Stat4DS / Homework 03 - Part B

Pierpaolo Brutti

Due Thursday, January 31, 2019, 23:00 PM on Moodle

General Instructions

I expect you to upload your solutions on Moodle as:

- a **single**, **running** `R Markdown` file (`.rmd`) + its `html` output, named with your surnames.
- an `.odc` file containing the requested OpenBUGS Doodle.

Please Notice

- Remember our **policy on collaboration**: *Collaboration on homework assignments with fellow students is **encouraged**. However, such collaboration should be clearly acknowledged, by listing the names of the students with whom you have had discussions concerning your solution. You may **not**, however, share written work or code after discussing a problem with others. The solutions should be written by **you**.*

Exercise: Deep-Sky-Bayes

For some strange turn in the history of science, (many) astrophysicists are real Bayesian "evangelists" that use, misuse and push the old Reverend into their data analytic games as soon as they possibly can. This little_tiny_exercise inspired by a not so old paper, is just one example of what can be done.

We all know what a star is, and we all know that they form and evolve in decently well-understood way. The Hertzsprung–Russell diagram where we graph each star in terms of its brightness against its temperature (color), was one of the tools that led astronomers to speculate about stellar evolution: stars, while fusing hydrogen in their cores, collapse from red giants to dwarf, and then, in the course of their lifetimes, move down along the so called main sequence, a prominent diagonal band that runs from the upper left to the lower right of the H-R diagram.

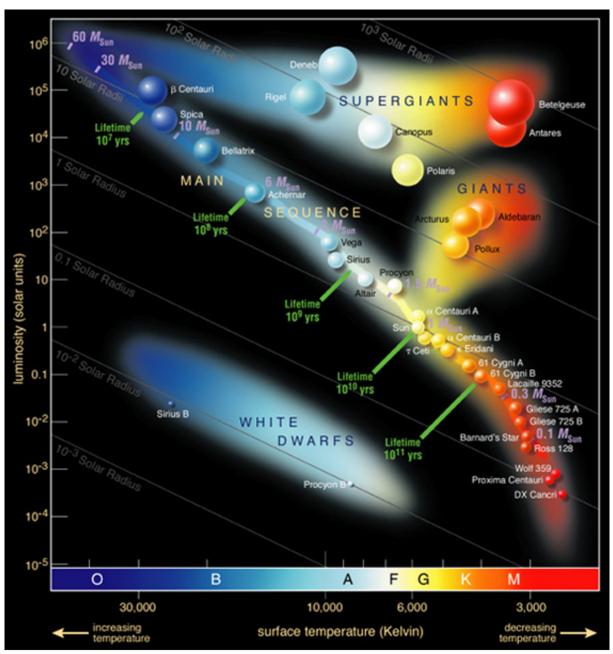

Figure 1: A Hertzsprung–Russell diagram.

Now, the problem we are tackling here is an old one, strictly related to these issues: determining the stellar initial mass function (IMF). The IMF plays an extremely crucial role in shaping stellar population: it determines the probability distribution of the mass at which a star enters the main sequence; it regulates the relative abundance of massive versus low-mass stars for each stellar generation; it influences most observable properties of stellar population and galaxies, and it is also required as input in some models of galaxy evolution.

For an astrophysicist the ultimate goal would clearly be to infer the IMF from first principles, but here we content ourselves with an empirical, data driven alternative: fit the IMF using a lognormal distribution as suggested by Zaninetti but in a Bayesian framework using OpenBUGS/JAGS.

The Data

The data available are from 208 stars located in the massive cluster NGC 6611 for which mass measurements are available (see Oliveira et al.).

Figure 2: The Eagle Nebula.

Now what?

Consider a lognormal model with location and scale parameters μ and σ for the stellar mass.

0. Explore the dataset with some standard plot (e.g. histogram, density plot, boxplot, etc) and summary statistics (location, variability, skewness and kurtosis). Briefly comment on the shape of the mass distribution.
1. Assuming non-informative[1] Gaussian and Gamma priors for the location and scale parameter (μ, σ), build the model as a Doodle in OpenBUGS and then save it as an `.odc` file that you will then upload on Moodle.
2. With this model at hand, initilize and then run 3 chains + a reasonable burn-in, to make inference on μ and σ. Report the relevant point estimates and credible intervals, qualitatively commenting on the mixing of the chain(s) by looking at the trace plots + autocorrelations (if you want, you can go deeper…)
3. Now repeat the analysis in JAGS_within_R writing down *explicitly* the model in the BUGS language (it should not be difficult, just a few lines). Conclude by overlaying to the basic histogram/density plot the "best" Bayesian counterpart(s) of your choice.
4. Perform a quick sensitivity analysis by slightly changing the priors and reporting on the impact you see on your inferential conclusions.

---

[1] Choosing suitable hyperparameters for the priors is up to you.