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 10

Tests for validation of estimates obtained on the basis of pairwise comparisons with random errors

10.1. Introduction

The estimators of relations of preference, equivalence and tolerance, presented in Chapters 3, 5, 6, 7, 8, and the test for the detection of relation type, presented in Chapter 4, are based on assumptions about distributions of pairwise comparisons and the crucial assumption – about the existence of the relation. The first group of assumptions is of probabilistic nature and can be verified with the use of known statistical tests. Verification of existence of the relation allows for rejection of the cases, when comparisons have been generated in a random way or elements of the set are equivalent or some other data structure characterizes them. The positive results of verification guarantee reliable estimates, negative ones indicate the direction of improvement or suggest rejection. Some of the tests proposed require having simulation results.

The problems of validation are discussed in the literature, e.g. for the preference relation - in David (1988), for the equivalence and tolerance relations - in Gordon (1999).

The tests useful for validation of assumptions about comparison errors are presented in statistical monographs, e.g. Sheskin (1997), Daniel (1990), Siegel, Castellan (1988), Domański (1979, 1990).

10.2 Verification of assumptions about comparison errors

The following assumptions are made concerning the comparison errors: • independence of the whole set of comparisons, • unimodality of distributions of comparison errors, • mode and median of the distributions of errors equal zero.

The independence of the whole set of comparisons $g_{\upsilon k}^{(\ell)}(x_i, x_j)$ is difficult to verify, because it requires multivariate tests. Construction of such tests is possible for large *N* only. Therefore, multivariate verification can be done, in fact, only for binary comparisons, in the case of identical distributions of all comparisons. Such a case is not typical in practice. For low or moderate *N*, it is admissible to verify independence of comparisons of individual pairs $(x_i, x_j) \quad (<i, j> \in R_m)$. Usually, it can be assumed that independence of comparisons of the same pair indicates independence required by the estimators proposed.

The independence of comparisons of the same pair, i.e., $g_{\upsilon,1}^{(\ell)}(x_i, x_j), ..., g_{\upsilon N}^{(\ell)}(x_i, x_j) \quad (<i, j> \in R_m)$ can be verified with the use of tests of randomness of a sample. The null hypothesis states randomness, the alternative one – non-randomness; the tests do not require the estimates $\hat{T}_{\upsilon}^{(\ell)}(x_i, x_j)$ or $\widetilde{T}_{\upsilon}^{(\ell)}(x_i, x_j)$. Some tests for verification of independence require the assumption about identical distribution of comparisons of an individual pair. The assumption can be tested with the use of well-known tests for homogeneity of a sample.

The hypotheses of unimodality and zero value of mode and median have to be verified with the use of different tests in the case of binary and multivalent comparisons.

In the case of binary comparisons the hypotheses mentioned are equivalent to the fact that the probability $1-\delta$ of correct comparison, $\{g_{bk}^{(\ell)}(x_i, x_j) = T_b^{(\ell)}(x_i, x_j)\} \quad (\ell \in \{e, \tau, p\};\ <i, j> \in R_m;\ k = 1, ..., N)$ is greater than ½. The basis for the hypotheses is constituted by the differences $\hat{T}_b^{(\ell)}(x_i, x_j) - g_{bk}^{(\ell)}(x_i, x_j)$ or $\widetilde{T}_b^{(\ell)}(x_i, x_j) - g_{bk}^{(\ell)}(x_i, x_j)$. The results of verification are valid in the case $\hat{T}_b^{(\ell)}(x_i, x_j) = T_b^{(\ell)}(x_i, x_j)$ or $\hat{T}_b^{(\ell)}(x_i, x_j) = T_b^{(\ell)}(x_i, x_j)$ and, therefore, the probabilities of errors in tests have to be corrected by the probability of errorless estimate, i.e. $P(\hat{\chi}_1^{(\ell)}, ..., \hat{\chi}_{\hat{n}}^{(\ell)} \equiv \chi_1^{*(\ell)}, ..., \chi_n^{*(\ell)})$ or $P(\widetilde{\chi}_1^{(\ell)}, ..., \widetilde{\chi}_{\hat{n}}^{(\ell)} \equiv \chi_1^{*(\ell)}, ..., \chi_n^{*(\ell)})$. The hypothesis under consideration can be verified on the basis of the binomial distribution. The null and alternative hypotheses assume the form, respectively: H_0: $1-\delta \leq ½$ and H_1: $1-\delta >½$. Rejection of the null hypothesis for all pairs (x_i, x_j) $(<i,j>\in R_m)$ does confirm the assumptions about the probability $1-\delta$. The unimodality of the distributions can be also verified on the basis of comparisons $g_{b,1}^{(\ell)}(x_i,x_j),..., g_{bN}^{(\ell)}(x_i,x_j)$ $(<i,j>\in R_m)$, i.e. without estimates $\hat{T}_b^{(\ell)}(\cdot)$ or $\widetilde{T}_b^{(\ell)}(\cdot)$. It is obvious that some number of opposite results (errors in tests) can occur too; the fraction of such results depends on the probabilities of errors in tests and the probability of obtaining the errorless estimate.

The case of multivalent comparisons is more complex, because the distributions of comparison errors are not determined, in general, by one parameter. The hypotheses about unimodality and values of mode and median equal zero have to be verified with the use of different tests.

Unimodality of distributions of comparison errors $g_{\mu k}^{(\ell)}(x_i,x_j)-T_{\mu}^{(\ell)}(x_i,x_j)$ $(\ell\in\{p,\tau\};\ <i,j>\in R_m;\ k=1,...,N)$ can be verified on the basis of differences $g_{\mu k}^{(\ell)}(x_i,x_j)-\hat{T}_{\mu}^{(\ell)}(x_i,x_j)$ or $g_{\mu k}^{(\ell)}(x_i,x_j)-\widetilde{T}_{\mu}^{(\ell)}(x_i,x_j)$ $(k=1,...,N)$. The results of verification have to be corrected by the probability of the errorless estimate. Tests for this purpose are based on multinomial distribution, statistics of such tests have asymptotic Gaussian distribution and verify, simultaneously, the value of mode (Domański, 1990). The hypothesis of unimodality of distribution can also be verified on the basis of comparisons $g_{\mu,1}^{(\ell)}(x_i,x_j)$, …, $g_{\mu N}^{(\ell)}(x_i,x_j)$ only.

The hypothesis that the median of comparison errors is equal zero can be verified a similar way.

10.3. Verification of existence of relation

Verification of existence of the relation (equivalence or preference or tolerance) in the set **X** has to be done after the positive results of tests for verification of properties of distributions of comparison errors. The null hypothesis H_0 assumes the form: the relation exists in the set **X**, the alternative H_1 – the relation does not exist in the set **X**. Non-existence of the relation means: randomness of comparisons or equivalence of elements of the set or existence of some other data structure. The basis for verification is constituted by the estimates ($\hat{T}_b^{(\ell)}(x_i,x_j)$ or $\widetilde{T}_b^{(\ell)}(x_i,x_j)$) of the relation and differences between the estimates and comparisons. The simulation approach is a useful tool for construction of tests.

The case of estimators based on binary comparisons is simpler than the one of multivalent comparisons; the simulation approach is also easier and more efficient in this case.

The basis for verification of null hypothesis are the sums of differences between estimates and comparisons, i.e.:

$$\frac{1}{N}\sum_{<i,j>\in R_m}\sum_{k=1}^{N}\left|\hat{T}_b^{(\ell)}(x_i,x_j)-g_{bk}^{(\ell)}(x_i,x_j)\right| \text{ or}$$

$$\frac{1}{N}\sum_{<i,j>\in R_m}\sum_{k=1}^{N}\left|\overset{\frown}{T}_b^{(\ell)}(x_i,x_j)-g_{bk}^{(\ell)}(x_i,x_j)\right|. \qquad (10.1)$$

In the case of known δ it is possible to determine the evaluation of the expected value and the variance of the sums (10.1) (in a similar way, as in Klukowski, 2008c, 2010b). The evaluations of expected value and variance of the first sum assume, in the case of errorless estimate and comparisons with values from the set $\{0, 1\}$, the form:

$$E\left(\frac{1}{N}\left(\sum_{<i,j>\in R_m}\sum_{k=1}^{N}\left|\hat{T}_b^{(\ell)}(x_i,x_j)-g_{bk}^{(\ell)}(x_i,x_j)\right|\right)\right)\le\frac{1}{2}m(m-1)\delta, \qquad (10.2)$$

$$E\left(\frac{1}{N}\left(\sum_{<i,j>\in R_m}\sum_{k=1}^{N}\left|\overset{\frown}{T}_b^{(\ell)}(x_i,x_j)-g_{bk}^{(\ell)}(x_i,x_j)\right|\right)\right)\le\frac{1}{2}m(m-1)\delta, \qquad (10.3)$$

$$Var\left(\frac{1}{N}\left(\sum_{<i,j>\in R_m}\sum_{k=1}^{N}\left|\hat{T}_b^{(\ell)}(x_i,x_j)-g_{bk}^{(\ell)}(x_i,x_j)\right|\right)\right)\le\frac{1}{2N}m(m-1)\delta(1-\delta), \qquad (10.4)$$

$$Var\left(\frac{1}{N}\left(\sum_{<i,j>\in R_m}\sum_{k=1}^{N}\left|\overset{\frown}{T}_b^{(\ell)}(x_i,x_j)-g_{bk}^{(\ell)}(x_i,x_j)\right|\right)\right)\le\frac{1}{2N}m(m-1)\delta(1-\delta). \qquad (10.5)$$

The case, when the assumption A3 is not satisfied, i.e. covariances between comparisons $g_{bk}^{(\ell)}(x_i,x_j)$, $g_{bk}^{(\ell)}(x_j,x_q)$ $(q\neq i)$ can be different from zero, is discussed in Klukowski (1990, 1994, 2002). The analysis is based on the fact that the values of such covariances are lower than $\delta(1-\delta)$ and the number of non-zero covariances can be determined (Klukowski, 1994, Lemma 2). Let us note that covariances can assume also negative values.

The evaluations (10.2), (10.3) allow for the construction of a test based of the Chebyshev inequality for expected value and variance.

The test is constructed in a similar way as in Klukowski (2008c, 2010b); the test statistic assumes, in the case of the estimator based on sums, the form:

$$\frac{1}{N}\Big(\sum_{<i,j>\in R_m}\sum_{k=1}^{N}\left|\hat{T}_b^{(\ell)}(x_i,x_j)-g_{bk}^{(\ell)}(x_i,x_j)\right|\Big), \tag{10.6}$$

i.e. it expresses the differences between an estimate and the comparisons. The test rejects the hypothesis about existence of the relation in the case of large (excessive) values of the statistic (10.6). The probabilities of errors in the test have to be corrected by the probability of errorless estimate.

The probability function of the sum (10.6) can be determined with the use of simulation approach (see Chapter 9). The basis for the approach is constituted by the estimates $\hat{T}_b^{(\ell)}(x_i,x_j)$ or $\widehat{T}_b^{(\ell)}(x_i,x_j)$ and the value of δ. The distributions allow for the construction of one-sided critical regions; therefore, tests based on simulation approach can guarantee lower probabilities of errors than those based on inequalities (10.2) - (10.5) and the Chebyshev inequality.

The case of multivalent comparisons is more complicated, because distributions of comparison errors are usually unknown; a precise estimation of these distributions is unrealizable in typical situations. Hence, it is suggested to replace the unknown distributions by the "quasi-uniform" distributions (see, e.g., Klukowski, 2007b, 2008b). In the case of the preference relation they are defined by the conditions:

- the case of $\hat{T}_\mu^{(p)}(x_i,x_j)\neq\pm(\hat{n}-1)$:

$$P(\hat{T}_{\mu}^{(p)}(x_i,x_j)-g_{\mu k}^{(p)}(x_i,x_j)<0)=P(\hat{T}_{\mu}^{(p)}(x_i,x_j)-g_{\mu k}^{(p)}(x_i,x_j)>0),$$
$$P(\hat{T}_{\mu}^{(p)}(x_i,x_j)-g_{\mu k}^{(p)}(x_i,x_j)=-1)=...=P(\hat{T}_{\mu}^{(p)}(x_i,x_j)-(\hat{n}-1)),$$
$$P(\hat{T}_{\mu}^{(p)}(x_i,x_j)-g_{\mu k}^{(p)}(x_i,x_j)=1)=...=P(\hat{T}_{\mu}^{(p)}(x_i,x_j)+(\hat{n}-1)),$$
$$P(\hat{T}_{\mu}^{(p)}(x_i,x_j)-g_{\mu k}^{(p)}(x_i,x_j)=0)=$$
$$=\max\{P(T_{\mu}^{(p)}(x_i,x_j)-g_{\mu k}^{(p)}(x_i,x_j)=1),P(\hat{T}_{\mu}^{(p)}(x_i,x_j)-g_{\mu k}^{(p)}(x_i,x_j)=-1)\},$$
$$\sum_{l=\hat{T}_{\mu}^{(p)}(\cdot)-(\hat{n}-1)}^{\hat{T}_{\mu}^{(p)}(\cdot)+(\hat{n}-1)} P(\hat{T}_{\mu}^{(p)}(x_i,x_j)-g_{\mu k}^{(p)}(x_i,x_j)=l)=1, \tag{10.7}$$

• the case of $\hat{T}_{\mu}^{(p)}(x_i,x_j)=\pm(\hat{n}-1)$:

$$P(\hat{T}_{\mu}^{(p)}(x_i,x_j)-g_{\mu k}^{(p)}(x_i,x_j)=0)=\tfrac{1}{2}+\varepsilon \quad (\varepsilon\in(0,\tfrac{1}{2})),$$
$$P(\hat{T}_{\mu}^{(p)}(x_i,x_j)-g_{\mu k}^{(p)}(x_i,x_j)=\pm l)=\tfrac{1-2\varepsilon}{4(\hat{n}-1)} \quad (l=1,...,2(\hat{n}-1)). \tag{10.8}$$

The distributions defined by (10.7), (10.8) satisfies the assumptions A1, A2 and has the following features: • quasi-symmetry, i.e. equal values of left and right tie, • equality of each probability in the left and right tie, • the variance close to maximal in the set of possible distributions of errors. The distributions defined by (10.7), (10.8) can be used as a basis for simulation of pairwise comparisons (similarly as in Chapter 9) and then for obtaining the distributions of sums:

$$\frac{1}{N}\sum_{<i,j>\in R_m}\sum_{k=1}^{N}\left|\hat{T}_{\mu}^{(p)}(x_i,x_j)-g_{\mu k}^{(p)}(x_i,x_j)\right| \text{ or } \sum_{<i,j>\in R_m}\sum_{k=1}^{N}\left|\widehat{T}_{\mu}^{(p)}(x_i,x_j)-g_{\mu k}^{(p)}(x_i,x_j)\right|. \tag{10.9}$$

The tests based on the distributions of sums (10.9) can be efficient in the case of good precision of estimation, i.e. the values of criterion functions close to zero. In the opposite case the validation ought to be completed with tests based on other properties of estimates.

The examples of such properties are: • positive correlation of comparisons $g_{\upsilon k}^{(\ell)}(x_i,x_r)$, $g_{\upsilon k}^{(\ell)}(x_j,x_r)$ $(x_i,x_j\in\hat{\chi}_q;\ 1\le q\le\hat{n};\ 1\le k\le N;\ r=1,...,m;)$, i.e. elements belonging to the same subset $\hat{\chi}_q^{(\ell)}$ and • positive correlation of comparisons $g_{\upsilon k}^{(\ell)}(x_i, x_r), ..., g_{\upsilon k}^{(\ell)}(x_i, x_r)$ and $g_{\upsilon l}^{(\ell)}(x_i, x_r), ..., g_{\upsilon l}^{(\ell)}(x_i, x_r)$ $(l \neq k;\ \ 1 \leq i \leq m;\ \ r = 1, ..., m)$. The null hypotheses H_0 state the lack of correlation (i.e. randomness of comparisons), the alternative ones, H_1 – positive correlation, close to one. Rejecting null hypotheses for all correlations validates the estimates of the relations. However, some fraction of non-positive correlations can occur, too. It is clear that the approach based on sums (10.9) verifies the "global" features of estimates, while those based on correlations – the "partial" features, i.e. corresponding to individual subsets of comparisons.

In the case of the preference relation, especially of $\hat{n} = m$, the estimates of ranks of individual elements can be applied for verification of existence of relation. In particular, the ranks of elements, obtained on the basis of individual subsets of comparisons $g_{\upsilon k}^{(p)}(x_i, x_j)$ and $g_{\upsilon l}^{(p)}(x_i, x_j)$ $(\upsilon \in \{b, \mu\},\ k \neq l;\ <i, j> \in R_m)$ have to be positively correlated. Denoting by $\hat{r}_{ik}^{(p)}$ $(i = 1, ..., m; 1 \leq k \leq N)$ the rank of the *i*-th element, obtained on the basis of the *k*-th subset of comparisons $g_{\upsilon k}^{(p)}(x_i, x_j)$ $(1 \leq k \leq N; <i, j> \in R_m)$ one can verify the hypotheses about positive correlation of ranks: $\hat{r}_{1,k}^{(p)}, \ldots, \hat{r}_{mk}^{(p)}$ and $\hat{r}_{1,l}^{(p)}, \ldots, \hat{r}_{ml}^{(p)}$ $(l \neq k)$. Moreover, it is also possible to verify the hypothesis about positive correlation (concordance) of the matrix of all ranks (see also Raghavachari, 2005):

$$\mathfrak{R}' = \begin{bmatrix} \hat{r}_{11}^{(p)} \ \cdots \ \hat{r}_{1,N}^{(p)} \\ \cdots \\ \hat{r}_{m,1}^{(p)} \ \cdots \ \hat{r}_{mN}^{(p)}. \end{bmatrix}, \tag{10.10}$$

where: $\boldsymbol{\Theta}'$ - matrix $\boldsymbol{\Theta}$ transposed.

The test for the concordance hypothesis is based on statistic *W* or *S*, defined as follows:

$$W = \frac{12S}{m^2(N^3 - N)}, \tag{10.11}$$

where:

$$S = \sum_{k=1}^{N}(R_k - \sum_{k=1}^{N} R_k \Big/ N)^2, \tag{10.12}$$

$R_k = \sum_{i=1}^{m} \hat{r}_{ik}$.

The null hypothesis H_0 states randomness, the alternative one, H_1, positive correlation. Rejection of the null hypotheses, especially in the case of correlation close to one, confirms the existence of the relation. The first statistic, (10.11), is appropriate for $m>7$, critical values are determined on the basis of (asymptotic) chi-square distribution. The second one, (10.12), is appropriate for $3 \le m \le 7$ and $3 \le p \le 20$; critical values are presented, e.g. in Siegel and Castellan (1988) (multivariate Kendall's rank test).

The tests described above can be also applied in the case of multiple estimates (i.e. multiple solutions of the optimization problems); they allow for detection of the best estimate instead of random selection.
Positive results of all tests applied in validation process confirm the existence of the relation, negative results - the opposite conclusion.

It is clear that the set of possible tests, useful in validation of estimates can be broadened; some of the tests presented in David (1988) and Gordon (1999) can also be used.

10.4. Tests for weak and strict form of the preference relation

Sometimes, knowledge of some features of a relation estimated may be incomplete; a typical problem of that type is distinguishing between weak and strict form of the preference relation. In particular, the strict form of the relation allows for selecting the best element of the set. Moreover, it suggests the application of an appropriate, multivalent, estimator. The test for verification to be used for this purpose is presented below.

The idea of a test for distinguishing between the strict and weak forms of the preference relation is close to those applied to equivalence and tolerance relations in the case of binary comparisons (Chapter 5).

The test statistic is a function of inconsistencies (differences) between the comparisons $g_{bk}^{(p)}(x_i, x_j)$ and estimates $\hat{T}_b^{(pw)}(x_i, x_j)$, $\hat{T}_b^{(ps)}(x_i, x_j)$ or $\widehat{T}_b^{(pw)}(x_i, x_j)$, $\widehat{T}_b^{(ps)}(x_i, x_j)$ - expressing, respectively, the weak ($\hat{T}_b^{(pw)}(x_i, x_j)$, $\widehat{T}_b^{(ps)}(x_i, x_j)$) and the strict ($\hat{T}_b^{(ps)}(x_i, x_j)$, $\widehat{T}_b^{(ps)}(x_i, x_j)$) form of the relation. The strict form is obtained on the basis of the tasks (7.2) or (7.3), with the feasible set satisfying the condition $n = m$. The weak form is obtained on the basis of the task with the feasible set satisfying the condition $n \le m-1$. The differences are determined for the set

$\hat{I}_p = \{<i,j> \mid \hat{T}_b^{(pw)}(x_i,x_j) \ne \hat{T}_b^{(ps)}(x_i,x_j)\}$ or

$\widehat{I}_p = \{<i,j> \mid \widehat{T}_b^{(pw)}(x_i,x_j) \ne \widehat{T}_b^{(ps)}(x_i,x_j)\}$.

The test statistic $S_N^{(\gamma)}$, corresponding to the estimates $\hat{T}_b^{(pw)}(x_i,x_j)$, assumes the form:

$$S_N^{(\gamma)} = \frac{1}{N}\frac{1}{\#(\hat{I}_p)} \sum_{<i,j>\in\hat{I}_p} \sum_{k=1}^{N} \eta(\hat{S}_{ijk}^{(\gamma)}), \tag{10.13}$$

where:

$$\hat{S}_{ijk}^{(\gamma)} = \left|\hat{T}_b^{(pw)}(x_i,x_j) - g_{bk}^{(p)}(x_i,x_j)\right| - \left|\hat{T}_b^{(ps)}(x_i,x_j) - g_{bk}^{(p)}(x_i,x_j)\right|$$

$$(1 \le k \le N;\ <i,j>\in \hat{I}_p), \tag{10.14}$$

and:

γ - index corresponding to the actual relation type – weak *w* or strict *s*,

$\#(\hat{I}_p)$ – the number of elements of the set $\hat{I}_p$,

$\eta(\hat{S}_{ijk}^{(\gamma)})$ - the function defined as follows:

$$\eta(\hat{S}_{ijk}^{(\gamma)}) = \begin{cases} -1 & if\ \hat{S}_{ijk}^{(\gamma)} \le 0; \\ 1 & if\ \hat{S}_{ijk}^{(\gamma)} > 0. \end{cases}$$

The properties of the statistics $S_N^{(\gamma)}$ depend on the actual relation type in the set **X**. In the case of weak form of the relation $\chi_1^{*(pw)},...,\chi_{n_w}^{*(pw)}$ and the errorless result of estimation, i.e. $\hat{\chi}_1^{(pw)},...,\hat{\chi}_{\hat{n}_w}^{(pw)} \equiv \chi_1^{*(pw)},...,\chi_{n_w}^{*(pw)}$, the equalities $\hat{T}_b^{(pw)}(x_i,x_j) = T_b^{(pw)}(x_i,x_j)$ $(<i,j>\in I_w)$ hold; the probability of such event is denoted $P(\hat{\chi}_1^{(pw)},...,\hat{\chi}_{\hat{n}_w}^{(pw)} \equiv \chi_1^{*(pw)},...,\chi_{n_w}^{*(pw)} \mid \mathbf{R}^{(pw)})$. In the case of the strict form of relation $\chi_1^{*(ps)},...,\chi_{n_s}^{*(ps)}$ and the errorless result of estimation, i.e. $\hat{\chi}_1^{(ps)},...,\hat{\chi}_{\hat{n}_s}^{(ps)} \equiv \chi_1^{*(ps)},...,\chi_{n_s}^{*(ps)}$, the equalities $\hat{T}_b^{(ps)}(x_i,x_j) = T_b^{(ps)}(x_i,x_j)$ $(<i,j> \in \hat{I}_p)$ hold; the probability of such event is denoted $P(\hat{\chi}_1^{(ps)},...,\hat{\chi}_{\hat{n}_s}^{(ps)} \equiv \chi_1^{*(ps)},...,\chi_{n_s}^{*(ps)} \mid \mathbf{R}^{(ps)})$. The variables $\eta(\hat{S}_{ijk}^{(\gamma)})$ $(<i,j> \in \hat{I}_p)$ assume values from the set $\{-1,1\}$. It can be shown, in the same way as in Klukowski (2006), that in the case of weak form of relation, the expected value and variance of the variables $\hat{S}_{ijk}^{(\gamma)}$ satisfy the inequalities:

$$E(\hat{S}_{ijk}^{(w)}) \le -1 + 2\delta, \tag{10.15}$$

$$Var(\hat{S}_{ijk}^{(w)}) \le 4\delta(1-\delta). \tag{10.16}$$

The expected value and the variance of the variable $S_N^{(w)}$ fulfill the relationships:

$$E(S_N^{(w)}) \le -1 + 2\delta, \tag{10.17}$$

$$Var(S_N^{(w)}) \le \tfrac{4}{N}(1 - 2L(\hat{I}_p)/(\#(\hat{I}_p))^2)\delta(1-\delta), \tag{10.18}$$

where:

$L(\hat{I}_p)$ - the number of non-correlated comparisons $g_{bk}^{(p)}(x_i, x_j)$ and $g_{bk}^{(p)}(x_r, x_s)$ $(r \ne i,j;\, s \ne i,j;\; i,j,r,s \in I_p)$ in the set $\hat{I}_p$.

In the case of strict relation and the errorless estimation result, the variance of the variable satisfies the inequality (10.18), while the expected value $E(S_N^{(s)})$ satisfies the condition:

$$E(S_N^{(s)}) \ge 1 - 2\delta. \tag{10.19}$$

The inequalities (10.17)–(10.19) and the Chebyshev inequality for the variance, imply the form of tests for both forms of the relation. The null and the alternative hypotheses for the weak preference relation, versus strict relation, assume the form:

$$H_{w,0} : E(S_N^{(w)}) \le -1 + 2\delta, \tag{10.20}$$

$$H_{w,1} : E(S_N^{(w)}) > 1 - 2\delta, \tag{10.21}$$

with the critical region:

$$\Lambda_N^{(w)} = \{S_N^{(w)} \mid S_N^{(w)} > 2\delta - 1 + \lambda\sigma_S\}\,, \tag{10.22}$$

where:
λ - the parameter obtained on the basis of Chebyshev inequality (28) (see Klukowski, 2006),
σ_S - square root of the expression $\frac{4}{N}(1-2L(I_w)/(\#(I_w))^2)\delta(1-\delta)$.

The hypotheses verifying strict relation, versus weak form, assume the form:

$$H_{s,0}: E(S_N^{(s)}) \geq 2\delta - 1\,, \tag{10.23}$$

$$H_{s,1}: E(S_N^{(s)}) < -1 + 2\delta\,, \tag{10.24}$$

with the critical region:

$$\Lambda_N^{(s)} = \{S_N^{(s)} \mid S_N^{(s)} < 1 - 2\delta - \lambda\sigma_S\}\,. \tag{10.25}$$

The evaluation of probability of the first type error of the test for weak preference relation assumes the form (see Klukowski, 2006):

$$1-(1-\alpha_w)P(\hat{\chi}_1^{(pw)},\ldots,\hat{\chi}_{\hat{n}_w}^{(pw)} \equiv \chi_1^{*(pw)},\ldots,\chi_{n_w}^{*(pw)} \mid \mathbf{R}^{(pw)})\,,$$

where: α_w is the probability resulting from the Chebyshev inequality.

The evaluation of probability of the second type error of the test for weak preference relation assumes the form (see Klukowski, 2006):

$$1-(1-\beta_w)P(\hat{\chi}_1^{(pw)},\ldots,\hat{\chi}_{\hat{n}_w}^{(pw)} \equiv \chi_1^{*(pw)},\ldots,\chi_{n_w}^{*(pw)} \mid \mathbf{R}^{(pw)})\,, \tag{10.26}$$

where: β_w is the probability resulting from the Chebyshev inequality.

The evaluations of probabilities of both errors require the probabilities of errorless estimation results, because the relationships (10.17)–(10.19) are valid in such case. The probabilities can be determined with the use of simulation (Chapter 9). It is reasonable to verify the null hypothesis about the relation form, providing the lower value of the criterion function.

The test based on medians from the comparisons of pairs is similar to the case of *N*=1, with appropriate modification of probabilities of comparison errors.

The idea of the test can be used for other values of the number of subsets *n* ($n \in \{1, ..., m\}$) ; the hypotheses verified assume the form:

$H_0: n = n_0$,
$H_1: n = n_0 - 1$ or $n = n_0 + 1$.

10.5. Summary

The tests presented here allow for a versatile verification of the assumptions about: • comparison errors, • existence of a relation in a set under consideration, and • relation type. In the case of negative result of verification it is possible to detect sources of errors or to reject an incorrect estimate. In particular, it possible to choose a correct estimate in the case of multiple solutions to the optimization tasks. Statistical tools, tests and simulations, allow for computer based inference. In the case of comparisons obtained from statistical tests or other decision procedures, the whole process of estimation and validation can be computerized. Thus, the approach proposed belongs to a common ground of statistical and data mining techniques.

Many tests proposed for validation of estimates are based on empirical comparison errors, i.e. resulting from the estimates ($\hat{T}_{\upsilon}^{(\ell)}(x_i, x_j)$ or $\widetilde{T}_{\upsilon}^{(\ell)}(x_i, x_j)$. Statistics based on such errors are, in fact, realizations of some mixtures of random variables. Therefore, the probabilities of errors in tests have to be corrected by the probability of errorless estimate. This probability can be determined on the basis of simulation. The concept of quasi-uniform distributions, introduced by the author, allows for application of simulation approach also in the case of unknown distributions of comparison errors. Thus, the estimators proposed can be applied in any rational scientific investigation. Statistical tools, based on simulation techniques, provide significant progress in comparison to heuristic approaches.

The estimates, which satisfy all tests used in validation process ought to be considered as trustworthy and reliable. An important feature of the approach (estimation and validation) proposed is its simplicity, which broadens the circle of potential users.

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