

Fall 2016 Astro 250: Stellar Populations

Instructor: Dan Weisz (dan.weisz@berkeley.edu)

Problem Set 1 – Intro to Probability and Stats

Assigned: 8/29/16

Due: 9/12/16

All problem sets should be completed in your public github repository. Problems that require written solutions should be completed using LaTeX. Coding exercises should be done in a jupyter notebook. For clarity, you may want to put each coding problem in a separate notebook.

Problem 1.

(a) In 1995, they introduced blue M&M's. Before then, the color mix in a bag of plain M&M's was 30% brown, 20% yellow, 20% red, 10% green, 10% orange, and 10% tan. Afterward, it was 24% blue, 20% green, 16% orange, 14% yellow, 13% red, 13% brown.

Suppose there are two bags of M&M's, one from 1994 and one from 1996 and you are randomly given one M&M from each bag. One is yellow, one is green. Using Bayes's theorem and a probability table to determine the relative probability that the yellow M&M came from the 1994 bag.

(b) Evaluate the "Evidence" and determine the normalized probability that the yellow M&M came from the 1994 bag.

Hint: This is similar to how we wrote out the Monty Hall problem in class.

Problem 2.

Using notes from class, code up your own Metropolis-Hastings (M-H) MCMC sampler. Although it is technically impossible to prove that an MCMC sampler definitively converges (this would require infinite runtime), a simple sanity test it is to sample from a one dimensional Gaussian distribution:

$$P(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(\mu-x)^2}{2\sigma^2}} \quad (1)$$

where μ and σ are values you choose, and x values are generated by your M-H sampler. The density of samples should trace the input distribution. For example, if you select $\mu = 5$ and $\sigma = 1$, the density of samples for x should be a 1d Gaussian with these values (perhaps modulo a normalization constant). Make plots that qualitatively demonstrate convergence of your sampler to a steady state (e.g., $\ln P$ vs. x , $\ln P$ vs. step number) and a plot that shows your samples relative to the true distribution. Your choice in step size should yield an acceptance fraction between ~ 0.25 and 0.5 .

Problem 3.

Using a probabilistic framework, write code that fits a straight line (i.e., $y = mx + b$) to fake data (i.e., data points and error bars). You will need to write a function that simulates fake data that includes Gaussian noise and an arbitrary number of points.

(a) Assume true values of $m = 5$, $b = -2$. Use the M-H MCMC sampler you wrote in problem 2 to infer the true values of m and b for 10, 100, and 1000 data points. Choose a modest amplitude for your uncertainties and clearly indicate your choice. For simplicity, you may assume top hat ('flat') priors for m and b . Make relevant diagnostic plots to indicate convergence, and plot your final results using `corner.py`.

(b) Repeat part (a), but replace your M-H sampler with `emcee`.

bonus problems:

(c) Repeat part (b), but replace the top hat priors with Gaussian priors on m and b . For a fixed, modest uncertainty size, show how your choice in prior affects the posterior in the limit of small N (e.g., $N = 10$) and large N (e.g., $N = 1000$).

(d) Assume that, in addition to Gaussian uncertainties on each data point, the data also have an intrinsic Gaussian scatter such that $\sigma^2 = \sigma_{\text{data}}^2 + \sigma_{\text{scatter}}^2$, where σ_{data} reflects the uncertainty on each data point, and σ_{scatter} is a single value applied to all data points. Modify your code from part (b) to infer the values of m , b , and σ_{scatter} . This will require modifying your fake data generation routine, likelihood function, priors, and calls to `emcee`. *Hint:* For modifying your priors, a so-called non-informative 'Jeffreys Prior' of $P(\sigma) \propto 1/\sigma$, may be helpful.

(e) In the limits of Gaussian uncertainties, σ_{scatter} has an alternative interpretation, i.e., is can be viewed as something other than intrinsic scatter in the data. What is it? (You can answer this in your jupyter notebook, as opposed to opening a new .tex file).