Detection of aggregative behaviour in binary choice experiments

Y. Brostaux

Department of Applied Statistics, Computer Science and Mathematics,
Gembloux Agro-Bio Tech, University of Liege (Belgium)
E-mail: *y.brostaux@ulg.ac.be*

Abstract

Aggregative behaviour can be defined as a natural tendency for individuals to cluster in space into groups of varying sizes. In entomology, this behaviour can be detected in binary choice experiments. Analysis of these experiments reveals specific characteristics (unequal counts between repetitions, dependence of individual choices) that make the traditional adjustment tests fail. We demonstrate that the use of generalized linear models can circumvent these pitfalls and deliver a reliable diagnosis on the social behaviour of the studied invertebrates. The strength of this behaviour can then be evaluated through the use of common correlation models.

Introduction

In agronomy, the ethological study of invertebrates is of primary interest for both pest and auxiliary species, as a step in the research of methods for either controlling or favouring them. Social behaviour is one of the first aspect studied in this context, because it leads to very different strategies for the management of the species. Social patterns can be characterized as neutral, aggregative or territorial. Aggregative behaviour is defined as a natural tendency for individuals to cluster in space into groups of varying sizes, as opposed to territorial behaviour which leads to maximal dispersion of individuals, the neutral pattern corresponding to a completely random spatial distribution, without any inter individual interaction.

Binary choice experiments

Social pattern can be detected in "binary choice experiments". In such experiments, individuals are introduced in a closed area containing two identical target zones, which are in some way attractive for the studied species, surrounded by a neutral area. After a pre established delay, the number of individuals having reached each target zone is counted and the whole process is repeated, each time with different individuals. The number of individuals in each target zone can then be modelled to test hypothesis about the type of social behaviour showed by the species in the context of the experiment.

If the social behaviour is neutral, the individual choices are independent and the number of individuals in one target zone can be modelled by a binomial distribution, given the assumption that the individual probability of the choices remains constant during the experiment. Any lack of fit of the observed numbers on this theoretical distribution can be interpreted as the results of either a territorial or an aggregative behaviour, depending the alternative pattern shown. With a territorial behaviour, individuals tend to choose the least crowded target zone, leading to a more uniform pattern than random, while individuals of an aggregative species will follow the group, resulting in very unbalanced distributions between the targets (Figure 1).

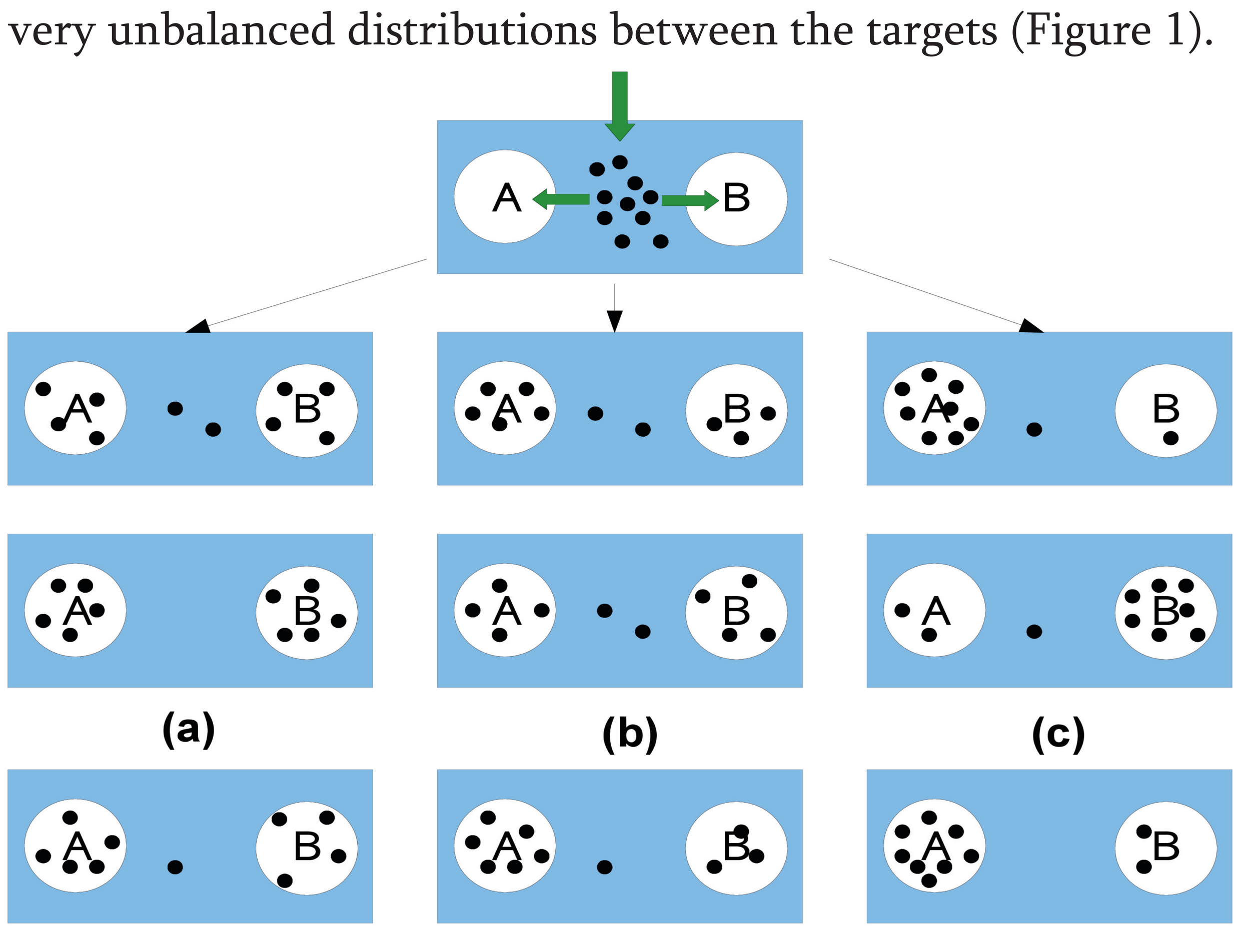

Figure 1 : Schematic representation of a binary choice experiment. N individuals are introduced in a neutral area with two identical target zones A and B. Territorial (a), neutral (b) or aggregative (c) behaviour.

Detecting aggregative behaviour

Analysis of these experiments to detect the type of social behaviour can be conducted through a goodness-of-fit test of the observed counts in the target zones on a binomial theoretical distribution (Figure 2). But observed data present specific characteristics that make the traditional adjustment tests (e.g. Chi-square GOF test) fail. Even if the total number of individuals introduced in the experimental area remains constant between replicates, the number of individuals reaching both target zones varies, as some individuals can be in the neutral area at the time of the observation. Each replicate has its own reference distribution, as the number of trials of the theoretical binomial distributions differs. The real null hypothesis is that the number of individuals in one target zone follows a family of binomial distributions of parameters (n_i, p), n_i being the sum of the counts in both target zones for replicate i.

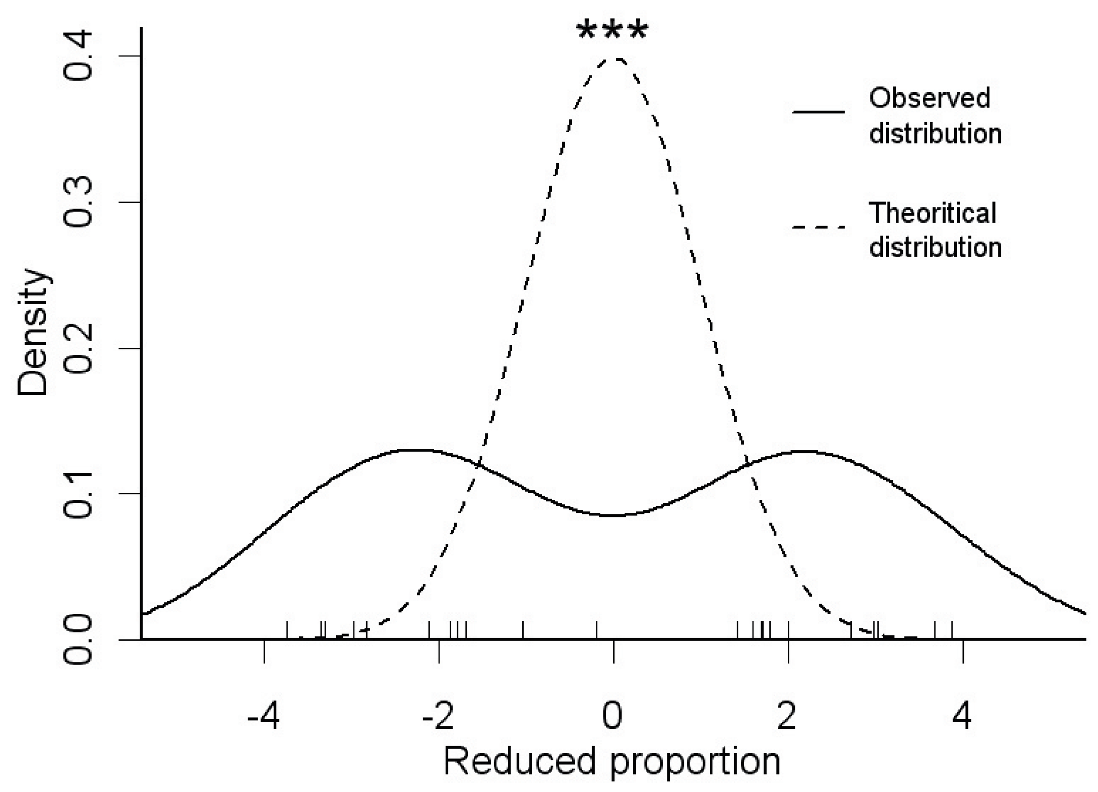

Figure 2 : Observed and theoretical distribution of the proportion of individuals on one target zone, showing strong agregative behaviour. Average proportion is still 50% but observed distribution shows strong departure from binomial assumption, with a bimodal observed distribution.

Generalized Linear Model

Nelder et Wedderburn (1972) defined the generalized linear model (GLMz) as en extension of the standard linear model to other error distributions than the Normal distribution.

In a binary choice experiment, the number of individuals reaching each target zone can easily be adjusted by a GLMz with binomial error distribution family and a logit link. As the two targets are identical, a null model (without any factor) is used, and the GLMz accept without any problem varying total numbers of individuals for each replicate of the experiment.

The significance test of the parameter of this model is not of primary interest here, but can be used to verify the equal attractiveness of both targets. To assess the social behaviour of the studied individuals, we use a property of the residual deviance of the model : if the model is well adjusted and the error distribution is well defined, the residual deviance of the GLMz should asymptotically follow a Chi-square distribution.

Then, a rejection of a goodness-of-fit test on the residual deviance of our model to a chi-square distribution indicates that the error distribution is not binomial, and that the individuals choices are not independant, hence showing some kind of social behaviour.

Further examination of the residual deviance can inform us about the type of social behaviour expressed (Williams, 1982) :

- residual deviance ≈ χ^2 : binomial distribution, independant choices;
- residual deviance <<< χ^2 : underdispersion, territorial choices;
- residual deviance >>> χ^2 : overdispersion, aggregative choices.

Common correlation model

Common correlation model interprets over- and underdispersion by a correlation exisiting between the responses of the members of a group. It can be used to evaluate the strength of the aggregative behaviour detected by GLMz model. Ridout, Demétrio & Firth (1999) showed that the intraclass correlation estimator ρ_{AOV} has good statistical properties while staying simple to calculate.

Conclusion

We now have both a statistical test to assess the social behaviour of groups of invertebrates and an estimator of the strength of this behaviour.

Robustness of the test to extreme experimental parameters (low numbers of individuals and/or replicates) is under investigation.

References

- Nelder, J.A., Wedderburn, R.W. (1972). Generalized linear models. Journal of the Royal Statistical Society Series A, 135 (3), 370–384.
- Ridout, M.S., Demétrio, C.G.B., & Firth, D. (1999). Estimating Intraclass Correlation for Binary Data. Biometrics, 55(1), 137-148.
- Williams, D.A. (1982). Extra-binomial Variation in Logistic Linear Models. Applied Statistics, 31(2), 144-148
- Zirbes, L., Deneubourg, J., Brostaux, Y., & Haubruge, E. (2010). A New Case of Consensual Decision: Collective Movement in Earthworms. Ethology, 116(6), 546-553.