39/2007

Raport Badawczy
Research Report

RB/41/2007

Estymacja relacji porządku na podstawie porównań parami z błędami losowymi

L. Klukowski

Instytut Badań Systemowych
Polska Akademia Nauk

Systems Research Institute
Polish Academy of Sciences

POLSKA AKADEMIA NAUK

Instytut Badań Systemowych

ul. Newelska 6

01-447 Warszawa

tel.: (+48) (22) 3810100

fax: (+48) (22) 3810105

Kierownik Pracowni zgłaszający pracę:
prof. dr hab. inż. Zbigniew Nahorski

Warszawa 2007

ESTYMACJA RELACJI PORZĄDKU NA PODSTAWIE PORÓWNAŃ PARAMI Z BŁĘDAMI LOSOWYMI

Leszek Klukowski
IBS PAN, Newelska 6, 01-447 Warszawa, e-mail: Leszek.Klukowski@ibspan.waw.pl

W pracy zaproponowano metodę estymacji relacji porządku w zbiorze skończonym, na podstawie wielokrotnych porównań parami, w postaci różnicy rang, zakłóconych błędami losowymi. Rozważono estymator oparty na koncepcji „najbliższego porządku" (ang. *nearest adjoining order* - NAO), zapoczątkowanej w pracy Slatera (1961); można go zaliczyć również do tzw. algorytmów kombinatorycznych. Estymowaną postać relacji otrzymuje się na podstawie optymalnego rozwiązania odpowiedniego zadania optymalizacji dyskretnej. Postać proponowanego estymatora określono na podstawie własności zdefiniowanej w pracy rodziny zmiennych losowych. Wykazano, że zmienna odpowiadająca rzeczywistej relacji ma najmniejszą wartość oczekiwaną w tej rodzinie oraz określono oszacowanie prawdopodobieństwa zdarzenia, że zmienna, odpowiadająca rzeczywistej postaci relacji, przyjmie wartość mniejszą niż zmienna odpowiadająca dowolnej innej (wyspecyfikowanej) postaci. Prawdopodobieństwo to dąży wykładniczo do jedności, wraz ze wzrostem liczby porównań do nieskończoności, przy słabych założeniach odnośnie rozkładów błędów porównań.

1. Wprowadzenie

Rozważany w niniejszej pracy problem estymacji relacji porządku stanowi kontynuację wcześniejszych prac autora, dotyczących estymacji trzech rodzajów relacji: porządku (zob. Klukowski (1994, 2000)), równoważności (zob. Klukowski (1990)) i tolerancji (zob. (Klukowski 2002, 2006)). Idea proponowanych estymatorów opiera się na koncepcji „najbliższego porządku" (zob. Slater (1961)); w literaturze tematu zalicza się je również do tzw. algorytmów kombinatorycznych (zob. Koronacki, Ćwik (2005), rozdz. 9). Algorytmy estymacji wymienionych relacji opierają się na rozwiązaniach optymalnych pewnych zadań optymalizacji dyskretnej. Możliwe jest analityczne określenie niektórych własności zmiennych losowych stanowiących podstawę estymacji, przy nie ograniczających założeniach odnośnie rozkładów błędów porównań. Błędy te nie muszą mieć zerowej wartości oczekiwanej, a ich rozkłady mogą być nieznane. Zakłada się jedynie, że są to rozkłady jednomodalne, a ich funkcje prawdopodobieństwa przyjmują maksimum i medianę w zerze; porównania wielokrotne każdej pary muszą być niezależnymi zmiennymi losowymi.

Przykładem problemu tego typu jest uporządkowanie klientów wg zarobków, na podstawie kwoty wydanej na zakupy, w przypadku wielokrotnych (niezależnych) zakupów. Przykłady zastosowań statystycznych algorytmów porządkowania podano m.in. w pracy David (1988).

Niniejsza praca składa się z czterech części oraz dodatku. W części drugiej podano niezbędne definicje i oznaczenia oraz sformułowanie problemu. Część trzecia zawiera postać proponowanego estymatora oraz twierdzenie określające jego własności. W części czwartej podsumowano treść pracy oraz wskazano kierunki dalszych badań. Dodatek zawiera ideę dowodu twierdzenia z części trzeciej.

2. Sformułowanie problemu

Poniższe sformułowanie jest rozwinięciem problemu z pracy Klukowskiego (2000) na przypadek wielokrotnych, niezależnych stochastycznie porównań każdej pary.

Zakłada się, że w skończonym zbiorze $\mathbf{X}=\{x_1, \ldots, x_m\}$, $m \geq 3$ istnieje, relacja słabej preferencji **R** o postaci:

$$\mathbf{R} = \mathbf{I} \cup \mathbf{P}, \tag{1}$$

gdzie:

I – relacja równoważności (zwrotna, przechodnia, symetryczna),

P – relacja mocnej preferencji (przechodnia, antysymetryczna).

Relacja **R** generuje z elementów zbioru **X** ciąg podzbiorów $\chi_1^*, \dots, \chi_n^*$ ($n \leq m$) taki, że (dowolny) element $x_i \in \chi_r^*$ jest preferowany w stosunku do elementu $x_j \in \chi_s^*$ ($r < s$) oraz każdy podzbiór χ_q^* ($1 \leq q \leq n$) zawiera elementy równoważne.

Relacja **R** może być scharakteryzowana przy wykorzystaniu funkcji $T : \mathbf{X} \times \mathbf{X} \to D$,
$D = \{-(n-1), \dots, 0, \dots, n-1\}$, zdefiniowanej w następujący sposób:

$$T(x_i, x_j) = d_{ij} \;\Leftrightarrow\; x_i \in \chi_k^*, \;\; x_j \in \chi_l^*, \;\; d_{ij} = k - l. \tag{2}$$

Wartość funkcji $T(x_i, x_j)$ jest różnicą rang (miejsc w uporządkowaniu) elementów x_i oraz x_j w relacji **R**. Wartość $T(x_i, x_j)=0$ oznacza równoważność obu elementów (należą do wspólnego podzbioru χ_q^*, $1 \leq q \leq n$). Jest oczywiste, że $T(x_i, x_j)=-T(x_j, x_i)$ dla $T(\cdot)\neq 0$.

Postać relacji $\chi_1^*, \dots, \chi_n^*$ należy określić (estymować) na podstawie porównań parami elementów zbioru **X**, zakłóconych błędami losowymi. Każda para $(x_i, x_j) \in \mathbf{X} \times \mathbf{X}$ jest porównywana N-krotnie ($N>1$), w sposób niezależny stochastycznie; wynik k-tego porównania ($k=1, \dots, N$) stanowi wartość funkcji:

$$g_k : \mathbf{X} \times \mathbf{X} \to D, \qquad D=\{-(m-1), \dots, 0, \dots, m-1\}. \tag{3}$$

Wynik $g_k(x_i, x_j)=c_{kij}$ jest oceną różnicy rang pary (x_i, x_j), w k-tym porównaniu. Zbiór wartości D zawiera liczby całkowite z zakresu: $-(m-1), \dots, m-1$, ponieważ nie zakłada się znajomości liczby podzbiorów n.

Ze względu na występowanie błędów losowych, każde porównanie $g_k(x_i, x_j)$ ($1 \leq k \leq N$) może przyjmować wartość różną od $T(x_i, x_j)$, co sprawia, że różnica $T(x_i, x_j)-g_k(x_i, x_j)$ może przyjmować z pewnymi prawdopodobieństwami wartości różne od zera. Zakłada się, że porównania $g_k(x_i, x_j)$ oraz $g_l(x_r, x_s)$ są niezależne, tj.:

$$P((g_k(x_i, x_j)=c_{kij})\cap(g_l(x_r, x_s)=c_{lrs}))=P(g_k(x_i, x_j)=c_{kij})P(g_l(x_r, x_s)=c_{lrs}) \quad (k\neq l). \tag{4}$$

Ponadto zakłada się, że funkcje prawdopodobieństwa błędów porównań każdej pary $(x_i, x_j)\in \mathbf{X}\times\mathbf{X}$, oznaczane symbolami $\alpha_{ijk})$, $\beta_{ijk}(l)$, $\gamma_{ijk}(l)$, zdefiniowane zależnościami:

$$\alpha_{ijk}(l)=Pr(T(x_i, x_j)-g_k(x_i, x_j)=l;\ T(x_i, x_j)=0),\ (-(m-1)\leq l\leq(m-1)), \tag{5}$$

$$\beta_{ijk}(l)=Pr(T(x_i, x_j)-g_k(x_i, x_j)=l;\ T(x_i, x_j)< 0),\ \ (-2(m-1)\leq l\leq 2(m-1)), \tag{6}$$

$$\gamma_{ijk}(l)=Pr(T(x_i, x_j)-g_k(x_i, x_j)=l;\ T(x_i, x_j)>0),\ (-2(m-1)\leq l\leq 2(m-1)), \tag{7}$$

spełniają warunki:

$$\sum_{l=-(m-1)}^{m-1} \alpha_{ijk}(l) = 1, \qquad \sum_{l=-2(m-1)}^{2(m-1)} \beta_{ijk}(l) = 1, \qquad \sum_{l=-2(m-1)}^{2(m-1)} \gamma_{ijk}(l) = 1, \tag{8}$$

$$\sum_{l\leq 0} Pr(T(x_i, x_j) - g_k(x_i, x_j) = l\,)>1/2, \tag{9}$$

$$\sum_{l\geq 0} Pr(T(x_i, x_j) - g_k(x_i, x_j) = l\,)>1/2, \tag{10}$$

$$Pr(T(x_i, x_j)-g_k(x_i, x_j)=l)\geq Pr(T(x_i, x_j)-g_k(x_i, x_j)=l+1); \qquad l\geq 0, \tag{11}$$

$$Pr(T(x_i, x_j)-g_k(x_i, x_j)=l)\geq Pr(T(x_i, x_j)-g_k(x_i, x_j)=l-1); \qquad l\leq 0. \tag{12}$$

Nierówności (9) – (12) implikują własności rozkładu błędów porównań każdej pary: • mediana równa zeru, • funkcja prawdopodobieństwa – jednomodalna, przyjmująca wartość maksymalną w zerze. Wartość oczekiwana każdego z błędów może być różna od zera (w szczególności, dla $T(x_i, x_j)=\pm n$ wartość ta jest niezerowa, jeśli $P(T(x_i, x_j)-g_k(x_i, x_j)=0)\neq 1$). Prawdopodobieństwa $\alpha_{ijk}(0)$, $\beta_{ijk}(0)$ oraz $\gamma_{ijk}(0)$ mogą być mniejsze niż ½; jest to osłabienie założenia z pracy Klukowskiego (1994), w której rozważono porównania wskazujące kierunek preferencji.

W celu uproszczenia rozważań, zakłada się, że rozkłady błędów $T(x_i, x_j)$-$g_k(x_i, x_j)$ ($(x_i, x_j)\in\mathbf{X}\times\mathbf{X}$) są identyczne dla wszystkich k, $1\leq k\leq N$; w konsekwencji porównania $g_1(x_i, x_j), \ldots, g_N(x_i, x_j)$ są realizacjami niezależnych zmiennych losowych o identycznych rozkładach. Z tego względu indeks k będzie pomijany w symbolach: $\alpha_{ijk}(l)$, $\beta_{ijk}(l)$, $\gamma_{ijk}(l)$. Osłabienie założenia o identyczności rozkładów nie stanowi problemu.

Stosując powyższe oznaczenia, problem estymacji relacji **R** na podstawie porównań parami można sformułować w następujący sposób. Określić relację **R** (ciąg $\chi_1^*, \ldots, \chi_n^*$ lub – równoważnie - wartości funkcji $T(x_i, x_j)$) na podstawie porównań $g_k(x_i, x_j)$ (k=1, ..., N), dla $(x_i, x_j)\in\mathrm{X}\times\mathrm{X}$.

Należy podkreślić, że estymacja relacji nie wymaga znajomości funkcji prawdopodobieństwa: $\alpha_{ij}(l)$, $\beta_{ij}(l)$, $\gamma_{ij}(l)$) oraz liczby podzbiorów n. Funkcje prawdopodobieństwa lub ich oszacowania są natomiast niezbędne do określenia własności wyników estymacji.

Dla dalszych rozważań niezbędne są poniższe oznaczenia:

- $t(x_i, x_j)$ - funkcja określająca różnice rang każdej pary $(x_i, x_j)\in\mathbf{X}\times\mathbf{X}$ dla dowolnej relacji porządku w zbiorze **X**, tzn.:

$$t(x_i, x_j) = d_{ij} \Leftrightarrow x_i \in \chi_k,\quad x_i \in \chi_l\ ;\quad d_{ij} = k - l; \tag{13}$$

- $I(\chi_1, \ldots, \chi_r)$, $P_1(\chi_1, \ldots, \chi_r)$, $P_2(\chi_1, \ldots, \chi_r)$, R_m – zbiory par indeksów $<i,j>$, ($j>i$) odpowiadające relacji $\chi_1, \ldots, \chi_r$, o następującej postaci:

$$I(\chi_1, \ldots, \chi_r)=\{<i,j> \mid t(x_i, x_j) = 0; j>i\}, \tag{14}$$

$$P_1(\chi_1, \ldots, \chi_r)=\{<i,j> \mid t(x_i, x_j) < 0; j>i\}, \tag{15}$$

$$P_2(\chi_1, \ldots, \chi_r)=\{<i,j> \mid t(x_i, x_j) > 0; j>i\}; \tag{16}$$

$$R_m= I(\chi_1, \ldots, \chi_r)\cup P_1(\chi_1, \ldots, \chi_r)\cup P_2(\chi_1, \ldots, \chi_r) =$$

$$=\{<i,j> \mid 1 \leq i, j \leq m \, ; j > i\}; \tag{17}$$

$$M = m(m-1)/2 = \#(R_m) \tag{18}$$

(symbol $\#(\Xi)$ oznacza liczbę elementów zbioru Ξ).

Podstawę do określenia własności rozważanego poniżej estymatora stanowią zmienne losowe $U_{ij}^{(k)}(\chi_1, \ldots, \chi_r)$, $V_{ij}^{(k)}(\chi_1, \ldots, \chi_r)$, $Z_{ij}^{(k)}(\chi_1, \ldots, \chi_r)$, $W^{(k)}(\chi_1, \ldots, \chi_r)$, zdefiniowane w następujący sposób:

$$U_{ij}^{(k)}(\cdot) = | g_k(x_i, x_j) | ; \qquad t(x_i, x_j) = 0, \tag{19}$$

$$V_{ij}^{(k)}(\cdot) = | t(x_i, x_j) - g_k(x_i, x_j) | ; \; t(x_i, x_j) < 0, \tag{20}$$

$$Z_{ij}^{(k)}(\cdot) = | t(x_i, x_j) - g_k(x_i, x_j) | ; \; t(x_i, x_j) > 0, \tag{21}$$

$$W^{(k)}(\cdot) = \sum_{<i,j> \in I(\cdot)} U_{ij}^{(k)}(\cdot) + \sum_{<i,j> \in P_1(\cdot)} V_{ij}^{(k)}(\cdot) + \sum_{<i,j> \in P_2(\cdot)} Z_{ij}^{(k)}(\cdot). \tag{22}$$

Zmienne losowe i symbole odpowiadające relacji **R** (bezbłędnemu wynikowi estymacji) będą oznaczane symbolami: $U_{ij}^{(k)*}$, $V_{ij}^{(k)*}$, $Z_{ij}^{(k)*}$, I^*, P_1^*, P_2^*, $W^{(k)*}$, natomiast odpowiadające dowolnej relacji $\widetilde{X}_1, \ldots, \widetilde{X}_{\widetilde{r}}$ - różnej od **R** - symbolami: $\widetilde{U}_{ij}^{(k)}, \widetilde{V}_{ij}^{(k)}, \widetilde{Z}_{ij}^{(k)}, \widetilde{I}$, $\widetilde{P}_1, \widetilde{P}_2, \widetilde{W}^{(k)}$. Dla każdego k ($1 \leq k \leq N$) różnica $W^{(k)*} - \widetilde{W}^{(k)}$ może być przedstawiona w postaci:

$$\begin{aligned} W^{(k)*} - \widetilde{W}^{(k)} = & \sum_{I^* \cap (\widetilde{P}_1 - P_1^*)} (U_{ij}^{(k)*} - \widetilde{V}_{ij}^{(k)}) + \sum_{I^* \cap (\widetilde{P}_2 - P_2^*)} (U_{ij}^{(k)*} - \widetilde{Z}_{ij}^{(k)}) + \\ & + \sum_{P_1^* \cap (\widetilde{I} - I^*)} (V_{ij}^{(k)*} - \widetilde{U}_{ij}^{(k)}) + \sum_{P_1^* \cap (\widetilde{P}_2 - P_2^*)} (V_{ij}^{(k)*} - \widetilde{Z}_{ij}^{(k)}) + \\ & + \sum_{(P_1^* \cap \widetilde{P}_1) \cap (T(\cdot) \neq \widetilde{T}(\cdot))} (V_{ij}^{(k)*} - \widetilde{V}_{ij}^{(k)}) + \sum_{P_2^* \cap (\widetilde{I} - I^*)} (Z_{ij}^{(k)*} - \widetilde{U}_{ij}^{(k)}) + \\ & + \sum_{P_2^* \cap (\widetilde{P}_1 - P_1^*)} (Z_{ij}^{(k)*} - \widetilde{V}_{ij}^{(k)}) + \sum_{(P_2^* \cap \widetilde{P}_2) \cap (T(\cdot) \neq \widetilde{T}(\cdot))} (Z_{ij}^{(k)*} - \widetilde{Z}_{ij}^{(k)}) \end{aligned} \tag{23}$$

lub – krócej – w postaci:

$$W^{(k)*} - \widetilde{W}^{(k)} = \sum_{\nu=1}^{8} \sum_{S\nu} Q_{ij}^{(k\nu)},$$

gdzie:

$$Q_{ij}^{(k,1)} = U_{ij}^{(k)*} - \widetilde{V}_{ij}^{(k)}, \quad S_1 = \{<i,j> \mid <i,j> \in I^* \cap (\widetilde{P}_1 - P_1^*)\}, \tag{24}$$

$$Q_{ij}^{(k,2)} = U_{ij}^{(k)*} - \widetilde{Z}_{ij}^{(k)}, \quad S_2 = \{<i,j> \mid <i,j> \in I^* \cap (\widetilde{P}_2 - P_2^*)\}, \tag{25}$$

$$Q_{ij}^{(k,3)} = V_{ij}^{(k)*} - \widetilde{U}_{ij}^{(k)}, \quad S_3 = \{<i,j> \mid <i,j> \in P_1^* \cap (\widetilde{I} - I^*)\}, \tag{26}$$

$$Q_{ij}^{(k,4)} = V_{ij}^{(k)*} - \widetilde{Z}_{ij}^{(k)}, \quad S_4 = \{<i,j> \mid <i,j> \in P_1^* \cap (\widetilde{P}_2 - P_2^*)\}, \tag{27}$$

$$Q_{ij}^{(k,5)} = V_{ij}^{(k)*} - \widetilde{V}_{ij}^{(k)}, \quad S_5 = \{<i,j> \mid <i,j> \in (P_1^* \cap \widetilde{P}_1) \cap (T(x_i, x_j) \neq \widetilde{T}(x_i, x_j))\}, \tag{28}$$

$$Q_{ij}^{(k,6)} = Z_{ij}^{(k)*} - \widetilde{U}_{ij}^{(k)}, \quad S_6 = \{<i,j> \mid <i,j> \in P_2^* \cap (\widetilde{I} - I^*)\}, \tag{29}$$

$$Q_{ij}^{(k,7)} = Z_{ij}^{(k)*} - \widetilde{V}_{ij}^{(k)}, \quad S_7 = \{<i,j> \mid <i,j> \in P_2^* \cap (\widetilde{P}_1 - P_1^*)\}, \tag{30}$$

$$Q_{ij}^{(k,8)} = Z_{ij}^{(k)*} - \widetilde{Z}_{ij}^{(k)}, \quad S_8 = \{<i,j> \mid <i,j> \in P_2^* \cap \widetilde{P}_2) \cap (T(x_i, x_j) \neq \widetilde{T}(x_i, x_j))\}. \tag{31}$$

3. Postać i własności estymatora relacji porządku

Podstawę do określenia postaci estymatora relacji **R** i jego własności stanowią uśrednione zmienne losowe $\overline{U}_{ij}(\cdot)$, $\overline{V}_{ij}(\cdot)$ oraz $\overline{Z}_{ij}(\cdot)$:

$$\overline{U}_{ij}(\cdot) = \frac{1}{N} \sum_{k=1}^{N} U_{ij}^{(k)}(\cdot), \tag{32}$$

$$\overline{V}_{ij}(\cdot) = \frac{1}{N} \sum_{k=1}^{N} V_{ij}^{(k)}(\cdot), \tag{33}$$

$$\overline{Z}_{ij}(\cdot) = \frac{1}{N} \sum_{k=1}^{N} Z_{ij}^{(k)}(\cdot). \tag{34}$$

Wariancje zmiennych (32) – (34) spełniają warunki:

$$\left.\begin{array}{l}\mathrm{Var}(\overline{U}_{ij}(\cdot))=\frac{1}{N}Var(U_{ij}^{(1)}(\cdot)),\\ \mathrm{Var}(\overline{V}_{ij}(\cdot))=\frac{1}{N}Var(V_{ij}^{(1)}(\cdot)),\\ \mathrm{Var}(\overline{Z}_{ij}(\cdot))=\frac{1}{N}Var(Z_{ij}^{(1)}(\cdot)).\end{array}\right\} \tag{35}$$

Suma $\overline{W}(\cdot)$ zmiennych (32) – (34) ma postać:

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

Uśrednione zmienne, odpowiadające relacji $\chi_1^*,\ldots,\chi_n^*$ będą oznaczane symbolami: $\overline{U}_{ij}^*$, $\overline{V}_{ij}^*$, $\overline{Z}_{ij}^*$, $\overline{W}^*$, natomiast zmienne odpowiadające dowolnej innej relacji $\tilde{\chi}_1,\ldots\tilde{\chi}_{\tilde{r}}$ - symbolami $\tilde{\overline{U}}_{ij}$, $\tilde{\overline{V}}_{ij}$, $\tilde{\overline{Z}}_{ij}$, $\tilde{\overline{W}}$.

Różnicę zmiennych losowych $\overline{W}^*$ - $\tilde{\overline{W}}$ można wyrazić w postaci:

$$\overline{W}^* - \tilde{\overline{W}} = \sum_{I^*\cap(\tilde{P}_1-P_1^*)}(\overline{U}_{ij}^*-\tilde{\overline{V}}_{ij})+\sum_{I^*\cap(\tilde{P}_2-P_2^*)}(\overline{U}_{ij}^*-\tilde{\overline{Z}}_{ij})+$$

$$+\sum_{P_1^*\cap(\tilde{I}-I^*)}(\overline{V}_{ij}^*-\tilde{\overline{U}}_{ij})+\sum_{P_1^*\cap(\tilde{P}_2-P_2^*)}(\overline{V}_{ij}^*-\tilde{\overline{Z}}_{ij})+\sum_{(P_1^*\cap\tilde{P}_1)\cap(T(\cdot)\neq\tilde{T}(\cdot))}(\overline{V}_{ij}^*-\tilde{\overline{V}}_{ij})+$$

$$\sum_{P_2^*\cap(\tilde{I}-I^*)}(\overline{Z}_{ij}^*-\tilde{\overline{U}}_{ij})+\sum_{P_2^*\cap(\tilde{P}_1-P_1^*)}(\overline{Z}_{ij}^*-\tilde{\overline{V}}_{ij})+\sum_{(P_2^*\cap\tilde{P}_2)\cap(T(\cdot)\neq\tilde{T}(\cdot))}(\overline{Z}_{ij}^*-\tilde{\overline{Z}}_{ij})=$$

$$=\sum_{\nu=1}^{8}\sum_{S_\nu}\overline{Q}_{ij}^{(\nu)}, \tag{37}$$

gdzie: $\overline{Q}_{ij}^{(\nu)}$ oraz S_ν - zdefiniowane analogicznie, jak w (24) – (31).

Własności różnicy zmiennych losowych $\overline{W}^*$ - $\tilde{\overline{W}}$ określa

Twierdzenie

Różnica $\overline{W}^*$ - $\tilde{\overline{W}}$ spełnia następujące warunki:

$$E(\overline{W}^*-\tilde{\overline{W}})<0, \tag{38}$$

$$P(\overline{W}^* - \tilde{\overline{W}} < 0) \geq 1 - \exp\left\{ -2N \frac{(\sum\limits_{\nu=1}^{8} \sum\limits_{S_\nu} E(Q_{ij}^{(1\nu)}))^2}{\tilde{g}^2 (2(m-1))^2} \right\}, \qquad (39)$$

gdzie: $\tilde{g} = \#(\bigcup\limits_{\nu=1}^{8} S_\nu)$.

Idee dowodów nierówności (38), (39) podano w dodatku.

Powyższe twierdzenie orzeka, że wartość oczekiwana zmiennej losowej $\overline{W}^*$, odpowiadającej relacji $\chi_1^*, \ldots, \chi_n^*$, jest mniejsza, niż zmiennej $\tilde{\overline{W}}$ odpowiadającej dowolnej innej relacji $\tilde{\chi}_1, \ldots, \tilde{\chi}_{\tilde{r}}$, tzn. zmienna $\overline{W}^*$ ma najmniejszą wartość oczekiwaną w zbiorze rozwiązań dopuszczalnych. Jeśli $N \rightarrow \infty$, to wartości każdej zmiennych $\overline{W}(\chi_1, \ldots, \chi_r)$ zbiegają stochastycznie (zob. zależności (35)) do swoich wartości oczekiwanych. Ponadto, prawdopodobieństwo zdarzenia $\{\overline{W}^* < \tilde{\overline{W}}\}$ jest nie mniejsze niż prawa strona nierówności (39); dąży ono wykładniczo do jedności, gdy $N \rightarrow \infty$. Oszacowanie to określa prawdopodobieństwo zdarzenia, że relacja $\chi_1^*, \ldots, \chi_n^*$, wygeneruje mniejszą wartość zmiennej $\overline{W}(\chi_1, \ldots, \chi_r)$, niż różna od niej relacja $\tilde{\chi}_1, \ldots, \tilde{\chi}_{\tilde{r}}$. Powyższe twierdzenie implikuje sposób rozwiązania problemu estymacji relacji $\chi_1^*, \ldots, \chi_n^*$: na podstawie porównań $g_k(x_i, x_j)$ ($k=1, \ldots, N$; $(x_i, x_j) \in \mathbf{X} \times \mathbf{X}$) wyznaczyć relację $\hat{\chi}_1, \ldots, \hat{\chi}_{\hat{n}}$, dla której wartość zmiennej losowej $\overline{W}(\chi_1, \ldots, \chi_r)$ jest najmniejsza (innymi słowy - określić relację, której poziom niezgodności z wynikami porównań jest najmniejszy). W tym celu należy rozwiązać zadanie optymalizacji dyskretnej o postaci:

$$\min_{\chi_1^{(t)},\ldots,\chi_r^{(t)}\in F_X} \{ \sum_{<i,\,j>\in R_m} \sum_{k=1}^{N} \left| t^{(t)}(x_i,x_j)-g_k(x_i,x_j) \right| \}, \tag{40}$$

gdzie:

F_X - zbiór rozwiązań dopuszczalnych, tj. rodzina wszystkich możliwych postaci relacji preferencji w zbiorze **X**;

$t^{(t)}(x_i, x_j)$ – funkcja charakteryzująca relację $\chi_1^{(t)},\ldots\chi_{r^{(t)}}^{(t)}$, element zbioru F_X.

Liczba rozwiązań zadania optymalizacyjnego (40) może przewyższać jeden. W przypadku rozwiązań wielokrotnych, oszacowanie (39) odnosi się do całego zespołu (alternatywy) uzyskanych relacji. Rozwiązanie jednoznaczne można wybrać w sposób losowy lub wprowadzając dodatkowe kryterium, np. wybierając relację minimalizującą wartość funkcji (40) na zbiorze $I(\hat{\chi}_1,\ldots,\hat{\chi}_{\hat{n}})$ lub na zbiorze $P_1(\hat{\chi}_1,\ldots,\hat{\chi}_{\hat{n}})\cup P_2(\hat{\chi}_1,\ldots,\hat{\chi}_{\hat{n}})$. Współczynnik $1/N$ został pominięty w funkcji (40), ponieważ nie wpływa on na postać rozwiązania optymalnego.

Określenie własności oceny $\hat{\chi}_1,\ldots,\hat{\chi}_{\hat{n}}$ może się opierać na wartościach (ocenach) odpowiadającej jej funkcji $\hat{T}(x_i,x_j)$ ($(x_i, x_j)\in\mathbf{X}\times\mathbf{X}$). Zbadanie nieobciążoności tych ocen oraz określenie ich łącznej wariancji w sposób analityczny wydaje się problemem trudnym; dyskusji wymaga też kwestia adekwatności wariancji, jako miernika precyzji oszacowania. Wykazanie zgodności rozważanego estymatora, tzn. stwierdzenia czy jest spełniony warunek:

$$\lim_{N\to\infty} P((\sum_{<i,\,j>\in R_m} \left|\hat{T}(x_i,x_j)-T(x_i,x_j)\right|)<\varepsilon)=0 \qquad (\varepsilon>0),$$

oraz określenie rodzaju (tempa) zbieżności, wydaje się realistyczne jedynie metodami symulacyjnymi. Przesłankami wskazującymi na zgodność tego estymatora są zależności: (35) i (39).

Określenie wartości prawej strony nierówności (39) wymaga znajomości wartości oczekiwanych $E(Q_{ij}^{(1\nu)})$; w celu ich wyznaczenia niezbędna jest m.in. znajomość funkcji prawdopodobieństwa rozkładów błędów porównań. W przypadku, gdy nie są one znane można je estymować lub aproksymować. W celu uzyskania realistycznych wyników estymacji tych funkcji, konieczne jest posiadanie odpowiedniej liczby porównań N. W przypadku, gdy liczba N jest niedostateczna, proponuje się wyznaczyć „górne oszacowania" tych funkcji, oparte na koncepcji symetrycznych rozkładów quasi-równomiernych (zob. Klukowski (2006)). Oszacowania takie wyznacza się na podstawie ocen: $\hat{T}(x_i, x_j)$ ($(x_i, x_i) \in \mathbf{X} \times \mathbf{X}$), rozwiązując jeden z podanych niżej układów równań. W przypadku $\hat{T}(x_i, x_j) \neq \pm(\hat{n}-1)$ układ taki ma postać:

$$\left.\begin{aligned}
&P(\hat{T}(x_i, x_j) - g_k(x_i, x_j) < 0) = P(\hat{T}(x_i, x_j) - g_k(x_i, x_j) > 0),\\
&P(\hat{T}(x_i, x_j) - g_k(x_i, x_j) = -1) =\\
&\qquad = P(\hat{T}(x_i, x_j) - g_k(x_i, x_j) = \hat{T}(x_i, x_j) - (\hat{n}-1)),\\
&P(\hat{T}(x_i, x_j) - g_k(x_i, x_j) = 1) =\\
&\qquad = P(\hat{T}(x_i, x_j) - g_k(x_i, x_j) = \hat{T}(x_i, x_j) + (\hat{n}-1)),\\
&P(\hat{T}(x_i, x_j) - g_k(x_i, x_j) = 0) =\\
&= \max\{P(\hat{T}(x_i, x_j) - g_k(x_i, x_j) = 1),\ \hat{T}(x_i, x_j) - g_k(x_i, x_j) = -1)\},\\
&P(\sum_{l=\hat{T}(\cdot)-(\hat{n}-1)}^{\hat{T}(\cdot)+(\hat{n}-1)} \hat{T}(x_i, x_j) - g_k(x_i, x_j) = l) = 1.
\end{aligned}\right\}(41)$$

W przypadku $\hat{T}(x_i, x_j) = \pm(\hat{n}-1)$ układ ten ma postać :

$P(\hat{T}(x_i, x_j) - g_k(x_i, x_j) = 0) = \frac{1}{2} + \varepsilon$ oraz

$P(\hat{T}(x_i, x_j) - g_k(x_i, x_j) = -l) = \frac{1-2\varepsilon}{2(2(\hat{n}-1))})$ $(l=1, \ldots, 2(\hat{n}-1))$ dla

$$\hat{T}(x_i, x_j) = -(\hat{n}-1),$$

$P(\hat{T}(x_i, x_j) - g_k(x_i, x_j) = l) = \frac{1-2\varepsilon}{2(2(\hat{n}-1))})$ $(l=1, \ldots, 2(\hat{n}-1))$ dla

$$\hat{T}(x_i, x_j) = \hat{n} - 1,$$

$$P(\sum_{l=\hat{T}(\cdot)+(\hat{n}-1)}^{\hat{T}(\cdot)-(\hat{n}-1)} \hat{T}(x_i, x_j) - g_k(x_i, x_j) = l) = 1,$$

gdzie:

ε - stała z przedziału $(0, \frac{1}{2})$ (np. $\varepsilon = \frac{1}{2(2(\hat{n}-1))}$ lub $\varepsilon = \frac{1}{2(2(\hat{n}-1))}$).

Układ równań (41) zakłada równość prawdopodobieństw z lewego i prawego ogona rozkładu (zob. pierwsze równanie układu). W przypadku asymetrii tego rozkładu (np. stwierdzonej na podstawie współczynnika asymetrii porównań $g_k(\cdot)$ $(k=1, \ldots, N)$) należy nadać pierwszemu równaniu postać:

$$P(\hat{T}(x_i, x_j) - g_k(x_i, x_j) < 0) = \lambda P(\hat{T}(x_i, x_j) - g_k(x_i, x_j) > 0),$$

gdzie: λ - stała zapewniająca spełnienie warunków (8) – (12).

Należy dodać, że oszacowanie (39) zależy od postaci relacji $\tilde{\chi}_1, \ldots, \tilde{\chi}_{\tilde{r}}$; relacje wykazujące wyższy „stopień niepodobieństwa" do relacji $\chi_1^*, \ldots, \chi_n^*$ implikują większą wartość oczekiwaną $\sum_{\nu=1}^{8}\sum_{S_\nu} E(Q_{ij}^{(1\nu)})$, i – w konsekwencji – większe prawdopodobieństwo $P(\overline{W}^* < \tilde{\overline{W}})$. Innymi słowy, otrzymanie (jako wyniku estymacji) relacji $\tilde{\chi}_1, \ldots, \tilde{\chi}_{\tilde{r}}$ wykazującej duże różnice w stosunku do $\chi_1^*, \ldots, \chi_n^*$ jest mniej prawdopodobne.

Określenie prawdopodobieństwa zdarzenia, że ocena $\hat{\chi}_1, \ldots, \hat{\chi}_{\hat{n}}$, wynikająca z rozwiązania zadania (40), jest identyczna z postacią relacji $\chi_1^*, \ldots, \chi_n^*$ wydaje się możliwe, jak już zaznaczono, metodami symulacyjnymi. Eksperyment symulacyjny, mający na celu jego określenie, może mieć następującą postać.

1^0. Rozwiązać zadanie (40) dla uzyskanych porównań $g_k(x_i, x_j) \in \mathbf{X} \times \mathbf{X}$ ($k = 1, \ldots, N$) oraz wyznaczyć relację $\hat{\chi}_1, \ldots, \hat{\chi}_{\hat{n}}$.

2^0. Na podstawie oceny $\hat{\chi}_1, \ldots, \hat{\chi}_{\hat{n}}$ oraz posiadanych funkcji prawdopodobieństwa błędów porównań (przybliżonych lub estymowanych) wygenerować w sposób symulacyjny κ zestawów ($\kappa >> 1$) wyników porównań $g_k(x_i, x_j)$ ($k=1, \ldots, N$; $(x_i, x_j) \in \mathbf{X} \times \mathbf{X}$).

3^0. Rozwiązać zadanie (40) dla każdego z κ zestawów porównań określonych w pkt 2^0 i wyznaczyć liczbę μ - bezbłędnych wyników estymacji.

4^0. Określić frakcję κ/μ i przyjąć ją jako ocenę prawdopodobieństwa uzyskania bezbłędnego rozwiązania problemu estymacji relacji porządku w całym zbiorze rozwiązań dopuszczalnych.

W celu realistycznej oceny prawdopodobieństwa uzyskania bezbłędnego wyniku estymacji relacji, liczba κ musi być odpowiednio duża, zwykle nie mniejsza niż 10^2. W przypadku uzyskania rozwiązania wielokrotnego zadania (40) (na podstawie porównań $g_k(\cdot)$) należy zrealizować symulacje 1^0 – 4^0 dla każdego z rozwiązań i przyjąć, jako ocenę prawdopodobieństwa, frakcję κ/μ o najmniejszej wartości. W przypadku pojedynczego rozwiązania zadania (40) (na podstawie rzeczywistych porównań), można też uznać za zasadne zrealizowanie eksperymentu symulacyjnego nie tylko dla relacji $\hat{\chi}_1, \ldots \hat{\chi}_{\hat{n}}$, ale również dla (kilku) relacji, odpowiadających wartościom funkcji (40) bliskim wartości minimalnej. Jako ocenę prawdopodobieństwa bezbłędnego wyniku należy przyjąć najmniejszą z otrzymanych frakcji. Postępowanie takie, wykorzystujące ideę technik *bootstrap*, ma przybliżony charakter, ale stanowi racjonalne rozwiązanie rozważanego problemu, przy przyjętych założeniach.

Należy dodać, że oszacowanie (39) nie zakłada znajomości liczby podzbiorów n relacji **R**; zakres wartości funkcji prawdopodobieństwa błędów porównań opiera się liczności m zbioru **X**. W przypadku, gdy $n<m$, prowadzi to do niedoszacowania prawdopodobieństwa (39). Dlatego też jeśli można uznać, że $\hat{n}$ jest równe n, celowe jest zastąpienie w nierówności (39) liczby m – oceną $\hat{n}$. Jako „zgrubny" wskaźnik zasadności tego przypuszczenia można wykorzystać minimalną wartość funkcji (40). Wartości bliskie zera lub zero (funkcja ta przyjmuje jedynie nieujemne wartości całkowite), oznaczające niski poziom „niezgodności" oceny $\hat{\chi}_1, \ldots \hat{\chi}_{\hat{n}}$ z porównaniami $g_k(x_i, x_j)$, wskazują na dobrą precyzję wyniku estymacji.

4. Zakończenie

Przedstawiona powyżej metoda estymacji relacji porządku - na podstawie porównań parami w postaci różnicy rang - stanowi rozwinięcie koncepcji przedstawionych w pracach Klukowskiego (1994, 2000). Otrzymane w niniejszej pracy wyniki opierają się na słabych założeniach; w szczególności nie wymaga się, aby wartość oczekiwana błędu porównania miała zerową wartość oczekiwaną oraz, aby prawdopodobieństwo zerowej wartości błędu porównania było większe od ½. Co więcej, stosowanie algorytmu nie wymaga znajomości rozkładów błędów porównań; wystarczające jest spełnienie założeń (9) – (12). Możliwe jest określenie pewnych własności tego estymatora, przy wykorzystaniu różnych technik statystycznych, w tym również symulacyjnych.

W trakcie dalszych prac wydaje się zasadne zbadanie własności algorytmu opartego na medianie z porównań, stanowiącego analogię do przypadku „zasady większości", rozważanego w pracy Klukowskiego (1994). Dla tego ostatniego algorytmu wykazano również wykładniczą zbieżność do jedności prawdopodobieństwa analogicznego do prawej strony nierówności (39), a odpowiada mu prostsze zadanie programowania dyskretnego. Ponadto celowe jest szersze zbadanie własności proponowanej metody estymacji (trudnych do analitycznego określenia) w sposób symulacyjny.

Dodatek

Idea dowodu nierówności (44):

$$E(\overline{W}^{*} - \tilde{\overline{W}}) < 0. \qquad \text{(A1)}$$

Rozważmy wartość oczekiwaną $E(Q_{ij}^{(k,1)})$. Zmienna losowa $Q_{ij}^{(k,1)}$ przyjmuje postać:

$$Q_{ij}^{(k,1)} = U_{ij}^{(k)*} - \tilde{V}_{ij}^{(k)} = |g_k(x_i, x_j)| - |\tilde{t}(x_i, x_j) - g_k(x_i, x_j)|$$

$$(<i,j> \mid <i,j> \in I^{*} \cap (\tilde{P}_1 - P_1^{*})). \qquad \text{(A2)}$$

Z definicji zmiennych losowych $U_{ij}^{(k)*}$ oraz $\widetilde{V}_{ij}^{(k)}$ wynika, że $T(\cdot)$=0 oraz $\widetilde{T}(\cdot)<0$; wynik porównania $g_k(x_i, x_j)$ może przyjąć wartość spełniającą jeden z warunków:

(i) $g_k(\cdot)\le\widetilde{T}(\cdot)$;

(ii) $\widetilde{T}(\cdot)<g_k(\cdot)<T(\cdot)$;

(iii) $g_k(\cdot)\ge T(\cdot)$.

W przypadku $g_k(\cdot)\le\widetilde{T}(\cdot)$ różnica $U_{ij}^{(k)*}$-$\widetilde{V}_{ij}^{(k)}$ jest równa: $-\widetilde{T}(\cdot)>0$ z prawdopodobieństwem: $\sum\limits_{l\le\widetilde{T}(\cdot)} P(g_k(\cdot)=l)<½$. W przypadku (iii), różnica (A2) jest równa: $\widetilde{T}(\cdot)<0$ z prawdopodobieństwem $\sum\limits_{l\ge T(\cdot)} P(g_k(\cdot)=l)>½$. Warunek (ii) implikuje, że wyrażenie (A2) jest równe $\widetilde{T}(\cdot)-2g_k(\cdot)$. Wyrażenie $\widetilde{T}(\cdot)-2g_k(\cdot)$ ($\widetilde{T}(\cdot)<g_k(\cdot)<T(\cdot)$) spełnia warunek:

$$\widetilde{T}(\cdot)<\widetilde{T}(\cdot)-2g_k(\cdot)<-\widetilde{T}(\cdot)$$

i przyjmuje wartości ze zbioru $\{\widetilde{T}(\cdot)+2, \ldots, -\widetilde{T}(\cdot)-2\}$, z prawdopodobieństwami $P(\widetilde{T}(\cdot)-2g_k=\iota)=P(g_k=(\widetilde{T}(\cdot)-\iota)/2)$. Wartości wyrażenia $\widetilde{T}(\cdot)-2g_k(\cdot)$ ($\widetilde{T}(\cdot)<g_k(\cdot)<T(\cdot)$) są położone symetrycznie wokół zera; ich prawdopodobieństwa spełniają warunki (zob. (9) – (12)):

$$P(\widetilde{T}(\cdot)-2g_k=-\iota)\ge P(\widetilde{T}(\cdot)-2g_k=\iota) \qquad (\iota>0);$$

ostatnia nierówność wynika z faktu, że w przypadku $\widetilde{T}(\cdot)-2g_k(\cdot)=-\iota$ różnica $T_1(\cdot)$- $g_k(\cdot)$ jest mniejsza (bliższa zeru), niż w przypadku $\widetilde{T}(\cdot)$-

$2g_k(\cdot)=t$ (innymi słowy wartość $g_k(\cdot)$ jest bliższa $T_1(\cdot)$). Koniunkcja powyższych faktów, tj.:

$P(Q_{ij}^{(k,1)}=\widetilde{t}(\cdot))=\sum_{l\geq T(\cdot)} P(g_k=l)>½,$

$P(Q_{ij}^{(k,1)}=-\widetilde{t}(\cdot))=\sum_{l\leq \widetilde{T}_1(\cdot)} P(g_k^{(1)}=l)<½,$

$P(\widetilde{t}(\cdot)-2g_k^{(1)}=-t)\geq P(\widetilde{t}(\cdot)-2g_k^{(1)}=t),$

implikuje:

$E(Q_{ij}^{(k,1)})<0.$

W analogiczny sposób wykazuje się nierówności: $E(Q_{ij}^{(k,\nu)})<0$, $\nu=$ 2, ..., 8. Implikują one ujemną wartość oczekiwaną $E(\overline{W}^* - \overline{\widetilde{W}})$.

□

Dowód nierówności (39):

$$P(\overline{W}^* < \overline{\widetilde{W}}) \geq 1\text{-}\exp\left\{-2N\frac{(\sum_{\nu=1}^{8}\sum_{S_\nu} E(Q_{ij}^{(1\nu)}))^2}{\widetilde{g}^2(2(m-1))^2}\right\}. \tag{A3}$$

Prawdopodobieństwo $P(\overline{W}^* < \overline{\widetilde{W}})$ można wyrazić w postaci:

$P(\overline{W}^* < \overline{\widetilde{W}}) = 1 - P(\overline{W}^* - \overline{\widetilde{W}} \geq 0)$

oraz

$$P(\overline{W}^* - \overline{\widetilde{W}} \geq 0) = P(\sum_{\nu=1}^{8}\sum_{S_\nu}\overline{Q}_{ij}^{(\nu)} \geq 0) =$$

$$= P(\sum_{\nu=1}^{8}\sum_{S_\nu}\frac{1}{N}\sum_{k=1}^{N} Q_{ij}^{(k\nu)} \geq 0) =$$

$$= P(\sum_{k=1}^{N}(\sum_{\nu=1}^{8}\sum_{S_\nu} Q_{ij}^{(k\nu)}) \geq 0). \tag{A4}$$

Nierówność (A4) może być przekształcona w następujący sposób:

$$P(\sum_{k=1}^{N}(\sum_{\nu=1}^{8}\sum_{S_\nu} Q_{ij}^{(k\nu)}) \geq 0) =$$

$$P(\sum_{k=1}^{N}(\sum_{\nu=1}^{8}\sum_{S_\nu} Q_{ij}^{(k\nu)}) - E(\sum_{k=1}^{N}(\sum_{\nu=1}^{8}\sum_{S_\nu} Q_{ij}^{(k\nu)})) \geq -E(\sum_{k=1}^{N}(\sum_{\nu=1}^{8}\sum_{S_\nu} Q_{ij}^{(k\nu)})) =$$

$$= P(\sum_{k=1}^{N}(\sum_{\nu=1}^{8}\sum_{S_\nu} Q_{ij}^{(k\nu)}) - N\sum_{\nu=1}^{8}\sum_{S_\nu} E(Q_{ij}^{(1\nu)}) \geq -N\sum_{\nu=1}^{8}\sum_{S_\nu} E(Q_{ij}^{(1\nu)})). \qquad \text{(A5)}$$

Nierówność (A5) wynika z założenia, że dla wszystkich k $(1 \leq k \leq N)$ rozkłady poszczególnych zmiennych losowych $U_{ij}^{(k)}(\cdot)$, $V_{ij}^{(k)}(\cdot)$ oraz $Z_{ij}^{(k)}(\cdot)$ są identyczne. Dlatego też wartości oczekiwane $E(Q_{ij}^{(k\nu)})$ $(1 \leq k \leq N)$ są również identyczne i równe $E(Q_{ij}^{(1,\nu)})$.

Prawdopodobieństwo (A5) może być obecnie oszacowane na podstawie nierówności Hoeffdinga:

$$P(\sum_{k=1}^{N} Y_i - \sum_{k=1}^{N} E(Y_i) \geq Nt) \leq \exp\{-2Nt^2/(b-a)^2\},$$

gdzie: Y_i $(i=1, \ldots, N)$ – niezależne zmienne losowe spełniające warunek

$P(a \leq Y_i \leq b) = 1$,

$(-\infty < a < b < \infty)$, t – stała dodatnia,

w następujący sposób:

$$P(\sum_{k=1}^{N}(\sum_{\nu=1}^{8}\sum_{S_\nu} Q_{ij}^{(k\nu)}) - N\sum_{\nu=1}^{8}\sum_{S_\nu} E(Q_{ij}^{(1,\nu)})) \geq -N\sum_{\nu=1}^{8}\sum_{S_\nu} E(Q_{ij}^{(1,\nu)})) \leq$$

$$\exp\left\{-2N\frac{(\sum_{\nu=1}^{8}\sum_{S_\nu} E(Q_{ij}^{(1,\nu)}))^2}{\widetilde{g}^2(2(m-1))^2}\right\}. \qquad \text{(A6)}$$

Postać mianownika wykładnika, występującego po prawej stronie nierówności (A6) wynika z faktu, że wartości każdej ze zmiennych losowych $Q_{ij}^{(k\nu)}$ $(k=1, \ldots, N;\ \nu=1, \ldots, 8)$ są zawarte w zbiorze $\{-(m-$

1), ..., $(m\text{-}1)\}$, a zatem ich suma $\sum_{\nu=1}^{8}\sum_{S_\nu} Q_{ij}^{(k\nu)}$, zawierająca $\widetilde{\vartheta}=\#(\bigcup_{\nu=1}^{8} S_\nu)$ składników, przyjmuje wartości ze zbioru $\{-\widetilde{\vartheta}\,2(m\text{-}1), ..., \widetilde{\vartheta}\,2(m\text{-}1)\}$.

$\square$

Literatura

David H. A. (1970) *Order Statistics*, J. Wiley, New York.

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., vol. 58, pp. 13-30.

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

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

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

Klukowski L. (2002) *Estimation of the tolerance relation on the basis of pairwise comparisons with random errors* (in Polish). In.: Bubnicki Z., Hryniewicz O., Kulikowski R. eds. *Methods and Technics of Data Analysis and Decision Support*. Academic Publishing House EXIT, Warsaw, pp. V-21 – V-35.

Klukowski L. (2006) *Tests for relation type - equivalence or tolerance - in finite set of elements.* Control and Cybernetics, vol. 35, pp. 369 - 384.

Klukowski L. (2006) *Estimation of Tolerance Relation on the Basis of Multiple Pairwise Comparisons with Random Errors*. Raport Badawczy RB/39/2006, IBS PAN Warszawa.

Koronacki J., Mielniczuk J. (2005) *Statystyczne systemy uczące się*. WNT, Warszawa.

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