ANIMAL BEHAVIOUR, 2008, **75**, 1093–1099
doi:10.1016/j.anbehav.2007.08.022

Precision and power in the analysis of social structure using associations

HAL WHITEHEAD
Department of Biology, Dalhousie University, Halifax

*(Received 5 December 2006; initial acceptance 15 January 2007;
final acceptance 14 August 2007; published online 28 January 2008; MS. number: A10639R2)*

I develop guidelines for assessing the precision and power of statistical techniques that are frequently used to study nonhuman social systems using observed dyadic associations. Association indexes estimate the proportion of time that two individuals are associated. Binomial approximation and nonparametric bootstrap methods produce similar estimates of the precision of association indexes. For a mid-range (0.4–0.9) association index to have a standard error of less than 0.1 requires about 15 observations of the pair associated, and for it to be less than 0.05, this rises to 50 observations. The coefficient of variation among dyads of the proportion of time that pairs of individuals are actually associated describes social differentiation (S), and this may be estimated from association data using maximum likelihood. With a poorly differentiated population ($S \sim 0.2$), a data set needs about five observed associations per dyad to achieve a correlation between true and estimated association indexes of $r = \sim 0.4$. It requires about 10 times as much data to achieve a representation with $r = \sim 0.8$. Permutation tests usually reject the null hypothesis that individuals have no preferred associates when $S^2 \times H > 5$, where H is the mean number of observed associations per individual. Thus most situations require substantial numbers of observations of associations to give useful portrayals of social systems, and sparse association data inform only when social differentiation is high.

Keywords: association; precision; social structure; statistical power

To describe and analyse the social systems of nonhumans, ethologists have assembled a body of quantitative techniques (Whitehead & Dufault 1999; Whitehead 2008). Some were specifically developed to meet the challenges of social analysis of nonhumans, others were imported from other areas of science. In a few areas, such as the analysis of dominance hierarchies, scientists have analysed the performance of measures and tests and sequentially developed improvements (e.g. de Vries 1995; Poisbleau et al. 2006; de Vries et al. 2006). However, they have rarely addressed the precision of statistics used to measure social structure, as well as the power of tests against null hypotheses. This is particularly the case for quantitative methods based upon observations of dyadic associations (but see Whitehead 2007, for methods of estimating the precision of lagged association rates), although models of animal social structure are often built upon observations of dyadic associations (Whitehead & Dufault 1999). It is not clear what can really be concluded from the results of such studies, or how much data are needed to reveal the attributes of a social system. Guidelines would be very useful when planning studies, when assessing the potential of an existing data set to reveal social structure, and in the evaluation of analyses. In this paper, I develop guidelines for assessing the precision and power of some techniques most often used in the study of nonhuman social systems using observed dyadic associations.

The dyadic association index is an estimate of the proportion of time that a pair of individuals is associated. Association indexes can be calculated using any of several formulae (Cairns & Schwager 1987), and association can be defined in a wide range of ways. However, from the perspective of Hinde's (1976) conceptual framework of nonhuman social structure, to use association indexes as the basis of a description or model of social structure, association should be closely related to the probability of behavioural interaction or communication between the two

Correspondence: H. Whitehead, Department of Biology, Dalhousie University, Halifax, Nova Scotia B4H 4J1, Canada (email: hal.whitehead@dal.ca).

0003–3472/08/$32.00/0

members of the dyad (Whitehead & Dufault 1999). If this condition holds, then an association index tells us something about the relationship between the pair of animals: it is a 'relationship measure' (Whitehead 1997). The precision of association indexes is the first issue addressed in this paper.

With association indexes calculated between all (or at least many) dyads in a study population, the analyst can abstract them (using Hinde's 1976 terminology) into models or displays of social structure using techniques such as cluster analysis and multidimensional scaling (Morgan et al. 1976). However, even when associations accurately reflect interaction rates and relationships, such displays, which are built upon many imperfect association indexes, may not well portray real social structures. The correlation between the patterns in the true association indexes (the proportion of time that dyads actually do spend together) and the estimated association indexes indicates the match between a real social structure and its model. This is the second topic of this paper.

Social analysis is also amenable to a hypothesis-testing perspective. A frequently useful null hypothesis is that individuals have no preferences for social partners, with the alternative that there are preferred and/or avoided associations between some pairs of individuals. Bejder et al. (1998) introduced a permutation test for the case in which dyadic association is defined based upon membership in a common group. The test controls for the structure of the data by holding constant both the number of individuals in each group as well as the number of groups in which each individual was observed. My colleagues and I have developed extensions to the Bejder et al. (1998) test, for instance, allowing definitions of association not based upon group membership (such as nearest neighbours) and controlling for demographic events such as birth, death and movements of animals into and out of the study area (Whitehead 1999; Whitehead et al. 2005). However, the power of such tests is unknown. This is the final major goal of this paper.

The second and third objectives required a measure of the variation in relationships among members of a study population. The coefficient of variation (CV) in the true association indexes, called the 'social differentiation' by Whitehead (2008), is suitable for this purpose and can be estimated with little bias using sparse or less than perfect data by the method of maximum likelihood (see Appendix).

The Precision of Association Indexes

Most dyadic association indexes are estimates of the proportion of time that the two individuals are associated (i.e. the true association index), and so range between 0.0 and 1.0 (an exception is Cole's (1949) index, which can be negative). To calculate association indexes, an analyst divides the study into 'sampling periods', which can be different time periods, surveys, or even encountered groups. The analyst then enumerates the number of periods in which the two animals, i and j, are associated (x), the number of periods in which both animals are observed but they are not associated (y_{ij}), and the numbers of periods in which only one animal is observed (y_i and y_j, respectively). Most association indexes (including the 'simple ratio', 'half-weight', 'twice-weight' and 'square-root') are of the form x/d where d is a function of x, y_{ij}, y_i and y_j (Cairns & Schwager 1987). For instance, the simple ratio index, championed by Ginsberg & Young (1992), is $\hat{\alpha} = x/(x + y_{ij} + y_i + y_j)$, and the half-weight is $\hat{\alpha} = x/(x + y_{ij} + (y_i + y_j)/2)$. The different association indexes try to compensate for different forms of sampling bias (such as a lower probability of identifying pairs of animals when associated than when alone) by using different forms of d, the denominator of the association index (see Cairns & Schwager 1987; Ginsberg & Young 1992).

Suppose α is the true association index, the proportion of time that the pair are actually in association. Then, if the sampling periods are assumed independent, x should be binomially distributed with parameters α and d. It follows from binomial theory that the standard deviation of x is:

$$\mathrm{SD}(x) = \sqrt{d \times \alpha \times (1 - \alpha)}$$

and the approximate standard error of the association index, $\hat{\alpha}$, is:

$$\mathrm{SE}(\hat{\alpha}) = \mathrm{SD}(x/d) = \sqrt{d \times \alpha \times (1 - \alpha)}\Big/d$$

As $\alpha \approx \hat{\alpha} = x/d$, an estimate of the standard error of the association index is:

$$\mathrm{SE}(\hat{\alpha}) = \sqrt{\hat{\alpha} \times (1 - \hat{\alpha})/d} = \hat{\alpha} \times \sqrt{(1 - \hat{\alpha})/x} \qquad (1)$$

Thus, equation (1) estimates the standard error of a calculated association index, $\hat{\alpha}$, from $\hat{\alpha}$ and d, the number of observations, or x, the number of observed associations. Figure 1 shows these relationships.

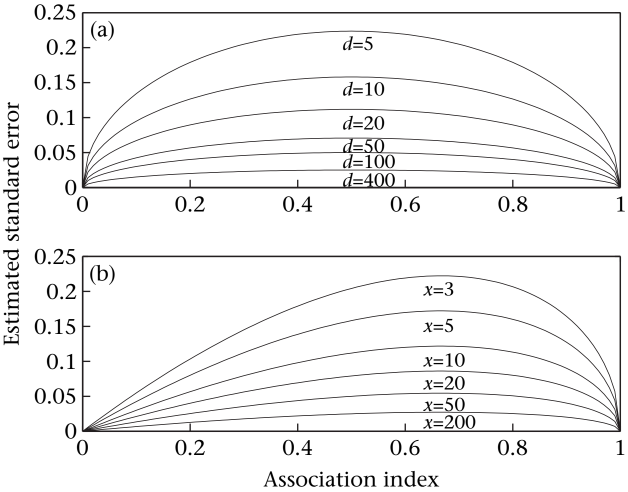

Figure 1. Expected standard errors of an association index (vertical axis) as a function of its value (horizontal axis) and (a) the number of observations (d) or (b) the number of observations of the pair associated (x), from equation (1).

To have a standard error in a mid-range (0.4–0.9) estimated association index of less than 0.1, there must be about 15 observations of the pair associated, and for it to be less than 0.05, this must be nearly 50, a considerable number of observations. If two individuals are observed associated five times with a sample size, *d*, of 15, the estimated association index is $\hat{\alpha} = 0.33$, but its estimated 95% confidence interval is 0.09–0.57. These animals could be spending 9% or 57% of their time together, quite different relationships.

An alternative method of estimating the precision of an association index $\hat{\alpha}$ is the nonparametric bootstrap (e.g. Efron & Gong 1983), in which bootstrap replicate samples are produced by sampling the sampling periods with replacement, and the standard deviation of the $\hat{\alpha}$s calculated from the bootstrap replicates is an estimate of the standard error of $\hat{\alpha}$. The distribution of estimates from bootstrap replicates can give confidence intervals (Efron & Gong 1983). In practice, estimates of standard errors of association indexes calculated using the nonparametric bootstrap and the binomial approximation (equation (1)) are usually in close agreement, as in Table 1, which shows association indexes among 10 bottlenose whales, *Hyperoodon ampullatus*, and their standard errors. The bootstrap estimates of standard error for this data set are a mean of 0.00025 higher than those from equation (1), and the mean absolute difference between the two is 0.00052.

Using the denominator of the association index, *d*, as the sample size for the binomial distribution is correct for the simple ratio index when its assumptions hold (see Ginsberg & Young 1992), as *d* is the integer number of sampling periods over which the individuals might have been associated. However, with other association indexes, *d* may be a noninteger, and it is only an approximation to the sample size. That it is usually a reasonable approximation is indicated by the very close correspondence for the bottlenose whale data between the standard errors calculated using equation (1) and by bootstrap methods when using the half-weight index (mean difference between standard errors from equation (1) and bootstrap = 0.00164; mean absolute difference = 0.00181), twice-weight index (mean difference = −0.00055; mean absolute difference = 0.00073) and square-root index (mean difference = 0.00193; mean absolute difference = 0.00213).

A drawback of both the binomial or bootstrap methods of estimating the precision of association indexes is that if the estimated association index is either $\hat{\alpha} = 0.0$ (i.e. $x = 0$; never seen associated), or $\hat{\alpha} = 1.0$ (i.e. $x = d$; always seen associated), then the estimated standard error is exactly 0.0 (see Fig. 1, Table 1), when, in fact, in these circumstances there may be considerable uncertainty about our estimate of either zero or complete association. For instance, if the true association index is $\alpha = 0.1$, the animals are associated 10% of the time and the sample size is $d = 10$, then there is a 35% probability that $x = 0$, resulting in $\hat{\alpha} = 0$ and $\mathrm{SE}(\hat{\alpha}) = 0$, and the appearance of certainty that they never associate. Similarly, if the true

Table 1. Simple ratio association indexes among 10 northern bottlenose whales, *Hyperoodon ampullatus*, and, beneath each, the estimated standard errors of the estimates from the binomial approximation (equation 1), in italics, and from 1000 bootstrap replicates, in bold*

Whale ID no.	1	3	37	45	54	102	251	409	531	824
1										
3	0.34 *0.08* **0.08**									
37	0.12 *0.05* **0.05**	0.07 *0.04* **0.04**								
45	0.12 *0.05* **0.05**	0.08 *0.04* **0.04**	0.04 *0.03* **0.03**							
54	0.05 *0.03* **0.03**	0.03 *0.03* **0.03**	0.00 *0.00* **0.00**	0.07 *0.04* **0.04**						
102	0.06 *0.03* **0.03**	0.10 *0.05* **0.05**	0.02 *0.02* **0.02**	0.08 *0.04* **0.04**	0.07 *0.04* **0.04**					
251	0.00 *0.00* **0.00**	0.03 *0.02* **0.02**	0.02 *0.02* **0.02**	0.10 *0.04* **0.04**	0.06 *0.04* **0.04**	0.04 *0.03* **0.03**				
409	0.02 *0.02* **0.02**	0.03 *0.03* **0.03**	0.00 *0.00* **0.00**	0.04 *0.03* **0.03**	0.03 *0.03* **0.03**	0.07 *0.04* **0.04**	0.03 *0.03* **0.03**			
531	0.02 *0.02* **0.02**	0.00 *0.00* **0.00**	0.05 *0.03* **0.03**	0.04 *0.03* **0.03**	0.03 *0.03* **0.03**	0.00 *0.00* **0.00**	0.08 *0.05* **0.04**	0.03 *0.03* **0.03**		
824	0.19 *0.06* **0.07**	0.03 *0.03* **0.03**	0.11 *0.05* **0.05**	0.06 *0.03* **0.03**	0.03 *0.03* **0.03**	0.00 *0.00* **0.00**	0.03 *0.03* **0.03**	0.00 *0.00* **0.00**	0.03 *0.03* **0.03**	

*See Gowans et al. (2001) for details of this study. Used here are only animals photo-identified on more than 20 days, with association defined as observed within 1 h of one another.

association index is 0.9, there is a 35% probability that $\hat{\alpha} = 1$ and $\mathrm{SE}(\hat{\alpha}) = 0$, and in this case, the appearance of certainty that they always associate. In such situations, it is probably more useful to estimate the confidence interval of the association index using a method such as Wilson's (1927) score for binomial proportions (here given for a 95% confidence interval):

$$\mathrm{CI}(\alpha) = \frac{\alpha + z_{0.975}^2/2d \pm z_{0.975}\sqrt{\left[\alpha(1-\alpha) + z_{0.975}^2/4d\right]/d}}{1 + z_{0.975}^2/d}$$

where $z_{0.975}$ is the 97.5th percentile of the normal distribution.

How Well Do Matrices of Association Indexes Reflect Social Structure?

As an association index indicates the relationship of a dyad, a matrix of association indexes among members of a community (such as Table 1) indicates their social structure. However, as shown in the previous section, the true and calculated association indexes may differ considerably. A measure of the utility of a matrix of estimated association indexes in describing social structure is the correlation coefficient (product-moment or Pearson correlation coefficient) between the true association indexes and estimated association indexes: $r(\{\alpha_{ij}\}, \{\hat{\alpha}_{ij}\}; i \neq j)$.

$$r(\{\alpha_{ij}\}, \{\hat{\alpha}_{ij}\}; i \neq j) = \frac{\sqrt{\mathrm{Var}(\alpha_{ij})}}{\sqrt{\mathrm{Var}(\hat{\alpha}_{ij})}} \qquad (2)$$

If S is the CV of the true association indexes (see Appendix), $\mathrm{Var}(\alpha_{ij}) = S^2 \times \mathrm{mean}(\alpha_{ij})^2$, and $\mathrm{mean}(\hat{\alpha}_{ij})$ approximates $\mathrm{mean}(\alpha_{ij})$. Then:

$$r(\{\alpha_{ij}\}, \{\hat{\alpha}_{ij}\}; i \neq j) = \frac{S}{\mathrm{CV}(\hat{\alpha}_{ij})} \qquad (3)$$

As S can be estimated by maximum likelihood (see Appendix), equation (3) provides an empirical method for assessing the efficacy by which a matrix of association indexes reflects real social structure for real data.

I examined how well equation (3) performs in practice using simulations. Each simulation used 10 individuals, with true association indexes among pairs of them (α_{ij}) being produced by the beta distribution with a mean 0.4 and a standard deviation equal to 0.1, 0.2, 0.3 or 0.4, for different sets of runs. There were 20 runs for each of these standard deviations and for means of 5, 15 and 45 samples per dyad (d_{ij}). In each run, the program chose actual numbers of samples per dyad (d_{ij}) from the discrete uniform distribution (from zero to twice the mean) and the numbers of observations of associations per dyad (x_{ij}) from the binomial distribution with parameters α_{ij} and d_{ij}, then calculated the simple ratio index, $\hat{\alpha}_{ij} = x_{ij}/d_{ij}$. The program then calculated the true correlation between $\{\alpha_{ij}\}$ and $\{\hat{\alpha}_{ij}\}$, and compared it with the estimated value from equation (3), with S estimated using the likelihood method (see Appendix).

These simulations suggest that equation (3) provides a reasonably good approximation of the real correlation coefficient between the true and estimated association indexes (Fig. 2), except for a small negative bias when $r > \sim 0.9$, which results from the negative bias in the estimation of S (see Appendix). Thus, equation (3) allows the evaluation of how well a matrix of association indexes, for instance Table 1, or a display of the matrix such a cluster analysis dendrogram or ordination using multidimensional scaling (Morgan et al. 1976), represents the real pattern of social relationships.

In the case of the bottlenose whale data, using the relationship in equation (3), the estimated correlation between the association indexes given in Table 1 and the true association indexes among the whales was $r = 0.60$ (bootstrap $\mathrm{SE} = 0.06$), indicating a moderately useful representation of social structure.

If we assume that $\{x_{ij}\}$ are Poisson distributed, which is reasonable if the observations of association are considerably less than the sample sizes ($x \ll d$), equation (2) is equivalent to:

$$r(\{\alpha_{ij}\}, \{\hat{\alpha}_{ij}\}) = \sqrt{\frac{1}{1 + \frac{1}{S^2 \times G}}} \qquad (4)$$

where G is the mean number of observed associations per dyad, $G = \sum_i \sum_{j \neq i} x_{ij}/n(n-1)$, where n is the number of individuals. This relationship (equation (4)) allows calculation of the number of observations per dyad required to achieve a desired correlation between the true and estimated association indexes, as is done in Table 2. Equation (4) assumes that effort is equally concentrated on all dyads. Unequal distribution of effort will reduce the ability to assess the pattern in the true association indexes.

The amount of data needed to give a representation with a given level of correlation to the true pattern varies greatly with the social differentiation (Table 2). A poorly

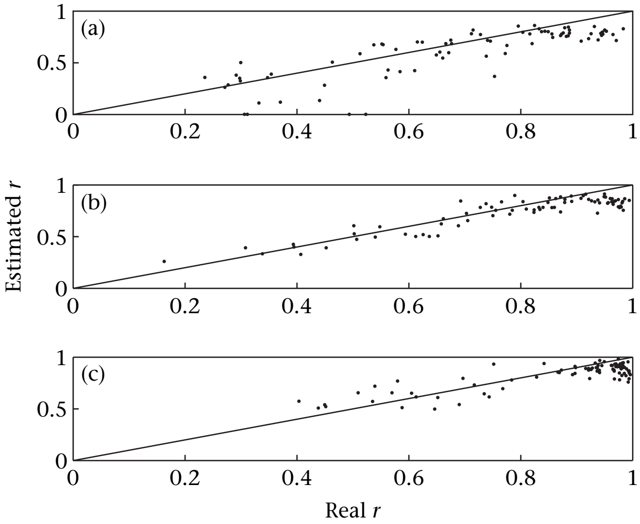

Figure 2. Performance of formula in equation (3) in estimating the correlation coefficient between the true association indexes and their estimated values. Each plot represents one run. The diagonal lines indicate ideal estimation. The mean number of observed associations per dyad (x_{ij}) was (a) 2, (b) 6 and (c) 18.

Table 2. Estimates of the quantity of data required, as expressed by the mean number of observed associations per dyad, to obtain correlation coefficients between the true and estimated association indexes of 0.4 or 0.8 from equation (4), and to reject the null hypothesis ($P < 0.05$) of no preferred or avoided companions using the Bejder et al. (1998) test (from equation (5))

Social differentiation (S)	Mean number of observed associations per dyad (G) for:		Mean observed associations per individual (H) for predicted rejection of null hypothesis of no preferred/avoided companionship
	r=0.4	r=0.8	
0.05	76.19	711.11	2000
0.2	4.76	44.44	125
0.8	0.30	2.78	7.8
2.5	0.03	0.28	0.8
10.0	0.002	0.02	0.05

differentiated population ($S <$ ~0.2) requires many associations per dyad to achieve even a somewhat representative pattern, whereas when social differentiation is high, the data requirements are much less. Generally, an analysis requires about 10 times as much data for a representation with $r = 0.8$ as for one with $r = 0.4$ (Table 2).

Power of Tests for Social Preference

I was unable to derive a useful analytical predictor of the power of Bejder et al.'s (1998) test for preferred or avoided association using statistical theory. Therefore I used simulated data sets to examine the power of this test. The data sets had four social differentiations: $S = 0.06$ ($\{\alpha_{ij}\}$ chosen randomly and uniformly in the range of 0.27–0.33); $S = 0.19$ ($\{\alpha_{ij}\}$ chosen randomly in the range of 0.2–0.6); $S = 0.58$ ($\{\alpha_{ij}\}$ chosen randomly in the range of 0.0–1.0); and $S = 9.95$ ($\alpha_{ij} = 1.0$ with probability 0.01, $\alpha_{ij} = 0.0$ with probability 0.99). The simulations used combinations of the following parameters: 10, 60, or 200 individuals in the population, 10, 60, or 200 sampling periods, and rates of identification per sampling period of 5%, 45% and 90%. The program produced three data sets for each value of S and all combinations of these parameters, although, in order to achieve sufficient associations to make the analysis meaningful, the 5% identification rate was used only with populations of 200 individuals, and with $S = 9.95$, only populations of 60 and 200 individuals and sampling rates of 45% and 90% were considered. The routine then tested each data set for preferred or avoided companions using the Bejder et al. (1998) test.

Figure 3 shows the results of these simulations. I looked, 'by eye', for a simple rule that would indicate the result of the Bejder et al. test. The test usually rejected ($P < 0.05$) the null hypothesis of no preferred companions if:

$$S^2 \times H > 5 \tag{5}$$

where H is the mean number of observed associations per individual (not per dyad as in the case of G; $H = \sum_i \sum_{j \neq i} x_{ij}/n$). Usually, the test did not reject the null hypothesis when the relationship in equation (5) did not hold (Fig. 3). For the 225 simulated data sets, this function (equation (5)) had 91% success in predicting the rejection of the null hypothesis (as compared to 90% for the multivariate discriminant function using log(S) and log(H) as predictor variables). In order to detect preferred companionship, a data set with low social differentiation ($S =$ ~0.05) needs thousands of observed associations per individual, one with intermediate social differentiation ($S =$ ~0.5) tens of observed associations, and one with high social differentiation ($S >$ ~2.5) less than one association per individual (Table 2).

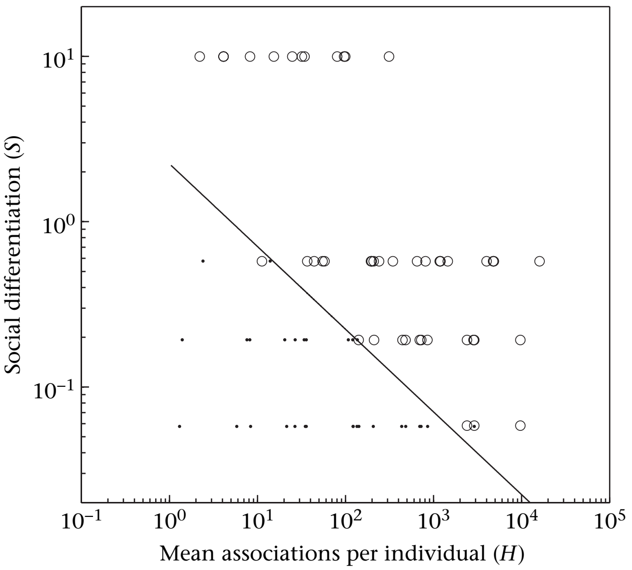

Figure 3. Ability of a data set to detect nonuniform probabilities of association using the Bejder et al. (1998) test. The social differentiation (CV of the true association indexes, S), is plotted against the mean number of observed associations per individual (H), with dots indicating that the results of permutation tests on three simulated data sets using the same parameters had mean(P) > 0.05 (so not rejecting the hypothesis of uniform association probabilities) and open circles indicating a significant result, with mean(P) < 0.05. The line shows the relationship $S^2 \times H = 5$, which approximately divides the significant from the nonsignificant results.

For the bottlenose whale data, whose association matrix is given in Table 1, $S^2 \times H = 5.05$, indicating borderline ability to reject the null hypothesis of no preferred/avoided companions.

Computer Program

SOCPROG, a package of MATLAB (the MathWorks, Natick, MA, U.S.A.) routines for analysing social structure (http://myweb.dal.ca/~hwhitehe/social.htm), carries out nearly all of the analyses described in this paper (Wilson's score is an exception). A compiled version of SOCPROG, for which the user does not need MATLAB, is also available at no cost.

Conclusions

The analyses carried out in this paper suggest that in most situations substantial quantities of data are needed to give even reasonably useful portrayals of social systems using association data. For instance, for an association index of 0.5 to have a 95% confidence interval spanning less than 0.4–0.6 (and so with $\mathrm{SE}(\hat{\alpha}_{ij}) < 0.05$), the index needs to be based upon at least 50 observations of the pair together (Fig. 1). And, for a population with an intermediate level of social differentiation ($S = 0.5$), a mean of at least 0.76 independent observations of associations per dyad produces only a somewhat representative set of association indexes with $r(\{\alpha_{ij}\}, \{\hat{\alpha}_{ij}\}) = 0.4$. Approximately 7.6 independent observations of associations per dyad achieve a much more reliable representation with $r \geq 0.8$. With $S = 0.5$, many pairs of dyads differ in their true rates of association by a factor of two or more. However, in this situation, the Bejder et al. (1998) test usually rejects the null hypothesis of no preferred or avoided companions only when at least 20 associations per individual have been observed. If the population is less socially differentiated so that differences in associate rates between dyads are more subtle, then these data requirements increase dramatically. Only in cases of highly socially differentiated populations can sparse data say much about social structure.

While the mean precision of association indexes and the validity of representations based upon association indexes are both functions of the number of observed associations per dyad, G (equations (1) and (4)), the power of tests for preferred associations is a function of the number of observed associations per individual, H. Given a level of observation effort, G will tend to decrease with population size and H will tend to increase. Thus, with smaller populations, association indexes will tend to be more precise and representations based upon them more valid, whereas tests of preferred association may be more powerful for larger populations. More generally, as population size increases, the level of detail revealed by social analyses will decline while the ability to make general inferences about the properties of the social structure improves.

It is probable that some published social analyses based on association indexes contain representations of social systems and conclusions about social features that have little validity because of poor analytical precision or power. Others (e.g. Sokal & Rohlf 1994; Ruxton & Colegrave 2006) have stressed previously the general importance of power and precision for quantitative studies of biological systems. The results of this paper, together with recent studies of the performance of other techniques (Poisbleau et al. 2006; Whitehead 2007), allow social analysts to start to estimate power and precision in some of their most common scenarios.

Acknowledgments

Thanks to David Lusseau and Wade Blanchard for statistical advice, and to Lorenz Gygax and three referees for useful comments. The study was funded by the Natural Sciences and Engineering Research Council of Canada.

References

Bejder, L., Fletcher, D. & Bräger, S. 1998. A method for testing association patterns of social animals. *Animal Behaviour*, **56**, 719–725.

Cairns, S. J. & Schwager, S. J. 1987. A comparison of association indices. *Animal Behaviour*, **35**, 1454–1469.

Cole, L. C. 1949. The measurement of interspecific association. *Ecology*, **30**, 411–424.

Efron, B. & Gong, G. 1983. A leisurely look at the bootstrap, the jackknife, and cross-validation. *American Statistician*, **37**, 36–48.

Ginsberg, J. R. & Young, T. P. 1992. Measuring association between individuals or groups in behavioural studies. *Animal Behaviour*, **44**, 377–379.

Gowans, S., Whitehead, H. & Hooker, S. K. 2001. Social organization in northern bottlenose whales (*Hyperoodon ampullatus*): not driven by deep water foraging? *Animal Behaviour*, **62**, 369–377.

Hinde, R. A. 1976. Interactions, relationships and social structure. *Man*, **11**, 1–17.

Morgan, B. J. T., Simpson, M. J. A., Hanby, J. P. & Hall-Craggs, J. 1976. Visualizing interaction and sequential data in animal behaviour: theory and application of cluster-analysis methods. *Behaviour*, **56**, 1–43.

Poisbleau, M., Jenouvrier, S. & Fritz, H. 2006. Assessing the reliability of dominance scores for assigning individual ranks in a hierarchy. *Animal Behaviour*, **72**, 835–842.

Ruxton, G. D. & Colegrave, N. 2006. *Experimental Design for the Life Sciences*. 2nd edn. Oxford: Oxford University Press.

Sokal, R. R. & Rohlf, F. J. 1994. *Biometry*. 3rd edn. New York: W. H. Freeman.

de Vries, H. 1995. An improved test of linearity in dominance hierarchies containing unknown or tied relationships. *Animal Behaviour*, **50**, 1375–1389.

de Vries, H., Stevens, J. M. G. & Vervaecke, H. 2006. Measuring and testing the steepness of dominance hierarchies. *Animal Behaviour*, **71**, 585–592.

Whitehead, H. 1997. Analysing animal social structure. *Animal Behaviour*, **53**, 1053–1067.

Whitehead, H. 1999. Testing association patterns of social animals. *Animal Behaviour*, **57**, F26–F29.

Whitehead, H. 2007. Selection of models of lagged identification rates and lagged association rates using AIC and QAIC. *Communications in Statistics—Simulation and Computation*, **36**, 1233–1246.

Whitehead, H. 2008. *Analyzing Animal Societies: Quantitative Methods for Vertebrate Social Analysis*. Chicago: University of Chicago Press.

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., Bejder, L. & Ottensmeyer, A. C. 2005. Testing association patterns: issues arising and extensions. *Animal Behaviour*, **69**, e1–e6.

Wilson, E. B. 1927. Probable inference, the law of succession, and statistical inference. *Journal of the American Statistical Association*, **22**, 209–212.

Appendix: Estimating Social Differentiation

Whitehead (2008) suggests that the CV of the true association indexes, $\mathrm{CV}(\alpha_{ij})$ $\{i \neq j\}$, can be used as a measure of social complexity, and calls it the social differentiation, S. The CV of the estimated association indexes cannot provide an estimate of social differentiation directly as the variance among estimated association indexes is a compound of the variance in the true association indexes and the sampling variance. However, the method of maximum likelihood separates these sources of variation.

We can consider x_{ij}, the number of observations of individuals *i* and *j* together, to be binomially distributed with coefficient α_{ij}, the true association index, and number of samples d_{ij}, the denominator of the estimated association index. Assuming that the α_{ij} are distributed according to the beta distribution, which gives values between 0 and 1, with mean μ and CV *S*, then the parameters of the beta distribution are $\beta_1 = \mu \times ((1-\mu)/(\mu \times S^2) - 1)$ and $\beta_2 = (1-\mu) \times ((1-\mu)/(\mu \times S^2) - 1)$, and the likelihood of the data, the $\{x_{ij}\}$, given μ and *S* is proportional to:

$$L = \prod_{ij} \int_0^1 \alpha_{ij}^{x_{ij}} \times \left(1 - \alpha_{ij}\right)^{\left(d_{ij} - x_{ij}\right)} \times B\left(\alpha_{ij}, \beta_1, \beta_2\right) \times d\left(\alpha_{ij}\right) \quad \text{(A1)}$$

where $B(\alpha_{ij}, \beta_1, \beta_2)$ is the probability density function of the beta distribution with parameters β_1 and β_2 at α_{ij}. Maximum likelihood then proceeds by choosing μ and *S* to maximize *L*. The integration, as well as the maximization, was carried out numerically using MATLAB7.1 (the MathWorks, Natick, MA, U.S.A.).

Simulation studies indicate that the maximum likelihood estimate (using equation (A1)) is approximately unbiased and of reasonable precision, except when $S \approx 1.0$, when there is a small negative bias, and when $S > \sim 2.0$, when the negative bias becomes more substantial (15% negative bias with $S \approx 2.0$, 23% negative bias with $S \approx 4.5$). The negative bias at high *S* may relate to the numerical methods used for evaluating the probability density function of the beta distribution at extreme parameter values. I suggest using the bootstrap procedure to estimate the precision of the estimate of *S*.