Problems for AstroHack Session on Bayesian Data Analysis

Problem 1: Probability:

Generate 1e4 values of (x,y) with the limits within (-10,10) using multivariate Gaussian centered at 0 and variances (1, 2). You can assume a correlation between them for an extra challenge. Make a scatter plot of the generated random numbers (x,y). You can make contours to highlight 68% and 90% enclosed points. Mark joint, marginal and conditional (for x=2) probabilities on the graph.

Problem 2 Distributions:

2a. Plot Poisson and Normal distributions to illustrate the Central Limit Theorem.
2b. Plot log-normal distribution
2c. Plot joint probability of two random variables p(x,y) and mark marginal and conditional probabilities
2d. Overplot Normal and t-distributions

Problem 3 Likelihood:

2a. You want to learn about the population of stars in a cluster. Assume that there are two classes of stars: Red Giants and White Dwarfs. What is the fraction of White Dwarfs in the population?

2b. Simulate an array of count data in an observed X-ray spectrum assuming a Poisson process. The energy range of the spectrum is [0.1-10] keV with 100 independent energy bins. The Poisson rate is a power law function of energy in a form of A*E^-(alpha), where A is an amplitude and alpha is a power law slope. Assume alpha = 1.5 for your simulations. Plot the simulated data for a total number of counts of 100 and 1000. How would you define the likelihood for this problem if alpha was a parameter of interest?

Problem 4 Posterior:

4.1 Normal distribution with unknown mean: A random sample of n radio sources is observed from a large population of sources, and their radio fluxes are measured. The average flux of the n sampled sources is y=150 mJy. Assume that the sources are normally distributed with unknown mean theta and standard deviation of 20 mJy. Suppose your prior distribution for theta is normal with mean 180 mJy and standard deviation of 40 mJy.

4a. Provide posterior distribution for theta.
4b. A new source is observed and has a flux m. Provide a posterior predictive distribution for m.
4c. For n=10 (and 100) stars give 95% posterior interval for theta and 95% posterior predictive interval for m.