Computational Statistics

Exercise sheet 8

To be handed in for marking until tutorial of 21/06/2017

Nelson Lima and Luca Amendola, ITP, Heidelberg

21/06/2017

www.thphys.uni-heidelberg.de/~amendola/teaching.html

Problem 1 - Maximum Likelihood analysis [10 points]

The goal is to perform a maximum likelihood analysis on a set of supernova data (see the dropbox folder named data on sheet 8). At the end, we want the position of the maximum of the likelihood. This gives us the most likely values of the matter density of the Universe, Ω_M, and the dark energy density, Ω_L.

First, define the following functions in R:

1. The comoving distance:

$$d_C(z, \Omega_M, \Omega_L) = \int_0^z \frac{dz'}{\sqrt{\Omega_M(1+z')^3 + \Omega_L}} \tag{1}$$

2. The luminosity distance:
If $\Omega_k = 0$: $d_L(z, \Omega_M, \Omega_L) = (1+z)d_C(z, \Omega_M, \Omega_L)$
If $\Omega_k > 0$: $d_L(z, \Omega_M, \Omega_L) = (1+z)\sinh(\sqrt{\Omega_k} d_C(z, \Omega_M, \Omega_L))/\sqrt{\Omega_k}$
If $\Omega_k < 0$: $d_L(z, \Omega_M, \Omega_L) = (1+z)\sin(\sqrt{-\Omega_k} d_C(z, \Omega_M, \Omega_L))/\sqrt{-\Omega_k}$
Here, $\Omega_k = 1 - \Omega_M - \Omega_L$.

3. The distance modules as predicted by the theory:

$$\mu(z, \Omega_M, \Omega_L) = 5\log_{10} d_L(z, \Omega_M, \Omega_L) \tag{2}$$

4. The auxiliary function S_n:

$$S_n(\Omega_M, \Omega_L) = \sum_i \frac{(m_i - \mu(z_i))^n}{\sigma_i^2} \tag{3}$$

Here, m_i is the distance modules of the i-th super nova at redshift z_i.

5. The log-Likelihood, after marginalization over a constant, is given by:

$$L(\Omega_M, \Omega_L) = -\frac{1}{2}(S_2 - \frac{S_1^2}{S_0}) \tag{4}$$

Next, find the position of the maximum for L in two cases: The flat case, where the curvature of the Universe vanishes, i.e. $\Omega_k = 0$, and the general case. In either case, both parameters should be between 0 and 1. Create a 2D grid of $N \times N$points on the domain $[0,1] \times [0,1]$ and evaluate L on each point. Then find where the maximum is. Do the same for the flat case on the interval $[0,1]$. Attention: In the general case, the computation time goes as N^2 and can be quite long (roughly an hour for $N = 100$). Choose small N's or use a small subset of the entire supernova catalog to test your code. Your result should be close to $(\Omega_M, \Omega_L) = (0.3, 0.7)$. Then you can do a more precise run. If you want, you can afterwards compare the performance with finding the maximum by using the R-package "maxLik".

Hints:

- If an R function expects a scalar, you can't pass a vector in its place and think R will apply the function to each element. Use *sapply* instead.
- You can't naively compare floating point numbers. To check if a number is zero, check if it is very small instead, i.e. if $abs(x) < 1e^{-12}$ or so. -
- *sapply*() or *integrate*() expect a function as an argument, that only takes one argument. Define a new function within the function you are calling *sapply* or *integrate* from to get rid of extra parameters.

Problem 2 - Confidence limits [5 points]

Using the posterior $L(\Omega_m)$ (flat case) that you coded in Problem 1, find the 95% confidence-level range for Ω_m, i.e. $\Omega_m^{min} \leq \Omega_m \leq \Omega_m^{max}$. You should find $\Omega_m^{min/max}$ by trial and error such that

$$L(\Omega_m^{min}) = L(\Omega_m^{max}) \int_{\Omega_m^{min}}^{\Omega_m^{max}} L(\Omega_m) d\Omega_m = .95 \int_0^1 L(\Omega_m) d\Omega_m \tag{5}$$

Problem 3 - Maximum Likelihood Estimators [5 points]

In this exercise, you are supposed to find the Maximum Likelihood estimators for the exponential and Poisson distributions.

- Suppose we have n independent and identically distributed data points x_{j} following an exponential distribution with parameter λ, such that $p(x_{\mathrm{j}}) = Ce^{-\lambda x_{\mathrm{j}}}$, with $0 \leq x_{\mathrm{j}} < \infty$. You need to:
  1. ensure the distribution is normalized by computing C;
  2. compute the Maximum likelihood estimator for λ.
- Now, suppose we again have n independent and identically distributed data points x_{j}, but now following a Poisson distribution with parameter λ, such that $p(x_{\mathrm{j}}) = \frac{\lambda^{x_{\mathrm{j}}} e^{-\lambda}}{x_{\mathrm{j}}!}$, with $x_j = 0, 1, 2, ...$ You have to:
  1. verify that the Poisson distribution is normalized;
  2. compute the maximum likelihood estimator for λ.