1, ..., \chi_r) +$$

$$\sum_{<i,j> \in J^{(e)}(\chi_1, ..., \chi_r)} V_{kij}^{(e)}(\chi_1, ..., \chi_r) \tag{85}$$

$$\overline{W}_N^{(e)}(\chi_1^{(e)}, ..., \chi_r^{(e)}) = \tfrac{1}{N} \sum_{k=1}^{N} W_k^{(e)}(\chi_1^{(e)}, ..., \chi_r^{(e)}) . \tag{86}$$

Własności zmiennych losowych $\overline{W}_N^{(e)*}$ oraz $\overline{\tilde{W}}_N^{(e)}$ odpowiadających relacjom – odpowiednio $\chi_1^{(e)*}, ..., \chi_n^{(e)*}$ oraz $\tilde{\chi}_1^{(e)}, ..., \tilde{\chi}_r^{(e)}$ - są określone przez

<u>Twierdzenie 5</u>.
Zmienne losowe $\overline{W}_N^{(e)*}$ oraz $\overline{\tilde{W}}_N^{(e)}$ spełniają nierówności:

$$E(\overline{W}_N^{(e)*}) < E(\overline{\tilde{W}}_N^{(e)}), \tag{87}$$

$$P(\overline{W}_N^{(e)*} < \overline{\tilde{W}}_N^{(e)}) \geq 1 - \exp\left\{-2N(\tfrac{1}{2} - \delta)^2\right\}. \tag{88}$$

Dowód nierówności (87), (88) analogiczny do dowodu nierówności z Twierdzenia 1.

Zadanie optymalizacyjne służące do wyznaczenia oceny $\hat{\chi}_1^{(e)}, ..., \hat{\chi}_n^{(e)}$ ma postać:

$$\min_{F_X^{(e)}} \sum_{<i,j> \in R_m} \sum_{k=1}^{N} \left| t_6^{(\iota)}(x_i, x_j) - g_{6k}(x_i, x_j) \right|, \tag{89}$$

gdzie:

$F_{\mathbf{X}}^{(e)}$ - zbiór rozwiązań dopuszczalnych (wszystkich relacji równoważności w **X**),

$t_6^{(\iota)}(x_i, x_j)$ - funkcja określająca ι-ty element zbioru $F_{\mathbf{X}}^{(e)}$.

Zadanie (89) również może mieć więcej niż jedno rozwiązanie.

6. Podsumowanie

Omówione w pracy estymatory relacji porządku, tolerancji i równoważności, charakteryzują się:
- dobrymi własnościami statystycznymi,
- szerokim zakresem zastosowań – ze względu słabe założenia,
- jednorodnością postaci,
- łatwością stosowania, a także intuicyjną formą.

Wyniki dotychczasowych badań empirycznych i symulacyjnych potwierdzają ich praktyczną przydatność; wskazują ponadto, że jest to pole dla dalszych wartościowych badań, zwłaszcza symulacyjnych.

Literatura

David H. A. (1988) *The Method of Paired Comparisons*. 2nd ed. Ch. Griffin, London.

Hoeffding W. (1963) Probability inequalities for sums of bounded random variables. *J. Amer. Statist. Assoc*., **58**, 13-30.

Klukowski L. (1990) Algorytm klasyfikacji prób w przypadku nieznanej liczby generujących je zmiennych losowych. *Przegląd Statystyczny* **XXXVII,** 167-177.

Klukowski L. (1994) Some probabilistic properties of the nearest adjoining order method and its extensions. *Annals of Operations Research*, **51**, 241-261.

Klukowski L. (2000) The nearest adjoining order method for pairwise comparisons in the form of difference of ranks. *Annals of Operations Research*, **97**, 357-378.

Klukowski L. (2002) Estymacja relacja tolerancji na podstawie porównań parami z błędami losowymi. W: Bubnicki Z., Hryniewicz O., Kulikowski R. red. *Metody i Techniki Analizy Informacji i Wspomagania Decyzji*. EXIT, Warszawa, V-21 – V-35.

Klukowski L. (2007a) Estimation of the Preference Relation on the Basis of Medians from Pairwise Comparisons in the Form of Difference of Ranks. W: Kurzynski M. i in., red., *Computer Recognition Systems* 2, *Advances in Soft Computing* 45. Springer, Berlin Heidelberg New York, 232 – 241.

Klukowski L. (2007b) Estimation of tolerance relation the basis of multiple pairwise comparisons with random errors. *Control and Cybernetics*, **36**, 443 - 466.

Klukowski L. (2007c) Estimation of the preference relation on the basis of multiple pairwise comparisons in the form of difference of ranks. Zgłoszone do *Control and Cybernetics*.

Slater P. (1961) Inconsistencies in a schedule of paired comparisons. *Biometrika*, 48, 303-312.

Olgierd Hryniewicz, Andrzej Straszak, Jan Studziński
red.

BADANIA OPERACYJNE I SYSTEMOWE: ŚRODOWISKO NATURALNE, PRZESTRZEŃ, OPTYMALIZACJA

Książka składa się z artykułów przedstawiających wyniki prac z dziedziny badań operacyjnych i systemowych, poświęconych środowisku naturalnemu i zarządzaniu nim, zwłaszcza w zakresie ochrony atmosfery, globalnego ocieplenia i walki z nim, jakości i zaopatrzenia w wodę. Tematyka ta jest rozszerzona o aspekty przestrzenne, regionalne i samorządowe, a także planowanie i funkcjonowanie infrastruktury. Tom zamykają prace metodyczne, dostarczające technik, będących podstawą prezentowanych zastosowań.

ISBN 83-894-7519-7
EAN 9788389475190

Instytut Badań Systemowych PAN
tel. (4822) 3810241 / 3810273 e-mail: biblioteka@ibspan.waw.pl