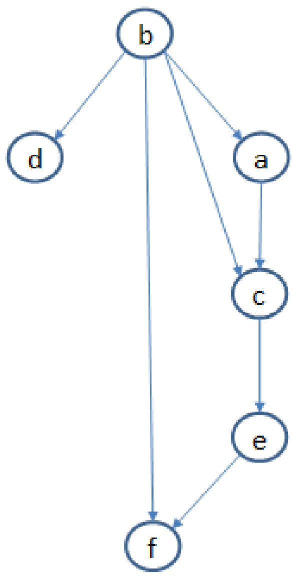
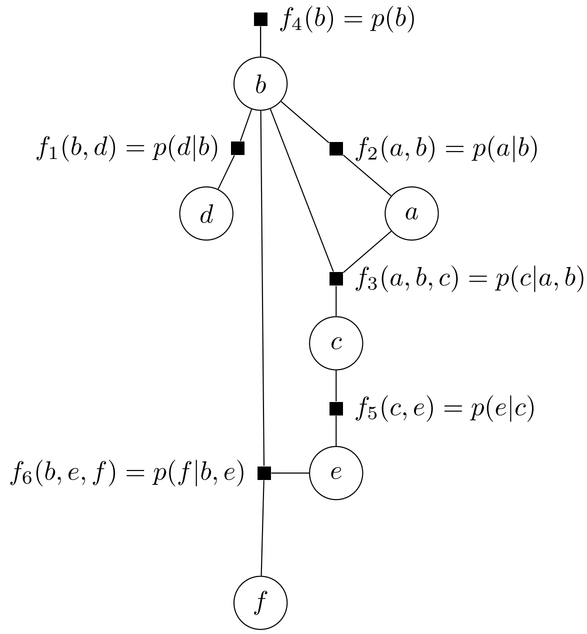
ECE521 Assignment 4 Example solutions

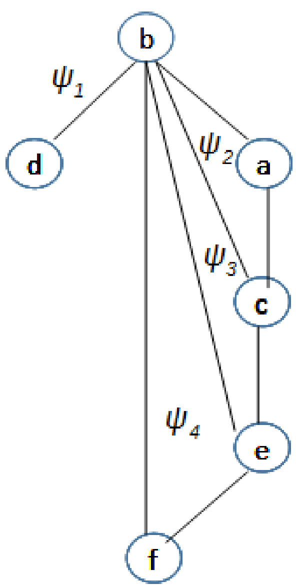
April 2017 - revised

Question 1.1

The Bayesian network, factor graph, and MRF are below. In the factor graph, f_2 and f_3 can be combined, and/or f_4 can be folded into other factors.



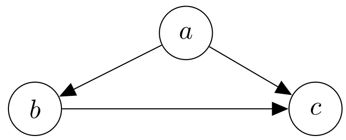




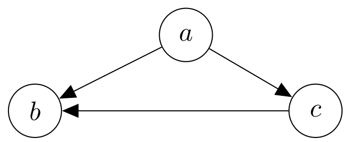
$$\psi_1 = \psi_2(b,d) = p(b)p(d|b)$$
$$\psi_2 = \psi_2(a,b,c) = p(a|b)p(c|a,b)$$
$$\psi_3 = \psi_3(b,c,e) = \psi_3(c,e) = p(e|c)$$
$$\psi_4 = \psi_4(b,e,f) = p(f|b,e)$$

Question 1.2

1.1 (a) Nodes a, b and c are all dependent on each other, and each node is conditionally dependent on either of the remaining nodes given the third. There are nine possibilities for the equivalent Bayesian networks. One possible BN is to let $f_2 \propto p(a)$ and $f_1 \propto p(c|a,b)p(b|a)$. The resulting BN graph looks like:

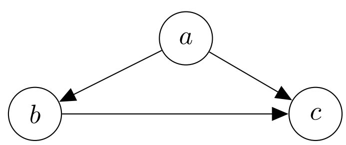


Another possible BN is to let $f_1 \propto p(c|a)p(b|a,c)$, and the graph is:



The global conditional independence properties of the above also match those for the given factor graph.

1.1 (b) Similar to factor graph (a), all three nodes in factor graph (b) (a,b and c) are condtionally dependent. The equivalent BNs for (b) are therefore the same as for factor graph (a), which has nine possible BN conversions. One possible equivalent BN is:



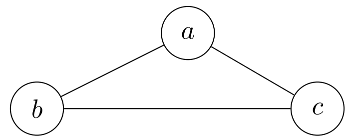
The conditional probabilities in general do not correspond to the factors from the factor graph. We start with the joint probability represented by the factor graph:

$$p(a,b,c) = \frac{1}{Z} f_1(b) f_2(c) f_3(a,b) f_4(a,c) f_5(b,c).$$

The conditional probabilities represented by the equivalent BN are obtained through mariginalization:

$$p(a) = \frac{1}{Z} \sum_{b,c} f_1(b) f_2(c) f_3(a,b) f_4(a,c) f_5(b,c)$$
$$p(b|a) = \frac{\sum_c f_1(b) f_2(c) f_3(a,b) f_4(a,c) f_5(b,c)}{\sum_{b,c} f_1(b) f_2(c) f_3(a,b) f_4(a,c) f_5(b,c)}$$
$$p(c|a,b) = \frac{f_1(b) f_2(c) f_3(a,b) f_4(a,c) f_5(b,c)}{\sum_c f_1(b) f_2(c) f_3(a,b) f_4(a,c) f_5(b,c)}$$

1.2 (a) Nodes a, b and c are all dependent on each other, and each node is conditionally dependent on either of the remaining nodes given the third. The MRF is:

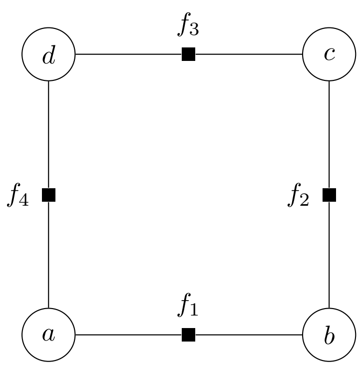


The global conditional independence properties (i.e. Markov blankets) of the above also match those of the given factor graph. The clique potential is $\psi_1(a, b, c) = f_1(a, b, c)f_2(a)$.

1.2 (b) The MRF is as in part (a), and the clique potential is

$$\psi_1(a, b, c) = \prod_{i=1}^{5} f_i = f_1(b)f_2(c)f_3(a, b)f_4(a, c)f_5(b, c)$$

2.



The factor graph is at left. An equivalent Bayesian network doesn't exist: in order to be a DAG (tree), there must exist a V structure. No UGM can model a V-structure in a DGM (see Murphy 19.2.3). For example, suppose the middle of the V occurs at node b, and the question is whether $a \perp\!\!\!\perp c|b, d$. Using the DGM, they are conditionally dependent. Using the UGM, they are not.

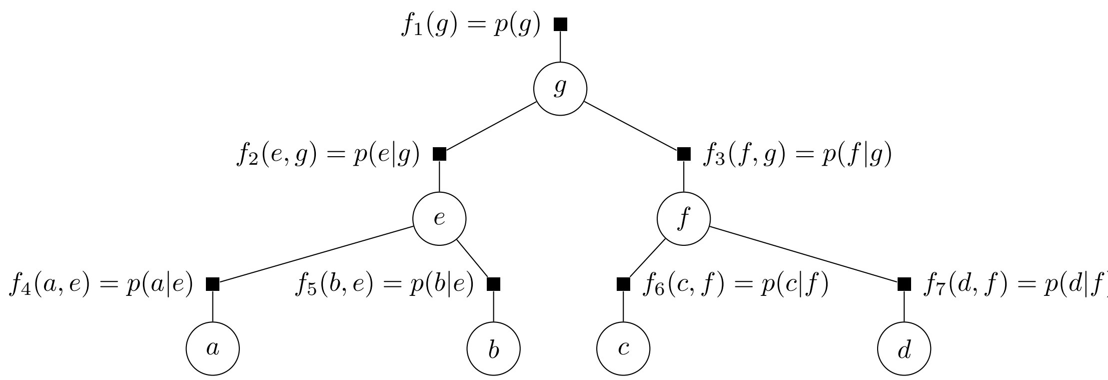
Question 1.3

1. $P(a, b, c, d, e, f) = P(a)P(e)P(b|a)P(c|e, b)P(d(c, b)P(f|b)$

2. No, yes, yes, yes, no.

Question 2

1. The factor graph is:



2. The messages emitted from leaf nodes a through d are $\{1, 1, 1, 1\}$. The same goes for the upward-bound messages from factors f_2 through f_7 — e.g. $\mu_{f_3 \to e}(e) = \{1, 1, 1, 1\}$ and so on. These 4-vectors correspond to the four possibilities for each random variable: 00, 01, 10, or 11.

The message from leaf factor f_1 is $\{0.5, 0.5\}$. The same goes for the downward-bound messages from factors f_2 and f_3 — e.g. $\mu_{g \to f_2}(g) = \{0.5, 0.5\}$.

The downward-bound message from factor f_2 (or, similarly, that from f_3) can be computed as $\mu_{f_2 \to e}(e) = \sum_{g' \in \mathbf{g}} p(e|g = g')\mu_{g \to f_2}(g') = \{0.42, 0.09, 0.09, 0.42\}$. Here, $\mathbf{g}$ represents the space of values g can assume, i.e. $\mathbf{g} = \{00, 11\}$.

The two downward-bound messages from variable e (or, similarly, those from f), are identical to one another. An example calculation is $\mu_{e \to f_4}(e) = \mu_{f_2 \to e}(e) \circ \mu_{f_5 \to e}(e) = \{0.42, 0.09, 0.09, 0.42\}$. The symbol $\circ$ represents the Hadamard product (elementwise multiplication).

The four downward-bound messages from factors f_4 through f_7 are also mutually identical. An example is $\mu_{f_4 \to a}(a) = \sum_{e' \in \mathbf{e}} p(a|e = e')\mu_{e \to f_4}(e') = \{0.3606, 0.1494, 0.1494, 0.3606\}$.

With the messages calculated, the formula in lecture 21, slide 21 can be used: $g(e) = \mu_{f_2 \to e}(e) \circ \mu_{f_4 \to e}(e) \circ \mu_{f_5 \to e}(e) = \{0.42, 0.09, 0.09, 0.42\} \circ \{1, 1, 1, 1\} \circ \{1, 1, 1, 1\} = \{0.42, 0.09, 0.09, 0.42\}$. Noting $Z = 1$, $p(e) = g(e)$.

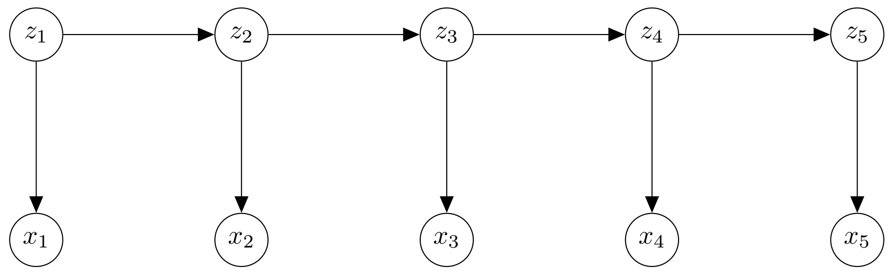
To calculate $p(e|a = 0.01)$, note $p(e|a = 0.01) = \frac{p(e, a=0.01)}{p(a=0.01)}$. The denominator can be calculated with the same formula used to find $g(e)$, and it can be shown again that $Z = 1$. The result is $p(a = 0.01) = \mu_{f_4 \to a}(a = 0.01) = 0.1494$.

The numerator can be calculated using the formula on slide 23 of lecture 21: $g(e, a = 0.01) = p(a = 0.01|e) \circ \mu_{f_4 \to e}(e) \circ \mu_{f_2 \to e}(e) = \{0.0378, 0.0729, 0.0009, 0.0378\}$. Since $Z = \sum_{e' \in \mathbf{e}} g(e') = \sum_{a' \in \mathbf{a}} g(a') = 1$, the result is $p(e, a = 0.01) = g(e, a = 0.01)$.

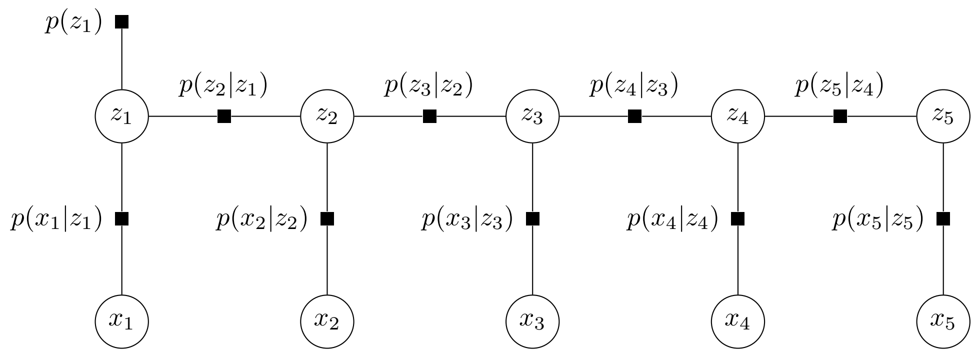
Dividing the two gives $p(e|a = 0.01) \approx \{0.2503, 0.4880, 0.0060, 0.2530\}$.

Question 3.1

1. The HMM BN is:



2. The HMM factor graph is:



Question 3.2

1. Using the usual equation for messages travelling from variables to factors, $\mu_{z_4 \to f_{z_3 z_4}}(z_4) = \mu_{f_{z_4 z_5} \to z_4}(z_4) \circ \mu_{f_{x_4 z_4} \to z_4}(z_4)$.

2. Similar to question 2.2, the marginal probability can be expressed as:

$$P(z_3 \,|\, x_1, x_2, x_3, x_4, x_5) = \frac{P(z_3, x_1, x_2, x_3, x_4, x_5)}{P(x_1, x_2, x_3, x_4, x_5)}$$

The numerator can be calculated using the equation from slide 22 of lecture 21:

$$P(x_i, X_s) = \frac{1}{Z} g(x_i, X_s), \quad \text{where} \;\; g(x_i, X_s) = \prod_{f_n \in Ne(x_i)} \mu_{f_n \to x_i}(x_i, X_s) \;\; \text{and} \;\; Z = \sum_{x_i} g(x_i).$$

The factor in the Z product is on slide 21:

$$g(x_i) = \prod_{f_n \in Ne(x_i)} \mu_{f_n \to x_i}(x_i)$$

Selecting z_3 to represent x_i in the above, then $X_s = \{x_1, x_2, x_3, x_4, x_5\}$ and:

$$g(x_i) = g(z_3) = \mu_{f_{z_3 z_2} \to z_3}(z_3) \circ \mu_{f_{z_4 z_3} \to z_3}(z_3) \circ \mu_{f_{x_3 z_3} \to z_3}(z_3)$$

$$Z = \sum_{z_3} g(z_3) = \sum_{z_3} \mu_{f_{z_3 z_2} \to z_3}(z_3) \circ \mu_{f_{z_4 z_3} \to z_3}(z_3) \circ \mu_{f_{x_3 z_3} \to z_3}(z_3)$$

$$
\begin{aligned}
g(z_3, x_1, x_2, x_3, x_4, x_5) =& \mu_{f_{z_3 z_2} \to z_3}(z_3, x_1, x_2, x_3, x_4, x_5) \ \circ \ \mu_{f_{z_4 z_3} \to z_3}(z_3, x_1, x_2, x_3, x_4, x_5) \circ \\
& \mu_{f_{x_3 z_3} \to z_3}(z_3, x_1, x_2, x_3, x_4, x_5)
\end{aligned}
$$

The denominator is calculated similarly, but x_i can be set to any of x_1, x_2, x_3, x_4, or x_5.

3. The marginal probability can be expressed as:

$$
P(x_6 \,|\, x_1, x_2, x_3, x_4, x_5) = \frac{P(x_1, x_2, x_3, x_4, x_5, x_6)}{P(x_1, x_2, x_3, x_4, x_5)} = \frac{\sum_{z_4} P(z_4, x_1, x_2, x_3, x_4, x_5, x_6)}{\sum_{z_4} P(z_4, x_1, x_2, x_3, x_4, x_5)}
$$

The same approach as in question 3.2.2 can be used, where z_4 is selected as the x_i, and in the numerator $X_s = \{x_1, x_2, x_3, x_4, x_5, x_6\}$. This yields:

$$
g(x_i) = g(z_4) = \mu_{f_{z_4 z_3} \to z_4}(z_4) \circ \mu_{f_{z_5 z_4} \to z_4}(z_4) \circ \mu_{f_{x_4 z_4} \to z_4}(z_4)
$$

$$
Z = \sum_{z_4} g(z_4) = \sum_{z_4} \mu_{f_{z_4 z_3} \to z_4}(z_4) \circ \mu_{f_{z_5 z_4} \to z_4}(z_4) \circ \mu_{f_{x_4 z_4} \to z_4}(z_4)
$$

$$
\begin{aligned}
g(z_4, x_1, x_2, x_3, x_4, x_5, x_6) =& \mu_{f_{z_4 z_3} \to z_4}(z_4, x_1, x_2, x_3, x_4, x_5, x_6) \ \circ \ \mu_{f_{z_5 z_4} \to z_4}(z_4, x_1, x_2, x_3, x_4, x_5, x_6) \circ \\
& \mu_{f_{x_4 z_4} \to z_4}(z_4, x_1, x_2, x_3, x_4, x_5, x_6)
\end{aligned}
$$

The denominator is calculated similarly, but $X_s = \{x_1, x_2, x_3, x_4, x_5\}$ as in the numerator in question 3.2.2.

Question 3.3

1. $\cdot$ is element-wise vector product. ^{T} is matrix transpose.

$$
\begin{aligned}
\mu_{f_{z_2 z_3} \to z_3} &= T \mu_{z_2 \to f_{z_2 z_3}} \\
&= T(\mu_{f_{z_1 z_2} \to z_2} \cdot \mu_{x_2 z_2 \to z_2}) \\
&= T((T \mu_{z_1 \to f_{z_1 z_2}}) \cdot (W^T x_2)) \\
&= T((T(\mu_{f_{x_1 z_1} \to z_1} \cdot \mu_{f_{z_1} \to z_1})) \cdot (W^T x_2)) \\
&= T((T(W^T x_1 \cdot \pi)) \cdot (W^T x_2))
\end{aligned}
$$

2. $\cdot$ is element-wise vector product. ^{T} is matrix transpose.

$$
\begin{aligned}
\mu_{z_3 \to f_{z_2 z_3}} &= \mu_{f_{z_3 z_4} \to z_3} \cdot \mu_{x_3 z_3 \to z_3} \\
&= (T^T \mu_{z_4 \to f_{z_3 z_4}}) \cdot (W^T x_3) \\
&= (T^T (\mu_{f_{z_4 z_5} \to z_4} \cdot \mu_{x_4 z_4 \to z_4})) \cdot (W^T x_3) \\
&= (T^T (T^T \mu_{z_5 \to f_{z_4 z_5}} \cdot (W^T x_4))) \cdot (W^T x_3) \\
&= (T^T (T^T \mu_{x_5 z_5 \to z_5} \cdot (W^T x_4))) \cdot (W^T x_3) \\
&= (T^T (T^T (W^T x_5) \cdot (W^T x_4))) \cdot (W^T x_3)
\end{aligned}
$$