CS 228, Winter 2011-2012
Problem Set #3

This assignment is due at 12 noon on Tuesday, February 21. Only 1 late day is allowed for this assignment. Submissions should be placed in the filing cabinet labeled "CS228 Homework Submission Box" located in the lobby outside Gates 187.

Assume that we have a calibrated clique tree T for a (unnormalized) distribution P_ϕ.

(a) **[3 points]** Describe how we can generate samples from $P(X_i)$ for any variable X_i . You may assume here and throughout that you have a subroutine for generating samples from any discrete distribution.

(b) **[15 points]** Now assume that our calibrated clique tree's underlying graph structure is a chain Markov Random Field given by X_1–X_2–...–X_n and that the clique tree is of the form C_1–C_2–...–C_{n-1}, where Scope$[C_i] = \{X_i,\ X_{i+1}\}$. We wish to generate a sample over all variables $X_1, X_2, \ldots, X_n$ from $P(X_1, X_2, \ldots, X_n)$. Design an efficient sampling algorithm to do this. (Your algorithm should not explicitly construct $P(X_1, X_2, \ldots, X_n)$ and then sample from it.) Prove that your algorithm correctly generates a sample from $P(X_1, X_2, \ldots, X_n)$; justify every step in your proof.

(c) **[7 points]** We now assume that each variable in the chain can take on d values and that sampling from a distribution with K entries in the joint probability table takes $O(K)$ time.

1. What is the time complexity of sampling from the chain X_1–X_2–...–X_n by explicitly constructing the full joint distribution over the n variables?
2. What is the time complexity of constructing and calibrating a clique tree over this chain?
3. Given that you have a calibrated and normalized clique tree, what is the time complexity of sampling from the chain using the method you described in part (b)?

(d) **[10 points]** Now describe how to generalize your answer in part (b) to the case of a fully general calibrated clique tree with M cliques. It is sufficient to provide pseudocode; you do not need to formally prove the correctness of your algorithm.

(e) **[10 points]** The block Gibbs sampling algorithm is an extension of the Gibbs sampling algorithm, in which we sample sets of variables (instead of a single variable). More precisely, we define sets of variables $\mathbf{X_c}$ and, at each point in the algorithm, sample all of the variables in $\mathbf{X_c}$ given the current assignment to the variables in $\mathbf{X_{-c}}$ (the set of all variables that are not in $\mathbf{X_c}$). In other words, we are sampling from $P(\mathbf{X_c}|\mathbf{X_{-c}})$. This approach is often applied to large sets $\mathbf{X_c}$, allowing us to take larger steps in the space (compared to the standard Gibbs sampling algorithm). For example, if we want to sample from a network that is a large grid, we might choose to sample an entire row at once instead of sampling one node at a time. However, such sets are often too large for an explicit representation of a joint distribution $P(\mathbf{X_c}|\mathbf{X_{-c}})$. In this case, we must come up with a way of generating samples

effectively from this distribution. The sampling algorithms you just constructed provide a standard solution to this difficulty.

Consider an $n \times n$ grid-structured pairwise MRF, where the variable in row i, column j is $X_{i,j}$. Each row has $n-1$ pairwise factors, and so does each column. Let $\phi_r(X_{i,j}, X_{i,j+1})$ be the row pairwise factor over variables $X_{i,j}$ and $X_{i,j+1}$, and let $\phi_c(X_{i,j}, X_{i+1,j})$ be the column pairwise factor over variables $X_{i,j}$ and $X_{i+1,j}$. We will assume that there are no singleton potentials. We want to use block Gibbs to sample from this network, where we will sample entire rows of variables at a time, given assignments to all the other variables in the network.

Explain precisely how you can use the clique tree sampling algorithm (from part (b) or part (d)) to efficiently sample the variables in row i (i.e., $X_{i,1}, X_{i,2}, \ldots, X_{i,n}$), given the assignments to all other variables. You may assume that $i > 1$ and $i < n$. You should clearly define the clique tree structure and state which factors are assigned to each clique. You should also define the scopes of the cliques and the scopes of the sepsets.

(f) **[5 points]** Is there any difficulty in using the block Gibbs strategy from part (e) for overlapping sets of variables if we wanted to sample from rows and then from columns? For example, say we have a 3×3 grid and want to sample row 1, then column 1, then row 2, then column 2, then row 3, and then column 3. Note that the rows and columns have overlapping sets of variables. Is this feasible? Why or why not?