A Critical Look at Compositional Analysis for Assessing Habitat Selection

L. Fattorini, C. Pisani

Abstract The analysis of habitat use by animals is a crucial issue of wildlife management and conservation and probably the main and most simple question to be addressed is if habitat types are used proportionately to their availability. To this purpose, the procedure by Aebischer et al .(1993), based on the use of compositional data analysis, is widely used. However, despite its merits, this procedure has some drawbacks which are likely to render unreliable any conclusion about habitat selection. In this paper the weakness of Aebischer et al. approach is enlightened and an alternative simple permutation solution is proposed.

1 Introduction

Habitat selection is now a burning theme of ecological research owing to the recent advances in GPS technology. However, the first and probably the main and most simple question to be addressed in habitat selection studies is if habitat types are used proportionately to their availability (the so called proportional or random habitat use) or if there is preference/avoidance of some habitat types. As pointed out by Johnson (1980), the analysis can be performed at different levels of choices. In this framework, Aebischer et al. (1993) give a procedure to compare: a) the proportion of each habitat within the home range vs the available proportion within a delineated study area (Johnson's second order selection); b) the proportion of each habitat use vs the corresponding proportion within the home range (Johnson's third order selection). Despite the rising of a plethora of sophisticated models to analyse habitat selection, the procedure by Aebischer et al. is still in wide use. This approach has the merit of viewing habitat selection analysis as the assessment of a system of statistical hypotheses regarding the animal population under study. As such, it proceeds at animal level, i.e. taking animals rather than radio locations as sample units and considering the proportion of animal trajectory (PAT) or the proportion of animal home range (PAHR) contained within each habitat type as the interest variables.

Accordingly, if the radio-tracked animals act independently, the approach completely removes any correlation problem among data which were instead present if radio locations were used as sample units. Despite these appealing features, this procedure suffers from some drawbacks which are mainly induced by the use of compositional data analysis (CODA) introduced by the authors in order to handle the fact PAT and PAHR data are linearly dependent, thus precluding the use of standard multivariate procedures. By means of CODA, log-ratio transforms are used instead of the original data, thus allowing the use of standard multivariate analysis. However, as pointed out by Aitchison (1994), hypotheses regarding compositional data should be consistently reformulated in terms of log-ratios before applying the standard tests, but unfortunately, in the Aebischer et al. procedure this is not performed.

2 Preliminaries and Notations

Given H habitat types, denote by $\mathbf{X}_U = \left[X_{U1}, \ldots, X_{UH}\right]^{\mathrm{T}}$ the random vector in which the random variable X_{Uj} is the proportion of the individual's use of habitat j and denote by $\mathbf{X}_A = \left[X_{A1}, \ldots, X_{AH}\right]^{\mathrm{T}}$ the random vector in which X_{Aj} is the proportion of the availability of habitat $j \left(j = 1, \ldots, H\right)$. If Johnson's second order selection is considered, then $\mathbf{X}_U$ is the H-dimensional random vector of PAHRs while $\mathbf{X}_A$ is a degenerate H-dimensional random vector invariably equal to the vector of H constants $\mathbf{a} = \left[a_1, \ldots, a_H\right]^{\mathrm{T}}$ where a_j is the proportion of habitat j available in the whole study area (which does not vary with animals). On the other hand, if Johnson's third order selection is under study, $\mathbf{X}_U$ is the H-dimensional random vector of PATs while $\mathbf{X}_A$ is the H-dimensional random vector of PAHRs. In both cases, the difference between use and availability is given by $\mathbf{D}_X = \mathbf{X}_U - \mathbf{X}_A = \left[D_{X1}, \ldots, D_{XH}\right]^{\mathrm{T}}$ where $D_{Xj} = X_{Uj} - X_{Aj}$. As positive values of D_{Xj} should mean animal's preference of habitat j while negative values should mean avoidance, the use of $\mathbf{D}_X$ should be, in our opinion, the most natural way for analysing habitat selection. To avoid constrained variables and singular variance-covariance matrices, CODA is based on the choice of a reference habitat, say $l \left(l = 1, \ldots, H\right)$, and on the use of the $\left(H - 1\right)$-variate vectors $\mathbf{Y}_U$ and $\mathbf{Y}_A$ where $Y_{Uj} = \ln\left(X_{Uj} / X_{Ul}\right)$ and $Y_{Aj} = \ln\left(X_{Aj} / X_{Al}\right)$, $j \neq l = 1, \ldots, H$. In this case, habitat selection is analysed by means of the difference vector $\mathbf{D}_Y = \mathbf{Y}_U - \mathbf{Y}_A$ but the differences are in this case less straightforwardly interpretable. Indeed, $D_{Yj} > 0$ entails $X_{Uj} / X_{Aj} > X_{Ul} / X_{Al}$

which means that, with respect to their availabilities, habitat j is used more intensively than habitat l.

3 A Critical Look at Compositional Analysis

In the parlance of Aebischer et al, the hypothesis to be assessed when dealing with habitat selection is that *the average member* of the population has a proportional habitat use. In a more formal framework, the null hypothesis (even if never explicitly mentioned by the authors) should be $H_{X0} : E(\mathbf{X}_U) = \mathbf{a}$ if PAHRs are compared with the constant vector of available proportions or $H_{X0} : E(\mathbf{X}_U) = E(\mathbf{X}_A)$ if PATs are compared with PAHRs. In both cases, the null hypothesis can be expressed as

$$H_{X0} : \boldsymbol{\mu}_X = \mathbf{0} \qquad (1)$$

where $\boldsymbol{\mu}_X = E(\mathbf{D}_X)$ and $\mathbf{0}$ denotes the vector of zeros of adequate dimension.

On the other hand, Aebicher et al. deal with the assessment of

$$H_{Y0} : \boldsymbol{\mu}_Y = \mathbf{0} \qquad (2)$$

where $\boldsymbol{\mu}_Y = E(\mathbf{D}_Y)$ by means of the familiar likelihood ratio test (LRT), which under H_{Y0} as well as under the assumption that $\mathbf{D}_Y$ has a multivariate normal distribution, is asymptotically distributed as a chi-square with $H - 1$ degrees of freedom. However, generally (1) does not coincide with (2). There are few peculiar situations in which (1) entails (2). The first case is when the components of $\mathbf{X}_U$ have the same marginal distribution and the same occurs for the components of $\mathbf{X}_A$ (if it is a random vector) or the components of $\mathbf{a}$ are all equal to $1/H$ while a second case occurs when $\mathbf{X}_U$ and $\mathbf{X}_A$ are random vectors with the same joint distribution. In more general situations, $\boldsymbol{\mu}_Y \neq \mathbf{0}$ even if $\boldsymbol{\mu}_X = \mathbf{0}$. In these cases, the LRT based on the log-ratio transformed data gives rise to an uncontrollable increase of the probability of rejecting (1) when it is true over the nominal level at which the assessment of (1) is performed. On the other hand, it may also happen that (1) is false while $\boldsymbol{\mu}_Y$ is near to $\mathbf{0}$, in such a way that the LRT based on the log-ratios tends to accept (2) with high probabilities. In these cases the procedure will suffer a remarkable loss of power, i.e. the probability of rejecting (1) when false will decrease toward the nominal level at which the assessment of (2) is performed. Moreover, Aebischer et al., in order to cope with the potential lack of normality of $\mathbf{D}_Y$, propose a permutation procedure which (tacitly) presumes $\mathbf{D}_Y$ simmetrically distributed around 0. However, nothing ensures that $\mathbf{D}_Y$ is simmetrically distributed around $\mathbf{0}$ and thus, even if less restrictive than the that based on the

multivariate normality assumption, the permutation procedure is likely to give unreliable evaluation of the *p*-values.

4 A Simple Permutation Solution

In order to overcome the drawbacks of the procedure proposed by Aebischer et al., the assessment of random use of habitat types may be alternatively formulated as

$$\mathrm{H}_{D0} : \bigcap_{j=1}^{H} \mathrm{H}_{0Dj}$$

where

$$\mathrm{H}_{Dj0} : \mathrm{P}(D_{Xj} > 0) = \frac{1}{2} \, .$$

The acceptance of the partial null hypothesis H_{Dj0} means that animals use habitat *j* in a proportion which has the same probability of being greater or less than the proportion of its availability, so that no habitat selection or avoidance may be claimed. Since X_{Uj} and X_{Aj} virtually constitute continuous random variable in $[0,1]$, D_{Xj} is a continuous random variable and the sign test can be adopted for assessing H_{Dj0} $(j=1,\dots,H)$. Then, the permutation procedure by Pesarin (2001) can be applied to combine the p-values resulting from the partial tests into a statistic for assessing the global hypothesis H_{D0}, thus giving rise to a completely nonparametric test.

5 A simulation study

In order to check the performance of the Aebischer et al. procedure and that of the nonparametric test alternatively proposed, a simple Monte Carlo study has been carried out in the framework of Johnson's first and second order selection by considering $H = 5$ habitat types and five different situations ranging from a completely even partition of the study area or of the home range respectively to a very unbalanced one. Animals available and used habitat proportions were generated by a Dirichlet distribution which represents the most familiar model to handle compositional data and the probability of rejecting H_{X0} and H_{0D} was empirically computed. Simulation results show that in some situations the Aebischer et al. procedure gives rise to a probability of rejecting H_{X0} dramatically larger than the nominal one, while the proposed permutation test is always conservative.

References

1. Aebischer N.J., Robertson P.A., Kenward R.E.: Compositional analysis of habitat use from animal radio-tracking data. Ecology, 74, 1315--1325 (1993)
2. Aitchison J.: The Statistical Analysis of Compositional Data. Chapman and Hall, London (1986)
3. Johnson D.S., Thomas D.L., Ver Hoef T.J., Christ A.: A general framework for the analysis of animal resource selection from telemetry data. Biometrics, 64, 968--976 (2008)
4. Pesarin F.: Multivariate Permutation Tests: With Applications in Biostatistics. Wiley, New York (2001)