NOTE ON MULTIDIMENSIONAL QUANTIFICATION OF DATA OBTAINED BY PAIRED COMPARISON

CHIKIO HAYASHI

(Received March 6, 1967)

This note is described as a supplement to the paper [1] published previously, to give remarks to which we must refer in applying our method to practical data obtained experimentally by paired comparison. In [1], we discussed the case where only m_{ij}'s (the number of judges who judge $O_i \succ O_j$) are known and the preference pattern by each judge between O_i and O_j, $i, j=1, \cdots, N$, $i \neq j$, N being the size of objects (we call PPEJ), is not available. If PPEJ is available, it is not always advantageous to apply the method directly. In such a case, it is desirable to classify the judges into several groups, by using the information of PPEJ's, so that within each group we find judges showing similar features (preference patterns) and that between different groups we find judges of different features, and then to apply the model B of [1] in each group. Thus we can reveal the group-difference through PPEJ's and make an inference concerning the preference structure constructed by judges and objects.

In grouping, it is useful to apply the e_{kl}-type-quantification [2]. Let e_{kl} be $\sum_{i=1}^{N} \sum_{\substack{j=1 \\ i \neq j}}^{N} \alpha_{ij}(k, l)$ where $\alpha_{ij}(k, l)$ means 1 when the kth judge and the lth judge have given the same judgement in comparison between O_i and O_j, i.e. both judge k and judge l observe $O_i \succ O_j$ or $O_i \prec O_j$; and 0 otherwise where $k, l=1, 2, \cdots, n$, n being the number of judges. That is, e_{kl} means the number of coinsidences of judgement in PPEJ for fixed k, l. Therefore, it could be said that the larger value e_{kl} shows, the more similar judgements judges k and l make. Thus, by the method of e_{kl}-type quantification, we classify the judges into several groups so that any two judges show large e_{kl} may belong to the same group and any two who show small e_{kl} may belong to different groups. We gave this example with success in [5].

We shall mention the theorems in the following.

Let x_k be the numerical values given to the kth judges $k=1, 2, \cdots, n$, by which we classify the judges into groups. x's are obtained so as to maximize the measure $G=-\sum\sum (x_k-x_l)^2 e_{kl}$ under the condition that the variance of x is a constant, for example, equal to 1. It is shown in [2], [3] that this idea is valid for our purpose.

THEOREM 1. (*unidimensional quantification* [2]) *Solve the following characteristic equation and find the characteristic vector corresponding to the maximum characteristic root.*

$$(1) \qquad \left\{-\sum_{\substack{l=1 \\ l\neq u}}^{n} a_{ul}\right\} x_u + \sum_{\substack{l=1 \\ l\neq u}}^{n} a_{ul} x_l = \lambda x_u , \qquad (u=1, 2, \cdots, n)$$

where $a_{kl}=e_{kl}+e_{lk}$, $\lambda=G/n\tau^2$, $\tau^2=\frac{1}{n}\sum_k^n x_k^2$. *x's which satisfy* (1) *have the property* $\sum_k^n x_k=0$, *i.e.* $\bar{x}=0$, *so* τ^2 *is equal to the variance of* x.

THEOREM 2. (*multidimensional quantification* [3]) *Take the measure* $Q=-\sum_{\substack{k=1 \\ k\neq l}}^{n}\sum_{l=1}^{n} e_{kl}\left\{\sum_s^S \frac{({}^sx_k-{}^sx_l)^2}{\tau_s^2}\right\}$, *where* ${}^sx_k$ *is to be given to the element* k *in the* s*th dimension. Maximize* Q *under the condition of* ${}^sx$*'s and* ${}^{s'}x$*'s* $(s\neq s')$ *being orthogonal, i.e.* $\sum_k^n {}^sx_k\, {}^{s'}x_k=0$ $(s, s'=1, 2, \cdots, S)$. *The solutions are given as the characteristic vectors corresponding to the characteristic roots* $\lambda_1>\lambda_2>\cdots>\lambda_s$ (*in the successive descending order of magnitude of characteristic roots*) *in the same characteristic equation* (1) *obtained in unidimensional case.*

τ_s^2's reduce to the variances of ${}^sx$'s. If we take $\tau_s^2=1$ for every s, we can show the meaning of Q more clearly. In the multidimensional case, we take the distance in S-dimensional Euclidean space under the constant variance instead of $(x_k-x_l)^2$'s in the unidimensional case.

THEOREM 3. [4] *Take* $e_{kl}+c=e'_{kl}$ *instead of* e_{kl}, *where* c *is a constant. Then, the solution* x'*'s concerning* G' *or* Q' *in which* e'_{kl}'s *are taken instead of* e_{kl}*'s in* G *or* Q *are quite the same to* x*'s concerning* G *and* Q, *except a constant multiplier, where* x*'s and* x'*'s correspond to the same order of magnitude of characteristic roots.*

By theorem 3, we can choose the constant c for the convenience of computation without any loss of validity.

Note. After the publication of [1], we found interesting papers concerning with multidimensional quantification (unfolding) of the data, [6], [7], [8].

THE INSTITUTE OF STATISTICAL MATHEMATICS

REFERENCES

[1] C. Hayashi, "Multidimensional quantification of the data obtained by the method of paired comparison," *Ann. Inst. Statist. Math.*, 16 (1964), 231-245.

[2] C. Hayashi, "On the prediction of phenomena from qualitative data and the quantification of qualitative data from the mathematico-statistical point of view," *Ann. Inst. Statist. Math.*, 2 (1952), 93-96.

[3] C. Hayashi, "Theory of quantification and its example (VI)," *Proc. Inst. Statist. Math.*, 9 (1961), 29-35.

[4] C. Hayashi, S. Takakura, M. Makita and S. Saito, "One method of quantification of attitude," *Proc. Inst. Statist. Math.*, 6 (1958), 1-40.

[5] S. Kondo and C. Hayashi, "On the preferences of sound quality," *Journal of the Acoustical Society of Japan*, 21 (1965), 216-226.

[6] P. Slater, "Inconsistencies in a schedule of paired comparisons," *Biometrika*, 48 (1961), 303-312.

[7] J. B. Kruskal, "Multidimensional scaling by optimizing goodness of fit to a nonparametric hypothesis," *Psychometrika*, 29 (1964), 1-27.
"Nonparametric multidimensional scaling; Numerical method," *Psychometrika*, 29 (1964), 115-129.

[8] L. Guttmam, "The nonparametric breakthrough for the behavioral sciences," Invited Address to the Automatic Data Processing Conference of the Information Processing Association of Israel, 5-6, 1966.