Working Paper Series, N. 1, January 2012

Department of Statistical Sciences
University of Padua
Italy

Hybrid pairwise likelihood analysis of animal behavior experiments

Manuela Cattelan
Department of Statistical Sciences
University of Padua
Italy

Cristiano Varin
D.A.I.S.
Ca' Foscari Univerity, Venice
Italy

Abstract: The study of the determinants of contests between animals is an important issue in understanding animal behavior. Tournament experiments among a set of animals are used by zoologists for this purpose. From a statistical point of view, the results of these tournament experiments are naturally analyzed by paired comparison models such as the Bradley-Terry and the Thurstone models. A major complication is the presence of dependence between the outcomes of couples of contests with an animal in common. Likelihood analysis of this type of animal behavior experiments in presence of interdependence between contests is computationally demanding. An alternative fitting method that mixes optimal estimation equations and pairwise likelihood inference is then suggested. The performance of the proposed methodology is investigated by simulation studies and then applied to a real data set about adult male Cape Dwarf Chameleons.

Keywords: Animal behavior experiments, Generalized estimating equations, Multivariate probit regression, Pairwise comparisons, Pairwise likelihood, Thurstone model.

UNIVERSITÀ DEGLI STUDI DI PADOVA
DIPARTIMENTO DI SCIENZE STATISTICHE

Final version (2012-01-24)

Contents

Department of Statistical Sciences
Via Cesare Battisti, 241
35121 Padova
Italy

tel: +39 049 8274168
fax: +39 049 8274170
http://www.stat.unipd.it

Corresponding author:
Manuela Cattelan
tel: +39 049 827 4124
manuela.cattelan@unipd.it

Hybrid pairwise likelihood analysis of animal behavior experiments

Manuela Cattelan
Department of Statistical Sciences
University of Padua
Italy

Cristiano Varin
D.A.I.S.
Ca' Foscari Univerity, Venice
Italy

Abstract: The study of the determinants of contests between animals is an important issue in understanding animal behavior. Tournament experiments among a set of animals are used by zoologists for this purpose. From a statistical point of view, the results of these tournament experiments are naturally analyzed by paired comparison models such as the Bradley-Terry and the Thurstone models. A major complication is the presence of dependence between the outcomes of couples of contests with an animal in common. Likelihood analysis of this type of animal behavior experiments in presence of interdependence between contests is computationally demanding. An alternative fitting method that mixes optimal estimation equations and pairwise likelihood inference is then suggested. The performance of the proposed methodology is investigated by simulation studies and then applied to a real data set about adult male Cape Dwarf Chameleons.

Keywords: Animal behavior experiments, Generalized estimating equations, Multivariate probit regression, Pairwise comparisons, Pairwise likelihood, Thurstone model.

1 Introduction

In nature it is common to observe fights between animals belonging to the same species. The results of such contests are of vital importance since they determine the access to food and the possibility to reproduce. Often, the competitions between animals are settled through aggressive behavior of the contestants and the withdraw of one animal, but sometimes they escalate to real contests. Experiments on contests between animals are performed with different aims, as for example finding the reasons of the escalation to real contests, analyzing the length of the contests, determining whether animals can assess the strength of the opponent and decide to withdraw early from the contest.

The outcomes of contests can be viewed as the results of paired comparisons from which a winner and a loser can be detected. Paired comparison data arise in many areas, from sport tournaments to psychometric analysis, from sensory testing to genetics, see Böckenholt (2006) and Cattelan (2011) for reviews on this topic. Stuart-Fox *et al.* (2006) employ a paired comparison model to study the association between the results of contests fought by chameleons and various covariates related to the size and the mass of the animals. Another work along this direction is Brown *et al.* (2006) that analyze the outcomes of fights between house crickets as a function of relative size and relative body mass of animals.

Stuart-Fox *et al.* (2006) list many advantages in designing the animal behavior study as a sort of tournament and using a paired comparison model for the analysis of the data. In particular, models for paired comparison data allow to include in the model multiple independent variables whose importance is assessed simultaneously and they take into account the participation of the same animal in more than one contest. Moreover, those models can be applied to incomplete tournaments data in which an animal fights against several other animals, but not against all other animals included in the study.

In tournament experiments, the same animal is used in different contests, hence inducing dependence between pairs of outcomes with an animal in common. In Stuart-Fox *et al.* (2006) the dependence in the observed data is accounted for through a covariate that counts the number of wins of the animals in previous contests. Also Kemp *et al.* (2006) recognize the problem of dependence since they observe animals in three subsequent contests. Kemp *et al.* employ an ordinal logistic model for the probability of an animal winning none, one, two or all three contests. Reference animals are identified and the results of animals competing with them are discarded, thus losing also a part of the total comparisons, namely only 102 out of 128 matches were used. The covariates included in the regression were computed as the value of the covariate for the reference animal minus the mean value of the three opponents.

In this manuscript, we suggest to analyze animal behavior experiments with a marginal Thurstone model designed in way to take into account the dependence of contests involving a common animal. The methodology is motivated and illustrated by the data analyzed in Stuart-Fox *et al.* (2006) and made publicly available through the `R` (R Development Core Team, 2011) package `BradleyTerry2` (Turner and Firth, 2011). The data consist in 106 outcomes of contests between 35 adult male Cape Dwarf Chameleons. The chameleons were grouped according to the size. Each animal fought against every other animal in the same quad. Then, animals competed with other chameleons of the next larger or smaller quad depending on whether they won or lost contests with animals in the same quad. The number of contests per animal varies from 3 to 9, the average number is 6.06. Ten animals lost all the contests in which they were involved and only two won all contests fought. Several covariates were recorded for each animal: the snout-vent length, body mass, tail length, jaw length and casque height. These measures were converted into size-free variables by taking the residuals of the variables regressed against the snout-vent length. A particular feature of the Cape Dwarf Chameleons is an irregular oval patch in the flank, which sometimes presents small peripheral patches. The extension of the entire flank patch and the main pink patch on the flank were computed as a proportion of the total flank area and then reported on the arcsine scale. See Table 1 for a list of the available covariates.

Table 1: Description of the covariates available for the male Cape Dwarf Chameleons data (Stuart-Fox *et al.*, 2006). Source: `BradleyTerry2` package (Turner and Firth, 2011).

SVL	snout-vent length
CH	residuals of casque height regression on SVL
JL	residuals of jaw length regression on SVL
TL	residuals of tail length regression on SVL
MASS	residuals of body mass regression on SVL
MP	proportion (arcsin transformed) of area of the flank occupied by the main pink patch on the flank
FP	proportion (arcsin transformed) of area of the flank occupied by the entire flank patch

The manuscript is organized as follows. Section 2 introduces basic models for the analysis of paired comparison data, such as the Bradley-Terry and the Thurstone models. In Section 3 a version of the Thurstone model that accounts for dependence in the data is described. Ordinary likelihood inference for this model is particularly cumbersome. In order to overcome the difficulties with the computation of the likelihood function, we suggest to resort to an iterative fitting method that cycles between optimal estimating equations and composite likelihood inference. Section 4 illustrates the finite sample performance of the proposed methodology by simulations, while Section 5 applies the proposed model to the chameleons data. Section 6 discusses the extension needed to analyze paired comparison experiments that allow for ties and Section 7 concludes. The `R` code implementing the analysis discussed in this manuscript is available upon request from the Authors.

2 The Thurstone independence model

Consider a pairwise experiment involving n different animals and let Y_{ijr} denote the binary random variable measuring the outcome of the rth contest between animals i and j,

$$\mathrm{Y}_{ijr} = \begin{cases} 1, & \text{if animal } i \text{ beats animal } j, \\ 0, & \text{if animal } j \text{ beats animal } i, \end{cases}$$

with $i, j = 1, \dots, n$ and $r = 1, \dots, m_{ij}$. Generally, animal behavior experiments are incomplete tournaments in the sense that not all contests between all the possible pairs that can be formed from the n animals are observed. Often these tournaments are also unbalanced because the number of contests vary from animal to animal. To account for the incompleteness and unbalanceness of the tournament, we let m_{ij} be zero if no contests between animals i and j are observed. The total number of contests is denoted by $m = \sum_{i<j} m_{ij}$.

We focus on situations where the interest lies in evaluating whether a p-dimensional vector of animal-specific covariates $\boldsymbol{x}_i$, $i = 1, \ldots, n$, has some predictive value about the contest outcome. In paired comparisons models, the expected value of the observed outcome in the rth paired comparison between object i and object j is expressed as

$$\pi_{ijr} = \mathrm{E}(\mathrm{Y}_{ijr}) = \mathrm{F}\left\{(\boldsymbol{x}_i - \boldsymbol{x}_j)^{\mathrm{T}}\boldsymbol{\beta}\right\}, \quad (1)$$

where $\boldsymbol{\beta}$ is a p-dimensional vector of unknown regressor coefficients which does not include an intercept term because this cannot be identified. Function $\mathrm{F}(\cdot)$ in (1) is the cumulative distribution function of a zero symmetric continuous variate. Popular choices for $\mathrm{F}(\cdot)$ are the logistic distribution leading to the Bradley-Terry model (Bradley and Terry, 1952) and the normal distribution leading to the Thurstone probit model (Thurstone, 1927). As it is well known, the similarity between the shapes of the normal and logistic distributions implies that logistic and probit regressions differ very slightly. However, for the subsequent developments in this manuscript, the probit Thurstone specification is more convenient and thus it is henceforth adopted. Accordingly, expectation (1) becomes

$$\pi_{ijr} = \Phi\left\{(\boldsymbol{x}_i - \boldsymbol{x}_j)^{\mathrm{T}}\boldsymbol{\beta}\right\},$$

where $\Phi(\cdot)$ is the cumulative distribution function of a standard normal variate.

The standard analysis is based on maximum likelihood estimation of $\boldsymbol{\beta}$ under the assumption of independence of the outcomes of the pairwise comparisons conditionally on the observed covariates. The *independence likelihood* for $\boldsymbol{\beta}$ is

$$\mathscr{L}_{\mathrm{ind}}(\boldsymbol{\beta}) = \prod_{i<j}^{n}\prod_{r=1}^{m_{ij}} \Phi\left\{(2y_{ijr} - 1)(\boldsymbol{x}_i - \boldsymbol{x}_j)^{\mathrm{T}}\boldsymbol{\beta}\right\}, \quad (2)$$

where it is implicitly intended that if there are no observed contests between animals i and j for some choice of indices i and j, then $m_{ij} = 0$ and the corresponding terms are dropped from the product. The value of the parameter that maximises the independence likelihood is denoted by $\hat{\boldsymbol{\beta}}_{\mathrm{ind}}$. Let $\boldsymbol{\pi}$ be the vector containing the model-based expectations for all the observed comparisons $\{\pi_{ijr} : i < j, r = 1, \ldots, m_{ij}\}$ and let $\boldsymbol{y}$ be the vector of the corresponding observations. Then, the maximum independence likelihood estimator $\hat{\boldsymbol{\beta}}_{\mathrm{ind}}$ is obtained by solving the likelihood equations

$$\mathbf{D}(\boldsymbol{\beta})^{\mathrm{T}}\mathbf{V}_{\mathrm{ind}}(\boldsymbol{\beta})^{-1}\{\boldsymbol{y} - \boldsymbol{\pi}(\boldsymbol{\beta})\} = \mathbf{0}, \quad (3)$$

where $\mathbf{D}(\boldsymbol{\beta})$ is the Jacobian of $\boldsymbol{\pi}$ and $\mathbf{V}_{\mathrm{ind}}(\boldsymbol{\beta}) = \mathrm{var}(\mathbf{Y})$ is the model-based variance of the outcomes under the independence assumption, that is a diagonal matrix with entries $\pi_{ijr}(\boldsymbol{\beta})\{1 - \pi_{ijr}(\boldsymbol{\beta})\}$. Under the independence assumption, $\hat{\boldsymbol{\beta}}_{\mathrm{ind}}$ has asymptotic normal distribution with mean $\boldsymbol{\beta}$ and variance $\mathbf{D}(\boldsymbol{\beta})^{-1}\mathbf{V}_{\mathrm{ind}}(\boldsymbol{\beta})\mathbf{D}(\boldsymbol{\beta})^{-\mathrm{T}}$.

3 Accounting for dependence between contests

The main reason of criticism of the above standard analysis is the assumption of independence between contests sharing an animal. The intent of this paper is to suggest an approach to handle dependence in paired comparisons data while keeping the simplicity of the marginal interpretation of the regression coefficients of the independence analysis.

The independence Thurstone model can be written as a censored linear regression model

$$\begin{aligned} \mathrm{Y}_{ijr} = 1 \quad &\leftrightarrow \quad \mathrm{Z}_{ijr} > 0, \\ \mathrm{Z}_{ijr} &= (\boldsymbol{x}_i - \boldsymbol{x}_j)^{\mathrm{T}} \boldsymbol{\beta} + \epsilon_{ijr}, \end{aligned} \tag{4}$$

where the hidden errors ϵ_{ijr} are independent and identically distributed standard normal variables, $\epsilon_{ijr} \overset{\text{i.i.d.}}{\sim} \mathrm{N}(0,1)$. A sensible extension of the independence model which takes into account the dependence among comparisons consists in replacing the independent hidden errors with the following correlated hidden errors

$$\begin{aligned} \epsilon_{ijr} &= u_i - u_j + \eta_{ijr}, \\ u_i &\overset{\text{i.i.d.}}{\sim} \mathrm{N}(0, \sigma^2), \\ \eta_{ijr} &\overset{\text{i.i.d.}}{\sim} \mathrm{N}(0, 1 - 2\sigma^2), \end{aligned} \tag{5}$$

where u_i and η_{ijr} are animal-specific and match-specific components, respectively. The variance of η_{ijr} is constrained to $1 - 2\sigma^2$ in order to retain the same scale of the hidden errors as in the independence Thurstone model and thus make the comparison of the results easier. This constraint causes no loss of generality because in any case the variance of the hidden errors must be constrained to ensure model identifiability. Clearly, this model specification requires that parameter σ^2 lies in the interval $(0, 1/2)$.

The hidden errors defined in (5) are not independent standard normal variates anymore but they follow a multivariate normal distribution with zero mean vector, unit variances and correlation matrix $\mathbf{P}$, whose non-diagonal entries are zero except for those corresponding to couples of contests with one or both the animals in common. More precisely, the correlation matrix $\mathbf{P}$ is

$$\mathbf{P} = \sigma^2 \mathbf{A}\mathbf{A}^{\mathrm{T}} + \left(1 - 2\sigma^2\right) \mathbf{I}_m, \tag{6}$$

where $\mathbf{A}$ is the $m \times n$ matrix that identifies the pairwise comparisons and $\mathbf{I}_m$ is the $m \times m$ identity matrix. The typical row of $\mathbf{A}$ corresponding to a contest between animals i and j has zero entries everywhere but at the ith and jth columns which have entries 1 and -1, respectively. Consequently, the entries of $\mathbf{P} = [\rho]$ depend on which animals are involved in the couple of contests, that is the correlation between the rth contest fought by animals i and j and the sth contest fought by animals k and l is

$$\rho = \begin{cases} -\sigma^2, & \text{if } i = l \text{ or } j = k, \\ \sigma^2, & \text{if } i = k \text{ and } j \neq l \text{ or } j = l \text{ and } i \neq k, \\ 2\sigma^2, & \text{if } i = k, j = l \text{ and } r \neq s, \\ 1, & \text{if } i = k,\ j = l \text{ and } r = s, \\ 0, & \text{otherwise.} \end{cases} \tag{7}$$

As an illustration, consider a small tournament with four animals and five contests observed, namely two contests between animals 1 and 2 and one for each of the following pairs: 1 vs 3, 3 vs 4 and 1 vs 4. Then, the tournament is identified by the following matrix $\mathbf{A}$

$$\mathbf{A} = \begin{pmatrix} 1 & -1 & 0 & 0 \\ 1 & -1 & 0 & 0 \\ 1 & 0 & -1 & 0 \\ 0 & 0 & 1 & -1 \\ 1 & 0 & 0 & -1 \end{pmatrix}$$

and the corresponding correlation matrix of the errors is

$$\mathbf{P} = \begin{pmatrix} 1 & 2\sigma^2 & \sigma^2 & 0 & \sigma^2 \\ 2\sigma^2 & 1 & \sigma^2 & 0 & \sigma^2 \\ \sigma^2 & \sigma^2 & 1 & -\sigma^2 & \sigma^2 \\ 0 & 0 & -\sigma^2 & 1 & \sigma^2 \\ \sigma^2 & \sigma^2 & \sigma^2 & \sigma^2 & 1 \end{pmatrix}.$$

The dependence model defined by equations (4)-(5) is, thus, characterized by:

(a) a marginal interpretation of the regression coefficients because $\pi_{ijr} = \Phi\left\{(\boldsymbol{x}_i - \boldsymbol{x}_j)^{\mathrm{T}}\boldsymbol{\beta}\right\}$ as in the independence model;

(b) cross-correlation between couples of contests sharing one or both the animals. In fact, the bivariate probability function of the proposed marginal dependence Thurstone model is

$$\mathrm{pr}(\mathrm{Y}_{ijr} = y_{ijr}, \mathrm{Y}_{kls} = y_{kls}) = \Phi_2\left\{(2y_{ijr} - 1)(\boldsymbol{x}_i - \boldsymbol{x}_j)^{\mathrm{T}}\boldsymbol{\beta}, (2y_{kls} - 1)(\boldsymbol{x}_k - \boldsymbol{x}_l)^{\mathrm{T}}\boldsymbol{\beta}; \rho\right\}, \tag{8}$$

with correlation ρ specified by equation (7).

The model described in this section will be called the marginal dependence Thurstone model.

3.1 Likelihood inference

The marginal dependence Thurstone model is an example of multivariate probit model. Inference in multivariate probit models is notoriously difficult because the likelihood requires the computation of the probability distribution function of a multivariate normal variate. Let $\boldsymbol{w}$ be the vector containing the terms $w_{ijr} = (2y_{ijr} - 1)(\boldsymbol{x}_i - \boldsymbol{x}_j)^{\mathrm{T}}\boldsymbol{\beta}$. Then, the likelihood of the marginal dependence Thurstone model is

$$\mathscr{L}_{\mathrm{dep}}(\boldsymbol{\beta}, \sigma^2) = \Phi_m\left(\boldsymbol{w}; \mathbf{P}\right), \tag{9}$$

where $\Phi_m(\cdot; \mathbf{P})$ is the probability distribution function of an m-dimensional multivariate standard normal variate with correlation matrix $\mathbf{P}$ as defined in equation (6). Clearly, if $\sigma^2 = 0$, then likelihood (9) reduces to the likelihood (2) of the independence Thurstone model.

Given the central role of the normal distribution, computation of multivariate normal probability functions has been extensively studied. For low dimensions, various deterministic numerical integration rules are available (Joe, 1995; Miwa *et* *al.*, 2003). However, these rules are subject to the curse of dimensionality and cannot be employed for moderate to large dimensions, as those occurring in the type of animal behavior experiments that motivated this paper. For example, in the Male Cape Dwarf Chameleons data later analyzed the number of contests is $m = 106$, a dimension large enough to prevent the use of deterministic numerical integration, at least with actual technologies. The alternative is, then, to resort to Monte Carlo or Quasi Monte Carlo integration rules; some useful references are Hajivassiliou *et al.* (1996), Chib and Greenberg (1998), Genz and Bretz (2002) and Jeliazkov and Lee (2010).

However, also Monte Carlo methods have some limitations because the computational cost for accurate approximation of the likelihood raises with the dimension, although not as heavily as for the deterministic rules. Hence, Monte Carlo approximations may not be appropriate if the intent is to identify a method that can be applied to a generic animal behavior experiment with a potentially large value of contests m. Furthermore, the statistical robustness of likelihood inference for the marginal dependence Thurstone model can be questioned because of the underlying multivariate normal assumption which seems very difficult to fully assess with any diagnostic method.

3.2 The hybrid pairwise likelihood approach

The optimal estimating equations for $\boldsymbol{\beta}$ under the dependence Thurstone model have the same form as the independence likelihood equations (3) but with a different variance matrix

$$\mathbf{D}(\boldsymbol{\beta})^{\mathrm{T}}\mathbf{V}_{\mathrm{dep}}(\boldsymbol{\beta},\sigma^2)^{-1}\left\{\boldsymbol{y}-\boldsymbol{\pi}(\boldsymbol{\beta})\right\}=\mathbf{0}, \tag{10}$$

where the variance matrix $\mathbf{V}_{\mathrm{dep}}(\boldsymbol{\beta},\sigma^2)$ is now computed under the assumed dependence model and thus it has entries of type

$$\begin{aligned}\mathrm{cov}(\mathrm{Y}_{ijr},\mathrm{Y}_{kls}) &= \mathrm{pr}(\mathrm{Y}_{ijr}=1,\mathrm{Y}_{kls}=1)-\mathrm{pr}(\mathrm{Y}_{ijr}=1)\mathrm{pr}(\mathrm{Y}_{kls}=1)\\ &= \Phi_2\left\{(\boldsymbol{x}_i-\boldsymbol{x}_j)^{\mathrm{T}}\boldsymbol{\beta},(\boldsymbol{x}_k-\boldsymbol{x}_l)^{\mathrm{T}}\boldsymbol{\beta};\rho\right\}-\Phi\left\{(\boldsymbol{x}_i-\boldsymbol{x}_j)^{\mathrm{T}}\boldsymbol{\beta}\right\}\Phi\left\{(\boldsymbol{x}_k-\boldsymbol{x}_l)^{\mathrm{T}}\boldsymbol{\beta}\right\}.\end{aligned}$$

The optimal estimating equations are not only theoretically attractive because they provide the optimal linear combination of the unbiased score equations $\{\mathbf{Y}-\boldsymbol{\pi}(\boldsymbol{\beta})\}$ but also appealing both from a computational point of view and for their robustness since only bivariate normal probabilities are involved.

Let $\hat{\boldsymbol{\beta}}_{\mathrm{dep}}(\sigma^2)$ denote the solution of the optimal estimating equations (10). The dependence parameter σ^2 appearing in the optimal estimating equations for $\boldsymbol{\beta}$ is not known and it is neither possible to estimate it from some estimating function that does not depend on $\boldsymbol{\beta}$. We choose, then, to resort to the *hybrid pairwise likelihood* approach by Kuk (2007). This is an iterative algorithm that cycles between optimal estimating equations for estimation of $\boldsymbol{\beta}$ given the current value of σ^2 and maximum pairwise likelihood estimation of σ^2 given the current value of $\boldsymbol{\beta}$. This hybrid method is particularly attractive in our context because it provides a receipt for estimation of σ^2 that again involves only bivariate distributional aspects.

Given the current value of $\boldsymbol{\beta}$, the hybrid parwise likelihood approach suggests to estimate σ^2 by maximization of the *pairwise likelihood* (Varin, Reid and Firth, 2011).

This is an example of a more general class of pseudolikelihoods called composite likelihoods (Lindsay, 1988) constructed by pooling together valid likelihoods for subsets of data. Composite likelihoods are used as surrogate of the full likelihood when this is difficult to compute or to specify. See, for example, Cox and Reid (2004), Molenberghs and Verbeke (2005) and the review by Varin, Reid and Firth (2011). The pairwise likelihood is the product of the likelihoods for all the couples of contests

$$\mathscr{L}_{\rm pair}(\sigma^2) = \prod_{i<j}^{n} \prod_{k<l}^{n} \prod_{r=1}^{m_{ij}} \prod_{s=1}^{m_{kl}} {\rm pr}({\rm Y}_{ijr} = y_{ijr}, {\rm Y}_{kls} = y_{kls}),$$

with the bivariate probabilities computed as in (8). Since only pairs of contests with one or both the animals in common are correlated, and thus informative on σ^2, then the pairwise likelihood reduces to

$$\mathscr{L}_{\rm pair}(\sigma^2) = \prod_{i<j}^{n} \prod_{r<s}^{m_{ij}} {\rm pr}({\rm Y}_{ijr} = y_{ijr}, {\rm Y}_{ijs} = y_{ijs}) \prod_{i<j<k}^{n} \prod_{r=1}^{m_{ij}} \prod_{s=1}^{m_{ik}} {\rm pr}({\rm Y}_{ijr} = y_{ijr}, {\rm Y}_{iks} = y_{iks}) \times$$
$$\times \prod_{i<k<j}^{n} \prod_{r=1}^{m_{ij}} \prod_{s=1}^{m_{kj}} {\rm pr}({\rm Y}_{ijr} = y_{ijr}, {\rm Y}_{kjs} = y_{kjs}) \prod_{i<j<k}^{n} \prod_{r=1}^{m_{ij}} \prod_{s=1}^{m_{jk}} {\rm pr}({\rm Y}_{ijr} = y_{ijr}, {\rm Y}_{jks} = y_{jks}).$$

Under regularity conditions, that involve essentially the correct specification of the univariate and bivariate marginal distributions, the maximum pairwise likelihood estimator $\hat{\sigma}^2_{\rm pair}$ is consistent and asymptotically normal. Let $\ell_{\rm pair}(\sigma^2) = \log \mathscr{L}_{\rm pair}(\sigma^2)$, then the asymptotic variance of $\hat{\sigma}^2_{\rm pair}$ has the "sandwich" form $\mathbf{H}^{-1}(\sigma^2)\mathbf{J}(\sigma^2)\mathbf{H}^{-1}(\sigma^2)$, where $\mathbf{H}(\sigma^2) = -{\rm E}\left\{\ell''_{\rm pair}(\sigma^2)\right\}$ and $\mathbf{J}(\sigma^2) = {\rm var}\left\{\ell'_{\rm pair}(\sigma^2)\right\}$, see Varin, Reid and Firth (2011).

Natural starting points for the hybrid pairwise likelihood algorithm are the maximum independence likelihood estimates $\hat{\boldsymbol{\beta}}_{\rm ind}$. Then, the algorithm proceeds by cycling between estimation of σ^2 by $\hat{\sigma}^2_{\rm pair}(\hat{\boldsymbol{\beta}}_{\rm dep})$ and estimation of $\boldsymbol{\beta}$ by $\hat{\boldsymbol{\beta}}_{\rm dep}(\hat{\sigma}^2_{\rm pair})$ until convergence is achieved. At convergence, $(\hat{\boldsymbol{\beta}}_{\rm dep}, \hat{\sigma}^2_{\rm pair})$ solve simultaneously the optimal estimating equations for the regression parameters and the pairwise likelihood for the dependence parameter. It is immediate to verify that if the first two moments are correctly specified, than the hybrid pairwise likelihood estimates for the regression coefficients $\hat{\boldsymbol{\beta}}_{\rm dep}$ are consistent and asymptotically normally distributed with the same asymptotic variance as if σ^2 was known, namely

$$\hat{\boldsymbol{\beta}}_{\rm dep} \dot{\sim} {\rm N}\left\{\boldsymbol{\beta}, \mathbf{D}(\boldsymbol{\beta})^{-1}\mathbf{V}_{\rm dep}(\boldsymbol{\beta}, \sigma^2)\mathbf{D}(\boldsymbol{\beta})^{-{\rm T}}\right\}.$$

Standard errors for $\hat{\boldsymbol{\beta}}_{\rm dep}$ are then computed by replacing the unknown $\boldsymbol{\beta}$ and σ^2 with their estimates $\hat{\boldsymbol{\beta}}_{\rm dep}$ and $\hat{\sigma}^2_{\rm pair}$.

3.2.1 Generalized estimating equations

Because of the difficulties in dealing with high-dimensional categorical distributions, various authors have proposed methods that involve only the mean vector and covariance matrix of the data. The most popular is likely the method of generalized estimating equations (GEEs) by Liang and Zeger (1986). At first glance, the optimal estimating equations (10) may appear a particular version of GEEs. However, there is a key difference. In GEEs, the variance matrix of the outcomes is modeled in terms of a working correlation matrix that does not depend on the parameters included in the mean. As stated by Nikoloulopoulos, Joe and Chaganty (2011) the working correlation matrices used in GEEs ignore the fact that correlations for non-normal data are constrained by the univariate margins. As a consequence, the consistency of GEEs may fail. See the introduction of Nikoloulopoulos, Joe and Chaganty (2011) for more discussion and references. In contrast, in the hybrid pairwise likelihood method discussed in the previous section, the variance matrix $\mathbf{V}_{\text{dep}}(\boldsymbol{\beta}, \sigma^2)$ is specified according to a probabilistic model and it depends also on univariate margins (and thus on the regression parameters $\boldsymbol{\beta}$). Hence, the hybrid pairwise likelihood method correctly takes into account the natural restrictions on the correlations in binary variables.

4 Simulations

Simulation studies are carried out in order to investigate the finite sample performance of the proposed methodology. Here, we present the results of just one of the studies, since all the others yield similar conclusions. The simulation study consists of several sets of simulated tournaments. Each tournament comprises 106 contests among 35 animals following the same schedule as the chameleons data analyzed in Section 5. We choose to mimic the chameleons data in way to better assess the importance of accounting for dependence in real data. The contests outcomes were generated according to the marginal dependence Thurstone model specified by equations (4)-(5) with a single continuous animal-specific covariate x_i simulated from a normal distribution with mean 0.5 and variance 0.25 using as regression coefficient the value $\beta = 1$. Various degrees of dependence as expressed by parameter σ^2 are considered. More specifically, $1,000$ simulated tournaments are generated for each of the following ten values of $\sigma^2 \in \{0, 0.05, 0.1, \ldots, 0.45\}$.

In our simulations the hybrid pairwise likelihood algorithm is numerically very stable, hence requiring very few cycles to reach convergence. In the majority of the cases convergence is reached in about 3 cycles. Table 2 shows the results of fitting the simulated tournaments with the independence and dependence models. As expected, the estimate of the regressor parameter is not much different under independence and dependence assumptions. However, simulation results clearly show that the independence model fails to take into account the increased variability as the dependence parameter σ^2 increases. In fact, the underestimation of uncertainty of the independence analysis is reflected in a poor coverage of the confidence intervals for the regression coefficient β. The empirical coverage of the confidence intervals for β based on the independence likelihood falls as σ^2 raises. On the contrary, the dependence model fitted by the hybrid pairwise likelihood method performs well for all the values of σ^2.

Table 2: Simulation study: Average simulated estimates (`est.`), average model-based standard errors (`s.e.`) and empirical coverage of 95% and 99% confidence intervals of the regression parameter estimated under the independence and the dependence assumptions for increasing values of the parameter σ^2. True value for the regression coefficient is 1.

	independence				dependence			
σ^2	est.	s.e.	0.95	0.99	est.	s.e.	0.95	0.99
0	1.033	0.204	0.953	0.996	1.033	0.210	0.956	0.997
0.05	1.032	0.204	0.933	0.988	1.032	0.217	0.944	0.993
0.10	1.024	0.204	0.930	0.984	1.021	0.223	0.952	0.991
0.15	1.036	0.205	0.896	0.976	1.034	0.235	0.942	0.989
0.20	1.035	0.205	0.913	0.981	1.034	0.242	0.958	0.994
0.25	1.042	0.207	0.893	0.965	1.039	0.256	0.962	0.989
0.30	1.059	0.209	0.854	0.953	1.053	0.267	0.940	0.989
0.35	1.063	0.210	0.859	0.955	1.056	0.276	0.950	0.990
0.40	1.055	0.210	0.810	0.939	1.049	0.286	0.946	0.991
0.45	1.080	0.213	0.827	0.928	1.072	0.299	0.949	0.988

5 Application to male Cape Dwarf Chameleons

The analysis of the outcomes of contests between male Cape Dwarf Chameleons is aimed at determining whether some covariates related to the size of the animals or the dimension of the flank patch influence the results of contests. Our intent is to illustrate that not accounting for the dependence among contests may result in imprecise inferential conclusions. For the purpose of illustration, we analyze in detail the independence Thurstone model selected on the basis of the Akaike Information Criterion. This model includes four covariates, namely tail length, snout-vent length, casque height and the proportion of the flank patch. The first three columns of Table 3 display the estimates and standard errors computed under the independence assumption. All four covariates are strongly significant according to the Wald test, as can be seen from the reported absolute z values defined as the absolute value of the ratio between estimates and standard errors. In particular, casque height and tail length are positively associated with victory while snout-vent length and the proportion of the flank patch are negatively associated.

The next step is to check whether the data support the presence of dependence or not. For this purpose we compute a first estimate of the dependence parameter σ^2 by maximizing the pairwise likelihood with $\boldsymbol{\beta}$ set equal to its estimate under the assumption of independence. This provides a consistent estimation of σ^2 under the dependence model because $\hat{\boldsymbol{\beta}}_{\text{ind}}$ is a consistent estimator of $\boldsymbol{\beta}$ also under the dependence model. The resulting estimate is $\hat{\sigma}^2_{\text{pair}}(\hat{\boldsymbol{\beta}}_{\text{ind}}) = 0.282$. The obvious question is whether this estimate is suggestive of the presence of dependence or if its relatively large value – remember that σ^2 is constrained to lie in the interval $(0, 0.5)$ – can instead be due to randomness. In other words, we are interested in validating

Table 3: Application to Male Cape Dwarf Chameleons: Estimates, standard errors and absolute z values for the best model according to AIC. Columns 1-3 report estimates obtained with the Thurstone independence model, while columns 4-6 report estimates obtained with the marginal dependence Thurstone model.

	indipendence			dependence		
	est.	s.e.	z-value	est.	s.e.	z-value
CH	0.289	0.121	2.388	0.311	0.166	1.873
TL	-0.069	0.021	3.286	-0.064	0.028	2.286
SVL	-0.092	0.034	2.706	-0.039	0.031	1.258
FP	0.030	0.013	2.308	0.023	0.018	1.278

the null hypothesis of independence $\mathrm{H}_0 : \sigma^2 = 0$ versus the alternative $\mathrm{H}_1 : \sigma^2 > 0$. Since the null hypothesis corresponds to the limiting case of σ^2 approaching zero, that is the border of the parameter space, then standard asymptotic results do not apply here. This difficulty can be overcome by relying on a parametric bootstrap assessment. We simulate $1,000$ bootstrap samples from the independence Thurstone model and compute the corresponding $1,000$ estimates $\hat{\sigma}^2_{\text{pair}}(\hat{\boldsymbol{\beta}}_{\text{ind}})$. The maximum of these simulated estimates is 0.22, that is none of the simulated estimates is greater than the estimate computed at the observed data. Hence, we conclude that it is extremely unlikely that the independence hypothesis holds.

We then fitted the dependence marginal Thurstone model by the hybrid pairwise likelihood approach. The algorithm converges in 7 cycles. At convergence, the estimated dependence parameter is $\hat{\sigma}^2_{\text{pair}} = 0.368$. Estimates, standard errors and z values for the regressor coefficients are displayed in the last three columns of Table 3. As expected, the standard errors of the model accounting for dependence are larger than those derived under the assumption of independence. Indeed, only the tail length covariate now appears to be significantly associated with the outcomes of the contests. The overall conclusion is that the effect of dependence in paired comparison experiments can be quite substantial, thus leading to considerably different inferential conclusions.

6 Extension to contests with ties

In the chameleons data, biologists terminated the interactions between animals once a clear winner was identified. However, in some instances it is not possible to detect a clear winner of the contest, hence a tie is observed. In this case the contest outcome Y_{ijr} is a categorical variable that assumes $\mathrm{Q} = 3$ ordered categories here arbitrarily coded as

$$\mathrm{Y}_{ijr} = \begin{cases} 0, & \text{if animal } j \text{ beats animal } i, \\ 1, & \text{if a tie is observed,} \\ 2, & \text{if animal } i \text{ beats animal } j. \end{cases}$$

Hence, to analyze data that include ties it is necessary to extend the Thurstone model described in this paper for ordinal observations. More generally, consider an outcome Y_{ijr} that can assume one of Q different values where 0 denotes the worst result for animal i and $\mathrm{Q}-1$ denotes the best possible result for animal i. The Thurstone model for ordinal data is derived by considering the corresponding censored linear regression model

$$\begin{aligned} \mathrm{Y}_{ijr} = q \quad &\leftrightarrow \quad \tau_q < \mathrm{Z}_{ijr} \leq \tau_{q+1}, \\ \mathrm{Z}_{ijr} &= (\boldsymbol{x}_i - \boldsymbol{x}_j)^{\mathrm{T}} \boldsymbol{\beta} + \epsilon_{ijr}, \end{aligned} \tag{11}$$

where parameters τ_q are threshold parameters such that $-\infty \equiv \tau_0 < \tau_1 < \ldots < \tau_{\mathrm{Q}-1} < \tau_{\mathrm{Q}} \equiv \infty$. Model identifiability requires that thresholds satisfy the following condition of symmetry: $\tau_q = -\tau_{\mathrm{Q}-q}$, $q = 0, \ldots, \mathrm{Q}-1$, and $\tau_{\mathrm{Q}/2} = 0$ if Q is even. To illustrate the need of this condition consider the win-tie-loss case ($\mathrm{Q} = 3$) and assume for simplicity that all the contests are observed only once, $m_{ij} = 1$ for any choice of indeces i, j. The model must assure that the probability that animal i beats animal j,

$$\mathrm{pr}\,(Y_{ij} = 2) = \Phi\left\{-\tau_2 + (\boldsymbol{x}_i - \boldsymbol{x}_j)^{\mathrm{T}} \boldsymbol{\beta}\right\},$$

is the same as the probability that animal j is beaten by animal i,

$$\mathrm{pr}\,(Y_{ji} = 0) = \Phi\left\{-\tau_1 - (\boldsymbol{x}_j - \boldsymbol{x}_i)^{\mathrm{T}} \boldsymbol{\beta}\right\},$$

and such a requirement is satisfied only if $\tau_1 = -\tau_2$.

Dependence between comparisons with common animals is introduced by assuming the same structure for the hidden errors as that proposed in formula (5) for binary outcomes. Hence, the correlation matrix of the hidden errors is equal to the one described in formula (6).

In order to employ the optimal estimating equations framework it is necessary to transform each categorical variable Y_{ijr} into a set of $\mathrm{Q}-1$ binary variables $\mathrm{Y}^*_{ijrq} = 1$ if $\mathrm{Y}_{ijr} \leq q$ and $\mathrm{Y}^*_{ijrq} = 0$ otherwise, for $q = 0, \ldots, \mathrm{Q}-2$. Let $\mathbf{Y}^*_{ijr} = (\mathrm{Y}^*_{ijr0}, \ldots, \mathrm{Y}^*_{ijr\,\mathrm{Q}-2})$ be the vector of the binary variables corresponding to Y_{ijr} and $\mathbf{Y}^* = (\mathbf{Y}^*_{ijr}, i < j, r = 1, \ldots, m_{ij})$ be the vector of all the binary variables for each contest. The relative observations are denoted by $\boldsymbol{y}^*$ and the corresponding vector of model based probabilities is $\boldsymbol{\pi}^*(\boldsymbol{\beta})$. Then, equation (10) becomes

$$\mathbf{D}^*(\boldsymbol{\beta})^{\mathrm{T}} \mathbf{V}^*{}_{\mathrm{dep}}(\boldsymbol{\beta}, \sigma^2)^{-1} \left\{\boldsymbol{y}^* - \boldsymbol{\pi}^*(\boldsymbol{\beta})\right\} = \mathbf{0}, \tag{12}$$

where $\mathbf{D}^*(\boldsymbol{\beta})$ is the Jacobian of $\boldsymbol{\pi}^*$ and $\mathbf{V}^*{}_{\mathrm{dep}}(\boldsymbol{\beta}, \sigma^2)$ is the covariance matrix of the $\mathbf{Y}^*$. The elements of the covariance matrix are

$$\mathrm{cov}(\mathrm{Y}^*_{ijrq}, \mathrm{Y}^*_{kls\tilde{q}}) = \mathrm{pr}(\mathrm{Y}^*_{ijrq} = 1, \mathrm{Y}^*_{kls\tilde{q}} = 1) - \mathrm{pr}(\mathrm{Y}^*_{ijrq} = 1)\mathrm{pr}(\mathrm{Y}^*_{kls\tilde{q}} = 1). \tag{13}$$

The univariate probabilities can be easily computed as $\mathrm{pr}(\mathrm{Y}^*_{ijrq} = 1) = \mathrm{pr}(\mathrm{Y}_{ijr} \leq q) = \Phi\{\tau_{q+1} - (\boldsymbol{x}_i - \boldsymbol{x}_j)^{\mathrm{T}} \boldsymbol{\beta}\} - \Phi\{\tau_q - (\boldsymbol{x}_i - \boldsymbol{x}_j)^{\mathrm{T}} \boldsymbol{\beta}\}$. Some care is needed in the computation of the bivariate probabilities in equation (13). In fact, these bivariate probabilities can involve either binary variables related to the same contest or variables related to different contests with one or two animals in common or variables related to different contests with no common animals. In the latter case the covariance (13) is zero. When variables refer to the same contest, then $\text{pr}(\text{Y}^*_{ijrq} = 1, \text{Y}^*_{ijr\tilde{q}} = 1) = \text{pr}(\text{Y}^*_{ijrt} = 1)$, where $t = \min\{q, \tilde{q}\}$. In the other cases, when $r \neq s$, then $\text{pr}(\text{Y}^*_{ijrq} = 1, \text{Y}^*_{kls\tilde{q}} = 1) = \Phi_2\{\tau_{q+1} - (\boldsymbol{x}_i - \boldsymbol{x}_j)^T\boldsymbol{\beta}, \tau_{\tilde{q}+1} - (\boldsymbol{x}_k - \boldsymbol{x}_l)^T\boldsymbol{\beta}; \rho\}$, where the correlation ρ is as described in (7).

7 Conclusions

We have presented a method to handle dependence in paired comparison experiments with specific attention to tournament data used in animal behavior experiments. The proposed methodology can be of interest for other areas where paired comparison data are collected. These include analysis of sport data, psychometric experiments about perception and genetics. The complicated cross-correlation structure in tournament-like data causes major difficulties in the computation of the full likelihood function. As a viable alternative, we have considered the combination of optimal estimating equations for the regressor parameters and pairwise likelihood for the dependence parameter following the suggestion by Kuk (2007). Simulation studies suggest that the hybrid pairwise likelihood method works well and a real application shows that the practical effect of accounting for dependence can have a substantial impact on inferential conclusions.

The issue of dependence has been considered also in other contexts in which paired comparison data arise. For example, in the analysis of preference data it is natural to assume that preferences expressed by the same person in different comparisons involving common objects are dependent (Thurstone, 1927; Dittrich *et al.*, 2002; Böckenholt and Tsai, 2007). When normality is assumed and the objects compared are more than four, different methods have been applied to overcome the computational difficulties of maximum likelihood estimation (Mosteller, 1951; Böckenholt and Tsai, 2001; Train, 2009). A viable alternative proposed in the psychometric literature is to rely on limited information estimation, that is a class of estimating functions constructed from low dimensional marginals. A popular limited information method used for preference data is due to Maydeu-Olivares (2001). This consists in minimizing some Mahalanobis distance between the model-based univariate and bivariate margins and the corresponding empirical proportions. Hence, the method described in Maydeu-Olivares (2001) requires independent replicates of the data which are unavailable in the context of animal behavior experiments considered in this manuscript. The hybrid pairwise likelihood method used in this paper can be viewed as a further type of limited information estimation since it requires only the first two moments. Application of the hybrid pairwise likelihood method to data that typically arise in the psychometric literature is a topic for future research.

Acknowledgements

The Authors acknowledge support from the PRIN 2008 grant of the Italian Minister of Instruction, University and Research (MIUR).

References

Böckenholt, U. (2006). Thurstonian-based analyses: past, present, and future utilities. *Psychometrika* **71**, 615–629.

Bočkenholt, U. and Tsai, R.-C. (2001). Individual differences in paired comparison data. *British Journal of Mathematical and Statistical Psychology* **54**, 265-277.

Bočkenholt, U. and Tsai, R.-C. (2007). Random-effects models for preference data. In C.R. Rao and S. Sinharay (Eds.), *Handbook of Statistics*, vol. 26, pp. 447-468. Elsevier Science, Amsterdam.

Bradley, R. A. and Terry, M. E. (1952). Rank analysis of incomplete block designs I: the method of paired comparisons. *Biometrika* **39**, 324–345.

Brown, W. D., Smith, A. T., Moskalik, B. and Gabriel, J. (2008). Aggressive contests in house crickets: size, motivation and the information content of aggressive songs. *Animal Behviour* **72**, 225-233.

Cattelan, M. (2011). Models for paired comparison data: a review with emphasis on dependent data. Working paper series n.8, Department of Statistical Sciences, University of Padua.

Chib, S. and Greenberg, E. (1998). Analysis of multivariate probit models. *Biometrika* **85**, 347-361.

Cox, D. and Reid N. (2004). A note on pseudolikelihood constructed from marginal densities. *Biometrika* **92**, 729–737.

Dittrich, R., Hatzinger, R. and Katzenbeisser, W. (2002). Modelling dependencies in paired comparison data. A log-linear approach. *Computational Statistics and Data Analysis* **40**, 39-57.

Genz, A. and Bretz, F. (2002). Methods for the computation of multivariate t-probabilities. *Journal of Computational and Graphical Statistics* **11**, 950-971.

Hajivassiliou, V. A., McFadden, D. and Ruud, P. (1996). Simulation of multivariate normal rectangle probabilities and their derivatives: theoretical and computational results. *Journal of Econometrics* **72** 85-134.

Jeliazkov, I. and Lee, E. H. (2010). MCMC perspectives on simulated likelihood estimation. *Advances in Econometrics* volume **26**: Maximum Simulated Likelihood Methods and Applications (edited by W.H. Greene and C. Hill). Emerald Group Publishing Limited, 3–40.

Joe, H. (1995). Approximations to multivariate normal rectangle probabilities based on conditional expectations. *Journal of the American Statistical Association* **90**, 957-964.

Kemp, D. J., Wiklund, C. and Van Dyck, H. (2006). Contest behavior in the speckled wood butterfly (*Pararge aegeria*): seasonal phenotypic plasticity and the functional significance of flight performance. *Behavioural Ecology and Sociobiology* **59**, 403-411.

Kuk, A. Y. C. (2007). A hybrid pairwise likelihood method. *Biometrika* **94**, 939–952.

Liang, K. Y. and Zeger, S. L. (1986). Longitudinal data analysis using generalized linear models. *Biometrika* **73**, 13–22.

Lindsay, B. G. (1988). Composite likelihood methods. *Contemporary Mathematics* **80**, 221-239.

Maydeu-Olivares, A. (2001). Limited information estimation and testing of Thurstonian models for paired comparison data under multiple judgement sampling. *Psychometrika* **66**, 209-228.

Miwa, A., Hayter J. and Kuriki, S. (2003). The evaluation of general non-centred orthant probabilities. *Journal of the Royal Statistical Society Series B* **65**, 223-234.

Molenberghs, G. and Verbeke, G. (2005). *Models for Discrete Longitudinal Data*. Springer.

Mosteller, F. (1951). Remarks on the method of paired comparisons. I. The least squares solution assuming equal standard deviations and equal correlations. *Psychometrika* **16**, 3-9.

Nikoloulopoulos, A. K., Joe, H. and Chaganty, N. R. (2011). Weighted scores method for regression models with dependent data. *Biostatistics* **12**, 653-665.

R Development Core Team (2011). R: A language and environment for statistical computing. *R Foundation for Statistical Computing*, Vienna, Austria. `http://www.R-project.org`.

Stuart-Fox, D. M., Firth, D., Moussalli, A. and Whiting, M. J. (2006). Multiple signals in chameleon contests: designing and analysing animal contests as a tournament. *Animal Behaviour* **71**, 1263-1271.

Thurstone, L. L. (1927). A law of comparative judgment. *Psychological Review* **79**, 281–299.

Train, K. E. (2009). *Discrete Choice Methods with Simulation*. Cambridge University Press, Cambridge.

Turner, H. and Firth, D. (2011). Bradley-Terry models in R: The BradleyTerry2 package. `http://CRAN.R-project.org/package=BradleyTerry`.

Varin, C., Reid, N., and Firth, D. (2011). An overview of composite likelihood methods. *Statistica Sinica* **21**, 5–42.

Acknowledgements

The Authors acknowledge support from the PRIN 2008 grant of the Italian Minister of Instruction, University and Research (MIUR).

Working Paper Series
Department of Statistical Sciences, University of Padua

You may order paper copies of the working papers by emailing wp@stat.unipd.it
Most of the working papers can also be found at the following url: http://wp.stat.unipd.it

UNIVERSITA DEGLI STUDI DI PADOVA
DIPARTIMENTO DI SCIENZE STATISTICHE

Department of Statistical Sciences
University of Padua
Italy