Brazilian Journal of Probability and Statistics (2002), 16, pp. 123–140.

THE LAW OF CATEGORICAL JUDGMENT REVISITED

Geraldo da Silva e Souza

Universidade de Brasília, Departamento de Estatística, 70910-990, Brasília, DF, Brazil. E-mail: Geraldo.Souza@embrapa.br

Summary
A referee, typical of a population, classifies each of $r \geq 2$ stimuli into one of $m \geq 2$ response categories which are ordered and mutually exclusive. The Law of Categorical Judgment specifies a set of equations relating stimuli and categories parameters to the probabilities of classification of stimuli into categories. This paper proposes the use of restricted maximum likelihood or restricted generalized least squares to estimate and test different parametric formulations of the Law of Categorical Judgment as well as a method to assess the relative importance of the stimuli. Both approaches allow the use of distributions other than the standard normal to model classification probabilities.

Key words: Categorical data analysis; generalized linear models; nonlinear models; Thurstone's law of categorical judgment.

1 Introduction

Consider a set of $r \geq 2$ stimuli $S = \{S_1, \ldots, S_r\}$ and a set of $m \geq 2$ categories $C = \{C_1, \ldots, C_m\}$. A referee or judge, randomly chosen from a population, is to classify each stimulus S_i into one of the categories C_j. The categories in C are mutually exclusive and ordered according to an underlying carachteristic of interest. In this context $C_1 < C_2 < \ldots < C_m$ represents the ordination in C, that is, relative to the characteristics of interest C_1 represents the least intense impulses and C_m the most intense impulses. Data sets generated from such processes are known as polytomous data with measurements on an ordinal scale. They are very common in biological, econometric, social, psychometric, and administrative work. See Souza and Ávila (2000), Rousseu et al (1999), Macedo (1997), Turoff and Hiltz (1996), Sousa (1993), and MacCullagh and Nelder (1989). A typical example obtains when each element of a sample of individuals, taken from a certain population, is asked to manifest his opinion, relative to some criterion, of each activity in a set of interest. Such stimuli could be research projects, research units of an institution or a set of actions for which we want to evaluate each element on a psychometric scale. Nelder and MacCullagh (1989, page 175) provides a simple instance of such a process where the stimuli are four cheese types (A,B, C, and D). The referee (taster) manifests the intensity relative to which he likes or dislikes a given cheese type in a ordinal scale from 1 to 10, where 1 means 'strong dislike' and 10 means 'excellent taste'. For those situations Thurstone (1927) proposes a general judgment model from which is possible to derive a set of equations relating stimuli and category parameters to frequencies in the contingency table of the referees evaluations. This set of equations is known as the Law of Categorical Judgment. The resulting statistical model, in its linear version, falls in the class of multinomial responses models which Nelder and MacCullagh (1989) discuss in Chapter 5.

Thurstonian models although originated a long time ago have been object of continous use and research. In this context it is worth mentioning the works of McFadden (1974, 2001) and Madeu-Olivares (1999, 2000), which discuss the use of Thurstone's proposal to model preferences in economics and psychology respectively. The best reference in regard to the philosophycal and mathematical aspects of the Law of Categorical Judgment is still Torgerson (1958). Some useful insights may also be obtained from Maydeu-Olivares (2000), Saaty (1994), Kotz and Johnson (1989), and Souza (1988).

In this paper we show how the Law of Categorical Judgment can be put into a framework similar to the one used by Grizzle, Stamer, and Koch (1969) in categorical data analysis and, alternatively, as a generalized nonlinear multinomial response model that can be analyzed via maximum likelihood. The statistical inference that follows is flexible enough to estimate stimulus and category and to test Thurstone's Law of Categorical Judgment as well. The approach adopted here, although well known in the statiscal literature, to the best knowledge of the author, does not appear in the standard psychometric literature and in the majority of the applications of the Law of Categorical Judgment which, typically, base analyses on the method of moments estimation.

The discussion carried out in the paper proceeds as follows. Section 2 deals with Thurstone's theory which leads to the Law of Categorical Judgment. Section 2 also shows how linear and nonlinear regression models can be put in use to fit the Law of Categorical Judgment via generalized least squares and maximum likelihood. Section 3 presents the classical approach to the Law of Categorical Judgment exploiting the discussion in Torgerson (1958). Section 4 illustrates Thurstone's theory with an application exploring some features of the statistical package SAS. Section 5 shows how to obtain a set of weights summing to one that serves the purpose of ranking stimuli. This approach is original and is competitive with Saaty's (1994) Analytical Hierarchy Process and the companion Thurstone's Law of Comparative Judgments when the number of stimuli and the size of the referees set are too large so that pairwise judgments become a nuisance to record and control. Finally in Section 6 a summary of the paper is presented including a brief discussion on the assumption of independence of stimuli judgments and the general applicability of the Law of Categorical Judgment.

2 The law of categorical judgment

The psychometric model proposed by Thurstone (1927) postulates the presence of a psychological continuum. Each time a referee faces a stimulus, a mental discriminal process is put into action and it generates a numerical value in the real line reflecting the stimulus intensity. Therefore, in this way, the stimuli translate in the psychological continuum into scale values $\mu_1, \ldots, \mu_r$. Likewise the categories translate into location values $\tau_1, \ldots, \tau_{m-1}$. These later quantities form a partition of the real line $(-\infty, \tau_1], (\tau_1, \tau_2], \ldots, (\tau_{m-1}, +\infty)$. The partition relates to stimuli S_i and categories C_j according to the following rule. The referee classifies stimulus S_i into $\bigcup_{l=1}^{j} C_l$ if and only if $\mu_i \leq \tau_j$. The process inherits randomness from the sampling scheme and from the fact that due to stochastic fluctuations in nature, a given stimulus and category when repeatedly evaluated by a referee do not generate the same scale and boundary values in the psychological continuum. Randomness leads one to assume that the μ_i are indeed means of random variables ξ_i with variance σ_i^2 and that τ_j are indeed means of random variables η_j with variances ϕ_j^2. The discussion imposes row independence and joint normality, that is, the ξ_i are uncorrelated and (ξ_i, η_j) are jointly normally distributed. In principle, one has primary interest in the differences $\mu_i - \mu_j$. These quantities may serve the purpose of assessing differences in intensity between stimuli. Section 5 treats the problem of measuring intensity and of ranking stimuli in more detail and offers an alternative and equivalent approach to measure differences in intensity.

Let π_{ij} denote the probability of locating stimulus S_i into one of the first j categories $C_1, C_2, \ldots, C_j$. We assume $\pi_{ij} > 0$. We have,

$$\begin{aligned} P\left\{S_i \in \bigcup_{l=1}^{j} C_l\right\} &= \pi_{ij} \;\; i = 1, \ldots, r, \;\; j = 1, \ldots, m-1. \\ &= P\{\xi_i \leq \eta_j\} \\ &= P\left\{Z \leq -\frac{\mu_i - \tau_j}{\sqrt{\mathrm{Var}\,(\xi_i - \eta_j)}}\right\} \end{aligned}$$

Let $g(.)$ denote the probit transformation. The assumption of joint normality lead to the equations

$$g(\pi_{ij}) = -\frac{\mu_i - \tau_j}{\sqrt{\mathrm{Var}\,(\xi_i - \eta_j)}} \;\; i = 1, \ldots, r \;\; j = 1, \ldots, m-1 \tag{2.1}$$

relating the cumulative probabilities π_{ij} to the parameters of Thurstone's model . Clearly it is possible to generalize the normal projection on the psychological continuum to other distributions. Any monotonic function may play the role of g(.). Typical alternatives in this context would be the logistic scale $g(x) = \ln\{x/(1-x)\}$ and the log-log scale $g(x) = \ln\{-\ln(1-x)\}$.

Suppose first that enough observations are available to estimate the probabilities π_{ij} in (2.1). In this context a sample version of the Law of Categorical Judgment is therefore

$$g(\hat{\pi}_{ij}) = -\frac{\mu_i - \tau_j}{\sqrt{\mathrm{Var}\,(\xi_i - \eta_j)}} + u_{ij} \quad i = 1, \dots, r \quad j = 1, \dots, m-1 \qquad (2.2)$$

where $\hat{\pi}_{ij}$ is the relative cumulative frequency of observations in category C_j. The vectors $u_i' = (u_{i1}, \dots, u_{im-1})$ are independently distributed with a distinct variance matrix for each i. Clearly,

$$\hat{\pi}_{ij} = \hat{p}_{i1} + \hat{p}_{i2} + \dots + \hat{p}_{ij}$$

where $\hat{p}_{il}$ represents the proportion of times the referees (sample) classify stimulus S_i into C_l. Let

$$G(\hat{\pi}) = (G_1'(\hat{\pi}_1), \dots, G_r'(\hat{\pi}_r))', \;\; \hat{\pi}' = (\hat{\pi}_1, \dots, \hat{\pi}_r)' \qquad (2.3)$$

where $G(\hat{\pi})$ is the response vector, $\hat{\pi}_i = (\hat{\pi}_{i1}, \dots, \hat{\pi}_{im-1})'$ and $G_i(\hat{\pi}_i)$ is the subvector of $G(\hat{\pi})$ formed with the quantities $g(\hat{\pi}_{ij})$, $j = 1, \dots, m-1$. The first order Taylor's expansion of $G_i(\hat{\pi}_i)$ about the true parameter $\pi_i = (p_{i1}, \dots, p_{im-1})'$ yields

$$G_i(\hat{\pi}_i) = G_i(\pi_i) + \qquad (2.4)$$
$$\begin{pmatrix} g'(\pi_{i1}) & 0 & 0 & \dots & 0 \\ g'(\pi_{i2}) & g'(\pi_{i2}) & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ g'(\pi_{im-1}) & g'(\pi_{im-1}) & \dots & \dots & g'(\pi_{im-1}) \end{pmatrix} \begin{pmatrix} \hat{p}_{i1} - p_{i1} \\ \vdots \\ \hat{p}_{im-1} - p_{im-1} \end{pmatrix}$$

where

$$g'(\pi_{ij}) = \sqrt{2\pi} \exp\left\{\frac{g^2(\pi_{ij})}{2}\right\}$$

for the probit link function,

$$g'(\pi_{ij}) = \frac{1}{\pi_{ij}(1 - \pi_j)}$$

for the logistic, and

$$g'(\pi_{ij}) = \frac{1}{\pi_{ij} \ln(1 - \pi_{ij})}$$

for the log-log scale.

Let H_i denote the lower triangular matrix in (2.5) and let V_i denote the variance matrix of $\hat{\pi}_i$. It is reasonable to assume that in the regression model (2.2) the residual vector has mean zero and variance

$$V = \text{diag}(H_1 V_1 H_1', \dots, H_r V_r H_r') \tag{2.5}$$

which can be estimated by

$$\hat{V} = \text{diag}(\hat{H}_1 \hat{V}_1 \hat{H}_1', \dots, \hat{H}_r \hat{V}_r \hat{H}_r') \tag{2.6}$$

using the quantities $\hat{p}_{ij}$ and $\hat{\pi}_{ij}$ to replace p_{ij} and π_{ij}, respectively.

With the level of generality above, the classification law that Thurstone proposes is not identifiable. However, depending on the assumptions regarding the components $\text{Var}(\xi_i - \eta_j)$, a solution to the Law of Categorical Judgment is viable. This section considers three distinct models which Torgenson (1958) labels Models B, C, and D. Models B and C are nonlinear. We begin our discussion with the simplest Model D.

2.1 Model D

Model D assumes $\text{Var}(\xi_i - \eta_j) = 1$ for any pair (i, j). Thus one obtains from (2.2)

$$E(g(\hat{\pi}_{ij})) = \tau_j - \mu_i \tag{2.7}$$

or, in matrix form

$$E(G(\hat{\pi})) = \begin{pmatrix} 1 & 0 & 0 & \dots & 0 & -1 & 0 & \dots & 0 \\ 0 & 1 & 0 & \dots & 0 & -1 & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & \dots & 0 & 0 & 1 & -1 & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & & \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & 0 & 0 & \dots & 0 & 0 & 0 & \dots & -1 \\ 0 & 1 & 0 & \dots & 0 & 0 & 0 & \dots & -1 \\ \vdots & \vdots & \vdots & & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & \dots & 0 & 0 & 1 & 0 & 0 & \dots & -1 \end{pmatrix} \begin{pmatrix} \tau_1 \\ \vdots \\ \tau_{m-1} \\ \mu_1 \\ \vdots \\ \mu_r \end{pmatrix} \tag{2.8}$$

Model D in (2.7) is not identified since the design matrix in (2.8) is not of full column rank. However all contrasts involving the μ_i are estimable. Since we are primarily interested in pairwise comparisons we may, without any loss of generality, impose the condition

$$\sum_i \mu_i = 0 \tag{2.9}$$

Under the restriction (2.9) the appopriate estimation process is restricted generalized least squares. The goodness of fit of the model may be assessed using the residual sum of squares and the chi-square test, with $(r-1)(m-2)$ degrees of freedom, proposed by Grizzle, Stamer, and Koch (1969).

2.2 Model B

Model B assumes $\text{Var}(\xi_i - \eta_j) = \delta_i$. From (2.2) this assumption generates the nonlinear regression model

$$g(\hat{\pi}_{ij}) = -\frac{\mu_i - \tau_j}{\delta_i} + \epsilon_{ij} \tag{2.10}$$

Notice that we must have $2r + m - 3 \leq r(m-1)$, i.e, the number of parameters should be at most the number of observations. The number of parameters is adjusted for two identifying restrictions:

$$\sum_{i=1}^{r} \frac{1}{\delta_i} = r, \text{ and } \sum_{i=1}^{r} \frac{\mu_i}{\delta_i} = 0. \tag{2.11}$$

Conditions in (2.10) are Torgerson (1958) restrictions. They generalize (2.9) imposed in the linear case. Alternative sets of restrictions have been suggested in the literature. Torgenson (1958), for example, mimics Gulliksen (1954) and imposes $\sum_j \tau_j = 0$ and $\sum_j \tau_j^2 = m - 1$. This section considers only the set of restrictions (2.11) since they seem to be more natural as a generalization of Model D ($\delta_i's = 1$). In some cases, however, Gulliksen's type restrictions may be easier to impose. It is worth mentioning that Model B is analogous to Equation 5.4 of McCullagh and Nelder (1989, p. 154) with the reparametrization $\delta_i = \exp(\omega_i)$.

Let $\eta_i = 1/\delta_i$, $\gamma_i = \mu_i/\delta_i$, and

$$\beta = (\eta_1\tau_1, \ldots, \eta_1\tau_{m-1}, \ldots, \eta_r\tau_1, \ldots, \eta_r\tau_{m-1}, \gamma_1, \ldots, \gamma_r)'$$

Then (2.10) can be written equivalently as

$$E(G(\hat{\pi})) = [I, A]\beta \tag{2.12}$$

where I is the identity of order $r(m-1)$ and A is the vertical concatenation of the r blocks

$$\begin{pmatrix} -1 & 0 & \ldots & 0 \\ \vdots & \vdots & & \vdots \\ -1 & 0 & \ldots & 0 \end{pmatrix}, \ldots, \begin{pmatrix} 0 & 0 & \ldots & -1 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \ldots & -1 \end{pmatrix}$$

each of dimension $(m-1)r$.

One can estimate Model B using generalized restricted nonlinear least squares. As before one may assess goodness of fit using the chi-square test of Grizzle, Stamer, and Koch (1969).

We may obtain convenient initial values for the nonlinear estimation making $\delta_i = 1$ and using Model D estimates for the parameters μ_i and τ_j.

2.3 Model C

Model C postulates that $\sqrt{\mathrm{Var}(\xi_i - \eta_j)} = \theta_j$. From (2.2) this assumption generates the nonlinear regression model

$$g(\hat{\pi}_{ij}) = -\frac{\mu_i - \tau_j}{\theta_j} + u_{ij} \tag{2.13}$$

The following two identifying restrictions are required:

$$\sum_{j=1}^{m-1} \frac{1}{\theta_j} = m - 1 \;\text{ and }\; \sum_{j=1}^{m-1} \frac{\tau_j}{\theta_j} = 0. \tag{2.14}$$

Restrictions (2.14) are dual of restrictions (2.11).

Obviously Model C also generalizes Model D although no restrictions are necessary to be imposed on the scale values μ_i. Notice that the effective number of parameters $r + 2(m-1)$ should be less then or equal to $r(m-1)$.

Let $\eta_j = 1/\theta_j$ and $\gamma_j = \tau_j/\theta_j$. Then from (2.10)

$$g(\pi_{ij}) = -\mu_i \eta_j + \gamma_j + u_{ij}$$

It follows that (2.13) is equivalent to

$$E(G(\hat{\pi})) = [-I, A]\,\beta \tag{2.15}$$

where

$$\beta = (\mu_1\eta_1, \ldots, \mu_1\eta_{m-1}, \ldots, \mu_r\eta_1, \ldots, \mu_r\eta_{m-1}, \gamma_1, \ldots, \gamma_{m-1})'$$

and A is the vertical concatenation of r identical blocks each being the identity of order $m - 1$.

We may obtain initial values for the nonlinear estimation of Model C making $\theta_j = 1$, using for μ_i the negative of the mean of the ith row of the matrix $(g(\hat{\pi}_{ij}))$, and using for τ_j the deviation of the mean of column j relative to the overall mean of the matrix.

If not enough replications are available to estimate all π_{ij} one may want to appeal to maximum likelihood instead of generalized least squares. Such is the case when some cells show zero frequencies and the usual link functions and their derivatives are not defined at zero. We should warn the reader however that sparse contigency tables may pose estimation problems to both methods. McCullagh and Nelder (1989) refer to sparseness when a large proportion of observed counts are small, that is, less than 5. These instances as well as the extreme case of zero frequencies may be solved partially by adding small constants to the cell frequencies. A few methods of doing this are given in Forthofer and Lehnen (1981).

For the maximum likelihood approach this section assumes the row totals m_i to be fixed. Let y_{ij} denote the frequency in cell (i, j). The total log-likelihood for the table is

$$\sum_{i=1}^{r} \ln\left(\frac{m_i!}{y_{i1}! \dots y_{im}!}\right) + \sum_{i=1}^{r}\sum_{j=1}^{m} y_{ij} \ln(p_{ij}) \tag{2.16}$$

where $p_{im} = 1 - \sum_{j=1}^{m-1} p_{ij}$ for every row i and, for Model B to fix ideas,

$$p_{i1} = g^{-1}\left(\frac{\tau_1 - \mu_i}{\delta_i}\right) \text{ and } p_{ij} = g^{-1}\left(\frac{\tau_j - \mu_i}{\delta_i}\right) - g^{-1}\left(\frac{\tau_{j-1} - \mu_i}{\delta_i}\right), \tag{2.17}$$

$j = 2, \dots, m-1$.

We seek to maximize the quantity $\sum_{i=1}^{r}\sum_{j=1}^{m} y_{ij} \ln(p_{ij})$ in (2.16), which from now on we refer simply as the log-likelihood, with respect to μ_i, δ_i, and τ_j, subject to (2.17) and Torgenson's restrictions.

Standard maximum likelihood theory applies to estimation and hypothesis testing and the analysis of deviance (McCullagh and Nelder, 1989) is the key to the assessment of goodness of fit and to find the best model, if any, fitting the data.

3 The classical approach to the law of categorical judgment

The classical estimation approach to the Law of Categorical Judgment is restricted to finding solutions to the set of equations (2.1) for the unknown parameters when one replaces π_{ij} with $\hat{\pi}_{ij}$, under the assumptions underlying Models D, B or C. The approach is basically derived from the method of moments. One computes these estimates as follow. Suppose that there are no zero frequencies in any cell and let $Z = (\hat{z}_{ij})$ be the matrix with the probit, or any other link function, values $\hat{z}_{ij} = g(\hat{\pi}_{ij})$. Let $\bar{z}_{i.}$ and $\tilde{z}_{i.}$ denote the mean and the standard error of row i respectively. Let $\bar{z}_{.j}$ and $\tilde{z}_{.j}$ denote these same quantities for column j. Let d be the standard error of the column means $\bar{z}_{.j}$, let e be the standard error of the row means $\bar{z}_{i.}$, and let $\bar{z}_{..}$ be the overall mean of Z.

For Model D, the solution for (2.1) subject to (2.9) is given by

$$\hat{\mu}_i = -\bar{z}_{i.} + \bar{z}_{..}, \quad \hat{\tau}_j = \bar{z}_{.j}. \tag{3.1}$$

This solution also minimizes the residual sum of squares for model (2.8) assuming ordinary least squares.

For Model B, the solution for (2.1) subject to (2.11) is given by

$$\hat{\tau}_j = \bar{z}_{.j}, \quad \hat{\delta}_i = d/\tilde{z}_{i.}, \quad \hat{\mu}_i = \bar{z}_{..} - \hat{\delta}_i \bar{z}_{i.} \tag{3.2}$$

Notice that when $2r + m - 3 = r(m-1)$ the generalized least squares estimates will coincide with the method of moments solution since the residual sum of squares function will be zero when evaluated at the method of moments estimates.

The method of moments estimates for Model C are dual of those of Model B. They are

$$\hat{\theta}_j = e/\tilde{z}.j, \quad \hat{\mu}_i = -\bar{z}_{i.}, \quad \hat{\tau}_j = -\bar{z}_{..} + \hat{\theta}_j \bar{z}_{.j} \tag{3.3}$$

When $r + 2(m-1) = r(m-1)$ generalized least squares and method of moments estimates will be equal.

Variances for the method of moments estimates (3.1), (3.2), and (3.3) can be computed using the fact that the estimates are all functions of $G(\hat{\pi})$ with $G(.)$ as in (2.3). If one uses the first order Taylor's series expansion of the estimates about the true values $z_{ij} = g(\pi_{ij})$, the variance matrix estimates are given by expressions of the form $L\hat{V}L'$ where L is a matrix with each row defined by a gradient vector and $\hat{V}$ as in (2.5). Let π be the population parameter vector corresponding to $\hat{\pi}$ of (2.3). For Model D, L has rows $\partial\hat{\mu}_i/\partial\pi'$ and $\partial\hat{\tau}_j/\partial\pi'$. The typical elements of these gradients are

$$\frac{\partial \hat{\mu}_i}{\partial z_{\nu l}} = \begin{cases} -\frac{r-1}{r(m-1)} & \text{if } i = \nu \\ \frac{1}{r(m-1)} & \text{if } i \neq \nu \end{cases} \tag{3.4}$$

and

$$\frac{\partial \hat{\tau}_j}{\partial z_{\nu l}} = \begin{cases} \frac{1}{r} & \text{if } j = l \\ 0 & \text{if } j \neq l \end{cases} \tag{3.5}$$

respectively. For Model B, L has rows $\partial\hat{\mu}_i/\partial\pi'$, $\partial\hat{\delta}_i/\partial\pi'$ and $\partial\hat{\tau}_j/\partial\pi'$. The typical elements of these vectors are

$$\frac{\partial \hat{\mu}_i}{\partial z_{\nu l}} = \begin{cases} \frac{1}{r(m-1)} - \frac{\hat{\delta}_i}{m-1} - \bar{z}_{i.}\left[\frac{\bar{z}_{.l} - \bar{z}_{..}}{r(m-2)d\tilde{z}i.} - \frac{d(z_{il} - \bar{z}_{i.})}{(m-2)(\tilde{z}_{i.})^3}\right] & \text{if } \nu = i \\ \frac{1}{r(m-1)} - \bar{z}_{i.}\frac{\bar{z}_{.l} - \bar{z}_{..}}{r(m-2)d\tilde{z}_{i.}} & \text{if } \nu \neq i \end{cases} \tag{3.6}$$

$$\frac{\partial \hat{\delta}_i}{\partial z_{\nu l}} = \begin{cases} \frac{\bar{z}_{.l} - \bar{z}_{..}}{r(m-2)d\tilde{z}_{i.}} - \frac{d(z_{il} - \bar{z}_{i.})}{(m-2)(\tilde{z}_{i.})^3} & \text{if } \nu = i \\ \frac{\bar{z}_{.l} - \bar{z}_{..}}{r(m-2)\tilde{z}_{i.}} & \text{if } \nu \neq i \end{cases} \tag{3.7}$$

and

$$\frac{\partial \hat{\tau}_j}{\partial z_{\nu l}} = \begin{cases} \frac{1}{r} & \text{if } j = l \\ 0 & \text{if } j \neq l \end{cases} \tag{3.8}$$

respectively. For Model C, L has rows $\partial\hat{\mu}_i/\partial\pi'$, $\partial\hat{\theta}_j/\partial\pi'$ and $\partial\hat{\tau}_j/\partial\pi'$. The typical elements of these vectors are

$$\frac{\partial\hat{\mu}_i}{\partial z_{\nu l}} = \begin{cases} -\frac{1}{m-1} & \text{if } \nu = i \\ 0 & \text{if } \nu \neq i \end{cases} \tag{3.9}$$

$$\frac{\partial\hat{\theta}_j}{\partial z_{\nu l}} = \begin{cases} \frac{\tilde{z}_{.l}(\bar{z}_{\nu.}-\tilde{z}_{..})}{(r-1)(m-1)e(\tilde{z}_{.l})^2} - \frac{e(z_{\nu l}-\tilde{z}_{.l})}{(n-1)(\tilde{z}_{.l})^3} & \text{if } j = l \\ -\frac{\bar{z}_{\nu.}-\tilde{z}_{..}}{(r-1)(m-1)e\tilde{z}_{.j}} & \text{if } j \neq l \end{cases} \tag{3.10}$$

and

$$\frac{\partial\hat{\tau}_j}{\partial z_{\nu l}} = \begin{cases} -\frac{1}{r(m-1)} + \frac{\hat{\theta}_l}{r} - \tilde{z}_{.l}\left[\frac{\bar{z}_{\nu.}-\tilde{z}_{..}}{(r-1)(m-1)e\tilde{z}_{.l}} + \frac{e(z_{\nu l}-\tilde{z}_{.l})}{(r-1)(\tilde{z}_{.l})^3}\right] & \text{if } j = l \\ -\tilde{z}_{.l}\frac{\bar{z}_{\nu.}-\tilde{z}_{..}}{(r-1)(m-1)e\tilde{z}_{.j}} - \frac{1}{r(m-1)} & \text{if } j \neq l \end{cases} \tag{3.11}$$

No formal statistical testing appears in the classical approach. The measure of goodness of fit suggested in applications relates to the ability of the models to reproduce the observed cell probabilities $\hat{p}_{ij}$. It is given by the mean absolute deviation

$$\text{mad} = \frac{1}{rm}\sum_{ij} \mid \tilde{p}_{ij} - \hat{p}_{ij} \mid \tag{3.12}$$

where $\tilde{p}_{ij}$ is a model based estimate.

3.1 Computational aspects

One can compute generalized least squares for both the linear (Model D) and nonlinear cases (Models B and C) in SAS[1] (Statistical Analysis System) using PROC IML and PROC MODEL or PROC NLIN. Firstly one computes the Cholesky decomposition of $\hat{V}^{-1}$ determining a matrix R such that $\hat{V}^{-1} = R'R$. Then one uses R to rotate the model to

$$RG(\hat{\pi}) = RX\beta + Ru$$

where X is the design matrix in (2.8), (2.12) or (2.3) depending on the model under study. With the vector of transformed responses $RG(\hat{\pi})$ and the new design matrix RX we may invoke PROC MODEL or PROC NLIN to fit the regression, using restrict statements in case of PROC MODEL and imposing restrictions to the model equation in case of PROC NLIN. One begins with Model D. This will provide convenient starting values for Model B. If one is using PROC MODEL, standard errors will need adjustments since the residual mean squares is not unit. PROC NLIN

[1]A macro SAS (%Thusrt) for this purpose is available from the author.

allows the option SIGMASQ=1 and will provide correct standard errors at the cost of additional computations to determine standard errors for the parameters left out of the model in the process of imposing restrictions. For maximum likelihood estimation one notices first that the function to be maximized is negative and therefore we may find its maximum minimizing its symmetric. Thus if we define the model response as being the square root of twice the negative of the log-likelihood for each row (stimulus) and zero as the response variable, then the quantity to be minimized is precisely the residual sum of squares for the corresponding nonlinear model. One can then use PROC NLIN to compute this minimum residual sum of squares with the options SIGMASQ=1 and METHOD=NEWTON.

More powerful and also more convenient than PROC NLIN to obtain maximum likelihood estimates in SAS is PROC NLMIXED. For PROC NLMIXED the same null response variable should be used with the objective function being the likelihood function itself.

The maximum likelihood estimation for the linear version of the Law of Categorical Judgment is a particular case of the more general linear models that PROC GENMOD can handle. See the documentation of SAS-STAT version 8. The parametrization PROC GENMOD uses is the same as in McCullagh and Nelder (1989, Chapter 5) and it differs from (2.9) but will produce the same contrast estimates.

For the method of moments estimates formulas (3.1)-(3.3) and (3.4)-(3.11) can also be computed without much effort in PROC IML.

4 An example

The contingency table defined in Table 1 shows frequencies of responses to ordinal categories 1, 2, 3, 4, and 5 of 5 stimuli, A, B, C, D and E. The table is taken from Torgenson (1958, p.211).

Table 1
Frequencies of responses to stimuli

Stimuli/Response	1	2	3	4	5	Total
A	100	38	49	11	2	200
B	84	27	47	23	19	200
C	13	32	110	39	6	200
D	62	14	32	23	69	200
E	4	9	49	58	80	200
Total	263	120	287	176	176	1000

We perform our analysis of the data in Table 1 fitting Models B, C and D with the probit link function and using method of moments, generalized least squares and maximum likelihood estimation. The probit provides the best fit, although one cannot reject the alternatives logistic and log-log. Table 2 shows the values of the mean absolute deviation (mad) for each of the estimation methods. Clearly Model B provides a superior fit than Models C and D. For Model B, method of moments, generalized least squares and maximum likelihood are essentially equivalent. The latter methods are only slightly superior to the method of moments. Formal tests of specifications can be performed with generalized least squares and maximum likelihood. Table 3 shows the proper statistics to this end. Model B is the only model not rejected by the chi-square test of Grizzle, Stamer, and Koch (1969) which applies to the generalized least squares residual sum of squares. These results are in close agreement with Table 2. It is interesting to report in terms of maximum likelihood estimation that the Pearson's generalized chi-square for Model B is 0,179 and that overdispersion is not present. Also Model B is the only model with an acceptable deviance value.

Table 2
Goodness of fit statistics - mad, for method of moments, maximum likelihood (ML), and generalized least squares (GLS).

Method/Model	D	B	C
Moments	0.064	0.002	0.074
ML	0,062	0.001	0.058
GLS	0,065	0.001	0.060

Table 3
Generalized least squares (GLS) and maximum likelihood goodness of fit statistics.

Model	DF	GLS Residual SS	$-\sum_{ij} y_{ij} \ln(p_{ij})$	Deviance
D	12	170.514	1413.316	174.840
B	8	0.160	1325.978	0.164
C	9	159.369	1412.012	172.232

Estimation results for Model B are shown in Table 4. As expected, the smallest standard errors are for generalized least squares and maximum likelihood estimates.

The primary interest in the analysis of the data in Table1 1 is to rank stimuli A, B, C, D and E. From the scale values estimates $\hat{\mu}_i$ in Table 4 we see that the induced order is E > D > C > B > A. Table 5 shows contrasts between pair of stimuli and serves the purpose of assessing the existence of real differences in this ordering. The quantities reported were derived for Model B generalized least squares estimates. At the 5% level the only nonsignificant difference is D-C. The contrast B-A is a boundary case.

Table 4

Model B method of moments (MM), maximum likelihood (ML), and generalized least squares (GLS) estimates. Values in parenthesis are standard errors.

Parameter	MM	ML	GLS
$\hat{\tau}_1$	-0.853	-0.847	-0.847
	(0.058)	(0.053)	(0.053)
$\hat{\tau}_2$	-0.388	-0.388	-0.388
	(0.046)	(0.045)	(0.045)
$\hat{\tau}_3$	0.536	0.537	0.537
	(0.047)	(0.046)	(0.046)
$\hat{\tau}_4$	1.234	1.225	1.225
	(0.073)	(0.064)	(0.064)
$\hat{\delta}_1$	0.898	0.908	0.909
	(0.082)	(0.068)	(0.068)
$\hat{\delta}_2$	1.381	1.370	1.370
	(0.114)	(0.107)	(0.107)
$\hat{\delta}_3$	0.614	0.611	0.611
	(0.033)	(0.030)	(0.030)
$\hat{\delta}_4$	2.322	2.314	2.315
	(0.242)	(0.234)	(0.234)
$\hat{\delta}_5$	0.906	0.909	0.909
	(0.069)	(0.064)	(0.064)
$\hat{\mu}_1$	-0.841	-0.844	-0.844
	(0.184)	(0.079)	(0.079)
$\hat{\mu}_2$	-0.577	-0.572	-0.571
	(0.178)	(0.104)	(0.104)
$\hat{\mu}_3$	0.076	0.075	0.075
	(0.066)	(0.047)	(0.047)
$\hat{\mu}_4$	0.308	0.305	0.304
	(0.405)	(0.170)	(0.170)
$\hat{\mu}_5$	0.995	0.993	0.993
	(0.091)	(0.076)	(0.076)

Table 5
Pairwise contrasts $\hat{\mu}_i - \hat{\mu}_\nu$ for generalized least squares Model B estimates.

Contrast	Estimate	Std error	z
B-A	0.273	0.137	1.993
C-A	0.919	0.093	9.882
D-A	1.148	0.198	5.798
E-A	1.837	0.110	16.700
C-B	0.646	0.121	5.339
D-B	0.875	0.213	4.108
E-B	1.564	0.135	11.585
D-C	0.229	0.188	1.218
E-C	0.918	0.091	10.088
E-D	0.689	0.196	3.515

5 Measuring stimuli intensity

Usually one would study differences in intensity between two impulses S_i and S_j examining for significance the contrast $\hat{\mu}_i - \hat{\mu}_j$. This is the idea behind the scaling process suggested by Thurstone - to transform an ordinal scale induced by the referees perception into an interval scale that allows comparisons among stimuli. In some applications a further analysis may be necessary. It may be of interest to measure the relative importance of the stimuli through a set of weights summing up to one. This would make the analysis comparable in many regards to Saaty's (1994) Analytic Hyerarchy Process which uses relative weights to rank stimuli. Working with weights reflecting the relative importance of each variable in a set, according to the perceptions of a group of referees, is a convenient way to assess a combined measure of performance when the stimuli are defined by variables comprising the performance dimension. An example in this regard is provided in Souza and Ávila (2000) where a set of weights is defined with the purpose to specify a combined measure of output for the production process of a research institution.

Motivated by Model D under the assumption that the psychological continuum is a projection of a log-normal distribution, we define the relative importance of stimulus i as

$$r_i = \frac{\exp(\mu_i)}{\sum_{\nu=1}^{n} \exp(\mu_\nu)} \tag{5.1}$$

It is seen from (5.1) that the ratios of relative importance r_i/r_j will be a monotonic transformation of the scale contrasts $\mu_i - \mu_j$ and therefore their inferences are equivalent.

Notice that under the assumption of a log-normal projection, that is, assuming that $\log(\xi_i) \sim N(\mu_i, \sigma^2)$ and $\log(\eta_j) \sim N(\tau_j, \phi^2)$ the r_i are the relative ratios of the means of the random variables ξ_i. In terms of the probabilities π_{ij} and assuming the moment equations (2.1), the weights r_i can expressed as follows:

$$
\begin{aligned}
&\frac{\exp(-g(\pi_{ij}))}{\sum_{\nu=1}^{r}\exp(-g(\pi_{\nu j}))} \quad \text{for Model } D\\
&\frac{\exp(-\delta_i g(\pi_{ij}))}{\sum_{\nu=1}^{r}\exp(-\delta_\nu g(\pi_{\nu j}))} \quad \text{for Model } B\\
&\frac{\exp(-\tau_j g(\pi_{ij}))}{\sum_{\nu=1}^{r}\exp(-\tau_j g(\pi_{\nu j}))} \quad \text{for Model } C
\end{aligned}
\tag{5.2}
$$

The formulas shown in (5.2) are particularly appealing for the logistic response function when they become simple functions of odds ratios what provides a further motivation for its use as a measure of stimuli relative intensity. Notice that in any case the ratios are the same for every j.

$$
\begin{aligned}
&\frac{\dfrac{1-\pi_{ij}}{\pi_{ij}}}{\sum_{\nu=1}^{r}\left(\dfrac{1-\pi_{\nu j}}{\pi_{\nu j}}\right)} \quad \text{for Model } D\\
&\frac{\left(\dfrac{1-\pi_{ij}}{\pi_{ij}}\right)^{\delta_i}}{\sum_{\nu=1}^{r}\left(\dfrac{1-\pi_{\nu j}}{\pi_{\nu j}}\right)^{\delta_\nu}} \quad \text{for Model } B\\
&\frac{\left(\dfrac{1-\pi_{ij}}{\pi_{ij}}\right)^{\theta_j}}{\sum_{\nu=1}^{r}\left(\dfrac{1-\pi_{\nu j}}{\pi_{\nu j}}\right)^{\theta_j}} \quad \text{for Model } C
\end{aligned}
\tag{5.3}
$$

From (5.3) we see that, for Model D, the larger r_i the more likely is to classify S_i into the higher response categories, increasing its importance. For Models B and Models C basically the same conclusion applies with likelyness being weighted by proper scale constants.

The variances of the relative importance $\hat{r_i}$ are easily computed expanding the estimates in Taylor series. They are given by the quantities $l_i' \text{Var}(\hat{\mu}) l_i$ where l_i has components $l_{i\nu}$, $\nu = 1, \ldots, r$ given by

$$l_{i\nu} = \begin{cases} \hat{r}_i(1 - \hat{r}_i) & \text{if } \nu = i \\ -\hat{r}_i \hat{r}_\nu & \text{if } \nu \neq i \end{cases} \tag{5.4}$$

Table 6 shows the relative intensity of the stimuli (5.1) together with standard errors derived from (33). The message is the same conveyed by the estimates $\hat{\mu}_i$. If the stimuli were performance variables the weights could be used to define a combined measure of performance taking into account the relative importance of each variable according to the perception of the population sampled.

Table 6
Intensities $r_i = \exp \mu_i / \sum_\nu \exp \mu_\nu$ for Model B using generalized least squares estimates.

Intensity	Std Err
0.070	0.006
0.092	0.011
0.176	0.012
0.221	0.032
0.441	0.027

6 Conclusions

The Law of Categorical Judgment derived from the work of Thurstone (1927) was revisited. The classical method of moments solution to this set of equations was discussed and proper standard errors computed for these estimators under general assumptions for the underlying distribution in the psychological continuum. Two major drawbacks of the classical theory are the asymptotic inefficiency of the estimates and the lack of a proper statistical framework allowing the test of Thurstone's formulations. In this context more adequate instruments of analysis were suggested with the use of maximum likelihood estimates and generalized least squares.

A measure of relative importance was proposed to assess the intensity of each stimulus. These values serve the purpose to define weights summing to one that can be used to define a combined measure of performance in some applications.

An important problem ignored throughout the paper until now in the discussion of the Law of Categorical Judgment is that observations corresponding to different stimuli may not always be independent. In many applications the same set of referees evaluate each stimuli. This will not invalidate any of the estimation methods but is likely to induce a positive correlation between the ratings of any two stimuli. The effect of a positive correlation is to reduce the variance of a given contrast between two stimuli and therefore the analysis in this context can be regarded as conservative since standard errors are overestimated.

Typically the Law of Categorical Judgment in one of the forms D, B, or C, provides a close fit to the data. The estimation approach, by any of the three methods, is also robust relative to the distribution postulated in the psychological continuum. Inferences will be similar whether one considers the probit, the logistic or the log-log scale.

vspace4mm

(Received October, 2000. Revised August, 2002)

References

Fothofer, R. N. and Lehnen, R. G. (1981). *Public Program Analysis A New Categorical Data Approach.* Belmont: Wadsworth.

Gulliksen, H. (1954). A least squares solution for successive intervals assuming unequal standard deviations. *Psychometrika*, **19**, 117–139.

Grizzle, J. E., Stamer, C. F. and Kock, G. G. (1969). Analysis of categorical data by linear models. *Biometrics*, **25**, 489–504.

Kotz, N. and Johnson, L. (1989). Thurstone's theory of comparative judgment. *Encyclopedia of Statistical Sciences*, **9**, 237–239. New York: Wiley.

Macedo, M. M. C. (1997). *The process of agricultural technology generation in Brazil: a social study.* PhD. Dissertation. England: University of Sussex.

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 preference data. Department of Psychology, University of Barcelona. To appear in Psychometrika.

McFadden, D. (1974). Conditional logit analysis of qualitative choice behavior. *Frontiers in Econometrics*, ed. Zarembka, P. New York: Academic Press.

McFadden, D. (2001). *Economic choices, economic decision of individuals.* Charter Day keynote address. Berkeley: University of California.

McCullagh, P. and Nelder, J. A. (1989). *Generalized Linear Models*, 2.ed. New York: Chapman & Hall.

Rousseau, B., Rogeaux, M. and O'Mahony M. (1999). Mustard discrimination by same-different and triangle tests: aspects of irritation, memory and τ criteria. *Food Quality and Preference*, **10**, 173–184.

Saaty, T. L. (1994). *The Analytic Hierarchy Process.* Pittsburgh: RWS.

Sousa, I. S. F. (1993). *A Sociedade, o Cientista e o Problema de Pesquisa.* São Paulo: Embrapa, Brasília/Hucitec.

Souza, G. (1998). *Introdução aos Modelos de Regressão Linear e Não-Linear.* Brasília: Embrapa.

Souza, G. and Ávila F. D. A. (2000). A psicometria linear da escalagem ordinal: uma aplicação na caracterização da importância relativa de atividades de produção em ciência e tecnologia. *Cadernos de Ciência e Tecnologia*, **17**, 11–27.

Souza, J. (1988). *Métodos de Escalagem Psicossocial.* Brasília: Thesaurus.

Thurstone, L. L. (1927). A law of comparative judgments. *Psychological Review*, **34**, 273–286.

Torgerson, W. S. (1958). *Theory and Methods of Scaling.* New York: Wiley.

Turoff, M. and Hiltz, S. R. (1996). Computer based delphi processes. *Gazing into the Oracle: The Delphi Method and its Applications to Social Policy and Public Health*, ed. Addler M. and Ziglio, E. London: Kingsley Publishers.