POLSKA AKADEMIA NAUK

Instytut Badań Systemowych

METHODS OF ESTIMATION OF RELATIONS OF: EQUIVALENCE, TOLERANCE AND PREFERENCE IN A FINITE SET

Leszek Klukowski

Warsaw 2011

SYSTEMS RESEARCH INSTITUTE
POLISH ACADEMY OF SCIENCES

Series: SYSTEMS RESEARCH
Volume 69

==

Series Editor:
Prof. dr hab. inż. Jakub Gutenbaum

Warsaw 2011

SYSTEMS RESEARCH INSTITUTE
POLISH ACADEMY OF SCIENCES

==

Leszek Klukowski

METHODS OF ESTIMATION OF RELATIONS OF: EQUIVALENCE TOLERANCE AND PREFERENCE IN A FINITE SET

Warsaw 2011

dr Leszek Klukowski
Systems Research Institute
Polish Academy of Sciences
Newelska 6, 01-447 Warsaw, Poland
email: Leszek.Klukowski@ibspan.waw.pl

Papers reviewers:

Prof. dr hab. inż. Ignacy Kaliszewski
Prof. dr hab. Tadeusz Trzaskalik

The work has been supported by the grant No N N111434937 of the Polish Ministry of Science and Higher Education

Printed in Polands
Systems Research Institute
Polish Academy of Sciences
Newelska 6, 01-447 Warsaw, Poland
www.ibspan.waw.pl

ISSN 0208-8029
ISBN 9788389475374

Chapter 8

Estimation of the preference relation – multivalent comparisons

8.1 Introduction

Multivalent comparisons express, in the case of the preference relation, differences of ranks, i.e. distance in ranking between the elements in a pair. The approach can be applied also in the case of multiple binary comparisons – as the second step. Estimates obtained in the first step have to satisfy the assumptions required by multivalent comparisons. The assumptions about distributions of comparison errors are weak, especially the probabilities of correct comparisons can be lower than ½.

8.2. Assumptions about multivalent comparisons

The preference relation, denoted $\chi_1^{(p)*}, ..., \chi_n^{(p)*}$ or $T_\mu^{(p)}(x_i, x_j)$, has to be estimated on the basis of comparisons $g_{\mu k}^{(p)}(x_i, x_j)$ $(k = 1, ..., N;$ $< i, j > \in R_m)$ defined as follows:

$g_{\mu k}^{(p)}(x_i, x_j) = d_{ijk}^{(p)}$ $(d_{ijk}^{(p)} \in \{0, ..., \pm(m-1)\})$, $d_{ijk}^{(p)}$ - evaluation of the difference of ranks $r - s$ (i.e. $T_\mu^{(p)}(x_i, x_j)$), disturbed by random error, which satisfy the assumptions A1, A2, A3 (Chapter 2),

$g_\mu^{(p,me)}(x_i, x_j)$ - the median in the set of comparisons: $\{g_{\mu,1}^{(p)}(x_i, x_j), ..., g_{\mu N}^{(p)}(x_i, x_j)\}$.

The set of values of $g_{\mu k}^{(p)}(x_i, x_j)$ is not the same as the set of values $T_\mu^{(p)}(x_i, x_j)$, because the number n is assumed to be unknown and is replaced by the number of elements of the set **X**, equal m. Clearly, that the maximum distance of elements has to be lower than m.

8.3. The form of estimators and their properties

The estimators $\hat{\chi}_1^{(p)},\dots,\hat{\chi}_n^{(p)}$ and $\widehat{\chi}_1^{(p)},\dots,\widehat{\chi}_n^{(p)}$ of the preference relation $\chi_1^{(p)*},\dots,\chi_n^{(p)*}$ are obtained on the basis of the following minimization problems:

$$\min_{F_{\mathbf{X}}^{(p)}}\{\sum_{<i,j>\in R_m}\sum_{k=1}^{N}\left|g_{\mu k}^{(p)}(x_i,x_j)-t_{\mu}^{(p)}(x_i,x_j)\right|\}, \tag{8.1}$$

where:

$F_{\mathbf{X}}^{(p)}$ - the feasible set, i.e. the family of all preference relations in the set **X**,

$t_{\mu}^{(p)}(x_i,x_j)$ - the function describing the elements of the set $F_{\mathbf{X}}^{(p)}$, defined in the same way as $T_{\mu}^{(p)}(x_i,x_j)$ (Chapter 2),

and

$$\min_{F_{X}^{(p)}}\{\sum_{<i,j>\in R_m}\left|g_{\mu}^{(p,me)}(x_i,x_j)-t_{\mu}^{(p)}(x_i,x_j)\right|\}, \tag{8.2}$$

where:

$g_{\mu}^{(p,me)}(x_i,x_j)$ - the sample median in the set $\{g_{\mu,1}^{(p)}(x_i,x_j),\dots,g_{\mu N}^{(p)}(x_i,x_j)\}$.

The properties of both estimators are based on properties of the random variables expressing differences between comparisons and the values determining the relation form - in the case of actual relation, $T_{\mu}^{(p)}(x_i,x_j)$, and for any relation, different from the actual one, $\widetilde{T}_{\mu}^{(p)}(x_i,x_j)$.

In the case of the actual relation, any difference can be expressed by one of the random variables:

$$U_{\mu k}^{(p)*}(x_i,x_j)=\begin{cases}0 \;\; \textit{if} \;\; g_{\mu k}^{(p)}(x_i,x_j)=T_{\mu}^{(p)}(x_i,x_j) \textit{ and } T_{\mu}^{(p)}(x_i,x_j)=0;\\ T_{\mu}^{(p)}(x_i,x_j)-g_{\mu k}^{(p)}(x_i,x_j) \textit{ if } \; g_{\mu k}^{(p)}(x_i,x_j)<T_{\mu}^{(p)}(x_i,x_j) \textit{ and}\\ T_{\mu}^{(p)}(x_i,x_j)=0;\\ g_{\mu k}^{(p)}(x_i,x_j)-T_{\mu}^{(p)}(x_i,x_j) \textit{ if } \; g_{\mu k}^{(p)}(x_i,x_j)>T_{\mu}^{(p)}(x_i,x_j) \textit{ and}\\ T_{\mu}^{(p)}(x_i,x_j)=0,\end{cases}$$

$$(8.3)$$

$$V_{\mu k}^{(p)*}(x_i,x_j)=\begin{cases} 0 \;\; if \;\; g_{\mu k}^{(p)}(x_i,x_j)=T_{\mu}^{(p)}(x_i,x_j) \; and \; T_{\mu}^{(p)}(x_i,x_j)<0; \\ T_{\mu}^{(p)}(x_i,x_j)-g_{\mu k}^{(p)}(x_i,x_j) \; if \; g_{\mu k}^{(p)}(x_i,x_j)<T_{\mu}^{(p)}(x_i,x_j) \; and \\ T_{\mu}^{(p)}(x_i,x_j)<0; \\ g_{\mu k}^{(p)}(x_i,x_j)-T_{\mu}^{(p)}(x_i,x_j) \; if \; g_{\mu k}^{(p)}(x_i,x_j)>T_{\mu}^{(p)}(x_i,x_j) \; and \\ T_{\mu}^{(p)}(x_i,x_j)<0, \end{cases} \tag{8.4}$$

$$Z_{\mu k}^{(p)*}(x_i,x_j)=\begin{cases} 0 \;\; if \;\; g_{\mu k}^{(p)}(x_i,x_j)=T_{\mu}^{(p)}(x_i,x_j) \; and \; T_{\mu}^{(p)}(x_i,x_j)>0; \\ T_{\mu}^{(p)}(x_i,x_j)-g_{\mu k}^{(p)}(x_i,x_j) \;\; if \; g_{\mu k}^{(p)}(x_i,x_j)<T_{\mu}^{(p)}(x_i,x_j); \\ T_{\mu}^{(p)}(x_i,x_j)>0; \\ g_{\mu k}^{(p)}(x_i,x_j)-T_{\mu}^{(p)}(x_i,x_j) \;\; if \; g_{\mu k}^{(p)}(x_i,x_j)>T_{\mu}^{(p)}(x_i,x_j); \\ T_{\mu}^{(p)}(x_i,x_j)>0, \end{cases} \tag{8.5}$$

The sum of such differences assumes, for any k $(1 \le k \le N)$, the form:

$$\sum_{<i,j>\in I^{(p)*}} U_{\mu k}^{(p)*}(x_i,x_j)+\sum_{<i,j>\in J_1^{(p)*}} V_{\mu k}^{(p)*}(x_i,x_j)+\sum_{<i,j>\in J_2^{(p)*}} Z_{\mu k}^{(p)*}(x_i,x_j)\,, \tag{8.6}$$

where:

$I^{(p)*}$ - the set of pairs $\{<i,j> \; \big| \, T_{\mu}^{(p)*}(x_i,x_j)=0\}$,

$J_1^{(p)*}$ - the set of pairs $\{<i,j> \; \big| \, T_{\mu}^{(p)*}(x_i,x_j)<0\}$,

$J_2^{(p)*}$ - the set of pairs $\{<i,j> \; \big| \, T_{\mu}^{(p)*}(x_i,x_j)>0\}$,

while the total sum of the differences - the form:

$$W_{\mu N}^{(p)*}=\sum_{k=1}^{N}\Big(\sum_{<i,j>\in I^{(e)*}} U_{\mu k}^{(p)}(x_i,x_j)+\sum_{<i,j>\in J_1^{(e)*}} V_{\mu k}^{(p)}(x_i,x_j)+\sum_{<i,j>\in J_2^{(e)*}} Z_{\mu k}^{(p)}(x_i,x_j)\Big)\,. \tag{8.7}$$

The same random variables, defined for any relation $\widetilde{\chi}_1^{(p)},..., \widetilde{\chi}_{\widetilde{n}}^{(p)}$, different from $\chi_1^{(p)*},..., \chi_n^{(p)*}$, assume the form:

$$\widetilde{U}_{\mu k}^{(p)}(x_i,x_j)=\begin{cases}0 \;\; if \;\; g_{\mu k}^{(p)}(x_i,x_j)=\widetilde{T}_{\mu}^{(p)}(x_i,x_j) \; and \; \widetilde{T}_{\mu}^{(p)}(x_i,x_j)=0;\\ \widetilde{T}_{\mu}^{(p)}(x_i,x_j)-g_{\mu k}^{(p)}(x_i,x_j) \; if \; g_{\mu k}^{(p)}(x_i,x_j)<\widetilde{T}_{\mu}^{(p)}(x_i,x_j) \; and\\ \widetilde{T}_{\mu}^{(p)}(x_i,x_j)=0;\\ g_{\mu k}^{(p)}(x_i,x_j)-\widetilde{T}_{\mu}^{(p)}(x_i,x_j) \; if \; g_{\mu k}^{(p)}(x_i,x_j)>\widetilde{T}_{\mu}^{(p)}(x_i,x_j) \; and\\ \widetilde{T}_{\mu}^{(p)}(x_i,x_j)=0,\end{cases} \tag{8.8}$$

$$\widetilde{V}_{\mu k}^{(p)}(x_i,x_j)=\begin{cases}0 \;\; if \;\; g_{\mu k}^{(p)}(x_i,x_j)=\widetilde{T}_{\mu}^{(p)}(x_i,x_j) \; and \; \widetilde{T}_{\mu}^{(p)}(x_i,x_j)<0;\\ \widetilde{T}_{\mu}^{(p)}(x_i,x_j)-g_{\mu k}^{(p)}(x_i,x_j) \;\; if \;\; g_{\mu k}^{(p)}(x_i,x_j)<\widetilde{T}_{\mu}^{(p)}(x_i,x_j) \; and\\ \widetilde{T}_{\mu}^{(p)}(x_i,x_j)<0;\\ g_{\mu k}^{(p)}(x_i,x_j)-\widetilde{T}_{\mu}^{(p)}(x_i,x_j) \;\; if \;\; g_{\mu k}^{(p)}(x_i,x_j)>\widetilde{T}_{\mu}^{(p)}(x_i,x_j) \; and\\ \widetilde{T}_{\mu}^{(p)}(x_i,x_j)<0;\end{cases} \tag{8.9}$$

$$\widetilde{Z}_{\mu k}^{(p)}(x_i,x_j)=\begin{cases}0 \;\; if \;\; g_{\mu k}^{(p)}(x_i,x_j)=\widetilde{T}_{\mu}^{(p)}(x_i,x_j) \; and \; \widetilde{T}_{\mu}^{(p)}(x_i,x_j)>0;\\ \widetilde{T}_{\mu}^{(p)}(x_i,x_j)-g_{\mu k}^{(p)}(x_i,x_j) \;\; if \;\; g_{\mu k}^{(p)}(x_i,x_j)<\widetilde{T}_{\mu}^{(p)}(x_i,x_j) \; and\\ T_{\mu}^{(p)}(x_i,x_j)>0;\\ g_{\mu k}^{(p)}(x_i,x_j)-\widetilde{T}_{\mu}^{(p)}(x_i,x_j) \;\; if \;\; g_{\mu k}^{(p)}(x_i,x_j)>\widetilde{T}_{\mu}^{(p)}(x_i,x_j) \; and\\ \widetilde{T}_{\mu}^{(p)}(x_i,x_j)>0;\end{cases} \tag{8.10}$$

The sum of such differences assumes, for any k $(1 \le k \le N)$, the form:

$$\sum_{<i,j>\in I^{(p)*}} U_{\mu k}^{(p)*}(x_i,x_j)+\sum_{<i,j>\in J_1^{(p)*}} V_{\mu k}^{(p)*}(x_i,x_j)+\sum_{<i,j>\in J_2^{(p)*}} Z_{\mu k}^{(p)*}(x_i,x_j)\,, \tag{8.11}$$

where:

$\widetilde{I}^{(p)}$ - the set of pairs $\{<i,j> \;\big|\; \widetilde{T}_{\mu}^{(p)}(x_i,x_j)=0\}$,

$\widetilde{J}_1^{(p)}$ - the set of pairs $\{<i,j> \;\big|\; \widetilde{T}_{\mu}^{(p)}(x_i,x_j)<0\}$,

$\widetilde{J}_2^{(p)}$ - the set of pairs $\{<i,j> \;\big|\; \widetilde{T}_{\mu}^{(p)}(x_i,x_j)>0\}$,

while the total sum of the differences the form:

$$\widetilde{W}_{\mu N}^{(p)} = \sum_{k=1}^{N}\Big(\sum_{<i,j>\in\widetilde{I}^{(p)}} \widetilde{U}_{\mu k}^{(p)}(x_i,x_j) + \sum_{<i,j>\in\widetilde{J}_1^{(e)}} \widetilde{V}_{\mu k}^{(p)}(x_i,x_j) + \sum_{<i,j>\in\widetilde{J}_2^{(e)}} \widetilde{Z}_{\mu k}^{(p)}(x_i,x_j)\Big) .(8.12)$$

The same random variables, corresponding to the median estimator, assume the form:

$$U_{\mu}^{(p,me)*}(x_i,x_j) = \begin{cases} 0 \ \ if \ \ g_{\mu}^{(p,me)}(x_i,x_j) = T_{\mu}^{(p)}(x_i,x_j) \ and \ T_{\mu}^{(p)}(x_i,x_j) = 0; \\ T_{\mu}^{(p)}(x_i,x_j) - g_{\mu k}^{(p,me)}(x_i,x_j) \ if \ g_{\mu k}^{(p,me)}(x_i,x_j) < T_{\mu}^{(p)}(x_i,x_j) \\ and \ T_{\mu}^{(p)}(x_i,x_j) = 0; \\ g_{\mu}^{(p,me)}(x_i,x_j) - T_{\mu}^{(p)}(x_i,x_j) \ if \ g_{\mu}^{(p,me)}(x_i,x_j) > T_{\mu}^{(p)}(x_i,x_j) \\ and \\ T_{\mu}^{(p)}(x_i,x_j) = 0, \end{cases}$$

(8.13)

$$V_{\mu}^{(p,me)*}(x_i,x_j) = \begin{cases} 0 \ \ if \ \ g_{\mu}^{(p,me)}(x_i,x_j) = T_{\mu}^{(p)}(x_i,x_j) \ and \ T_{\mu}^{(p)}(x_i,x_j) < 0; \\ T_{\mu}^{(p)}(x_i,x_j) - g_{\mu}^{(p,me)}(x_i,x_j) \ if \ g_{\mu}^{(p,me)}(x_i,x_j) < T_{\mu}^{(p)}(x_i,x_j) \\ and \ T_{\mu}^{(p)}(x_i,x_j) < 0; \\ g_{\mu}^{(p,me)}(x_i,x_j) - T_{\mu}^{(p)}(x_i,x_j) \ if \ g_{\mu k}^{(p)}(x_i,x_j) > T_{\mu}^{(p)}(x_i,x_j) \\ and \ T_{\mu}^{(p)}(x_i,x_j) < 0, \end{cases}$$

(8.14)

$$Z_{\mu}^{(p,me)*}(x_i,x_j) = \begin{cases} 0 \ \ if \ \ g_{\mu}^{(p,me)}(x_i,x_j) = T_{\mu}^{(p)}(x_i,x_j) \ and \ T_{\mu}^{(p)}(x_i,x_j) > 0; \\ T_{\mu}^{(p)}(x_i,x_j) - g_{\mu}^{(p,me)}(x_i,x_j) \ \ if \ g_{\mu}^{(p,me)}(x_i,x_j) < T_{\mu}^{(p)}(x_i,x_j) \\ and \ T_{\mu}^{(p)}(x_i,x_j) > 0; \\ g_{\mu}^{(p,me)}(x_i,x_j) - T_{\mu}^{(p)}(x_i,x_j) \ \ if \ g_{\mu}^{(p,me)}(x_i,x_j) > T_{\mu}^{(p)}(x_i,x_j) \\ and \ T_{\mu}^{(p)}(x_i,x_j) > 0, \end{cases}$$

(8.15)

The total sum of such differences assumes the following form:

$$W_{\mu}^{(p,me)*}(x_i,x_j)=\sum_{<i,j>\in I^{(p)*}} U_{\mu}^{(p,me)*}(x_i,x_j)+\sum_{<i,j>\in J_1^{(p)*}} V_{\mu}^{(p,me)*}(x_i,x_j)+ \sum_{<i,j>\in J_2^{(p)*}} Z_{\mu}^{(p,me)*}(x_i,x_j). \tag{8.16}$$

The random variables, defined for any relation $\tilde{\chi}_1^{(p)},...,\tilde{\chi}_{\tilde{n}}^{(p)}$, different from $\chi_1^{(p)*},...,\chi_n^{(p)*}$ assume the form:

$$\tilde{U}_{\mu}^{(p,me)}(x_i,x_j)=\begin{cases}0 \ \ if \ \ g_{\mu}^{(p,me)}(x_i,x_j)=\tilde{T}_{\mu}^{(p)}(x_i,x_j) \ and \ \tilde{T}_{\mu}^{(p)}(x_i,x_j)=0;\\ \tilde{T}_{\mu}^{(p)}(x_i,x_j)-g_{\mu}^{(p,me)}(x_i,x_j) \ if \ g_{\mu}^{(p,me)}(x_i,x_j)<\tilde{T}_{\mu}^{(p)}(x_i,x_j) \ and \\ \tilde{T}_{\mu}^{(p)}(x_i,x_j)=0;\\ g_{\mu}^{(p,me)}(x_i,x_j)-\tilde{T}_{\mu}^{(p)}(x_i,x_j) \ if \ g_{\mu}^{(p,me)}(x_i,x_j)>\tilde{T}_{\mu}^{(p)}(x_i,x_j) \ and \\ \tilde{T}_{\mu}^{(p)}(x_i,x_j)=0,\end{cases} \tag{8.17}$$

$$\tilde{V}_{\mu}^{(p.me)}(x_i,x_j)=\begin{cases}0 \ \ if \ \ g_{\mu}^{(p,me)}(x_i,x_j)=\tilde{T}_{\mu}^{(p)}(x_i,x_j) \ and \ \tilde{T}_{\mu}^{(p)}(x_i,x_j)<0;\\ \tilde{T}_{\mu}^{(p)}(x_i,x_j)-g_{\mu}^{(p,me)}(x_i,x_j) \ \ if \ g_{\mu}^{(p,me)}(x_i,x_j)<\tilde{T}_{\mu}^{(p)}(x_i,x_j) \ and \\ \tilde{T}_{\mu}^{(p)}(x_i,x_j)<0;\\ g_{\mu}^{(p,me)}(x_i,x_j)-\tilde{T}_{\mu}^{(p)}(x_i,x_j) \ \ if \ g_{\mu}^{(p,me)}(x_i,x_j)>\tilde{T}_{\mu}^{(p)}(x_i,x_j) \ and \\ \tilde{T}_{\mu}^{(p)}(x_i,x_j)<0;\end{cases} \tag{8.18}$$

$$\tilde{Z}_{\mu k}^{(p,me)}(x_i,x_j)=\begin{cases}0 \ \ if \ \ g_{\mu k}^{(p,me)}(x_i,x_j)=\tilde{T}_{\mu}^{(p)}(x_i,x_j) \ and \ \tilde{T}_{\mu}^{(p)}(x_i,x_j)>0;\\ \tilde{T}_{\mu}^{(p)}(x_i,x_j)-g_{\mu k}^{(p,me)}(x_i,x_j) \ \ if \ g_{\mu k}^{(p,me)}(x_i,x_j)<\tilde{T}_{\mu}^{(p)}(x_i,x_j) \ and \\ T_{\mu}^{(p)}(x_i,x_j)>0;\\ g_{\mu k}^{(p,me)}(x_i,x_j)-\tilde{T}_{\mu}^{(p)}(x_i,x_j) \ \ if \ g_{\mu k}^{(p,me)}(x_i,x_j)>\tilde{T}_{\mu}^{(p)}(x_i,x_j) \ and \\ \tilde{T}_{\mu}^{(p)}(x_i,x_j)>0;\end{cases} \tag{8.19}$$

The total sum of such differences assumes, for any k $(1\le k\le N)$, the following form:

$$\widetilde{W}_{\mu}^{(p,me)}(x_i,x_j)=\sum_{<i,j>\in I^{(p)*}}\widetilde{U}_{\mu k}^{(p,me)}(x_i,x_j)+\sum_{<i,j>\in J_1^{(p)*}}\widetilde{V}_{\mu k}^{(p,me)}(x_i,x_j)+ \sum_{<i,j>\in J_2^{(p)*}}\widetilde{Z}_{\mu k}^{(p,me)}(x_i,x_j). \tag{8.20}$$

It can be shown that:

Theorem 5

The following relationships are true:

$$E(W_{\mu N}^{(p)*}-\widetilde{W}_{\mu N}^{(p)})<0\,, \tag{8.21}$$

$$E(W_{\mu N}^{(p,me)*}-\widetilde{W}_{\mu N}^{(p,me)})<0\,, \tag{8.22}$$

$$\lim_{N\to\infty} Var(\tfrac{1}{N}W_{\mu N}^{(p)*})=0\,, \tag{8.23}$$

$$\lim_{N\to\infty} Var(\tfrac{1}{N}\widetilde{W}_{\mu N}^{(p)})=0\,, \tag{8.24}$$

$$\lim_{N\to\infty} Var(W_{\mu N}^{(p,me)*})=0\,, \tag{8.25}$$

$$\lim_{N\to\infty} Var(\widetilde{W}_{\mu N}^{(p,me)})=0\,. \tag{8.26}$$

The probability $P(W_{\mu N}^{(p)*}<\widetilde{W}_{\mu N}^{(p)})$ satisfies the inequality:

$$P(W_{N}^{(p)*}<\widetilde{W}_{N}^{(p)})\geq 1- \exp\{-2N\frac{\left(\sum_{<i,j>\in R_m}\sum_{k=1}^{N}E\left(\left|g_{\mu k}^{(p)}(x_i,x_j)-T_{\mu}^{(p)}(x_i,x_j)\right|-\left|g_{\mu k}^{(p)}(x_i,x_j)-\widetilde{T}_{\mu}^{(p)}(x_i,x_j)\right|\right)\right)^2}{(2\vartheta(m-1))^2}\} \tag{8.27}$$

where:

ϑ - the number of elements of the set $\{(x_i,x_j)\mid T_{\mu}^{(p)}(x_j,x_j)\neq\widetilde{T}_{\mu}^{(p)}(x_j,x_j)\}$, which implies

$$\lim_{N\to\infty} P(W_N^{(p)*} < \widetilde{W}_N^{(p)}) = 1.$$

In the case of the median estimator the following relationship is true:

$$\lim_{N\to\infty} P(W_{\mu N}^{(\tau,me)*} < \widetilde{W}_{\mu N}^{(\tau,me)}) = 1. \tag{8.28}$$

The proofs of the relationships (8.21)–(8.27) are given in Klukowski (2008b). The inequality (8.27) is based on Hoeffding's (1963) inequality for bounded random variable.

The convergence of the variances (8.25), (8.26) and the probability (8.28) results from the formula establishing the distribution of the sample median (David, 1970):

$$P(g_\mu^{(\tau,me)}(x_i,x_j) - T_\mu^{(\tau,me)}(x_i,x_j) = 0) = \frac{N!}{(((N-1)/2)!)^2} \int_{G(-1)}^{G(0)} t^{(N-1)/2}(1-t)^{(N-1)/2} dt\,, \tag{8.29}$$

where:

$G(\cdot)$ - the cumulative distribution function of the comparison error for the pair (x_i, x_j).

Inequality (8.28) based on the relationship (8.29) has been obtained in Klukowski (2008b).

The existence of distributions of both estimators (8.2), (8.3) can be proven in the same way as in Chapter 2.

The minimization problems (8.2), (8.3) require for solving more complex algorithms than those for binary problems. For $m \le 12$ complete enumeration can be applied. Validation of estimates is discussed in Chapter 10.

The convergence for (8.27), (8.28) is obvious in the case of identical distributions of comparisons of each pair, but is also valid for different distributions.

Thus, the estimators based on optimal solutions of the tasks (8.2), (8.3) are consistent.

Inequality (8.27) is applicable for known distributions of comparison errors and for fixed relation $\tilde{\chi}_1^{(p)}, ..., \tilde{\chi}_{\tilde{n}}^{(p)}$. If the distributions are unknown, they can be replaced by quasi-uniform (upper bound) distributions, similar to those for the tolerance relation and multivalent comparisons (Chapter 7). An example of such distribution is given in Klukowski (2008b). For a sufficient *N*, at least several, the distributions of comparison errors can be estimated.

The properties of the estimators have been examined through simulation; the results are presented in the next chapter. In general, the survey shows good efficiency of the estimators proposed and the advantage of the estimator based on sums of differences between the comparisons and the relation form. Especially, the first estimator requires 5 to 7 comparisons in order to guarantee the frequency of errorless comparison close to one. The second estimator requires approximately two comparisons more.

The approach can be also developed for the case of partial orders, i.e. the relations which are not complete. For this purpose it is necessary to introduce an additional result of comparison, corresponding to incomparable elements.

8.4. Summary

The estimators of the preference relation, based on multivalent comparisons, have good statistical properties, obtained in the analytical way. Additional properties, expressing the precision of estimates, have been investigated in the simulation study, presented in Chapter 9. The results confirm good efficiency and allow for determining the respective parameters, especially the number of comparisons *N*, guaranteeing an appropriate precision of estimates.

The book presents the estimators of three relations: equivalence, tolerance, and preference in a finite set of data items, based on multiple pairwise comparisons, assumed to be disturbed by random errors. The estimators were developed by the author. They can refer to binary (qualitative), multivalent (quantitative) and combined comparisons. The estimates are obtained on the basis of solutions to the discrete programming problems. The estimators have been developed under weak assumptions on the distributions of comparison errors; in particular, these distributions can have non-zero expected values. The estimators have good statistical properties, including, especially importantly, consistency. Therefore, they produce good results in cases when other methods generate incorrect estimates. The precision of the estimators has been established with the use of simulation methods. The estimates can be validated in a versatile way. The whole estimation process, i.e. comparisons, estimation and validation can be computerized. The approach allows also for inference about the relation type – equivalence or tolerance, on the basis of binary data. Thus, it has features of data mining methods.

The estimators have been applied for ranking and grouping of data from some empirical sets. In particular, estimation of the tolerance relation (overlapping classification) was applied for determination of homogenous shapes of functions expressing profitability of treasury securities and was used for forecasting purposes.

ISSN 0208-8029
ISBN 9788389475374

SYSTEMS RESEARCH INSTITUTE
POLISH ACADEMY OF SCIENCES
Phone: (+48) 22 3810246 / 22 3810277 / 22 3810241 / 22 3810273
email: biblioteka@ibspan.waw.pl