Class 14 Agenda: Markov Chains II

1 Announcements

- HW7 due next Wednesday.

2 Questions/Lecture Recap

Any questions or reflections from the quiz or minilectures? (Definitions about Markov chains; stationary distributions; fundamental theorem of Markov chains; Markov Chain Monte Carlo)

3 Queues

In this exercise we'll practice setting up a Markov chain, and analyzing its stationary distribution.

Group Work

Suppose that $\mu, \lambda \in (0, 1)$ so that $\mu + \lambda \leq 1$. Consider a queue of maximum length n, which works like this:

- If there are $< n$ items in the queue, then an item joins the queue with probability λ.
- If there are > 0 items in the queue, then an item gets served and leaves the queue with probability μ.
- With any remaining probability, nothing happens.

1. Let X_t be the length of the queue at time t, and suppose that $X_0 = 0$. Draw the diagram and transition matrix for the Markov chain $X_0, X_1, \ldots$.
2. What is the stationary distribution of this Markov chain, in terms of λ and μ?
   ***Hint**: If π is the stationary distribution, and $X_t \sim \pi$, then for any partition of the states into two sets A and B, we have $\Pr[X_t \in A, X_{t+1} \in B] = \Pr[X_t \in B, X_{t+1} \in A]$. (That is, in the stationary distribution, the probability of crossing from A to B is the same as crossing from B to A....why?). Apply this for $A = \{0, 1, \ldots, i\}$ and $B = \{i+1, \ldots, n\}$ to show that for all i, $\pi_i \lambda = \pi_{i+1} \mu$.*
3. What is the expected amount of time, starting with an empty queue, until the queue is empty again? How does this behave as $n \to \infty$ if (a) $\lambda < \mu$ and (b) $\lambda > \mu$?

4 Gibbs Sampling

In this exercise, we'll explore a special case of MCMC, called "Gibbs Sampling" which arises in many settings in machine learning and language modeling.

Group Work

1. Suppose that π is a joint distribution on X and Y. Suppose that it is hard to sample from π, but relatively easy to sample from $\pi(X|Y=y)$ or $\pi(Y|X=x)$ for any x, y in the support of X and Y respectively.

   Consider the following way to set up a Markov chain $(X_0, Y_0), (X_1, Y_1), \ldots$:

   - Suppose $(X_t, Y_t) = (x, y)$.
   - Draw $x' \sim \pi(X|Y=y)$.
   - Draw $y' \sim \pi(Y|X=x')$.
   - Set $(X_{t+1}, Y_{t+1}) = (x', y')$.

   That is, we first condition on $Y=y$ and draw a new value x' for X, and then we condition on that value x' for X and draw a new value y' for Y.

   Show that π is a stationary distribution for this Markov chain. That is, if P is the transition matrix, show that $\pi P = \pi$.

2. Under what conditions on π does the Fundamental Theorem of Markov chains hold? (That is, what do you need to assume about π to ensure that π is the unique stationary distribution of the Markov chain above, and that the Markov chain will converge to π as $t \to \infty$?)

3. Assuming that those conditions are met, how would you interpret this result in the context of Markov Chain Monte Carlo?

4. Suppose your goal is to create a language model that allows you to sample a uniformly random 7-word sentence from the distribution of naturally occurring 7-word sentences. How could you use Gibbs sampling to do this, and what would the challenges be? [Note: The setup described above for sampling from a joint distribution (X, Y) when you know the univariate distributions $\pi(X|Y=y)$ and $\pi(Y|X=x)$ directly extends to joint distributions of multiple random variables, e.g. $(X^1, X^2, \ldots, X^7)$ provided you know (or can sample from) the univariate distributions $\pi(X^1|X^2=x^2, \ldots, X^7=x^7), \pi(X^2|X^1=x^1, X^3=x^3, \ldots, X^7=x^7)$, etc.]