

Estimating co-occurrences index in finite populations

S. A. Gattone, T. Di Battista, F. Fortuna

Università G. d'Annunzio Chieti-Pescara

gattone@unich.it

June 6, 2019



Overview

- 1 Introduction
- 2 Spatial auto-correlation of co-occurrence patterns
- 3 Simulation
- 4 Conclusions

What is species co-occurrence

- Species co-occurrence is simply a measurement of the frequency at which two or more species co-occur among a set of spatial locations
- It is important to assess whether that frequency is essentially random, positive, or negative.

To determine whether a particular pair of species is aggregated, segregated or random in occurrence

Why species co-occurrence

Analysis and interpretation of species co-occurrence leads to:

- understand the interaction between species and their coexistence mechanisms
- hypotheses about the structure of species assemblages

Several important implications for biodiversity conservation

Different spatial scales of interest

The distribution of a given species community is dependent:

- to their substrate for example the single tree or a part of it
- groups of trees, a grove (by microclimate characterization and short-range dispersion)
- forest/landscape (by medium and long-range dispersion).

It is important to identify the minimal spatial scale which ensures the minimum living conditions of the species of interest

“I will never be able to protect an entire forest. If I save just some right tree, would then I still be able to preserve the biological populations of interest?”

Methods to study species co-occurrence

The data are organized as a presence–absence matrix where each row represents a species and each column represents a site.

A given **co-occurrence metric** is computed species-by-species and/or aggregated at matrix level

The expected co-occurrence can be obtained through randomization of the matrix (Gotelli, 2000) or through basic probability theory (Veech, 2013)

The co-occurrence is analysed at unit level: spatial patterns in species–environment relationships are not considered

Methods to study species co-occurrence

The data are organized as species abundance values observed in each plot/site of the study area (fully censused).

Codispersion analysis (Cuevas et al, 2013) to characterize spatial patterns in species co-occurrences (Buckley et al, 2016)

Codispersion quantifies covariation of two or more spatial patterns as a function of both spatial lag and direction

Sampling design for finite population framework

The proposal: to study co-occurrences spatial patterns in a sampling design framework

Population $U = \{u_1, u_2, \dots, u_N\}$ of N units

Let W a symmetrical $N \times N$ matrix describing the spatial contiguity existing among the different zones

W reflects some spatial relationship of interest (it may depend on some environmental explanatory variable)

We denote with x_i^h and x_i^k two binary variables indicating at unit i the presence or the absence of species h and k , respectively

Co-occurrence autocorrelation index

We use the following comparison criterion between x_i^h and x_j^k in a generic pair of zones:

If $x_i^h = x_j^k$, $d(x_i^h, x_j^k) = 1$ and if $x_i^h \neq x_j^k$, $d(x_i^h, x_j^k) = -1$

The proposed index is essentially the mean of all the comparisons $d(x_i^h, x_j^k)$ among the pairs

$$r_W = \frac{\sum_{i,j} d(x_i^h, x_j^k) w_{ij}}{\sum_{i,j} w_{ij}} \quad (1)$$

r_W evaluates the dependence between species h and k at the spatial level of interest given by the spatial contiguity matrix W with components w_{ij}

Estimation of r_W

Select a sample of n units by SRSWOR $U_0 = \{u_i : i \in s\}$ where s denotes the set of labels selected in the initial sample

Survey the spatial network, say A_i , i.e. all the units contiguous to u_i , $i = 1, \dots, n$

Let define for each element in the sample the following quantities:

$$c_{l.} = \sum_j d(x_l^h, x_j^k) w_{lj} \quad (2)$$

$$w_{l.} = \sum_j w_{lj} \quad (3)$$

Estimation of r_W

We propose an estimator based on the Horvitz-Thompson ratio estimator

$$\hat{r}_W = \frac{\sum_{l=1}^L \frac{c_l}{\alpha_l}}{\sum_{l=1}^L \frac{w_l}{\alpha_l}} \quad (4)$$

where

- L is the total number of distinct spatial networks in the sample
- α_l is the first-order inclusion probability, i.e. the probability the initial sample U_0 intersects the spatial network A_l

Variance of \hat{r}_W

The approximate variance of \hat{r}_W is the variance of the Horvitz-Thompson estimator of the variable $z_l = c_l - r_W w_l$, given by (Dryver and Chao, 2007)

$$V(\hat{r}_W) = \sum_{j=1}^K \sum_{k=1}^K \frac{z_j z_k (\alpha_{jk} - \alpha_j \alpha_k)}{\alpha_j \alpha_k} \quad (5)$$

where

- K is the number of distinct spatial networks in the population
- α_{jk} is the 2^{nd} order inclusion probability, i.e. the probability that spatial networks A_j and A_k are both intersected in the initial sample U_0

An unbiased estimator of $V(\hat{r}_W)$ is given by

$$\hat{V}(\hat{r}_W) = \sum_{l=1}^L \sum_{m=1}^L \frac{z'_j z'_k (\alpha_{lm} - \alpha_l \alpha_m)}{\alpha_l \alpha_m} \quad (6)$$

where

- $z'_j = c_{.j} - \hat{r}_W w_{.j}$.

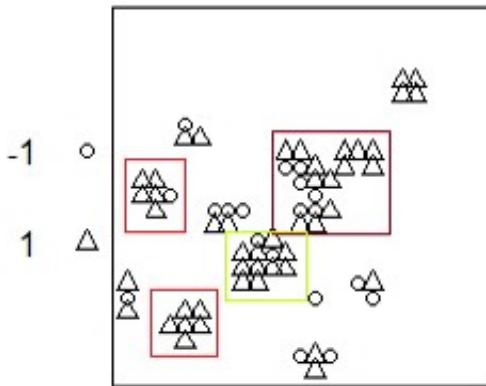
A simulation study

- The spatial contiguity is obtained by simulating a Matern cluster process on a 30×30 grid with $N = 900$ plots with parameters $\lambda = 20$, $\mu = 20$, $\phi = 0.05$
- Different co-occurrence patterns: neutral, aggregation
- Initial sample size $n=10, 20, 40$

Evaluate

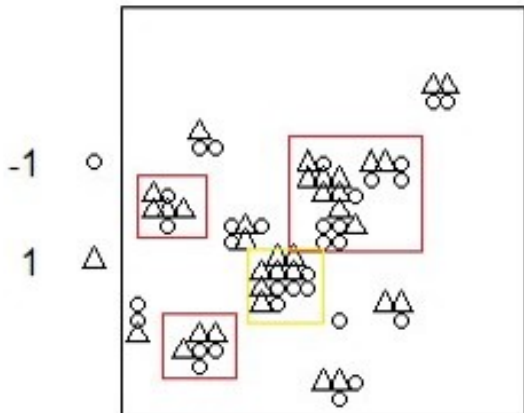
- the bias of the estimator
- the variance of the estimator
- the bias of the variance estimator

An example



An example where there is aggregation at spatial network level:
 $r_W = 0.52$

An example



An example where there is a low aggregation at spatial network level:
 $r_W = -0.06$

n	Bias(\hat{r}_w)%	$\sqrt{V(\hat{r}_w)}$	$\sqrt{E\{\hat{V}(\hat{r}_w)\}}$	Bias($\sqrt{\hat{V}(\hat{r}_w)}$)%
10	5%	0.202	0.152	-24%
20	3%	0.130	0.114	-12%
40	0%	0.074	0.069	-5%

Table: Simulation results with $r_w = -0.06$

n	Bias(\hat{r}_w)%	$\sqrt{V(\hat{r}_w)}$	$\sqrt{E\{\hat{V}(\hat{r}_w)\}}$	Bias($\sqrt{\hat{V}(\hat{r}_w)}$)%
10	7%	0.158	0.113	-28%
20	3%	0.102	0.090	-11%
40	1%	0.054	0.049	-8%

Table: Simulation results with $r_w = 0.52$

Conclusions and further work

- Application to real data (the survey is ongoing at Foresta delle Lame - Northern Appenines, *Lobarion pulmonariae* community)
- Provide tools for a design-based inference
- Extensions to other spatial index such as the Moran's index

References

- H. Buckley, B. Case, and A. Ellison. Using codispersion analysis to characterize spatial patterns in species co-occurrences. *Ecology*, 97(1),32–39, (2016).
- F. Cuevas, E. Porcu, and R. Vallejos. Study of spatial relationships between two sets of variables: a nonparametric approach. *Journal of Nonparametric Statistics*, 25, 695–714, (2013).
- Dryver AL, Chao CT. Ratio estimators in adaptive cluster sampling. *Environmetrics* 18,607–620, (2007).
- N. Gotelli. Null model analysis of species co-occurrence patterns. *Ecology*, 81,2606–2621 (2000).
- J. Veech. A probabilistic model for analysing species co-occurrence. *Global Ecology and Biogeography*, 22, 252–260, (2013).

Thank you
for
your Attention!