A 2-dimensional extension of the Bradley–Terry model for paired comparisons

ARTICLE *in* JOURNAL OF STATISTICAL PLANNING AND INFERENCE · DECEMBER 2005

Impact Factor: 0.68 · DOI: 10.1016/j.jspi.2004.05.008

2 AUTHORS:

David Causeur
Agrocampus Ouest

Francois Husson
Agrocampus Ouest

Available online at www.sciencedirect.com
SCIENCE DIRECT®
Journal of Statistical Planning and Inference 135 (2005) 245–259
journal of statistical planning and inference
www.elsevier.com/locate/jspi

A 2-dimensional extension of the Bradley–Terry model for paired comparisons

David Causeur[a, b, *], François Husson[a]

[a] *Laboratoire de Mathematique Appliquees, Pole d'Enseignement Superieur et de Recherche Agronomique de Rennes, 65 rue de St-Brieuc, CS 84215, 35042 Rennes cedex, France*
[b] *CREST-ENSAI, Campus de Ker Lann, Bruz, France*

Received 23 October 2003; accepted 13 May 2004
Available online 23 July 2004

Abstract

The Bradley–Terry model is widely and often beneficially used to rank objects from paired comparisons. The underlying assumption that makes ranking possible is the existence of a latent linear scale of merit or equivalently of a kind of transitiveness of the preference. However, in some situations such as sensory comparisons of products, this assumption can be unrealistic. In these contexts, although the Bradley–Terry model appears to be significantly interesting, the linear ranking does not make sense. Our aim is to propose a 2-dimensional extension of the Bradley–Terry model that accounts for interactions between the compared objects. From a methodological point of view, this proposition can be seen as a multidimensional scaling approach in the context of a logistic model for binomial data. Maximum likelihood is investigated and asymptotic properties are derived in order to construct confidence ellipses on the diagram of the 2-dimensional scores. It is shown by an illustrative example based on real sensory data on how to use the 2-dimensional model to inspect the lack-of-fit of the Bradley–Terry model.

MSC: primary 62J12, 62J15; secondary 62H15

Keywords: Bradley–Terry models; Multidimensional scaling; Paired comparison models; Structured interaction models

* Corresponding author. Laboratoire de Mathematique Appliquees, Pole d'Enseignement Superieur et de Recherche Agronomique de Rennes, 65 rue de St-Brieuc, CS 84215, 35042 Rennes cedex, France. Tel.: +33-22348-5884; fax: +33-22348-5871.

E-mail addresses: david.causeur@agrocampus-rennes.fr (D. Causeur), husson@agrorennes.educagri.fr (F. Husson).

0378-3758/$ - see front matter © 2004 Elsevier B.V. All rights reserved.
doi:10.1016/j.jspi.2004.05.008

1. Introduction

Paired comparisons of a set of objects are encountered in a wide variety of applications, most of the time aiming at a ranking of these objects. There is a broad literature both on the practical situations in which paired comparisons are usual and on the properties of the models devoted to these situations. For instance, the book by David (1988) provides a detailed review of paired comparison models. Probably, the most cited among the applied uses of paired comparisons is the tournament analysis in which the objects are players or teams competing with each other on pairs. In the citation analysis examined by Stigler (1994), the relationships between journals are similarly investigated on the basis of a square table displaying the number of citations of a journal by another one. In the context of sensory studies also, some trials consist of comparing food products by means of tasting experiments. In this context, the preference data analysis is expected to yield a ranking of the products that is to be related with the preference of the consumers.

More generally, ranking by paired comparisons models is usually achieved by means of a score supposed to reflect a kind of merit of the items on a linear scale. The most popular and in many situations very useful way of ranking in the context of paired comparisons is due to Bradley and Terry (1952). Although the pioneering works on the so-called Bradley–Terry model were not explicitly connected with analysis of variance models, Agresti (1990) and more recently McCullagh (2000) have clearly considered this model as a special case of a logistic analysis of variance model for binomial data. That point of view obviously gives more insight to the algorithmic aspects of the estimation issue but also to the practical limitations of this model. Some of these practical limitations have of course called for extensions, mainly by introducing non-linearity. One of the first reviews of such extensions can be found in Davidson and Farquhar (1976). For instance, Rao and Kupper (1967) and Davidson (1970) have proposed non-linear Bradley–Terry models that account for ties. What can be called the home-field advantage is accounted for in other extensions by Agresti (1990). As shown by Hunter (2004), most of these non-linear Bradley–Terry models belong to a class of models for binomial data for which some algorithms, called Minorization–Maximization or MM algorithms, are shown to have desirable properties. It can also be noticed that both in the context of sensory studies and psychometrics, such as in De Soete and Carroll (1983) or Dittrich et al. (1998), the introduction of covariates in the model have been suggested, for instance, to allow for interaction between preference and external information concerning the experiments.

As the Bradley–Terry model assumes a linear scale of merit that makes the ranking of the objects possible, it can be inappropriate when these merits are not transitively related. For instance, in the sensory context mentioned above, it can be observed that a product, say A, is markedly better than B, B is also markedly better than C; however C is preferred to A. In that case, obviously, no linear scale of merit can reflect properly the preference. As mentioned by Hunter (2004), a directed graph in a 2-dimensional space in which the distances between the objects reflect the absolute difference in preference and the directions of the edge that point out of the preferred object is much more appropriate. Our purpose is to propose a model that extends the Bradley–Terry model by accounting for this special form of interaction between the objects.

Section 2 is devoted to the definition of the model and the motivations for extending the Bradley–Terry model. The properties of the maximum-likelihood estimation are examined in Section 3 and testing procedures are provided for the relevance of the 2-dimensional model relative to its uni-dimensional version. The illustration of the present model by real data consisting in paired comparisons of cornflakes with respect to their crunchy is provided in Section 4. The data analysis show especially how the 2-dimensional model can be used to give more insight to the initial ranking provided by the Bradley–Terry model. Finally, Section 5 is dedicated to the concluding remarks.

2. A 2-dimensional Bradley–Terry model

Consider the paired comparison of $n,\ n\geqslant 3$, objects by means of repeated independent experiments. The response outcome that is observed at each comparison is a binomial variable, thereafter denoted N_{ij}, that counts the number of times the item *i* beats the item *j*

$$N_{ij} \sim \mathscr{B}(m_{ij};\pi_{ij}),$$

where $m_{ij}\geqslant 1$ is the number of pairings of the objects *i* and *j* and $0\leqslant \pi_{ij}\leqslant 1$ stands for the probability that *i* beats *j*.

Many attempts to rank the objects with respect to their performances along the comparisons with the other objects are based on more or less sophisticated modelling of the probability π_{ij}. Since Bradley and Terry (1952), such problems of modelling are indeed investigated by introducing a score vector $\lambda=(\lambda_1,\lambda_2,\ldots,\lambda_n)'$ that is used for ranking the objects. The so-called Bradley–Terry model can be seen as a logistic analysis of variance model for binomial data:

$$\mathrm{logit}(\pi_{ij}) = \lambda_i - \lambda_j.$$

The idea behind the previous model is to represent the merit of an item, say *i*, by a score λ_i on a linear scale. Identifiability of the scores is usually assessed by assuming that $\sum_{i=1}^{n}\lambda_i=0$.

Example. Consider the virtual paired comparison data displayed in Table 1 in which 5 objects are compared by 100 pairings. The cell (i,j) in this table gives the proportion of

Table 1
Virtual illustrative data and deviance residuals of Bradley–Terry model

	Data				
	1	2	3	4	5
1		0.25	0.25	0.27	0.12
2			0.27	0.62	0.25
3				0.38	0.25
4					0.27

Deviance residuals				
1	2	3	4	5
	−0.68	0.16	0.26	0.38
		−3.77	2.94	0.16
			−2.94	−0.68
				0.26

Bradley–Terry scores				
−1.06	−0.10	0.10	0.00	1.06

pairings in which i has beaten j. In this example, regarding the data, there is undoubtedly a linear scale of merit that reflects the following ranks: $(1) < (2, 3, 4) < 5$. The data shows clearly a situation of intransitiveness among 2, 3 and 4 since 4 is beaten by 2, 2 is beaten by 3 and 3 is beaten by 4.

The Bradley–Terry scores are also given in Table 1 together with the deviance residuals as defined by Pierce and Schafer (1986). The largest residuals point out the pairings for which the Bradley–Terry model fails to reflect the observed differences in terms of preference. Obviously, regarding these deviance residuals and the Bradley–Terry scores, the paired comparisons of the objects (2, 3, 4) show a marked difference between these objects that is not captured by the Bradley–Terry model.

Therefore, in some situations, although the Bradley–Terry model provides a significant ranking of the objects that are compared by pairs, some lack-of-fit problems may occur. This lack-of-fit can be due to the non-relevance of a linear scale of merit and as a consequence of an overall ranking. Our aim is to account for possible interactions between the objects that could affect the linear representation of preference.

The following non-linear logistic model is now assumed:

$$\operatorname{logit}(\pi_{ij}) = \sigma_{ij}\sqrt{(\lambda_{i,1}-\lambda_{j,1})^2 + (\lambda_{i,2}-\lambda_{j,2})^2}, \tag{1}$$

where $\sigma_{ij} = \pm 1$ is a sign parameter, with $\sigma_{ij} = -\sigma_{ji}$, that reflects the preference between the objects i and j, Σ is the $n \times n$ matrix with generic term σ_{ij}, $\boldsymbol{\lambda}_1 = (\lambda_{1,1}, \lambda_{2,1}, \ldots, \lambda_{n,1})'$ and $\boldsymbol{\lambda}_2 = (\lambda_{1,2}, \lambda_{2,2}, \ldots, \lambda_{n,2})'$ are n-vectors that satisfy the following identifiability restrictions: $\sum_{i=1}^n \lambda_{i,1} = \sum_{i=1}^n \lambda_{i,2} = 0$ and $\sum_{i=1}^n \lambda_{i,1}\lambda_{i,2} = 0$. In the following, $\boldsymbol{\lambda}$ will denote the complete vector of parameters $(\boldsymbol{\lambda}_1', \boldsymbol{\lambda}_2')'$.

In other words, $\operatorname{logit}(\pi_{ij})$ is now considered as a signed Euclidean distance in $\mathbb{R}^2$, which is very similar to the multidimensional scaling issue usually defined for normal data. By analogy with the classical multidimensional scaling, orthogonality between $\boldsymbol{\lambda}_1$ and $\boldsymbol{\lambda}_2$ is assumed to face invariance of the likelihood function relative to rotations.

Incidentally, estimating $\boldsymbol{\lambda}_1$ and $\boldsymbol{\lambda}_2$ and displaying them on a diagram provide a kind of mapping of the preference in the sense that the distances between the points on this diagram can be interpreted as differences in preference between the objects. Note that if a linear scale of preference actually exists, namely if $\boldsymbol{\lambda}_2 = 0$ and the Bradley–Terry ranking is consistent with the observed preference, model (1) gives the same representation of these objects as the Bradley–Terry model.

Example. The virtual paired comparison data in Table 1 have been created according to an exact 2-dimensional display of $\operatorname{logit}(\pi_{ij})$. Fig. 1 provides a representation of the exact distances between the objects and those figured by the Bradley–Terry scores. Note that the x-axis in the 2-dimensional mapping of the preference does still reflect the main trend of preference identified by the Bradley–Terry model. However, the diagram makes more obvious in terms of merit the differences between the objects $(2, 3, 4)$. More generally, comparing the linear and the 2-dimensional representation of the objects can indeed be used as a graphical lack-of-fit diagnostic for the Bradley–Terry model.

In the next section, the maximum-likelihood estimation procedure is described and asymptotic variances are given in order to construct confidence ellipses.

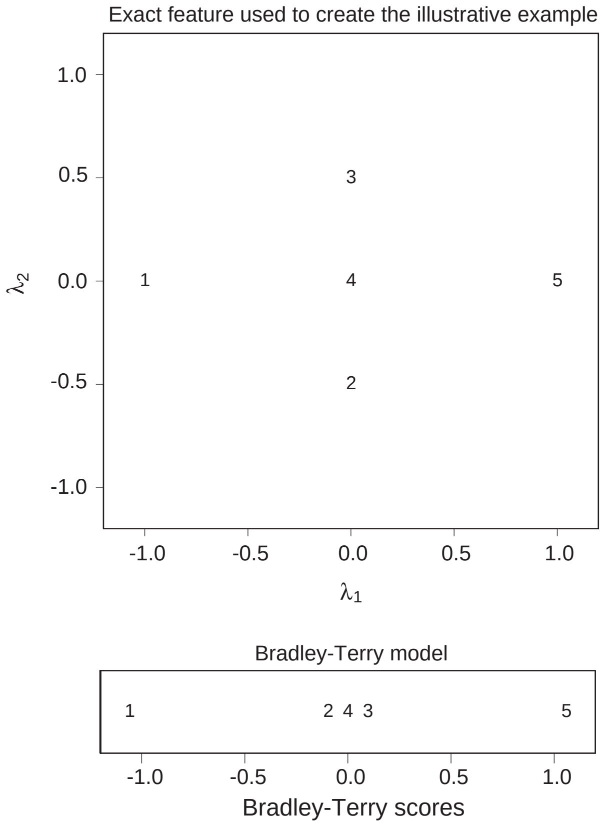

Fig. 1. Comparison of the Bradley–Terry scores and the exact feature used to create the illustrative data given in Table 1.

3. Maximum-likelihood estimation and testing

First, let us derive the log-likelihood $\mathscr{L}(\lambda, \Sigma)$, or $\mathscr{L}$ for convenience, of model (1) under the identifiability restrictions

$$\mathscr{L} = \sum_{i=1}^{n} \sum_{j=i+1}^{n} \log \mathbb{P}(N_{ij} = n_{ij}) + a_1 \sum_{i=1}^{n} \boldsymbol{\lambda}_{i,1} + a_2 \sum_{i=1}^{n} \boldsymbol{\lambda}_{i,2} + a_3 \sum_{i=1}^{n} \boldsymbol{\lambda}_{i,1}\boldsymbol{\lambda}_{i,2},$$

$$\sim \sum_{i=1}^{n} \sum_{j \neq i, j=1}^{n} n_{ij} \log \pi_{ij} + a_1 \sum_{i=1}^{n} \boldsymbol{\lambda}_{i,1} + a_2 \sum_{i=1}^{n} \boldsymbol{\lambda}_{i,2} + a_3 \sum_{i=1}^{n} \boldsymbol{\lambda}_{i,1}\boldsymbol{\lambda}_{i,2},$$

$$\sim \mathscr{L}^* + a' \varphi(\boldsymbol{\lambda}),$$

where $\sim$ stands for the equality upto an additive constant, $a=(a_1,a_2,a_3)'$ are Lagrange multipliers and $\varphi(\boldsymbol{\lambda})$ the following functions of $\boldsymbol{\lambda}$:

$$\varphi(\lambda)=\begin{bmatrix}\sum_{i=1}^{n}\boldsymbol{\lambda}_{i,1}\\ \sum_{i=1}^{n}\boldsymbol{\lambda}_{i,2}\\ \sum_{i=1}^{n}\boldsymbol{\lambda}_{i,1}\boldsymbol{\lambda}_{i,2}\end{bmatrix}.$$

The identifiability restrictions on model (1) are equivalently expressed by $\varphi(\boldsymbol{\lambda})=0$. Therefore, $\mathscr{L}^*$ is the unrestricted log-likelihood and the following notations are also introduced in the previous expression of $\mathscr{L}$, for $j<i$, $\pi_{ij}=1-\pi_{ji}$ and $n_{ij}=m_{ij}-n_{ji}$.

The maximization relative to the $n(n-1)/2$ sign parameters consists in maximizing separately the quantities $\gamma_{ij}=n_{ij}\log(\pi_{ij})+n_{ji}\log(1-\pi_{ji})$. It is straightforwardly checked that γ_{ij} is maximized by $\hat{\sigma}_{ij}=+1$ if $n_{ij}>n_{ji}$ and by $\hat{\sigma}_{ij}=-1$ otherwise. Note that, if $n_{ij}=n_{ji}$, γ_{ij} does not depend on the sign of σ_{ij}. In that case, it can be decided for instance that $\hat{\sigma}_{ij}=+1$.

Concerning the merit vectors, the estimated signs $\hat{\sigma}_{ij}$ are now substituted in the log-likelihood. Iterative algorithms dedicated to the maximization of the log-likelihood function $\mathscr{L}$ have been proposed by many authors applied to non-linear extensions of the Bradley–Terry model. A wide review of the different algorithmic approaches for a broad variety of models is provided by Hunter (2004). Hunter (2004) also shows that the MM algorithms are powerful tools to deal with some Bradley–Terry type models especially in the case of a large number of parameters. In the present case, as it is usually implemented for logistic regression models, we will derive the sequence $(\boldsymbol{\lambda}^{(k)})_{k\geqslant 0}$ of estimates for $\boldsymbol{\lambda}$ by a Newton–Raphson approach.

3.1. Starting point for the iterative algorithm

A crucial point for an iterative algorithm is the intelligent choice of a starting point. In the present situation, this can be obtained by a first-order normal approximation of the binomial model. First, it is well known that, for a large number m_{ij} of pairings, the observed proportion N_{ij}/m_{ij} of wins for *i* against *j* is distributed as a normal variable with mean π_{ij} and variance $\pi_{ij}(1-\pi_{ij})/m_{ij}$. Asymptotic normality is therefore also valid for $\mathrm{logit}(N_{ij}/m_{ij})$. Moreover, it can be deduced from a Taylor series expansion of $\mathrm{logit}(N_{ij}/m_{ij})$ around π_{ij} that

$$\mathbb{E}\left(\mathrm{logit}\,\frac{N_{ij}}{m_{ij}}\right)=\mathrm{logit}\,\pi_{ij}+\frac{\pi_{ij}(1-\pi_{ij})}{2m_{ij}}\frac{\partial^2\,\mathrm{logit}\,x}{\partial x^2}(\pi_{ij})+o(m_{ij}^{-1}).$$

Therefore, $\mathrm{logit}(N_{ij}/m_{ij})$ can approximately be seen as a normal variable whose expectation is a signed Euclidean distance in $\mathbb{R}^2$. According to Ramsay (1982), this normal framework enables the derivation of the coordinates $\boldsymbol{\lambda}_1$ and $\boldsymbol{\lambda}_2$ by a classical multidimensional scaling approach. This method aims at approximating, in the sense of least-squares, the $n\times n$ matrix with generic term $[\mathrm{logit}(N_{ij}/m_{ij})]^2$ by a matrix whose generic term is a squared Euclidean distance between n points in a 2-dimensional space.

Let Q denote the $n\times n$ square table with generic term $q_{ij}=[\mathrm{logit}(N_{ij}/m_{ij})]^2$, if $i\neq j$ and $q_{ii}=0$. Let $Q_c=[I_n-(1/n)J_n]Q[I_n-(1/n)J_n]$ denote the matrix obtained by a row–column centering of Q, where I_n stands for the $n\times n$ identity matrix and J_n for the $n\times n$ matrix with all elements equal to 1. Let us also assume that $-1/2\ Q_c$ has q positive eigenvalues $d_1\geqslant d_2\geqslant\cdots\geqslant d_q$ with $2\leqslant q\leqslant n-1$. Call v_i the normalized eigenvector associated to the eigenvalue d_i, then the following starting point for an iterative estimation algorithm is proposed:

$$\boldsymbol{\lambda}_i^{(0)}=\sqrt{d_i}\,v_i,\quad i=1,2.$$

If $q=2$ and $d_i=0$, for $i>2$, it can be deduced from Gower (1966) for instance that Q is exactly a matrix whose generic term is the Euclidean distance between the n points M_i with x-coordinates in $\boldsymbol{\lambda}_1^{(0)}$ and y-coordinates in $\boldsymbol{\lambda}_2^{(0)}$. If $q\geqslant 2$, with possibly negative eigenvalues $d_i,\ i>2$, the matrix of the squared distances between the points M_i is the least-squares approximation of Q by a matrix of squared distances. Moreover, as a nice consequence of this particular choice for a starting point, $\boldsymbol{\lambda}_1^{(0)}$ and $\boldsymbol{\lambda}_2^{(0)}$ satisfy the identifiability restrictions.

Note that for each eigenvector v_i, the opposite $-v_i$ can be chosen as a vector of coordinates. This has of course no impact on the distances between the points but will result in a different 2-dimensional plot, namely the symmetric image with respect to the x-axis or the y-axis. In practice, in order to connect the 2-dimensional model with the Bradley–Terry model, the signs for v_1 shall be chosen in order to reflect the main trends in the Bradley–Terry ranking.

3.2. The Newton–Raphson iterative algorithm

A standard method of solving the likelihood equations is Newton–Raphson's method. This method requires the calculations of the first- and second-order derivatives of the log-likelihood $\mathscr{L}$ relative to $\boldsymbol{\lambda}$. The calculations for the derivatives of $\mathscr{L}^*$ are provided in Appendix A. Let $D_{\mathscr{L}^*}(\boldsymbol{\lambda})$ denote the $2n$-vector of first-order derivatives of $\mathscr{L}$ relative to $\boldsymbol{\lambda}$ and $D^2_{\mathscr{L}^*}(\lambda)$, the $2n\times 2n$ matrix of second-order derivatives of $\mathscr{L}^*$ relative to $\boldsymbol{\lambda}$.

With these notations, the Newton–Raphson sequence $(\boldsymbol{\lambda}^{(k)})_{k\geqslant 0}$ of estimates for $\boldsymbol{\lambda}$ are defined as follows:

$$\begin{bmatrix}\boldsymbol{\lambda}^{(k+1)}\\ a^{(k+1)}\end{bmatrix}=\begin{bmatrix}\boldsymbol{\lambda}^{(k)}\\ a^{(k)}\end{bmatrix} - \begin{bmatrix} D^2_{\mathscr{L}^*}(\boldsymbol{\lambda}^{(k)})+a_3^{(k)}\begin{pmatrix}\mathbf{0}_{n,n} & \mathbf{I}_n\\ \mathbf{I}_n & \mathbf{0}_{n,n}\end{pmatrix} & D_\varphi(\boldsymbol{\lambda}^{(k)})\\ D'_\varphi(\boldsymbol{\lambda}^{(k)}) & \mathbf{0}_{3,3}\end{bmatrix}^{-1} \times\begin{bmatrix} D_{\mathscr{L}^*}(\boldsymbol{\lambda}^{(k)})+D_\varphi(\boldsymbol{\lambda}^{(k)})a^{(k)}\\ \varphi(\boldsymbol{\lambda}^{(k)})\end{bmatrix},$$

where

$$D_\varphi(\boldsymbol{\lambda})=\begin{pmatrix}\mathbf{1}_n & \mathbf{0}_n & \boldsymbol{\lambda}_2\\ \mathbf{0}_n & \mathbf{1}_n & \boldsymbol{\lambda}_1\end{pmatrix},$$

is the $2n \times 3$ matrix of first-order derivatives for $\varphi(\boldsymbol{\lambda})$ relative to $\boldsymbol{\lambda}$, $\mathbf{1}_n$ and $\mathbf{0}_n$ are n-vectors with all elements equal to 1 and 0, respectively, and $\mathbf{0}_{k,n}$ is a $k \times n$ matrix with all elements equal to 0. The initial value $a^{(0)}$ for the Lagrange multipliers a can be chosen arbitrarily without major consequences on the convergence.

Standard results that can be found for instance in Silvey (1979) about the likelihood theory under restrictions on the parameters can be applied here to claim that, if the sequence $(\boldsymbol{\lambda}^{(k)})$ converges to a limit $\hat{\boldsymbol{\lambda}}$, then $\hat{\lambda}$ satisfies the identifiability restrictions and is a maximizer of the likelihood function.

3.3. Asymptotic properties of the maximum-likelihood estimator

The usual results about the asymptotic properties of the maximum-likelihood estimator cannot be applied here because the sign parameters do not belong to an open and continuous parametric set. In the sequel, the asymptotic properties of the estimators of the merit vectors are first derived as if the sign parameters σ_{ij} were known. Thereafter, the impact of estimating the signs on the properties of the estimators of the score parameters are investigated by simulations.

3.3.1. Asymptotics when the sign parameters are supposed to be known

In the following, the asymptotic results are derived under the standard assumption that $m_{ij} \to \infty$ and the number n of objects is held fixed. In tournament applications, other asymptotic conditions are sometimes studied, in which the number of pairings are small whereas n is large. Further details can be found in Simons and Yao (1999).

As presented in details in Silvey (1979) for instance, for large numbers of pairings m_{ij}, the maximum-likelihood estimator $\hat{\boldsymbol{\lambda}}$ is normally distributed with mean $\boldsymbol{\lambda}$ and variance-covariance $V_{\hat{\boldsymbol{\lambda}}}$ defined as the $2n \times 2n$ upper diagonal block of the following matrix:

$$\begin{bmatrix} D^2_{\mathscr{L}^*}(\boldsymbol{\lambda}^{(k)}) + a_3^{(k)} \begin{pmatrix} \mathbf{0}_{n,n} & \mathbf{I}_n \\ \mathbf{I}_n & \mathbf{0}_{n,n} \end{pmatrix} & D_{\varphi}(\boldsymbol{\lambda}^{(k)}) \\ D'_{\varphi}(\boldsymbol{\lambda}^{(k)}) & \mathbf{0}_{3,3} \end{bmatrix}^{-1}. \tag{2}$$

This makes it possible to construct confidence regions for the true points $(\lambda_{i,1}, \lambda_{i,2})$. Standard normal distribution theory indeed gives the confidence ellipse with probability level α as delimited by the equation

$$\begin{bmatrix} \hat{\lambda}_{i,1} - \lambda_{i,1} \\ \hat{\lambda}_{i,2} - \lambda_{i,2} \end{bmatrix}' \mathbb{V} \begin{bmatrix} \hat{\lambda}_{i,1} \\ \hat{\lambda}_{i,2} \end{bmatrix}^{-1} \begin{bmatrix} \hat{\lambda}_{i,1} - \lambda_{i,1} \\ \hat{\lambda}_{i,2} - \lambda_{i,2} \end{bmatrix} = c(\alpha),$$

where $c(\alpha)$ is the threshold of a χ^2_2 variate with the probability level α.

Moreover, a multiple comparison procedure can also be defined by testing the significance of the terms logit π_{ij}. This is achieved by calculating the asymptotic variance of logit $\hat{\pi}_{ij}$. Call $J_{ij} = J_{ij}(\lambda_{i,1}, \lambda_{i,2}, \lambda_{j,1}, \lambda_{j,2})$ the 4×1 Jacobian of the derivatives of the predicting transformation given by expression (1) relative to $(\lambda_{i,1}, \lambda_{i,2}, \lambda_{j,1}, \lambda_{j,2})$ and V_{ij}, the 4×4 sub-matrix of $V_{\hat{\boldsymbol{\lambda}}}$ containing the variances and covariances of $\hat{\lambda}_{i,1}$, $\hat{\lambda}_{i,2}$, $\hat{\lambda}_{j,1}$ and $\hat{\lambda}_{j,2}$. An

Table 2
Sequential analysis of deviance table for testing the significance of the 2-dimensional Bradley–Terry model

Models	Deviance	Degrees of freedom
(M_0)	$-2\mathscr{L}(0)+2\mathscr{L}_{\max}$	$\frac{n(n-1)}{2}$
(M_1)	$-2\mathscr{L}(0)+2\mathscr{L}(\hat{\lambda}_{M_1})$	$n-1$
Model (1)	$-2\mathscr{L}(\hat{\lambda}_{M_1})+2\mathscr{L}(\hat{\lambda})$	$n-2$
Residual	$-2\mathscr{L}(\hat{\lambda})+2\mathscr{L}_{\max}$	$\frac{n(n-1)}{2}-(2n-3)$

Table 3
Score vectors and probabilities π_{ij} used for the simulations

Scores				
	1	2	3	4
First dim.	−0.5	0.5	0.5	−0.5
Second dim.	−0.5	−0.5	0.5	0.5

Probabilities π_{ij}				
	1	2	3	4
1		0.27	0.20	0.73
2			0.27	0.20
3				0.27

explicit expression for the Jacobian J_{ij} is easily deduced from the calculations provided in Appendix A. The asymptotic variance $\mathbb{V}[\operatorname{logit}\hat{\pi}_{ij}]$ can now be derived as follows:

$$\mathbb{V}[\operatorname{logit}\hat{\pi}_{ij}]=J'_{ij}V_{ij}J_{ij}.$$

Finally, the relevance of a 2-dimensional display of the preference relative to 1-dimensional sub-models can be tested by the usual likelihood ratio tests. In the present context, the log-likelihood $\mathscr{L}_{\max}$ of the saturated model equals, upto an additive constant, $\sum_{i=1}^{n}\sum_{j=1,\ j\neq i}^{n} n_{ij}\log n_{ij}/m_{ij}$.

It is especially interesting to compare model (1) to the following sub-models:

$$\operatorname{logit}\pi_{ij}=0 \qquad (\mathrm{M}_0)$$

$$\operatorname{logit}\pi_{ij}=\sigma_{ij}|\lambda_i-\lambda_j|. \qquad (\mathrm{M}_1)$$

Of course, model (M_0) is the null model obtained from model (1) by assuming $\boldsymbol{\lambda}=0$. (M_1) is the 1-dimensional model obtained from model (1) by assuming $\boldsymbol{\lambda}_2=0$. It coincides with the Bradley–Terry model if and only if there are no contradictions between the observed preferences and the ranks deduced from the Bradley–Terry scores.

The sequential testing statistics are given in Table 2 together with the degrees of freedom of the corresponding χ^2 distributions.

3.3.2. *Simulation study when the sign parameters are not supposed to be known*

The merit vectors used for the simulation feature are chosen to figure a square whose edges have length 1. They are given in Table 3 together with the probabilities π_{ij}. For each of three situations corresponding to $m=10,\ 30$ or 100 pairings, 1000 data tables have been simulated according to model (1).

Table 4
Biases and root mean squared errors (RMSE) from simulations and asymptotic standard deviation (ASD) from expression (2)

		λ_1				λ_2			
		1	2	3	4	1	2	3	4
$m=10$	Bias	0.045	−0.042	−0.033	0.030	0.047	0.044	−0.047	−0.044
	RMSE	0.332	0.330	0.334	0.329	0.322	0.327	0.324	0.321
	ASD	0.336	0.336	0.336	0.336	0.336	0.336	0.336	0.336
$m=30$	Bias	0.023	−0.016	−0.025	0.018	0.016	0.028	−0.035	−0.009
	RMSE	0.226	0.215	0.215	0.215	0.211	0.215	0.222	0.221
	ASD	0.194	0.194	0.194	0.194	0.194	0.194	0.194	0.194
$m=100$	Bias	0.001	0.001	−0.004	0.002	0.006	0.007	−0.009	−0.004
	RMSE	0.114	0.115	0.112	0.112	0.114	0.111	0.115	0.119
	ASD	0.106	0.106	0.106	0.106	0.106	0.106	0.106	0.106

The biases and root mean squared errors for the estimation of the scores are derived from the simulations. The results are given in Table 4 together with the asymptotic standard deviations given by expression (2). First, note that, in non-asymptotic conditions, the biases of the estimators of the scores seem to be quite important but they tend to disappear for a large number of pairings. Moreover, asymptotically, the asymptotic standard deviations calculated from expression (2) are very close to the root mean squared errors calculated from the simulations.

On the right column of graphs in Fig. 2, histograms of the estimated 1st scores for the 1st point are provided together with the density of a normal distribution with the same expectation and standard deviation as those estimated from the results of the simulations. For $m = 10$, the empirical distribution seems to be quite asymmetric but this asymmetry becomes less important for $m = 30$ and even less for $m = 100$. Similar properties have been observed for the distributions of the other estimated scores. Bivariate histograms not provided here suggest also that, for large values of m, the scores on the x- and the y-axis for a given object are approximately jointly distributed according to a bivariate normal distribution.

The mean-squared error matrix derived from the simulations is now used to calculate confidences ellipses for the score parameters at level 95%. These confidence ellipses are shown on the left column of graphs in Fig. 2 together with the confidence ellipses derived from expression (2). It appears here that the confidence ellipses derived from the simulations are rather close to the ellipses calculated from expression (2).

4. Illustration

The data set used to illustrate the contribution of the 2-dimensional model to the analysis of paired comparisons consists in tasting experiments conducted in the National Higher Institute for Education in the Food Industry in France in 2002. 100 assessors participated

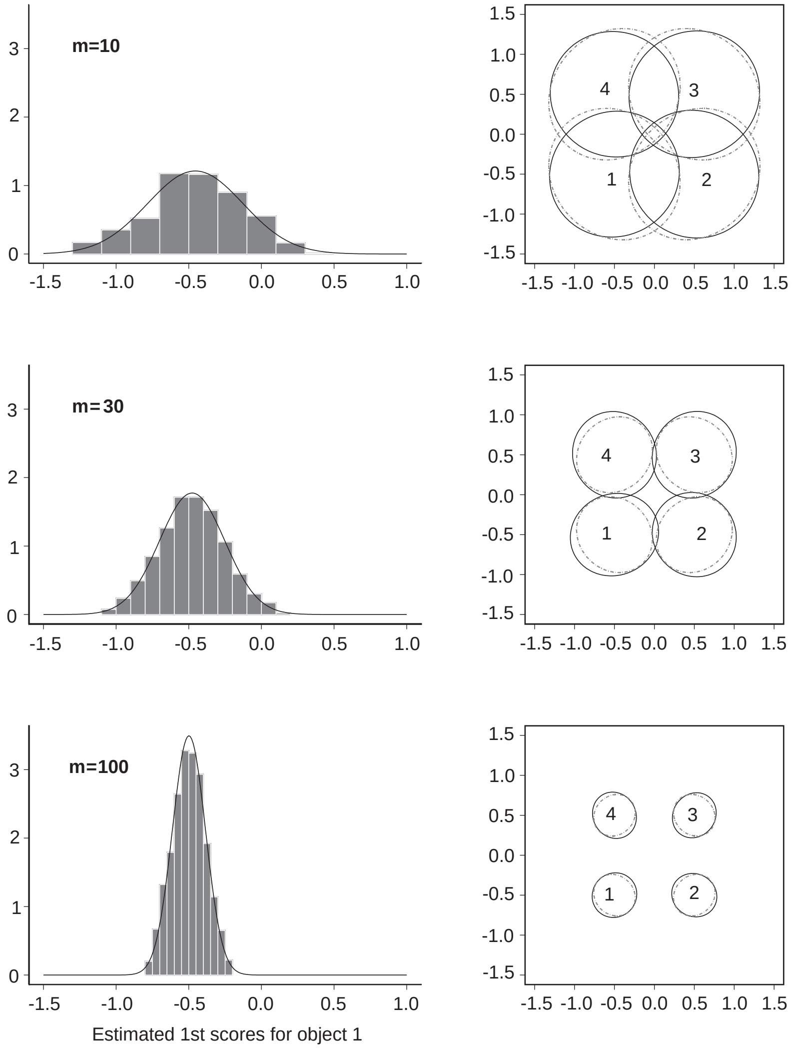

Fig. 2. Left column: histograms of the estimated 1st scores for point 1. Right column: confidence ellipses at level 95% for the score parameters. The solid ellipses are derived from the mean squared error matrix derived from the simulations. The dashed ellipses are derived from expression (2).

in this sensory experiment aiming at an analysis of the preference of cornflakes with regard to their crunchiness. Seven types of cornflakes were used for this experiment. The square table giving the numbers of assessors that have found that the product i was more crunchy than the product j is reproduced in Table 5.

First, a Bradley–Terry model appears to be highly significant for the present data. Moreover, the Bradley–Terry scores given in Table 7 suggests the following ranking of the preference: $6 < 3 < (4, 1, 5) < (7, 2)$.

Table 5
Results from paired comparisons of 7 types of cornflakes by 100 assessors. The cell (i, j) contains the number of assessors that have evaluated the cornflakes i as more crunchy than the cornflakes j

	1	2	3	4	5	6	7
1	0	39	64	40	61	76	46
2	61	0	65	59	55	85	60
3	36	35	0	31	25	41	35
4	60	41	69	0	41	80	28
5	39	45	75	59	0	71	37
6	24	15	59	20	29	0	18
7	54	40	65	72	63	82	0

Table 6
Sequential analysis of deviance table for the cornflakes data

Models	Deviance	Degrees of freedom	p-value
(M_0)	306.90	21	
(M_1)	269.08	6	0.00
Model (1)	25.59	5	1.07×10^{-4}
Residual	12.23	10	

Table 7
Estimated Bradley–Terry scores and estimated coordinates $\boldsymbol{\lambda}_1$ and $\boldsymbol{\lambda}_2$ of model (1)

	Types of cornflakes						
	1	2	3	4	5	6	7
Bradley–Terry scores	0.16	0.53	−0.60	0.12	0.16	−0.85	0.47
$\boldsymbol{\lambda}_1$	0.14	0.44	−0.44	0.27	0.30	−0.95	0.25
$\boldsymbol{\lambda}_2$	0.18	−0.02	0.10	−0.40	−0.16	−0.09	0.39

The analysis of deviance in Table 6 gives the sequential tests listed in Table 2. It is deduced from this table that model (1) is also highly significant relative to its one-dimensional version. The estimated coordinates of each type of cornflakes are provided in Table 7. The main trend reflected by the Bradley–Terry ranking is also obvious from the vector $\boldsymbol{\lambda}_1$ of coordinates.

This is confirmed by Fig. 3 that displays both the Bradley–Terry scores and the 2-dimensional representation together with confidence ellipses at the confidence level 95%. Whereas the 4 groups $6 < 3 < (4, 1, 5) < (7, 2)$ can be distinguished from the Bradley–Terry analysis, the 2-dimensional representation also points out that 6 and 3 have been significantly less preferred than the others, but gives more insight to the sensory distances within the set of products $(1, 2, 4, 5, 7)$.

For example, the Bradley–Terry model superimposed the products 1, 4 and 5. In fact, a look at the data in Table 5 shows that 5 beats 4 in 59% of their paired comparisons, 4 beats

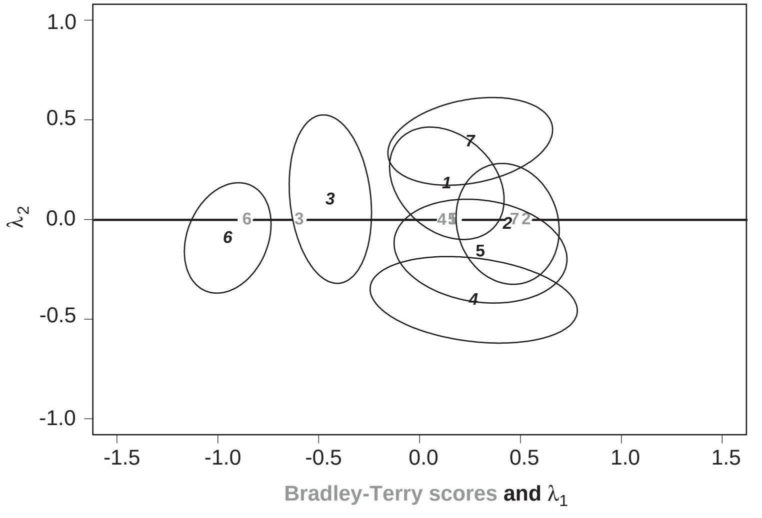

Fig. 3. Bradley–Terry scores and the 2-dimensional representation with confidence ellipses at level 95%.

1 for 60% of their paired comparisons and 1 beat 5 for 61% of their paired comparisons. The 2-dimensional model makes this situation of intransitiveness more obvious.

As another example, 2 beats 7 for 60% of their paired comparisons, 2 and 5 are almost equally preferred (2 beats 5 for 55% of their paired comparisons); 7 beats 5 for 63% of their paired comparisons. But 2 and 7 seem similar on the 1-dimensional graph and 2 seems markedly better than 5. Here, the 2-dimensional graph represents 2 and 5 as closer and separates 2 from 7.

5. Concluding remarks

The extension of the Bradley–Terry model we propose can be seen as a logistic multidimensional scaling model for binomial data. By the way, although the 2-dimensional model has the advantage that the relationships among objects can be readily visualized via a 2-dimensional plot, higher dimensional extensions can be imagined. As recalled by Hunter (2004), estimation in such non-linear logistic models can however be cumbersome as soon as the number of objects is large. Fortunately, in the present situation, the connection with the classical multidimensional scaling issue in the normal framework provides a relevant strategy to choose a starting point for the Newton–Raphson algorithm.

Strictly speaking, modelling paired comparisons by model (1) does not yield a ranking of the objects. However, as it has been shown by the illustrative examples, it can be expected in most situations that the x-axis of the 2-dimensional representation should be partly consistent with the Bradley–Terry ranking. The 2nd dimension is then used to give insight into the relevance of an additive way of modelling the preference. This also means that the analysis of deviance table provides testing tools for the relevance of the Bradley–Terry ranking.

From a practical point of view, at least in sensory analysis, it is almost impossible to obtain a complete paired comparison data table when the numbers of products is too large. Incomplete designs are then used that consist for instance in presenting a set of objects to some assessors and another partly different set of objects to some other assessors. From such designs, complete tables are available for subsets of products but all pairs of objects are not compared. Our current work is to construct predictions of the unobserved paired comparisons by means of the 2-dimensional modelling.

Acknowledgements

We thank the referees for providing us with new references and for their comments that have led to major improvements in the paper.

Appendix A. First- and second-order derivatives of $\mathscr{L}^*$

For all $k = 1, \ldots, n$, for $r = 1, 2$

$$\frac{\partial \mathscr{L}^*}{\partial \lambda_k^{(r)}} = \sum_{i=1}^{n} \sum_{j=i+1}^{n} (n_{ij} - m_{ij}\pi_{ij}) \frac{\partial \operatorname{logit} \pi_{ij}}{\partial \lambda_k^{(r)}},$$

where

$$\frac{\partial \operatorname{logit}(\pi_{ij})}{\partial \lambda_k^{(r)}} = \begin{cases} 0, & \text{if } k \neq i, \quad k \neq j, \\ \sigma_{ij} \dfrac{\lambda_i^{(r)} - \lambda_j^{(r)}}{\sqrt{(\lambda_i^{(1)} - \lambda_j^{(1)})^2 + (\lambda_i^{(2)} - \lambda_j^{(2)})^2}}, & \text{if } k = i, \\ -\sigma_{ij} \dfrac{\lambda_i^{(r)} - \lambda_j^{(r)}}{\sqrt{(\lambda_i^{(1)} - \lambda_j^{(1)})^2 + (\lambda_i^{(2)} - \lambda_j^{(2)})^2}}, & \text{if } k = j. \end{cases}$$

For all $k, l = 1, \ldots, n$, for all $r = 1, 2$, for all $s = 1, 2$

$$\frac{\partial^2 \mathscr{L}^*}{\partial \lambda_k^{(r)} \lambda_l^{(s)}} = \sum_{i=1}^{p} \sum_{j=i+1}^{n} -m_{ij}\pi_{ij}(1 - \pi_{ij}) \frac{\partial \operatorname{logit} \pi_{ij}}{\partial \lambda_k^{(r)}} \frac{\partial \operatorname{logit} \pi_{ij}}{\partial \lambda_l^{(s)}} + \sum_{i=1}^{p} \sum_{j=i+1}^{n} (n_{ij} - m_{ij}\pi_{ij}) \frac{\partial^2 \operatorname{logit} \pi_{ij}}{\partial \lambda_k^{(r)} \lambda_l^{(s)}},$$

where

$$\frac{\partial^2 \operatorname{logit}(\pi_{ij})}{\partial\lambda_k^{(r)}\partial\lambda_l^{(s)}}=\frac{\partial}{\partial\lambda_l^{(s)}}\left[\frac{\partial \operatorname{logit}(\pi_{ij})}{\partial\lambda_k^{(r)}}\right],$$

$$= \begin{cases} 0, & \text{if } k \neq i, \quad k \neq j, \\ 0, & \text{if } l \neq i, \quad l \neq j, \\ \sigma_{ij}\dfrac{u_{r,s}(i,j)}{\left[(\lambda_i^{(1)}-\lambda_j^{(1)})^2+(\lambda_i^{(2)}-\lambda_j^{(2)})^2\right]^{3/2}}, & \text{if } k=l=i, \\ -\sigma_{ij}\dfrac{u_{r,s}(i,j)}{\left[(\lambda_i^{(1)}-\lambda_j^{(1)})^2+(\lambda_i^{(2)}-\lambda_j^{(2)})^2\right]^{3/2}}, & \text{if } k=i, \quad l=j, \\ -\sigma_{ij}\dfrac{u_{r,s}(i,j)}{\left[(\lambda_i^{(1)}-\lambda_j^{(1)})^2+(\lambda_i^{(2)}-\lambda_j^{(2)})^2\right]^{3/2}}, & \text{if } k=j, \quad l=i, \\ \sigma_{ij}\dfrac{u_{r,s}(i,j)}{\left[(\lambda_i^{(1)}-\lambda_j^{(1)})^2+(\lambda_i^{(2)}-\lambda_j^{(2)})^2\right]^{3/2}}, & \text{if } k=l=j, \end{cases}$$

with $u_{r,s}(i,j) = (\lambda_i^{(2)} - \lambda_j^{(2)})^2$ if $r = s = 1$, $u_{r,s}(i,j) = (\lambda_i^{(1)} - \lambda_j^{(1)})^2$ if $r = s = 2$ and $u_{r,s}(i,j) = -(\lambda_i^{(1)} - \lambda_j^{(1)})(\lambda_i^{(2)} - \lambda_j^{(2)})$ if $r \neq s$.

References

Agresti, A., 1990. Categorical Data Analysis. Wiley, New York.

Bradley, R.A., Terry, M.A., 1952. Rank analysis of incomplete block designs. Biometrika 39, 324–345.

David, H.A., 1988. The Method of Paired Comparisons. second ed. Oxford University Press, New York.

Davidson, R.R., 1970. On extending the Bradley–Terry model to accommodate ties in paired comparison experiments. J. Amer. Statist. Assoc. 65, 317–328.

Davidson, R.R., Farquhar, P.H., 1976. A bibliography on the method of paired comparisons. Biometrics 32, 241–252.

De Soete, G., Carroll, J.D., 1983. A maximum likelihood method for fitting the wandering vector model. Psychometrika 48, 553–566.

Dittrich, R., Hatzinger, R., Katzenbeisser, W., 1998. Modelling the effect of subject-specific covariates in paired comparison studies with an application to university rankings. Appl. Statist. 47 (4), 511–525.

Gower, J.C., 1966. Some distance properties of latent root and vector methods used in multivariate analysis. Biometrika 53.3, 325–338.

Hunter, D.R., 2004. MM algorithms for generalized Bradley–Terry models. Ann. Statist. 32, 386–408.

McCullagh, P., 2000. Invariance and factorial models (RSS discussion paper, Oct 13 1999) reply. J. Roy. Statist. Soc. B 62, 209–256.

Pierce, D.A., Schafer, D.W., 1986. Residuals in generalized linear models. J. Amer. Statist. Assoc. 81, 977–986.

Ramsay, J.O., 1982. Some statistical approaches to multidimensional scaling data. J. Roy. Statist. Soc. A 145 (3), 285–312.

Rao, P.V., Kupper, L.L., 1967. Ties in paired-comparison experiments: a generalization of the Bradley–Terry model. J. Amer. Statist. Assoc. 62, 194–204.

Silvey, S.D., 1979. Statistical Inference, third ed. Chapman & Hall, London, 191pp.

Simons, G., Yao, Y.C., 1999. Asymptotics when the number of parameters tends to infinity in the Bradley–Terry model for paired comparisons. Ann. Statist. 27, 1041–1060.

Stigler, S.M., 1994. Citation patterns in the journals of statistics and probability. Statist. Sci. 9, 94–108.