



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

Editorial Board

Series: SYSTEMS RESEARCH

Prof. Olgierd Hryniewicz - chairman

Prof. Jakub Gutenbaum – series editor

Prof. Janusz Kacprzyk

Prof. Tadeusz Kaczorek

Prof. Roman Kulikowski

Prof. Marek Libura

Prof. Krzysztof Malinowski

Prof. Zbigniew Nahorski

Prof. Marek Niezgódka

Prof. Roman Słowiński

Prof. Jan Studziński

Prof. Stanisław Walukiewicz

Prof. Andrzej Weryński

Prof. Antoni Żochowski



**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

**Copyright © by Systems Research Institute
Polish Academy of Sciences
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

Main notations

$\mathbf{X} = \{x_1, \dots, x_m\}$ ($3 \leq m < \infty$) – set of elements

$\mathbf{R}^{(e)}$ - equivalence relation

$\mathbf{R}^{(\tau)}$ – tolerance relation

$\mathbf{R}^{(p)}$ – preference relation

$\chi_1^{(\ell)*}, \dots, \chi_n^{(\ell)*}$ ($\ell \in \{p, e, \tau\}; n \geq 2$) - relation form (family of subsets)

$T_v^{(\ell)}(x_i, x_j)$ - the function expressing relation $\chi_1^{(\ell)*}, \dots, \chi_n^{(\ell)*}$

$\tilde{T}_v^{(\ell)}(x_i, x_j)$ - the function expressing relation $\tilde{\chi}_1^{(\ell)}, \dots, \tilde{\chi}_n^{(\ell)}$

$g_{vk}^{(\ell)}(x_i, x_j)$ ($\ell \in \{e, \tau, p\}; v \in \{b, \mu\}; k = 1, \dots, N$) - pairwise comparisons

b – binary comparisons

μ - multivalent comparisons

$g_v^{(\ell, me)}(x_i, x_j)$ - median from the comparisons of a pair (x_i, x_j)

$F_X^{(\ell)}$ - the feasible set of optimization problem

$\hat{\chi}_1^{(\ell)}, \dots, \hat{\chi}_n^{(\ell)}$ - the estimator of the relation based on the sum of differences between the comparisons and the relation form

$\tilde{\chi}_1^{(\ell)}, \dots, \tilde{\chi}_n^{(\ell)}$ - the estimator of the relation based on the sum of differences between medians from comparisons for each pair and the relation form

$\hat{T}_v^{(\ell)}(x_i, x_j)$ - the function expressing an estimate $\hat{\chi}_1^{(\ell)}, \dots, \hat{\chi}_n^{(\ell)}$

$\tilde{T}_v^{(\ell)}(x_i, x_j)$ - the function expressing an estimate $\tilde{\chi}_1^{(\ell)}, \dots, \tilde{\chi}_n^{(\ell)}$

$\hat{\Delta}_v^{(\ell)} = \sum_{\langle i, j \rangle \in R_m} \left| \hat{T}_v^{(\ell)}(x_i, x_j) - T_v^{(\ell)}(x_i, x_j) \right|$ - one-dimensional estimation error (estimator based on the sums of differences)

$\tilde{\Delta}_v^{(\ell)} = \sum_{\langle i, j \rangle \in R_m} \left| \tilde{T}_v^{(\ell)}(x_i, x_j) - T_v^{(\ell)}(x_i, x_j) \right|$ - one-dimensional estimation error (estimator based on medians from comparisons)

$W_{vN}^{(\ell)*}$ - random variable expressing differences between comparisons

$\mathcal{G}_{vk}^{(\ell)}(x_i, x_j)$ and values $T_v^{(\ell)}(x_i, x_j)$

$\tilde{W}_{vN}^{(\ell)}$ - random variable expressing differences between comparisons

$\mathcal{G}_{vk}^{(\ell)}(x_i, x_j)$ and values $\tilde{T}_v^{(\ell)}(x_i, x_j)$

List of tables

Table 1. The efficiency of estimators based on binary comparisons, $n=9$ subsets (linear order) 90

Table 2. The efficiency of estimators based on binary comparisons, $n=6$ subsets; number of elements: 1, 2, 1, 2, 1, 2. 91

Table 3. The efficiency of estimators based on binary comparisons, $n=3$ subsets; number of elements: 2, 3, 4. 92

Table 4. The efficiency of estimators based on multivalent comparisons, variant I of distribution function; $n=9$ subsets (linear order) . . . 96

Table 5. The efficiency of estimators based on multivalent comparisons, $n=6$ subsets, variant I of distribution function; number of elements: 1, 2, 1, 2, 1, 2. 97

Table 6. The efficiency of estimators based on multivalent comparisons, variant I of distribution function $n=3$ subsets; number of elements: 2, 3, 4. 98

Table 7. The efficiency of estimators based on multivalent comparisons, variant II of distribution function; $n=9$ subsets (linear order) . . . 99

Table 8. The efficiency of estimators based on multivalent comparisons, $n=6$ subsets, variant II of distribution function; number of elements: 1, 2, 1, 2, 1, 2. 100

Table 9. The efficiency of estimators based on multivalent comparisons, variant II of distribution function $n=3$ subsets; number of elements: 2, 3, 4. 101

Table 10. Frequencies of estimation errors; variant I of the probability function; $n=9$ subsets, $\alpha=0,3334$ 102

Table 11. Frequencies of estimation errors; variant I of the probability function; $n=9$ subsets, $\alpha=0,4167$ 102

Table 12. Frequencies of estimation errors; variant I of the probability function; $n=9$ subsets, $\alpha=0,5000$ 103

Table 13. Frequencies of estimation errors; variant I of the probability function; $n=6$ subsets, $\alpha=0,3334$ 103

Table 14. Frequencies of estimation errors; variant I of the probability function; $n=6$ subsets, $\alpha=0,4167$ 104

Table 15. Frequencies of estimation errors; variant I of the probability function; $n=6$ subsets, $\alpha=0,5000$	104
Table 16. Frequencies of estimation errors; variant I of the probability function; $n=3$ subsets, $\alpha=0,3334$	105
Table 17. Frequencies of estimation errors; variant I of the probability function; $n=3$ subsets, $\alpha=0,4167$	105
Table 18. Frequencies of estimation errors; variant I of the probability function; $n=3$ subsets, $\alpha=0,5000$	106

List of graphs

Graph 1. Frequencies of errors of the estimator based on sums of differences, for $N=1, 3, 5, 7, 9$ (serie 1 – 5); $n=3, \alpha=0,3334$ 107

Graph 2. Frequencies of errors of the estimator based on medians, for $N=1, 3, 5, 7, 9$ (serie 1 – 5); $n=3, \alpha=0,3334$ 107

Graph 3. Frequencies of errors of the estimator based on sums of differences, for $N=1, 3, 5, 7, 9$ (serie 1 – 5) $n=3, \alpha=0,5000$ 107

Graph 4. Frequencies of errors of the estimator based on medians, for $N=1, 3, 5, 7, 9$ (serie 1 – 5); $n=3, \alpha=0,5000$ 108

Graph 5. Frequencies of errors of the estimator based on sums of differences, for $N=1, 3, 5, 7, 9$ (serie 1 – 5); $n=9, \alpha=0,3334$ 108

Graph 6. Frequencies of errors of the estimator based on medians, for $N=1, 3, 5, 7, 9$ (serie 1 – 5); $n=9, \alpha=0,3334$ 108

Graph 7. Frequencies of errors of the estimator based on sums of differences, for $N=1, 3, 5, 7, 9$ (serie 1 – 5); $n=9, \alpha=0,5000$ 109

Graph 8. Frequencies of errors of the estimator based on medians, for $N=1, 3, 5, 7, 9$ (serie 1 – 5); $n=9, \alpha=0,500$ 109

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