Bayesian Analysis of Niche Overlap in Ecology*

Manuel Mendoza R.

Departamento de Estadística
Instituto Tecnológico Autónomo de México

*Joint work with E. Gutiérrez-Peña, G. Núñez-Antonio, A. Contreras-Cristán y E. Mendoza

Outline

Niche overlap

- The study of the interaction among species is an active area of research in Ecology.
- In particular, it is of interest to evaluate the overlap of their ecological niches.
- The niche is the multidimensional hypervolume in which a species maintains a viable population (Hutchinson, 1957).
- The space where the niche is defined includes the spatial and temporal dimensions as well as other environmental variables.

- Temporal activity is one of the axes of the niche most commonly used to explore ecological segregation among animal species.
- Many contributions focus on the overlap of this variable.
- If we record times around a 24-hour clock, then we have a circular variable.
- This is the case that will be analyzed here.

- As for the original objective, it is possible to think, at least theoretically, in terms of the multivariate overlap.
- In practice, however, the overlap is usually calculated separately for each axis of the niche.
- The resulting quantities are then combined to provide an idea of the 'global' overlap.
- There exist several proposals for the combination rule (Geange et al. 2011).

Data

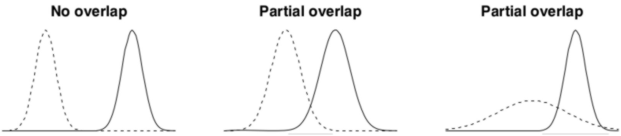
- Collecting information on animal activity patterns in the wild is not an easy task and can be expensive.
- Camera-trapping systems have allowed ecologists to generate useful information to address this issue.
- Some cameras are distributed over the research area.
- Every time a sensor detects some movement, a photo is taken (including date and time of the day).

- At some point the photos are collected and the species that appear in each image are identified.
- As a result we get, for each species of interest, a collection of times of the day $\{X_1, X_2, \ldots, X_m\}$.
- These times are regarded as a random sample from an unknown distribution.
- Given the samples from two species: $\{X_1, X_2, \ldots, X_m\}$ and $\{X_1^*, X_2^*, \ldots, X_n^*\}$, we have to estimate the overlap of the activity patterns.

The Camera-trapping system

Measures of niche overlap

- For a continuous variable X, if the probability density functions $f(x)$ and $g(x)$, corresponding to the species A and B are known, the overlap can be measured in a variety of ways

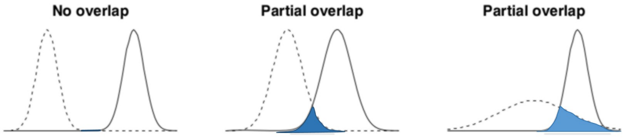


- There are symmetric and asymmetric measures of overlap.
- Two examples of symmetric measures (Ridout & Linkie 2009):

$$\Delta(f,g) = \int \min\{f(x), g(x)\}\,\mathrm{d}x \quad \text{and} \quad \rho(f,g) = \int \sqrt{f(x)g(x)}\,\mathrm{d}x.$$

Here we focus on the *overlap coefficient* $\Delta(f,g)$.

- Overlap coefficient: $\Delta(f, g) = \int \min\{f(x), g(x)\}\, \mathrm{d}x$



- The idea is to estimate Δ using a sample from $f(x)$ and a sample from $g(x)$; (f and g circular distributions).

Parametric models

- In practice, the densites $f(x)$ and $g(x)$ are unknown and must be estimated.
- As in many other applications, the first attempts were based on parametric models, specifically normal distributions.
- Just like any other parametric model, this may be quite a restrictive assumption.
- Alternative: non-parametric estimation of $f(x)$ and $g(x)$.

Non-parametric models

- Kernel density estimation (Silverman 1986).
- For circular data (such as activity times around a 24-hour clock) we can replace the usual Gaussian kernels by distributions defined on the circle.
- The are a number of models available. Two of the most popular are the von Mises and the projected normal distributions.

R package **overlap**

Provides:

- Functions to fit kernel density functions to data on temporal activity patterns of animals;
- Estimates for the coefficient $\Delta(f, g)$;
- Approximate confidence intervals (bootstrap) for $\Delta(f, g)$.

Specific to temporal (circular, cyclic) data.
(Meredith & Ridout 2016).

Bayesian inference

- We propose a Dirichlet Process Mixture model (Escobar & West 1995).
- We can produce non-parametric inferences about $f(x)$, $g(x)$ and any functional thereof.
- In particular, we can obtain the posterior distribution for $\Delta(f, g)$.
- We may use von Mises or projected normal distributions instead of the usual Gaussian kernels.

The DPM model

- For each species, the distribution of X is modelled as a DPM of $F(\cdot|\Psi)$,

$$\begin{aligned} X|\Psi &\sim F(\cdot|\Psi) \\ \Psi|H &\sim H \\ H &\sim DP(\alpha, H_0), \end{aligned}$$

  where the c.d.f. H follows a Dirichlet process with parameters H_0 and α.

- Given a sample $X_1, \ldots, X_n$ from X, we can use sampling-based methods to obtained 'realizations' of the posterior distribution of the density of the variable X.

The DPM model

- For a large integer M, we then get a sample of pairs of estimated densities: $\{(f_i, g_i)\}_{i=1}^{M}$.
- From these, a sample $\{\Delta_i\}_{i=1}^{M}$ from the corresponding posterior distribution of the overlap coefficient Δ, can be readily obtained ($\Delta_i = \Delta(f_i, g_i)$).
- We can then produce any inference of interest regarding Δ (in particular, pointwise estimates and probability intervals) .

Choices for $F(\cdot|\Psi)$

- von Mises distribution

$$F(\theta|\mu,\kappa) = \frac{\exp\{\kappa\cos(\theta-\mu)\}}{2\pi I_0(\kappa)}$$

with $\Psi = (\mu,\kappa)$ and where $I_0(\cdot)$ is a modified Bessel function of the first kind and order 0.

This distribution is also known as the Circular Normal distribution because of its similarities to the Normal distribution on the real line.

Choices for $F(\cdot|\Psi)$

- Projected normal distribution

$$F(\theta|\boldsymbol{\mu}) = \frac{1}{2\pi} \exp\left\{-\frac{1}{2}\boldsymbol{\mu}^t\boldsymbol{\mu}\right\} \left[1 + \frac{\boldsymbol{v}^t\boldsymbol{\mu}}{\phi(\boldsymbol{v}^t\boldsymbol{\mu})}\Phi(\boldsymbol{v}^t\boldsymbol{\mu})\right]$$

where $\boldsymbol{v} = (\cos\theta, \sin\theta)$, and $\phi(\cdot)$ and $\Phi(\cdot)$ denote the p.d.f. and the c.d.f. of a standard normal distribution, respectively.

This is the marginal distribution of the angle θ that we get when a bivariate normal distribution $N(\boldsymbol{X}|\boldsymbol{\mu}, \boldsymbol{I})$ is transformed into polar coordinates.

Choices for $F(\cdot|\Psi)$

- Here, we discuss the results for the DPM of projected normal distributions. This model entails challenges and advantages.
- We only observe the angles $\{\theta_i\}_{i=1}^n$, and the corresponding radial coordinates $\{r_i\}_{i=1}^n$ should be treated as missing.
- A Dirichlet Process Mixture of Projected Normals is the same as a Projected Dirichlet Process Mixture of Normals. Then, many calculations are simplified.

Specific model

- For each species, we define

$$\begin{aligned} \boldsymbol{Y}|\boldsymbol{\mu} &\sim N(\cdot|\boldsymbol{\mu}, \boldsymbol{I}) \\ \boldsymbol{\mu}|H &\sim H \\ H &\sim DP(\alpha, H_0) \end{aligned}$$

  H_0, a bivariate normal distribution $N(\cdot|\boldsymbol{\mu}_0, \boldsymbol{\Sigma}_0)$; $\alpha \sim Ga(a_0, b_0)$.

- The variable of interest (the time of the day X), is given by:

  θ, the angle of the transformation to polar coordinates

$$\boldsymbol{Y} \rightarrow (r, \theta).$$

Prior Distribution

- In the case of a nonparametric analysis, the specification of the prior distribution is usually a rather complicated task.
- There are no general rules to choose a noninformative prior.
- Here, we use a predictive argument.
- A prior is noninformative if the corresponding predictive distribution gives us little information regarding the location of future observations.

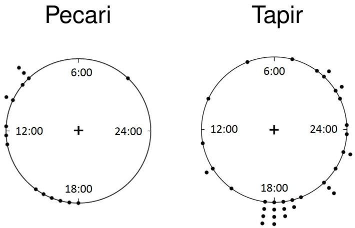
- The parameters of the prior are chosen to produce a prior predictive distribution for θ, that is Uniform over the unit circle.
- Specifically, we adopt the prior:

$$\boldsymbol{\mu} \sim N(\mathbf{0}, \boldsymbol{I}) \quad \text{and} \quad \alpha \sim Ga(2, 4).$$

- The first part guarantees a Uniform predictive *a priori*, and also conditional realizations of the density of θ not far from this average.
- The prior for α, describes rather vague knowledge regarding the mixture model.

Real data

- We have a sample of size $n = 16$ of pecari activity times and a sample of size $m = 35$ of tapir activity times.



- but first...

Simulated data

We considered the following two mixtures:

$$f(\varphi) = 0.1\, \phi^{PN}(\varphi \mid \boldsymbol{\mu}_{11}) + 0.2\, \phi^{PN}(\varphi \mid \boldsymbol{\mu}_{12}) + 0.4\, \phi^{PN}(\varphi \mid \boldsymbol{\mu}_{13}) + 0.3\, \phi^{PN}(\varphi \mid \boldsymbol{\mu}_{14})$$

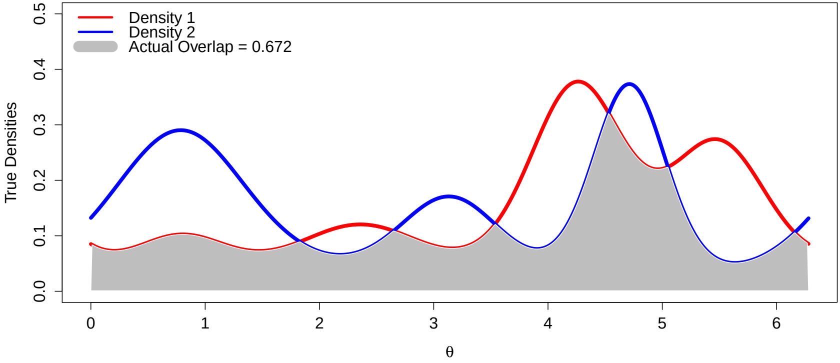
and

$$g(\psi) = 0.5\, \phi^{PN}(\psi \mid \boldsymbol{\mu}_{21}) + 0.2\, \phi^{PN}(\psi \mid \boldsymbol{\mu}_{22}) + 0.3\, \phi^{PN}(\psi \mid \boldsymbol{\mu}_{23}),$$

where $\boldsymbol{\mu}_{11} = (1.5, 1.5)^t$, $\boldsymbol{\mu}_{12} = (-1, 1)^t$, $\boldsymbol{\mu}_{13} = (-1, 2)^t$, $\boldsymbol{\mu}_{14} = (1.5, -1.5)^t$

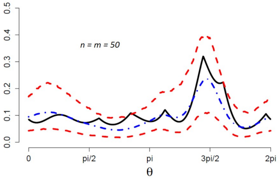
and $\boldsymbol{\mu}_{21} = (1, 1)^t$, $\boldsymbol{\mu}_{22} = (-2, 0)^t$, $\boldsymbol{\mu}_{23} = (0, -3)^t$.

Examples::Simulated data

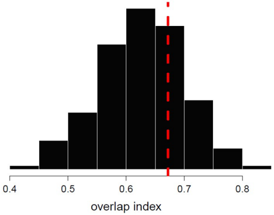


Analysis

- We simulated samples of sizes n and m, respectively, from these models.
- We then estimated each of the densities f and g, as well as the corresponding overlap coefficient $\Delta(f, g)$.
- As an example, for the case $n = m = 50$, here we show the estimation of $\min(f, g)$.



- Posterior distribution of the overlap coefficient (true value in red):

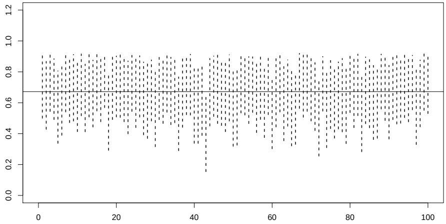


- The corresponding 95% credible interval is given by $(0.499, 0.765)$.

Sample size		Interval
f	g	
10	20	(0.534, 0.887)
10	50	(0.557, 0.832)
10	100	(0.566, 0.828)
20	50	(0.506, 0.800)
20	100	(0.602, 0.824)
50	100	(0.565, 0.783)

Table: 95% credible intervals for Δ (true value = 0.672).

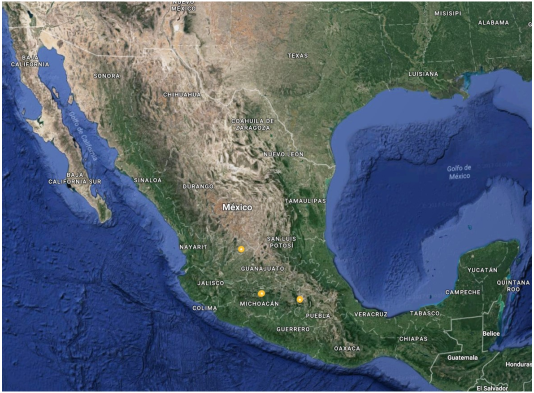
- 95% credible intervals for the overlap coefficient based on 100 pairs of samples ($n = m = 10$):



- The true value of the overlap coefficient is captured by 99 of the 100 intervals.

Real data

El Triunfo Biosphere Reserve in Chiapas, Mexico.



Real data

El Triunfo Biosphere Reserve in Chiapas, Mexico.

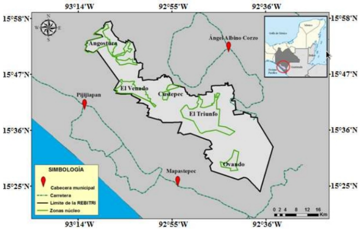
Real data

El Triunfo Biosphere Reserve in Chiapas, Mexico.

Real data

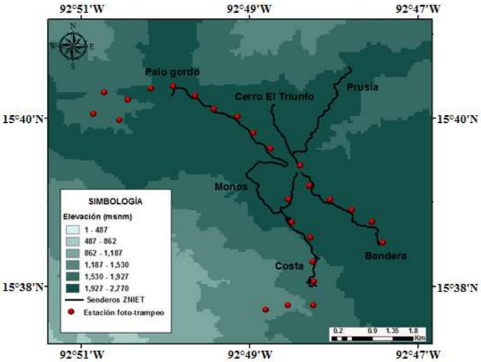
Data concerning three mammalian species (Red Brocket Deer, Baird's Tapir, Collared Pecari) inhabiting the 'El Triunfo' Biosphere Reserve in Chiapas, Mexico.





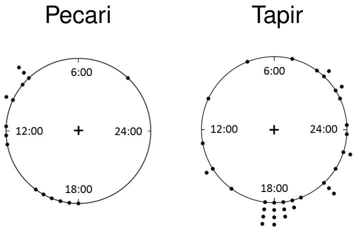
Data collection

Camera-trapping was used to obtain the data.

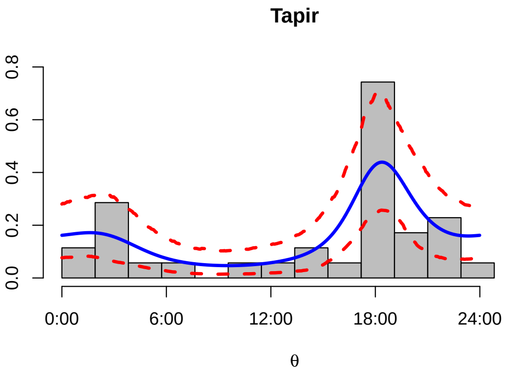


Analysis

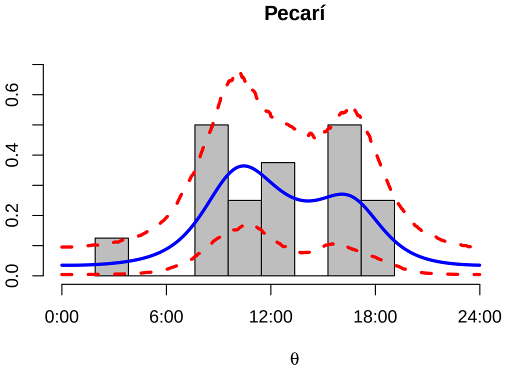
- For this illustration we use a sample of size $n = 16$ of pecari activity times and a sample of size $m = 35$ of tapir activity times.



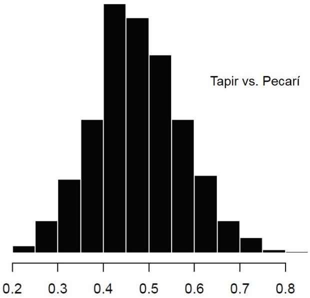
- Posterior inference for the distribution of tapir activity times



- Posterior inference for the distribution of pecari activity times



- Posterior distribution of the overlap coefficient:



- The corresponding 95% credible interval is given by $(0.328, 0.687)$.

Concluding remarks

- Our approach yields the full posterior distribution of the overlap coefficient.
- Moreover, our proposal allows us to make inferences about any other characteristic of the densities f and g.
- In particular, we can evaluate other measures of overlap.

- Comparisons with the results obtained with the R package **`overlap`** suggest that our procedure is more precise.

- Their (approximate) confidence intervals are usually larger than our probability intervals.

- They fail to capture the true value of Δ more often than ours.

- We are currently exploring the use of DPMs of von Mises distributions.

- Extensions to mixed multivariate settings (including both circular and linear variables) are also of interest.

References

- Escobar and West (1995). Bayesian density estimation using mixtures. *Journal of American Statistical Association*, **90**, 577–588.
- Geange, Pledger, Burns and Shima (2011). A unified analysis of niche overlap incorporating data of different types. *Methods in Ecology & Evolution*, **2**, 175–184.
- Hutchinson (1957), Concluding remarks. *Cold Spring Harbor Symposia on Quantitative Biology*, **22**, 415–427.
- Meredith and Ridout (2016). `overlap`: Estimates of coefficient of overlapping for animal activity patterns. R package version 0.2.6.
- Ridout and Linkie (2009). Estimating overlap of daily activity patterns from camera trap data. *Journal of Agricultural, Biological, and Environmental Statistics*, **14**, 322–337.
- Silverman (1986), *Density estimation for Statistics and data analysis*. Chapman and Hall, New York.