CS265, Fall 2023

Class 14 Agenda: Markov Chains II

1 Announcements

- HW7 should be out soon, due after break.

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 Gibbs Sampling

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

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.

Group Work

1. With the setup above, show that π is a stationary distribution for this Markov chain.

   Hint: Recall that you want to show that for all x, y,

$$\pi(x, y) = \sum_{x', y'} \pi(x', y') \Pr[(x', y') \to (x, y)]$$

(Why?)

2. Does the Fundamental Theorem of Markov Chains automatically apply in this setting? If not, what additional assumptions do you need to make?
3. This procedure is called "Gibbs Sampling." If it's easy to sample from the marginal distributions, but difficult to sample from π itself, explain why the previous two parts (assuming your assumptions in the previous part are met) give us an algorithm to approximately sample from π. (Don't worry about how efficient the algorithm is for now...)
4. What happens if you apply (an appropriately multivariate) form of Gibbs sampling to the problem of sampling a random proper coloring of a graph? Do you get the same algorithm as we saw in the minilecture, or a different algorithm?

   Note: To do this question you'll have to think about how to extend what we did above to more than two variables.
5. (This one is a bit more open-ended...) 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: Same note as the previous question.
6. Has anyone in your group encountered MCMC before? In what context? If not, what else can you think of that Gibbs sampling or MCMC more generally might be useful for?