CORRECTIONS TO CLASSICAL PROCEDURES FOR ESTIMATING THURSTONE'S CASE V MODEL FOR RANKING DATA

Alberto Maydeu Olivares
Instituto de Empresa
Marketing Dept.
C/Maria de Molina 11-15
28006 Madrid – Spain
Alberto.Maydeu@ie.edu

Abstract

The classical method (Mosteller, 1951) for estimating Thurstone's Case V model for ranking data consists in a) transforming the observed ranking patterns to patterns of binary paired comparisons, b) obtaining the normal deviate corresponding to the men of each binary variable, and c) estimate the model parameters from these deviates by least squares. However, classical procedures do not take into account the dependencies among the deviates and as a result, asymptotic standard errors (SEs) and goodness of fit (GOF) test are incorrect.
We provide formulae that provide correct asymptotic SEs and GOF in this situation. A small simulation study shows that adequate standard errors and goodness of fit tests can be obtained with rather small sample sizes even in very large models.

Keywords:
categorical data analysis, preference data, random utility

1. Introduction

To model ranking data, Thurstone (1931) proposed transforming the observed ranking patterns to patterns of binary paired comparisons and fitting his paired comparisons model (Thurstone, 1927) to the transformed data. The classical method for estimating this model (Mosteller, 1951a; Torgerson, 1958) consists in obtaining the normal deviate corresponding to each paired comparisons mean, and then estimate the model parameters from these deviates by least squares. This is a mean structure approach to estimating Thurstone's Case V model as it only uses univariate information (the means) from the paired comparisons. Most Thurstonian models for paired comparisons and ranking data are not identified when estimated as a mean structure. Estimating them as a mean structure requires introducing unnecessary identification restrictions on the models. The most notable exception is Thurstone's Case V model for ranking data. This model is identified if estimated only from the means of the paired comparisons. Here, we provide asymptotically correct standard errors and goodness of fit test for this model when it is estimated as a mean structure using the classical estimation procedure described.

2. Mean structure estimation of Thurstone's Case V ranking model

Suppose a random sample of N individuals ranks n stimuli according to some preference criterion. To transform the rankings to paired comparisons we construct a dichotomous variable y_l for each ordered pairwise combination of stimuli to indicate which stimulus was ranked above the other

$$y_l = \begin{cases} 1 & \text{if stimulus } i \text{ is ranked above stimulus } i' \\ 0 & \text{if stimulus } i \text{ is ranked below stimulus } i' \end{cases}, \tag{1}$$

where $l \equiv (i,i'),(i = 1,...,n - 1;\ i' = i + 1,...,\ n)$. With n stimuli there are $\tilde{n} = \binom{n}{2} = \frac{n(n-1)}{2}$ paired comparisons.

Maydeu-Olivares (1999) showed that the probability of any such binary pattern under Thurstone's model is

$$\Pr\left(\bigcap_{l=1}^{\tilde{n}} y_l\right) = \int \underset{\mathbf{R}}{\cdots} \int \phi_{\tilde{n}}\left(\mathbf{z}^* : \mathbf{0}, \boldsymbol{P}\right) d\mathbf{z}^* \tag{2}$$

where $\phi_{\tilde{n}}(\bullet)$ denotes a $\tilde{n}$-variate normal density function, and $\mathbf{R}$ is a rectangular region with limits (τ_l, ∞) if $y_l = 1$ and $(-\infty, \tau_l)$ if $y_l = 0$. Let $\boldsymbol{\tau}$ be a vector obtained by stacking all thresholds τ_l in lexicographic order, Thurstone's model imposes the following restrictions on $\boldsymbol{\tau}$ and on the correlation matrix $\mathbf{P}$ (Maydeu-Olivares, 1999)

$$\boldsymbol{\tau} = -\mathbf{D}\mathbf{A}_n\boldsymbol{\mu} \qquad\qquad \boldsymbol{P} = \mathbf{D}\left(\mathbf{A}_n\boldsymbol{\Sigma}\mathbf{A}_n'\right)\mathbf{D}. \tag{3}$$

In (3) $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ are the mean vector and covariance matrix of the unobserved preferences (*discriminal processes*) assumed by Thurstone's model, $\mathbf{A}_n$ is a $\tilde{n} \times n$ design matrix where each column corresponds to one of the stimuli, and each row to one of the paired comparisons, and $\mathbf{D} = \left(\mathrm{Diag}\left(\mathbf{A}_n\boldsymbol{\Sigma}\mathbf{A}_n'\right)\right)^{-\frac{1}{2}}$.

Thurstone's Case V model is a popular restricted form of Thurstone's general model in which it is assumed that $\boldsymbol{\Sigma} = \sigma^2\mathbf{I}$. Not all parameters of model (2) subject to the Case V restrictions are identified. Given the comparative nature of the data, we need to introduce one restriction among the elements of $\boldsymbol{\mu}$. Also, σ^2 is not identified. To identify Thurstone's Case V ranking model it is convenient to set $\mu_n = 0$ and $\sigma^2 = ½$. With these identification restrictions, (3) simplifies to

$$\boldsymbol{\tau} = -\mathbf{A}_n\boldsymbol{\mu}, \tag{4}$$

$$\boldsymbol{P} = \frac{1}{2}\mathbf{A}_n\mathbf{A}_n', \tag{5}$$

where $\mathbf{P}$ is a correlation matrix whose elements, $\rho_{ll'}$, can take the values ½, -½, or 0.

Now, since each of the variables $\mathbf{y}$ is binary, their mean is simply $\pi_l := \Pr(y_l = 1)$ which we shall denote by π_l. Under Thurstone's Case V model,

$$\pi_l := \Pr\left(y_l = 1\right) = \int_{\tau_l}^{\infty} \phi_1\left(z_l^* : 0,1\right) dz_l^* = \Phi_1\left(-\tau_l\right), \tag{6}$$

where $\Phi_1(\bullet)$ denotes a univariate standard normal distribution function. Let $\boldsymbol{\pi}$ and $\mathbf{p}$ denote the $\tilde{n}$ dimensional vectors obtained by stacking all population means and sample means, respectively, in lexicographic order. We observe in (6) that the relation between τ and $\boldsymbol{\pi}$ is one-to-one.

Let $\tilde{\boldsymbol{\mu}}' = \left(\mu_1, \cdots, \mu_{n-1}\right)$ denote the vector of identified parameters in Thurstone's Case V model for ranking data. Then, (4) can be written as $\boldsymbol{\tau} = -\mathbf{K}\tilde{\boldsymbol{\mu}}$ where, $\mathbf{K}$ is a $\tilde{n} \times (n-1)$ matrix of full column rank obtained by deleting the last column of $\mathbf{A}_n$. Thus, Thurstone's Case V model for ranking data can be estimated as a mean structure model, as the means $\boldsymbol{\pi}$ of the binary variables $\mathbf{y}$ suffice to identify this model. The simplest estimation approach for this mean structure is (Mosteller, 1951a; Torgerson, 1958) to first estimate each threshold τ_l separately from the sample mean p_l using $\hat{\tau}_l = -\Phi_1^{-1}\left(p_l\right)$ and then estimate the model parameters $\tilde{\boldsymbol{\mu}}$ by least squares as

$$\hat{\tilde{\boldsymbol{\mu}}} = -\mathbf{H}\hat{\boldsymbol{\tau}}\,, \qquad \mathbf{H} = \left(\mathbf{K}'\mathbf{K}\right)^{-1}\mathbf{K}'\,. \tag{7}$$

We shall next provide asymptotic standard errors for these estimates of $\tilde{\boldsymbol{\mu}}$, as well as a goodness of fit test of the model.

3. Standard errors and goodness of fit tests

Let $\mathbf{e} := (\mathbf{p} - \boldsymbol{\pi})$. First, we notice that $\sqrt{N}\mathbf{e} \xrightarrow{d} N(\mathbf{0}, \boldsymbol{\Gamma})$ where $\xrightarrow{d}$ denotes convergence in distribution. $\boldsymbol{\Gamma}$ has diagonal elements $N\,\mathrm{Var}\left(\pi_l\right) = \pi_l\left(1 - \pi_l\right)$ and off diagonal elements $N\,\mathrm{Cov}\left(\pi_l, \pi_{l'}\right) = \pi_{ll'} - \pi_l\pi_{l'}$. Under the model,

$$\pi_{ll'} := \Pr\left[\left(y_l = 1\right) \cap \left(y_l = 1\right)\right] = \int_{\tau_l}^{\infty}\int_{\tau_{l'}}^{\infty} \phi_2\left(z_l^*, z_{l'}^* : 0, 0, 1, 1, \rho_{ll'}\right) dz_l^* dz_{l'}^* = \Phi_2\left(-\tau_l, -\tau_{l'}, \rho_{ll'}\right),$$

where $\rho_{ll'}$ denotes an element of $\mathbf{P}$ in (5). Thus, the sample means $\mathbf{p}$ need not be statistically independent.

We shall now obtain the asymptotic distribution of $\hat{\tilde{\boldsymbol{\mu}}}$ in (7). Let $\boldsymbol{\Delta} = \dfrac{\partial \boldsymbol{\pi}}{\partial \boldsymbol{\tau}'}$ be a diagonal matrix with elements $-\phi_1\left(\tau_i : 0, 1\right)$. By the multivariate delta theorem, $\sqrt{N}\left(\hat{\boldsymbol{\tau}} - \boldsymbol{\tau}\right) \stackrel{a}{=} \boldsymbol{\Delta}^{-1}\sqrt{N}\mathbf{e}$, where $\stackrel{a}{=}$ denotes asymptotic equality. Coupling this result with (7) $\sqrt{N}\left(\hat{\tilde{\boldsymbol{\mu}}} - \tilde{\boldsymbol{\mu}}\right) \stackrel{a}{=} -\mathbf{H}\boldsymbol{\Delta}^{-1}\sqrt{N}\mathbf{e}$. Finally, $\sqrt{N}\left(\hat{\tilde{\boldsymbol{\mu}}} - \tilde{\boldsymbol{\mu}}\right) \xrightarrow{d} N\left(\mathbf{0}, \mathbf{H}\boldsymbol{\Delta}^{-1}\boldsymbol{\Gamma}\boldsymbol{\Delta}^{-1}\mathbf{H}'\right)$. Thus, letting $\hat{\boldsymbol{\Delta}}$ and $\hat{\boldsymbol{\Gamma}}$ denote $\boldsymbol{\Delta}$ and $\boldsymbol{\Gamma}$ evaluated at $\hat{\tilde{\boldsymbol{\mu}}}$,

$$\mathrm{SE}\left(\hat{\tilde{\boldsymbol{\mu}}}\right) = \sqrt{\mathrm{VecDiag}\left(\mathbf{H}\hat{\boldsymbol{\Delta}}^{-1}\hat{\boldsymbol{\Gamma}}\hat{\boldsymbol{\Delta}}^{-1}\mathbf{H}'\right)/N} \tag{8}$$

To test the goodness of fit of the model, consider the residual vector $\hat{\mathbf{e}} := (\mathbf{p} - \hat{\boldsymbol{\pi}})$, where $\hat{\boldsymbol{\pi}} := \boldsymbol{\pi}\left(\hat{\tilde{\boldsymbol{\mu}}}\right)$. By a Taylor expansion, $\sqrt{N}\left(\hat{\boldsymbol{\pi}} - \boldsymbol{\pi}\right) \stackrel{a}{=} -\boldsymbol{\Delta}\mathbf{K}\sqrt{N}\left(\hat{\tilde{\boldsymbol{\mu}}} - \tilde{\boldsymbol{\mu}}\right)$. Since

$\hat{\mathbf{e}} = \mathbf{e} - (\hat{\boldsymbol{\pi}} - \boldsymbol{\pi})$, $\sqrt{N}\hat{\mathbf{e}} \stackrel{a}{=} \left(\mathbf{I} + \boldsymbol{\Delta}\mathbf{K}\mathbf{H}\boldsymbol{\Delta}^{-1}\right)\sqrt{N}\mathbf{e}$. Let now $\tilde{\boldsymbol{\Delta}}_c$ be an orthogonal complement to $\tilde{\boldsymbol{\Delta}} := \boldsymbol{\Delta}\mathbf{K}$, that is, $\tilde{\boldsymbol{\Delta}}_c \tilde{\boldsymbol{\Delta}} = \mathbf{0}$. Then, $\tilde{\boldsymbol{\Delta}}_c \sqrt{N}\hat{\mathbf{e}} \stackrel{a}{=} \tilde{\boldsymbol{\Delta}}_c \sqrt{N}\mathbf{e}$, and $\tilde{\boldsymbol{\Delta}}_c \sqrt{N}\mathbf{e} \xrightarrow{d} N\left(\mathbf{0}, \tilde{\boldsymbol{\Delta}}_c \boldsymbol{\Gamma} \tilde{\boldsymbol{\Delta}}_c'\right)$. Thus,

$$T_B = N\hat{\mathbf{e}}' \hat{\tilde{\boldsymbol{\Delta}}}_c{}' \left(\hat{\tilde{\boldsymbol{\Delta}}}_c \hat{\boldsymbol{\Gamma}} \hat{\tilde{\boldsymbol{\Delta}}}_c{}'\right)^{-1} \hat{\tilde{\boldsymbol{\Delta}}}_c \hat{\mathbf{e}} \xrightarrow{d} \chi^2_{\tilde{n}-n+1}. \tag{9}$$

Now, partition the $\tilde{n} \times \tilde{n}$ diagonal matrix $\boldsymbol{\Delta}$ as $\boldsymbol{\Delta} = \begin{pmatrix} \boldsymbol{\Delta}_{n-1} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Delta}_{\tilde{n}-n+1} \end{pmatrix}$. Then it can be readily verified that

$$\tilde{\boldsymbol{\Delta}}_c = \left(\boldsymbol{\Delta}_{\tilde{n}-n+1}\mathbf{A}_{n-1}\boldsymbol{\Delta}^{-1}_{n-1} \;\vdots\; \mathbf{I}_{\tilde{n}-n+1}\right). \tag{10}$$

Then, $\hat{\tilde{\boldsymbol{\Delta}}}_c$ is simply obtained by evaluating (10) at $\hat{\boldsymbol{\mu}}$.

3. Some numerical results

To illustrate the present discussion, we asked 56 Psychology undergraduate students to express their preferences using a ranking experiment for the following Psychology career areas {A = Academic, C = Clinical, E = Educational, and I = Industrial}. Their rankings were transformed to 6 paired comparisons variables using (1). The proportion of respondents that chose the first career in each of the following pairs {{A, C}, {A, E}, {A, I}, {C, E}, {C, I}, {E, I}} was $\mathbf{p}'$ = (4, 6, 15, 41, 41, 31)/56. We estimated Thurstone's Case V ranking model from these univariate proportions obtaining T_b = 6.03 on 3 d.f., p = 0.11. Thus, the model reproduces well these proportions. The estimated mean preferences with asymptotic standard errors in parentheses were $\hat{\mu}_A$ = -0.80 (0.17), $\hat{\mu}_C$ = 0.71 (0.17), and $\hat{\mu}_E$ = 0.22 (0.16). The mean preference for an industrial career was fixed to zero for identification purposes. Thus, we conclude that the most preferred career path among these Psychology undergraduates is in Clinical Psychology, followed by Educational, then Industrial, and finally in Academia. Given that the model assumes that $\boldsymbol{\Sigma} = \sigma^2\mathbf{I}$, we also conclude from these results that the preferences for these career areas may be independent. They need not be independent, however, to be consistent with the data. Tsai (2000) showed that there is a set of covariance structures on $\boldsymbol{\Sigma}$ that do not assume that preferences are independent (i.e. where $\boldsymbol{\Sigma}$ is not diagonal) that yield equivalent models to Thurstone's Case V ranking model in the sense as providing the same set of probabilities (2).

The sample sized used in this example was rather small. Thus, it is of interest to investigate whether the standard errors and goodness of fit tests obtained using asymptotic theory can be trusted when only small samples are available. To investigate the small sample performance of the procedures described in this paper we performed a simulation study. We generated 1000 replications with sample size $N = 50$ from a Thurstone's Case V ranking model for $n = 10$ stimuli. The values used to generate the data were similar to those estimated in the numerical example, $\boldsymbol{\mu}' = (-.8,.7,.2,-.8,\cdots,.2,0)$, $\sigma^2 = ½$. The results suggest that adequate parameter estimates, standard errors and goodness of fit tests for this model can be obtained with as few as 50 observations. The relative bias of the parameter estimates ranged from 0% to 4%, and the relative bias of the standard errors ranged from -3% to 3%. Furthermore, with 36 degrees of freedom available for testing, the mean and variance of the T_b statistic across replications were 35.8 and 69.3. We also computed a Kolmogorov-Smirnoff one-sample test to investigate the match of the empirical distribution of the T_b statistic to its reference chi-square distribution, obtaining $D_{KS} = 0.80$ which is less than the critical value of 1.35 at the 5% level of significance.

4. Concluding remarks

We have introduced formulae for obtaining standard errors and goodness of fit tests when Thurstone's Case V model for ranking data (Thurstone, 1931) is estimated using the classical least squares procedure for paired comparisons data (Mosteller, 1951a; Torgerson, 1958). Our proposed test statistic is analogous to a statistic proposed by Browne (1984) in the context of covariance structure analysis. The classical least squares estimation procedure is a computationally very attractive procedure to estimate Thurstone's Case V ranking model. Our small simulation study reveals that adequate parameter estimates, standard errors and goodness of fit tests can be obtained for large models with very few observations.

Mosteller (1951b) introduced a goodness of fit test for this estimation procedure that should not be used when the paired comparisons are obtained from ranking patterns. This is because Mosteller's test assumes that the paired comparisons are statistically independent, an assumption that is violated in the case of ranking data. When applied to ranking data, Mosteller's test is overly optimistic. For instance, the value of Mosteller's test for the numerical example presented here is 0.59 on 3 d.f., $p = 0.90$. Mosteller's test should not be applied to paired comparisons data when the individuals respond to all paired comparisons (i.e., multiple judgment paired comparisons) as again the paired comparisons are not statistically independent.

For that matter, the classical estimation procedure described here should not be used to estimate Thurstone's Case V model from multiple judgment paired comparisons data as in this case σ^2 is identified, but it is not identified from univariate information alone. To estimate Thurstone's Case V model from multiple judgment paired comparisons data one must at least use univariate and bivariate information on the paired comparisons (Maydeu-Olivares, 2001). Estimation methods for Thurstonian models using univariate and bivariate information are available using Mplus (Muthén & Muthén, 2001). For an overview see Maydeu-Olivares and Böckenholt (2005).

References

Browne, M.W. (1984). Asymptotically distribution free methods for the analysis of covariance structures. British Journal of Mathematical and Statistical Psychology, 37, 62-83.

Maydeu-Olivares, A. (1999). Thurstonian modeling of ranking data via mean and covariance structure analysis. Psychometrika, 64, 325-340.

Maydeu-Olivares, A. (2001). Limited information estimation and testing of Thurstonian models for paired comparison data under multiple judgment sampling. Psychometrika, 66, 209-228.

Maydeu-Olivares, A. & Böckenholt, U. (2005). Structural equation modeling of paired comparisons and ranking data. Psychological Methods, 10, 285-304.

Mosteller, F. (1951a). Remarks on the method of paired comparisons: I. The least squares solution assuming equal standard deviations and equal correlations. Psychometrika, 16, 3-9.

Mosteller, F. (1951b). Remarks on the method of paired comparisons: III. A test of significance for paired comparisons when equal standard deviations and equal correlations are assumed. Psychometrika, 16, 207-218.

Muthén, L. & Muthén, B. (2001). Mplus. Los Angeles, CA: Muthén & Muthén.

Torgerson, W.S. (1958). Theory and methods of scaling. New York: Wiley.

Thurstone, L.L. (1927). A law of comparative judgment. Psychological Review, 79, 281-299.

Thurstone, L.L. (1931). Rank order as a psychological method. Journal of Experimental Psychology, 14, 187-201.

Tsai, R.-C. (2000). Remarks on the identifiability of Thurstonian ranking models: Case V, Case III or neither? Psychometrika, 75, 233-240.