Inbar Maayan
OEB201 Final Project
November 29, 2017

R Script: <Maayan_script.R>
Data files: <BaseData.csv>, <AllPoints.csv>

Introduction and aim:

The overall aim of the project is to understand how the number and morphology of co-occurring species is related to differences in shape and size between males and females (sexual dimorphism, Fig. 1) across Caribbean anole species. The project was motivated by the observation that species living on depauperate islands tend to exhibit greater differences between males and females. This trend is thought to be the result of the dynamic relationship between intersexual niche partitioning within species and interspecific competition between species making up anole communities.

Figure 1 In many species of *Anolis* lizard, differences in size and shape exist between the male (R) and the female (L). These are the result of different selective forces acting on each sex. The ratio of male-to-female morphology is called sexual dimorphism.

The aim for the term project is to model – and develop a preliminary understanding of – the relationship between species number (community size) and morphology (degree of sexual dimorphism). To do this, I obtained morphological and distributional data on several anole species in Caribbean, built models to simulate fake data under a range of parameters, compared data simulated on the basis of my real data and the model to my real data, and evaluated the fit of my model. Briefly, I found that the simple model I constructed does not have much explanatory power, an observation likely driven by the weak relationship between the predictor I used and the actual data. Increasing the number of species and the number of sampling sites helped, but a number of other factors must be taken into consideration in future versions of the model.

Methods:

Data Acquisition

Morphological measurements of several characters (including toepad lamellae counts, linear measurements of appendicular and axial skeletal characters, head dimensions, and

eye diameter) in both males and females of 63 anole species were provided by L. Sims 2009). These data were cleaned and reduced to 60 species for which data were available for a minimum of five males and five females.

Sexual dimorphism (SD) was calculated two different ways: either by dividing mean male size by mean female size (mean-based), or by dividing maximum male size by maximum female size (max-based). The first method is common practice, and likely captures a value closer to the size distribution found in natural populations. However, since reptiles continue to grow (asymptotically) past maturity, using max values captures the potential size that younger specimens may have reached had they not been collected (it is difficult to gauge the exact age of an adult lizard).

The geographic distribution of each species was mapped using georeferenced museum specimens found on VertNet (www.vertnet.org). I used the R package rvertnet to query and download data on each of the 60 species, which I then merged into a single file (with uninformative columns subsequently removed). These data were plotted onto a map of the Caribbean using the R packages maps and mapdata, and inspected visually (and cross referenced with other sources) to delineate anole community composition. Community size was conservatively capped at 15 species, as 12 is the greatest number of species that have been recorded at one site.

Statistical Methods

While standard practice is to log-transform morphological variables (and this is recommended by Gelman), I chose not to do so with my real data so that it would be more intuitively interpretable. In later stages I will likely do this, especially as I expand to other morphological variables aside from body size, the variable used here.

To generate fake test data in order to test that the model I set up works, I designated a number of "true" parameters as the basis of the simulation. My model explores the relationship between sexual dimorphism (response) and community size (predictor) while accounting for each species' island of origin. As such, I stipulated the number of islands, each of which is home to between 1 and 15 different anole species. I included island of

origin as a random effect because I expect that islands differ from each other (in characteristics that I, admittedly, did not include in the model but would likely do in a more complex version); therefore, I chose a random intercept model, and simulated intercepts for each island. My simulation for island intercepts and for the sexual dimorphism values of lizard species found on the islands was based on specification of parameters μ_α, σ_α, β, and σ_y. These correspond to the mean and standard deviation of SD values, the relationship between SD and community size, and the error not explained by predictors in the model, respectively. To make SD more interpretable, since it is a ratio, I centered values on the mean I had specified prior. Because my real data includes only a subset of the species found on the larger islands, I also experimented with deleting a random subset of species to mirror this incompleteness. In my fake data, number of species on each island is equivalent to number of species in an island community, which will be broken down when I incorporate island size (and thus, number of communities per island) into the model at a later stage.

To quickly run models, I used lme4; I then ran a subset of the models in rstanarm using the stan_lmer model function. I examined my models critically using the shinystan interface in order to see how model simulations compared with each other, and later, how they compared with my real data.

Results and Findings

Mapping distributions of anole species showed that some species were very widespread and occurred together with many congeners, while others were restricted to only a small part of a large island (such as Cuba), or to just one small island (such as St. Lucia) (For maps, see supplement).

A strong relationship between species number and sexual dimorphism is clearly lacking, at least with the data at hand (Figure 2). The negative relationship found elsewhere is potentially the result of a few outliers with particularly high SD, as well as over-estimation of community size (Schoener 1977).
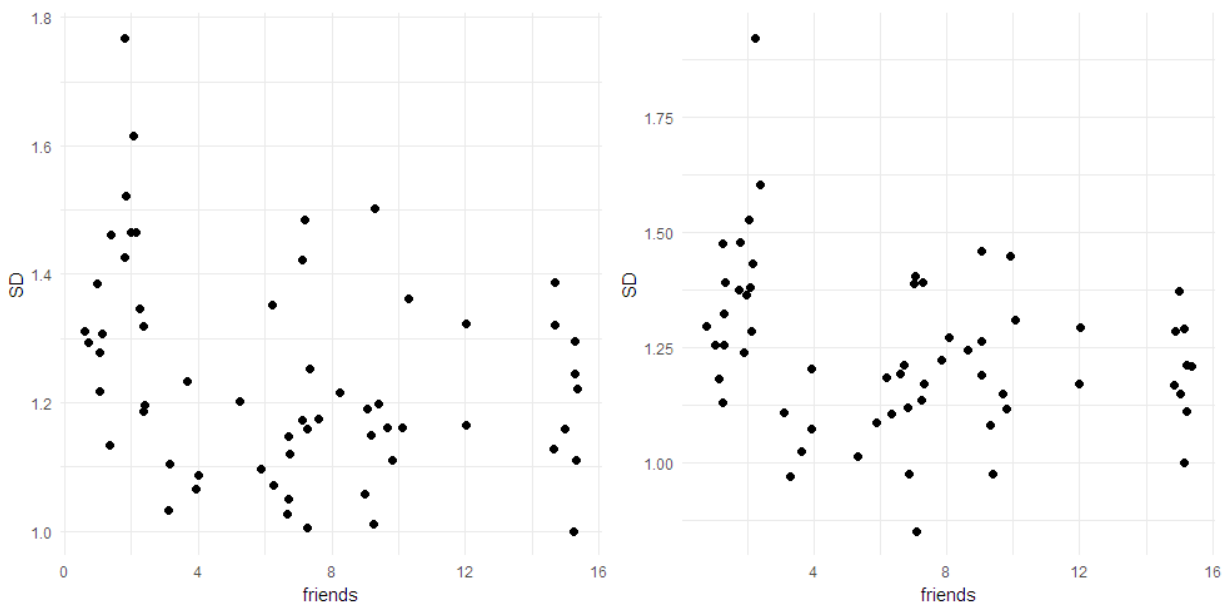
Figure 2 The relationship between number of congeners ("friends") and sexual dimorphism (Left, mean-based; Right, max-based) appears to follow a slight negative trend, with species occurring in smaller communities having greater SD; however, this relationship is clearly weak.

As Figure 3 shows, the model fit the max-based SD data more successfully (B), though for both real-data models many of the simulations deviated considerably; thus, the model could be improved. Notably, much of the variation found in the data could not be explained by the model. To see how increasing the number of islands and species would affect the fit of the model, I used fake data on the basis of each of these real-data sets (mean and max-based) (C) and (E). These models produced very similar outcomes to each other, and more of the simulations agreed with each other, so increasing the amount of input data likely helped.
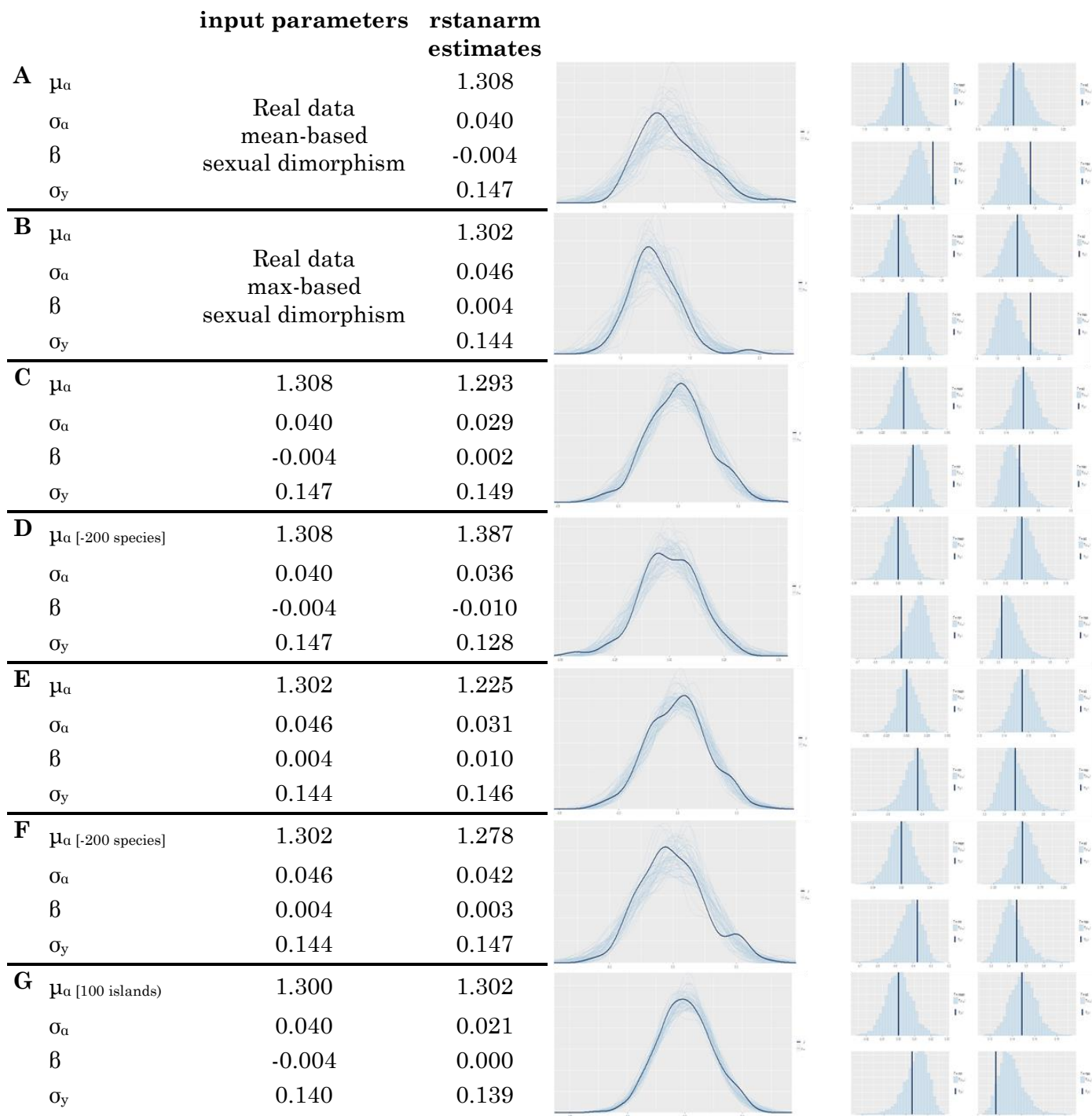
		input parameters	rstanarm estimates	
A	μ_α		1.308	
	σ_α	Real data mean-based sexual dimorphism	0.040	
	β		-0.004	
	σ_y		0.147	
B	μ_α		1.302	
	σ_α	Real data max-based sexual dimorphism	0.046	
	β		0.004	
	σ_y		0.144	
C	μ_α	1.308	1.293	
	σ_α	0.040	0.029	
	β	-0.004	0.002	
	σ_y	0.147	0.149	
D	μ_α [-200 species]	1.308	1.387	
	σ_α	0.040	0.036	
	β	-0.004	-0.010	
	σ_y	0.147	0.128	
E	μ_α	1.302	1.225	
	σ_α	0.046	0.031	
	β	0.004	0.010	
	σ_y	0.144	0.146	
F	μ_α [-200 species]	1.302	1.278	
	σ_α	0.046	0.042	
	β	0.004	0.003	
	σ_y	0.144	0.147	
G	μ_α [100 islands)	1.300	1.302	
	σ_α	0.040	0.021	
	β	-0.004	0.000	
	σ_y	0.140	0.139	

Figure 3 Estimated parameters from real and fake data models, with distributions generated by posterior predictive checks in shinystan. Clear variation exists in the success of the models under different parameter values.

Since my real data are incomplete, I also tested the effect of excluding a subset of the species from the analysis (3D) and (3F). Doing so caused a greater mismatch between the distribution and the simulations, which provides insight into the effect of having missing species. When I increased the number of islands from 50 to 100 (3G), fit of the simulated distributions improved greatly, indicating that sampling across more communities would likely improve my estimates; however, this also included more species total, so it is possible that both changes led to this shift.

In future iterations of the model, it is clear that more factors should be included. I would especially like to look at how including phylogenic relationships among the species will impact the result of the model – if at all. If sexual dimorphism is more affected by environment than by phylogenetic inertia, then it is reasonable to expect that environmental variables would explain the patterns among species in a better way. As such, I would also like to examine how including island size as well as climactic and environmental variables impacts the explanatory power of the model. It will also be interesting to examine traits aside from body size, as some of these might show a clearer signal with regards to one or more potential predictors. At a later stage, it will likewise be important to differentiate between populations within species, since some populations (e.g. *Anolis sagrei* in the Bahamas) have only a few congeners, while other populations (e.g. *Anolis sagrei* in Cuba) occur in very different habitats and come into contact with a variety of other species.

References

- Malhotra, Anita; Thorpe, Roger S. (1999), Reptiles & Amphibians of the Eastern Caribbean, Macmillan Education.
- Schoener, Tom W. (1978) "Competition and the Niche" in Gans and Tinkle, eds. *Biology of the Reptilia. Volume 7. Ecology and Behaviour A.* Academic Press, London, New York, and San Francisco.
- Sims, Lela M. (2009) "Effects of Resource Competition on Sexual Dimorphism of Size and Shape in Caribbean *Anolis* Lizards". Undergraduate thesis, Harvard University.