MASTER IN ASTRONOMY/PHYSICS - 2020/2021

Data Analysis in Astronomy/Physics

Bayesian Statistical Inference

Previous information: Spectral lines in discrete absorption or emission spectra are never strictly monochromatic. Even at very high resolution, one observes a spectral distribution $I(\nu)$ of the absorbed or emitted intensity around the central frequency $\nu_0 = (E_i - E_k)/h$ corresponding to an atomic or molecular transition with energy difference $\Delta E = E_i - E_k$ between upper and lower levels. The function $I(\nu)$ in the vicinity of ν_0 is called the line profile. There are several causes for the broadening of spectral lines, some internal to the atom or molecule, others external, and each produces its characteristic profile. Some types of profile, for example, have a broad core and small wings; others have a narrow core and extensive, broad wings. While both natural broadening, due to radiation damping, and pressure broadening give rise to lorentzian line profiles, thermal broadening produces (normalised) gaussian line profiles

$$G(\nu) = 1 \pm \frac{1}{g}\sqrt{\frac{\ln 2}{\pi}}\, e^{-\left[\frac{\ln 2 (\nu-\nu_0)^2}{g^2}\right]}$$

where the full width at half maximum (FWHM) is equal to $2g$ and proportional to the square root of the temperature of the material responsible for the thermal broadening. The continuum has been normalized to one, and an emission line corresponds to an addiction to the continuum, while an absorption line corresponds to a subtraction to the continuum. Assuming each broadening mechanism acts independently from the others, the observed line profile will be a convolution of the line profiles associated with each mechanism, yielding a Voigt profile. However, in what follows, we will assume that thermal broadening clearly dominates over all other line broadening mechanisms, as a consequence of prior information that leads us to believe that the object observed has a very high temperature.

Question A: Perform the necessary calculations to determine the probability of the existence of an **emission** spectral line with a gaussian profile and $g = 2$, taking into account the data. This consists of one dataset containing 101 independent (normalized) flux measurements. They can be found in the ASCII tables provided. The columns (from left to right) have the following meaning: (1) frequency, ν (in MHz); (2) the most probable value for the amplitude of the line profile as a function of ν, defined as $G_\mu(\nu)$, after continuum subtraction. The measurement uncertainty about $G_\mu(\nu)$ should be assumed to be well described by a Normal probability distribution

$$P[G(\nu)] = \frac{1}{\sigma\sqrt{2\pi}}\, e^{-\left\{\frac{\left[G(\nu)-G_\mu(\nu)\right]^2}{2\sigma^2}\right\}}$$

where σ is equal to 0.05 for all measurements.

Question B: Assume now *a priori* that an **emission** spectral line with a gaussian profile and $g = 2$ exists. Determine the posterior probability distribution for ν_0 and make a plot of it. Determine also its mode, median, mean and the smallest interval in ν_0 that contains 95% of the posterior probability. Construct a plot that allows the visualization of the values observed for $G(\nu)$, including the uncertainty associated with each point, as a function of the frequency, ν, as well as the theoretical line profile, assuming that ν_0 is equal to the mode previously obtained, and to the lower and upper limits of the smallest interval in ν_0 that contains 95% of the posterior probability.

Question C: Perform again the necessary calculations to determine the probability of the existence of an **emission** spectral line with a gaussian profile, taking into account the data, but using solely the output of Monte Carlo Markov Chains and assuming the value of g to be unknown *a priori*. Then, assuming *a priori* that such a spectral line exists, determine and plot the (marginal) posterior probability distribution for each of the parameters ν_0 and g. Characterize them by determining their mode, mean, and the smallest interval in ν_0 and g that contains 95% of the associated (marginal) posterior probability. Finally, plot their joint posterior probability distribution.

.

The **results of this project** should be **delivered** as a ***R notebook*** or ***Jupyter notebook*** **until the 16th of May**.

Note: You should implement in **R** any algorithms used to solve the questions proposed. There are several **R** *packages* that contain algorithms for the generation of Monte Carlo Markov Chains. Two simple such *packages* are

mcmc (http://cran.r-project.org/web/packages/mcmc/)
adaptMCMC (http://cran.r-project.org/web/packages/adaptMCMC/)

while

BayesianTools (http://cran.r-project.org/web/packages/BayesianTools/)
LaplacesDemon (http://cran.r-project.org/web/packages/LaplacesDemon/)

offer more complete inference frameworks. You can use one of such *packages* or develop your own algorithm for generating Monte Carlo Markov Chains. The installation of the **R** *package*

pracma (http://cran.r-project.org/web/packages/pracma/)

may also allow for a faster and more precise estimation of integrals when needed.