ANIMAL BEHAVIOUR, 2008, **76**, 1441–1449
doi:10.1016/j.anbehav.2008.05.030

Available online at www.sciencedirect.com
ScienceDirect

Exact conditional tests for log-linear models: application to animal dominance

JOHN M. ROBERTS, JR
University of New Mexico

(Received 27 October 2007; initial acceptance 11 January 2008; final acceptance 5 May 2008; published online 23 August 2008; MS. number: AS-1318R)

Keywords: chi-square; Markov chain Monte Carlo; model fit; statistical method

Analyses of summary data on the outcomes of animal dominance contests can be roughly classified as either ranking animals without appeal to an explicit probability model or fitting such a model, with characterization of the dominance structure based on the form of the model and estimates of its parameters (Tufto et al. 1998; de Vries 1998; Adams 2005). The straightforward description of hierarchy obtained with ranking approaches is very valuable, but working within the context of an explicit probability model also offers some important benefits. First, this framework includes formal mechanisms for evaluating a model's fit to the observed data. If a model does not fit, and is therefore an unsatisfactory description of the observed data, behavioural interpretations of estimated model parameters would be untrustworthy. For instance, model parameters could be used to estimate individuals' fighting strengths, and these estimates in turn used to examine the association between strength and such characteristics as body size or age, or the propensities for stronger and weaker animals to engage in contests. The resulting conclusions are more defensible when the model gives an adequate description of the observed dominance data than when it does not. With formal assessment of model fit, we know whether we are basing interpretations on the parameters of well-fitting models. Similarly, when a model embodies an idea about underlying dominance structure, assessing the model's fit helps us decide whether that idea is reasonable. Second, estimation under an explicit probability model allows for consideration of uncertainty in parameter estimates, with the observed data used to calculate estimates viewed as one, but not the only, possible realization from the assumed model.

Of course ranking and modelling approaches are not mutually exclusive, as rankings may be obtained from estimates of a probability model; with the Bradley & Terry (1952) model, for instance, estimated strengths are obtained along with an assessment of the model's fit to the observed data. But there are desirable features of the modelling framework even when it is used to determine a ranking. We can compare fit of other models, implying different sets of underlying outcome probabilities, to that of Bradley–Terry, and possibly decide that estimates under an alternative model should be used to determine the ranking. Also, we can conduct formal hypothesis tests, perhaps of the hypothesis that two particular animals have the same strength. In any event, I focus here on methods for fitting probability models to dominance data.

Log-linear models are an important subset of probability models for dominance data. Log-linear models developed as general methods for analysing frequency tables in which observations are cross-classified according to two (or more) categorical variables (Bishop et al. 1975). These models are written in terms of the logged expected cell counts in the table, with the counts themselves assumed to be distributed as Poisson. Typical models consider factors such as the table's marginal totals and the nature of the association between the row and column variables in explaining the pattern of cell counts. This family of models is very flexible in allowing different specifications representing various substantive hypotheses.

Because dominance data on a set of K animals are often displayed in a $K \times K$ table reporting the number of wins by each animal in all the pairs, log-linear models are a natural analytic choice. A great variety of log-linear models can be constructed, with different conceptions of the nature of

Correspondence: J. M. Roberts, Jr, Department of Sociology, MSC05 3080,University of New Mexico, Albuquerque, NM 87131, U.S.A. (email: jmrob@unm.edu).

0003–3472/08/$34.00/0

dominance structure expressed through the outcome probabilities implied by the model. Rather than, say, positing a general concept like 'linear hierarchy', a model would specify the form of outcome probabilities (which could, of course, be consistent with 'linear hierarchy' or some other general notion). For example, a simple model might be built on the idea that there is really no structure, and each dominance contest is like a coin flip. Or, as in Bradley–Terry, the model might follow from the idea that probabilities of outcomes are a function of a numerical strength for each animal. Still another possibility is that there is some ranking of animals, and that in each contest there is a fixed probability of the higher-ranked animal defeating the lower. (I examine models like these in the examples below.) A model's fit to data helps decide whether the hypothesized dominance structure underlying the model is plausible, with parameter estimates giving further insight into the dominance relations.

A likely unrealistic assumption in fitting log-linear models is that the observed contest outcomes are independent of each other. The outcome of one contest may well influence the outcome of later contests, in that pair or others. Hsu et al. (2006) reviewed many theoretical and empirical studies of experience effects (and possible mechanisms for these effects) on dominance outcomes. Experience effects can be associated with the winning and losing histories of the animals involved in a particular contest, and may influence the emergence and stability of dominance hierarchies. While noting that experience effects may vary by species, and can be complicated by differences across groups in the distribution of dominance-relevant characteristics such as size, Hsu et al.'s review shows that dominance contests are not in general strictly independent. In addition to individual experience effects, animals' awareness of contests between others (observer effects) and of their own earlier encounters with specific others may contribute to nonindependence. Hierarchies may develop differently if individual recognition leads to avoidance of contests with known stronger animals, or better assessment of opponents' abilities (Hsu et al. 2006). Studies have shown an effect of individual recognition on dominance even in such animals as lobsters (Karavanich & Atema 1998) and swordtail fish (Morris et al. 1995).

Given these indications of nonindependence in sets of observed dominance outcomes, log-linear models for summary dominance tables would probably not be preferred if detailed information on the temporal ordering (or physical proximity) of dominance contests were available. However, information on the sequence (or location) of contests is quite rare in published accounts. Jameson et al. (1999) also suggested that in data collected over a period that is brief relative to the length of the group's existence, violations of the independence assumption may not be so severe. Note too that all descriptive or modelling efforts involve some approximation to the full richness of the phenomenon being studied. When only summary data are available, then, log-linear models offer a fruitful approach to understanding patterns in dominance data.

As discussed above, assessment of a model's fit to observed data is critical to this enterprise, both as a key to deciding which specification of the dominance structure is most consistent with the data, and to base substantive interpretations on parameter estimates from well-fitting models. Standard methods for assessing fit, however, are questionable when the data are sparse, with many small counts in the table's cells. This issue is critical for applications to animal dominance, as the fact that dominance data tables usually have many cells with small (often zero) counts could be a serious obstacle to the use of log-linear models. However, recent decades have seen considerable development of statistical methods for assessing model fit in sparse tables. 'Exact conditional tests' take advantage of both increasing computer power and continuing theoretical advances, with the result that sparseness is not as great a problem as it was in the past.

After reviewing log-linear models for dominance data, I discuss some recent methods from the statistical literature for assessing fit of these models in sparse tables. I then apply these methods to two example data sets, with results of traditional methods for assessing fit compared to those from exact conditional tests.

Log-linear Models

Classic references on log-linear models for tables cross-classifying observations by two or more categorical variables include Bishop et al. (1975), Haberman (1979), Fienberg (1980) and Agresti (2002). These works provide many details that are omitted in the brief review here. With $\boldsymbol{\mu}$ representing the vector of expected counts in the table's cells, a log-linear model has the form $\log \boldsymbol{\mu} = \mathbf{X}\,\boldsymbol{\lambda}$, where $\boldsymbol{\lambda}$ is a vector of parameters and $\mathbf{X}$ is a matrix representing the model. Estimates of $\boldsymbol{\mu}$ and $\boldsymbol{\lambda}$ are obtained via the vector $\mathbf{f}$ of observed cell counts. Observed counts are assumed to be distributed as Poisson, or in another way that implies the same likelihood function. (In a dominance data table, the diagonal cells are not meaningful, so for the present purposes $\boldsymbol{\mu}$ and $\mathbf{f}$ represent the vectors of expected and observed counts in the off-diagonal cells.) In this notation $\boldsymbol{\lambda}$ refers to a set of nonredundant parameters; in general, more accessible ways of writing the model will indicate more parameters, but constraints allow these additional parameters to be obtained from those in the vector $\boldsymbol{\lambda}$.

The $\mathbf{X}$ matrix defines the model, but typically a variety of equivalent constructions exist for a given model. The key is that $\mathbf{X}$ indicates the model's sufficient statistics. In this setting, these are summary statistics that will, under maximum likelihood estimation of the model, be equated in the observed and expected counts. That is, $\mathbf{X}'\,\boldsymbol{\mu}^* = \mathbf{X}'\,\mathbf{f}$, where $\boldsymbol{\mu}^*$ denotes the (maximum likelihood) estimated expected cell counts. For example, in a two-way table the sufficient statistics for the log-linear model of independence (or no association between row and column variables) are the row and column totals, and the table of estimated expected counts will have the same row and column totals as the table of observed counts. The independence model can be represented by different constructions of the $\mathbf{X}$ matrix, but in all cases $\mathbf{X}'\,\mathbf{f}$ will give nonredundant summary statistics that (1) allow recovery of the row and column totals and (2) equal the