Markov Chain Monte-Carlo: Metropolis-Hastings Method

The Metropolis algorithm for MCMC is covered in the book. Essentially, it creates a Markov chain on the statespace S by constructing a graph $G = (S, E)$, and then using the following procedure:

- If $X_n = s_i$, then select s_j from s_i's d_i neighbors uniformly. This is called the *selection step*.
- We decide to use the selection s_j with probability
$$\min\left(1, \frac{\pi(s_j)d_i}{\pi_(s_i)d_j}\right)$$
This is called the *acceptance step*.
- If we accepted, then $X_{n+1} = s_j$, otherwise $X_{n+1} = s_i$ (we stay).

We showed that this leads to a Markov chain that has π as its stationary distribution.

The Metropolis algorithm is useful, but there is a generalization which, by providing more freedom in the selection step, allows us to create better algorithms. This called Metropolis-Hastings, and is the same as the usual Metropolis algorithm, but we allow the choice of neighbor in the selection step to be non-uniform. Formally we define it like this:

Let H be a stochastic matrix on S (that is, it has positive elements and its rows sum to one). This is called the *selection kernel*, and must have the properties that if $H(i, j) > 0$ then $H(j, i) > 0$, and that if H was the transition kernel for the Markov chain, that Markov chain would be irreducible. Note that for $s_j \in S$, $H(j, \cdot)$ is a distribution. The Metropolis-Hastings chain is created as follows:

- Selection phase: If $X_n = s_i$ draw s_j from $H(i, \cdot)$.
- Acceptance phase: Accept s_j with probability:
$$\min\left(1, \frac{\pi(s_j)H(j,i)}{\pi(s_i)H(i,j)}\right)$$
- If we accepted, then $X_{n+1} = s_j$, otherwise $X_{n+1} = s_i$ (we stay).

Easy problem: Show that the Metropolis algorithm is a special case Metropolis-Hastings with:

$$H(i,j) = \begin{cases} \frac{1}{d_i} & \text{if } s_i \sim s_j \\ 0 & \text{otherwise.} \end{cases}$$

Slightly harder problem: Show that the Gibbs sampler is a special case of Metropolis Hastings on the space S^V where for $\xi, \xi' \in S^V$:

$$H(\xi, \xi') = \begin{cases} \frac{1}{k}\pi_v(\xi'(v)|\xi) & \text{if } \xi \text{ and } \xi' \text{ differ only at site } v \\ \frac{1}{k}\sum_{v \in V} \pi_v(\xi(v)|\xi) & \text{if } \xi = \xi' \\ 0 & \text{otherwise.} \end{cases}$$

where $k = |V|$ and $\pi_v(s|\xi)$ is the notation I used in class for the π-probability of site v having value s conditioned on all other sites having the value given by ξ.