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 7

Estimation of the preference relation – binary comparisons

7.1. Introduction

This chapter presents the problem of estimation of the preference relation on the basis of binary comparisons. They can assume three values $\{0, \pm 1\}$, but the form of estimators is similar to those constructed for the equivalence and tolerance relations.

7.2. Assumptions about binary comparisons

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

$$g_{bk}^{(p)}(x_i, x_j) = \begin{cases} -1\ \textit{if}\ \ k - th\ comparison\ indicates\ x_i \in \chi_r^{(p)*}, x_j \in \chi_s^{(p)*}, r < s; \\ 0\ \text{ if } k - \text{th comparison indicates that } (x_i, x_j)\ are\ included \\ \quad in\ the\ same\ subset\ \chi_q^{(\tau)*}; \\ 1\ \textit{if}\ \ k - th\ comparison\ indicates\ x_j \in \chi_r^{(p)*}, x_i \in \chi_s^{(p)*}, r < s. \end{cases} \qquad (7.1)$$

The comparisons $g_{bk}^{(p)}(x_i, x_j)$ have to satisfy the assumptions A1, A2 and A3 (Chapter 2), i.e. the probability of correct comparison, $1-\delta$, has to be higher that the sum of probabilities of incorrect results. In the case of equivalent elements, it is sufficient to assume that zero is the mode and median of comparisons errors, i.e. the same as in the case of multivalent comparisons. The assumption about independence of comparisons can be relaxed in such way that comparisons of different pairs, i.e. $g_{bk}^{(p)}(x_i,x_j)$ and $g_{bk}^{(p)}(x_r,x_s)$ $(i \neq r,s;\ j \neq r,s)$ are independent (see Klukowski, 1994).

7.3. The form of estimators and their properties

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

$$\min_{F_{\mathbf{X}}^{(p)}}\{\sum_{<i,j>\in R_m}\sum_{k=1}^{N}\left|g_{bk}^{(p)}(x_i,x_j)-t_b^{(p)}(x_i,x_j)\right|\}, \tag{7.2}$$

where:

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

$t_b^{(p)}(x_i,x_j)$ - the function describing a relation from the set $F_{\mathbf{X}}^{(p)}$,

and

$$\min_{F_X^{(\tau)}}\{\sum_{<i,j>\in R_m}\left|g_b^{(p,me)}(x_i,x_j)-t_b^{(p)}(x_i,x_j)\right|\}, \tag{7.3}$$

where:

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

The differences $g_{bk}^{(p)}(x_i,x_j)-t_b^{(p)}(x_i,x_j)$ and $g_b^{(p,me)}(x_i,x_j)-t_b^{(p,me)}(x_i,x_j)$ can assume five values, i.e. {-2, -1, 0, 1, 2}, and generate more complicated optimization tasks than the binary tasks. Therefore, it is rational to transform these variables into the binary form - in the following way:

$$\Theta(g_{bk}^{(p)}(x_i,x_j)-t_b^{(p)}(x_i,x_j))=\begin{cases}0\ if\ g_{bk}^{(p)}(x_i,x_j)=t_b^{(p)}(x_i,x_j);\\1\ if\ g_{bk}^{(p)}(x_i,x_j)\neq t_b^{(p)}(x_i,x_j),\end{cases} \tag{7.4a}$$

$$\Theta(g_b^{(p,me)}(x_i,x_j)-t_b^{(p)}(x_i,x_j))=\begin{cases}0\ if\ g_b^{(p,me)}(x_i,x_j)=t_b^{(p)}(x_i,x_j);\\1\ if\ g_b^{(p,me)}(x_i,x_j)\neq t_b^{(p)}(x_i,x_j).\end{cases} \tag{7.4b}$$

In this case the estimators $\hat{\chi}_1^{(p)},...,\hat{\chi}_n^{(p)}$ and $\widetilde{\chi}_1^{(p)},...,\widetilde{\chi}_n^{(p)}$ are obtained on the basis of the binary problems:

$$\min_{F_{\mathbf{X}}^{(p)}} \{ \sum_{<i,j>\in R_m} \sum_{k=1}^{N} \Theta(g_{bk}^{(p)}(x_i, x_j) - t_b^{(p)}(x_i, x_j)) \}, \tag{7.5}$$

$$\min_{F_X^{(\tau)}} \{ \sum_{<i,j>\in R_m} \Theta(g_b^{(p,me)}(x_i, x_j) - t_b^{(p)}(x_i, x_j)) \}. \tag{7.6}$$

The properties of the estimators based on original and transformed comparisons are similar (Klukowski, 1990b).

The properties of the estimators (7.5), (7.6) are based on properties of the random variables:

$$W_{bN}^{(p)*} = \sum_{<i,j>\in R_m} \sum_{k=1}^{N} \Theta(g_{bk}^{(p)}(x_i, x_j) - T_b^{(p)}(x_i, x_j)), \tag{7.7}$$

$$W_{bN}^{(p,me)*} = \sum_{<i,j>\in R_m} \Theta(g_b^{(p,me)}(x_i, x_j) - T_b^{(p)}(x_i, x_j)), \tag{7.8}$$

corresponding to the actual relation, $\chi_1^{(p)*}, ..., \chi_n^{(p)*}$, and the random variables:

$$\widetilde{W}_{bN}^{(p)} = \sum_{<i,j>\in R_m} \sum_{k=1}^{N} \Theta(g_{bk}^{(p)}(x_i, x_j) - \widetilde{T}_b^{(p)}(x_i, x_j)), \tag{7.9}$$

$$\widetilde{W}_{bN}^{(p,me)} = \sum_{<i,j>\in R_m} \Theta(g_b^{(p,me)}(x_i, x_j) - \widetilde{T}_b^{(p)}(x_i, x_j)), \tag{7.10}$$

corresponding to a relation $\widetilde{\chi}_1^{(p)}, ..., \widetilde{\chi}_n^{(p)}$, different from the actual one.
It can be shown (Klukowski, 1994), that

Theorem 4

The following relationships are true:

$$E(W_{bN}^{(p)*} < \widetilde{W}_{bN}^{(p)}) < 0, \tag{7.11}$$

$$E(W_{bN}^{(p,me)*} < \widetilde{W}_{bN}^{(p,me)}) < 0, \tag{7.12}$$

$$\lim_{N\to\infty} Var(\tfrac{1}{N} W_{bN}^{(p)*}) = 0, \tag{7.13}$$

$$\lim_{N\to\infty} Var(\tfrac{1}{N}\widetilde{W}_{bN}^{(p)}) = 0\,, \tag{7.14}$$

$$\lim_{N\to\infty} Var(W_{bN}^{(p,me)*}) = 0\,, \tag{7.15}$$

$$\lim_{N\to\infty} Var(\widetilde{W}_{bN}^{(p,me)}) = 0\,. \tag{7.16}$$

The probabilities $P(W_{bN}^{(p)*} < \widetilde{W}_{bN}^{(p)})$, $P(W_{bN}^{(p,me)*} < \widetilde{W}_{bN}^{(p,me)})$ satisfy the inequalities:

$$P(W_{bN}^{(p)*} < \widetilde{W}_{bN}^{(p)}) \geq 1 - \exp\{-2N(\tfrac{1}{2}-\delta)^2\}\,, \tag{7.17}$$

$$P(W_{bN}^{(p,me)*} < \widetilde{W}_{bN}^{(p,me)}) \geq 1 - 2\exp\{-2N(\tfrac{1}{2}-\delta)^2\}\,. \tag{7.18}$$

The proofs of relationships (7.11) – (7.18), based on Hoeffding inequality and Chebyshev inequality for expected value, are given in Klukowski (1994).

The properties expressed by relationships (7.11) – (7.18), which constitute the basis for estimators (7.5), (7.6), are similar to the case of the equivalence and tolerance relations.

The existence of distributions of both estimators (7.5), (7.6) can be proven in the same way as in Chapter 2.

The tasks (7.5), (7.6) can be solved using the binary algorithms discussed in Chapter 2. Validation of estimates obtained is discussed in Chapter 10.

The properties of estimates obtained with the use of simulation are presented in Chapter 9; they allow for determining the parameters, especially the number of comparisons *N*, guaranteeing the probability of errorless estimates close to one.

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 the additional result of comparison, corresponding to incomparable elements.

7.4 Summary

The estimators of the preference relation have similar form and properties as the estimators of the equivalence and tolerance relations for binary comparisons. The properties of estimates for binary and multivalent comparisons have been examined using simulation; they are discussed in Chapter 9. Estimates obtained on the basis of binary comparisons can be used in two-stage estimation. In the first step it is necessary to determine N estimates of the relation form, in the next step we determine the differences of ranks and finally we apply the estimator based on multiple multivalent comparisons, presented in Chapter 8. Simulation survey shows good efficiency of multivalent estimators and, therefore, the second stage can improve the initial binary estimates.

A significant advantage of efficiency of the estimator based on the sum of differences should be emphasized. The number of comparisons guaranteeing good precision of estimates is moderate, from 3 to 7; the median estimator requires the number of comparisons increased by two.

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