Class 14 Agenda: Markov Chains II

1 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)

2 Gibbs Sampling

In this group work, we'll explore a special case of MCMC, called "Gibbs Sampling" which arises in lots of applications.

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 on next page...

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?)

   Hint: Once you recall that you want to show the above, here are some steps to show that.

   - Write down an expression for $\Pr[(x', y') \to (x, y)]$ in terms of $\pi(\cdot, \cdot)$, using the definition of Gibbs sampling.
   - You'll end up with some double-sums. Can you move those sums around productively?

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, and don't overthink this part...].

4. How would you generalize Gibbs sampling to more than two variables? Use your generalization to design an algorithm to sample a uniformly proper coloring of a random graph. Does this give the same Markov chain as the algorithm in the mini-lecture?

5. Has anyone in your group encountered MCMC in general or Gibbs sampling in particular before? If so, in what context? If not, what else can you think of that Gibbs sampling or MCMC more generally might be useful for?