UNIVERSITÀ DEGLI STUDI DI SIENA

QUADERNI DEL DIPARTIMENTO
DI ECONOMIA POLITICA E STATISTICA

Lorenzo Fattorini
Caterina Pisani
Francesco Riga
Marco Zaccaroni

A Permutation-based Combination of Sign Tests for Assessing Habitat Selection

n. 622 – Ottobre 2011

Abstract - The analysis of habitat use in radio-tagged animals is approached by comparing the portions of use vs the portions of availability observed for each habitat type. Since data are linearly dependent with singular variance-covariance matrices, standard multivariate statistical test cannot be applied. To overcome the problem, compositional data analysis is customary performed via log-ratio transform of sample observations. The procedure is criticized in this paper, emphasizing the many drawbacks which may arise from the use of compositional analysis. An alternative nonparametric solution is proposed in the framework of multiple testing. The habitat use is assessed separately for each habitat type by means of the sign test performed on the original observations. The resulting *p*-values are combined in an overall test statistic whose significance is determined permuting sample observations. The theoretical findings of the paper are checked by simulation studies. Applications to some case studies are considered.

Key words: compositional data analysis, Johnson's second order selection, Johnson's third order selection, Monte Carlo studies, multiple testing, random habitat use.

JEL Classification: C120

Lorenzo Fattorini, Dipartimento di Economia Politica e Statistica, Università di Siena, Siena, Italy
Caterina Pisani, Dipartimento di Economia Politica e Statistica, Università di Siena, Siena, Italy
Francesco Riga, Istituto Sup. per la Protezione e la Ricerca Amb.le (ISPRA), Ozzano Emilia, Italy
Marco Zaccaroni, Dipartimento di Biologia Evoluzionistica, Università di Firenze, Firenze, Italy

Introduction

The analysis of habitat use by animals is a crucial issue of wildlife management and conservation. Habitat selection is now a burning theme of ecological research owing to the recent advances in GPS technology which render available considerable amounts of telemetry data. Manly et al. (2002) provide a general introduction to habitat selection analysis while the special issue of the Journal of Wildlife Management (Strickland & Mc Donald 2006) gives a more updated review of habitat selection issues. More recently, general frameworks for the statistical analysis of habitat selection are furnished by Johnson et al. (2008), Kooper & Manseau (2009) and Kneib, Knauer & Küchenhoff, H. (2011) through the use of weighted distributions, generalized estimating equations and categorical regression, respectively.

The first and probably the main and most simple question to be addressed in habitat selection studies is if habitat types are all used proportionately to their availability (the so called *proportional* or *random habitat use*, henceforth RHU) 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, Robertson & Kenward (1993) give a procedure to compare: a) the portion of each habitat within the home range vs the available portion within a delineated study area (Johnson's second order selection); b) the portion of each habitat use vs the corresponding portion 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, Robertson & Kenward (1993) is still in wide use, as can be checked from the number of citations in impacted journals (see the web site apps.isiknowledge.com).

The pioneering approach by Aebischer, Robertson & Kenward (1993) 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 *portion of animal trajectory* (PAT) or the *portion of animal home range* (PAHR) within each habitat type as the interest variables. Since the trajectory of a single animal is unknown and is approximated by the sequence of radio-tracking data achieved for the animal at discrete times, if radio-tracking times are sufficiently frequent and suitably distributed throughout the monitoring time, the relative frequency of radio locations in each habitat constitutes an unbiased estimator of PAT in the habitat. At the same time, the areal distribution of radio locations, extrapolated by suitable statistical techniques (e.g. kernel smoothing, bivariate normal ellipses or minimum convex hull) constitutes an estimator of the animal home range from which PAHRs can be subsequently derived. In this context, serial correlation among radio tracking data of single animals may constitute a problem only for the estimation of PATs and PAHRs. Following Aebischer, Robertson & Kenward (1993), the actual values of PATs and PAHRs are left undistinguished from their estimates achieved from the radio tracking data, supposing that the number of radio locations adopted for each animal are sufficiently large to give stable and accurate estimates of these quantities. Accordingly, if the radio-tracked animals act independently (e.g. they do not belong to the same flock or herd), the approach completely removes any correlation problem among data which would be instead present if radio locations were used as sample units.

Despite these appealing features, the procedure by Aebischer, Robertson & Kenward (1993) suffers from some drawbacks which are likely to render unreliable any conclusion about habitat selection. The main problems are induced by the use of *compositional data analysis* (henceforth CODA) adopted by the authors in order to handle the fact that PAT and PAHR data recorded from a sample of radio-collared animals are vectors of positive components subject to a unit-sum constraint. Thus, as is well recognized in compositional literature (e.g. Aitchison, 1986, 1994), data are linearly dependent and spurious correlations are induced by the constraint giving rise to singular variance-covariance matrices which, in turn, preclude the use of standard multivariate procedures such as MANOVA or other likelihood ratio tests. On the other hand, by means of CODA, log-ratio transforms are used instead of the original data, thus achieving variance-covariance matrices which are positive definite with probability one and allowing for 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. Unfortunately, in the framework of habitat selection analysis, the RHU hypothesis cannot be generally reformulated in terms of log-ratio expectations and then assessed by the familiar *likelihood ratio test* (LRT) as actually proposed by Aebischer, Robertson & Kenward (1993). As a consequence, the LRT performed on log-ratio data does not necessarily assess the RHU hypothesis. Beside this main problem, the whole procedure tacitly presumes, at least, the symmetry of the distributions of log-ratios around their expectations, which does not necessarily holds. Moreover, in presence of null values of PATs and PAHRs, the use of log ratios necessitates the introduction of very arbitrary solutions.

The purpose of this paper is to propose a pure nonparametric statistical procedure which avoids the use (and the problems) of CODA. The proposed procedure is simply based on the original data and on sign test. While sign test is adopted for assessing RHU for each single habitat, the permutation procedure by Pesarin (2001) is applied to combine the p-values resulting from the single tests for obtaining an overall statistic adopted for the simultaneous RHU assessment in all habitat types. The proposed procedure readily overcomes the problems entailed by the use of CODA only presuming a minimal set of assumptions on PAT and PAHR data.

Materials and methods

Preliminaries and notation

Given K habitat types, denote by $\mathbf{X}_U = [X_{U1}, \ldots, X_{UK}]^{\mathrm{T}}$ the random vector in which the random variable X_{Uj} is the portion of the individual's use of habitat j and denote by $\mathbf{X}_A = [X_{A1}, \ldots, X_{AK}]^{\mathrm{T}}$ the random vector in which X_{Aj} is the portion of the availability of habitat j $(j = 1, \ldots, K)$.

If Johnson's second order selection is analysed, then $\mathbf{X}_U$ is the K-dimensional random vector of PAHRs while $\mathbf{X}_A$ is a degenerate K-dimensional random vector invariably equal to the vector of K constants $\mathbf{a} = [a_1, \ldots, a_K]^{\mathrm{T}}$ in which $a_j > 0$ represents the portion of habitat j available in the whole study area. On the other hand, if Johnson's third order selection is under study, $\mathbf{X}_U$ is the K-dimensional random vector of PATs while $\mathbf{X}_A$ is the K-dimensional random vector of PAHRs. In both cases, the difference between use and availability is given by the random vector $\mathbf{D}_X = \mathbf{X}_U - \mathbf{X}_A = [D_{X1}, \ldots, D_{XH}]^{\mathrm{T}}$, 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.

Owing to the compositional nature of $\mathbf{X}_U$ and $\mathbf{X}_A$, their components are subject to the unit-sum constraints $\mathbf{1}^{\mathrm{T}}\mathbf{X}_U = \mathbf{1}^{\mathrm{T}}\mathbf{X}_A = 1$ where $\mathbf{1}$ is the vector of ones of adequate dimension. Accordingly, the components of $\mathbf{D}_X$ are obviously subject to the zero-sum constraint

$$\mathbf{1}^{\mathrm{T}}\mathbf{D}_X = 0 \qquad (1)$$

As to the nature of the random variables X_{Uj} s and X_{Aj} s, they may virtually take all the values in the closed interval $\left[0,1\right]$ but do not generally constitute continuous random variables in $\left[0,1\right]$. For example, when X_{Uj} represents the PAT in the habitat *j* which is customary estimated by the relative frequency of animal's radio locations in the habitat, then X_{Uj} necessarily takes discrete fractional values in the set $\left\{0/r, 1/r, \ldots, r/r\right\}$ where *r* is the number of radio locations adopted to approximate the animal's trajectory. Moreover, when X_{Uj} or X_{Aj} represent the PAHR in the habitat *j* which is customary achieved by spatial smoothing techniques performed on animal's radio locations, then it may happen $X_{Uj} = 0$ or $X_{Aj} = 0$ if no location of the animal is observed in habitat *j*. On the other hand, the constants a_j s may take all the values in the open interval $(0,1)$, as no available habitat proportion can obviously be 0 (which would mean absence of the habitat) or 1 (which would mean presence of a unique habitat). As a consequence of these considerations the D_{Xj} s are not necessarily continuous random variables in $\left[-1,1\right]$.

Now suppose a sample of *n* radio-collared animals and denote by $\mathbf{x}_{Ui} = \left[x_{U1i}, \ldots, x_{UKi}\right]^{\mathrm{T}}$ the vector in which x_{Uji} is the portion of the use of habitat *j* for animal *i* and by $\mathbf{x}_{Ai} = \left[x_{A1i}, \ldots, x_{AKi}\right]^{\mathrm{T}}$ the vector in which x_{Aji} is the portion of the availability of habitat *j* for animal *i* $(i = 1, \ldots, n)$, in such a way that $\mathbf{d}_{Xi} = \mathbf{x}_{Ui} - \mathbf{x}_{Ai} = \left[d_{X1i}, \ldots, d_{XKi}\right]^{\mathrm{T}}$ where $d_{Xji} = x_{Uji} - x_{Aji}$ constitutes the difference vector. Obviously, in the case of Johnson's second order selection, $\mathbf{x}_{Ai} = \mathbf{a}$ for all *i*. Owing to relation (1), $\mathbf{1}^{\mathrm{T}}\mathbf{d}_{Xi} = 0$ for all $i = 1, \ldots, n$, i.e. the $\mathbf{d}_{Xi}$ s lie in a $(K-1)$ hyperplane. Accordingly, their mean vector, say $\bar{\mathbf{d}}_X = \left[\bar{d}_{X1}, \ldots, \bar{d}_{XK}\right]^{\mathrm{T}}$ is such that $\mathbf{1}^{\mathrm{T}}\bar{\mathbf{d}}_X = 0$ while the variance-covariance matrix, say $\mathbf{S}_X$, is of rank smaller than K, i.e. $\det(\mathbf{S}_X) = 0$.

In order to avoid constrained variables and singular variance-covariance matrices, CODA is based on the arbitrary choice of a reference habitat, say *k*, and on the use of the log-ratio transforms $\mathbf{Y}_U = lrt_k(\mathbf{X}_U)$ and $\mathbf{Y}_A = lrt_k(\mathbf{X}_A)$, where $\mathbf{Y}_U = \left[Y_{U1}, \ldots, Y_{UK}\right]^{\mathrm{T}}$ and $\mathbf{Y}_A = \left[Y_{A1}, \ldots, Y_{AK}\right]^{\mathrm{T}}$ are $(K-1)$ vectors having as components the log-ratios $Y_{Uj} = \ln(X_{Uj}/X_{Uk})$ and $Y_{Aj} = \ln(X_{Aj}/X_{Ak})$, respectively $(j \neq k = 1, \ldots, K)$. In this case, the habitat selection analysis proceeds by means of the difference vector

$\mathbf{D}_Y = \mathbf{Y}_U - \mathbf{Y}_A = \left[D_{Y1}, \ldots, D_{YK} \right]^{\mathrm{T}}$ where $D_{Yj} = Y_{Uj} - Y_{Aj}$, even if the differences are less straightforwardly interpretable. Indeed, $D_{Yj} > 0$ is equivalent to $X_{Uj} / X_{Aj} > X_{Uk} / X_{Ak}$ which means that, with respect to their availabilities, habitat *j* is used more intensively than the reference habitat *k*. It is at once apparent that $\mathbf{Y}_U$ and $\mathbf{Y}_A$ depend on the choice of *k*. However, for simplicity of notation, throughout the paper any mention of the reference habitat is avoided if not essential.

As the X_{Uj}s and X_{Aj}s are random variables on $\left[0,1\right]$ or constants on $(0,1)$, the Y_{Uj}s and Y_{Aj}s are random variables on the real axis. Moreover, no linear relation exists among them, in such a way that the D_{Yj}s constitute a set of linearly independent random variables. Thus, given a sample of *n* radio-collared animals, denote by $\mathbf{y}_{Ui} = \left[y_{U1i}, \ldots, y_{UKi} \right]^{\mathrm{T}}$ the transformed vector $\mathbf{y}_{Ui} = lrt_k(\mathbf{x}_{Ui})$ in which $y_{Uji} = \ln(x_{Uji} / x_{Uli})$ and by $\mathbf{y}_{Ai} = \left[y_{A1i}, \ldots, y_{AKi} \right]^{\mathrm{T}}$ the transformed vector $\mathbf{y}_{Ai} = lrt_k(\mathbf{x}_{Ai})$ in which $y_{Aji} = \ln(x_{Aji} / x_{Ali})$ in such a way that $\mathbf{d}_{Yi} = \mathbf{y}_{Ui} - \mathbf{y}_{Ai} = \left[d_{Y1i}, \ldots, d_{YKi} \right]^{\mathrm{T}}$ where $d_{Yji} = y_{Uji} - y_{Aji}$ constitutes the difference vector. Owing to the linear independence among the components of the $\mathbf{d}_{Yi}$s, these vectors lies in the full $(K-1)$ Euclidean space, in such a way that their mean vector, say $\bar{\mathbf{d}}_Y = \left[\bar{d}_{Y1}, \ldots, \bar{d}_{YK} \right]^{\mathrm{T}}$ is unconstrained while the variance-covariance matrix, say $\mathbf{S}_Y$, is full of rank with a strictly positive determinant $\det(\mathbf{S}_Y)$.

A Critical look at compositional analysis

Theoretical considerations

Usually, statistical hypotheses deal with some aspects of the statistical distribution generating the quantities of interest (e.g expectation, median, distribution function) which are assessed on the basis of a random sample of individuals from the population. In the present case, the hypothesis to be assessed is that the *average member* of the population (in the parlance of Aebischer, Robertson & Kenward 1993) uses habitats proportionately of their availability. In a more formal framework, the null hypothesis (even if never explicitly mentioned by the authors) should be $\mathrm{H}_{X0} : \mathrm{E}(\mathbf{X}_U) = \mathbf{a}$ if PAHRs are compared with the constant vector of available proportions or $\mathrm{H}_{X0}:\mathrm{E}(\mathbf{X}_U)=\mathrm{E}(\mathbf{X}_A)$ if PATs are compared with PAHRs. In both cases, the null hypothesis can be expressed as

$$\mathrm{H}_{X0}:\boldsymbol{\mu}_X=\mathbf{0} \tag{2}$$

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

On the other hand, Aebicher et al. (1993) propose a CODA-based procedure in which the hypothesis

$$\mathrm{H}_{Y0}:\boldsymbol{\mu}_Y=\mathbf{0} \tag{3}$$

is assessed by means of the LRT statistic $-2\ln\lambda$, where $\boldsymbol{\mu}_Y$=E($\mathbf{D}_Y$) and $\lambda=\det(\mathbf{S}_Y)/\det(\mathbf{S}_Y+\overline{\mathbf{d}}_Y\overline{\mathbf{d}}_Y^T)$. Under H_{Y0} and under the assumption that $\mathbf{D}_Y$ has a multivariate normal distribution, $-2\ln\lambda$ is asymptotically (n large) distributed as a chi-square with $K-1$ degrees of freedom. Thus, H_{Y0} is rejected at a level α if $1-F_{K-1}(-2\ln\lambda)\leq\alpha$, where F_m denotes the chi-square distribution function with m degrees of freedom. The fact that a reference habitat k is used as divisor in log-ratios does not cause problems as the LRT (as other multivariate techniques) is invariant under the choice of k (Aitchison, 1986, Chapter 6).

However, as proven in Appendix 1 , (3) does not coincide with the RHU hypothesis of type (2). There are some peculiar situations in which (2) and (3) are equivalent. The first situation occurs in second order selection, when the components of $\mathbf{a}$ are all equal to $1/K$; another situation occurs in third order selection when the components of $\mathbf{X}_U$ are identically distributed random variables and the same occurs for the components of $\mathbf{X}_A$. In more general (and more realistic) situations, $\boldsymbol{\mu}_Y\neq\mathbf{0}$ even if $\boldsymbol{\mu}_X=\mathbf{0}$. In these cases, the LRT based on the $\mathbf{d}_{Yi}$s gives rise to an uncontrollable increase of the probability of rejecting (2) when it is true over the nominal level α at which the assessment of (3) is performed. Obviously, such a probability tends to inflate as $\boldsymbol{\mu}_Y$ differs from $\mathbf{0}$. Accordingly, the unreliability of assessing (2) via the assessment of (3) can be roughly quantified by the Euclidean norm of $\boldsymbol{\mu}_Y$, say $\|\boldsymbol{\mu}_Y\|$, when $\boldsymbol{\mu}_X=\mathbf{0}$. However, since $\boldsymbol{\mu}_Y$ varies with the choice of the reference habitat k, while the probability of rejecting (3) does not depend on k (as the LRT is invariant with respect to k) , a more objective measure of the unreliability of the CODA-based procedure is the averaged norm

$$\Delta = \left\{ \frac{1}{K} \sum_{k=1}^{K} \left\| \mu_{Y/k} \right\|^{2} \right\}^{1/2} \tag{4}$$

where, with obvious notation, $\boldsymbol{\mu}_{Y/k}$ here denotes the expectation of $\mathbf{D}_Y$ when the reference habitat is *k*. Henceforth, Δ will be referred for brevity to as the unreliability measure of CODA-based procedure.

A further problem of the CODA-based procedure is that the determination of *p*-values by means of the chi-square distribution holds asymptotically only if the $\mathbf{d}_{Yi}$s come from a multivariate normal distribution. As nothing ensures multivariate normality of $\mathbf{D}_Y$, the authors propose a permutation procedure which (tacitly) presumes $\mathbf{D}_Y$ simmetrically distributed around $\boldsymbol{\mu}_Y$. If symmetry holds, under (3) $\mathbf{D}_Y$ and $-\mathbf{D}_Y$ are identically distributed in such a way that each difference $\mathbf{d}_{Yi}$ can be randomized by attaching the scalar 1 or -1 with probability 1/2 (or by permuting $\mathbf{y}_{Ui}$ with $\mathbf{y}_{Ai}$, equivalently). Thus, for each data set $\mathbf{d}_{Y1},\ldots,\mathbf{d}_{Yn}$ there are $Q = 2^n$ permutations of these data which may occur with the same probability, from which the permutation distribution of $-2\ln\lambda$ can be determined. Then the *p*-value of the test statistic achieved on the real data set can be obtained from the permutation distribution. Since for *n* large, 2^n permutations may be prohibitive to be considered, the permutation distribution is usually estimated by a random sample of *q* permutations out of the 2^n. However, once again, nothing ensures that $\mathbf{D}_Y$ is simmetrically distributed around $\boldsymbol{\mu}_Y$. A very peculiar case in which symmetry occurs is when $\mathbf{X}_U$ and $\mathbf{X}_A$ are identically and independently distributed. In this case the two vectors are exchangeable in such a way that $\mathbf{D}_Y$ and $-\mathbf{D}_Y$ are equivalent. Thus, even if less restrictive than the procedure based on the assumption of multivariate normality, the permutation procedure may give unreliable evaluation of the *p*-values.

Simulation studies

In order to confirm these theoretical considerations, two Monte Carlo studies were carried out. Firstly, in the framework of second order selection, $K = 5$ habitat types were presumed to partition the study area in accordance with a constant vector **a**. Five different situations were considered, ranging from a completely even partition of the study area into habitats of equal availability to a very unbalanced partition with a dominant habitat covering the 70% of the study area and the remaining ones covering small percentages of 10 and 5% (see Table 1). As the Dirichlet distribution represented the most familiar model to handle with compositional data (see Appendix 2), the vector $\mathbf{X}_U$ was presumed to follow a Dirichlet distribution with parameter $\delta_U \mathbf{a}$ where $\delta_U = 1,10,100$ was an inverse index of variability of the marginal distributions of $\mathbf{X}_U$ (see Appendix 2). In this way, $\mathrm{E}(\mathbf{X}_U) = \mathbf{a}$ irrespective of δ_U, i.e. the RHU hypothesis of type (2) was satisfied for each δ_U.

Then, a sample of $n = 15$ radio-collared animals was presumed and, for each of the five situations and for each value of δ_U, 100,000 samples of size 15 were generated from the Dirichlet distribution with parameter $\delta_U \mathbf{a}$. Then for each sample, the LRT statistic $-2\ln\lambda$ was computed. The function *compana* (with parameters $\mathrm{nrep} = 1000$ e $\mathrm{rnv} = 10^{-18}$) of the package *adehabitat* (version 1.8.3) available in the R software (version 2.12.1) was used to assess H_{Y0} at the nominal levels $\alpha = 0.10,0.05,0.01$ by means of both parametric and permutation procedures (Calenge, 2006). Accordingly, H_{Y0} was rejected if $1 - F_4(-2\ln\lambda) \le \alpha$ when the LRT statistic was compared with the chi-square distribution (parametric test) or if $-2\ln\lambda$ was greater than the $1-\alpha$ quantile of the permutation distribution based on $q = 1000$ permutations (permutation test). Finally, the probability of rejecting H_{X0} was empirically determined as the fraction of times H_{Y0} was rejected. As the LRT statistic was invariant with respect to the choice of the reference habitat, results did not depend on this choice.

A similar Monte Carlo study was repeated in the framework of third order selection. Also in this case, $K = 5$ habitat types were presumed. Then the vector $\mathbf{X}_A$ was presumed to follow a Dirichlet distribution with parameter $\delta_A \mathbf{a}$, where $\delta_A = 100$ and **a** varies in accordance with the five situations considered in the previous experiment (see Table 2), while the vector $\mathbf{X}_U$ was presumed to be independent to $\mathbf{X}_A$ with a Dirichlet distribution with parameter $\delta_U \mathbf{a}$, where $\delta_U = 1,10,100$. In this way, $\mathrm{E}(\mathbf{X}_U) = \mathrm{E}(\mathbf{X}_A) = \mathbf{a}$ irrespective of δ_A and δ_U, i.e. the RHU hypothesis of type (2) was satisfied for each pair δ_A, δ_U, even if for $\delta_U = 1,10$ the variables quantifying habitat use had a greater variability than those quantifying habitat availability.

Then, a sample of $n = 15$ radio-collared animals was presumed and, for each **a** and for each value of δ_U, 100,000 samples of size 15 were generated from the Dirichlet distribution with parameter $100\mathbf{a}$ (availabilities) and coupled with samples of the same size independently generated from the Dirichlet distribution with parameter $\delta_U \mathbf{a}$ (uses). For each couple of samples, the LRT statistic $-2\ln\lambda$ was computed. Once again the probability of rejecting H_{X0} was empirically determined as the fraction of times H_{Y0} was rejected.

During the simulation, Dirichlet random vectors were generated using the function *rdirichlet* available in the *MCMCpack* package (version 1.0-11) of the R software (version 2.12.1).

For each combination of $\mathbf{a}$ and δ_U, Table 1 and 2 report the unreliability measure Δ theoretically determined by means of relations (A2.2) or (A2.3) respectively, as well as the frequency of rejection of (2) corresponding to type 1 errors $\alpha = 0.10, 0.05, 0.01$ at which the assessment of (3) is performed for both parametric and permutation tests.

As expected, the simulation results completely confirm the concerns about the CODA-based procedure:

i) when $\Delta = 0$, i.e. hypotheses (2) and (3) are equivalent, the rejection probabilities of (2) tend to be quite similar to the nominal type 1 errors at which (3) is assessed even if some discrepancies are still observed when the parametric test is used, owing to the lack of multivariate normality of the $\mathbf{d}_{Yi}$s (see lines 1, 6, 11 of Table 1 and lines 1-5, 6 and 11 of Table 2); this problem is considerably reduced by the use of permutation test but discrepancies still remains owing to the lack of symmetry in the $\mathbf{d}_{Yi}$s (see lines 1,6,11 of Table 1 and 6, 11 of Table 2); as theoretically argued, the rejection probabilities of (2) coincide with the nominal type 1 error for (3) when $\mathbf{X}_U$ and $\mathbf{X}_A$ are independently and identically distributed (as for the first five cases of Table 2);

ii) apart from these peculiar cases, when $\Delta \neq 0$, as generally happens in practical situations, the rejection probabilities of (2) turn out to be considerably greater than the nominal type 1 error of (3) and the differences tend to be more and more marked as Δ increases; practically speaking, when the availability of habitat types (fixed or expected) is uneven and when $\mathbf{X}_U$ and/or $\mathbf{X}_A$ show a marked variability (as may occur when a limited number of radio locations are adopted to quantify PATs and/or PAHRs) (3) is rejected all the times even if RHU is true (see the last lines of Table 1 and 2).

Further concerns

As already pointed out in the Introduction, practical problems occur for the CODA-based procedure in presence of 0s. Indeed, as emphasized in the previous section, the X_{Uj}s and X_{Aj}s are customary quantified in the field by radio-tracking data in such a way that they may be 0 when no animal's location is observed in the habitat *j*. In these cases, Aebischer, Robertson & Kenward (1993) suggest substituting zeros with *a "small positive value, less than the smallest recorded non zero proportion*, as *a zero numerator or denominator in the log-ratio transformation is invalid*". The solution seems quite arbitrary and it is likely to heavily impact on the assessment results when the presence of 0s is non negligible.

A simple permutation solution

Theoretical background

The problems induced by the CODA-based procedure suggest using alternative assessments of the RHU hypothesis directly operating on $\mathbf{D}_X$. To this purpose a multivariate nonparametric test for assessing (2) is requested, which avoids unrealistic distributional assumptions on $\mathbf{D}_X$. At least to our knowledge no test of this type is available in literature, as nonparametric assessments on mean vectors invariably involve the symmetry of distributions around the mean vector as a minimal requirement (e.g. Pesarin, 2001, section 3.5 emphasizes that these tests actually constitute multivariate tests of symmetry).

In order to avoid distributional assumptions, the RHU hypothesis must be rephrased in such a way to render necessary only a minimal set of realistic assumptions. As to these assumptions, it is worth noting that in the case of second order selection, the X_{Uj}s represent the PAHRs quantified by spatial smoothing techniques performed on animal's radio locations. As previously pointed out they may be 0 when no radio location is found on the habitat *j* but it is quite difficult that they may coincide with the available portion $a_j > 0$. Accordingly it can be realistically assumed that

$$\Pr(X_{Uj} = a_j) = \Pr(D_{Xj} = 0) = 0 \tag{5}$$

On the other hand, in the case of third order selection, the X_{Uj}s represent the PATs quantified by the relative frequency of animal's radio locations in the habitats while PAHRs play in this case the role of X_{Aj}s. Thus, if $X_{Aj} > 0$, it is quite difficult that it may coincide with the used portion X_{Uj}. Accordingly it can be realistically assumed that

$$\Pr(X_{Uj} = X_{Aj} \mid X_{Aj} > 0) = \Pr(D_{Xj} = 0 \mid X_{Aj} > 0) = 0 \tag{6}$$

As opposite, if no location is observed in the habitat, it may happens that $X_{Aj} = 0$, in which case it obviously also happens that $X_{Uj} = 0$. Hence, $\Pr(D_{Xj} = 0 \mid X_{Aj} = 0) = 1$.

On the basis of these considerations, a suitable hypothesis to be used for both second and third order selection is given by

$$\mathrm{H}^{\circ}_{X0} : \bigcap_{j=1}^{K} \left\{ \pi_j = 0.5 \right\} \tag{7}$$

where $\pi_j = \Pr(D_{Xj} > 0 \mid X_{Aj} > 0)$ and, in case of second order selection, the event $X_{Aj} > 0$ has probability one. Since π_j represents the probability that habitat j, if available, is used more intensively than its availability, the π_js are quantities between 0 and 1 with $\pi_j > 0.5$ when habitat j is preferred, $\pi_j < 0.5$ when habitat j is avoided and $\pi_j = 0.5$ in case of random use. Thus, the obvious sense of (7) is that each habitat type, when available, is used for a portion which has the same probability of being greater or less than the available portion. Even is (7) does not coincides in general with (2) , no habitat selection or avoidance can be claimed for any habitat type if (7) is true. Thus (7) can be suitably taken as the RHU hypothesis to be assessed.

Combination of sign tests

Since (7) is given by the intersection of the *K* partial hypotheses regarding each habitat use, say $\mathrm{H}^{\circ}_{X0j} : \pi_j = 0.5$, the assessment of the partial hypotheses can be straightforwardly performed by means of the sign test, without no assumptions except (5) or (6). Thus, for each habitat j denote by n_j the number of animals for which $x_{Aji} > 0$ (note that in the case of second order selection the n_js are invariably equal to *n*) and by n_j^+ the number of d_{Xji}s strictly greater than 0 and adopt the quantity $t_j = \max(n_j^+, n_j - n_j^+)$ as the test statistic. Under H°_{X0j}, n_j^+ is the realization of a binomial random variable with parameters n_j and 1/2 in such a way that t_j ranges from $n_j/2$ to n_j for n_j even and from $(n_j+1)/2$ to n_j for n_j odd while large values of t_j denote failure of H°_{X0j}. Accordingly the *p*-value corresponding to each t_j is given by

$$p_j = 2^{-n_j+1} \sum_{t=t_j}^{n_j} \binom{n_j}{t} \tag{8}$$

in such a way that H°_{X0j} is rejected at level α when $p_j \le \alpha$. Since the test statistic t_j is discrete, it has a finite number of available *p*-values, usually referred to as *natural p-values* of the test. Actually, if H°_{X0j} is rejected when $p_j \le \alpha$, the test is conservative, in the sense that the true level at which the test is performed coincides with the nearest natural *p*-value smaller than or equal to α. By performing the randomization of the test, any α-level of interest could be achieved. However, as pointed out by Randles & Wolfe (1979), "*this would not be a desirable practice*". It is also worth noting that the fraction $f_j = n_j^+ / n_j$ constitutes an unbiased and consistent (as n_j increases) estimator of π_j. Indeed, the sign test based on t_j is equivalent to the test based on the statistic $\left| f_j - 0.5 \right|$.

Now the key problem is the assessment of the whole hypothesis H°_{X0} at the same prefixed significance level α at which each H°_{X0j} has been assessed. Westfall and Young (1993) investigate the use of the minimum *p*-value, say

$$p = \min(p_1, \dots, p_K) \tag{9}$$

as an overall test statistic to assess H°_{X0}. Subsequently, Pesarin (2001) proposes a more general procedure for multiple testing, considering a wide class of combining functions and referring to (9) as the Tippet combination algorithm.

Accordingly, using Tippet combination, the crucial point reduces to determine the distribution of the minimum *p*-value under H°_{X0}. Indeed, the analytical determination is prohibitive owing to the unknown dependence structure existing among the partial tests. Pesarin (2001, section 5.3) suggests the use of a permutation approach. The approach considers an equally likely random choice of the sign to be attributed to each difference $\mathbf{d}_{Xi}$ in such a way that the random sign affects in the same way all the K differences related to the same animal, thus preserving their dependence relations. Also in this case, there are $Q = 2^n$ possible sign choices with the same probability. Accordingly, denote by t^*_{jv} the value of the sign test adopted for assessing the partial hypothesis H°_{X0j} computed on the *v*-th choice of signs, from which the corresponding *p*-value, say p_{jv}^{*}, can be achieved by means of (8). Then, the sequence of minimum *p*-values $p_{v}^{*} = \min(p_{1v}^{*}, \ldots, p_{Kv}^{*})$ for $v = 1, \ldots, Q$ determines the permutation distribution of (9), from which the overall *p*-values for assessing H_{X0}° turns out to be

$$\tilde{p} = \frac{1}{Q} \sum_{v=1}^{Q} I(p \geq p_{v}^{*})$$

where $I(\bullet)$ is equal to 1 if $\bullet$ is true and 0 otherwise. Accordingly, H_{X0}° is rejected at a level a if $\tilde{p} \leq \alpha$. When *Q* is too large, $\tilde{p}$ can be approximated by using the same procedure performed on a random sample of *q* permutations out of *Q*.

Simulation studies

In order to check the performance of the procedure based on the combination of sign tests as well as to perform comparisons with the CODA-based procedure, two Monte Carlo studies were carried out. In the framework of second order selection, $K = 5$ habitat types were presumed with the same availability vectors **a** considered in the previous simulations. Thus, the random vector $\mathbf{X}_U$ was generated having **a** as the vector of expectations and medians of the X_{Uj} s, in such a way that both the RHU hypotheses H_{X0} and H_{X0}° were true. Since this feature cannot be ensured by Dirichlet distributions, $\mathbf{X}_U$ was generated as $\mathbf{a} + \mathbf{U}$ where $\mathbf{U} = \left[U_1, \ldots, U_K\right]^{\mathrm{T}}$ was a random vector in which the first $K - 1$ components were independent Beta random variables symmetrically distributed in the range $(-w, w)$ with

$$w = \min(a_1, \ldots, a_{K-1}, \frac{a_K}{K-1}),$$

and shape parameter $\beta = 0.10, 0.25, 1$ which constitutes an inverse index of variability, while the last component was given by $U_K = -(U_1 + \ldots + U_{K-1})$ (see Appendix 3). During the simulation, Beta random variables were generated using the function *rbeta* available in the *stats* package (version 2.12.1) of the R software (version 2.12.1).

Then, a sample of $n = 15$ radio-collared animals was presumed and, for each of the five situations and for each value of β, one hundred thousand samples of size 15 were generated.

For each sample, the LRT statistic $-2\ln\lambda$ was used (as in the previous simulation studies) to assess H_{Y0} at the nominal levels $\alpha = 0.10, 0.05, 0.01$ by means of both parametric and permutation procedures and the probability of rejecting H_{X0} was empirically determined as the fraction of times H_{Y0} was rejected. At the same time, for each sample, the *p*-values of the sign tests performed for each partial hypothesis H°_{X0j} was computed by means of (8) together with the overall *p*-values $\tilde{p}$ determined on the basis of a random sample of $q = 1000$ permutations out of $Q = 2^{15}$.

A similar Monte Carlo study was repeated in the framework of third order selection. Once again, $K = 5$ habitat types were presumed to partitioning the study area and the vector $\mathbf{X}_A$ was presumed to follow a Dirichlet distribution with parameter $\delta_A \mathbf{a}$, where $\delta_A = 100$ and $\mathbf{a}$ varies in accordance with the five situations considered in the previous simulations, while the vector $\mathbf{X}_U$ was obtained as $\mathbf{X}_U = \mathbf{X}_A + \mathbf{U}$ where $\mathbf{U}$ was the vector of Beta variables adopted in the previous simulation with shape parameters $\beta = 0.10, 0.25, 1$. The unique exception was the range of the Beta variables, which in this case was given by the random variable

$$W = \min(X_{U1}, \ldots, X_{UK-1}, \frac{X_{UK}}{K-1}) .$$

As shown in Appendix 3, both the vectors $\mathbf{X}_U$ and $\mathbf{X}_A$ had $\mathbf{a}$ as the vector of expectations and medians in such a way that both the RHU hypotheses H_{X0} and H°_{X0} were true.

Then, a sample of $n = 15$ radio-collared animals was presumed and, for each of the five situations and for each value of β, 100,000 samples of size 15 were generated from the Dirichlet distribution with parameter 100$\mathbf{a}$ (availabilities) and coupled with samples of the same size generated by adding the U_j s to the X_{Aj} s (uses). For each couple of samples, the LRT statistic $-2\ln\lambda$ was computed and the probability of rejecting H_{X0} was empirically determined as the fraction of times H_{Y0} was rejected. Moreover, for each sample the *p*-values of the sign tests performed for each partial hypothesis were computed together with the overall *p*-value determined on the basis of a random sample of $q = 1000$ permutations.

For each combination of $\mathbf{a}$ and β, Table 3 and 4 report the unreliability measure Δ together with the frequency of rejection of (2) corresponding to type 1 errors $\alpha = 0.10, 0.05, 0.01$ at which the assessment of (3) is performed for both parametric and permutation tests as well as the frequency of rejection of (7) for the same type 1 errors. As to Δ, since its analytical determination was prohibitive, it was empirically determined by the Monte Carlo counterpart of (4). Moreover, since the rejection rates of the partial hypotheses by means of the sign test turns out to be very similar to those of the overall hypothesis (with differences at third decimal digit) they are omitted for brevity.

While simulation results prove the adequacy of the procedure based on the combination of sign tests, they once again confirm the unreliability of the CODA-based procedure. Indeed:

i) since Δ is invariably greater than zero, the CODA-based procedure shows rejection probabilities invariably greater than the nominal levels with discrepancies which tend to increase with Δ;

ii) the procedure based on the combination of sign tests turns out to be conservative, showing rejection rates for both overall and partial hypotheses invariably smaller than the nominal type 1 error; it is worth noting that the discrepancies between nominal and actual levels are only due to the discrete nature of the sign tests statistic; indeed, the whole simulation was repeated by using the randomized version of the sign test and the resulting rejection rates (rounded at the second decimal digit) turned out to be invariably equal to the nominal type 1 errors.

Ordering habitat by use

When the hypothesis of proportional habitat use is rejected, Aebischer, Robertson & Kenward (1993) propose a next step for ranking the habitat type in order of use. Even if not explicitly mentioned, the ranking criterion adopted by to the authors is based on the number of times, say τ_j, in which $\mathrm{E}\left\{\ln(X_{Uj}/X_{Uk})\right\}$ turns out to be greater than $\mathrm{E}\left\{\ln(X_{Aj}/X_{Ak})\right\}$ for $k \neq j = 1,\dots,K$. The τ_js are integers between 0 and $K-1$ that should rank the habitats in order of what the authors call the *increasing relative use* where 0 is the worst and $K-1$ is the best. As these quantities are actually unknown, the ranking is based on their sample counterparts, say r_j, i.e the number of times in which $\bar{y}_{U(j/k)} = \frac{1}{n}\sum_{i=1}^{n}\ln(x_{Uji}/x_{Ukj})$ turns out to be greater than $\bar{y}_{A(j/k)} = \frac{1}{n}\sum_{i=1}^{n}\ln(x_{Aji}/x_{Akj})$. Unfortunately, the ranking procedure suffers from the same drawbacks pointed out for the CODA-based assessment of RHU. Indeed, owing to the lack of nice results about expectation of ratios and logarithms, inconsistent ranking may take place comparing $\mathrm{E}\left\{\ln(X_{Uj}/X_{Uk})\right\}$ vs $\mathrm{E}\left\{\ln(X_{Aj}/X_{Ak})\right\}$ rather than $\mathrm{E}(X_{Uj}/X_{Uk})$ vs $\mathrm{E}(X_{Aj}/X_{Ak})$ or $\mathrm{E}(X_{Uj})/\mathrm{E}(X_{Uk})$ vs $\mathrm{E}(X_{Aj})/\mathrm{E}(X_{Ak})$. Moreover, as the CODA-based assessment, the procedure suffer from presence of zeros, which must be substituted by arbitrary small values or discarded.

Once again a simple alternative solution can be found with the untransformed data, ranking habitats in accordance with the π_j s. If , the RHU hypothesis H°_{X0} is accepted and no other assessment is performed as the π_j are invariably equal to 0.5. On the other hand, if $\tilde{p} \le \alpha$ and H°_{X0} is rejected, the *p*-value of each partial hypothesis H°_{X0j} are considered in such a way that the whole set of *K* habitat types is partitioned into three disjoint sets: the set of habitat types for which $p_j \le \alpha$ and $f_j > 0.5$ which will be refereed to as the set of preferred habitats or P-habitats; the set of habitat types for which $p_j \le \alpha$ and $f_j < 0.5$ which will be refereed to as the set of avoided habitats or A-habitats and the set for which $p_j > \alpha$ which will be referred to as the randomly used habitats or R-habitats. Practically speaking the partition induces a sort of habitat ordering based on the π_j s, i.e the P-habitats having π_j s greater than 0.5, the R-habitats having π_j s all equal to 0.5 and the A-habitats with π_j s smaller than 0.5.

Since no ordering is necessary within the R-habitats, a further less formal ordering is suitable only within P- and A-habitats, conditional to the partition achieved by the assessment of H°_{X0} and without adjusting *p*-level for multiple testing. The ordering can be performed by assessing the hypothesis $\mathrm{H}^{\circ}_{jk} : \pi_j = \pi_k$ can be performed for each $k \ne j$ in the P- and A-sets by means of the test statistic $t_{jk} = \left| f_j - f_k \right|$. Once again, the *p*-value corresponding to t_{jk} , say p_{jk}, can be determined by using the permutations of sample already adopted to determine $\tilde{p}$, as the fraction of permutations giving rise to a test statistic greater than t_{jh}. If $p_{jk} \le \alpha$ and $f_j > f_k$ than habitat *j* has an higher rank than *k* among P- or A- habitats, the opposite if $f_j < f_k$.

Results

The procedure based on the combination of sign tests was adopted to assess habitat selection in two novel investigations performed in Italy regarding a population of European Brown Hares *(Lepus europaeus)* and Corsican Red Deer *(Cervus elaphus corsicanus)* and on the data set from Aebischer, Robertson & Kenward (1993, Appendix 1) related to thirteen radio-tagged Ring-necked Pheasants (*Phasanius colchicus*) and seventeen radio-tagged Gray Squirrels (*Sciurus corolinensis*), in such a way to compare the results with those achieved by the CODA-based procedure. For all the case studies, assessments were performed at type 1 error $\alpha = 0.05$, thus rejecting random use only in presence of strong empirical evidence. Computations were performed using code in Fortran 77 compiled with Fortran Power Station 4.0.

European Brown Hare

The study was carried out in the protected area of Spicciano (Tuscany, Central Italy). In January 2008, fourteen hares (seven females and seven males) were captured by means of nets. Hares were marked with gps collars (Tellus mini – Televilt, weighting 74 g) scheduled to acquire animal location every 2 hours for 98 days. Hares were released in the same place of capture. Location errors of gps collars were evaluated at about 15 m, in such a way that circles of radius 15 m centred on the recorded positions were likely to cover most true locations. Land use data were recorded by means of field surveys. Vegetation was classified into seven categories: woodland, scrub land and hedges, winter cereals, extensive fruits crops (i.e. vineyards and orchards with cover-crops inter-row), intensive fruit crops (i.e. vineyards and orchards without inter-row cover-crops), meadows, fallow fields. Classification and analysis of GIS data were obtained using AcrView 3.2 (ESRI, Redlands, California). The whole reference area was determined by a minimum convex polygon obtained by pooling the localizations of all hares plus a buffer zone outlining the polygon of width 256 m. The radius of the buffer zone was determined on the basis of the average size of home ranges which turned out to be of 20.5 ha, corresponding to a circle of radius 256 m. Home ranges were determined using 95% fixed kernel ranges (Worton, 1989). Each collected location was attributed to the larger habitat type present in the circle of 15 m radius around the recorded GPS position. Table 5 reports the habitat composition within the study area, the composition of home ranges (PAHR) and the relative frequencies of radio locations (PAT) within each habitat types collected during day and night.

The comparisons of PAHRs vs available area rejected the RHU hypotheses with an overall *p*-values of $\tilde{p} = 0.001$ determined on the basis of all the possible $2^{14} = 16{,}384$ permutations of the sample observations. From the assessment of each habitats, winter cereals were preferred, fallow fields and woodland were avoided and the remaining were randomly used: winter cereals>scrub lands and edges=intensive fruit crops$=$meadows=extensive fruits crops>fallow fields>woodland. (see Table 6a). The comparisons of day PATs vs. PAHRs was not significant, so the hypothesis of RHU was accepted, while the comparison of night PATs vs. PAHRs rejected the RHU hypotheses with an overall *p*-values of $\tilde{p} = 0.009$. Winter cereals were preferred and scrub land and edges were avoided at night, while the remaining habitats were selected at random. The habitat ordering gave rise to: winter cereals>fallow fields=intensive fruit crops =meadows=woodland=extensive fruit crops>scrub land and edges (see Table 6b).

Corsican Red Deer (Cervus elaphus corsicanus)

The study was a part of a more general conservation initiative performed by Ente Foreste Sardegna and by the Italian Institute for Environmental Protection and Research (ISPRA) aimed to create many interconnected populations of the endangered Corsican Red Deer throughout Sardinia (Italy) and to establish a meta-population potentially more persistent to extinction risks due to genetic factors and ecological catastrophes. The reintroduction started in January 2009. So far, thirty individuals were released from Seui fence (centre of Sardinia) and Costa Verde (south western coast) wild populations. Twenty-two animals were reintroduced to Ulassai (south eastern coast) and eight were released in the Montarbu Oasis (centre). A total of twenty individuals were fitted with GPS/GSM radio-collars, scheduled to acquire animal location every 6 hours, while a continuous session (2 fix per hour) was performed every 15 days. A pilot study was performed to assess the GPS collar efficiency and to evaluate location error in each habitat. Data on land-use/land-cover were obtained from the Sardinian regional geographic information system, and reclassified in four main habitat: woodland, scrub land and Mediterranean maquis, pastures, intensive agriculture. Classification and analysis of GIS data were obtained using AcrMap 9.2 (ESRI, Redlands, California). The reference area was defined by a minimum convex polygon obtained by pooling the locations of all the deer plus a buffer zone of width 500 m around the polygon. Home range size was calculated using 95% fixed kernel ranges (Worton, 1989). Data on habitat and home range composition (PAHR) and frequencies of radio locations (PAT) are reported in Table 7.

The reintroduced deer habitat preference was assessed by comparing PAHRs vs. available area. As woodland PAHRs were invariably greater then woodland availability in the study area, while PAHRs of the remaining habitat types were invariably smaller then the corresponding availabilities (see Table 7), RHU hypothesis was rejected with an overall *p*-values of $\tilde{p} = 0.000$ (determined on the basis of all the possible $2^{20} = 1{,}048{,}576$ permutations of the sample observations). Woodland was preferred while the remaining habitats were avoided. (see Table 8a). Interestingly, the comparison of PATs vs PAHRs gave rise to completely opposite results. Indeed, the RHU hypothesis was rejected with an overall *p*-values of $\tilde{p} = 0.000$, but woodland was in this case avoided and scrub land and Mediterranean maquis and pastures were preferred, while intensive agriculture was used at random. Since comparison of scrub land and

Mediterranean maquis vs pasture was not significant, the ranking of habitats was: scrub lands and Mediterranean maquis=pastures>intensive agriculture>woodland (see Table 8b).

Ring-necked Pheasants and Gray Squirrels

Habitats type for pheasants were: scrub, broadleaved woodland, coniferous woodland, grassland, cropland. The comparison of PAHRs vs available area rejected the RHU hypotheses H°_{X0} with an overall *p*-value $\tilde{p} = 0.001$ determined on the basis of all the possible $2^{13} = 8{,}192$ permutations of sample data. Scrub was preferred while the remaining habitats were used at random (see Table 9a). On the other hand, the CODA-based procedure rejected H_{Y0} (permutation *p*-value achieved by $q = 999$ permutations smaller than 0.001) without giving no habitat responsible for the rejection, while habitat ordering gave rise to: crub=broadleaf>conifer=grassland>cropland (Aebischer, Robertson & Kenward 1993).

The comparison of PATs vs PAHRs rejected the RHU hypotheses H°_{X0} with an overall *p*-value $\tilde{p} = 0.000$. Scrub and broadleaf were preferred, grassland was avoided and coniferous and cropland were selected at random. The comparison of scrub vs broadleaf was not significant, so that the ordering was: scrub=broadleaves>coniferous=cropland>grassland (see Table 9b). As to the CODA-based procedure, in order to avoid zeros the analysis was carried out on three habitat types always available for twelve individuals. The procedure rejected H_{0Y} (permutation *p*-value equal to 0.003) and the habitat ordering gave rise to: scrub=broadleaf> grassland (Aebischer, Robertson & Kenward 1993).

As to squirrel study, habitats type were: young beech and spruce plantation, *Thuja* plantation, larch plantation, mature deciduous woodland, open ground. The comparison of PAHRs vs available area rejected the RHU hypotheses H°_{X0} with an overall *p*-value $\tilde{p} = 0.000$ determined on the basis of all the possible $2^{17} = 131{,}072$ permutations of sample data. Larch and mature were preferred, *Thuja* and open were avoided, young was used at random. Since comparisons of larch vs mature and *Thuja* vs open were not significant the ordering was: larch=mature>young>*Thuja*=open (see Table 10a). The CODA-based procedure rejected H_{0Y} (permutation *p*-value smaller than 0.001) giving the same ordering achieved by the combination of sign test (Aebischer, Robertson & Kenward 1993).

As to the comparison of PATs vs PAHRs, *Thuja* plantation was available for two individuals only and hence was excluded by the analysis. The H°_{X0} hypothesis was rejected with an overall *p*-value $\tilde{p} = 0.002$. Mature was preferred, young, larch were avoided and open was selected at random. The comparison between young and larch was not significant, so that the ordering was: mature>open>young=larch (see Table 10b). Also in the CODA-based procedure, *Thuja* plantation was not considered and a modification of the procedure was performed in order to handle the presence of zeros, which occurred in ten animals out of seventeen. The procedure rejected H_{0Y} (permutation *p*-value equal to 0.012) while the habitat ordering gave rise to inconsistent ranking results (Aebischer, Robertson & Kenward 1993).

Discussion

As opposite to CODA-based procedure the procedure based on the combination of sign tests assess the RHU hypothesis without presuming unrealistic assumptions (such as multivariate normality or symmetry) about sample observations. The combination of sign tests is able to handle the presence of zeros in both availability and use data, without involving arbitrary reconstructions of sample data. Interestingly, the use of multiple testing allow to reject at a pre-fixed significance level the overall hypothesis of RHU also determining at the same significance level which habitat types are responsible for rejection. Simulation studies prove that actual significance levels are invariably near to nominal levels, with negligible discrepancies which are only due to the discrete nature of sign test statistic. That is not true for the CODA based procedure in which, for some situations, the actual rejection rates turn out to be much greater than the nominal level.

At the end of the proposed procedure the set of habitat types is partitioned to preferred habitats, avoided habitats and randomly used habitats. Further ordering among preferred and avoided habitats are attempted even in a less formal way.

Acknowledgements

The authors thank Luca Pratelli for his helpful suggestions in the theoretical aspects of the work.

Appendix 1

Writing and rewriting an hypothesis of random habitat use

The RHU hypothesis (2) actually constitutes a multivariate hypothesis which can be rewritten as

$$\mathrm{H}_{X0} : \bigcap_{j=1}^{K} \left\{ \mathrm{E}\,(X_{Uj} - X_{Aj}) = 0 \right\} \tag{A1.1}$$

where $\mathrm{E}(X_{Uj} - X_{Aj}) = 0$ is the univariate hypothesis that the expected use of habitat *j* coincides with its expected (or constant) availability. The obvious sense of (A1.1) is that H_{X0} is true if all the univariate hypotheses are true. In turn, chosen a reference habitat *k*, (A1.1) is equivalent to

$$\mathrm{H}_{X0} : \bigcap_{j\neq k=1}^{K} \left\{ \frac{\mathrm{E}(X_{Uj})}{\mathrm{E}(X_{Uk})} = \frac{\mathrm{E}(X_{Aj})}{\mathrm{E}(X_{Ak})} \right\} \tag{A1.2}$$

Indeed, if (2) is true, than for any habitat j it follows from (A1.1) that $\mathrm{E}(X_{Uj}) = \mathrm{E}(X_{Aj})$ from which $\mathrm{E}(X_{Uj}) / \mathrm{E}(X_{Aj}) = 1$.

Accordingly, for the reference habitat k and for each $j \neq k$, it follows that $\mathrm{E}(X_{Uj}) / \mathrm{E}(X_{Aj}) = \mathrm{E}(X_{Uk}) / \mathrm{E}(X_{Ak})$ from which $\mathrm{E}(X_{Uj}) / \mathrm{E}(X_{Ul}) = \mathrm{E}(X_{Ak}) / \mathrm{E}(X_{Ak})$. As to the reverse, if (A1.2) is true, then for the reference habitat *k* and for each $j \neq k$ it holds that $\mathrm{E}(X_{Uj}) / \mathrm{E}(X_{Uk}) = \mathrm{E}(X_{Aj}) / \mathrm{E}(X_{Ak})$ or equivalently $\mathrm{E}(X_{Uj}) / \mathrm{E}(X_{Aj}) = \mathrm{E}(X_{Uk}) / \mathrm{E}(X_{Ak})$, i.e. for each $j = 1,\ldots,K$ it holds that $\mathrm{E}(X_{Uj}) / \mathrm{E}(X_{Aj}) = c$ or equivalently $\mathrm{E}(X_{Uj}) = c\mathrm{E}(X_{Aj})$.

But since $\sum_{j=1}^{K} \mathrm{E}(X_{Uj}) = \sum_{j=1}^{K} \mathrm{E}(X_{Aj}) = 1$, then $c = 1$, which obviously implies (2).

In a similar way, chosen a reference habitat *k*, (3) constitutes a multivariate hypothesis which is equivalent to

$$\mathrm{H}_{Y0} : \bigcap_{j\neq k=1}^{K} \left\{ \mathrm{E}\,(Y_{Uj} - Y_{Aj}) = 0 \right\}$$

or, more explicitly, to

$$\mathrm{H}_{Y0} : \bigcap_{j \neq k=1}^{K} \left\{ \mathrm{E}\left(\ln \frac{X_{Uj}}{X_{Uk}} \right) = \mathrm{E}\left(\ln \frac{X_{Aj}}{X_{Ak}} \right) \right\}. \tag{A1.3}$$

From (A1.2) and (A1.3), it is at once apparent that (3) is equivalent to (2) if

$$\ln\left\{ \frac{\mathrm{E}(X_{Uj})}{\mathrm{E}(X_{Uk})} \right\} - \ln\left\{ \frac{\mathrm{E}(X_{Aj})}{\mathrm{E}(X_{Ak})} \right\} = \mathrm{E}\left\{ \ln\left(\frac{X_{Uj}}{X_{Uk}} \right) \right\} - \mathrm{E}\left\{ \ln\left(\frac{X_{Aj}}{X_{Ak}} \right) \right\} \tag{A1.4}$$

for each $j \neq k = 1, \ldots, K$. Since $\mathrm{E}\{\ln(X)\}$ generally differs from $\ln \mathrm{E}(X)$, relation (A1.4) does not generally hold.

Appendix 2

Dirichlet distributions and log-ratio transforms

The Dirichlet distribution is probably the most familiar model adopted for positive random vectors $\mathbf{X} = [X_1, \ldots, X_K]^{\mathrm{T}}$ subject to the constraint $\mathbf{1}^{\mathrm{T}}\mathbf{X} = 1$. A *K*-variate random vector **X** is said to have a Dirichlet distribution with parameters $\delta > 0$ and $\boldsymbol{\theta} = [\theta_1, \ldots, \theta_K]^{\mathrm{T}}$ with $\theta_j > 0$ for each $j = 1, \ldots, K$ if the joint probability density function at $\mathbf{x} = [x_1, \ldots, x_K]^{\mathrm{T}}$ with $\mathbf{1}^{\mathrm{T}}\mathbf{x} = 1$ is given by

$$f(\mathbf{x}) = \frac{\Gamma(\delta\theta)}{\prod_{j=1}^{K} \Gamma(\delta\theta_j)} \prod_{j=1}^{K} x_j^{\delta\theta_j - 1}$$

where $\theta = \mathbf{1}^{\mathrm{T}}\boldsymbol{\theta}$. As is well known (e.g. Fang et al., 1990), each marginal variable X_j has a beta distribution on $[0,1]$ with shape parameters $\delta\theta_j$ and $\delta(\theta - \theta_j)$ in such a way that

$$\mathrm{E}(X_j) = \frac{\theta_j}{\theta}$$

and

$$\mathrm{V}(X_j) = \frac{\theta_j(\theta - \theta_j)}{\theta^2(\delta\theta + 1)} .$$

Accordingly, marginal expectations do not depend on δ and marginal variances increase as δ decreases. In the framework of habitat selection analysis, δ obviously accounts for the variability of portions of animal trajectories or home ranges within habitat types. However, when these quantities are estimated on the field by means of animal's radio locations, δ also accounts for the number of radio locations adopted in the study, since marginal variances decrease as the r_i s increase and estimates become close to the real values.

If **X** has a Dirichlet distribution with parameters δ and $\boldsymbol{\theta}$, the log-ratio transform $\mathbf{Y} = lr_k(\mathbf{X})$ is a random vector on R^{K-1} whose *j*-th marginal random variable $Y_j = \ln(X_j / X_k)$ has a logistic distribution of type IV (also referred to as the exponential generalized beta distribution of type II) with expectation

$$\mathrm{E}(Y_j) = \varphi(\delta\theta_j) - \varphi(\delta\theta_k) \qquad \text{(A2.1)}$$

where $\varphi(x) = \partial \ln \Gamma(x) / \partial x$ denotes the digamma function (e.g. Johnson and Kotz, 1970, p. 142, Fang et al., 1990, Problem 1.5).

In the case of Johnson's second order selection, denote by **a** the vector of portions of habitat types in the study area and suppose that $\mathbf{X}_U$ has a Dirichlet distribution with parameters δ_U and **a,** in such a way that H_{X0} is true. Thus, in accordance with (A2.1), the squared value of the unreliability measure of CODA-based procedure turns out to be

$$\Delta^2 = \frac{1}{K} \sum_{k=1}^{K} \sum_{j \neq k} \left\{ \phi(\delta_U a_j) - \phi(\delta_U a_k) - \ln(a_j / a_k) \right\}^2 \qquad \text{(A2.2)}$$

In a similar way, in the case of Johnson's third order selection, suppose that $\mathbf{X}_U$ and $\mathbf{X}_A$ have Dirichlet distributions with the same parameter **a** and variability parameters δ_U and δ_A, respectively, in such a way that H_{X0} is true. From (A2.1), the squared value of unreliability measure is

$$\Delta^2 = \frac{1}{K}\sum_{k=1}^{K}\sum_{j\neq k}\left\{\phi(\delta_U a_j) - \phi(\delta_U a_k) - \phi(\delta_A a_j) - \phi(\delta_A a_k)\right\}^2 \qquad \text{(A2.2)}$$

Appendix 3

Generating dependent compositional data

It is worth noting that $\mathbf{X}_U$ and $\mathbf{X}_A$ arise from the choice of the same animal and as such they should be realistically presumed as dependent random vectors. However, the general problem of constructing dependent random vectors $\mathbf{X}_1 = [\,X_{11},\ldots,X_{1K}\,]^{\mathrm{T}}$ and $\mathbf{X}_2 = [\,X_{21},\ldots,X_{2K}\,]^{\mathrm{T}}$ subject to the constraint $\mathbf{1}^{\mathrm{T}}\mathbf{X}_1 = \mathbf{1}^{\mathrm{T}}\mathbf{X}_2 = 1$ is difficult to solve in the framework of Dirichlet model since any couple of subvectors $\mathbf{X}_1, \mathbf{X}_2$ partitioning a vector $\mathbf{X}$ with a Dirichlet distribution turn out to be independent with marginal Dirichlet distributions (see Fang et al., 1990, Theorem 1.4).

To this purpose, it is convenient to consider one vector, say $\mathbf{X}_1$, distributed as a Dirichlet random vector with parameters $\delta > 0$ and $\boldsymbol{\theta}$ in such a way that $\mathbf{1}^{\mathrm{T}}\mathbf{X}_1 = 1$, and then obtaining $\mathbf{X}_2$ by means of $\mathbf{X}_1 + \mathbf{U}$, where $\mathbf{U}$ is a random vector in which $K-1$ components, say $U_1,\ldots,U_{K-1}$, are random variables in the range $-W, W$ with

$$W = \min(X_{11},\ldots,X_{1K-1},\frac{X_{1K}}{K-1})$$

and $U_K = -(U_1 + \ldots + U_{K-1})$. Indeed, after a straightforward algebra it can be proven that $0 < X_{2j} < 1$ for each $j = 1,\ldots,K$ while $\mathbf{1}^{\mathrm{T}}\mathbf{X}_2 = 1$ by construction. Obviously $\mathrm{E}(X_{2j}) = \mathrm{E}(X_{1j}) + \mathrm{E}(U_j)$, while $\mathrm{V}(X_{2j}) = \mathrm{V}(X_{1j}) + \mathrm{V}(U_j)$, providing that $\mathbf{X}_1$ and $\mathbf{U}$ are independent. If $\mathrm{E}(\mathbf{U}) = \mathbf{0}$, then $\mathbf{X}_1$ and $\mathbf{X}_2$ are dependent with the same mean vector. Moreover, if the U_js are symmetrically distributed around 0, than $\Pr(X_{2j} > X_{1j}) = 0.5$ for each $j = 1,\ldots,K$. These two last features can be readily achieved if the U_js are independent beta variables on $-W, W$ with shape parameters both equal to $\beta > 0$ in such a way that they turn out to be symmetric around 0, with variance

$$\mathrm{V}(U_j) = \frac{1}{4(\beta^2+1)}$$

Accordingly the U_js inflate the variances of the X_{1j} by a term which increases as β approaches 0.

If $\mathbf{X}_1$ coincides with the vector of constants **a**, then if $\mathrm{E}(\mathbf{U}) = \mathbf{0}$ and the U_js are symmetrically distributed around 0, $\mathrm{E}(\mathbf{X}_2) = \mathbf{a}$, $\Pr(X_{2j} > a_j) = 0.5$ and $\mathrm{V}(X_{2j}) = \mathrm{V}(U_j)$ for each $j = 1,\ldots,K$. Obviously, in this case the U_js varies on $-w, w$ with $w = \min(a_1,\ldots,a_{K-1},\frac{a_K}{K-1})$.

Table 1. Type 1 errors of the hypothesis of random habitat use H_{X0}, in the case of Johnson's second order selection for the CODA-based parametric and permutation tests in terms of habitat type availabilities (**a**), variability index (δ_U), unreliability measure (Δ) and nominal type 1 errors (α).

a	δ_U	Δ	Parametric test			Permutation test		
			α			α		
			0.10	0.05	0.01	0.10	0.05	0.01
(0.20, 0.20, 0.20, 0.20, 0.20)	100	0.00	0.19	0.11	0.03	0.10	0.05	0.01
(0.25, 0.25, 0.20, 0.15, 0.15)		0.02	0.19	0.11	0.03	0.10	0.05	0.01
(0.40, 0.25, 0.15, 0.10, 0.10)		0.05	0.21	0.12	0.04	0.11	0.06	0.01
(0.60, 0.10, 0.10, 0.10, 0.10)		0.05	0.23	0.14	0.04	0.13	0.07	0.01
(0.70, 0.10, 0.10, 0.05, 0.05)		0.12	0.28	0.17	0.06	0.16	0.09	0.02
(0.20, 0.20, 0.20, 0.20, 0.20)	10	0.00	0.19	0.11	0.03	0.11	0.05	0.01
(0.25, 0.25, 0.20, 0.15, 0.15)		0.22	0.22	0.14	0.04	0.13	0.07	0.01
(0.40, 0.25, 0.15, 0.10, 0.10)		0.58	0.38	0.26	0.09	0.25	0.14	0.04
(0.60, 0.10, 0.10, 0.10, 0.10)		0.62	0.58	0.43	0.19	0.42	0.28	0.09
(0.70, 0.10, 0.10, 0.05, 0.05)		1.46	0.83	0.70	0.39	0.70	0.53	0.22
(0.20, 0.20, 0.20, 0.20, 0.20)	1	0.00	0.21	0.13	0.04	0.13	0.07	0.02
(0.25, 0.25, 0.20, 0.15, 0.15)		3.25	0.38	0.25	0.09	0.26	0.16	0.05
(0.40, 0.25, 0.15, 0.10, 0.10)		8.40	0.89	0.77	0.46	0.77	0.63	0.32
(0.60, 0.10, 0.10, 0.10, 0.10)		8.97	1.00	0.98	0.85	0.98	0.94	0.74
(0.70, 0.10, 0.10, 0.05, 0.05)		20.11	1.00	1.00	0.99	1.00	0.99	0.93

Table 2. Type 1 errors of the random habitat use hypothesis H_{X0} in the case of Johnson’s third order selection for the CODA-based parametric and permutation tests in terms of use and availability expectations (**a**), variability index (δ_U), unreliability measure (Δ) and nominal type 1 errors (α).

a	δ_U	Δ	Parametric test			Permutation test		
			α			α		
			0.10	0.05	0.01	0.10	0.05	0.01
(0.20, 0.20, 0.20, 0.20, 0.20)	100	0.00	0.19	0.11	0.03	0.10	0.05	0.01
(0.25, 0.25, 0.20, 0.15, 0.15)		0.00	0.19	0.11	0.03	0.10	0.05	0.01
(0.40, 0.25, 0.15, 0.10, 0.10)		0.00	0.19	0.11	0.03	0.10	0.05	0.01
(0.60, 0.10, 0.10, 0.10, 0.10)		0.00	0.19	0.11	0.03	0.10	0.05	0.01
(0.70, 0.10, 0.10, 0.05, 0.05)		0.00	0.18	0.11	0.03	0.10	0.05	0.01
(0.20, 0.20, 0.20, 0.20, 0.20)	10	0.00	0.19	0.11	0.03	0.10	0.05	0.01
(0.25, 0.25, 0.20, 0.15, 0.15)		0.20	0.21	0.13	0.04	0.12	0.06	0.01
(0.40, 0.25, 0.15, 0.10, 0.10)		0.53	0.34	0.22	0.08	0.21	0.12	0.03
(0.60, 0.10, 0.10, 0.10, 0.10)		0.57	0.50	0.35	0.14	0.34	0.21	0.06
(0.70, 0.10, 0.10, 0.05, 0.05)		1.34	0.74	0.59	0.29	0.59	0.41	0.15
(0.20, 0.20, 0.20, 0.20, 0.20)	1	0.00	0.21	0.13	0.04	0.13	0.07	0.02
(0.25, 0.25, 0.20, 0.15, 0.15)		3.23	0.38	0.25	0.09	0.25	0.16	0.05
(0.40, 0.25, 0.15, 0.10, 0.10)		8.35	0.88	0.77	0.45	0.76	0.62	0.31
(0.60, 0.10, 0.10, 0.10, 0.10)		8.92	1.00	0.98	0.85	0.98	0.94	0.73
(0.70, 0.10, 0.10, 0.05, 0.05)		19.99	1.00	1.00	0.99	1.00	0.99	0.92

Table 3. Type 1 errors of the random habitat use hypotheses H_{X0} and H°_{X0} in the case of Johnson's second order selection for the CODA-based parametric and permutation tests and for the combination of sign tests in terms of habitat type availabilities (**a**), variability index (β), unreliability measure (Δ) and nominal type 1 errors (α).

a	β	Δ	Parametric test			Permutation test			Combination of Sign Tests		
			α			α			α		
			0.10	0.05	0.01	0.10	0.05	0.01	0.10	0.05	0.01
(0.20, 0.20, 0.20, 0.20, 0.20)	1.00	0.05	0.20	0.12	0.04	0.10	0.05	0.01	0.03	0.03	0.00
(0.25, 0.25, 0.20, 0.15, 0.15)		0.04	0.20	0.13	0.04	0.11	0.06	0.01	0.03	0.03	0.00
(0.40, 0.25, 0.15, 0.10, 0.10)		0.40	0.30	0.19	0.07	0.18	0.10	0.02	0.03	0.03	0.00
(0.60, 0.10, 0.10, 0.10, 0.10)		0.36	0.30	0.19	0.06	0.19	0.10	0.03	0.03	0.03	0.00
(0.70, 0.10, 0.10, 0.05, 0.05)		0.43	0.29	0.18	0.06	0.17	0.09	0.02	0.03	0.03	0.00
(0.20, 0.20, 0.20, 0.20, 0.20)	0.25	0.13	0.20	0.12	0.04	0.11	0.05	0.01	0.04	0.04	0.00
(0.25, 0.25, 0.20, 0.15, 0.15)		0.11	0.22	0.14	0.05	0.12	0.06	0.01	0.04	0.04	0.00
(0.40, 0.25, 0.15, 0.10, 0.10)		2.24	0.60	0.42	0.17	0.43	0.26	0.07	0.04	0.04	0.00
(0.60, 0.10, 0.10, 0.10, 0.10)		1.93	0.69	0.51	0.20	0.58	0.40	0.14	0.04	0.04	0.00
(0.70, 0.10, 0.10, 0.05, 0.05)		2.33	0.56	0.38	0.14	0.39	0.22	0.05	0.03	0.04	0.00
(0.20, 0.20, 0.20, 0.20, 0.20)	0.10	0.26	0.20	0.12	0.04	0.11	0.06	0.01	0.04	0.04	0.00
(0.25, 0.25, 0.20, 0.15, 0.15)		0.23	0.23	0.14	0.05	0.13	0.07	0.02	0.04	0.04	0.00
(0.40, 0.25, 0.15, 0.10, 0.10)		6.61	0.78	0.60	0.25	0.72	0.52	0.17	0.04	0.04	0.00
(0.60, 0.10, 0.10, 0.10, 0.10)		5.55	0.92	0.80	0.40	0.88	0.75	0.38	0.04	0.04	0.00
(0.70, 0.10, 0.10, 0.05, 0.05)		6.77	0.71	0.51	0.20	0.62	0.40	0.10	0.04	0.04	0.00

Table 4. Type 1 errors of the random habitat use hypotheses H_{X0} and H°_{X0} in the case of Johnson's third order selection for the CODA-based parametric and permutation tests and for the combination of sign tests in terms of use and availability expectations (**a**), variability index (β), unreliability measure (Δ) and nominal type 1 errors (α).

a	β	Δ	Parametric test			Permutation test			Combination of Sign Tests		
			α			α			α		
			0.10	0.05	0.01	0.10	0.05	0.01	0.10	0.05	0.01
(0.20, 0.20, 0.20, 0.20, 0.20)	1.00	0.01	0.19	0.12	0.04	0.11	0.05	0.01	0.04	0.04	0.01
(0.25, 0.25, 0.20, 0.15, 0.15)		0.01	0.20	0.12	0.04	0.11	0.06	0.01	0.04	0.04	0.01
(0.40, 0.25, 0.15, 0.10, 0.10)		0.24	0.25	0.15	0.05	0.15	0.08	0.02	0.04	0.04	0.01
(0.60, 0.10, 0.10, 0.10, 0.10)		0.18	0.23	0.14	0.04	0.14	0.07	0.02	0.04	0.04	0.01
(0.70, 0.10, 0.10, 0.05, 0.05)		0.27	0.25	0.14	0.04	0.15	0.08	0.02	0.04	0.04	0.01
(0.20, 0.20, 0.20, 0.20, 0.20)	0.25	0.12	0.20	0.12	0.04	0.11	0.06	0.01	0.04	0.04	0.01
(0.25, 0.25, 0.20, 0.15, 0.15)		0.12	0.22	0.14	0.04	0.13	0.06	0.02	0.04	0.04	0.01
(0.40, 0.25, 0.15, 0.10, 0.10)		1.04	0.36	0.23	0.07	0.24	0.13	0.03	0.03	0.03	0.01
(0.60, 0.10, 0.10, 0.10, 0.10)		0.70	0.34	0.21	0.06	0.25	0.14	0.04	0.03	0.03	0.01
(0.70, 0.10, 0.10, 0.05, 0.05)		1.24	0.38	0.23	0.06	0.26	0.14	0.03	0.04	0.04	0.01
(0.20, 0.20, 0.20, 0.20, 0.20)	0.10	0.24	0.20	0.12	0.04	0.11	0.05	0.01	0.03	0.03	0.00
(0.25, 0.25, 0.20, 0.15, 0.15)		0.22	0.22	0.14	0.04	0.13	0.07	0.01	0.04	0.04	0.00
(0.40, 0.25, 0.15, 0.10, 0.10)		2.66	0.43	0.27	0.09	0.29	0.16	0.04	0.03	0.03	0.00
(0.60, 0.10, 0.10, 0.10, 0.10)		1.68	0.43	0.26	0.07	0.31	0.18	0.05	0.04	0.04	0.00
(0.70, 0.10, 0.10, 0.05, 0.05)		3.33	0.47	0.29	0.09	0.31	0.17	0.04	0.03	0.03	0.00

Table 5. Habitat composition within the study area and within home ranges (PAHR) and relative frequencies of radio locations (PAT) within each habitat types collected during day and night for a sample of 14 radio collared hares in the area of Spicciano, Tuscany (Central Italy).

animal number	% kernel home range							% radio locations – day (above) and night (below)						
	wood-land	scrub land, hedges	winter cereals	extensive fruit crops	intensive fruit crops	meadows	fallow fields	wood-land	scrub land, hedges	winter cereals	extensive fruit crops	intensive fruit crops	meadows	fallow fields
1	13.54	4.74	40.00	30.75	8.73	0.90	1.34	10.87 1.71	13.59 1.50	32.61 47.54	42.39 36.83	0.00 11.99	0.00 0.21	0.54 0.21
2	0.86	10.53	56.08	24.88	1.15	1.53	4.98	0.00 1.24	32.35 2.90	34.97 70.95	25.16 15.77	0.00 0.62	0.00 1.87	7.52 6.64
3	0.12	7.26	52.91	8.45	0.00	7.46	23.80	0.00 0.00	26.47 2.86	57.65 61.59	7.06 4.76	0.00 0.00	1.76 5.08	7.06 25.71
4	2.19	15.15	30.91	37.73	3.45	4.71	5.87	7.21 1.27	74.17 8.03	5.41 33.19	11.71 45.24	0.30 1.27	0.60 1.69	0.60 9.30
5	15.02	7.72	40.47	11.42	16.76	0.71	7.89	24.32 1.57	39.46 8.74	28.11 35.43	2.16 15.02	0.00 34.53	0.00 0.00	5.95 4.71
6	32.77	2.89	25.58	24.18	7.51	3.57	3.49	62.50 4.15	3.13 1.55	34.38 39.38	0.00 22.80	0.00 27.46	0.00 2.07	0.00 2.59
7	1.33	15.45	46.10	29.07	1.93	2.50	3.63	0.00 0.42	40.80 4.60	24.75 80.13	33.11 11.51	0.00 1.46	1.00 0.84	0.33 1.05
8	0.75	6.60	58.21	20.53	0.70	8.97	4.25	0.00 1.43	24.12 0.61	22.83 66.40	39.23 12.02	0.00 0.20	11.25 12.83	2.57 6.52
9	3.99	12.37	26.31	18.88	20.24	10.97	7.25	9.82 1.31	34.97 4.60	4.91 41.79	3.07 10.07	18.40 17.72	15.95 20.13	12.88 4.38
10	1.24	12.72	48.57	14.35	2.80	7.72	12.60	0.34 0.21	51.19 8.70	28.47 50.93	15.25 7.66	0.34 2.69	0.34 24.64	4.07 5.18
11	2.70	5.40	47.48	26.28	5.07	8.42	4.66	0.65 1.05	9.48 1.27	30.72 53.59	23.86 20.68	1.31 5.27	15.69 11.60	18.30 6.54
12	3.61	13.49	27.23	27.92	10.48	9.66	7.62	2.75 3.32	9.17 9.59	5.50 50.18	13.76 21.77	29.36 1.48	2.75 2.21	36.70 11.44
13	0.35	10.19	50.29	16.84	0.73	11.75	9.86	12.64 2.35	2.23 1.17	28.62 64.79	1.86 12.68	3.35 5.63	37.17 6.57	14.13 6.81
14	8.92	7.75	47.68	14.86	5.42	4.96	10.41	0.33 0.85	12.87 2.97	43.23 77.75	1.98 6.99	0.00 0.21	3.30 1.06	38.28 10.17
% of study area	29.99	6.04	17.21	25.05	4.93	5.95	10.82							

Table 6a. Combination of the sign tests for the assessment of random habitat use from the sample of 14 radio-tagged hares in the area of Spicciano, Tuscany (Central Italy). Home range vs habitat composition. Type 1 error $\alpha = 0.05$.

overall *p*-value $\tilde{p} = 0.001$

habitat type	winter cereals	scrub land, hedges	intensive fruit crops	meadows	extensive fruit crops	fallow fields	woodland	*decision*
winter cereals (1.000)	0.000							*preferred*
scrub land, hedges (0.786)		0.057						*randomly used*
intensive fruit crops (0.500)			1.000					
meadows (0.500)				1.000				
extensive fruit crops (0.357)					0.424			
fallow fields (0.143)						0.013	1.000	*avoided*
woodland (0.071)							0.002	

(values in columns represent the f_j s, diagonal values represent the *p*-values for each individual hypothesis, non diagonal values represent *p*-values for paired comparisons among habitat types; significant result are marked in grey)

Table 6b. Combination of the sign tests for the assessment of random habitat use from the sample of 14 radio-tagged hares in the area of Spicciano, Tuscany (Central Italy). Fraction of radio locations vs home range (night). Type 1 error $\alpha = 0.05$.

overall *p*-value $\tilde{p} = 0.009$

habitat type	winter cereals	fallow fields	intensive fruit crops	meadows	extensive fruit crops	woodland	scrub land, hedges	*decision*
winter cereals (0.929)	0.002							*preferred*
fallow fields (0.429)		0.791						*randomly used*
intensive fruit crops (0.385)			0.581					
meadows (0.357)				0.424				
extensive fruit crops (0.214)					0.057			
woodland (0.214)						0.057		
scrub land, hedges (0.071)							0.002	*avoided*

(values in columns represent the f_j s, diagonal values represent the *p*-values for each individual hypothesis, non diagonal values represent *p*-values for paired comparisons among habitat types; significant result are marked in grey)

Table 7. Habitat composition within the study area and within home ranges (PAHR) and relative frequencies of radio locations (PAT) within each habitat types for a sample of 20 radio-tagged Corsican deer in two areas of Sardinia (Italy).

animal Number	% MCP home range				% radio locations			
	woodland	scrub land, Mediterranean maquis	pastures	intensive agricolture	woodland	scrub land, Mediterranean maquis	pastures	intensive agricolture
1	86.02	13.27	0.42	0.29	76.60	22.95	0.43	0.03
2	82.82	16.05	0.78	0.35	67.91	31.32	0.77	0.00
3	98.36	0.89	0.43	0.32	75.11	16.82	7.85	0.22
4	88.59	9.56	1.61	0.23	63.47	34.98	1.55	0.00
5	86.53	11.89	1.34	0.24	47.11	51.02	1.84	0.03
6	97.76	1.05	1.19	0.00	55.45	10.36	34.19	0.00
7	98.02	1.05	0.94	0.00	55.13	6.13	38.74	0.00
8	87.46	12.09	0.45	0.00	75.00	24.53	0.47	0.00
9	86.16	13.07	0.77	0.00	46.39	51.63	1.90	0.07
10	90.64	8.90	0.47	0.00	55.13	6.13	38.74	0.00
11	80.32	18.37	1.28	0.04	73.71	25.07	1.15	0.07
12	89.28	10.02	0.47	0.23	69.73	28.91	1.24	0.12
13	76.71	22.12	0.69	0.48	51.95	46.78	1.27	0.00
14	84.23	15.01	0.71	0.05	61.99	35.63	2.38	0.00
15	88.43	10.74	0.41	0.41	75.04	17.30	1.07	6.59
16	98.36	0.79	0.81	0.04	82.69	7.03	10.25	0.03
17	95.16	2.72	2.03	0.08	71.92	8.14	19.94	0.00
18	96.82	1.38	1.72	0.08	82.16	6.21	11.63	0.00
19	97.15	1.04	1.72	0.08	82.05	7.50	10.46	0.00
20	89.39	9.76	0.59	0.27	72.28	26.39	1.03	0.31
% of study area	52.21	37.78	5.58	4.42				

Table 8a. Combination of the sign tests for the assessment of random habitat use from the sample of 20 radio-tagged Corsican deer in two areas of Sardinia (Italy). Home range vs habitat composition. Type 1 error $\alpha = 0.05$.

overall *p*-value $\tilde{p} = 0.000$

habitat type	woodland	scrub land, M. maquis	pastures	intensive agricolture	*decision*
woodland (1.000)	0.000				*preferred*
scrub land, M. maquis (0.000)		0.000	1.000	1.000	*avoided*
pastures (0.000)			0.000	1.000	
intensive agricolture (0.000)				0.000	

(values in columns represent the f_j s, diagonal values represent the *p*-values for each individual hypothesis, non diagonal values represent *p*-values for paired comparisons among habitat types; significant result are marked in grey)

Table 8b. Combination of the sign tests for the assessment of random habitat use from the sample of 20 radio-tagged Corsican deer in two areas of Sardinia (Italy). Percentage of radio locations vs home range. Type 1 error $\alpha = 0.05$.

overall p-value $\tilde{p} = 0.000$

habitat type	scrub land, M. maquis	pastures	intensive agricolture	woodland	*decision*
scrub land, M. maquis (0.950)	0.000	0.625			*preferred*
pastures (0.850)		0.003			
intensive agricolture (0.200)			0.077		*randomly used*
woodland (0.000)				0.000	*avoided*

(values in columns represent the f_j s, diagonal values represent the p-values for each individual hypothesis, non diagonal values represent p-values for paired comparisons among habitat types; significant result are marked in grey)

Table 9a Combination of the sign tests for the assessment of random habitat use from the sample of 13 radio-tagged pheasants in Lion Estate (County Kildare, Ireland). Home range vs habitat composition. Type 1 error $\alpha = 0.05$.

overall *p*-value $\tilde{p} = 0.001$

habitat type	scrub	broadleaf	grassland	coniferous	crop	*decision*
scrub (1.000)	0.000					*preferred*
broadleaf (0.769)		0.092				*randomly used*
grassland (0.615)			0.581			
coniferous (0.308)				0.267		
crop (0.231)					0.092	

(values in columns represent the f_j s, diagonal values represent the *p*-values for each individual hypothesis, non diagonal values represent *p*-values for paired comparisons among habitat types; significant result are marked in grey)

Table 9b Combination of the sign tests for the assessment of random habitat use from the sample of 13 radio-tagged pheasants in Lion Estate (County Kildare, Ireland). Percentage of radio locations vs home range. Type 1 error $\alpha = 0.05$.

overall *p*-value $\tilde{p} = 0.000$

habitat type	scrub	broadleaf	coniferous	crop	grassland	*decision*
scrub (0.846)	0.022	1.000				*preferred*
broadleaf (0.846)		0.022				
coniferous (0.750)			0.625			*randomly used*
crop (0.250)				0.625		
grassland (0.000)					0.000	*avoided*

(values in columns represent the f_j s, diagonal values represent the *p*-values for each individual hypothesis, non diagonal values represent *p*-values for paired comparisons among habitat types; significant result are marked in grey)

Table 10a Combination of the sign tests for the assessment of random habitat use from the sample of 17 radio-tagged squirrels in Elton Estate (Northamptonshire, UK). Home range vs habitat composition. Type 1 error $\alpha = 0.05$.

overall *p*-value $\tilde{p} = 0.000$

habitat type	larch	mature	young	thuja	open	*decision*
larch (0.765)	0.049	1.000				*preferred*
mature (0.765)		0.049				
young (0.588)			0.629			*randomly used*
thuja (0.059)				0.000	1.000	*avoided*
open (0.000)					0.000	

(values in columns represent the f_j s, diagonal values represent the *p*-values for each individual hypothesis, non diagonal values represent *p*-values for paired comparisons among habitat types; significant result are marked in grey)

Table 10b Combination of the sign tests for the assessment of random habitat use from the sample of 17 radio-tagged squirrels in Elton Estate (Northamptonshire, UK). Percentage of radio locations vs home range. Type 1 error $\alpha = 0.05$.

overall *p*-value $\tilde{p} = 0.002$

habitat type	mature	open	young	larch	*decision*
mature (0.882)	0.002				*preferred*
open (0.471)		1.000			*randomly used*
young (0.167)			0.039	0.999	*avoided*
larch (0.154)				0.022	

(values in columns represent the f_j s, diagonal values represent the *p*-values for each individual hypothesis, non diagonal values represent *p*-values for paired comparisons among habitat types; significant result are marked in grey)

References

Aebischer, N.J., Robertson, P.A. & Kenward, R.E. (1993) Compositional analysis of habitat use from animal radio-tracking data. *Ecology*, 74, 1315-1325.

Aitchison, J. (1986) *The Statistical Analysis of Compositional Data*. Chapman and Hall, London.

Calenge, C. (2006) The package “adehabitat” for the R software: A tool for the analysis of space and habitat use by animal. *Ecological Modelling*, 197, 516-519.

Johnson, D.H. (1980) The comparison of usage and availability measurements for evaluating resource preference. *Ecology*, 61, 65-71.

Johnson, D.S., Thomas, D.L., Ver Hoef, T.J. & Christ, A. (2008) A general framework for the analysis of animal resource selection from telemetry data. *Biometrics*, 64, 968-976.

Kneib, T., Knauer, F. & Küchenhoff, H. (2011) A general approach to the analysis of habitat selection. *Environmental and Ecological Statistics*, 18, 1-25.

Kooper, N. & Manseau, M. (2009) Generalized estimating equations and generalized linear mixed-effects models for modelling resource selection. *Journal of Applied Ecology*, 46, 590-599.

Manly, B.F.J., Mc Donald, L.L., Thomas, D.L. & Erickson, W.P. (2002) *Resource Selection by Animals*. Kluver, Dordrecht.

Pesarin, F. (2001) *Multivariate Permutation Tests: With Applications in Biostatistics*. Wiley, New York.

Randles, R.H. & Wolfe, D.A. (1979) *Introduction to the Theory of Nonparametric Statistics*. Wiley, New York.

Strickland, M.D. & McDonald, L.L. (2006) Introduction to the special section on resource selection. *Journal of Wildlife Management*, 70, 321-323.

Worton, B.J. (1989) Kernel methods for estimating the utilization distribution in home-range studies. *Ecology*,70,164-168.