RANKING IN FRACTIONAL TRIAD COMPARISONS

BY

S.C. RAI

Institute of Agricultural Research Statistics, New Delhi

(Received in March, 1970)

1. INTRODUCTION

In sensory test, the number of samples a judge can assess is often limited by sensory fatigue. The difficulty of obtaining satisfactory quantitative measures of treatment effects usually entails assessment by ranking. The analysis of experiments based on ranking items has received considerable attention in Statistical methodology. Ranking methods are generally used where quantitative observations cannot be obtained easily or sometimes we follow this method in order to reduce the labour of computation or to get a rapid result.

Various authors have proposed different methods for the rank analysis. The analysis of paired comparisons has attracted the attention of many authors including Thurstone (1927), Guttman (1946), Kendall and Babington Smith (1940), Bradley and Terry (1952). Recently Rai and Sadasivan have proposed an extended model of Bradley and Terry for fractional paired comparisons. The analysis of experiments involving ranking in triple comparisons has been developed by Pendergrass and Bradley (1960). Simplification to this model has been proposed by Park (1961). Here we shall develop a method of analysis of ranking involving fractional triad comparisons as an extension of the Pendrgrass-Bradley model. A test procedure has been developed where a mathematical model involving treatment parameters has been proposed. The estimation of these parameters and investigation of properties of the model have been discussed. The methods employed are parallel to those of Rai and Sadasivan (7) proposed for fractional paired comparisons.

2. Mathematical model

Let us consider t treatments in an experiment involving triple comparisons. It is supposed that the treatments $T_1, T_2, \ldots T_t$ have true ratings $\pi_1, \ldots\ldots \pi_t$ on a particular subjective continuum which satisfies the following conditions :

$$\pi_i \geqslant 0 \quad \text{and} \quad \sum_{i=1}^{t} \pi_i = 1$$

The total number of triplets formed out of t treatments is tc_3. If we are interested in only one treatment say T_1 and we want to compare this treatment against the remaining $(t-1)$ treatments in triple comparisons, then we can study only those triplets where this particular treatment appears. The number of such triplets will be $\frac{(t-1)(t-2)}{2}$. The members of each of $\frac{(t-1)(t-2)}{2}$ triplets will be ranked in order of acceptability. In a triplet the best treatment will be given rank 1, the second one rank 2 and the third will have rank 3. In triplets having treatments T_1, T_i and T_j ; $i \neq j$, $=2, \ldots t)$ we have,

$$P(T_1 > T_i > T_j) = \pi_1^2 \, \pi_i / \triangle_{1ij} \qquad \ldots(1)$$

where $P(T_1 > T_i > T_j)$ represents the probability that treatment T_1 is rated top, T_i central and T_j bottom and

$$\triangle_{1ij} = \pi_1^2(\pi_i + \pi_j) + \pi_i^2(\pi_1 + \pi_j) + \pi_j^2(\pi_1 + \pi_i) \qquad \ldots(2)$$

3. The likelihood function

We may obtain the likelihood function assuming the probability independence for different triplets and for different replications. The ranks of T_1, T_i and T_j in the kth comparison will be denoted by $r_{1k \cdot ij}$; $r_{ik \cdot 1j}$ and $r_{jk \cdot 1i}$ respectively where $k=1, \ldots n$. Tied ranks are not permitted in the model. The probability of a specified ranking in the kth repetition is given by

$$\pi_1^{3-r_{1k \cdot ij}} \; \pi_i^{3-r_{ik \cdot 1j}} \; \pi_j^{3-r_{jk \cdot 1i}} / \triangle_{1ij} \qquad \ldots(3)$$

because if T_1 obtains the top rank, T_i is judged as second and T_j as third, then $r_{1k \cdot ij} = 1$; $r_{ik \cdot 1j} = 2$ and $r_{jk \cdot 1i} = 3$ and the expression (3) becomes $\pi_1^2 \pi_i / \triangle_{1ij}$. Similarly if T_i obtains rank 1, T_1 rank 2 and T_j rank 3 then the expression (3) becomes $\pi_i^2 \; \pi_1 / \triangle_{1ij}$ and so on. Multiplying the appropriate expression for all comparisons within a repetition and for all n repetitions, we reach the likelihood function in the general form

$$L=\frac{\pi_1^{\frac{3n}{2}(t-1)(t-2)-\sum\limits_{i<j=2}^{t}\sum\limits_{k=1}^{n} r_{1k,\,ij}}\ \prod\limits_{i=2}^{t}\pi_i^{3n(t-2)-\sum\limits_{i<j}\sum\limits_{k=1}^{n} r_{ik,\,1j}}}{\prod\limits_{1<i<j}(\triangle_{1ij})^n} \qquad \ldots(4)$$

When repetitions of the design are performed by groups with distinct parameters, the likelihood function will be product over the groups of functions of the form (4).

4. Likelihood Ratio Tests and Estimation

We can test the significance of the equality of treatment effects. Consider

H_0 : $\pi_1=\ldots=\pi_t=1/t$ against the alternative

H_a : $\pi_i\neq\pi_j$ for some $i\neq j$, $i, j=1,\ldots\ldots t$

The maximum likelihood estimators $p_1, \ldots p_t$ are obtained by maximising $\log L$ with respect to $\pi_1,\ldots\pi_t$ subject to the condition that $\sum\limits_{i=1}^{t}\pi_i=1$. The resulting normal equations are

$$\frac{a_1}{p_1}=n\sum_{i<j}\left\{\left[2p_1(p_i+p_j)+p_i^2+p_j^2\right]\Big/D_{1ij}\right\} \qquad \ldots(5)$$

and

$$\frac{a_i}{p_i}=n\sum_{j}\left\{\left[2p_i(p_1+p_j)+p_1^2+p_j^2\right]\Big/D_{1ij}\right\} \qquad \ldots(6)$$

$$i\neq j=(i=2,\ldots\ldots t)$$

where

$$a_1=\frac{3n}{2}(t-1)(t-2)-\sum_{i<j=2}^{t}\sum_{k=1}^{n} r_{1k,ij} \qquad \ldots(7)$$

$$a_i=3n(t-2)-\sum_{i<j=2}^{t}\sum_{k=1}^{n} r_{ik,1j} \qquad \ldots(8)$$

and

$$D_{1ij}=p_1^2(p_i+p_j)+p_i^2(p_1-p_j)+p_j^2(p_1+p_i) \qquad \ldots(9)$$

Solutions of these equations will give the values of $p_1,\ldots p_t$.

The normal equations are solved by iterative methods. The iteration proceeds as follows :

Let $p_1^{(o)}, \ldots p_t^{(o)}$ be first trial values for $p_1, \cdot p_t$. Second trial values are obtained by putting the first trial values in the following equations.

$$\frac{CO_1}{p_1^{(1)}} = n \sum_{i<j} \left\{ \left[2p_1^{(0)} (p_i^{(0)} + p_j^{(0)}) + p_i^{(0)2} + p_j^{(0)2} \right] \Big/ D_{1ij}^{(0)} \right\} \quad \ldots(10)$$

and $$\frac{CO_i}{p_i^{(1)}} = n \sum_{j} \left\{ \left[2p_i^{(0)} (p_1^{(0)} + p_j^{(0)}) + p_1^{(0)2} + p_j^{(0)2} \right] \Big/ D_{1ij}^{(0)} \right\} \quad \ldots(11)$$

$$i \neq j, = 2 \ldots t$$

where C is eliminated through the assumption that $\sum_{i=1}^{t} p_i = 1$ and $D_{1ij}^{(o)}$ is the value of D_{1ij} evaluated by using $p_1^{(o)} \ldots p_t^{(o)}$. The procedure indicated is continued through repeated use of (10) and (11) until the process converges to the required accuracy. The rapidity of the convergence is good if the initial trial values are good. The values of $p_1, p_2, \ldots\ldots p_t$ in the initial trial are taken in proportion to

$$\frac{n(t-1)(t-2)}{2} \left\{ (\Sigma r_2) \ldots (\Sigma r_t) \right\} : n(t-2) \left\{ (\Sigma r_1)(\Sigma r_3) \ldots (\Sigma r_t) \right\}$$

$$: n(t-2)\{(\Sigma r_1) \ldots (\Sigma r_{t-1})\} \text{ where } \Sigma r_1, \Sigma r_2, \ldots \Sigma r_t$$

are the sums of ranks for treatments $T_1, T_2, \ldots T_t$ respectively over all repetitions. These values are good first approximations in most cases. In case of extreme sets of values of the sums of ranks Σr_j where a particular treatment (say T_j) is always given the rank 1 in all the comparisons, the corresponding value of p_j is taken as 1. Similarly when a particular treatment is always rated as third in all the comparisons, the value of p for this treatment is taken as zero.

Now the estimated values of $\pi_1, \ldots \pi_t$ are obtained under the hypothesis H_a. The likelihood function L given by (4) is used to obtain the likelihood ratio λ and Z which is given by $Z = -2 \log_e \lambda$. Therefore

$$Z = n(t-1)(t-2) \log_e 6 + 2a_1 \log_e p_1 + 2 \sum_{i=2}^{t} a_i \log_e p_i$$

$$-2n \sum_{1<i<j} \log_e (D_{1ij}) \quad \ldots(12)$$

For large n, Z may be taken to have the x^2 distribution with $(t-1)$ *d.f.* (Wilks) under the null hypothesis H_o.

Tables for the distribution of Z for small sample sizes may be developed but these would be extremely voluminous. The procedure for developing such tables is similar to that used by Rai and Sadasivan (7). An example of such tables is given below in Table 1 wherein Z_o indicate the values of Z for specified sets of sums of ranks in the table.

TABLE 1

Distribution of $Z=-2 \log_e \lambda$

Number of treatment=4, Number of replication=1

Rank Sums				*Estimates of π_i*				*Distribution*	
Σr_1	Σr_2	Σr_3	Σr_4	p_1	p_2	p_3	p_4	z_0	$p(z \geqslant z_0)$
3	4	5	6	1	—	—	—	10·75	·405
5	2	5	6	—	1	—	—	10·75	
6	2	4	6	—	1	—	—	10·75	
6	2	5	5	—	1	—	—	10·75	
7	2	3	6	—	1	—	—	10·75	
7	2	4	5	—	1	—	—	10·75	
8	2	3	5	—	1	—	—	10·75	
8	2	4	4	—	1	—	—	10·75	
9	2	3	4	—	1	—	—	10·75	
4	3	5	6	·46	·42	·12	—	6·91	·465
4	4	4	6	·48	·26	·26	—	5·21	·495
6	3	3	6	·22	·39	·39	—	5·19	·525
5	3	4	6	·33	·42	·25	—	4·77	·585
8	3	3	4	·08	·35	·35	·22	2·64	·615
4	4	5	5	·46	·26	·14	·14	2·60	·645
7	3	3	5	·15	·37	·37	·11	2·27	·675
5	3	5	5	·32	·38	·15	·15	2·05	·705
6	3	4	5	·24	·39	·24	·13	1·23	·745
5	4	4	5	·34	·26	·26	·14	0·96	·885
7	3	4	4	·16	·38	·23	·23	0·92	·960
6	4	4	4	·25	·25	·25	·25	0·00	1·000

5. Combination of results

Sometimes, triple comparisons may be completed in groups of repetitions by different judges at different times or under different circumstances. The experiment may be considered as one with g groups of repetitions, the uth of which has n_u repetitions. Then $n=\sum_{u=1}^{g} n_u$. The failure of treatment parameters $\pi_{1u},\ldots\pi_{tu}$ to be the same for each group, represents a group X treatment interaction or lack of agreement. We now propose a test to detect such interactions.

Consider

$$H_o : \pi_{iu}=1/t \text{ for all } i \text{ and } u$$

and $$H_a : \pi_{iu}\neq 1/t \text{ for some } i \text{ and } u$$

If λ_c is the likelihood ratio in this case then

$$Z_c=-2 \log_e \lambda_c=\sum_{u=1}^{g} Z_u \qquad \ldots(13)$$

where Z_u is the value of Z given by (12) computed for the uth group. For large value of n_u, Z_c has the x^2 distribution with $g\ (t-1)$ degrees of freedom. This test is designated as the "combined test" of the treatment equality. The test of interaction is a test of null hypothesis.

$$H_o : \pi_{iu}=\pi_i \quad i=1,\ldots t;\ u=1,\ \ldots,\ g$$

against the alternative

$$H_a : \pi_{iu}\neq\pi_i \text{ for some } i \text{ and } u.$$

The likelihood ratio test of H_o and H_a depends on Z_c-Z and has the x^2 distribution with $(g-1)\ (t-1)$ degrees of freedom for large values of n_u. Z_c is calculated as defined in (13) and Z as in (12) based on pooling the sums of ranks obtained from all the repetitions.

6. A test of the model

The most general model for triple comparisons is formed by postulating for each triplet the existence of positive parameters π_{lij}, π_{lji}, π_{ilj}, π_{ijl}, π_{jli} and π_{jil} corresponding to the probabilities of occurrence of six possible rankings of T_l, T_i, T_j. Thus π_{lij} is the probability that T_l, T_i, T_j receive the rank 1, 2, 3 respectively in a triplet. The six parameters for each triplet add to unity and their maximum likelihood estimators are f_{lij}/n, f_{lji}/n, f_{ilj}/n, f_{ijl}/n, f_{jli}/n and f_{jil}/n for the n comparisons of this triplet where f_{lij} is the number of times the ranking 1, 2 and 3 for T_l, T_i and T_j respectively occurs in the n triplets.

The basic model for triple comparisons implies that

$$H_o : \pi_{lij} = \pi_1^2 \ \pi_i / \triangle_{lij} \ ; \ i \neq j \ ; \ i, j = 2, \ldots t$$

against the alternative

$$H_a : \pi_{lij} \neq \pi_1^2 \ \pi_i / \triangle_{lij} \text{ for some } i, j$$

The general likelihood function for triple comparisons is

$$L(\pi_{lij}) = \prod_{i<j} \pi_{lij}^{f_{lij}} \quad \ldots(14)$$

Let us define f'_{lij} as the expected frequency corresponding to the observed frequency f_{lij}, then the estimate of the expected frequency under H_o is given by

$$f'_{lij} = np_1^2 p_i / D_{lij} \quad \ldots(15)$$

The likelihood ratio statistic for testing H_o is given in terms of frequencies by

$$-2 \log_e \lambda = 2 \sum_{\substack{i,j=2 \\ i<j}}^{t} f_{lij} \log_e [f_{lij} / f'_{lij}] \quad \ldots(16)$$

For large n, this statistic has a x^2 distribution with

$$\left[\frac{5(t-1)(t-2)}{2} - (t-1) \right] \text{degree of freedom.}$$

Now in equation (16), we take $f_{lij}/f'_{lij} = 1 + e_{lij}$ where e_{lij} may have either positive or negative values. Then

$$-2 \log_e \lambda = 2 \sum_{\substack{i,j=2 \\ i<j}} f'_{lij} (1 + e_{lij}) \log_e (1 + e_{lij})$$

Expanding the logarithmic series in powers of e_{lij} and ignoring the terms greater than e_{lij}^2, we have

$$-2 \log_e \lambda \approxeq 2 \sum_{\substack{i,j=2 \\ i<j}}^{t} f'_{lij} (1 + e_{lij}) \left(e_{lij} - \frac{e_{lij}^2}{2} \right). \quad \ldots(17)$$

The errors committed in ignoring the higher terms of e_{lij} in the expansion of the series, will not be large if $|e_{lij}|$ is small. We notice that $\Sigma f'_{lij}\, e_{lij}=0$ and equation (17) takes the form

$$-2 \log_e \lambda \approx \Sigma f'_{lij}\, e_{lij}^2$$

After putting the value of e_{lij}, we have the final result in the following form:

$$-2 \log_e \lambda \approx \Sigma (f_{lij}-f'_{lij})^2 / f'_{lij} \qquad \ldots(18)$$

Thus the statistic $-2 \log_e \lambda$ is transformed to the usual χ^2 test of goodness of fit.

7. AN ILLUSTRATIVE EXAMPLE :

We shall demonstrate some of the procedures by a numerical example given below.

TABLE 2

Frequencies of rankings with $t=4$ and $n=40$

f 123=10	(8·86)	f 124=12	(11·22)	f 134=10	(11·29)
f 132= 8	(7·63)	f 142= 8	(7·06)	f 143= 8	(8·26)
f 213= 8	(7·56)	f 214= 8	(9·57)	f 341= 8	(4·45)
f 231= 6	(5·56)	f 241= 6	(5·13)	f 314= 8	(8·29)
f 312= 4	(5·61)	f 412= 4	(3·79)	f 413= 4	(4·45)
f 321= 4	(4·78)	f 421= 2	(3·23)	f 431= 2	(3·26)

Here only those triplets are retained where treatment T_1 appears.

From the above table we obtain the following preference matrix.

TABLE 3

Preference Matrix and sum of ranks

Treatment Nos.	*Number of times ranked as*			*Sum of ranks Σr_i*	a_i
	First	*Second*	*Third*		
1	56	36	28	212	148
2	28	28	24	156	84
3	24	26	30	166	74
4	12	30	38	186	54

We now obtain the values of $p_1,...p_4$. Successive approximations of these values along with the corresponding value of Z are presented in the following table.

TABLE 4

Successive approximations to p_1, ...p_4 and corresponding values of z

Approximations	p_1	p_2	p_3	p_4	z
1	·264	·258	·242	·216	13·03
2	·318	·273	·237	·172	17·30
3	·320	·273	·235	·172	17·32

The successive approximations show the convergence of the estimates of p_1, ...p_4 and of Z values. The final Z taken as χ^2 with 3 degrees of freedom indicates highly significant treatment main effects.

The goodness of fit test may be applied after appropriate substitution in (18). The different values of f'_{lij} are obtained from (15) and have been shown in parentheses in Table 3. Using (16), we find that $-2 \log_e \lambda = 5{\cdot}45$ and this is taken to a value from χ^2 distribution with 12 degrees of freedom. Use of the form (18) yields the comparable value of $-2 \log_e \lambda$ to be 5·41. The above values indicate that the proposed model is quite satisfactory for these data.

8. Discussion and summary

A method of analysis of ranking in fractional triad comparisons are discussed which permits tests of hypotheses of general class and the estimation of treatment ratings or preferences. In the null hypothesis we assume that the treatment ratings are equal whereas the alternative hypothesis makes no assumption regarding the equality of the treatment ratings. The probability of the sums of ranks $P(r_1 < r_i < r_j)$ involves three paired comparisons consisting of pairs of treatments (T_l, T_i) ; (T_l, T_j) and (T_i, T_j). These comparisons must be consistent in which $r_l < r_i$; $r_1 < r_j$ and $r_i < r_j$

The approach here may be used for the generalisations of ranking in blocks of size greater than 3. In subjective testing involving tastes or odours, paired or triple ranking will satisfy most of the requirements of the experimenter. Fractional triad comparison is efficient as compared to triple comparison involving all the triplets in the sense that it contains only a fraction of the triplets. This reduces the number of items to be presented to the judges for ranking in sensory tests and experiments can be conducted efficiently and satisfactorily. Fractional triad comparison involves a triple comparison consisting of all the triplets for one treatment and paired comparisons for remaining treatments. When the number of treatments is large in a sensory test then fractional triad comparison may be advocated in place of triple comparison involving all the triplets because the number of triplets taken for study will be reduced in a fractional triad comparison and the judges can give their opinion without much difficulty.

REFERENCES

1. Bradley. R.A. (1954) : Incomplete block rank analysis : On the appropriateness of the model for a method of paired comparisons. Biometrics, 10 ; 375-39?.
2. Bradley, R.A. and Terry, M.E. (1952) : The rank analysis of incomplete block designs I. The method of paired comparisons. Biometrika, 39 ; 324-345.
3. Fisher, R.A. and Yates, F. (1963) : Statistical Tables for Biological, Agricultural and Medical Research, 6th ed. Edinburgh : Oliver and Boyd.
4. Kendall, M.G. and Babington Smith, B. (1940) : On the method of paired comparisons. Biometrika, 31 ; 324-345.
5. Park, G.T. (1961) : Sensory testing by triple comparisons. Biometrics, 17 ; 251-260.
6. Pendergrass, R.C. and Bradley, R.A. (1960) : Ranking in triple comparisons. Contributions to Prabability and Statistics. Edited by I. Olkin, et. al., Stanford Univ. Press, Stanford, California. 331-351.
7. Rai, S.C. and Sadasivan, G. (1969) : A Bradley-Terry model for partial pairs. Abstract. Jour. Ind. Soc. Agr. Stat. 21 ; 155.
8. Wilks, S.S. (1946) : *Mathematical Statistics*. Princeton, N.J. : Princeton Univ. Press.