Question Set 4 — Metropolis and Nested Sampling

Question 1

Consider doing basic 'linear regression' - fitting of a straight line to data $\left\{\{(x_i, y_i)\}_{i=1}^{N}, N\right\}$. Make the following assumptions. First, the probability distribution for the data given the parameters:

$$y_i | m, b, \sigma \sim \text{Normal}(mx_i + b, \sigma^2) \tag{1}$$

This is the assumption of 'gaussian noise' around the straight line. Astronomers often assume σ is known and different for each data point, perhaps given as a third column in the data file (the 'error bars'). In this question I'm assuming a constant σ applies to all data points and it is unknown.

Assume the following naive wide priors for the three unknown parameters:

$$m \sim \text{Normal}(0, 1000^2) \tag{2}$$
$$b \sim \text{Normal}(0, 1000^2) \tag{3}$$
$$\ln \sigma \sim \text{Uniform}(-10, 10) \tag{4}$$

Follow these steps to implement this model for the Metropolis algorithm:

1. Copy `transit_model.py` to a new file called `straightline.py`, and work on that.
2. Modify `num_params` and `data` appropriately, and delete un-needed variables.
3. Rewrite `from_prior` and `log_prior`, and modify `jump_sizes`, so they reflect the above prior distributions.
4. Edit `log_likelihood` to be appropriate for this problem.

Run `plain_metropolis.py` and look at trace plots for the parameters. Use the resulting posterior samples to summarise the inferences about m, b, and σ.

For more on fitting straight lines, and the kinds of assumptions that are useful in an astronomy context, see the famous paper by David Hogg.

Question 2

Modify the straight line model so that the noise distribution is potentially heavy tailed:

$$y_i | m, b, \sigma, \nu \sim t(mx_i + b, \sigma^2, \nu) \tag{5}$$

This requires the extra ν parameter. Give it this prior:

$$\ln \nu \sim \text{Uniform}(0,\, 5). \tag{6}$$

Do your inferences about m, b, and σ change much?

Question 3

Run the two different models in Nested Sampling and get their marginal likelihoods. Interpret the conclusions. You might also want to look at the posterior distribution for ν under Model 2's assumptions.