Additional remarks on Markov Chains and MCMC

As a follow-up to class yesterday, I wanted to make some additional remarks about Markov chains and MCMC, especially in response to some of the excellent questions that were asked.

First of all, regarding Markov chains on graphs, as in the example of the drunk, I commented that the stationary distribution is controlled by the in-degree at each node. This is not quite correct. What I was thinking of was the case of directed graphs in which the in-degree $d(v)$ is equal to the out-degree at every node v of the graph. For such graphs, it is possible to show that the stationary distribution is proportional to the degree $d(v)$ at each node. (I'm asking you to prove something similar for undirected graphs on the next homework.)

This result applies to the example of the drunk where $d(v) = 2$ at every node v, implying that in the long run, the drunk will be in each state with equal probability. Someone asked in class what happens if the drunk always moves one step away from the wall (rather than staying where he is) when he hits the wall. In this case, $d(-2) = d(+2) = 1$, and $d(-1) = d(0) = d(+1) = 2$, so the stationary distribution gives probability $1/8$ to each of the states -2 and $+2$, and probability $1/4$ to each of the other three states.

A stationary distribution exists if the Markov chain satisfies a technical condition called ergodicity. Roughly, this condition rules out certain kinds of degenerate behavior, such as cycling between states, or certain states being unreachable from other states.

There has been a lot of research on how fast a Markov chain "mixes," i.e., reaches its stationary distribution. In general, a Markov chain can be described by a matrix of probabilities giving the probability of moving from state i to state j in a single step. In many cases of interest, the speed with which a Markov chain mixes is controlled by the second largest eigenvalue of this matrix.

Regarding MCMC, here is a different, hopefully more intuitive, way of seeing that $P(\mathbf{X}|\mathbf{e})$ is the stationary distribution for the Markov chain associated with the MCMC algorithm. To show that this distribution is stationary, we need to show that if the current state (vector of assignments to the non-evidence random variables) is already distributed according to $P(\mathbf{X}|\mathbf{e})$, then the next state also will have this same distribution.

So suppose that the current vector of assignments $\mathbf{x}$ to the variables $\mathbf{X}$ already is distributed according to $P(\mathbf{X}|\mathbf{e})$. The MCMC algorithm chooses one component i, and changes X_i according to the distribution $P(X_i|\mathbf{x}_{-i}, \mathbf{e})$. Let us call the resulting vector $\mathbf{x}'$. Given our assumptions, we can think of $\mathbf{x}'$ being constructed according to the following process: first choose $\mathbf{x}'_{-i} = \mathbf{x}_{-i}$, which, by assumption, is distributed according to $P(\mathbf{X}_{-i}|\mathbf{e})$. Then select x'_i according to $P(X_i|\mathbf{x}_{-i}, \mathbf{e})$. But this is exactly the same as choosing $\mathbf{x}'$ according to $P(\mathbf{X}|\mathbf{e}) = P(\mathbf{X}_{-i}|\mathbf{e})P(X_i|\mathbf{X}_{-i}, \mathbf{e})$. So $\mathbf{x}'$ has the same distribution $P(\mathbf{X}|\mathbf{e})$.