UCF Physics: AST 5765/4762: (Advanced) Astronomical Data Analysis

Fall 2019 Homework 4

Due Tuesday 24 September 2019

Work:
Become sufficiently familiar with Python to:

1. Read and write FITS files
2. Read, alter, and add header items
3. Read, parse, and write ASCII data files
4. Produce PDF and PNG versions of plots or images
5. Display and explore an astronomical image with `ds9`

Become sufficiently familiar with probability and error analysis to:

1. Understand all the terms used as headings or italicized in the reading assignments.
2. Calculate the mean, median, mode, range, standard deviation, and variance of a group of measurements.
3. Understand the Gaussian and Poisson distributions, their parameters, and their application in astronomical observation.
4. Memorize the formula for the Gaussian distribution.

Resources:

1. The DS9 documentation:
   `http://hea-www.harvard.edu/RD/ds9/`, also in the class online space.
2. Chapters 1 and 2 of Bevington (**DUE** before class Tuesday, 17 September 2019)
3. Chapters 3 and 4 of Bevington (**DUE** before class Tuesday, 24 September 2019)
4. Chapter 6 of Bevington (**DUE** before class Tuesday, 1 October 2019)

Hand in:

1. (10 points) Find a Python function that returns a sample consisting of `N` random draws from a Gaussian distribution of width `sigma` and mean `cx`. Use it to get a 10,000-draw sample from a population with $\sigma = 13$ and $\bar{x} = 55$. Watch your datatypes and be sure to follow good coding habits here and throughout.

Write a series of Python commands that test your function:

2. (10 points) Plot the histogram of your sample. The histogram should have bins from 0 to 100 that are 1 unit wide. Be sure the plot has axis labels and a title. Save it as a PNG file.

3. (10 points) Overplot a Gaussian with the same parameters. Save it as a PNG file. Remember that the analytic Gaussian distribution, integrated over each bin, gives you the expected number of draws in each bin for a SINGLE draw from the parent population (i.e., less than one draw per bin, on average). You'll need to multiply by the number of draws to get the expected number of draws per bin for a larger sample. Please approximate the integration over the bin boundaries by evaluating the Gaussian at the center of each bin.

   Write a series of Python commands that explores probability. You can do the next three problems separately, or (better) you may write a single block of code that does the job. Either gets full credit if done correctly. Regardless of how you approach it, use loops or array math to handle the calculations, rather than repeatedly writing out the same commands. Most people will write a nested loop, but it is possible, with some NumPy trickery, to use just one `for` loop.

4. (10 points) Use your function to create a sample with 10 draws from the population given above. Do this 10 times. For each sample, record the sample number, the sample mean, and the sample standard deviation in a table (array) with one row per sample (so there are three columns and 10 rows). Do not include the samples in your array. Write the array into a file called `hw4_`*username*`.txt`. In your file, before the table, include a comment stating the number of draws. Hint: If you can't easily figure out how to write to a file, Google may be your friend.

5. (10 points) Do the same (10 repetitions for each sample size) for samples with 100, 1,000, 10,000, 100,000, and 1,000,000 draws, recording each set of results in a separate 2D array and appending them to the file above.

6. (10 points) Find the standard deviation of the means for *each* of the six sample sizes. Don't do this from first principles, but rather use a function on the data you have generated. Print the sample sizes and these new numbers on the screen and in your `.txt` file with an appropriate comment.

7. (10 points) Make a log-log plot of the standard deviation of the mean *vs.* sample size. Make the sample size be the independent variable. Save as PNG.

   You may do the following longhand and include a photo or scan in your homework handin directory. Do not hand in paper. Please express the final answers as decimal numbers.

8. (2 points) What are $\begin{array}{r} 10 \pm 2 \\ +30 \pm 4 \\ \hline \end{array}$

9. (2 points) What are

$$\begin{array}{r} 10 \pm 2 \\ -30 \pm 4 \\ \hline \end{array}$$

10. (2 points) What are

$$\begin{array}{r} 10 \pm 2 \\ \times 30 \pm 4 \\ \hline \end{array}$$

11. (2 points) What are

$$\begin{array}{r} 10 \pm 2 \\ \div\ 30 \pm 4 \\ \hline \end{array}$$

12. (10 points) Include a copy of your class log file in your handin. Print the Git log for your main homework file.
13. **AST 5765 only:** (10 points) Show analytically that the full width at half maximum (FWHM) of a Gaussian distribution is 2.354σ (See Bevington, p. 28). This can be done either in a PDF file or in comments in your homework file. If it's in a PDF file, please refer to the filename from the homework file.
14. **AST 5765 only:** (10 points) A line on a log-log plot has a specific meaning, and is the main purpose of log-log plots. Derive the relationship of x, y, and the slope and intercept of a line in a log-log plot.