Problem Set 3
Midterm Exam

Prof. Jim Rehg
CS 7636 Computational Perception
College of Computing
Georgia Institute of Technology

March 2, 2004

Solutions must be submitted at beginning of class on Wednesday, March 3, 2004. This is an open-book exam. Hand-written answers are fine as long as they are legible and organized. Please show all work. I’ve tried to leave space on this handout for your solutions. The midterm is worth 15% of your total grade.
1. [5% of grade]
Consider the following joint probability distribution \( p(x, y, z, w) \) for discrete 3-valued random variables:

<table>
<thead>
<tr>
<th>( p )</th>
<th>( x )</th>
<th>( y )</th>
<th>( z )</th>
<th>( w )</th>
</tr>
</thead>
<tbody>
<tr>
<td>1/3</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>1/3</td>
<td>1</td>
<td>2</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>1/3</td>
<td>2</td>
<td>2</td>
<td>1</td>
<td>3</td>
</tr>
<tr>
<td>0</td>
<td>all other tuples</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Let \( G = (V, E) \) be the undirected graphical model with \( V = \{x, y, z, w\} \) and \( E = \{(x, y), (y, w), (w, z), (x, z)\} \).

(a) Show that for disjoint subsets \( A, B, C \) of \( V \), \( A \) sep \( B \mid C \) \( \Rightarrow \ A \perp \perp B \mid C \). Show that \( G \) is the minimal undirected graph with this property (i.e. no edge can be removed from \( E \)).

(b) Show that \( p(x, y, z, w) \) cannot be expressed as the product of potential functions defined over the cliques of \( G \).
(c) Draw a junction tree for $p$ and express $p$ as a product of the junction tree potentials.

(d) Draw a Bayesian network model for $p$ and give the parameters.
2. [5% of grade]
The boundary of a node \( \alpha \), written \( \text{bd}(\alpha) \), is the set of parents and neighbors of \( \alpha \).
In other words, if the graph contains the directed edge \( \beta \to \alpha \) or the undirected edge \( \beta \sim \alpha \), then \( \beta \in \text{bd}(\alpha) \). The closure of a node \( \alpha \) is defined as \( \text{cl}(\alpha) = \alpha \cup \text{bd}(\alpha) \). The Markov blanket of a node \( \alpha \) in a directed graph, written \( \text{bl}(\alpha) \), is the set consisting of the parents of \( \alpha \), the children of \( \alpha \), and the nodes sharing a child with \( \alpha \). Note that for an undirected graph we have \( \text{bl}(\alpha) = \text{bd}(\alpha) \).

(a) For the following graphs, identify \( \text{cl}(A) \) and \( \text{bl}(A) \), and show the corresponding moralized graph.
(b) Identify whether the following statements are true or false. If true, explain why. If false, explain by giving a counter-example. Assume in all cases that \( \alpha \) is a node in the graph, and that the graph and its associated distribution obey the appropriate global Markov property (i.e. for a directed graph, d-separation implies conditional independence; for an undirected graph, separation implies conditional independence).

i. For an undirected graph, \( \alpha \perp \perp V \setminus \text{cl}(\alpha) \mid \text{bd}(\alpha) \)

ii. For a directed graph, \( \alpha \perp \perp V \setminus \text{cl}(\alpha) \mid \text{bd}(\alpha) \)

iii. For a directed graph, \( \alpha \perp \perp V \setminus \{\alpha \cup \text{bl}(\alpha)\} \mid \text{bl}(\alpha) \)
iv. For a directed graph, let $M$ denote the corresponding moralized graph. Then
\[ \alpha \perp \perp V \setminus \text{cl}_M(\alpha) \mid \text{bd}_M(\alpha), \]
where $\text{bd}_M(\alpha)$ and $\text{cl}_M(\alpha)$ refer to the boundary and closure of $\alpha$ with respect to the moralized graph $M$. 
3. [5% of grade]
Suppose that the color pixel value \( z \) is distributed according to:

\[
(1) \quad z = \bar{z} + z_0 \\
(2) \quad p(\bar{z}|\theta) = \sum_{i=1}^{m} \alpha_i N(\mu_i, \Sigma) \\
(3) \quad p(z_0|\theta) = N(\mu_0, \Sigma_0)
\]

This is a standard mixture density for color, with the addition of an offset term \( z_0 \).

(a) List the specific parameters in the vector \( \theta \) from the model. What is the total number of parameters (i.e. how many distinct numbers need to be specified)?
Draw a Bayesian network that models the process of generating training data \( \{z_i\}_{i=1}^{N} \) from Equation 1. Include the parameters in your model.

(b) Given a large number of iid pixel observations \( \{z_i\}_{i=1}^{N} \), we wish to obtain maximum likelihood estimates of the model parameters. Derive the E and M steps of an EM algorithm for estimating \( \theta \). Show and justify all of the steps in your derivation, beginning from first principles (i.e. identification of hidden variables, observed variables, etc.)
(c) Write the lower bounding auxiliary function $\mathcal{L}(q, \theta)$ for this problem (e.g. substitute the parameteric form of the above model into the general expression from Equation 11.8 in the text).

(d) As in the usual mixture density case, the M-step can be solved in closed form. In other words, your answer to part (b) provides an explicit formula for $\theta^* = \arg \max_\theta \mathcal{L}(q, \theta)$. Explain the fallacy in the following (incorrect) argument:

"We have that $\mathcal{L}(q^*, \theta) = l(\theta; x)$ for all possible $\theta$. It follows in particular that $\mathcal{L}(q^*, \theta^*) = l(\theta^*; x)$. Since we can solve in closed-form for the $\theta^*$ which is the global maximum of $\mathcal{L}(q^*, \theta)$, only one iteration of the M-step is required. The result is the global maximum of $l(\theta, x)$."
4. [Extra Credit: +1% of grade]
Let \( P(X) \) be a probability distribution and let \( G = (V, E) \) be a connected undirected graph over the variables \( X \). We say that the distribution \( P(X) \) is \textit{pairwise Markov} (which we write \( P \)) with respect to \( G \) if any two nodes \( \alpha \) and \( \beta \) that are not directly connected in \( G \) are conditionally independent given the rest of the nodes. More compactly, if \( P \), then \((\alpha, \beta) \notin E \implies \alpha \perp \perp \beta \mid V \setminus \{\alpha, \beta\}\).

(a) Show that the global Markov property implies \( P \). Suppose that \( A \) sep \( B \mid C \implies A \perp \perp B \mid C \) for all disjoint sets \( A, B, C \). Show that if \( P(\alpha \mid X \setminus \alpha) \neq P(\alpha \mid X \setminus \{\alpha, \beta\}) \), then \((\alpha, \beta) \in E \) (pairwise Markov property), where \( \alpha \) and \( \beta \) are individual nodes. (This should be straightforward).

(b) In section 4.5 of chapter 4, it is shown that the pairwise Markov property implies the global Markov property. Here you are given the easier task of proving that the pairwise Markov property implies the local Markov property. You may use similar arguments, but don’t simply cite the book’s results (e.g. pairwise \( \Rightarrow \) global \( \Rightarrow \) local).

Let \( P(X) \) be a \textit{positive} probability distribution (\( P(X) > 0 \ \forall X \)) and let \( G = (X, E) \) be a graph with the property that if \( P(\alpha \mid X \setminus \alpha) \neq P(\alpha \mid X \setminus \{\alpha, \beta\}) \), then \((\alpha, \beta) \in E \). Show that any variable \( \alpha \) is independent of \( X \setminus \text{cl}(\alpha) \) given \( \text{bd}(\alpha) \), where \( \text{bd}(\alpha) \) is the set of neighbors of \( \alpha \) and \( \text{cl} = \alpha \cup \text{bd}(\alpha) \).