Sampling tutorial (2019)

Exercise 1: Building blocks of sampling algorithms ....

To better understand the accept/reject step of the Metropolis-Hastings algorithm, implement a random walk along the positive and negative x-axis as follows. The starting position of your random walker is arbitrary, you may choose the origin.

1. Provide storage space for a sequence of numbers which will represent the position on the x-axis.
2. Invent an algorithm *A1* which generates random numbers which are either -1 or +1 with the same frequency.
3. Invent an algorithm *A2* which accept a number proposed to it with probability $p \in [0, 1]$.
4. Combine *A1* and *A2* or invent an *A3*. The aim is to accept -1s and +1s with probabilities $1 - p$ and p.
5. Add the accepted random number to the current position of the walker.
6. For different values of p, how does the walker drift aliong the axis? Plot its path: By plotting position against index number, you can see how it walks.
7. Is this random walk Markovian?

Exercise 2: MCMC of Supernova data; measuring Ω_m, w_0, w_a ....

Supernovae of type Ia occur in the late Universe and can be used to constrain dark matter and dark energy. The files 'Supernova_Seed176.txt' and 'Supernova_Seed177.txt' contain two Monte Carlo Markov Chains. These sampled the posterior $\mathcal{P}(\Omega_m, w_0, w_a|\mathbf{x})$, where Ω_m is the dark matter density (it can theoretically take values between 0 and 1). The parameters w_0 and w_a are the equation-of-state parameters for Dark Energy today (subscript zero), and the first Taylor-order of a potential cosmic evolution (if $w_a = 0$, then Dark Energy doesn't evolve). If $w_0 = -1$, then one speaks of 'Dark Energy being a cosmological constant', which is indeed the currently preferred model.
The given chains are already cleaned. For this exercise scatter plots (dots on a canvas) and histograms are sufficient.

1. Identify which of the 4 columns stores the log-likelihood.
2. Plot the posterior in 3d (splot in gnuplot). If you use tiny dots, you will be able to distinguish regions of high and low density. What do these regions indicate?
3. How many 2d marginals are there for 3 parameters? Plot all of them.
4. Identify which of the columns stores which parameter, using the text above. Label the axes accordingly.
5. Find the maximum of the posterior. Which are the most likely values for Ω_m, w_0, w_a?
6. Plot the marginal distribution of Ω_m, i.e. plot $\mathcal{P}(\Omega_m|\mathbf{x})$.

7. For $\Omega_m = 0.1$, judge by eye which values of w_0 are compatible with the data.

8. According to measurements of the cosmic microwave background (CMB), Ω_m is already rather well known: the CMB yields the contraints $\Omega_m = 0.315 \pm 0.017$, with a very Gaussian posterior. Reweigh the supernova chain to combine the experiments. Plot the new, joint posterior.

Exercise 3: Uncleaned MC chain from a cosmic microwave background analysis ....

The file 'Bad_NeutrinoViscosity_Chain.txt' contains an uncleaned Monte Carlo Markov Chain. When opening the file, you will see that it contains more columns than the supernova chain. This is because more parameters were varied this time. A real analysis would contain 10,15,20... further columns, so this is still a very small chain.

1. The chain was created by a sampler which weights repeated points by integer weights. The sampler also stores the log-likelihood. Which columns store the integer weights? Why does MCMC need those weights?

2. Which column stores the log-likelihood?

3. Plot the log-likelihood against interation number. What do the steep drop and the plateau tell you about the quality of the chain? Describe what the sampler does in these two phases.

4. Plot the joint density of 2 arbitrary parameters. Why does the chain look the way it does?

Exercise 4: Monte Carlo and the curse of dimensionality ....

To understand why sampling becomes quickly more efficient than grid computations, we compute all the empty space where a grid would evaluate the likelihood in vain. For this, let d be the dimension of an $\mathbb{R}^d$.

1. Compute the volume $V_c(d)$ of the unit hyper-cube in dimension d (i.e. the hyper-cube of side length unity).

2. Compute the volume $V_s(d)$ of the unit hyper-sphere in dimension d.

3. Plot $V_s(d)/V_c(d)$ as a function of d. What happens to the sphere?

Exercise 5: Gibbs sampling the straight line with dual errors (optional) ....

In the tutorial session on Bayesian inference, we computed the posterior for the slope m of a straight line when estimated from data points with uncertainties in both x and y directions. Here, we shall now Gibbs sample this posterior and plot it. You may use $\hat{x} = 10, \hat{y} = 15$ as synthetic dataset.

1. Why does it make sense to plot a joint posterior of x and m?

2. Code up a Gibbs sampler and generate samples of the joint posterior for x, m. (Reminder: x was a latent variable, which was unobservable. We treated it as a nuisance parameter and marginalized over it to get the posterior of m.)

a) The Gibbs sampler draws alternatingly from the conditional distributions. Here, we have two conditional distributions: conditional on an m-value, and conditional on an x-value. Code up a function that draws random numbers at fixed x; the conditional distribution is here

$$m \sim \mathcal{P}(m|\hat{x}, \hat{y}, x) = \mathcal{G}\left(\frac{\hat{y}}{x}, \frac{1}{x^2}\right). \tag{1}$$

This equation describes that values for m are drawn, for fixed x. But because the sampler will walk around in x, m-space, your function must be able to handle an input-x. So the x is 'fixed per draw', not 'fixed till the end of time'. (If you wish to verify this conditional distribution, you need to take the posterior from the Bayesian Inference tutorial, and pull x out of the brackets in the exponential.)

b) Code up a function that draws random numbers from the second conditional

$$x \sim \mathcal{P}(x|\hat{x}, \hat{y}, m) = \mathcal{G}\left(\frac{\hat{x} + \hat{y}m}{1 + m^2}, \frac{1}{1 + m^2}\right). \tag{2}$$

This function describes that random values for x are generated, keeping m fixed to some input value. Again, the walker will move in x, m-space, hence your function needs to use m as an input variable.

c) Combine your two functions into the Gibbs algorithm: You need to draw alternatingly from x and m, feeding the current values to the next iteration.

3. Collect all points of x, m in a chain. Make a scatter plot of them, and a 2-d histogram $h(i, j)$.

4. Plot a heat-map or any other 2-d map of the posterior for $\mathcal{P}(m, x)$. This will display a banana-shaped joint posterior of x and m.

5. The posterior of m is then the integral over all x-values. You can generate it by ignoring the x-values in the chain, i.e. making a one-dimensional histogram of all m-values, without resolving the m-dependency.

Exercise 6: Computing confidence contours (optional) ....

Confidence contours, or credibility contours, cut a likelihood or posterior at a certain height. If the likelihood/posterior is sampled, then the samples have first to be converted into histograms, and the histogram bins then form the basis of where to cut the likelihood/posterior. Implement this task as follows.

1. Vectorize the 2-d histogram bins, i.e. sort them into an auxilliary array which has only one index $h(k)$ where $k = f(i, j)$ uniquely assignes a single value k to each pair i, j.

2. Rearrange the auxilliary array $h(k)$ such that it is monotonically decreasing with k. Stepping from small to increasing k is now akin to 'walking down the posterior in a spiral'.

3. Compute the total posterior likelihood thus stored in the array (i.e. sum over the array).

4. Compute which heights $h(k)$ demark the level at which 68% and 95% of the total posterior volume have been summed up.

5. Plot confidence contours at those levels.