

Midterm problem

Understanding how organisms use and compete for resources is a central question in ecology. Space is a limiting resource for sessile organisms, including terrestrial plants as well as marine algae and invertebrates. In temperate subtidal ecosystems, sessile taxa recruit to and then spread on bare rock and encrusting algae - the sum of the latter two are defined as available space (Elahi and Sebens 2012). Space is renewed when sessile taxa senesce and die, or when they are removed by mobile grazers.

Here we use an observational dataset to answer the following question - **what are the direct and indirect causes of available space in rocky subtidal habitats?** More specifically, we can separate this into two questions, and I list the relevant data in our case study:

- Does sessile biodiversity affect resource use?
 - resource is available space (**space**)
 - richness of sessile taxa (**richness**)
 - Simpson's diversity of sessile taxa (**simpsondiv**)
 - all three metrics from quadrats sampled in July 2008
- Do grazers affect resource use?
 - mean number of urchins per transect (**urchin_mean**)
 - mean number of chitons per quadrat (**chiton_mean**)
 - both metrics are the average over 3 sampling dates (December 2007, March 2008, July 2008)

The following DAG represents the hypothesized causal relationships between the predictors, with coefficients estimated using a structural equation model. Use this to frame your interpretation of the results.

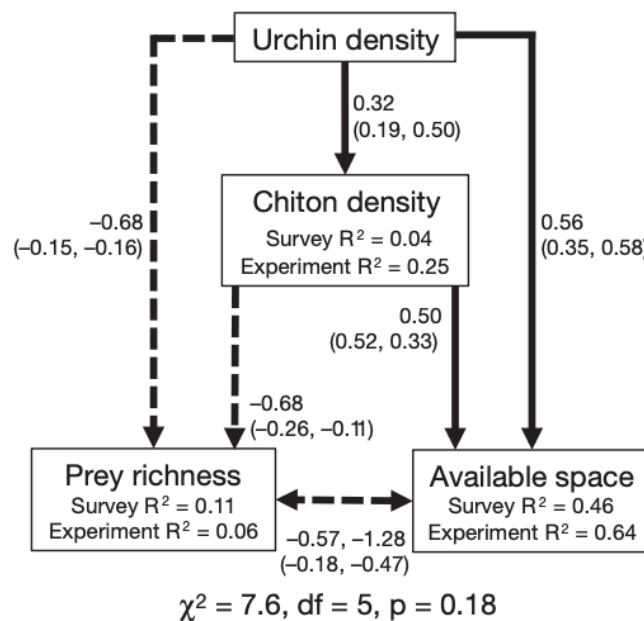


Figure 1: Path diagram from Elahi and Sebens 2012 (Fig. 6)

(consider prey richness to cause available space, rather than a bidirectional arrow for the purposes of this exercise)

Your challenge is to recreate the following table using **Stan** and **rethinking**:

Table 1. Linear mixed-effects models used to predict percent cover of available space (logit-transformed) in permanent quadrats surveyed in 2008. Transect was nested within site; both were treated as random effects. Candidate models are listed with the number of parameters (K), corrected AIC (AIC_c), the difference in AIC between the candidate model and the best model (Δi), and Akaike weights (w_i). The best model, including only the densities of lined chitons *Tonicella* spp. and red urchins *Strongylocentrotus franciscanus*, is shown in **bold**

Model	K	AIC_c	Δi	w_i
Prey richness (PR)	5	247.66	35.03	0.0000
Grazer richness (CR)	5	234.93	22.30	0.0000
Chiton density (CD)	5	215.50	2.88	0.1772
Urchin density (UD)	5	235.27	22.65	0.0000
PR + CR	6	237.45	24.82	0.0000
PR + CD	6	221.94	9.31	0.0071
PR + UD	6	238.48	25.85	0.0000
CR + CD	6	217.52	4.89	0.0647
CR + UD	6	231.60	18.98	0.0001
CD + UD	6	212.63	0	0.7469
PR + CR + CD + UD	8	223.08	10.45	0.0040

Figure 2: Table 1 from Elahi and Sebens 2012

(you can ignore grazer richness (CR); that column is missing in the data for some reason ... which will reduce your work considerably anyway)

Interpret your results in the context of the DAG above, and highlight differences between your WAIC table and the published one.

The data

Load the data:

```
library(rethinking)
library(tidyverse)
# Set the web address where R will look for files from this repository
repo_url <- "https://raw.githubusercontent.com/elahi/elahi.github.io/master/"
dat <- read.csv(paste(repo_url, "data/meps_2012_survey_data.csv", sep = ""))
glimpse(dat)
```

Rows: 72

Columns: 9

```
$ X          <int> 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17,~
$ site       <chr> "ON", "ON", "ON", "ON", "ON", "ON", "ON", "ON", "ON", "ON"~
$ transect   <chr> "ON0", "ON0", "ON0", "ON0", "ON1", "ON1", "ON1", "ON1", "O~
```

```

$ quadrat      <chr> "ONO_45", "ONO_46", "ONO_48", "ONO_49", "ON1_17", "ON1_84"~
$ chiton_mean  <dbl> 0.0000000, 1.0000000, 0.0000000, 0.0000000, 1.0000000, 0.0~
$ space        <dbl> 0.14213721, 0.22878590, 0.12508222, 0.24288815, 0.73034557~
$ richness     <int> 10, 12, 8, 10, 10, 9, 9, 7, 8, 8, 9, 8, 10, 13, 10, 15, 13~
$ simpsondiv   <dbl> 0.7004303, 0.7497803, 0.6160904, 0.7580468, 0.7172267, 0.7~
$ urchin_mean  <dbl> 0.6666667, 0.6666667, 0.6666667, 0.6666667, 1.0000000, 1.0~

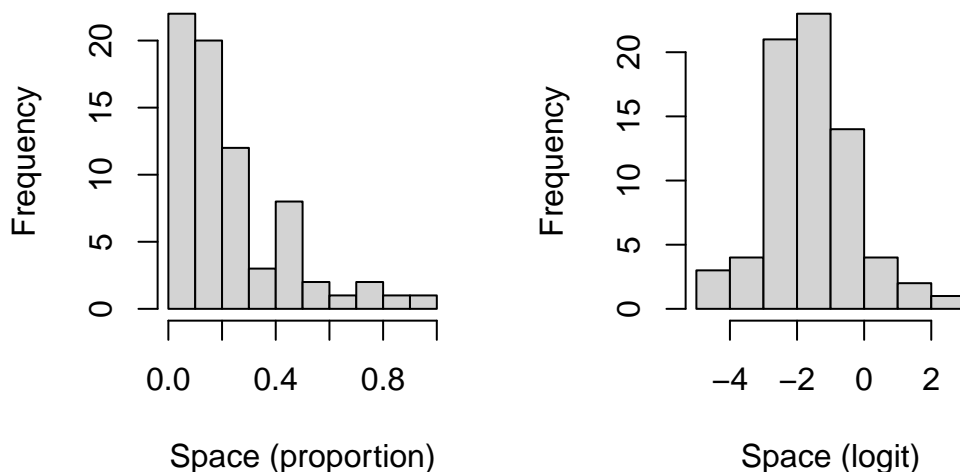
```

Transform the proportional cover of space so that we can model it using a normal distribution:

```

# Define logit function
logit <- function(p) log(p / (1 - p))
dat <- dat %>%
  mutate(space_logit = logit(space))
# Plot
par(mfrow = c(1,2))
hist(dat$space, main = "", xlab = "Space (proportion)")
hist(dat$space_logit, main = "", xlab = "Space (logit)")

```



Importantly, the response is no longer bounded between 0 and 1. It is *not* important that $\text{logit}(\text{space})$ appears normally distributed! It just so happens that this is the case here. For now, ignore the effect of site and transect.

Suggested workflow

You will skip the generative simulation for this exercise. Here is a suggested workflow:

1. Visualize the data to make sense of the patterns
2. Standardize the predictors
3. Come up with sensible priors by plotting prior predictive distributions (see HW4-key or SR2 Fig 5.8 for example). Justify your choice of priors.

4. Run the models using `ulam`. Diagnose your chains - what are the criteria you checked to ensure convergence? See SR2 9.4-9.5 for help.
5. Compare WAIC for the models.
6. Interpret in the context of the DAG and the published results.
7. Plot the counterfactual effect of one predictor, holding any other predictors constant. Justify your choice of the model used for these posterior predictions. See SR2 Fig. 5.9 for example of a counterfactual plot from a multiple regression. Interpret.

References

- Elahi, Robin, and Kenneth P Sebens. 2012. “Consumers Mediate Natural Variation Between Prey Richness and Resource Use in a Benthic Marine Community.” *Marine Ecology Progress Series* 452: 131–43.