1 Question 1

Consider the "Alarm" Bayesian network from problem 2 of Assignment 3, with the given structure and CPTs. The task is also the same, computing probability of Earthquake given that JohnCalls is True and MaryCalls is False, but by using a sampling approach this time. Of course, the network is small enough that exact inference, by computing the joint distribution, variable elimination or belief propagation is easy; this assignment is thus just to practice with the tools of sampling which may be required when the networks get much larger and more complex.

Write a program to generate a Markov Chain by Gibbs sampling(Algorithm 12.4, page 506 in the KF book). Your program need not be general; it can be specific to the current network, even specific to the given evidence values. Note that the Gibbs sampler requires computation of the probability of a node given its Markov blanket; however, this is straightforward (see equation 12.23) and does not require you to implement a complex exact inference algorithm such as the sum-product algorithm.

Solution:
Exact solutions in order T, F
$P(A|J = T, M = F) = \{0.03025, 0.96975\}$
$P(B|J = T, M = F) = \{0.01227, 0.98773\}$
$P(E|J = T, M = F) = \{0.00462, 0.99538\}$
$P(N|J = T, M = F) = \{0.25921, 0.74079\}$
$P(T|J = T, M = F) = \{0.85153, 0.14847\}$

2 Question 2

Consider a modified Hidden Markov Model where the state transition is a function of not just the previous state but of two previous states (except for the state at t=1). Thus, we are given $P(\mathbf{X}^{(t+1)}|\mathbf{X}^{(t)}, \mathbf{X}^{(t-1)})$. Let the observation model still be a function of the current state only, i.e. we are given $P(\mathbf{O}^{(t)}|\mathbf{X}^{(t)})$. These two distributions, along with the initial distribution, $P(X^{(0)})$ provide a complete parameterization for this model.

1. Show a clique-tree for this network that could be used to make inferences in the rolled-out network would like (as in Figure 15.1, but without the messages that are passed).You only need to show nodes

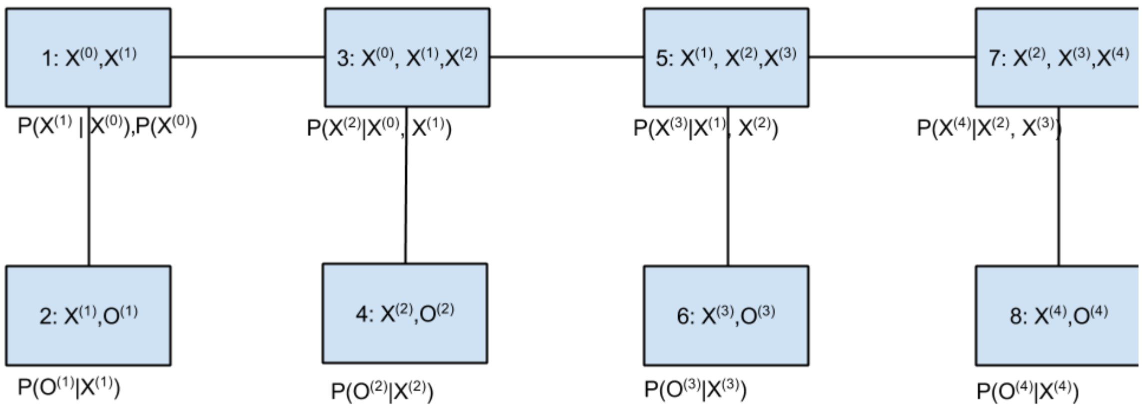

Figure 1: corresponding clique tree

up to some time period, say up to $t = 4$.

Solution: See Figure 1. Notice that clique $X^{(0)}, X^{(1)}$ could be merged with clique $X^{(0)}, X^{(1)}, X^{(2)}$ as well. Both are valid. In this case, both observations at time $t = 1$ (clique $X^{(1)}, O^{(1)}$) and $t = 2$ (clique $X^{(2)}, O^{(2)}$) would be connected to clique $X^{(0)}, X^{(1)}, X^{(2)}$.

2. Derive recursive equations for filtering for this model, i.e.equations for computing $\sigma^{(t)}(\mathbf{X}^{(t)}|\mathbf{O}^{(1:t)})$ in terms of beliefs at earlier time slices. It is reasonable to expect that these equations will be similar to equations (15.1) and (15.2) but a function of two earlier belief states. You may derive these equations using the clique-tree derived in part (a) above or directly by simplifying the probability distribution, following

a procedure similar to that used to derive equations (15.1) and (15.2).

Solution:
$\sigma^{(0)} = P(X^{(0)})$
$\sigma^{(1)}(X^{(1)}, X^{(0)}) = P(X^{(1)}, X^{(0)}|o^{(1)}) \propto P(X^{(1)}|X^{(0)})\sigma^{(0)}(X^{(0)})P(o^{(1)}|X^{(1)})$
$\sigma^{(t+1)}(X^{(t+1)}, X^{(t)}) = P(X^{(t+1)}, X^{(t)}|o^{(1:t+1)}) \propto$
$\Sigma_{x^{(t-1)}} P(X^{(t+1)}|X^{(t)}, x^{(t-1)})\sigma^{(t)}(X^{(t)}, x^{(t-1)})P(o^{(t+1)}|X^{(t+1)})$
If we choose to split up the computation, it will be as follows:
State Prediction:
$\sigma^{(.t+1)}(X^{(t+1)}, X^{(t)}) = P(X^{(t+1)}, X^{(t)}|o^{(1:t)}) =$
$\Sigma_{x^{(t-1)}} P(X^{(t+1)}|X^{(t)}, x^{(t-1)})\sigma^{(t)}(X^{(t)}, x^{(t-1)})$
Observation correction:
$\hat{\sigma}^{(t+1)}(X^{(t+1)}, X^{(t)}) \propto P(o^{(t+1)}|X^{(t+1)})\sigma^{(.t+1)}(X^{(t+1)}, X^{(t)})$
Renormalizing $\hat{\sigma}^{(t+1)}(X^{(t+1)}, X^{(t)})$ gives us $\sigma^{(t+1)}(X^{(t+1)}, X^{(t)})$.