Statistics 251/551
2004: Solutions to sheet 3

(3.1) In general,

$$\mathbb{P}\{X = x \mid Y = y\} = \sum\nolimits_z \mathbb{P}\{X = x \mid Y = y, Z = z\}\mathbb{P}\{Z = z \mid Y = y\}.$$

If $\mathbb{P}\{X = x \mid Y = y, Z = z\} = g(x, y)$ for all z, it factorizes out from each summand on the right-hand side leaving $g(x, y)\sum_z \mathbb{P}\{Z = z \mid Y = y\} = g(x, y)$.

(3.2) The specification of the V_A functions is a little misleading. If you take it literally, you will find two distinct configurations (all 0's or all 1's) each with probability 1. Of course, we should divide by the constant

$$K = \sum\nolimits_x \prod\nolimits_{A\in\mathcal{C}} V_A(x)$$

to make the Gibbs specification a proper probability distribution. The proof of the Hammersley-Clifford theorem shows that we can always choose the V_A so that $K = 1$. If the V_A are specified as in the Problem then the constant is needed. Fortunately, in calculations of conditional probabilities the constant cancels out. For example,

$$\begin{aligned}\mathbb{P}\{X_5 = 1 \mid X_2 = 0, X_4 = 0, X_6 = 0, X_8 = 0\}& \\ &= \frac{V_{2,5}(0,1)V_{4,5}(0,1)V_{5,6}(1,0)V_{5,8}(1,0)}{\sum_{b=0}^{1} V_{2,5}(0,b)V_{4,5}(0,b)V_{5,6}(b,0)V_{5,8}(b,0)} \\ &= \frac{0.5^4}{1^4 + 0.5^4} = \frac{1}{17}\end{aligned}$$

and

$$\mathbb{P}\{X_9 = 1 \mid X_6 = 0, X_8 = 0\} = \frac{V_{6,9}(0,1)V_{8,9}(0,1)}{\sum_{b=0}^{1} V_{6,9}(0,b)V_{8,9}(0,b)} = \frac{0.5^2}{1^2 + 0.5^2} = \frac{1}{5}$$

(3.3) Suppose the probabilities were given by a Gibbs distribution. The following table shows the arguments of the V_A for each of the four cliques for each of the eight configurations with nonzero probability. I have omitted entries that duplicate an earlier entry in the same column, in order that it be quite clear that all four pairs of 0's and 1's appear as arguments to each V_A.

x	$V_{1,2}$ args	$V_{2,3}$ args	$V_{3,4}$ args	$V_{4,1}$ args
0000	00	00	00	00
1000	10			01
1100	11	10		
1110		11	10	
0001			01	10
0011		01	11	
0111	01			
1111				11

It follows that we would have $V_A(i, j) > 0$ for all A all i, j. However, that would force

$$\mathbb{P}\{X_1 = 1, X_2 = 0, X_3 = 1, X_4 = 0\} = V_{1,2}(1,0)V_{2,3}(0,1)V_{3,4}(1,0)V_{4,1}(0,1) > 0,$$

which is inconsistent with the assigned probabilities.

To verify the Markov property we need to check a big bunch of equalities of the form

$$(*)\qquad \mathbb{P}\{X_1 = i_1 \mid X_2 = i_2, X_3 = i_3, X_4 = i_4\} = \mathbb{P}\{X_1 = i_1 \mid X_2 = i_2, X_4 = i_4\}$$

Strictly speaking, it is only necessary to consider those triples (i_2, i_3, i_4) for which $\mathbb{P}\{X_2 = i_2, X_3 = i_3, X_4 = i_4\} > 0$. Also, it is enough to consider the case $i_1 = 1$, because the analogous assertions for $i_1 = 0$ then follow by subtraction from 1.

For $(i_2, i_4) = (0, 0)$ or $(i_2, i_4) = (1, 1)$ assertion $(*)$ is trivial because

$$\{X_2 = 0, X_3 = 0, X_4 = 0\} = \{X_2 = 0, X_4 = 0\}$$
$$\{X_2 = 1, X_3 = 1, X_4 = 1\} = \{X_2 = 1, X_4 = 1\}$$

For $(i_2, i_4) = (1, 0)$ or $(i_2, i_4) = (0, 1)$ assertion $(*)$ is also trivial, because

$$\mathbb{P}\{X_1 = 1 | X_2 = 1, X_4 = 0\} = 1 = \mathbb{P}\{X_1 = 0 | X_2 = 0, X_4 = 1\}$$

I should also check the analogous assertions for $\mathbb{P}\{X_2 = 1 \mid X_1 = i_1, X_3 = i_3, X_4 = i_4\}$, and so on.

It is possible to reduce the amount of tedious detail by checking an equivalent form of the Markov property. In general, random variables X and Y are said to be ***conditionally independent*** given another random quantity Z if

$$(**)\qquad \mathbb{P}\{X = x, Y = y \mid Z = z\} = \mathbb{P}\{X = x \mid Z = z\}\mathbb{P}\{Y = y \mid Z = z\}$$

for all x, y and z for which $\mathbb{P}\{Z = z\} > 0$. If $(**)$ holds $\mathbb{P}\{Y = y, Z = z\} > 0$, then

$$\begin{aligned}
\mathbb{P}\{X = x \mid Y = y, Z = z\} &= \frac{\mathbb{P}\{X = x, Y = y, Z = z\}}{\{\mathbb{P}\{Y = y, Z = z\}} \\
&= \frac{\mathbb{P}\{X = x, Y = y \mid Z = z\}\mathbb{P}\{Z = z\}}{\mathbb{P}\{Y = y \mid Z = z\}\mathbb{P}\{Z = z\}} \\
&= \frac{\mathbb{P}\{X = x \mid Z = z\}\mathbb{P}\{Y = y \mid Z = z\}}{\mathbb{P}\{Y = y \mid Z = z\}} \qquad \text{by } (**) \\
&= \mathbb{P}\{X = x \mid Z = z\}
\end{aligned}$$

The MRF property for X_1, X_2, X_3, X_4 is equivalent to a pair of conditional independences: X_1 independent of X_3 given the random vector (X_2, X_4); and X_2 independent of X_4 given the random vector (X_1, X_3). That is, we need to check

$$\mathbb{P}\{X_1 = i_1, X_3 = i_3 \mid X_2 = i_2, X_4 = i_4\} = \mathbb{P}\{X_1 = i_1 \mid X_2 = i_2, X_4 = i_4\}\mathbb{P}\{X_3 = i_3 \mid X_2 = i_2, X_4 = i_4\}$$
$$\mathbb{P}\{X_2 = i_2, X_4 = i_4 \mid X_1 = i_1, X_3 = i_3\} = \mathbb{P}\{X_2 = i_2 \mid X_1 = i_1, X_3 = i_3\}\mathbb{P}\{X_4 = i_4 \mid X_1 = i_1, X_3 = i_3\}$$

In every case, the factorization is trivial because one of the random variables has a degenerate conditional distrbution. For example, $\mathbb{P}\{X_3 = 0 \mid X_2 = 0, X_4 = 0\} = 1$.

(3.4) It is important to realize what is not being suggested in this problem. If a particular graph structure on $\mathcal{G}_X \cup \mathcal{G}_Y$ is to make (X, Y) a MRF, then a particular set of conditional independences is required. However, we could have a MRF without requiring particular dependences. For example, if all components of (X, Y) are independent then it is a MRF for any choice of edges for the graph, including the extreme case where there are no edges or the other extreme where every pair of nodes is connected by an edge.

Consider once more the general situation of a MRF $W_1 = (Z_1, Z_2, \ldots, Z_n)$ on a graph with nodes $\mathcal{G}_1 = \{g_1, g_2, \ldots, g_n\}$ and edges $\mathcal{E}_1$. The random variables $W_2 = (Z_2, \ldots, Z_n)$ can be identified with the nodes $\mathcal{G}_2 = \{g_2, \ldots, g_n\}$. If we remove from $\mathcal{E}_1$ all edges with g_1 as a vertex we are left with a set of edges $\mathcal{E}_1'$ joining nodes in $\mathcal{G}_1$. As explained in class, it need not be true that W_2 is a MRF for $(\mathcal{G}_2, \mathcal{E}_1')$. However, W_2 is a MRF for a new graph $(\mathcal{G}_2, \mathcal{E}_2)$, where $\mathcal{E}_2$ is obtained by adding to $\mathcal{E}_1'$ new edges $\{g_i, g_j\}$ for all nodes with $g_i \sim_1 g_1 \sim_1 g_j$, where the symbol $\sim_1$ refers to an edge in the original $\mathcal{E}_1$. That is, we define $\{g_i, g_j\}$, with $i, j \geq 2$, to be an $\mathcal{E}_2$-edge if either (maybe both) of the following conditions is satisfied: (i) $g_i \sim_1 g_j$ or (ii) $g_i \sim_1 g_1 \sim_1 g_j$.

Now consider $W_3 = (Z_3, \ldots, Z_n)$ with nodes $\mathcal{G}_3 = \{g_3, \ldots, g_n\}$. We can be sure that W_3 is a MRF for $(\mathcal{G}_3, \mathcal{E}_3)$ if we define $\mathcal{E}_3$ to contain all those $\{g_i, g_j\}$, with $i, j \geq 3$, for which: (i) $\{g_i, g_j\} \in \mathcal{E}_2$, or (ii) $\{g_i, g_2\} \in \mathcal{E}_2$ and $\{g_2, g_j\} \in \mathcal{E}_2$.

Suppose $\{g_i, g_j\}$ is a newly created $\mathcal{E}_3$-edge. Then we know both nodes were connected to g_2 in $\mathcal{E}_2$, that is $g_i \sim_2 g_2 \sim_2 g_j$. By definition of $\mathcal{E}_2$-edges, either $g_i \sim_1 g_2$ or $g_i \sim_1 g_1 \sim_1 g_2$, with a similar pair of possibilities for g_j. If you track through all four pairs of possibilities you will see that one of the following must be true:

$$g_i \sim_1 g_2 \sim_1 g_j \text{ or } g_i \sim_1 g_2 \sim_1 g_1 \sim_1 g_j \text{ or } g_i \sim_1 g_1 \sim_1 g_2 \sim_1 g_j \text{ or } g_i \sim_1 g_1 \sim_1 g_j$$

And so on.

A formal inductive argument would hypothesize that $g_i \sim_k g_j$ if and only if either $g_i \sim_1 g_j$ or there is a chain $g_i \sim_1 a_1 \sim_1 \ldots \sim_1 a_m \sim_1 g_j$, where $\{a_1, \ldots, a_m\} \subseteq \{g_1, \ldots, g_k\}$. If (g_i, g_j) were newly created as an $\mathcal{E}_{k+1}$-edge we would have $g_i \sim_k g_k$ and $g_k \sim_k g_j$. By the inductive hypothesis, there exist nodes from $\{g_1, \ldots, g_{k-1}\}$ for which

$$g_i \sim_1 a_1 \sim_1 \ldots \sim_1 a_m \sim_1 g_k \sim_1 b_1 \sim_1 \ldots \sim_1 b_\ell \sim_1 g_j$$

Thus g_i and g_j are joined by a chain through $\{g_1, \ldots, g_k\}$. And so on.

The first part of the question differs only notationally from the problem I have just solved.

The story for $\mathbb{P}\{X = x \mid Y = y\}$ is much easier. We may write the original Gibbs distribution as

$$\mathbb{P}\{X = x, Y = y\} = \prod\nolimits_{A \in \mathcal{C}} V_A(x, y)$$

where $\mathcal{C}$ is the set of cliques for the graph on $\mathcal{G}_X \cup \mathcal{G}_Y$. The conditional distribution is

$$\mathbb{P}\{X = x \mid Y = y\} = \prod\nolimits_{A \in \mathcal{C}} V_A(x, y) / \mathbb{P}\{Y = y\}$$

Remember that we treat y as a constant. As a function of x, the factor $V_A(x, y)$ depends on x only through the values for nodes in $\mathcal{G}_X \cap A$. Moreover, each pair of nodes in $\mathcal{G}_X \cap A$ is connected by an edge. Thus, the conditional distribution is already expressed as a Gibbs distribution, a product over complete subsets of $\mathcal{G}_X$. We do not need to create any extra edges between nodes in $\mathcal{G}_X$ to make the conditional probability a MRF for each fixed y.