EP 4130/PH 6130 Assignment 1

Deadline **16** January 2023 before **23:59 hrs**

All problems (Except the last) have equal weightage of 10 points each. The last problem has 20 points. Please show the source code used for each of the problems.

1. Create 1000 draws from a normal distribution of mean of 1.5 and standard deviation of 0.5. Plot the pdf. Calculate the sample mean, variance, skewness, kurtosis as well as standard deviation using MAD and σ_G of these samples.

2. Plot a Cauchy distribution with μ=0 and γ=1.5 superposed on the top of a Gaussian distribution with μ=0 and σ=1.5. Use two different line styles to distinguish between the Gaussan and Cauchy distribution on the plot and also indicate these in the legends.

3. Plot Poisson distribution with mean of 5, superposed on top of a Gaussian distribution with mean of 5 and standard deviation of square root of 5. Use two different line styles for the two distributions and make sure the plot contains legends for both of them.

4. The following were the measurements of mean lifetime of K meson (as of 1990) (in units of 10^{-10} s) :
$0.8920 \pm 0.00044; 0.881 \pm 0.009; 0.8913 \pm 0.00032; 0.9837 \pm 0.00048; 0.8958 \pm 0.00045$. Calculate the weighted mean lifetime and uncertainty of the mean.

5. Download the eccentricity distribution of exoplanets from the exoplanet catalog `http://exoplanet.eu/catalog/`. Look for the column titled e, which denotes the eccentricity. Draw the histogram of this distribution. Then redraw the same histogram after Gaussianizing the distribution using Box-transformation either using `scipy.stats.boxcox` or from first principles using the equations shown in class or in arXiv:1508.00931. Note that exoplanets without eccentricity data can be ignored.

EP 4130/PH 6130 Assignment 2

Deadline **23** Jan 2023 before **23:59 hrs**

All problems have equal weightage of 20 points each. Please show the source code used for each of the problems. Please don't hesitate to ask questions if you're stuck anywhere. The final assignment must be submitted as a pdf file.

1. In the class, we demonstrated the Central Limit Theorem for a sample drawn from a uniform distribution. Reproduce a similar plot for a sample drawn the from chi-square distribution with degrees of freedom equal to 3, for samples drawn once, 5 times, and 10 times. Either plot all of these on one multipanel figure similar to AstroML figure 3.20. (20 points)
   (Hint: look up `numpy.random.chisquare` and show the distribution of x from 0 to 10)

2. The luminosity and redshift of galaxy clusters from XMM-BCS survey (details available at arXiv:1512.01244) can be downloaded `http://www.iith.ac.in/~shantanud/test.dat`. Plot the luminosity as a function of redshift on a log-log scale. By eye, do you think the datasets are correlated? Calculate the Spearman, Pearson and Kendall-tau correlation coefficients and the p-value for the null hypothesis. (20 points)

3. Wind speed data from the Swiss Wind Power data website can be found at `http://wind-data.ch/tools/weibull.php`. Using the data provided on the website, plot the probability distribution and overlay the best-fit Weibull distribution (with the parameters shown on the website). (20 points)
   (Hint : A on the website is same as λ, which was used in class to parameterize the Weibull distribution.)

4. Generate two arrays of size 1000 drawn from a Gaussian distribution of mean of zero and standard deviation of one. Calculate Pearson correlation coefficient and its $p-$value using `scipy` module. Also check if the $p-$ value agrees with that calculated using the Student-t distibution.

EP 4130/PH 6130 Assignment 3

Deadline **30** January 2023 before **23:59 hrs**

First two problems have equal weightage of 25 points each. Last problem has a weightage of 10 points. Please show the source code used for each of the problems.

1. In class, we showed histograms of standard deviation and σ_G of bootstrap samples drawn from a Gaussian distribution with mean equal to 0 and standard deviation equal to 1. Draw a similar histogram of median of 10,000 bootstrap samples drawn from the same Gaussian distribution. According to `http://tinyurl.com/h6p43o8`, the standard deviation of the sample median of a Gaussian distribution is equal to $\sqrt{\frac{\pi}{2n}}$. Overlay a Gaussian distribution on top of the histogram with mean equal to the mean of the generated data sample and standard deviation equal to the standard deviation of the median (Hint: Look up `astroML.stats.median_sigmaG`. Also note that you don't have to draw 10,000 histograms, but only one histogram consisting of 10,000 bootstrap resamples.)

2. arXiv:1008.4686, Exercise 1 on Page 5, except the last sentence of the question related to σ_m^2. (Hint : Use χ^2 minimization to obtain best-fit values of b and m, instead of linear algebra. You can look up `curve_fit` function in scipy.)

3. Calculate the p-value for the four chi-square values for the plot shown in class from astroMl book which can be found at `https://www.astroml.org/book_figures_1ed/chapter4/fig_chi2_eval.html`.
(Hint : You can read off the χ^2 values from the graph by multiplying by D.O.F.)

EP 4130/PH 6130 Assignment 4

Deadline **14** February 2023 before **23:59 hrs**
Please show source code for all the problems (except the third)

1. Download the data corresponding to x, y, and σ_y from `http://www.iith.ac.in/~shantanud/testdata.dat`. Find the best-fit values after fitting the data to linear, quadratic, and cubic polynomials. Find out which model fits the data best from frequentist model comparison as well as using AIC and BIC. For frequentist model comparison, using the linear model as the null hypothesis, find out the p value corresponding to the preferred model. (or if the linear model is the prefered model, then compare it to the quadratic model). Also show a plot overlaying the data with best fit solutions from linear, quadratic and cubic functions with different line styles for each of the fits. [30 pts]

2. For the model comparison problem shown in class on JVDP's blog, calculate AIC and BIC for the linear and quadratic models. Do these results agree with the frequentist model comparison results shown on the blog? Also mention the qualitative significance using strengt of evidence rules. [10 pts]

3. Find out one paper in research literature which uses the Kolmogorov-Smirnov test and explain briefly how it was used in that paper. Is K-S test used incorrectly (in this paper) as per the warnings on the Penn State website discussed in class? [10 pts]
(Hint : Use google scholar or enter "Kolmogorov-Smirnov" or "K-S" test in the abstract tab in the astro-ph or hep-ex section of arXiv. Also indicate the paper reference.)

4.
   - Calculate the significance in terms of no of sigmas of the Higgs boson discovery claim from the p value given in the abstract of the ATLAS discovery paper, arXiv:1207.7214.
   - Do the same for the LIGO discovery of GW150914, for which the p value $= 2 \times 10^{-7}$. (Hint : look up `norm.isf`)
   - From the Super-K discovery paper for neutrino oscillations (hep-ex/9807003), calculate the χ^2 GOF using the best-fit $\nu_\mu \leftrightarrow \nu_\tau$ oscillation solution. (Hint : check page 4 of the paper, second column, last paragraph) [10 pts]

EP 4130/PH 6130 Assignment 5

Deadline **20** Feb 2023 before **23:59 hrs**
Please show the source code for each of the problems.

1. Download the asteroid dataset from `http://astrostatistics.psu.edu/datasets/asteroid_dens.dat`. Apply the Shapiro-Wilk test to both the asteroid density values and the natural logarithm of the density values. From the p values, which of these is closer to a Gaussian distribution? Verify this by plotting histograms of both density and its logarithm and overlaying the best-fit normal distribution (Look up `stats.norm.fit`) (25 points)

2. Download the Hipparcos star catalog from `http://iith.ac.in/~shantanud/HIP_star.dat`. Detailed explanation of the columns in this dataset can be found in `http://astrostatistics.psu.edu/datasets/HIP_star.html` under "Dataset". Calculate using two-sample t-test whether the color (B-V) of the Hyades stars differs from the non-Hyades ones. The Hyades stars have Right Ascension between 50° and 100°, declinations between 0 and 25°, proper motion in RA between 90 and 130 mas/year, proper motion in DEC between -60 and -10 mas/year. Any other star which does not satisfy any of the above conditions is considered a non-Hyades star. (25 points)

3. The T90 distribution for Beppo-Sax T90 data can be found at `http://www.iith.ac.in/~shantanud/beppoSax.txt`. Apply GMM to $\log_{10}$ of T90 data and find the optimum number of components using AIC and BIC by plotting BIC as a function of number of componts (20 points)
(Hint: Look at the source code for astroML figure 6.6)

EP 4130/PH 6130 Assignment 6

Deadline **10** March 2023 before **17:00 hrs**
Please show the source code for the second and third problems.

1. In 1919, two expeditions sailed from Britain to test if the light deflection from stars agrees with Einstein's General Theory of Relativity. Einstein's theory predicts a value of 1.74 arc-seconds, whereas Newtonian gravity predicts a value exactly half of that. The team by Eddington obtained a value of 1.61 ± 0.40 arc-seconds, while the team by Crommelin reported a value of 1.98 ± 0.16 arc-seconds. Calculate the Bayes factor between General Relativity and Newtonian gravity from those data, assuming Gaussian likelihoods. (10 points)
   (For more information about these expeditions and associated controversies, check out arXiv:0709.0685)

2. For exercise 1 in arXiv:1008.4686, calculate the 68% and 95% joint confidence intervals on b and m.
   (Hint : Either use `emcee` followed by `plot_mcmc` code in `astroML.plotting` or use the `corner` module. Alternately, use the techniques of linear algebra and using the example shown in class during the discussion on frequentist analysis) (20 points)

3. Fit the data in Table 1 of arXiv:1008.4686 to a straight line, after including all the data points, (after ignoring σ_x and ρ_{xy}) using both maximum likelihood analysis and using a Bayesian analysis to identify the outliers, using the same procedure as in the second of Jake VanDerPlas blog article. Show graphically the best fit line using both maximum likelihood analysis and also using Bayesian analysis, including the outlier points. (30 points)

EP 4130/PH 6130 Assignment 7

Deadline **24** March 2023 before **23:59 hrs**
Please show the source code.

1. Download the SPT f_{gas} data from `http://iith.ac.in/~shantanud/fgas_spt.txt`. Fit the data to $f_0(1 + f_1 z)$ where f_0 and f_1 are unknown constants. Determine the best fit values of f_0 and f_1 including 68% and 90% credible intervals using `emcee` and `corner.py` . The priors on f_0 and f_1 should be $0 < f_0 < 0.5$ and $-0.5 < f_1 < 0.5$. (30 pts)

2. Calculate the Bayes factor for the linear and quadratic model for the example given on fifth blog article of the Pythonic Perambulations Series using `dynesty` or `Nestle`. Do the values agree with what's on the blog (obtained by integrating the emcee samples).? (30 points)

3. Download the SDSS quasar dataset from `http://astrostatistics.psu.edu/datasets/SDSS_quasar.dat`. Plot the KDE estimate of the quasar redshift distribution (the column with the title z) using a Gaussian and also an exponential kernel (with bandwidth=0.2) from -0.5 to 5.5. (20 points)
   (Hint: Look at the KDE help page in scikit-learn or use the corresponding functions in astroML module by looking at source code of astroML figures 6.3 and 6.4)

EP 4130/PH 6130 Assignment 8

Deadline **4** April 2023 before **23:59 hrs**
Please show the source code.

1. Calculate the angular two-point correlation function of galaxies (including errors obtained by 10 bootstrap resamples) using subset of data from the Blanco Cosmology Survey with r-band magnitude between 17 and 20, and using 16 logarithmic-spaced angular bins from $1/60^\circ$ to 1°. Use a linear scale for Y-axis. Galaxies in Blanco Cosmology Survey have spread_model > 0.002. This data can be downloaded from `http://www.iith.ac.in/~shantanud/BCS05hr_reduced.txt` (30 points)
   (Hint : Look at the astroML source code for Figure 6.17)