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

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 4

Estimation of the tolerance relation – binary comparisons

4.1. Introduction

The problem of estimation of the tolerance relation is discussed first for binary comparisons. The estimators of the relation are constructed in a way similar to the estimators of equivalence relation.

4.2. Assumptions about binary comparisons

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

$$g_{bk}^{(\tau)}(x_i,x_j)=\begin{cases}0 & \text{if } k-\text{th comparison indicates that a pair } (x_i,x_j) \\ & \text{belongs to } \chi_q^{(\tau)*}\cap\chi_s^{(\tau)*} \ (1\le q,s\le m-1); \\ 1 & \text{otherwise.}\end{cases} \tag{4.1}$$

The comparisons $g_{bk}^{(\tau)}(x_i,x_j)$ $(<i,j>\in R_m)$ evaluate values $T_b^{(\tau)}(x_i,x_j)$ have to satisfy the assumptions A1, A2, A3, i.e. the probability of correct comparison $1-\delta$ have to be higher than probability of incorrect comparison δ, and all comparisons are stochastically independent. The assumption about independence can be relaxed in a similar way as in the case of equivalence relation; the relaxation is important if comparisons are the results of statistical test.

The number of subsets n is assumed unknown and has to be bounded by the number of elements m. The number of subsets of the tolerance relation, with at least one intersection, has to be lower than the number of elements, i.e. lower than m.

4.3. The form of estimators and their properties

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

$$\min_{F_{\mathbf{X}}^{(\tau)}} \{ \sum_{<i,j>\in R_m} \sum_{k=1}^{N} \left| g_{bk}^{(\tau)}(x_i, x_j) - t_b^{(\tau)}(x_i, x_j) \right| \}, \tag{4.2}$$

where:

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

and

$$\min_{F_{X}^{(\tau)}} \{ \sum_{<i,j>\in R_m} \left| g_b^{(\tau,me)}(x_i, x_j) - t_b^{(\tau)}(x_i, x_j) \right| \}, \tag{4.3}$$

where:

$g_b^{(\tau,me)}(x_i, x_j)$ - the sample median in the set $\{g_{b,1}^{(\tau)}(x_i, x_j), ..., g_{bN}^{(\tau)}(x_i, x_j)\}$.

The existence of distributions of both estimators (4.2), (4.3) can be proven in the same way as in Chapter 2.

The properties of both estimators are based on properties of the random variables $W_{bN}^{(\tau)*}$, $W_{bN}^{(\tau,me)*}$, corresponding to actual relation $\chi_1^{(\tau)*}, ..., \chi_n^{(\tau)*}$, and the variables $\widetilde{W}_{bN}^{(\tau)}$, $\widetilde{W}_{bN}^{(\tau,me)}$, corresponding to any other relation $\widetilde{\chi}_1^{(\tau)}, ..., \widetilde{\chi}_n^{(\tau)}$. The variables are sums of binary random variables:

$$\begin{aligned} W_{bN}^{(\tau)*} &= \sum_{<i,j>\in R_m} \sum_{k=1}^{N} U_{bk}^{(\tau)*}(x_i, x_j) + \sum_{<i,j>\in R_m} \sum_{k=1}^{N} V_{bk}^{(\tau)*}(x_i, x_j) = \\ & \sum_{<i,j>\in R_m} \sum_{k=1}^{N} \left| g_{bk}^{(\tau)}(x_i, x_j) - T_b^{(\tau)}(x_i, x_j) \right|, \end{aligned} \tag{4.4}$$

$$W_{bN}^{(\tau,me)*} = \sum_{I^{(\tau)*}} U_b^{(\tau,me)*}(x_i,x_j) + \sum_{J^{(\tau)*}} V_b^{(\tau,me)*}(x_i,x_j) = \sum_{<i,j>\in R_m} \left| g_b^{(\tau,me)}(x_i,x_j) - T_b^{(\tau)}(x_i,x_j) \right|, \tag{4.5}$$

where:

$$U_{bk}^{(\tau)*}(x_i,x_j) = \begin{cases} 0 \;\; if \;\; g_{bk}^{(\tau)}(x_i,x_j) = T_b^{(\tau)}(x_i,x_j); \;\; T_b^{(\tau)}(x_i,x_j) = 0; \\ 1 \;\; if \;\; g_{bk}^{(\tau)}(x_i,x_j) \neq T_b^{(\tau)}(x_i,x_j); \;\; T_b^{(\tau)}(x_i,x_j) = 0, \end{cases} \tag{4.6}$$

$$V_{bk}^{(\tau)*}(x_i,x_j) = \begin{cases} 0 \;\; if \;\; g_{bk}^{(\tau)}(x_i,x_j) = T_b^{(\tau)}(x_i,x_j); \;\; T_b^{(\tau)}(x_i,x_j) = 1; \\ 1 \;\; if \;\; g_{bk}^{(\tau)}(x_i,x_j) \neq T_b^{(\tau)}(x_i,x_j); \;\; T_b^{(\tau)}(x_i,x_j) = 1, \end{cases} \tag{4.7}$$

$$U_b^{(\tau,me)*}(x_i,x_j) = \begin{cases} 0 \;\; if \;\; g_b^{(\tau,me)}(x_i,x_j) = T_b^{(\tau)}(x_i,x_j); \;\; T_b^{(\tau)}(x_i,x_j) = 0; \\ 1 \;\; if \;\; g_b^{(\tau,me)}(x_i,x_j) \neq T_b^{(\tau)}(x_i,x_j); \;\; T_b^{(\tau)}(x_i,x_j) = 0, \end{cases} \tag{4.8}$$

$$V_b^{(\tau,me)*}(x_i,x_j) = \begin{cases} 0 \;\; if \;\; g_b^{(\tau,me)}(x_i,x_j) = T_b^{(\tau)}(x_i,x_j); \;\; T_b^{(\tau)}(x_i,x_j) = 1; \\ 1 \;\; if \;\; g_b^{(\tau,me)}(x_i,x_j) \neq T_b^{(\tau)}(x_i,x_j); T_b^{(\tau)}(x_i,x_j) = 1, \end{cases} \tag{4.9}$$

$I^{(\tau)*}$ - the set of pairs $\{<i,j> \;\left|\, T_b^{(\tau)}(x_i,x_j) = 0\right.\}$,
$J^{(\tau)*}$ - the set of pairs $\{<i,j> \;\left|\, T_b^{(\tau)}(x_i,x_j) = 1\right.\}$,
and

$$\widetilde{W}_{bN}^{(\tau)} = \sum_{<i,j>\in I^{(\tau)*}} \sum_{k=1}^{N} \widetilde{U}_{bk}^{(\tau)}(x_i,x_j) + \sum_{<i,j>\in J^{(\tau)*}} \sum_{k=1}^{N} \widetilde{V}_{bk}^{(\tau)}(x_i,x_j) = \sum_{<i,j>\in R_m} \sum_{k=1}^{N} \left| g_{bk}^{(\tau)}(x_i,x_j) - \widetilde{T}_b^{(\tau)}(x_i,x_j) \right|, \tag{4.10}$$

$$\widetilde{W}_{bN}^{(\tau,me)} = \sum_{<i,j>\in \widetilde{I}^{(\tau)}} \widetilde{U}_b^{(\tau,me)}(x_i,x_j) + \sum_{<i,j>\in \widetilde{J}^{(\tau)}} \widetilde{V}_b^{(\tau,me)}(x_i,x_j) = \sum_{<i,j>\in R_m} \left| g_b^{(\tau,me)}(x_i,x_j) - \widetilde{T}_b^{(\tau)}(x_i,x_j) \right| \tag{4.11}$$

where:

$$\widetilde{U}_{bk}^{(\tau)}(x_i, x_j) = \begin{cases} 0 \ \ if \ \ g_{bk}^{(\tau)}(x_i, x_j) = \widetilde{T}_b^{(\tau)}(x_i, x_j); \ \widetilde{T}_b^{(\tau)}(x_i, x_j) = 0; \\ 1 \ \ if \ \ g_{bk}^{(\tau)}(x_i, x_j) \neq \widetilde{T}_b^{(\tau)}(x_i, x_j); \ \widetilde{T}_b^{(\tau)}(x_i, x_j) = 0, \end{cases} \tag{4.12}$$

$$\widetilde{V}_{bk}^{(\tau)}(x_i, x_j) = \begin{cases} 0 \ \ if \ \ g_{bk}^{(\tau)}(x_i, x_j) = \widetilde{T}_b^{(\tau)}(x_i, x_j); \ \widetilde{T}_b^{(\tau)}(x_i, x_j) = 1; \\ 1 \ \ if \ \ g_{bk}^{(\tau)}(x_i, x_j) \neq \widetilde{T}_b^{(\tau)}(x_i, x_j); \ \widetilde{T}_b^{(\tau)}(x_i, x_j) = 1, \end{cases} \tag{4.13}$$

$$\widetilde{U}_{b}^{(\tau,me)}(x_i, x_j) = \begin{cases} 0 \ \ if \ \ g_{b}^{(\tau,me)}(x_i, x_j) = \widetilde{T}_b^{(\tau)}(x_i, x_j); \ \widetilde{T}_b^{(\tau)}(x_i, x_j) = 0; \\ 1 \ \ if \ \ g_{b}^{(\tau,me)}(x_i, x_j) \neq \widetilde{T}_b^{(\tau)}(x_i, x_j); \ \widetilde{T}_b^{(\tau)}(x_i, x_j) = 0, \end{cases} \tag{4.14}$$

$$\widetilde{V}_{b}^{(\tau,me)}(x_i, x_j) = \begin{cases} 0 \ \ if \ \ g_{b}^{(\tau,me)}(x_i, x_j) = \widetilde{T}_b^{(\tau)}(x_i, x_j); \ \widetilde{T}_b^{(\tau)}(x_i, x_j) = 1; \\ 1 \ \ if \ \ g_{b}^{(\tau,me)}(x_i, x_j) \neq \widetilde{T}_b^{(\tau)}(x_i, x_j); \ \widetilde{T}_b^{(\tau)}(x_i, x_j) = 1, \end{cases} \tag{4.15}$$

$\widetilde{I}^{(\tau)}$ - the set of pairs $\{< i, j > \ | \widetilde{T}_b^{(\tau)}(x_i, x_j) = 0\}$,
$\widetilde{J}^{(\tau)}$ - the set of pairs $\{< i, j > \ | \widetilde{T}_b^{(\tau)}(x_i, x_j) = 1\}$.

It can be shown that

Theorem 2

The following relationships hold:

$$E(W_{bN}^{(\tau)*} < \widetilde{W}_{bN}^{(\tau)}) < 0, \tag{4.16}$$

$$E(W_{bN}^{(\tau,me)*} < \widetilde{W}_{bN}^{(\tau,me)}) < 0, \tag{4.17}$$

$$\lim_{N\to\infty} Var(\tfrac{1}{N} W_{bN}^{(\tau)*}) = 0, \tag{4.18}$$

$$\lim_{N\to\infty} Var(W_{bN}^{(\tau,me)*}) = 0, \tag{4.18b}$$

$$\lim_{N\to\infty} Var(\tfrac{1}{N} \widetilde{W}_{bN}^{(\tau)}) = 0 \tag{4.19}$$

$$\lim_{N\to\infty} Var(\widetilde{W}_{bN}^{(\tau,me)}) = 0\,. \tag{4.19b}$$

and:

$$P(W_{bN}^{(\tau)*} < \widetilde{W}_{bN}^{(\tau)}) \geq 1 - \exp\{-2N(\tfrac{1}{2} - \delta)^2\}\,, \tag{4.20}$$

$$P(W_{bN}^{(\tau,me)*} < \widetilde{W}_{bN}^{(\tau,me)}) \geq 1 - 2\exp\{\tfrac{1}{2} - \delta)^2\}\,. \tag{4.21}$$

The proof of (4.16) – (4.21) is the same as in the case of the equivalence relation (see also Klukowski, 2002, 2011a).

The interpretation of the relationships (4.16) – (4.21) is similar to the case of the equivalence relation.

The tasks (4.1), (4.2) can be solved with the use of the algorithms similar to those presented in Chapter 2. The tests for validation of the estimates are presented in Chapter 10.

An example of application of the approach is presented in Klukowski (2008c). It presents overlapping classification of shapes of functions expressing profitability of treasury bonds.

4.4. Summary

The estimators of the equivalence relation have good statistical properties and a simple form – similarly as in the case of equivalence relation. Moreover, there exist efficient algorithms for solving the respective optimization tasks.

The next chapter presents the tests for detecting the relation type. They are useful in the case of unknown relation type (overlapping or non-overlapping) and comparisons indicating equivalence or non-equivalence of elements in a pair. In other words, the test recognizes the correct model of data.

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