

Activity 01 – Normal Distribution

September 13, 2022

0.1 Activity 01

This activity is meant for you to start practicing with the ideas of inference, simulations, posteriors and Bayesian best practices.

The Panoramic Survey Telescope & Rapid Response System ([Pan-STARRS](#)) is a wide-field imaging telescope that is observing the sky repeatedly to conduct a number of Galactic and Extragalactic studies. The data are released periodically, and the catalogs includes measurements in five filters ([grizy](#)) covering 30,000 square degrees of the sky, with typically ~12 epochs for each filter.

The archive [interface](#) allows to search the second data release (DR2) and download all the multi-epoch observations for any object covered by the Survey.

In this activity, we will accurately estimate the flux of the standard star BD+28 4211. Standard stars are objects whose flux does not change with time (over sufficiently long timescales) and knowing their precise flux is extremely important in astronomy as it is used to remove instrumental effects from astronomical observations.

The file BD+284211.csv includes repeated flux measurements for BD+28 4211 in the i filter, as a function of time. The columns are as follows:

- objID : object identifier
- filterID : filter used for the observations
- obsTime : time of the observation (Modified Julian Date at the midpoint of the observation) in days
- ra : Right Ascension of the object
- dec : Declination of the object (RA and Dec are analog of Latitude and longitude)
- psfFlux : flux of the star in Jy
- psfFluxErr : measurement error
- infoFlag2 : flag on the measurement [0 = good measurement, 32 = fitted with 2 PSF, problematic]

You will use these data to estimate the flux of BD+28 4211, and report the result.

1 Perform a visual inspection of the data. Do you see anything unusual?

Have the fluxes remained constant during the observation period?

Anything special about the fluxes that have flags different from 0?

Plot the distribution of fluxes.

2 Clearly explain the statistical model you use, including the likelihood function and the prior that you chose. Clearly explain how you compute the constants for the prior.

3 Write a code to compute the model. 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 to 5731.

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

5 Draw a random sample of size $N=100$ from the posterior distribution. Plot the distribution of this sample and compare it with the analytical expression. Compute the arithmetic mean and the difference between this and the mean of the posterior. Are they the same?

Now consider 5 increasing values of N , from 10^2 to 10^5 , linear in log step. For each sample compute the difference between the arithmetic mean and posterior mean. Comment on the trend between the difference and N .

6 Describe and justify any decisions you made. [E.g., what do you do with the flagged measurements?]

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

[]: