

week14_coursework

February 6, 2025

1 PHYS465: Coursework Exercise 4: Part 1

1.0.1 Deadline Monday 17th Feb @ 4pm

1.0.2 This coursework assesses the learning outcomes from Week 14.

The final coursework exercise for PHYS465 is made up of 2 parts. This is part 1; part 2 will be released on Wed 12th Feb. Your submission should include both exercises

1.1 Instructions

- Submit your work via Moodle.
- You must submit a fully compiled `.ipynb` file which includes all codes required to replicate your results **and** a `.pdf` version.
 - **Dont forget to run every cell before submitting**
 - You must also respond to the mandatory GenAI self-assessment questionnaire.
- Your submission must include text (in markup format) that describes what each cell does and summarises the conclusions
- The estimated workload for this is 4-6 hours.

1.1.1 Tips

- The last question of this exercise asks you identify a key result. **To do this you do not have to have completed all exercises.** This assessment is designed to test your reflections on the problem undertaken.
- Don't worry too much about how your code looks - while some marks will be given for sensible coding, the focus of this assessment is your approach used in solving the problem, your reasoning, explanation and answer.
- As data visualisation is a key outcome, marks will be given for well presented plots
- Explain all your reasoning for each step. A *significant fraction* of the marks are given for explanations and discussion, as they evidence understanding of the analysis.
- Include all relevant lines of code including import statements and read statements. As part of the assessment your code will be run offline.

1.1.2 WARNING

- This submission must be your own work. Please note the university's policy on plagiarism.
- While it is acceptable (and indeed encouraged) to share ideas, you must ensure that you do not use other people's code or text, and that the reflections are your own.

- It is acceptable to use GenAI tools for guidance on how to approach this exercise, but you must ensure that all code is written by you.
 - Should you use GenAI in this work, then answer yes to the GenAI self-assessment. You will not be penalised for this. ***

1.2 The Problem

Dataset 1:

For this exercise, I have generated some data that mimics the results of an experiment. The file of interest can be found https://raw.githubusercontent.com/MatSmithAstro/phys465_resources/main/coursework/datasets/week1

The file contains 53 measurements (x, y), each with an associated uncertainty (y_{err})

This data is expected to be well fitted by a linear function ($y = \text{inter} + \text{slope} \times x$). It is your job to recover the value of inter and slope ***

Dataset 2:

Distances to type Ia supernova (SNeIa) are used to measure the size of the Universe as a function of time. The state of the art in this field is the Dark Energy Survey (DES). The latest results can be downloaded from https://raw.githubusercontent.com/MatSmithAstro/phys465_resources/main/coursework/datasets/DES-S

This dataset contains 1829 SNeIa each with a measured redshift (z_{HD}), distance (MU) and uncertainty (MUERR_FINAL).

In cosmology we are interested in constraining the expansion rate of the universe today (the ‘Hubble constant’, H_0), the fraction of the Universe that is matter (Ω_{m}), and the fraction that is dark energy (Ω_{λ}).

In this exercise, we will use the measured distances to estimate these parameters

1.3 Part 1: Algorithm Development

1. Write a function that defines a linear function
[1 mark]
2. Write a function to calculate the log likelihood between this model and the dataset
 - *Hint*: the equation for this can be found in the lecture notes.
 - *Hint*: not all terms in this equation are necessary, but the more that are included, the more accurate the results will be
 [3 marks]
3. From previous experience, we assume a uniform prior of $-50 < \text{inter} < 50$ and $0 < \text{slope} < 100$. Write a function that describes the log prior. A normalisation of $1e-2$ is reasonable. Combine these two functions to estimate the log posterior.
[3 marks]
4. Write your own Metropolis-Hastings algorithm to conduct a random walk in this (slope, intercept) parameter space
 - *Hint*: recall that we have defined the log posterior: recall what happens to a ratio when logged

- *Hint*: given that we are working in log space, the random uniform number must also be logged
 - *Hint*: remember to store all steps taken
[6 marks]
5. Read in dataset 1, and run an MCMC chain with 10,000 steps to constrain `inter` and `slope`. From the results calculate the maximum likelihood for each parameter.
 - *Hint*: good initial conditions are `slope = 1.2` and `inter = 4.5`. A step size of 0.05 for both parameters is also appropriate
[3 marks]
 6. From this output, plot the chain values against the step number to determine how many steps should be removed
[3 marks]
 7. After removing ‘the burn’, plot marginalised histograms and a 2D distribution for this chain.
 - *Hint*: the `corner` library is useful for this.
 - *Hint*: the 2D distribution can be either a contour or a scatter plot
[5 marks]

Maximum available: 24 ***

1.4 Part 2: Estimating the cosmological parameters with MCMC

to estimate the cosmological parameters we can use the `astropy.cosmology` library. In particular we can load either the `astropy.cosmology.FlatLambdaCDM` or `astropy.cosmology.LambdaCDM` objects.

These allow us to calculate distances given different values of z , H_0 , Ω_m and Ω_Λ . (NB: the `flatLambdaCDM` enforces that $\Omega_\Lambda = 1 - \Omega_m$). i.e. `LambdaCDM(H0=h, Om0=Omega_m).distmod(mu_data.zHD.values).value` will return the theoretical distance to this event.

Using this module, we will determine the best-fitting values of H_0 , Ω_m and Ω_Λ . From experience, we know that $(0.2 < \Omega_m < 0.5)$ & $(65. < H_0 < 75.)$.

-
1. Read in the SN dataset, and make a plot of the terms of interest (zHD ; redshift and μ distance modulus).
[2 marks]
 2. Write a function to calculate the log likelihood between the `FlatLambdaCDM` model and this dataset that uses the `astropy.cosmology` module.
 - *Hint*: it is perfectly acceptable to draw on your experience in Section 1
[4 marks]
 3. Write a function to calculate the log prior. Combine this with the log likelihood to estimate the log posterior.
[4 marks]
 4. Use either the `emcee` or `pymc` library to produce an MCMC chain for this problem.
 - *Hint*: Recall that for these two libraries the data is input slightly differently.
 - *Hint*: For `emcee` using the Ensemble sampler, 100 walkers is appropriate. For this sampler you must also provide initial conditions for each walker.
[7 marks]

5. Using the `corner` package, produce a `corner` plot of your results.
[5 marks]
6. **Extension:** Either:
 - Expand this analysis to the `LambdaCDM` module.
 - Make a plot of the data with a selection of the viable model parameters overplotted.
[6 marks]

Maximum available : 28 ***

1.5 Part 3: Concluding remarks

1. **Summary Statement.** Write a short reflective statement (200 words max) summarising a key result from cosmological analysis in Part 2. You may include a maximum of one figure. Key topics to consider are the existence of dark energy, the consequences of removing parameters from our analysis, limitations in the dataset. A significant fraction of the marks are awarded for reflective thinking: what did you learn? If you did not make it to Part 2, then reflect on your learning from Part 1
[8 marks]

Additional Marks Marks will be awarded for notebooks, codes and plots that are well explained and well formatted. In particular, attention will be given to sensible variable names, easy to follow comments and notebook structure.

[6 marks]

Maximum available: 14 * Maximum possible** 66
