

New Mathematical Model for the Law of Comparative Judgment

Mai Zhou
University of Kentucky, Lexington, Kentucky
Chengwu Cui
Lexmark International Inc., Lexington, Kentucky

Abstract

The law of comparative judgment provides a useful approach for psychophysical scaling in subjective printing quality measurement and evaluation. It links the scales, through a set of equations, to the proportion of times that one stimulus is judged greater than another in terms of perceived strengths of stimuli. The equations can be further simplified with additional assumptions on the perceptual dispersion of each stimulus and on their correlations to various forms that can be solved, via least squares method, for practical applications. However, the normal deviates can approach the value of (plus or minus) infinity. Further, because the equations for the law of comparative judgment deal with the normal deviates, solutions are generally not optimized for the proportions of choice. The method of modeling the correlation between two stimuli can also be improved to reflect the underlying mechanism of perception. To overcome these problems, we propose to directly model the discriminal dispersion variation and use a maximum likelihood model to describe the law of comparative judgment. The model describes the law of comparative judgment directly in the proportion of choice domain. Solutions are also given for actual paired comparison data. Other useful information such as the standard error of predictions can also be obtained easily based on the simplified forms of the model.

Introduction

Psychometric measurement methodologies are based on the assumptions that human's perceptual response to physical stimuli can be modeled like any other physical process, and the measurement follows the same statistical rules as that of other physical measurement theories. With these assumptions, many different methods have been used to measure the so-called psychometric continuums or the perceptual quantities.[1,2] One of the most often used methods is the method of paired comparison. Paired comparison, in its original sense, is a fundamental technique in our everyday decision making process. Given a pair of stimuli, if we compare them enough times for the specific perceptual attribute of interest, we can conclude with a certain degree of confidence that one is stronger than the other is. Such a comparison process is the so-called psychometric measurement when the process is under controlled conditions. The direct measurement response, in this case, is the proportion of choice. We can convert the proportion of choice into its normal deviate based on a discriminal dispersion in the form of a normal distribution. If we have more than two stimuli and have each possible pair combination compared, we will need an optimization procedure to construct a scale and place all stimuli on that scale.

To process paired comparison data, many models are proposed based on the specific fields of application, measurement goals, and available computation resources.[1,2] The most well known models are the Thurstone model and the Bradley-Terry model.[2,3] The Thurstone model is the most complete and often referred to as the law of comparative judgment. Based on certain underlying principles, Thurstone modeled paired comparison in a general form and then simplifies the model based on the degree of variation of the discriminal dispersions. When all dispersions are equal, we have the simplest form of the law of comparative judgment - Thurstone's case V. The variation of the discriminal dispersion is modeled via the correlation between the discriminal dispersions of the two stimuli. Although modeling the distribution of the difference between two normal distributions is a common practice in statistics, whether the same advantage can be gained when applied to the paired comparison process is not clear. In this paper, we propose to model the dispersion variation in a different way and rewrite the law of comparative judgment accordingly.

Various models based on simplified forms of Thurstone's law of comparative judgment can be solved with minimum computation requirements. The conversion from the proportion of choice to the normal deviates does not exist at the two tail ends of the normal distribution. This will frequently happen when a stimulus is consistently regarded as stronger than another. Because of this problem, special treatment is needed. Also Thurstone's model solution process is dealt with in the normal deviates domain, the optimal solution may not be optimal in the proportion of choice domain.

The Bradley-Terry[3] model proposes the use of a simple ratio of the scaled values, or true ratings, to represent the probability of the outcome of the comparison. Using the maximum likelihood estimate, the model is able to estimate the scaled values based on measured proportion of choice data. It can be proved that the Bradley-Terry model is actually close to the Thurstone's case V model.[3] Use of the Maximum likelihood estimate in the Bradley-Terry model avoids the computational pitfall of using the normal deviates; it also provides optimization in the proportion of choice domain. However, the Bradley-Terry model does not address the issue of potential discriminal dispersion variations.

In this paper, we propose to directly model the discriminal dispersions, instead of using the traditional difference of two dispersions (distributions) approach, and use a maximum likelihood model to rewrite the law of comparative judgment.

Theory

Thurstone's Law of Comparative Judgment

Assuming stimulus j and k belong to a set of n stimuli, when j and k are compared, Thurstone's law of comparative judgment specifies,

$$s_k - s_j = x_{jk}\sigma_{jk} \tag{1}$$

where s_k and s_j are the scaled values (response strength) of the stimulus k and j, respectively; x_{jk} is the normal deviate of the proportion of choice; σ_{jk} is the distribution or dispersion of the discriminal process of the comparison process. Assume σ_j and σ_k are the discriminal dispersions of stimulus j and k, respectively, and r_{jk} is the correlation between the pairs of discriminal dispersions, we have,

$$\sigma_{jk}{}^2 = \sigma_j{}^2 + \sigma_k{}^2 - 2r_{jk}\sigma_j\sigma_k \tag{2}$$

Equation (1) is often referred to as the Case I of Thurstone's law of comparative judgment. There is no unique solution to Equation (1). The use of Equation (2) allows Thurstone's model to be reduced to various solvable forms by assuming uniform dispersion, no correlation, or fixed correlation. Equation (2) is often used to calculate the difference distribution of two distributions with the correlation also computable once we know the two difference distributions.

Newer Form for the Law of Comparative Judgment

What can cause the potential correlation between the response to the two stimulus? It is known that biological response, especially the human perceptual process, is highly adaptive and is at its best when used as a null tester. Therefore, the discriminal dispersion should depend on the difference of response strength of the two stimuli. If the goal is to reduce unknowns in a model, it seems that modeling the dispersion directly as a function of stimulus response strength is more appropriate as opposed to using Equation (2). After all, we are less interested in the discriminal dispersion to each stimulus than we are interested to the discriminal dispersion of the comparison process. Let p_{jk} be the total number stimulus j is judged greater than k out of N times of comparisons, we may write,

$$p_{jk} = \alpha \int_{-\infty}^{s_j - s_k} \Phi(x, \sigma_{jk})dx \tag{3}$$

where,

$$\Phi(x, \sigma_{jk})$$

is the dispersion distribution function with σ_{jk} as the standard deviation if the normal distribution is used (not necessary); and

$$\alpha = \frac{1}{\int_{-\infty}^{\infty} \Phi(x,1)dx}.$$

Solution to Equation (3) can be achieved by maximizing the following likelihood function,

$$MLL = \sum_{j=1}^{n} \sum_{k=j+1}^{n-1} \left\{ \begin{array}{l} p_{jk} \log\left[1 - \alpha \int_{-\infty}^{s_j - s_k} \Phi(x, \sigma_{jk})dx\right] \\ + (N - p_{jk}) \log\left[\alpha \int_{-\infty}^{s_j - s_k} \Phi(x, \sigma_{jk})dx\right] \end{array} \right\} \tag{4}$$

Equation (4) can be optimized when the number of variables is sufficiently small relative to the number of comparisons.

Solution and Error Analysis

Equation 4 can be solved with the assistance of most statistical software packages using their corresponding generalized linear model solvers. The solution of the generalized linear model can also provide the variance matrix for estimation of the computed parameters, in this case, the scaled values and the dispersions. The variance matrix computation is based on the second order partial derivatives of Equation 4 following standard statistic inference procedures. If the available software package does not allow the computation of the variance matrix, an alternative approach is to use the likelihood ratio and Wilks theorem to obtain the 95% confidence intervals for parameters via.[4] We hereby illustrate the method for a confidence interval for parameter s_j. The method is asymptotically equivalent to the variance matrix or normal approximation method. The method is to, 1) find the maximum of the MML function given in Equation 4 by allowing all parameters to change; 2) find the maximum of MML function by allowing all parameters to change except s_j; 3) repeat 2) with a set of values in the vicinity of the s_j value obtained in 1) until $s_j = t$ such that,

*2 [Maximum of MML (all parameters change) -
Maximum of MML (all parameters change, but $s_j = t$)]
< 3.84* *(5)*

The collection of t is the 95% confidence interval for s_j.

Simplification of the Model

Apparently, when

$$\Phi(x, \sigma_{jk})$$

takes the form of the standard normal distribution, $\sigma_{jk}=1$, then Equation (4) becomes Thurstone's Case V and also approximates the Bradley-Terry model. If we assume normal distribution and let,

$$\sigma_{jk} = f(s_j, s_k) \qquad (6)$$

Equation (6) can be written in various forms based on certain assumptions in regard to the discrimination process of the comparison. Here we only give one example. We can assume the function in Equation (6) is only relevant to the difference of the two stimuli, we can write,

$$\sigma_{jk} = 1 + \beta \,|\, s_j - s_k \,|^{\gamma} \qquad (7)$$

where β and Υ are parameters to be determined. Equation (7) implies that discriminal dispersion for the comparison process is only a function of the difference between the two stimuli. Other forms of functions f in Equation 6 are also possible.

Application to Experimental Data

Hevner's Data Set

A well-known set of paired comparison data on psychophysical scaling was reported by Hevner on the degree of handwriting excellence. [5] The paired comparison data of the 370 pairs for 20 handwriting samples given by Hevner are used here to demonstrate the method.

The computation was carried out in Mathcad™ (ver. 8, professional edition) via its "Maximize" function call using the conjugate gradient method. The dispersion function used was given by Equation 7. The computation took 90 seconds on a Pentium™ II 333Mhz mobile processor with 128 MB RAM. The computed scaled values were normalized to and compared with Hevner's results as shown in Fig. 1. Hevner used Thurstone's tabulation method and dropped the values between 0 and 0.03 and between 0.97 and 1.

Table 1. Comparison of average and maximum residue errors by three different models.

	Ave (%)	Max. (%)
Hevner	3.2	20
Equation 7	2.5	15
Case V	2.5	15

The average and maximum residue (proportion of choice) errors are listed in Table 1.

The dispersion distribution determined by β and γ obtained is also shown in Fig. 2.

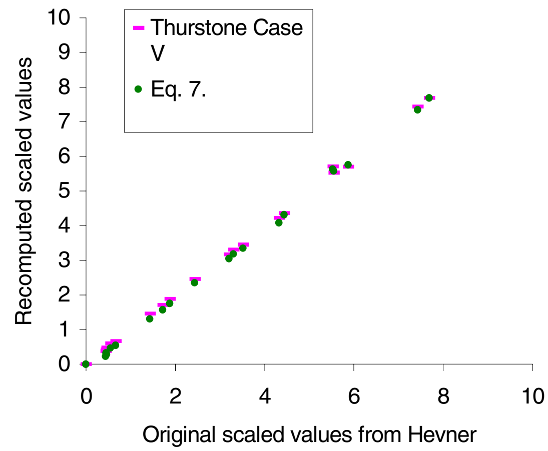

Figure 1. Recomputed excellence of handwriting scaled values of the 20 samples based on the paired comparison data by Hevner. The dots represent scaled values recomputed based on Equation 4 and 7; "-" represent recomputed scaled values based on Equation 4 assuming Thurstone's case V condition.

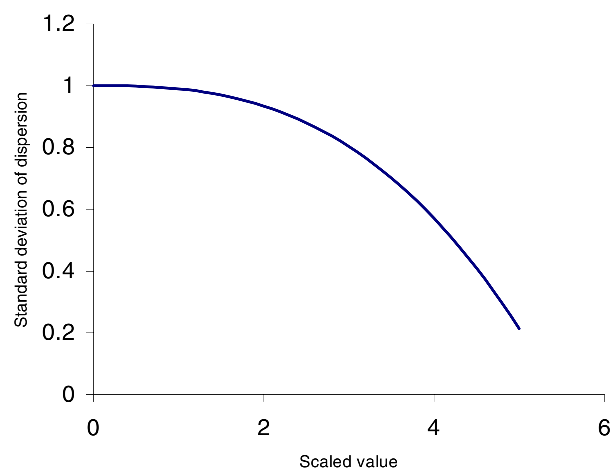

Figure 2. Plot of Equation 7 with fitted parameter β and γ for Hevner's handwriting excellence data.

However, further analysis proved that β and γ were insignificant for this set of data. When β and γ were set to zero, the model given by Equation 4 worked equally well. The scaled values (normalized to Hevner's values) are also shown in Fig. 1. The average and maximum prediction errors are given in Table 1.

Data From Colorfulness Scaling Test

In an experiment to scale the perceived colorfulness of a set of nine green color samples, the following paired comparison proportion of choice matrix as shown in Table 2 was obtained with 50 subjects.[6]

Table 2. Paired comparison data from colorfulness scaling experiment.

	1	2	3	4	5	6	7	8	9
1	/	5	4	37	6	3	34	7	4
2	45	/	7	45	28	6	43	42	15
3	46	43	/	46	40	27	44	43	36
4	13	5	4	/	4	3	34	5	3
5	44	22	10	46	/	6	42	42	10
6	47	44	23	47	44	/	45	45	35
7	16	7	6	16	8	5	/	7	4
8	43	8	7	45	8	5	43	/	5
9	46	35	14	47	40	15	46	45	/

The data were processed in the same way as done for Hevner's data. The model using the dispersion model given by Equation 7 demonstrates a better fit than the Case V model as shown in Table 3.

Table 3. Comparison of average and maximum residue errors by the model using Equation 7 and Thurstone's Case V mode.

	Ave (%)	Max. (%)
Equation 7	6.3	15
Case V	7.1	19

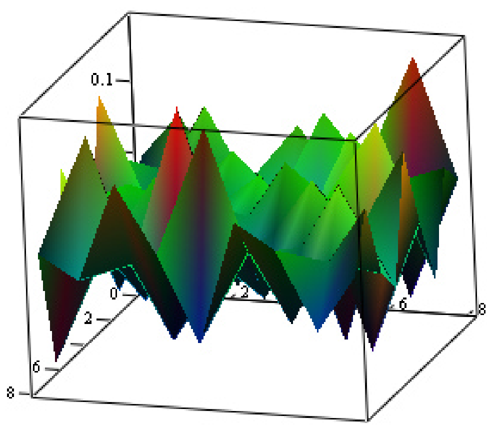

Figure 3. Prediction error (fit residue) by the Case V model.

Fig. 3 and 4 show the residue of fit by Thurstone's case V model and the model defined by Equation 7, respectively. In Fig. 3 and 4, the vertical axis is the prediction error (proportion of choice, in %) and the two horizontal axes define the pair combinations.

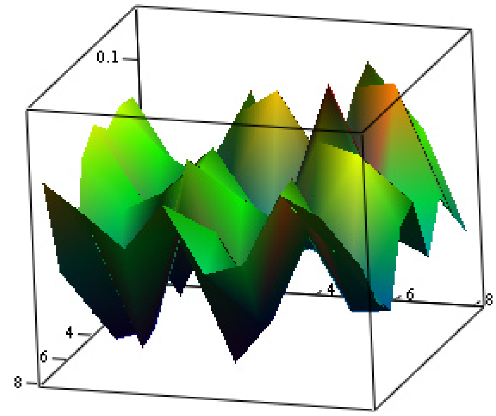

Figure 4. Prediction error (fit residue) by the model using Equation 7 to define dispersion.

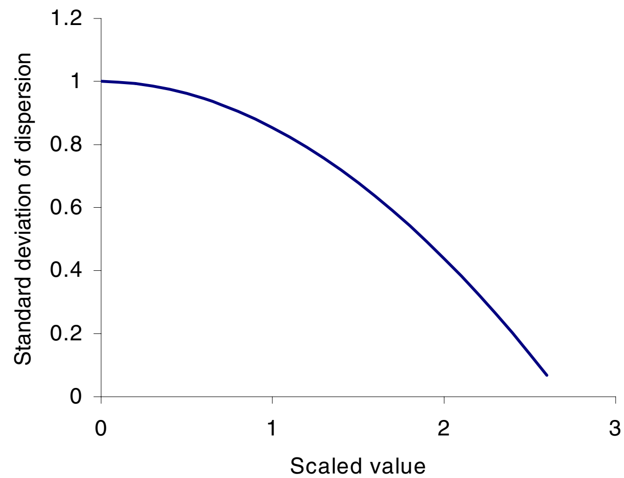

Figure 5. Plot of Equation 7 with fitted parameter β and γ for the colorfulness data.

The scaled values obtained between the two models seemed to differ by a factor of 2. The 95% confidence intervals for the scaled values were determined following Equation 5, and they were found to be typical of 0.2 for all stimuli.

The dispersion distribution is shown in Fig. 5.

Discussion

Assisted with modern software packages, the newer model for the law of comparative judgment can be easily solved as demonstrated by Hevner's data. The newer model also gave improved results over the traditional method used by Hevner as shown by the average and maximum residue error given in Table 1. In the case of Hevner's data, the newer model is believed to be equivalent to the Bradley-Terry model.

The colorfulness scaling test data set proves that a simplification to the newer model, as given by Equation 7, can fit the data better. The fit residues shown in Fig. 3 and 4 along with the average and maximum residue error all prove that the dispersion model given by Equation 7 is better than the Case V model. The dispersion distribution shows a decrease with the difference between the two stimuli increase as shown in Fig. 5.

The 95% confidence interval calculation shows the error analysis can be effectively obtained and useful for measurement error analysis.

Summary

We have described the law of comparative judgment in a newer form and propose to use the maximum likelihood estimate to estimate the scaled values. The newer model can be easily implemented in current available mathematical and statistical software packages. This approach also allows the estimation of the 95% confidence interval of the estimated scaled values.

Acknowledgement

We would like to thank Dr. Shaun Love at Lexmark International Inc. for helpful discussions.

References

1. Peter, G. Engeldrum, Psychometric scaling, A toolkit for imaging system development, Imcotek press. (2000)
2. Torgerson, W.S. Theory and methods of scaling, John Wiley & Sons, Inc. (1958)
3. Bradley, R.A. and Terry, M. E. “Rank analysis of incomplete block design. I. The method of paired comparisons.” *Biometrika* **39,** 324-345. (1952)
4. P. Hall and B. La Scala, Methodology and algorithms of empirical likelihood, International Statistical Review, **58**, 109-127. (1990)
5. Kate Hevner, An empirical study of three psychophysical methods, J. Gen. Psychol., **4**, 191-212. (1930)
6. To be published.

Biography

Mai Zhou received his BS degree in mathematics from Zhejiang University, MS and Ph.D. in statistics from Columbia University. He has been a faculty member at Statistics Department at University of Kentucky since 1989. His research interest include survival analysis, empirical likelihood methods, computer intensive statistical inference, and regression models with censored data.