ICIC DATA ANALYSIS WORKSHOP 2016: PRE-WORKSHOP EXERCISES

ALAN HEAVENS, ANDREW JAFFE, DANIEL MORTLOCK, JONATHAN PRITCHARD, ELENA SELLENTIN

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. Basically this means you have some programming language installed and you are comfortable using it (it's your choice what you use; if you want to try R, which is used extensively in the statistics community, download from http://www.r-project.org/). You will need to do some programming, read in data files, plot data, and draw some random numbers from uniform and gaussian distributions; if you don't do these, you won't be able to take advantage of the hands-on part of the workshop (or you will be spending your time in the workshop catching up). The tasks themselves are pretty straightforward - the theory section is here for completeness, but actually all that is needed from this section is to be able to program a couple of the formulae (specifically equations $5-7$).

2. Background

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 the hands-on exercises.

3. Cosmology with Supernovae Ia: theory

The flux from a supernova of luminosity L is given by

$$(1) \qquad f = \frac{L}{4\pi D_L^2}$$

[1]sweeping many complications under the carpet.

where D_L is the *Luminosity Distance*. In Big Bang cosmology it is given as a function of the supernova redshift z by

$$(2) \qquad D_L = \frac{(1+z)c}{H_0\sqrt{|1-\Omega|}} S_k(r),$$

where

$$(3) \qquad r(z) = \sqrt{|1-\Omega|} \int_0^z \frac{dz'}{\sqrt{\Omega_{\rm m}(1+z')^3 + \Omega_{\rm v} + (1-\Omega)(1+z')^2}}.$$

and $S_k(r) = \sin r, r, \sinh r$, depending on whether $\Omega \equiv \Omega_{\rm m} + \Omega_{\rm v}$ is $> 1, = 1$, or < 1, and z is the observed redshift of the supernova[2]. $\Omega_{\rm m}, \Omega_{\rm v}$ and H_0 are the density parameters (today) in matter, vacuum energy, and the Hubble constant. It is beyond the scope of these notes to derive this, but it is standard material for an undergraduate cosmology course.

For a flat Universe ($\Omega = 1$), this simplifies to

$$(4) \qquad D_L(z) = 3000h^{-1}(1+z) \int_0^z \frac{dz'}{\sqrt{\Omega_{\rm m}(1+z')^3 + 1 - \Omega_{\rm m}}} \text{ Mpc},$$

where $H_0 = 100h\,{\rm km}\ {\rm s}^{-1}\ {\rm Mpc}^{-1}$. To avoid evaluating integrals to calculate D_L, we can use an accurate fitting formula (valid for flat universes only), given by U.-L. Pen, ApJS, 120:4950, 1999:

$$(5) \qquad D_L(z) = \frac{c}{H_0}(1+z) \left[\eta(1, \Omega_{\rm m}) - \eta\left(\frac{1}{1+z}, \Omega_{\rm m}\right) \right]$$

where

$$(6) \qquad \eta(a, \Omega_{\rm m}) = 2\sqrt{s^3+1} \left[\frac{1}{a^4} - 0.1540\frac{s}{a^3} + 0.4304\frac{s^2}{a^2} + 0.19097\frac{s^3}{a} + 0.066941s^4 \right]^{-1/8}$$

and $s^3 \equiv (1-\Omega_{\rm m})/\Omega_{\rm m}$. This is claimed to be accurate to better than 0.4% for $0.2 \le \Omega_{\rm m} \le 1$.

Fluxes are usually expressed in magnitudes, where $m = -2.5\log_{10} F+$constant. The *distance modulus* is $\mu = m - M$, where M is the absolute magnitude[3], which is the value of m if the source is at a distance 10pc:

$$(7) \qquad \mu = 25 - 5\log_{10} h + 5\log_{10}\left(\frac{D_L^*}{\rm Mpc}\right)$$

The Hubble constant has been factored out of D_L: $D_L^* \equiv D_L(h = 1)$.

If we have measurements of μ, then we can use Bayesian arguments to estimate the parameters $\Omega_{\rm m}, \Omega_{\rm v}, h$. For anyone unfamiliar with cosmology, these numbers are somewhere between 0 and 1.

---

[2]Note that for $\Omega = 1$, the $\sqrt{|1-\Omega|}$ terms in D_L and $r(z)$ cancel.

[3]There is a simplification in the exercise here; we assume we know what the absolute magnitude (or luminosity) of type I supernovae are, but in fact unless we have supernovae with known distances, we don't. M and h are degenerate, since M is set from low-redshift supernovae where we assume Hubble's law to give us the distance. For the purpose of this exercise, we will cheat.

4. Data

The data file (from the JLA sample:
`http://supernovae.in2p3.fr/sdss_snls_jla/ReadMe.html#sec-1`)
consists of measurements of the redshift (assumed precisely accurate), and an average distance modulus μ_i, for a collection of 740 supernovæ binned into 31 redshift bins. The sample ASCII file (`jla_mub.txt`) contains the following:

z	μ
0.010	32.953886976
0.012	33.8790034661
0.014	33.8421407403

. . .

The errors are given as a 31×31 covariance matrix, supplied as a single list of numbers in the ASCII file `jla_mub_covmatrix.txt`.

5. Exercises

Using the analytic approximation for luminosity distance in flat universes given above, do the following:

(1) Plot distance modulus vs redshift for $\Omega_{\rm m} = 0.2$, 0.3, 0.4, 0.5 for the range $0 < z < 2$, for $H_0 = 70\,{\rm km\,s^{-1}\,Mpc^{-1}}$.
(2) Plot the observational data in the file (`jla_mub.txt` from the workshop website) on the graph.
(3) Generate a random set of 20 supernovae with redshifts drawn uniformly from $0 < z < 2$ and with gaussian errors in their apparent magnitudes with r.m.s. 0.1 mag, for $\Omega_{\rm m} = 0.3$ in a flat Universe.
(4) Plot these data on the graph, with error bars.
(5) (Optional). Write a routine which computes D_L^* from the integral form for general (non-flat) universes as well, tabulates it as a function of $\Omega_{\rm m}$, $\Omega_{\rm v}$, and z, and stores it in a file. Write a routine to read the table in, and write an interpolating routine to compute it quickly (or alternatively compute it with fine spacings (e.g. 0.001) so that going to the nearest tabulated point is very accurate).