ICIC DATA ANALYSIS WORKSHOP

PRE-WORKSHOP EXERCISE -

ALAN HEAVENS

1. Introduction

The purpose of this preparatory exercise is to ensure you are are set up on your laptop to do a few of the basic things which will be needed for the hands-on exercises in the workshop, especially within the colab environment, which we encourage you to do. Separate links to colab tutorials will be provided for those who are unfamiliar with it. You may also use your local machine if you prefer.

You need to have access to some programming language (of your choice) and be comfortable using it. We assume most students will use python on colab. You will need to do some programming, read in data files, plot data, plot functions, and draw some random numbers from uniform and gaussian distributions; if you don't do these, you may waste time in the hands-on part of the workshop installing software, remembering how to run codes etc. Although this preliminary exercise is connected with one of the three hands-on projects, the skills needed will be useful for all three.

2. Background: cosmology with Supernovae Ia

This exercise is based around the relationship between the apparent brightness and distance of objects of a standard brightness (so-called 'standard candles'). It is useful as the theoretical relation depends on the amount of Dark Matter, Dark Energy and on the curvature of the Universe. Type I supernova are thought to arise from collapse of white dwarf stars when the Chandrasekhar limit is reached, and are thus[1] standard candles, and the resulting "Supernova Hubble Diagram" was famously used to infer the acceleration of the Universe, culminating in the award of the Nobel prize to Schmidt, Perlmutter and Riess. We will build on this example in one of the hands-on exercises.

3. Data vs. Theory

Here is some minimal information. We have some data, and a theoretical model, which has some parameters. A very typical problem is to infer the model parameters from the data. Here you will read in data, plot it, overplot some theoretical curves (whose formula is slightly involved), and draw some simulated data at random.

Data:

We have a dataset of measurements of the 'distance modulus' (a logarithmic measure of distance) μ at various different redshifts z. So the data consists of pairs of values, $z_i, \mu_i; i = 1, \ldots 31$. You will read these in and plot them later.

[1]sweeping many complications under the carpet.

Theory:

At this stage, you do not need to understand these formulae, but they given theoretical curves for μ as a function of z, for a model with two parameters in it, h (Hubble constant, in units of 100 km/s/Mpc) and $\Omega_{\rm m}$ (matter density parameter). The equation is a little involved, and requires 4 lines of code to compute it, but you don't need to understand it, just code it. The so-called Distance Modulus μ (unit: magnitudes) at redshift z is

$$\mu(z) = 25 - 5\log_{10} h + 5\log_{10}\left(D_L^*(z)\right) \tag{1}$$

where

$$D_L^*(z) = 3000(1+z)\left[\eta(1,\Omega_{\rm m}) - \eta\left(\frac{1}{1+z},\Omega_{\rm m}\right)\right] \tag{2}$$

and (approximately)

$$\eta(a,\Omega_{\rm m}) = 2\sqrt{s^3+1}\left[\frac{1}{a^4} - 0.1540\frac{s}{a^3} + 0.4304\frac{s^2}{a^2} + 0.19097\frac{s^3}{a} + 0.066941 s^4\right]^{-1/8} \tag{3}$$

and

$$s^3 \equiv (1-\Omega_{\rm m})/\Omega_{\rm m}. \tag{4}$$

If we have measurements of μ_i, then in the course we will see how we can use Bayesian arguments to infer the parameters h and $\Omega_{\rm m}$. For anyone unfamiliar with cosmology, these numbers are somewhere between 0 and 1.

4. Data

The data file (from the JLA supernova sample:
`http://supernovae.in2p3.fr/sdss_snls_jla/ReadMe.html#sec-1`)
consists of 31 pairs of redshift z and distance modulus μ. The supplied ASCII file (`jla_mub.txt`) contains the following:

z	μ
0.010	32.953886976
0.012	33.8790034661
...	

5. Exercises

(1) Plot the theoretical distance modulus formula against redshift, for parameter values $\Omega_{\rm m} = 0.2$, 0.3, 0.4, 0.5 for the range $0 < z < 2$. Set $h = 0.7$.

(2) Plot the observational data in the file (`jla_mub.txt`) on the graph.

(3) Now generate some simulated data. Make a random set of 20 supernovae with redshifts drawn uniformly from $0 < z < 2$. Assuming that $\Omega_{\rm m} = 0.3$ and $h = 0.7$, compute μ for each of them.

(4) Find a Gaussian random number generator routine, and add an error to μ for each supernova, drawn from a Gaussian error distribution with mean zero and r.m.s. 0.1 (magnitudes).

(5) Plot these data on a graph, with error bars, and overplot theory curves for $\Omega_m = 0.3$ and $h = 0.6, 0.7$ and 0.8). They should scatter around the correct middle curve.