Statistics Mini-course Problem Set 1

Due on Thu. Mar 31

We will do some of the exercises in Hogg, Bovy, & Lang (HBL; 2010) (`1008.4686`), with some slight variations. You should solve these exercises on a computer and the best way to hand in the problem set is as an `ipython notebook`. Rather than sending me the notebook, you can upload it to `GitHub`, which will automatically render the notebook. Rather than starting a repository for a single notebook, you can upload your notebook as a `gist`, which are version-controlled snippets of code.

If you want to upload your notebook as a gist from the command-line, you can use the package at this http URL and use it as follows. Log into your `GitHub` account:

```
gist --login
```

and then upload your notebook `statminicourse_2016_PS1_YOURNAME.ipynb` as

```
gist statminicourse_2016_PS1_YOURNAME.ipynb
```

If you want to make further changes, you can clone your gist in a separate directory and use it as you would any other git repository.

Problem 1: Do exercise 1 in BHL.

Problem 2: Do exercise 1, but assuming that the errors σ_y of neighboring data points in x are correlated with a correlation coefficient of $\rho = 0.5$. E.g., data points 15 and 16 have y measurements whose uncertainty is described by a covariance matrix $\begin{pmatrix} \sigma_{y,15}^2 & 0.5\,\sigma_{y,15}\,\sigma_{y,16} \\ 0.5\,\sigma_{y,15}\,\sigma_{y,16} & \sigma_{y,16}^2 \end{pmatrix}$. Note that the data points in Table 1 are not sorted on x! How does the uncertainty variance σ_m^2 on the slope change?

Problem 3: Write a Metropolis-Hastings sampler for a general one-dimensional probability distribution $p(x)$ with a Gaussian proposal distribution (characterized by a width parameter that should be passed to the code) that returns a sampling and the acceptance fraction. Test it with a Gaussian with zero mean and unit variance: plot a normalized histogram of the samples and compare it to the analytical PDF. Then apply it to sample a probability distribution consisting of the sum of two Gaussians with equal weights, unit variance for each, and means 0 and 10 (again plot a histogram of the samples and the analytical PDF). Try to find a relatively high acceptance fraction.

Problem 4: Solve exercise 6 in HBL using MCMC sampling with `emcee`.