Homework 3 Solutions: *Hirak Bandyopadhyay*

Problem 1:

For this problem, we have to group the galaxies first. I first get all the data for the coordinates of the galaxies, the right ascension and declination. Now, I divide the sphere of earth in two halves and each half in 4 quadrants. Hence I say is the declination is positive, then it is in northern hemisphere (denoted by +) and if negative, then I denote it by -. Now, once I do that I define each 6 hours of right ascension to be 90deg and hence I have sorted 8 scopes in 3D space. I plot all the galaxies like this and the sorting is given in the 'sorted.dat' file. Now, I utilize the proximity and distance for the galaxy clusters and sort them into appropriate groups. Now, I take average distance and velocity for each group and find out the least square fit using the 'hubbleme.py' code. The terminal output gives me this,

Hubble's constant slope b = 525.14 +- 182.77 km/s/Mpc

Intercept with r axis a = -98.51 +- 127.80 km/s

Estimated v error bar sigma = 211.1 km/s

Now, we know that the value from the Hubble paper is 513 +- 60 km/s/Mpc, which is really close to the value I got.

So, the age of the universe is

$$t_0 = \frac{1}{H} = \frac{1}{525.14}\ Mpc.\frac{s}{km} = \frac{1}{525.14} * 3.086 * 10^{19} * \frac{1}{3600*24*365}\ year = 1.863\ Gyr.$$

This estimate is really poor.

Problem 2:

For this problem, we have to solve the least square fit first for the quadratic equation. The solutions are attached in the end of this document.

Now, as we have measured all the necessary quantities now we can proceed to the code. First, we use the template to read the data file of co2 concentration. We understand we are looking for a CO2 concentration vs year plot, so we sort the data and the dates to two separate arrays. Now, once that is done, we initialize all the parameters needed and the use a for loop to sum up all those quantities. The read method gives us the values in strings, so we convert those in float for our purpose. Now, once we have all the summed up values, we define delta as denominator, and a, b, c accordingly. Now, for uncertainties, we again initialize all the parameters, and then first find out the variance using the designated formulae from lecture notes.

$$\sigma^2 \equiv \frac{f(a,b)}{\nu} = \frac{1}{n-2}\sum_{i=0}^{n-1}\left(y_i - y(x_i)\right)^2 \ .$$

Once we use that formulae to sum up the values using another for loop, we move on to find the uncertainties of a, b, c using the following formulae,

$$\sigma_f^2 = \sum_{i=1}^{N} \sigma_i^2 \left(\frac{\partial f}{\partial y_i}\right)^2$$

We use another for loop to calculate all these quantities, and then impose the condition that if the number of data points is more than 2 then we can set the uncertainties of a, b, c to be the calculated ones, otherwise, we set all to zero.

Now, with all these values, we use another program to do create the graphs of the fitting and extrapolation, and we take the data from the actual co2 data file and create another file called co2.dat. The gnuplot program takes values from that file and output the fitting parameters in terminal and create two png files for present day fitting and future prediction.

Now, to extrapolate this into future, we solve the quadratic equation to find out the death time to be **3914 Year** Co2 concentration becomes toxic for any animal.

These are the final set of parameters from gnuplot

```
Final set of parameters            Asymptotic Standard Error
=======================            ==========================

a               = 0.012576         +/- 0.0003072    (2.442%)

b               = -48.4646         +/- 1.221        (2.52%)

c               = 46994.8          +/- 1214         (2.583%)
```

We can see that the gnuplot fitting error percentage is less than our quadratic fitting.

Problem 3:

For this problem, we repeat the same with exponential plot. We take a natural logarithm on both the sides and then follow the exact same procedure of problem 2. The calculated quantities are scanned and attached at the end of the document.

For this case,

$$y = Ae^{bx}$$
$$Taking\ natural\ log,$$
$$lny = lnA + bx$$

Now, in our calculation $b = B\ and\ a = \exp(A)$

The final fit parameters from gnuplot is given below,

```
Final set of parameters            Asymptotic Standard Error
=======================            ==========================

a               = 0.0560982        +/- 0.002254     (4.019%)

b               = 0.00439823       +/- 2.01e-05     (0.457%)
```

We see that the error bar is again lesser than the calculated value.

For this case, we solve the linear equation of natural logarithm to find out the death time to be **3115 Year** which is significantly closer to the time predicted in problem2.

$$f(a,b) = \sum_{i=0}^{N-1} \left(\frac{y_i - ax_i^2 - bx_i - c}{\sigma_i}\right)^2$$

So, $\frac{\partial f}{\partial a} = 0$ (1)

$$\Rightarrow \sum \frac{2}{\sigma_i^2}\left(y_i - ax_i^2 - bx_i - c\right)\left(-x_i^2\right) = 0$$

$$\Rightarrow \sum \left(x_i^2 y_i - ax_i^4 - bx_i^3 - cx_i^2\right) = 0$$

We define,

$$S_{2xy} = \sum x_i^2 y_i \qquad S_{3x} = \sum x_i^3$$

$$S_{4x} = \sum x_i^4 \qquad S_{2x} = \sum x_i^2$$

$$S = \sum \frac{1}{\sigma_i^2}$$

and, $\frac{\partial f}{\partial b} = 0$ (2)

$$\Rightarrow \sum \frac{2}{\sigma_i^2}\left(y_i - ax_i^2 - bx_i - c\right)\left(-x_i\right) = 0$$

$$\Rightarrow \sum \left(x_i y_i - ax_i^3 - bx_i^2 - cx_i\right) = 0$$

We define, $S_{xy} = \sum x_i y_i$

and, $S_x = \sum x_i$

and,

$$\frac{\partial f}{\partial c} = 0 \qquad (3)$$

$$\Rightarrow \sum \frac{2}{\sigma_i^2}\left(y_i - a x_i^2 - b x_i - c\right)\left(\frac{-1}{\sigma_i^2}\right) = 0$$

$$\Rightarrow \sum \left(y_i - a x_i^2 - b x_i - c\right) = 0$$

we define, $S_y = \sum y_i$

So, from (1),

$$S_{2xy} - a S_{4x} - b S_{3x} - c S_{2x} = 0 \quad \text{—— (I)}$$

from (2),

$$S_{xy} - a S_{3x} - b S_{2x} - c S_x = 0 \quad \text{—— (II)}$$

from (3),

$$S_y - a S_{2x} - b S_x - c S = 0 \quad \text{—— (III)}$$

So,
$$\left(\begin{array}{ccc|c} S_{4x} & S_{3x} & S_{2x} & S_{2xy} \\ S_{3x} & S_{2x} & S_x & S_{xy} \\ S_{2x} & S_x & S & S_y \end{array}\right)$$

$$\hat{a} = \begin{pmatrix} S_{2xy} & S_{3x} & S_{2x} \\ S_{xy} & S_{2x} & S_x \\ S_y & S_x & S \end{pmatrix}$$

$$\hat{b} = \begin{pmatrix} S_{4x} & S_{2xy} & S_{2x} \\ S_{3x} & S_{xy} & S_x \\ S_{2x} & S_y & S \end{pmatrix} \qquad \hat{c} = \begin{pmatrix} S_{4x} & S_{3x} & S_{2xy} \\ S_{3x} & S_{2x} & S_{xy} \\ S_{2x} & S_x & S_y \end{pmatrix}$$

$$\Delta = \begin{pmatrix} S_{4x} & S_{3x} & S_{2x} \\ S_{3x} & S_{2x} & S_x \\ S_{2x} & S_x & S \end{pmatrix}$$

So,

$$\det \Delta = \Big[S_{4x}\left(S_{2x} S - S_x^2\right) - S_{3x}\left(S_{3x} S - S_x S_{2x}\right) + S_{2x}\left(S_{3x} S_x - S_{2x}^2\right) \Big]$$

$$a = \frac{\det \hat{a}}{\Delta}$$

$$= \frac{S_{2xy}\left(S_{2x} S - S_x^2\right) - S_{3x}\left(S_{xy} S - S_x S_y\right) + S_{2x}\left(S_{xy} S_x - S_{2x} S_y\right)}{\Delta}$$

$$b = \frac{S_{4x}\left(S_{xy} S - S_x S_y\right) - S_{2xy}\left(S_{3x} S - S_{2x} S_x\right) + S_{2x}\left(S_{3x} S_y - S_{xy} S_{2x}\right)}{\Delta}$$

$$c = \frac{S_{4x}\left(S_{2x} S_y - S_{xy} S_x\right) - S_{3x}\left(S_{3x} S_y - S_{xy} S_{2x}\right) + S_{2xy}\left(S_{3x} S_x - S_{2x}^2\right)}{\Delta}$$

$$\sigma_f^2 = \sum_{i=0}^{N-1} \sigma_i^2 \left(\frac{\partial f}{\partial y_i}\right)^2$$

$$\sigma_a^2 = \sum_{i=0}^{N-1} \sigma_i^2 \left(\frac{\partial a}{\partial y_i}\right)^2$$

$$\cancel{\frac{\partial a}{\partial y_i} = \frac{S_{2x}\left(S_{2x} S - S_x^2\right) - S_{3x}\left(S_x S - S_x\right) + S_{2x}\left(S_x^2 - S_{2x}\right)}{\Delta}}$$

$$\frac{\partial a}{\partial y_i} = \frac{x_i^2\left(S_{2x} S - S_x^2\right) - S_{3x}\left(x_i S - S_x\right) + S_{2x}\left(x_i S_x - S_{2x}\right)}{\Delta}$$

$$\frac{\partial b}{\partial y_i} = \frac{S_{4x}\left(x_i S - S_x\right) - x_i^2\left(S_{3x} S - S_{2x} S_x\right) + S_{2x}\left(S_{3x} - x_i S_{2x}\right)}{\Delta}$$

$$\frac{\partial c}{\partial y_i} = \frac{S_{4x}\left(S_{2x} - x_i S_x\right) - S_{3x}\left(S_{3x} - x_i S_{2x}\right) + x_i^2\left(S_{3x} S_x - S_{2x}^2\right)}{\Delta}$$

$$f(a,b,c) = \sum \left(\frac{y_i - a e^{b x_i} - c}{\sigma_i} \right)^2$$

Problem 3

So, $\dfrac{\partial f}{\partial a} = \sum 2 \left(y_i - a e^{b x_i} - c \right)\left(-e^{b x_i} \right) = 0.$

So, $\sum \left(e^{b x_i} y_i - a e^{2 b x_i} - c e^{b x_i} \right) = 0.$

$$\left.\begin{aligned} S_{2xy} &= \sum (x_i^2 y_i) \\ S_{yly} &= \sum (y_i \ln y_i) \\ S_{xy} &= \sum (x_i y_i) \\ S_{xyly} &= \sum (x_i y_i \ln y_i) \end{aligned}\right\} a$$

$S_y = \sum y_i$ $\qquad A = (S_y S_{2xy} - S_{xy}^2)$

$$a = \frac{S_{2xy} S_{yly} - S_{xy} S_{xyly}}{S_y S_{2xy} - S_{xy}^2}$$

So, $\dfrac{\partial f}{\partial y} =$ $\qquad \dfrac{\partial}{\partial y_i} \cancel{(x_i y_i \ln y_i)}$

$$\frac{\partial}{\partial y_i}\left(y_i \ln y_i\right) = \left(\ln y_i + 1\right)$$

So, $\dfrac{\partial a}{\partial y_i} = \rule{8cm}{0.4pt}$

$$\frac{\partial a}{\partial y_i} = \frac{x_i^2 S_{yly} + S_{2xy}\left(1 + \ln y_i\right) - x_i S_{xyly} - S_{xy} x_i \left(\ln y_i + 1\right)}{S_y S_{2xy} - S_{xy}^2}$$

$$- \frac{\left(S_{2xy} S_{yly} - S_{xy} S_{xyly}\right)}{\left(S_y S_{2xy} - S_{xy}^2\right)^2}\left[S_{2xy} + x_i^2 S_y - 2 S_{xy} x_i\right]$$

$$b = \frac{S_y S_{xyly} - S_{xy} S_{yly}}{S_y S_{2xy} - S_{xy}^2}$$

So, $$\frac{\partial b}{\partial y_i} = \frac{S_{xyly} + S_y . x_i\left(\ln y_i + 1\right) - x_i S_{yly} - S_{xy}\left(\ln y_i + 1\right)}{S_y S_{2xy} - S_{xy}^2}$$

$$- \frac{\left(S_y S_{xyly} - S_{xy} S_{yly}\right)}{\left(S_y S_{2xy} - S_{xy}^2\right)^2}\left[S_{2xy} + x_i^2 S_y - 2 x_i S_{xy}\right]$$