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

Preface

A pairwise comparison states equivalence or non-equivalence of two elements taken from some (finite) set, or indicates preference within a pair; the result of comparison can be disturbed by a random error. Such comparisons can be obtained with the use of statistical tests, on the basis of expert opinions or through other decision procedures.

The methods of paired comparisons are used for extracting information from data. They are applied in many fields of application: economy, finance, marketing, medicine, psychometrics, sport, and others. The results obtained in this area concern mainly the ranking of elements from some set (David, 1988, Bradley, 1976, 1984, Brunk, 1960, Davidson, 1976). These results are obtained with the use of quite restrictive assumptions.

The present work develops and broadens the approach in the area. The following directions are of special importance:

- classification problems – with non-overlapping and overlapping subsets, on the basis on binary (qualitative) and multivalent (quantitative) comparisons;
- relaxation of assumptions about random errors of pairwise comparisons in ranking and classification problems; this is especially important from the practical point of view;
- development of the homogenous concept of estimation of all types of relations considered - based on the discrete programming tasks; two types of estimators have been defined – appropriate for typical requirements of users; they allow also for combination of different kinds of comparisons (binary and multivalent);
- development of the properties of the estimators – both analytical and based on simulation approach; the main property is consistency of estimates; the precision of estimates can be determined also in the case of unknown distributions of comparison errors;
- development of the statistical tools (tests), allowing for validation of estimates, especially the original tests for determination of relation type – equivalence or tolerance;

- indication of the algorithms for solving the optimization tasks, necessary for obtaining the estimates; this comprises exact algorithms and different types of approximate algorithms.

The methods presented in the book are based on the idea of the *NEAREST ADJOINING ORDER* (NAO), introduced by Slater (1961), and summarize the results obtained by the author in the domain. The NAO methods are classified among the combinatorial algorithms (see e.g. David, 1988, Ch. 2, Koronacki, Mielniczuk, 2005, Ch.9). The main idea of the approach is to determine an estimate featuring the minimum of inconsistencies with a sample (comparisons). The estimates are obtained on the basis of discrete optimization tasks, and they usually involve a significant computation cost.

The results of the work have been achieved within the project *Methods of estimation of relations of: preference, equivalence, and tolerance*, financed by the Ministry of Science and Higher Education, No N N111 434937, carried out at the Systems Research Institute Polish Academy of Sciences, in the years 2009 - 2011. The papers on this subject have been published during the work at the Systems Research Institute in the years 2006 – 2011 and at the Ministry of Finance of the Republic of Poland in the years 1993 - 2006.

The work is intended for scholars interested in ranking and classification problems, especially based on pairwise statistical tests, expert opinions and other procedures, which are prone to generating random errors. The estimates have good statistical properties, not attainable in other approaches.

The work can be useful for researchers and practitioners having basic knowledge in statistics, simulation techniques and discrete optimization. The work is oriented on statistical and data mining approaches; i.e. algorithms of estimation, identification and validation realized entirely with a computer. A user of such results has to perform the interpretation of the results obtained.

The work presents main theoretical results. The proofs of theorems are presented in the articles; some ideas are shown in the Appendices. They are based on the original ideas of the author, known probabilistic inequalities and properties of order statistics. Some of the proofs are analytically complicated and cumbersome. Examples of applications are described in journal papers and conference papers (see Bibliography).

The author is grateful to the scientists who promoted the idea of the work and have made fruitful comments.

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