Assignment #2: Probabilities, Bayes' theorem, and Fitting with Bivariate Errors (total 10 points), due by 11:59 pm Tuesday, 15 October 2024

Exercise 1: Bayes' on test accuracy. An emergency beacon for hiking (also known as a personal locator beacon, or PLB) is a small handheld device that will send out a distress signal in the event of an emergency. The Altitude Manufacturing Company makes 81% of the PLBs, the Bright Company makes 15% of them, and the Camping Company makes the other 4%. The PLBs made by Altitude have a 4% rate of defects, the Bright PLBs have a 6% rate of defects, and the Camping PLBs have a 9.5% rate of defects.

(a) If a PLB is randomly selected from the general population of all PLBs, find the probability that it was made by the Altitude Manufacturing Company. **(1 point)**

(b) If a randomly selected PLB is then tested and is found to be defective, find the probability that it was made by the Altitude Manufacturing Company. **(1 point)**

You can write your answers for this question using any text editor of your preference, but, please, set the document layout with the following features:

- Page orientation: portrait;
- Page size: 8.5" × 11" (letter);
- Margins: 2.54 cm (top, bottom, left, right);
- Line spacing: between 1.0 and 1.15 pt;
- No double-column format. Write your answers as a 1-column text;
- Paragraph alignment can be 'align left' or 'justify';
- Font type and size: Times New Roman, 12 pt, non-italic, non-bold;
- Use *italic*/**bold**/<u>underline</u> only to emphasize brief pieces of text if needed.
- ONLY `pdf` FILES ARE ACCEPTED. Do not submit editable files such as `.docx`.

Alternatively, you can write your answers on paper and scan your solutions. However, **make sure your answers can be read clearly and the handwriting is legible.** The scanned file to be submitted needs to be in `pdf` format.

*****

Exercise 2: Linear Regression with Bivariate Errors. The data file `pressureRatio.dat` contains the measurement of the thermal energy of certain sources using two independent methods labelled Method 1 and Method 2, and their ratio. The error bars indicate the 68% or 1σ, confidence intervals, although the fact that most intervals are asymmetric indicates that the measurements do not follow exactly a Gaussian distribution. Note that all quantities are in arbitrary units, meaning you do not need to worry about units in this exercise. The file contains 25 rows ($N = 25$ different sources) and 13 columns, with each column corresponding to the following quantities:

- (Column 1) `Num`: integer indexing a source
- (Column 2) `Radius`: radius of the source
- (Column 3) `R_e+`: upper uncertainty in the radius of the source
- (Column 4) `R_e-`: lower uncertainty in the radius of the source
- (Column 5) `E1`: energy of the source measured via Method 1
- (Column 6) `E1_e+`: upper uncertainty in `E1`
- (Column 7) `E1_e-`: lower uncertainty in `E1`
- (Column 8) `E2`: energy of the source measured via Method 2
- (Column 9) `E2_e+`: upper uncertainty in `E2`
- (Column 10) `E2_e-`: lower uncertainty in `E2`
- (Column 11) `ratio`: ratio (`E1`/`E2`)
- (Column 12) `r_e+`: upper uncertainty in `ratio`
- (Column 13) `r_e-`: lower uncertainty in `ratio`

The way to read and report asymmetric uncertainties is, for instance, for the ratio of source 1, we write $0.86\pm^{0.08}_{0.07}$ or $0.86^{+0.08}_{-0.07}$. For the present analysis, one can consider the intervals – with approximately 68% probability – to include the true values of the studied quantities. This means that the intervals can be expressed as a $\pm\sigma$ interval assuming that the measurements are normally distributed. For more details about the data, check M. Bonamente, et al., New J. Phys. **14**, 025010 (2012).

(a) Calculate the weighted average of the ratios (taken between the two methods) and its standard deviation, assuming that the errors are Gaussian and equal to the average of the asymmetric errors. **(1 point)**

(b) Calculate the linear average of the ratios and explain why it is larger than the weighted average from (a). **(1 point)**

Consider all 25 measurements of `ratio` in the data file. Assume that an additional ± 0.1 is to be added linearly to the statistical error of each measurement reported on the file.

(c) Show that the addition of this source of uncertainty results in a weighted average of 0.95 ± 0.04. **(1 point)**

(d) Compare with the standard weighted average (with no additional uncertainty added) and explain the reason for the difference. **(1 point)**

(e) For the data on the file, calculate the linear average, the weighted average, and the median of each quantity (`radius`, `E1`, `E2`, and `ratio`). You may assume that the error of each measurement is the average of the two errors reported on the file. **(1 point)**

(f) Calculate the weighted logarithmic average of the quantity `ratio` in the file, and its uncertainty, and then convert the results back to linear scale. **(1 point)**

For linear regression with bivariate error, we need to consider the uncertainties in both variables X and Y in the fitting. For that, you will need the **Linear Regression for Data with Measurement Errors and Intrinsic Scatter (BCES)** package (see `https://pypi.org/project/bces/`).

(g) Use the bivariate error data of `E1` and `E2` from the file, with the errors obtained as the average of the given uncertainties, and calculate the best-fit parameters and errors of the linear model $y = bx + a$ with x being `E1` and y being `E2`. Present a plot of the best-fit curve with the data points (including the error bars in both x and y). **(1 point)**

(h) Use the bivariate error data of `Radius` and `ratio` from the file, with the errors obtained as the average of the given uncertainties, and calculate the best-fit parameters and errors of the linear model $y = bx + a$ with x being `Radius` and y being `ratio`. Present a plot of the best-fit curve with the data points (including the error bars in both x and y). **(1 point)**

*****

Submission Information

Name the `pdf` file you generated for question 1 as `question1_assignment_2.pdf` and the code for question 2 as `code2_assignment_2.ipynb`. Note that only Jupyter Notebook files are accepted for the coding part. You will upload the `pdf` file of your written answer and your code to the Gradescope platform. **It is important that you log in to Gradescope.ca and not Gradescope.com.** Your code needs to be documented, i.e., introduce comment lines to explain their main procedures. Pure lines of code without explanatory comments will have reduced marks. You can upload multiple files to the Gradescope platform and you can resubmit your work until the due date. Upload all files at once. We will test-run your submitted codes for syntax errors and check if they generate the requested outcomes. We will also check if the results/figures presented in your assignment are 'paper-like' quality and that the quantitative predictions are scientifically/mathematically founded.

ONLY `pdf` (for the written reports), `jpg/jpeg` (for figures), and `.ipynb`, (for codes) FILES ARE ACCEPTED. This means we are only accepting Python Notebooks.

IMPORTANT: In your codes include instructions on how to run it and, if applicable, include testing values for the initial conditions or settings that you attempted so we can reproduce the results. We will test other initial conditions and parameter variations to evaluate the robustness of your code. But we need a place to start. Include also information about the version of your coding platform. If this information is not included, points will be deducted.

IMPORTANT: Your codes need to output all required numbers of the exercises and they should be printed clearly. The printed values should be properly identified with text saying what is being printed. We can only assess and consider marks for numerical answers printed by your code (we cannot give marks to what we cannot see). We cannot debug your code and/or search the answers inside your code for you. This is a graded component of the course, therefore, we can ONLY evaluate what is uploaded and see printed on the screen. We will test other parameter variations and settings in your code, but we will not alter its structure and logistics. Interpret your code as any other graded component as 'in paper', meaning what you submitted is your final answer and what is not provided (or printed) in terms of answers, we cannot consider for marks.

IMPORTANT: (Follow-up from the item above) If an exercise asks you to write a code to compute or calculate a value, we will need to see the calculated output value and not simply a number manually printed using the `print` command. For instance, if an exercise asks you to write a code to obtain the velocity of a particle, we need to see the output velocity of the algorithm you implemented and not simply `print(2.34, 'km/s')`.

*****