

Estimating dyad association probability under imperfect and heterogeneous detection: a Bayesian approach

M. J. Klaich^{a,b}, P.G. Kinas^c, S.N. Pedraza^{a,b}, M.A. Coscarella^{a,b} and E.A. Crespo^{a,b}

^a Centro Nacional Patagónico, Boulevard Brown 2915, Puerto Madryn, Chubut, Argentina

^b Universidad Nacional de la Patagonia San Juan Bosco, Boulevard Brown 3150, Sede Puerto Madryn, Chubut, Argentina

^c Universidade Federal do Rio Grande, Avenida Itália, km 8, Rio Grande, RS, Brazil

Abstract

In social studies, individual association indices estimate the proportion of time two individuals (i.e. a dyad) spend together. However, traditional association indices rely on the assumptions that individual detection probabilities (p) are either one (at least approximately) or independent of the association states. Based on marked individuals, we develop a model to estimate the probability a dyad is in associated state (ψ) accounting for detectability less than one and varying according to association states. Our model allows for both individual and dyad missing observation and can easily be extended to incorporate covariate information for modeling detectability and dyad association probability. Parameter estimates are obtained as posterior means via Monte Carlo Markov Chain. A simulation study showed that our model-based approach yield unbiased estimates, even for low and heterogeneous detection probabilities, while, in contrast, standard indices showed moderate to strong biases.

Key words: Social analysis, Bayesian inference, detection probabilities, individual association index, dyad association probability.

Introduction

The most common approach to study social relationships among individually identified animals is to use association indices (AI) (Cairns & Schwager 1987; Ginsberg & Young 1992). In terms of dyads, all AI can be interpreted as estimators of the probability that a dyad is in associated state at a given time t (Cairns & Schwager 1987; Whitehead 2008b).

In practical terms, missing values are the immediate result when individual detection probabilities are very low. Concerning AI, when detection is perfect, the Simple Index (SI) is the unbiased estimator of both the probability that two individuals (i.e. a dyad) are associated and the proportion of time they spend together (Cairns & Schwager 1987; Ginsberg & Young 1992). When individual detectability is less than one, and considering the potential sources of bias on SI, two types of missing data can occur, when only one individual from the dyad is observed and when the complete dyad is not observed. In both cases we cannot say if the dyad is in associated state or not. Under an inferential framework, Cairns & Schwager (1987) developed two different models for estimating via maximum likelihood method the AI as the probability that a dyad is in association state at given time t (ψ). Despite a clear improvement over the traditional AI, the maximum likelihood estimator is rarely used; the simplicity in calculating SI and the Half Weight Index (HWI) may explain its enduring preference over the likelihood approximation in most social studies. Lusseau et al. (2008) pointed out the importance of sampling effects on AI and included that form of uncertainty when weighted network measures were estimated. However, the effect of imperfect

or/and heterogeneous individual detection on quantitative social analyses remain a pending issue.

Here, we develop a model to estimate the probability that a dyad is in associated state while accounting for individual detection probability less than one and heterogeneity among association states in the detection probability. Our Bayesian approach provides a flexible framework allowing for missing data, time-specific parameters and the inclusion of covariate information at individual, dyad and group levels.

First, we show how to build the association history for a dyad. This is analogous to an individual encounter history used in mark-recapture models (Williams et al. 2002) and detection history for sites in occupancy models (MacKenzie et al. 2002). Second, we introduce the model to estimate the probability that a dyad is in associated state and an estimation procedure is proposed via the Monte Carlo Markov Chain under a Bayesian framework. Finally, we discuss possible extensions of the model and its applicability in social and ecological studies.

Methods

Suppose a population in which a proportion of animals are individually identified by either artificial or natural marks. Then, a mark-recapture experiment is conducted with 4 sampling occasions in which an individual named “*a*” is detected on occasions 1, 3 and 4 and an individual named “*b*” is detected on occasions 1 and 4. This can be expressed by two encounter histories (*eh*) as:

$$\begin{aligned} eh_{(a)}: & 1\ 0\ 1\ 1 \\ eh_{(b)}: & 1\ 0\ 0\ 1 \end{aligned}$$

Both individuals are simultaneously detected on occasions 1 and 4 while at occasion 3 only individual *a* was detected. Assuming that the co-occurrence of individuals *a* and *b* are socially driven (Whitehead & Dufault 1999), we define a state A for which the individuals are associated at a given sampling occasion. However, it is possible that even being identified at the same occasion two individuals are not associated, in which case we define a second state B for which this dyad is not associated. A third situation, named I can be also defined for the occasions when only one individual of the dyad is detected. Letter I is an indicator that a dyad was partially observed and that the actual association state remains unknown. Definitions of both ‘associated’ (i.e. A) and ‘not associated’ (i.e. B) states must be established by the researcher prior the analysis and for the particular study (Michener 1980). Then, assuming the individuals *a* and *b* were observed in an associated state at first occasion and they were not associated at occasion 4, the history for the dyad can be written as:

$$A\ 0\ I\ B$$

The likelihood formulation of our model is based on the fundamentals of mark-recapture models (Williams et al. 2002) and occupancy models (MacKenzie et al. 2005). We focused on populations which are closed regarding birth, immigration, mortality and emigrations. We also consider the dyad as the equivalent of an individual in mark-recapture analyses or as a site in occupancy analyses.

Assuming time-dependent parameters, the contribution to the whole data likelihood of the dyad with association history *A0IB* is:

$$\psi_1 (p_1^A)^2 \left[\psi_2 (1 - p_2^A)^2 + (1 - \psi_2) (1 - p_2^B)^2 \right] \left\{ 2 \left[\psi_3 p_3^A (1 - p_3^A) + (1 - \psi_3) p_3^B (1 - p_3^B) \right] \right\} (1 - \psi_4) (p_4^B)^2 \quad (1)$$

The product of terms with subscript 1 denote the probability that a dyad is in associated state and both individuals are detected given they are in associated state at sampling occasion 1. The expression between brackets with subscript 2, indicate the

two possible non-detections of both individuals in the dyad. The first term is the probability that a dyad is in associated state and the non detection of both individuals in the dyad given they are in associated state at sampling occasion 2, whereas the second term is the probability a dyad is in a non associated state and the non detection of both individuals in the dyad given they are in non associated state at sampling occasion 2. The expression between braces represents all possible events when only one individual in the dyad was detected. Finally, the last product of terms with subscript 4 represents the probability that the dyad is in not associated state and both individuals in the dyad are detected given they are in not associated state at sampling occasion 4. The above likelihood formulation assumes homogeneous detection within each association state and instantaneous detection of the individuals in the dyad.

Assuming dyads are detected independently and parameters being time specific, the likelihood of the data is:

$$L(\boldsymbol{\psi}, \mathbf{p}^A, \mathbf{p}^B) = \prod_{j=1}^J \prod_{i=1}^t \left[\psi_i (p_i^A)^2 \right]^{In(A)_{j,i}} \left[(1-\psi_i)(p_i^B)^2 \right]^{In(B)_{j,i}} \left\{ 2 \left[\psi_i p_i^A (1-p_i^A) + (1-\psi_i) p_i^B (1-p_i^B) \right] \right\}^{In(I)_{j,i}} \dots \\ \times \left[\psi_i (1-p_i^A)^2 + (1-\psi_i)(1-p_i^B)^2 \right]^{In(O)_{j,i}} \quad (2)$$

where t is the number of sampling occasions, J is the total number of dyads observed across the t sampling occasions, $In(A)_{j,i}$, $In(B)_{j,i}$, $In(I)_{j,i}$ and $In(O)_{j,i}$ are the usual indicator functions for the j th dyad and occasion i for events A , B , I and O respectively. For instance, if the j th dyad is observed in association state A at occasion i the indicator function $In(A)_{j,i}$ will take value 1 and $In(A)_{j,i} = 0$ otherwise.

When parameters are assumed to be constant across time, the likelihood can be simplified into:

$$L(\psi, p^A, p^B) = \prod_{j=1}^J \left[\psi (p^A)^2 \right]^{x_{A,j}} \left[(1-\psi)(p^B)^2 \right]^{x_{B,j}} \left\{ 2 \left[\psi p^A (1-p^A) + (1-\psi) p^B (1-p^B) \right] \right\}^{(y_{a,j}+y_{b,j})} \dots \\ \times \left[\psi (1-p^A)^2 + (1-\psi)(1-p^B)^2 \right]^{z_j} \quad (3)$$

where $x_{A,j}$ is the number of times the j th dyad was observed in state A ; $x_{B,j}$ is number of times the j th dyad was observed in state B ; $y_{a,j}$ is the number of times only individual a was observed from the j th dyad; $y_{b,j}$ is the number of times only individual b was observed from the j th dyad; and z_j is the number of times neither a nor b were observed from the j th dyad.

Within a Bayesian framework, and assuming uniform prior distributions on all three parameters, the posterior distribution is proportional to the likelihood function.

Simulation study

A simulation study was undertaken for evaluating the performance of the proposed method for estimating ψ and p . The data were simulated without time effects for association and detection parameters. The effects of five factors were investigated: $t = 5$ and 10 ; the population size under study: $N = 20$ and 40 ; $\psi = 0.5, 0.7$ or 0.9 ; $p^A = 0.1, 0.5$ or 0.7 and $p^B = 0.1, 0.5$ or 0.7 . For each of the 108 scenarios, 200 data sets were simulated. The data simulation process was divided in two steps: 1) assign the association state to each dyad across the sampling occasions and 2) apply the effect of individual detection given the dyad association state. For each sampling occasion, an independent Bernoulli trial with parameter ψ was generated for each dyad in the population of all possible dyads when the size population was N . If the generated number was 0, then the dyad was considered as not associated; if the generated number

was 1 the dyad was considered as associated. Individual detection effects were assigned next by generating the number of detected individuals in each dyad from a binomial distribution with parameters 2 and p^A for dyads in associated state or 2 and p^B for dyads in not associated state. Finally, a 0 indicates that both individuals from the dyad were not detected, 1 indicates that only one individual was detected and 2 indicates that both individuals in the dyad were detected.

Estimating posterior distributions via Markov Chain Monte Carlo

For each parameter ψ , p^A and p^B , a uniform distribution on interval [0,1] was used as a prior distribution. As estimators we used the posterior mean from each simulated data set across the 108 scenarios. The posterior distributions were estimated via MCMC using the free software OpenBUGS via the R2WinBUGS and BRUGS packages in the free software R (R Development Core Team 2009). The bugs function was set using one Markov chain; the total number of iterations was 18000, a burn-in of 15000 iterations and thinning of 1. The 200 posterior medians of ψ , p^A and p^B for each scenario were summarized and compared with their corresponding traditional estimators of ψ means and variances of SI and HWI were estimated across all 200 simulated data set for each of the 108 scenarios.

Results

In general, the proposed method provided reasonable estimates of the probability that a dyad is in associated state. When individual state specific detection was either homogeneous ($p^A = p^B$) or heterogeneous ($p^A < p^B$ or $p^A > p^B$) the method provided reasonable unbiased estimates of ψ in most scenarios considered. Biased estimates of ψ were observed for combinations ($\psi = 0.5$, $p^A = 0.5$, $p^B = 0.7$); ($\psi = 0.5$, $p^A = 0.7$, $p^B = 0.5$) and ($\psi = 0.7$, $p^A = 0.7$, $p^B = 0.5$). Concerning the performance of conventional association indices, SI and HWI, as estimators of the probability a dyad is associated, they were shown to be biased with imperfect individual detection (p^A and $p^B < 1$), for either homogeneous ($p^A = p^B$) or heterogeneous ($p^A \neq p^B$) state specific individual detections.

Discussion

When an identification process is involved in social studies, it is difficult to insure a perfect individual detection. Moreover, for species like tree-dwelling monkeys (Jack & Fedigan 2004) or cetaceans (Williams & Thomas 2009) the individual detection is sometimes a difficult and complex task. Our simulations showed strong bias in conventional estimates of the intensity of association in scenarios with either imperfect or heterogeneous detection. Only when individual detections were as high as 0.9, the traditional association indices were almost unbiased estimates of ψ . It is clear that a negative correlation exists between the difficulty in calculation and implementation of a model and its bias for estimating the probability that a dyad is in associated state.

The model presented here assumes both geographic and demographic closure. However, the association between two individuals could be affected by demographic processes like survival or migration along years or seasons (Whitehead 2008a). A more elaborate model for estimating association probability could be developed incorporating individual survival and individual temporary migration (Kendall et al. 1997; Schaub et al. 2004). In addition, our model could be extended to consider the event a dyad is in associated state as a first-order Markov chain (see Schwarz et al. (1993) for an

application in the mark-recapture modeling framework). This would allow to estimate the transition probability that a dyad is in associated state at occasion i given it was in association state at occasion $i-1$ (ψ_i^{AA}) and the transition probability that a dyad is in associated state at occasion i given it was in not association state at occasion $i-1$ (ψ_i^{BA}). A model with $\psi_i^{AA} = \psi_i^{BA}$ would assume that transitions between association states are random, while a model with $\psi_i^{AA} \neq \psi_i^{BA}$ would assume that transitions between association states are not random (i.e. Markovian process). Hypotheses regarding preferred or avoided companions could also be assessed, a dyad tending to remain more associated than no associated if $\psi_i^{AA} > \psi_i^{BB}$ and vice versa if $\psi_i^{BB} > \psi_i^{AA}$. Evidence for either can be calculated from the joint posterior distribution.

Conclusion

When individual detection is not accounted for when estimating the probability that a dyad is in associated state, biased estimates are obtained if conventional association indices are used. However, we showed that a model-based approach along with a Bayesian inferential framework which models the imperfect and heterogeneous individual detection explicitly, yield less biased estimates for the probability that a dyad is in associated state. The application of the method on a data set of marked individuals from either terrestrial or aquatic social species is possible because of the general formulation of the proposed model. The extensions of the proposed model allows for the inclusion of demographic parameters such as survival, movement and migration rates, which represents a new and versatile tool for further quantitative social studies.

References

- Cairns, S. J. & Schwager, S. J. 1987. A comparison of association indices. *Animal Behaviour*, 35, 1454–1469.
- Ginsberg, J. R. & Young, T. P. 1992. Measuring association between individuals or groups in behavioural studies. *Animal Behaviour*, 44, 377–379.
- Jack, K. M. & Fedigan, L. 2004. Male dispersal patterns in white-faced capuchins, *Cebus capucinus* Part 1: patterns and causes of natal emigration. *Animal Behaviour*, 67, 761–769
- Kendall, W. L., Nichols, J. D., & Hines, J. E. 1997. Estimating temporary emigration using capture–recapture data with Pollock’s robust design. *Ecology*, 78, 563–578.
- Lusseau, D., Whitehead H. & Gero, S. 2008. Incorporating uncertainty into the study of animal social networks. *Animal Behaviour*, 75, 1809–1815.
- MacKenzie, D. I., Nichols, J. D., Lachman, G. B., Droege, S., Royle, J.A. & Langtimm, C. A. 2002. Estimating site occupancy rates when detection probabilities are less than one. *Ecology*, 83, 2248–2255.
- MacKenzie, D. I., Nichols, J. D., Royle, J. A., Pollock, K. H., Bailey, L. L. & Hines, J. E. 2005. *Occupancy Estimation and Modeling: Inferring Patterns and Dynamics of Species Occurrence*. San Diego, CA: Elsevier.
- Michener, G. R. 1980. The measurement and interpretation of interaction rates: an example with Richardson's ground squirrels. *Biology of Behaviour*, 5, 371–384.
- R Development Core Team (2009). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL <http://www.R-project.org>.

- Schaub, M., Gimenez, O., Schmidt, B. R. & Pradel, R. 2004. Estimating survival and temporary emigration in the multistate capture-recapture framework. *Ecology*, 85, 2107-2113.
- Schwarz, C. J., Schweigert, J. F. & Arnason, A. N. 1993. Estimating migration rates using tagrecovery data. *Biometrics*, 49, 177-193
- Whitehead, H. & Dufault, S. 1999. Techniques for analyzing vertebrate social structure using identified individuals: review and recommendations. *Advances in the Study of Behavior*, 28, 33-74.
- Whitehead, H. 2008a. *Analyzing Animal Societies: Quantitative Methods for Vertebrate Social Analysis*. Chicago: University of Chicago Press.
- Whitehead, H. 2008b. Precision and power in the analysis of social structure using association. *Animal Behaviour*, 75, 1093-1099
- Williams, B. K., Nichols, J. D. & Conroy, M. J. 2002. *Analysis and management of animal populations*. San Diego, California, USA: Academic Press.
- Williams, R. & Thomas, T. 2009. Cost-effective abundance estimation of rare animals: Testing performance of small-boat surveys for killer whales in British Columbia. *Biological Conservation*, 142, 1542-1547.