

Activity 02 – Binomial Distribution

September 13, 2022

0.1 Activity 02

In this activity you will continue practicing with the ideas of inference, simulations, posteriors and Bayesian best practices.

Today, you will work with galaxies. More specifically, with a particular class of objects, called [Green Peas](#). Green Peas are small galaxies (1 to 10% the mass of the Milky Way), but are forming stars at a much faster rate compared to our own galaxy. It is not clear why Green Peas are forming stars at such a fast pace. One possibility is that the formation of new stars is triggered by interactions with nearby companions.

To test this hypothesis, Laufman et al. (2022) used new and archival data to search for companions around a sample of 23 Green Pea galaxies, and around a sample of 43 normal galaxies, matched in stellar mass and distance to the sample of Green Peas (but with normal star-formation rates). The data were obtained with the [MUSE spectrograph](#) on the [Very large Telescope](#) in Chile.

The results of the analysis are collected in the file `Laufman_Table2.dat`, which includes the following columns:

- Galaxy : object identifier
- has_companion : a flag that identify whether or not the galaxy has a companion (1= has a companion, 0 =does not have a companion)
- GP : Is it a Green Pea or a Normal Galaxy ? 1=GP 2= Normal Galaxy

You will use the data to estimate the probability that a galaxy has a companion. You will make this estimate for both samples, the Green Peas and the normal galaxies.

1 Perform a visual inspection of the data, for both samples. What kind of data do you have? How do they differ from the data you had last week?

2 Build a statistical model as discussed in lecture. Clearly explain it. Including a motivation for the choice of the probability of observing the data, given the unknown parameter, and the probability of the parameter. Clearly explain how you compute the constants for the prior.

3 Write a code to compute the model. As usual, follow best practices when writing your code [you will share the code with us]. Make sure the code is well commented and can easily be read by humans. Check that the code's paths are not specific to your machine. For reproducibility, set the seed for the pseudo-random number generator explicitly.

4 Plot the posterior distribution functions, and compute the position of the posterior's maximum and its 95% credible interval.

5 Perform and describe a sensitivity analysis (i.e., discuss how the choice of the prior influences the result).

6 Using the analysis you performed above, would you be comfortable stating “The Green Pea galaxies are clearly different from the normal galaxy population”? Why yes or why not?

[]: