PH 4130/PH 6130 Assignment 2

Deadline **31** January 2018 before **17:00 hrs**

All problems have equal weightage of 30 points each (A bonus question worth 5 points in first question). Please show the source code used for each of the problems. Use of ipython notebooks is encouraged to do assignments. Please don't hesitate to ask questions if you're stuck anywhere.

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 multiplanel 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)
   (a) Are there any distributions such that the samples drawn from those distributions don't follow the Central Limit Theorem?
   (b) Try and check for yourself if samples drawn from Cauchy distribution obey the Central Limit Theorem. Check out `numpy.random.standard_cauchy` to draw samples from Cauchy distribution. (10 points)
   (c) Bonus: How do you sample from a distribution given any arbitrary but well-defined probability density function? It is the standard techinque to sample from exotic distributions. (5 points)

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. (25 points)
   (a) Every algorithm has a time complexity with it. It is used to give the notion of time required for the algorithm as the function of input sizes. In a very loose sense, time complexity can be understood as the time required when the algorithm is run on an infinitely powerful machine.
   (b) Find the time complexities of these "correlation computing routines". Also report the "fastest" of them all. (5 points)
   (c) (Hint: Refer to the documentation)

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). (30 points)
   (Hint : A on the website is same as λ, which was used in class to parameterize the Weibull distribution)
   (a) Further Reading: Given a set of samples and a distribution parameterized by parameters(such as λ of Weibull), how do we find the best parameters?
   (b) Hint: Check out Maximum Likelihood Estimate(MLE)