CS 228T QUIZ 1

1. The idea behind basic Monte Carlo integration is to replace intractable expectations $\mathbb{E}_p[f(X)]$ with sums $\sum_{i=1}^n f(x_i)$, where x_i are independent samples from $p(x)$. What potential difficulty in this basic method does Markov chain Monte Carlo address?

2. Consider the following two statements:

i. A Markov chain with transition operator T satisfies detailed balance with respect to a distribution p, *i.e.*,

$$p(x)T(x \to x') = p(x')T(x' \to x)$$

holds for all possible outcomes x, x'.

ii. The distribution p is a stationary distribution for T.

Assume here that the Markov chains mentioned are regular and have finite state space, so they can get from any state to any other state in a finite number of steps with positive probability.

Which of the following is true?

(a) (i) implies (ii), but not vice versa
(b) (ii) implies (i), but not vice versa
(c) (i) if and only if (ii)
(d) (i) and (ii) are unrelated

3. Consider the Bayesian network $X \to Y \to Z$. If the current sample is (x_0, y_0, z_0), and the first substep of Gibbs sampling is to sample y, with what probability will the first subsample be (x_0, y_1, z_0)?

(a) $p(y_1 \mid x_0)$
(b) $p(y_1 \mid x_0, z_0)$
(c) $p(y_1 \mid z_0)$
(d) $p(y_1)$

4. Suppose you have a bipartite MRF with two sets of variables, $\{X_1, \ldots, X_n\}$ and $\{Y_1, \ldots, Y_n\}$. Assume that n is large, e.g. $n = 1000$. Each X_i is connected to each Y_j, none of the Y_j's are connected to each other, and the X_i's are internally connected using a tree structure. Assume that the edges in the tree structure connecting the X_i's induce very strong correlations between the X_i's that they connect. If you are applying collapsed Gibbs sampling to this MRF, which variables should you use as the sampled variables?

(a) The X_i's.
(b) The Y_j's.
(c) All the variables.
(d) Either the X_i's or the Y_j's, it doesn't matter.

Date: September 29, 2020.

5. A difficulty with MCMC methods is that it can be difficult to know when the chain has mixed. There are a number of heuristic strategies that attempt to mitigate this problem. Which of the following is generally a good default approach? (Again, assume we are sampling in high dimensions and that the state space is very large.)

(a) Run one chain for a long time.
(b) Run a large number of chains for a short time, each initialized at different starting values.
(c) Run a few chains for a medium time, each initialized at different starting values.

6. When running an MCMC method, is it better to have to sample more variables or fewer variables?