ON A COEFFICIENT OF UNIDIMENSIONAL ORDERING FOR THE INDIVIDUALS' ATTITUDES

By Hirosi Hudimoto

(Received July 18, 1960)

1. Introduction and summary

The problem treated in this paper is to grade the intensity of an individual's attitude to some object A which may be psychological or sociological (such as circumstances, concept or topic). Suppose that we have m items (or questions) $(\Pi_1, \Pi_2, \cdots, \Pi_m)$ which characterize A in some sense, and we can obtain responses $\{e_{ik}\}$ of n individuals to those items, such that e_{ik} takes on one or zero according as the kth individual takes positive or negative attitude to the ith item. The grading of the individual's intensities of attitude to A is then carried through by considering the total score $\sum_{i=1}^{m} e_{ik}$. In section 2, adequacy of this procedure is discussed, and in section 3, a coefficient of tendency to the unidimensional ordering is given. In section 4, the case where responses to different items are statistically independent is treated.

2. Adequacy of unidimensional ordering

We intend to grade the intensity of an individual's attitude to a psychological or sociological object A using responses of each individual to a set of m items (or questions) $(\Pi_1, \Pi_2, \cdots, \Pi_m)$ related with A. We consider a set of dichotomous items to which an individual's judgement can be expressed by bipolar scales (in positive or negative, favorable or unfavorable).

Now, we shall give score 1 or 0 according to individual's response of favor or unfavor to each item. Then, suppose the degree of intensity of individual's attitude to the object A can be expressed by the total score of these respondent values for the items. We call such a procedure adequate, when the relation as shown in the following scheme (Fig. 1) holds. Now, suppose that the number of respondent individuals n is larger than or equal to $(m+1)$ and that the relation in this scheme precisely holds. Denote the proportion (relative frequency) of the respondent individuals with score 1 to item Π_i by p_i. Then the proportions of respondent individuals corresponding to patterns

Fig. 1.

items	Π_1	Π_2	$\cdots$	Π_{m-1}	Π_m
The ideal pattern of individuals' responses	1	1	$\cdots$	1	1
	1	1	$\cdots$	1	0
	$\vdots$	$\vdots$		$\vdots$	$\vdots$
	1	0	$\cdots$	0	0
	0	0	$\cdots$	0	0

$(0, 0, \cdots, 0)$, $(1, 1, \cdots, 1)$ and $(\overbrace{1, 1, \cdots, 1}^{i}, \overbrace{0, \cdots, 0}^{m-i})$ in the scheme (Fig. 1) are $(1-p_1)$, p_m and (p_i-p_{i+1}) $(i=1, \cdots, m-1)$, when $(\Pi_1, \Pi_2, \cdots, \Pi_m)$ is rearranged in order of magnitude of p_i's, i.e. $p_1 \geqq p_2 \geqq \cdots \geqq p_m$. Denote these by $\tau_1=(1-p_1)$, $\tau_2=p_1-p_2$, $\cdots$, $\tau_i=p_{i-1}-p_i$, $\cdots$, and $\tau_{m+1}=p_m$, respectively. Such a situation is shown in Fig. 2.

Fig. 2.

Π_1	Π_2	$\cdots$	Π_i	$\cdots$	Π_{m-1}	Π_m	
p_1	p_2	$\cdots$	p_i	$\cdots$	p_{m-1}	p_m	
1	1	$\cdots$	1	$\cdots$	1	1	$\tau_{m+1}=p_m$
1	1	$\cdots$	1	$\cdots$	1	0	$\tau_m=p_{m-1}-p_m$
$\vdots$	$\vdots$		$\vdots$		$\vdots$	$\vdots$	
1	0	$\cdots$	0	$\cdots$	0	0	$\tau_2=p_1-p_2$
0	0	$\cdots$	0	$\cdots$	0	0	$\tau_1=1-p_1$

However, the situation as described above will not usually occur because of individuals' misjudgements or of inadequacy of the questions, and 2^m kinds of individuals' response patterns can occur if n is larger than or equal to 2^m. But, if there should exist the property that Π_i's are mutually related, perhaps the individuals' response patterns in the scheme (Fig. 1) would have higher proportions of responses than the remaining patterns. Therefore, when n is larger than 2^m, it will be reasonable to compare the set of actually occurred patterns with the ideal set of patterns in the scheme.

3. A coefficient of unidimensional ordering

Let us consider the case that the number of respondent individuals

n is comparatively larger than 2^m. In this case, it will be able to classify the responses into 2^m possible kinds of actual response patterns. Then, we compare the set of actually occurred patterns with the ideal set of patterns in the scheme by the ratio of variances calculated from the individuals' total scores.

Let t_i be the proportion of the individuals' responses on the ith pattern from the pattern $(0, 0, \cdots, 0)$ in order of occuring of actual patterns. The ratio as described above is expressed by

$$(1) \qquad \frac{B_m}{A_m}=\frac{\sum_{k'=1}^{2^m}\left[\sum_{i=1}^{m}(e_{ik}'-p_i)\right]^2 t_{k'}}{\sum_{k=1}^{m+1}\left[\sum_{i=1}^{m}(e_{ik}-p_i)\right]^2 \tau_k},$$

where τ_k has been defined in section 2 and

$$(2) \qquad e_{ik}=\begin{cases}1 & \text{If score 1 is given to the } i\text{th item in the } k\text{th response pattern,}\\ 0 & \text{otherwise.}\end{cases}$$

Then, under the condition that $\{p_1, p_2, \cdots, p_m\}$ *has occurred, it holds that*

$$(3) \qquad 0\leqq\frac{B_m}{A_m}\leqq 1$$

for all $\{t_1, t_2, \cdots, t_{2^m}\}$.

In the case $m=2$, the relation holds because

$$\begin{aligned} A_2&=\sum_{k=1}^{3}\left[\sum_{i=1}^{2}(e_{ik}-p_i)\right]^2\tau_k\\ &=\sum_{k'=1}^{4}\left[\sum_{i=1}^{2}(e_{ik'}-p_i)\right]^2 t_{k'}+2p_{(0,1)}\\ &=B_2+2p_{(0,1)}\ , \end{aligned}$$

and $p_{(0,1)}\geqq 0$, where $p_{(0,1)}$ denotes the value of $t_{k'}$ on pattern (0, 1).

Assume that the relation $A_m\geqq B_m$ holds in the case m, one more item is added and $t_{k'}$'s are divided into $t^0_{k'}$'s and $t^1_{k'}$'s according to scores 0 and 1 taken in the $(m+1)$st item.

Then, using the relations $\sum_{k'=1}^{2^m}\left\{\sum_{i=1}^{m}\xi_{ik'}\right\}t_{k'}=0$ and $\sum_{k'=1}^{2^m}t_{k'}=1$ we have

$$(4) \qquad B_{m+1}=\sum_{k'=1}^{2^m}\left[\left\{\sum_{i=1}^{m}\xi_{ik'}-p_{m+1}\right\}^2 t^0_{k'}+\left\{\sum_{i=1}^{m}\xi_{ik'}+(1-p_{m+1})\right\}^2 t^1_{k'}\right]$$

$$=\sum_{k'=1}^{2^m}\left(\sum_{i=1}^{m}\xi_{ik'}\right)^2 t_{k'}+2\sum_{k'=1}^{2^m}\sum_{i=1}^{m}\xi_{ik'}t_{k'}^1+p_{m+1}(1-p_{m+1})$$
$$=B_m+2\sum_{k'=1}^{2^m}\sum_{i=1}^{m}\xi_{ik'}t_{k'}^1+p_{m+1}(1-p_{m+1})\ ,$$

where $\xi_{ik'}$ denotes $(e_{ik'}-p_i)$. We shall apply the above relation to finding relation (3). On the other hand, we get by the same procedure

(4′)
$$A_{m+1}=\sum_{k=1}^{m}\left\{\sum_{i=1}^{m}\xi_{ik}-p_{m+1}\right\}^2\tau_k+\left\{\sum_{i=1}^{m}\xi_{i,m+1}-p_{m+1}\right\}^2(p_m-p_{m+1})$$
$$+\left\{(m+1)-\sum_{i=1}^{m+1}p_i\right\}^2 p_{m+1}$$
$$=\sum_{k=1}^{m+1}\left(\sum_{i=1}^{m}\xi_{ik}\right)^2\tau_k+2\left(m-\sum_{i=1}^{m}p_i\right)p_{m+1}+p_{m+1}(1-p_{m+1})$$
$$=A_m+2\left(m-\sum_{i=1}^{m}p_i\right)p_{m+1}+p_{m+1}(1-p_{m+1})\ .$$

Therefore, when relation $A_m \geqq B_m$ holds, we get

$$\left(m-\sum_{i=1}^{m}p_i\right)p_{m+1}=\left(m-\sum_{i=1}^{m}p_i\right)\sum_{k'=1}^{2^m}t_{k'}^1\geqq\sum_{k=1}^{2^m}\sum_{i=1}^{m}\xi_{ik'}t_{k'}^1\ ,$$

and relation (3) is seen to hold by mathematical induction.

Thus, the ratio B_m/A_m will be considered to be a coefficient that shows closeness of actual response patterns to the ideal set of response patterns in the scheme. However, if we want to give zero when actual response patterns do not show closeness to the ideal set of response patterns in the scheme, i.e., e_{ik} takes independently score 1 or 0 for each item, it is not suitable to adopt the ratio B_m/A_m. Therefore, we take up

(5)
$$U_n=\frac{b_m}{a_m}=\frac{\sum_{i\neq j}^{\binom{m}{2}}\sum_{k'=1}^{2^m}(e_{ik'}-p_i)(e_{jk'}-p_j)t_{k'}}{\sum_{i\neq j}^{\binom{m}{2}}\sum_{k=1}^{m+1}(e_{ik}-p_i)(e_{jk}-p_j)\tau_k}$$
$$=\frac{\sum_{k'=1}^{2^m}\left\{\sum_{i=1}^{m}(e_{ik'}-p_i)\right\}^2 t_{k'}-\sum_{i=1}^{m}p_i(1-p_i)}{\sum_{k=1}^{m+1}\left\{\sum_{i=1}^{m}(e_{ik}-p_i)\right\}^2\tau_k-\sum_{i=1}^{m}p_i(1-p_i)}\ .$$

It can be seen that (5) does not essentially differ from (1), and U_m is easily calculated from the last expression of (5). On the other hand, we can also change (5) into a slightly different form by the following procedure. For the numerator b_m, we have from (4)

$$B_m = B_{m-1} + 2\sum_{k'=1}^{2^{(m-1)}} \sum_{i=1}^{m-1} \xi_{ik'} t^1_{k',m} + p_m(1-p_m) ,$$

$$B_{m-1} = B_{m-2} + 2\sum_{k'=1}^{2^{(m-2)}} \sum_{i=1}^{m-2} \xi_{ik'} t^1_{k',m-1} + p_{m-1}(1-p_{m-1}) ,$$

$$\vdots \qquad \vdots \qquad \vdots \qquad \vdots$$

$$B_2 = B_1 + 2\sum_{k'=1}^{2} \xi_{1k'} t^1_{k',2} + p_2(1-p_2)$$

$$B_1 = p_1(1-p_1) ,$$

and

$$b_m = B_m - \sum_{i=1}^{m} p_i(1-p_i) = 2\sum_{j=1}^{m-1} \sum_{k'=1}^{2^j} \sum_{i=1}^{j} \xi_{ik'} t^1_{k',j+1}$$

$$= 2\sum_{j=1}^{m-1} \sum_{k'=1}^{2^j} \sum_{i=1}^{j} (e_{ik'} - p_i) t^1_{k',j+1} ,$$

where $\sum_{k'=1}^{2^j} t^1_{k',j+1} = p_{j+1}$, and for the denominator a_m, we have

$$a_m = \sum_{i \neq j}^{\binom{m}{2}} \sum_{k=1}^{m+1} (e_{ik} - p_i)(e_{jk} - p_j)\tau_k$$

$$= \sum_{i \neq j}^{\binom{m}{2}} \left\{ \sum_{k=1}^{m+1} e_{ik} e_{jk} \tau_k - p_i p_j \right\} = \sum_{k=1}^{m+1} \sum_{i \neq j}^{\binom{m}{2}} e_{ik} e_{jk} \tau_k - \sum_{i \neq j}^{\binom{m}{2}} p_i p_j$$

$$= \binom{m}{2} p_m + \binom{m-1}{2}(p_{m-1} - p_m) + \cdots + \binom{m-i}{2}(p_{m-i} - p_{m-i+1})$$

$$+ \cdots + (p_2 - p_3) - \sum_{i \neq j}^{\binom{m}{2}} p_i p_j$$

$$= \sum_{i=1}^{m-1} (m-i) p_{m-i+1} - \sum_{i \neq j}^{\binom{m}{2}} p_i p_j$$

$$= \sum_{i=1}^{m-1} \left\{ (m-i) - \sum_{j=1}^{m-i} p_j \right\} p_{m-i+1} ,$$

and

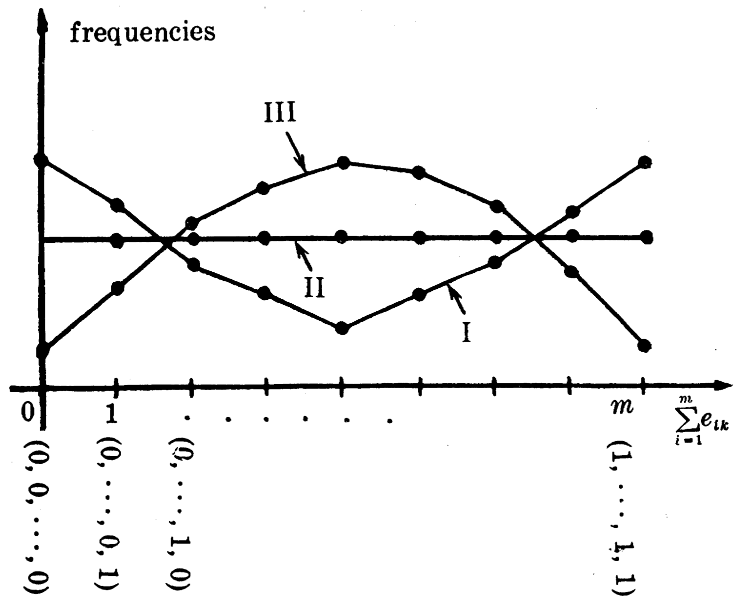
$$a_m = A_m - \sum p_i(1-p_i)$$

$$= 2\sum_{j=1}^{m-1} p_{j+1} \sum_{i=1}^{j} (1-p_i) .$$

Therefore, (5) is written as follows:

$$U_m = \frac{\sum_{j=1}^{m-1} \sum_{k'=1}^{2^j} \sum_{i=1}^{j} (e_{ik'} - p_i) t^1_{k',j+1}}{\sum_{j=1}^{m-1} p_{j+1} \sum_{i=1}^{j} (1-p_i)} \tag{6}$$

and *the denominator* a_m *of* U_m *is always positive as far as some* p_i *takes neither zero nor one.*



But numerator b_m in (6) may be negative. Especially, when the individuals' response patterns are arranged in an ascending order of the total respondent values $\sum_{i=1}^{m} e_{ik}$, where an arrangement of the same total scores is arbitrary, if the distribution of responses concentrates toward such response patterns as is located about the middle of this arrangement, such a situation is seen to be liable to occur from the form of (6).

Fig. 3 shows three kinds of typical distributions of frequencies on individuals' response patterns in order to illustrate the above described situation. Then, U_m will be liable to take a positive or negative value corresponding to the distribution of the type I or III in this figure, and $U_m=0$ for II. For instance, when the proportion of individuals that take the favorable response to item $\prod_1$, p_1 is near one and p_m to item $\prod_m$ near zero, the case of distribution of the type III will occur. However, in such a case, $\prod_1$ and $\prod_m$ are useless for our purpose because of no ability to discriminate the individuals' attitudes. Therefore, excepting such items, we may remove the value of U_m from negative to positive. We shall give an illustrative example of the case of three items in Fig. 4.

4. Case where responses to different items are statistically independent

In this section, we shall assume that e_{ik}'s independently take score 1 or 0 of each item ($i=1, 2, \cdots, m$), and rewrite (5) as

$$
(7) \qquad U_m = \frac{\sum\limits_{i \neq j}^{\binom{m}{2}} \sum\limits_{k'=1}^{n} (e_{ik'} - p_i)(e_{jk'} - p_j)}{n \sum\limits_{j=1}^{m-1} p_{j+1} \sum\limits_{i=1}^{j} (1 - p_i)},
$$

Fig. 4.

Π_1	Π_2	Π_3
p_1	p_2	p_3
$\frac{17}{20}$	$\frac{1}{2}$	$\frac{3}{20}$
$\frac{7}{10}$	$\frac{1}{2}$	$\frac{3}{10}$
$\frac{5}{8}$	$\frac{1}{2}$	$\frac{3}{8}$
$\frac{11}{20}$	$\frac{11}{20}$	$\frac{9}{20}$
$\frac{1}{2}$	$\frac{1}{2}$	$\frac{1}{2}$
$\frac{1}{2}$	$\frac{1}{2}$	$\frac{1}{2}$
$\frac{13}{20}$	$\frac{1}{2}$	$\frac{7}{20}$
$\frac{9}{16}$	$\frac{1}{2}$	$\frac{7}{16}$

frequencies

		Π_1	Π_2	Π_3								
actual patterns	t_8	1	1	1	$\frac{1}{4}$	$\frac{2}{10}$	$\frac{2}{10}$	$\frac{1}{8}$	$\frac{1}{20}$	0	0	0
	t_7	1	1	0	$\frac{1}{8}$	$\frac{2}{10}$	$\frac{1}{10}$	$\frac{1}{8}$	$\frac{2}{10}$	$\frac{1}{4}$	$\frac{3}{10}$	$\frac{2}{5}$
	t_6	1	0	1	$\frac{1}{16}$	$\frac{1}{20}$	$\frac{1}{10}$	$\frac{1}{8}$	$\frac{1}{10}$	$\frac{1}{8}$	$\frac{1}{10}$	$\frac{1}{20}$
	t_5	0	1	1	$\frac{1}{16}$	$\frac{1}{20}$	$\frac{1}{10}$	$\frac{1}{8}$	$\frac{2}{10}$	$\frac{1}{8}$	$\frac{1}{10}$	$\frac{1}{20}$
	t_4	1	0	0	$\frac{1}{8}$	$\frac{2}{10}$	$\frac{1}{10}$	$\frac{1}{8}$	$\frac{2}{10}$	$\frac{1}{4}$	$\frac{3}{10}$	$\frac{2}{5}$
	t_3	0	1	0	$\frac{1}{16}$	$\frac{1}{20}$	$\frac{1}{10}$	$\frac{1}{8}$	$\frac{1}{10}$	$\frac{1}{8}$	$\frac{1}{10}$	$\frac{1}{20}$
	t_2	0	0	1	$\frac{1}{16}$	$\frac{1}{20}$	$\frac{1}{10}$	$\frac{1}{8}$	$\frac{1}{10}$	$\frac{1}{8}$	$\frac{1}{10}$	$\frac{1}{20}$
	t_1	0	0	0	$\frac{1}{4}$	$\frac{2}{10}$	$\frac{2}{10}$	$\frac{1}{8}$	$\frac{1}{20}$	0	0	0
ideal patterns	τ_4	1	1	1	$\frac{7}{16}$	$\frac{7}{20}$	$\frac{1}{2}$	$\frac{1}{2}$	$\frac{9}{20}$	$\frac{3}{8}$	$\frac{3}{10}$	$\frac{3}{20}$
	τ_3	1	1	0	$\frac{1}{16}$	$\frac{3}{20}$	0	0	$\frac{1}{10}$	$\frac{1}{8}$	$\frac{2}{10}$	$\frac{7}{20}$
	τ_2	1	0	0	$\frac{1}{16}$	$\frac{3}{20}$	0	0	0	$\frac{1}{8}$	$\frac{2}{10}$	$\frac{7}{20}$
	τ_1	0	0	0	$\frac{7}{16}$	$\frac{7}{20}$	$\frac{1}{2}$	$\frac{1}{2}$	$\frac{9}{20}$	$\frac{3}{8}$	$\frac{3}{10}$	$\frac{3}{20}$
coefficients of ordering				U_3	0.47	0.37	0.20	0	−0.22	−0.45	−0.54	−0.74

where $n(\geqq m+1)$ is the sample size. In this case, if the sample size n is large, U_m will approximately have the normal distribution with mean 0 and variance

$$\frac{\sum_{i\neq j}^{\binom{m}{2}} p_i p_j (1-p_i)(1-p_j)}{(n-1)\left(\sum_{j=1}^{m-1} p_{i+1} \sum_{i=1}^{j} (1-p_i)\right)^2} .$$

We shall give the first four moments of U_m under the condition that $\{p_1, p_2, \cdots, p_m\}$ has occurred. First, we have, writing E for expected values,

$$\begin{aligned} &E(\xi_{ik})=0 \\ &E(\xi_{ik}^2)=p_i(1-p_i) \\ &E(\xi_{ik}^3)=p_i(1-p_i)(1-2p_i) \\ &E(\xi_{ik}^4)=p_i(1-p_i)\{1-3p_i(1-p_i)\} \end{aligned}$$

where $\xi_{ik}=e_{ik}-p_i$. Therefore, the mean and variance of U_m are easily obtained as follows:

$$E(U_m)=0$$

$$E(U_m^2)=\frac{\sum_{i\neq j}^{\binom{m}{2}} p_i p_j (1-p_i)(1-p_j)}{(n-1)\left(\sum_{j=1}^{m-1} p_{j+1} \sum_{i=1}^{j} (1-p_i)\right)^2} .$$

We must evaluate the numerator of (7) in order to find the third and fourth moments, but the procedure can be carried out in the same ways as in Kendall [2, pp. 107–09]. Actually we have

$$\begin{aligned} E(U_m^3)=&\frac{1}{\left(\sum_{i=1}^{m-1} p_{j+1} \sum_{i=1}^{j} (1-p_i)\right)^3}\left\{\frac{6}{(n-1)^2}\sum p_i p_j p_k (1-p_i)(1-p_j)(1-p_k)\right. \\ &\left.+\frac{n-3}{(n-2)(n-1)(n+1)}\sum p_i p_j (1-p_i)(1-p_j)(1-2p_i)(1-2p_j)\right\} , \\ E(U_m^4)=&\frac{1}{\left(\sum_{j=1}^{m-1} p_{j+1} \sum_{i=1}^{j} (1-p_i)\right)^4}\left[\frac{3}{(n-1)^2}(\sum p_i p_j (1-p_i)(1-p_j))^2\right. \\ &-\frac{6}{(n-1)^2(n+1)}(\sum p_i^2 p_j^2 (1-p_i)^2(1-p_j)^2)+\frac{72}{(n-1)^3} \\ &\times(\sum p_i p_j p_k p_l (1-p_i)(1-p_j)(1-p_k)(1-p_l)) \end{aligned}$$

$$+\frac{12}{(n-1)^2(n-2)}(\sum p_i p_j p_k(1-p_i)(1-p_j)(1-p_k)(1-2p_i)(1-2p_j))$$

$$+\frac{1}{(n-1)(n-2)(n-3)n^3}\sum\left\{(n+1)p_i(1-p_i)-6np_i^2(1-p_i)^2\right\}$$

$$\times\left\{(n+1)p_j(1-p_j)-6np_j^2(1-p_j)^2\right\}\Bigg],$$

$i, j, k, l=1, 2, \cdots, m(i\neq j\neq k\neq l)$.

THE INSTITUTE OF STATISTICAL MATHEMATICS.

REFERENCES

[1] L. Guttman, "The basis for scalogram analysis," S. A. Stouffer et al., *Measurement and prediction*, Princeton Univ. Press, 1950, pp. 60-90.
[2] M. G. Kendall, *Rank correlation methods*, Edition 2, Charles Griffin and Company Limited, 1955.
[3] R. Jardine, "Ranking methods and the measurement of attitudes," *J. Amer. Statist. Assn.*, Vol. 53, (1958), pp. 720-728.
[4] Leo A. Goodman, "Simple statistical methods for scalogram analysis," *Psychometrika*, Vol. 24, (1959), pp. 29-43.