Animal Behaviour 84 (2012) 1523–1531

Contents lists available at SciVerse ScienceDirect

Animal Behaviour

journal homepage: www.elsevier.com/locate/anbehav

Using Markov chain Monte Carlo (MCMC) to visualize and test the linearity assumption of the Bradley–Terry class of models

Aaron Shev [a], Fushing Hsieh [a], Brianne Beisner [b], Brenda McCowan [b,*]

[a] *Department of Statistics, University of California, Davis, Davis, CA, U.S.A.*
[b] *Department of Population Health & Reproduction, California National Primate Research Center, University of California, Davis, Davis, CA, U.S.A.*

ARTICLE INFO

Article history:
Received 7 February 2012
Initial acceptance 19 April 2012
Final acceptance 10 August 2012
Available online 27 October 2012
MS. number: A12-00121R

Keywords:
Bradley–Terry
goodness of fit
linearity
paired comparisons
ranking
rhesus macaque

The construction of dominance hierarchies for animal societies is an important aspect of understanding the nature of social relationships, and the models to calculate dominance ranks are many. However, choosing the appropriate model for a given data set may appear daunting to the average behaviourist, especially when many of these models assume linearity of dominance. Here, we present a method to test whether or not a data set fits the assumption of linearity using the Bradley–Terry model as a representative of the class of models that assume linearity. Our method uses the geometry of a posterior distribution of possible rankings given the data by using a random walk on this distribution. This test is intuitive, efficient, particularly for large number of individuals, and represents an improvement over previous linearity tests because it takes into account all information (i.e. both linear and apparently circular or nonlinear information) from the data with few restrictions due to high dimensionality. Such a test is not only useful in determining whether a linear hierarchy is relevant to a given animal society, but is necessary in justifying the results of any analysis for which the assumption of linearity is made, such as the Bradley–Terry model. If the assumption of linearity is not met, other methods for ranking, such as the beta random field method proposed by Fushing et al. (2011, *PLoS One*, **6**, e17817) should be considered.

Consider a society or group of individuals vying for a position of dominance or stature in a series of pairwise conflicts. Each decisive interaction in which a winner and a loser are identified holds information about the hierarchical structure of this group. The intuitive approach is to use this information to order the individuals from greatest competitive ability to least, as is often done for many sports to determine placement in playoff tournaments. Ranks assigned to teams in this manner are certainly transitive if team A outranks team B and team B outranks team C, then team A outranks team C. But cyclic relations in the data, such as A > B > C > A, can cause problems in determining the appropriate ranking if such a ranking even exists. In animal societies, for example, decisive agonistic interactions among group members are used to reconstruct a social hierarchy. For instance, if one individual threatens another and the second individual runs away, this is an indication of dominance and carries information about the rank of each individual. As with the sports tournament example, all agonistic interactions between pairs of individuals are aggregated and used to determine a rank order. However, the dominance structure in animal societies may be more complex than a simple linear hierarchy. For example, rhesus macaques, *Macaca mulatta*, live in large multimale, multifemale social groups in which both individuals and family groups (matrilines) are constantly competing for status in the society, so conflicting information can cause misleading results in ranking when using a method that assumes transitivity in dominance. Indeed, several captive groups of rhesus macaques at the California National Primate Research Center (CNPRC) exhibit nonlinear, corporative hierarchical structure (Beisner et al. 2011; Fushing et al. 2011b). Here we develop a new approach for estimating ranks that does not require linearity (Fushing et al. 2011a, b).

Constructing a dominance hierarchy from observed agonistic interactions in a group of animals was first described by Schjelderup-Ebbe (1922) for chickens and, since then, there has been much discussion over how to best determine the order for a hierarchy. Most methods fall into one of two classes. The first class of methods finds the appropriate ranking by reordering the rows and columns of the win/loss matrix, (w_{ij}), where w_{ij} is the number of times individual i has prevailed over individual j. The reordering is done by minimizing some numerical criterion calculated for the

* Correspondence: B. McCowan, Department of Population Health & Reproduction, California National Primate Research Center, University of California, Davis, One Shields Avenue, Davis, CA 95616, U.S.A.
E-mail address: bjmccowan@ucdavis.edu (B. McCowan).

0003-3472/$38.00
http://dx.doi.org/10.1016/j.anbehav.2012.09.026

win/loss matrix. This first class is perhaps best characterized by de Vries's (1998) I&SI method, which orders the matrix by minimizing the number of times a pair of individuals has a rank ordering that does not match their empirical dominance, denoted as an inconsistency, and the absolute difference in ranks for an inconsistent pair, denoted as the strength of an inconsistency. More recently this class has been characterized by the beta random field method proposed by Fushing et al. (2011a), which provides a more sophisticated criterion for ordering the matrix than the I&SI method.

The second class of models are based on Thurstone's (1927) method of paired comparison, which utilizes a cardinal dominance index to rank individuals. The method of paired comparisons ranks individuals by the proportion of individuals dominated, which fails to model the randomness in the data and does not use all the information provided. The Bradley–Terry model (Bradley & Terry 1952) solves the issues of the Thurstone model by linking the cardinal dominance indices to the binomial probabilities of dominance through the expit function,

$$p_{ij} = \frac{e^{d_i}}{e^{d_i} + e^{d_j}}. \tag{1}$$

Boyd & Silk (1983) modified the Bradley–Terry model by generalizing this structure and linking the binomial probabilities of dominance to any monotone increasing function, F, of the difference between cardinal dominance indices:

$$p_{ij} = F(d_i - d_j). \tag{2}$$

Thus the Bradley–Terry model becomes a special case of the Boyd & Silk model.

Of the two classes of models, we claim that the first is generally the most reasonable approach to finding an appropriate ordering for a hierarchy. The primary disadvantage of the second class of models is the assignment of a cardinal dominance index to every individual, which inherently assumes that dominance is transitive. Consider a circular relationship among three individuals in which, when laid out in a triangular formation, each individual will win with probability 1 against the competitor directly clockwise. If all dyads have played an equal number of games, it is impossible to determine a reasonable ranking. Complex relationships such as this circular triad are more likely to arise in larger data sets, and thus, the simple assumption of linear structure is less likely to fit the data.

As it stands, Bradley–Terry/Boyd & Silk approach is still the most widely used methodology for ranking inside and outside of academia. For instance, the very popular and widely applied Elo-rating is just another form of this approach (see the recent paper by Neumann et al. 2011). Its popularity is partly built on its easy interpretation and partly on its effective computation. Another important feature of this approach is that no matter whether the critical linearity assumption is satisfied or not, this approach always provides a ranking sequence. That is to say, it becomes extremely convenient because of its easy application, even though the resultant ranking sequence might be poorly supported by the data. We will show that, when the linearity assumption is violated, the Bradley–Terry model may produce counterintuitive orderings. Thus, a rigorous linearity check is essential, especially in animal behavioural researches, which typically contain ranking as a very common and important part of scientific tasks.

We propose a goodness-of-fit test for the Bradley–Terry model to test whether the assumption of dominance transitivity is violated in the data and to show the effect of circular dominance on rankings. We choose the Bradley–Terry model as a representative of the second (Thurstonian) class of models from which to draw conclusions for the entire class for a couple of reasons. First, the deterministic models (i.e. those models that do not incorporate the randomness of the data) are a poor choice because they do not properly weigh the outcomes of likely events (e.g. a strong individual winning) versus unlikely events (e.g. a weak individual winning). Second, precisely which model is chosen as representative is somewhat arbitrary because all models in this class make the assumption of dominance transitivity. For example, the Boyd & Silk model is a generalized version of other models in this class, and it assumes transitive dominance regardless of the monotone increasing function, F, that is chosen. Thus, the only effective difference between models of this type is sensitivity to the assumption. If it is shown that a violation of this assumption produces poor results for a given method, it must be true for all other methods with the same assumption, although the severity of the violation required to produce poor results may vary by method. Finally, given that the choice of the representative of the Thurstonian class of models does not affect the conclusion, we chose the Bradley–Terry model specifically as it is perhaps the most widely used case of Boyd & Silk's model. It has been well studied and generalized in a number of ways including modifications to handle ties by Rao & Kupper (1967), to account for home field advantage by Agresti (1990), and for use in the context of classification by Hastie & Tibshirani (1998). In addition, Adams (2005) proposed a Bayesian extension to the Bradley–Terry model, which is utilized in this paper and described in more detail below.

Other tests have previously been proposed but each of these is limited in scope. Appleby (1983) proposed a test that counts the number of circular triads in the data and compares this to the expected number, but this method is limited because it only identifies circular dominance paths of length three and because empirical data almost never come close to the random expectation of cyclicity (Shizuka & McDonald 2012). Kasuya (1995) also proposed a test that counts all possible data sets possessing fewer violations of linearity, but this test is hindered by computation time when large numbers of individuals are competing. Here, we will visually demonstrate why the Bradley–Terry model fails when the assumption of linearity is not met, and propose a test that is more thorough in evaluation of the assumption, is broadly applicable to data sets of various sizes, and is relatively efficient and intuitive to use.

In the field of animal behaviour, researchers are now more aware of the potential flaws due to violation of the linearity assumption. Although simple linear ranking methods are appealing for reasons of parsimony, large data sets (i.e. those with many subjects) may often violate the assumption and may require the more complex nonparametric approaches (de Vries 1998; Fushing et al. 2011a).

THE BRADLEY–TERRY MODEL AND AN EXTENSION

The Bradley–Terry Model

Suppose we observe a series of independent pairwise competitions among N individuals. For a given dyad (a group of two competitors), denote n_{ij} as the number of competitions between competitors i and j, and denote w_{ij} as the number of competitions in which individual i won. The Bradley–Terry model offers an explicit probability model for estimation of the dominance indices through maximum likelihood. Let p_{ij} be the true probability that player i dominates player j, then the probability that individual i wins w_{ij} times against individual j is given by