Please return this homework in class, to Transportation 117 by 4:30 pm, or by email to swoh@illinois.edu.

Problem 2.1

In this problem, we will show that when the distribution \( \mu(x) \) is not strictly positive (i.e. \( \mu(x) = 0 \) for some \( x \)), then the I-map for this distribution is not unique. Consider a distribution of 4 binary random variables \( x_1, x_2, x_3, \) and \( x_4 \) such that \( \mu(x_1 = x_2 = x_3 = x_4 = 1) = 0.5 \) and \( \mu(x_1 = x_2 = x_3 = x_4 = 0) = 0.5 \). The following two undirected graphical models are both minimal I-maps for this distribution, hence it is not unique.

\[
\begin{array}{ccc}
3 & 4 \\
2 & 1 \\
\end{array}
\quad
\begin{array}{ccc}
3 & 4 \\
2 & 1 \\
\end{array}
\]

(a) Prove that the two undirected graphical models above are minimal I-maps for the distribution \( \mu(x) \).

You need to show that both graphs are I-maps for the given distribution \( \mu(x) \) and that removing any edge results in introducing independencies that are not implied by the distribution \( \mu(x) \).

(b) Now, we show that starting with a complete graph and eliminating edges that are pairwise conditionally independent does not always give you an I-map (minimal or not). Start with a complete graph \( K_4 \). For each pair of nodes, eliminate the edge between this pair if they are conditionally independent given the rest of the nodes in the graph. Continue this procedure for all pairs of nodes and examine the resulting graph. Is this an I-map of the distribution \( \mu(x_1, x_2, x_3, x_4) \)?

Recall from class, that a distribution over \( x \) is (globally) Markov with respect to \( G = (V, E) \) if, for any disjoint subsets of nodes \( A, B, C \) such that \( B \) separates \( A \) from \( C \), \( x_A \perp x_B \mid x_C \) is satisfied. Recall two other notions of Markovity. A distribution is pairwise Markov with respect to \( G \) if, for any two nodes \( i \) and \( j \) not directly linked by an edge in \( G \), the corresponding variables \( x_i \) and \( x_j \) are independent conditioned on all of the remaining variables, i.e. for all \( (i, j) \notin E \),

\[
x_i \perp x_{V \setminus \{i,j\}} \mid x_j
\]

A distribution is locally Markov with respect to \( G \) if any node \( i \), when conditioned on the variables on the neighbors of \( i \), is independent of the remaining variables, i.e. for all \( i \in V \),

\[
x_i \perp x_{\partial i} \mid x_{V \setminus \{i, \partial i\}}
\]

(c) Using the example of distribution on 4 random variables as a counter example, prove that a distribution is pairwise Markov w.r.t. \( G \) does not always imply that it is locally Markov w.r.t. the same graph \( G \). (However, if the distribution is positive, pairwise Markovity implies local and global Markovity.)
(d) Using the definitions of Markov properties, prove that if a distribution is globally Markov with respect to $G$, then it is locally Markov with respect to $G$.

(e) (Optional) Using the definitions of Markov properties, prove that if a distribution is locally Markov with respect to $G$, then it is pairwise Markov with respect to $G$.

**Problem 2.2**

Consider the (parallel) sum-product algorithm on an undirected tree $T = (V, E)$ with compatibility functions $\psi_{ij}$ such that $\mu(x) = \prod_{(i,j) \in E} \psi_{ij}(x_i, x_j)$. Consider any initialization of messages, which is denoted by $\nu_{i \to j}^{(0)}(x_i)$ for all directions $i \to j$ and all states $x_i$. Messages at step $t \geq 1$ are denoted by $\nu_{i \to j}^{(t)}(x_i)$. In this problem, we will prove by induction that the sum-product algorithm, with the parallel schedule, converges in at most diameter of the graph iterations. (Diameter of the graph is the length of the longest path.)

(a) For $D = 1$, the result is immediate. Consider a graph of diameter $D$. At each time step the message that each of the leaf nodes sends out to its neighbors is constant because it does not depend on messages from any other nodes. Construct a new undirected graphical model $T' = (V', E')$ by stripping each of the leaf nodes from the original graph $T$. Let $\psi_{ij}^{(t)}(x_i, x_j)$ be the compatibility functions for the new graphical model, and $\nu_{i \to j}^{(t)}(x_i)$ be the messages of (parallel) sum-product algorithm on the new graphical model. Let $\mathcal{L}$ be the set of leaves in $T$ and $\mathcal{L}'$ be the set of nodes that is adjacent to a node in $\mathcal{L}$. For the new graphical model, we add, for all $i \in \mathcal{L}'$,

$$\psi_{ij}^{(t)}(x_i) = \psi_i(x_i) \prod_{k \in \partial i \cap \mathcal{L}} \sum_{x_k} \nu_{k \to i}^{(0)}(x_k) \psi_{k \to i}(x_k, x_i)$$

where $\psi_i(x_i) = 1$ if $\psi_i(x_i)$ is not defined for the original graph $G$ and for all other edges we keep the original compatibility functions

$$\psi_{ij}^{(1)}(x_i, x_j) = \psi_{ij}(x_i, x_j).$$

Also we initialize the messages as

$$\nu_{i \to j}^{(0)}(x_i) = \nu_{i \to j}^{(1)}(x_i).$$

Show that $\nu_{i \to j}^{(t)}(x_i) = \nu_{i \to j}^{(t+1)}(x_i)$ for all $(i, j) \in E'$ and all $t \geq 0$.

(b) Argue that $T'$ has diameter strictly less than $D - 1$.

(c) By the induction assumption that the parallel sum-product algorithm converges to a fixed point after at most $d$ time steps when the diameter is $d \leq D - 1$, the sum-product algorithm on $T'$ converges after at most $D - 2$ time steps. Show that if we add back the leaf nodes into $T'$ and run (parallel) sum-product algorithm for one more time step, then all messages will have converged to a fixed point.

**Problem 2.3**

For $\ell \in \mathbb{N}$, let $G_\ell = (V_\ell, E_\ell)$ be an $\ell \times \ell$ two-dimensional grid. We consider an Ising model on $G_\ell$ with parameters $\theta = \{\theta_{ij}, \theta_i : (i, j) \in E_\ell, i \in V_\ell\}$. This is the probability distribution over $x \in \{+1, -1\}^{V_\ell}$

$$\mu(x) = \frac{1}{Z_G} \exp \left\{ \sum_{(i,j) \in E_\ell} \theta_{ij} x_i x_j + \sum_{i \in V_\ell} \theta_i x_i \right\}$$  \hspace{1cm} (1)$$

\text{1} Namely $V_\ell = [\ell] \times [\ell]$ and, for any two vertices $i, j \in V_\ell$, $i = (i_1, i_2)$, $j = (j_1, j_2)$, $i_1, i_2, j_1, j_2 \in [\ell]$, $(i, j) \in E_\ell$ if and only if $i_1 = j_1$ and $|i_2 - j_2| = 1$, or $i_2 = j_2$ and $|i_1 - j_1| = 1$. $2$
(1) Write the belief propagation (BP) update equations for this model. Also write the update equation for the log-likelihood ratio

\[
L_{i \rightarrow j}^{(t)} = \frac{1}{2} \log \left( \frac{\nu_{i \rightarrow j}^{(t)}(+1)}{\nu_{i \rightarrow j}^{(t)}(-1)} \right)
\]

(2) Write a program that implements these update. You are requested to return a printout of the code (Matlab, C, C++, Java, . . . , are accepted). Feel free to download and start from the skeleton in bp.m from the course website.

(3) Consider the case \( \ell = 10 \) (and hence \( n = 100 \) nodes). For each \( \beta \in \{0.2, 0.4, \ldots, 2.8, 3.0\} \), generate an instance by drawing \( \theta_i, \theta_{ij} \) uniformly random in \([0, \beta]\). Run the BP iteration and monitor convergence by computing the quantity

\[
\Delta(t) \equiv \frac{1}{|\tilde{E}_\ell|} \sum_{(i,j) \in \tilde{E}_\ell} \left| \nu_{i \rightarrow j}^{(t+1)}(+1) - \nu_{i \rightarrow j}^{(t)}(+1) \right|
\]

Here \( \tilde{E}_\ell \) denotes the set of directed edges in \( G_\ell \), in particular \( |\tilde{E}_\ell| = 2 |E_\ell| \).

Plot \( \Delta(t = 15) \) and \( \Delta(t = 25) \) versus \( \beta \), for the random instances generated with \( \beta \in \{0.2, 0.4, \ldots, 2.8, 3.0\} \). Comment on the results.

(4) Repeat the calculation at the precious point, with now \( \theta_i, \theta_{ij} \) uniformly random in \([-\beta, +\beta]\), with \( \beta \in \{0.2, 0.4, \ldots, 2.8, 3.0\} \). Comment on the results.

**Problem 2.4**

In this problem, you will implement the sum-product algorithm on a line graph and analyze the behavior of S&P 500 index over a period of time. The following figure shows the price of S&P 500 index from January 2, 2009 to September 30, 2009 (http://finance.yahoo.com).

![Price of S&P 500 index from January 2, 2009 to September 30, 2009](http://finance.yahoo.com)

For each week, we measure the price movement relative to the previous week and denote it using a binary variable (+1 indicates up and 1 indicates down). The price movements from week 1 (the week of January 5) to week 39 (the week of September 28) are plotted below:

Consider a hidden Markov model in which \( x_t \) denotes the economic state (good or bad) of week \( t \) and \( y_t \) denotes the price movement (up or down) of the S&P 500 index. We assume that \( x_{t+1} = x_t \) with probability 0.8, and \( P_{Y_t \mid X_t}(y_t = +1 \mid x_t = \text{‘good’}) = P_{Y_t \mid X_t}(y_t = -1 \mid x_t = \text{‘bad’}) = q \). In addition, assume
that \( \mathbb{P}(X_1 = \text{'bad'}) = 0.8 \). Download the file `sp500.mat` from course website, and load it into MATLAB. The variable `price.move` contains the binary data above. Implement the (sequential) sum-product algorithm and submit a hardcopy of the code (you dont need to include the code for loading data, generating figures, etc.).

(a) Assume that \( q = 0.7 \). Plot \( \mathbb{P}(X_t = \text{'good'}|y) \) for \( t = 1, 2, \ldots, 39 \). What is the probability that the economy is in a good state in the week of September 28, 2009 (week 39)?

(b) Repeat (a) for \( q = 0.9 \). Compare the results of (a) and (b).