Exercise Session 4

1) Given a pairwise singly connected Markov Network of the form:

$$p(\mathbf{x}) = \prod_{i \sim j} \phi(x_i, x_j). \tag{1}$$

Explain how to efficiently compute the normalisation factor (also called the partition function) Z as a function of the potentials ϕ.

2) Consider a pairwise Markov network defined on binary variables:

$$p(\mathbf{x}) = \phi(x_1, x_{100}) \prod_{i=1}^{99} \phi(x_i, x_{i+1}) \tag{2}$$

Is it possible to efficiently compute $\max_{x_1,\ldots,x_{100}} p(x)$?

3) Consider the distribution $P(a, b, c, d) = P(a)P(b|a)P(c|a)P(d|b, c)$.

a) Draw the factor graph for this distribution.

b) This graph is multiply-connected, which precludes max-product inference to compute $\max_{a,b,c,d} P(a, b, c, d)$. However, if we know the value of a, we can compute $\max_{b,c,d} P(b, c, d|a)$ using the max-product algorithm.

Draw the factor graph in which to perform max-product inference and explain in three lines why the max-product does apply to this factor graph.

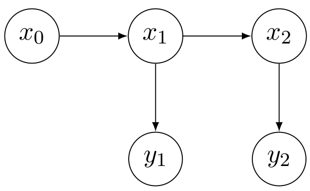
c) List all the messages (in the notation of the book) sent by the max- product algorithm. Show how these messages are computed, and explain how to compute $\max_{b,c,d} P(b, c, d|a)$ based on these messages.

d) Explain in three lines how you can obtain the state of a that maximizes $P(b, c, d|a)$ from this (also referred to as a^*).

4) A hamster is busy in an obstacle course. The obstacle course is divided in three stages A, B and C. It is not always possible to observe the stage in which the hamster is.

- variable $x_t \in \{A, B, C\}$ denotes the stage the hamster is in at time t.
- variable $y_t \in \{0, 1\}$ denotes whether or not we can observe the hamster at time t.

The belief network that we use is the following HMM.



The transition probabilities for the HMM are:

$$\begin{array}{lll} p(x_t = A|x_{t-1} = A) = 0.25 & p(x_t = A|x_{t-1} = B) = 0.25 & p(x_t = A|x_{t-1} = C) = 0.5 \\ p(x_t = B|x_{t-1} = A) = 0.5 & p(x_t = B|x_{t-1} = B) = 0.25 & p(x_t = B|x_{t-1} = C) = 0.25 \\ p(x_t = C|x_{t-1} = A) = 0.25 & p(x_t = C|x_{t-1} = B) = 0.5 & p(x_t = C|x_{t-1} = C) = 0.25 \end{array} \tag{3}$$

The probabilities of observing the hamster are conditioned on the stage in which the hamster is.

$$p(y_t = 1|x_t = A) = 0.5 \quad p(y_t = 1|x_t = B) = 0.25 \quad p(y_t = 1|x_t = C) = 0.25 \tag{4}$$

The initial probabilities of the Hamster being in state A, B, C are:

$$p(x_0 = A) = 0.5 \;\; p(x_0 = B) = 0.25 \;\; p(x_0 = C) = 0.25 \tag{5}$$

Given that the observations $y_1 = 0$ and $y_2 = 1$, what is the most probable value assignment for the hidden states x_0, x_1, x_2?

5) Consider the factor graph shown in Figure 1 and use the sum-product algorithm to compute.

   a) Z (the partition function of $p(A, B, C, D, E, F)$)

   b) $p(C = 0)$ and $p(C = 1)$

   c) $p(A = 0)$ and $p(A = 1)$

   All the variables represented are binary. The factors can be evaluated according to the following tables.

B	$f_1(B)$
0	3
1	1

A	B	C	$f_2(A, B, C)$
0	0	0	5
0	0	1	4
0	1	0	5
0	1	1	8
1	0	0	1
1	0	1	2
1	1	0	1
1	1	1	3

C	D	$f_3(C, D)$
0	0	10
0	1	10
1	0	1
1	1	1

D	E	$f_4(D, E)$
0	0	1
0	1	10
1	0	10
1	1	1

E	$f_5(E)$
0	10
1	1

D	F	$f_6(D, F)$
0	0	1
0	1	10
1	0	10
1	1	1

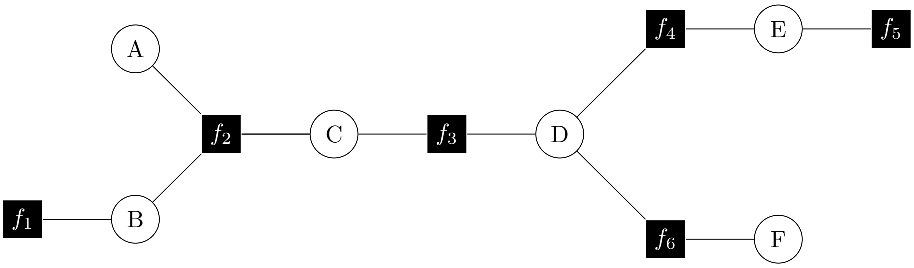


Figure 1: Factor graph for Question 5.

6) Use bucket elimination to compute $p(g)$ in the network from Figure 2. Sketch the steps that lead to the final solution and list the formulas you use. Use the graphical notation and table that we have employed in the course. The variable ordering is alphabetical.

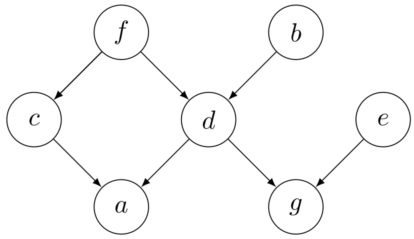
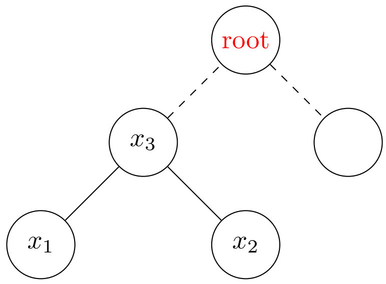


Figure 2: Belief network for Question 6.

Exercise Session 4 – Solutions

1. Because the Markov network is singly connected, there's EXACTLY one path between each pair of nodes. Note that, an undirected graph can be singly connected if and only if it is acyclic. Thus, we can then use the sum product algorithm (without the need for loop-cutting).



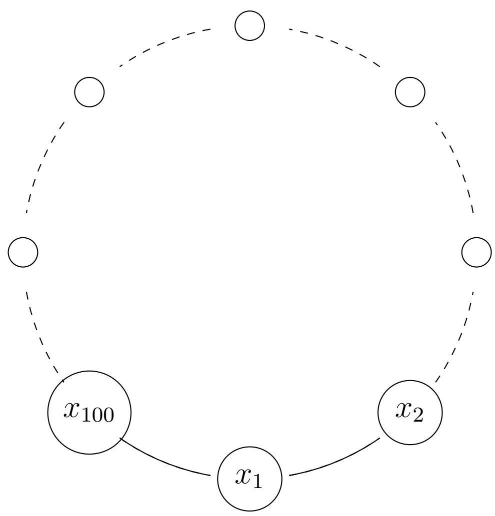
The normalization constant, Z, is defined as follows:

$$Z = \sum_{\mathbf{x}} p(\mathbf{x}) = \sum_{\mathbf{x}} \prod_{i \sim j} \phi(x_i, x_j)$$

where $i \sim j$ denotes that there is an edge between the nodes i and j.

To efficiently compute Z using the sum product algorithm, we must compute messages from leaves to the root node. This amounts to *pushing* the summations inside or starting from the leaves and working up the the root, which all effectively describe the same algorithm.

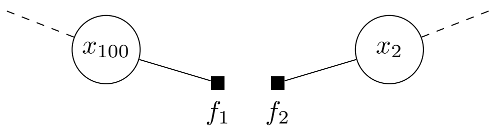
2. We notice that the network is a loop of 100 nodes, so it would look like the following:



The **loop-cut** algorithm can be used to efficiently compute this. By conditioning on x_1, we effectively have to compute

$$\max_{x_1} \max_{x_2,\ldots,x_{100}} p(\mathbf{x})$$

where – given x_1 – the network for $(x_2, \ldots, x_{100})$ is singly connected and thus sum-product can be used. The factor graph becomes

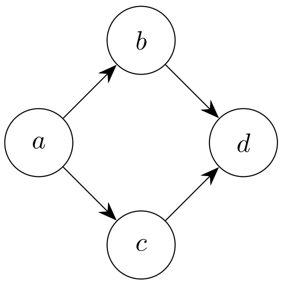


with the factors defined as $f_1(x_{100}) = \phi(x_{100}, x_1 = A)$ and $f_2(x_2) = \phi(x_2, x_1 = A)$ for a fixed value of A – as done by the outer max. In other words, we need to find

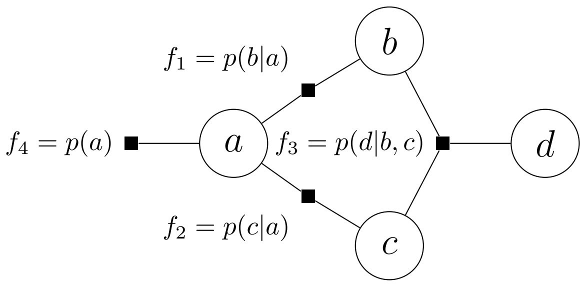
$$\max(\max_{x_2,\ldots,x_{100}} p(x_1 = True, x_2, \ldots, x_{100}), \quad \max_{x_2,\ldots,x_{100}} p(x_1 = False, x_2, \ldots, x_{100}))$$

So, the first part will have factors defined as $f_1(x_{100}) = \phi(x_{100}, x_1 = True)$ and $f_2(x_2) = \phi(x_2, x_1 = True)$ and the second part will have factors defined as $f_1(x_{100}) = \phi(x_{100}, x_1 = False)$ and $f_2(x_2) = \phi(x_2, x_1 = False)$.

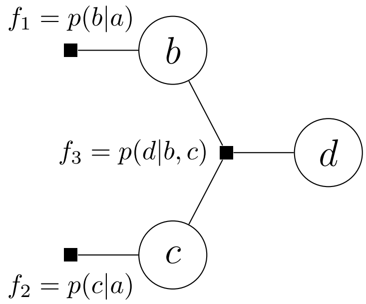
3. a) Based on the factorization we know that a has no parents; a is the only parent of both b and c ; and b and c are parents of d. This will result in the following Bayesian Network:



Now we can convert this to the following factor graph. We could have also done that directly just by looking at the arguments of each of the factors.



b) In order to compute $\max_{a,b,c,d} P(a, b, c, d)$, we would have to remove the loop in the factor graph shown above. This can be done by loop cutting at the node a. As a does not have any parents, it is easier to cut the loop at a as opposed to cutting it at any other node. Now, we get the following factor graph:



This factor graph is suitable for performing the max-product inference as it is acyclic.

c) The messages are as follows (for a fixed value of a).

- $\mu_{f_1 \to b} = f_1 = p(b|a)$
- $\mu_{f_2 \to c} = f_2 = p(c|a)$
- $\mu_{b \to f_3} = \mu_{f_1 \to b}$
- $\mu_{c \to f_3} = \mu_{f_2 \to c}$
- $\mu_{f_3 \to d} = \max_{b,c} f_3(d,b,c)\mu_{b \to f_3}\mu_{c \to f_3} = \max_{b,c} p(d|b,c)p(b|a)p(c|a)$

In order to compute $\max_{b,c,d} P(b,c,d|a)$ we then

i. Compute $\max_d \mu_{f_3 \to d}(d)$ – which is possible because using the messages to d we have marginalised over all other variables.

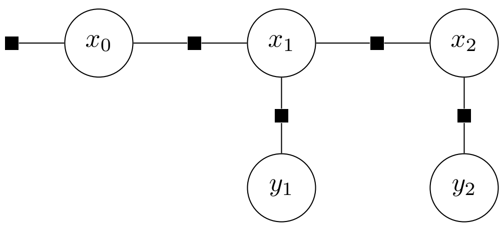
ii. Fill in the d which maximises the previous expression in the message $\mu_{f_3 \to d}(d)$ in order to maximise over b and c and get the final result.

What we're effectively doing is first determining the most probable state for d, given that we know nothing else about states of other variables. After this is determined, we compute the combined state of b and c that is most probable to have generated this d, as d depends on both of them.

A thorough example is given in the solution for Exercise 5.

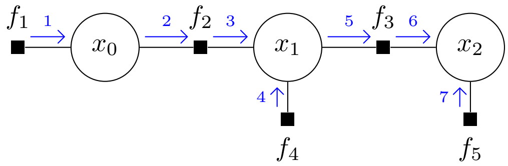
d) To obtain the state of a that maximizes $p(b,c,d|a)$, we look at $p(b^*,c^*,d^*|a)$ [1] and choose $a^* = True$ if $p(b^*,c^*,d^*|a = True) > p(b^*,c^*,d^*|a = False)$ or choose $a^* = False$ otherwise.

4. The original factor graph is as follows



but as the y_i values are observed, the bottom variables turn to factors as well, resulting in the following factor graph.

[1] b^*, c^* and d^* are the most probable states for b, c, and d that maximize $p(b,c,d|a)$ which are calculated in c) part.



This is because the factors f_4 and f_5 can now only depend on the value of x_1 and x_2 respectively. The sequence of messages are shown and x_2 is chose as the root. These messages can now be computed as follows:

1. $\mu_{f_1 \to x_0}(x_0) = f_1(x_0) = p(x_0)$
2. $\mu_{x_0 \to f_2}(x_0) = \mu_{f_1 \to x_0}(x_0) = p(x_0)$
3. $\mu_{f_2 \to x_1}(x_1) = \max_{x_0} \mu_{x_0 \to f_2}(x_0) f_2(x_0, x_1) = \max_{x_0} p(x_0) p(x_1|x_0)$

$$\begin{aligned}
&\mu_{f_2 \to x_1}(A) = \max(1/2 \cdot 1/4; 1/4 \cdot 1/4; 1/4 \cdot 1/2) = 1/8 \\
&\mu_{f_2 \to x_1}(B) = 1/4 \\
&\mu_{f_2 \to x_1}(C) = 1/8
\end{aligned}$$

4. $\mu_{f_4 \to x_1}(x_1) = f_4(x_1) = p(y_1 = 0|x_1)$
5. $\mu_{x_1 \to f_3}(x_1) = \mu_{f_2 \to x_1}(x_1) \mu_{f_4 \to x_1}(x_1)$

$$\begin{aligned}
&\mu_{f_4 \to x_1}(A) = 1/8 \cdot 1/2 = 1/16 \\
&\mu_{f_4 \to x_1}(B) = 3/16 \\
&\mu_{f_4 \to x_1}(C) = 3/32
\end{aligned}$$

6. $\mu_{f_3 \to x_2}(x_2) = \max_{x_1} \mu_{x_1 \to f_3}(x_1) f_3(x_1, x_2) = \max_{x_1} \mu_{x_1 \to f_3}(x_1) p(x_2|x_1)$

$$\begin{aligned}
&\mu_{f_3 \to x_2}(A) = 3/64 \\
&\mu_{f_3 \to x_2}(B) = 3/64 \\
&\mu_{f_3 \to x_2}(C) = 3/32
\end{aligned}$$

7. $\mu_{f_5 \to x_2}(x_2) = f_5(x_2) = p(y_2 = 1|x_2)$

Now we have everything to compute the most probable assignment for x_2.

$$\begin{aligned}
x_2^* &= \text{argmax}_{x_2} \mu_{f_5 \to x_2}(x_2) \mu_{f_3 \to x_2}(x_2) \\
&= \text{argmax}_{x_2} (\underbrace{3/64 \cdot 1/2}_{A}; \underbrace{3/64 \cdot 1/4}_{B}; \underbrace{3/32 \cdot 1/4}_{C}) \\
&= \{A, C\}
\end{aligned}$$

Out of these, let's continue with $x_2 = A$. We backtrack to message 6 and get

$$\begin{aligned}
\mu_{f_3 \to x_2}(x_2 = A) &= \max_{x_1} \mu_{x_1 \to f_3}(x_1) p(x_2 = A|x_1) \\
&= \max_{x_1} (\underbrace{1/16 \cdot 1/4}_{A}, \underbrace{3/16 \cdot 1/4}_{B}, \underbrace{3/32 \cdot 1/2}_{C}) \\
&= 3/64
\end{aligned}$$

for $x_1^* = \{B, C\}$. Picking $x_1 = B$ to continue, we backtrack to 3 and find

$$\begin{aligned}\mu_{f_2 \to x_1}(x_1 = B) &= \max_{x_0} p(x_0) p(x_1 = B | x_0) \\ &= \max_{x_0}(\underbrace{1/2 \cdot 1/2}_{A}, \underbrace{1/4 \cdot 1/4}_{B}, \underbrace{1/4 \cdot 1/1/4}_{C}) \\ &= 1/4\end{aligned}$$

for $x_0^* = A$. One most probable value assignment of (x_0, x_1, x_2) is then (A, B, A). Picking $x_1 = C$, we backtrack to 3 and find

$$\begin{aligned}\mu_{f_2 \to x_1}(x_1 = C) &= \max_{x_0} p(x_0) p(x_1 = C | x_0) \\ &= \max_{x_0}(\underbrace{1/2 \cdot 1/4}_{A}, \underbrace{1/4 \cdot 1/2}_{B}, \underbrace{1/4 \cdot 1/1/4}_{C}) \\ &= 1/8\end{aligned}$$

for $x_0^* = \{A, B\}$. Two most probable value assignments of (x_0, x_1, x_2) are then (A, C, A) and (B, C, A).

Let us now consider $x_2^* = C$ and backtrack to message 6.

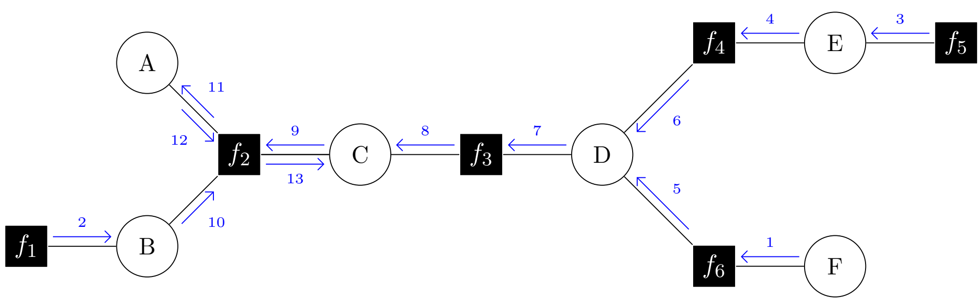
$$\begin{aligned}\mu_{f_3 \to x_2}(x_2 = C) &= \max_{x_1} \mu_{x_1 \to f_3}(x_1) p(x_2 = C | x_1) \\ &= \max_{x_1}(\underbrace{1/16 \cdot 1/4}_{A}, \underbrace{3/16 \cdot 1/2}_{B}, \underbrace{3/32 \cdot 1/4}_{C}) \\ &= 3/32\end{aligned}$$

Thus $x_1^* = B$ in this case. Now, we backtrack to 3 and find

$$\begin{aligned}\mu_{f_2 \to x_1}(x_1 = B) &= \max_{x_0} p(x_0) p(x_1 = B | x_0) \\ &= \max_{x_0}(\underbrace{1/2 \cdot 1/2}_{A}, \underbrace{1/4 \cdot 1/4}_{B}, \underbrace{1/4 \cdot 1/1/4}_{C}) \\ &= 1/4\end{aligned}$$

Thus $x_0^* = A$ in this case. So, another most probable value assignment of (x_0, x_1, x_2) is then (A, B, C).

5. We pick A as a root and compute all messages up to A. This means that we're marginalising over all other variables and effectively computing $\mu_{f_2 \to A}(A) = \sum_{B,C,D,E,F} p(A, B, C, D, E, F)$.



The messages can now be computed as follows:

1. $\mu_{F \to f_6}(0) = \mu_{F \to f_6}(1) = 1$

2. $$\begin{array}{lll} \mu_{f_1\to B}(B) & = & f_1(B) \\ \mu_{f_1\to B}(0) & = & 3 \\ \mu_{f_1\to B}(1) & = & 1 \end{array}$$

3. $$\begin{array}{lll} \mu_{f_5\to E}(E) & = & f_5(E) \\ \mu_{f_5\to E}(0) & = & 10 \\ \mu_{f_5\to E}(1) & = & 1 \end{array}$$

4. $$\begin{array}{lllll} \mu_{E\to f_4}(E) & = & \mu_{f_5\to E}(E) & & \\ \mu_{E\to f_4}(0) & = & \mu_{f_5\to E}(0) & = & 10 \\ \mu_{E\to f_4}(1) & = & \mu_{f_5\to E}(1) & = & 1 \end{array}$$

5. $$\begin{array}{lllllll} \mu_{f_6\to D}(D) & = & \sum_F \mu_{F\to f_6}(F) f_6(D,F) & & & & \\ \mu_{f_6\to D}(0) & = & \mu_{F\to f_6}(0) f_6(0,0) + \mu_{F\to f_6}(1) f_6(0,1) & = & 1\cdot 1 + 1\cdot 10 & = & 11 \\ \mu_{f_6\to D}(1) & = & \mu_{F\to f_6}(0) f_6(1,0) + \mu_{F\to f_6}(1) f_6(1,1) & = & 1\cdot 10 + 1\cdot 1 & = & 11 \end{array}$$

6. $$\begin{array}{lllllll} \mu_{f_4\to D}(D) & = & \sum_E \mu_{E\to f_4}(E) f_4(D,E) & & & & \\ \mu_{f_4\to D}(0) & = & \mu_{E\to f_4}(0) f_4(0,0) + \mu_{E\to f_4}(1) f_4(0,1) & = & 10\cdot 1 + 1\cdot 10 & = & 20 \\ \mu_{f_4\to D}(1) & = & \mu_{E\to f_4}(0) f_4(1,0) + \mu_{E\to f_4}(1) f_4(1,1) & = & 10\cdot 10 + 1\cdot 1 & = & 101 \end{array}$$

7. $$\begin{array}{lllllll} \mu_{D\to f_3}(D) & = & \mu_{f_4\to D}(D)\cdot \mu_{f_6\to D}(D) & & & & \\ \mu_{D\to f_3}(0) & = & \mu_{f_4\to D}(0)\cdot \mu_{f_6\to D}(0) & = & 11\cdot 20 & = & 220 \\ \mu_{D\to f_3}(1) & = & \mu_{f_4\to D}(1)\cdot \mu_{f_6\to D}(1) & = & 11\cdot 101 & = & 1111 \end{array}$$

8. $$\begin{array}{lllllll} \mu_{f_3\to C}(C) & = & \sum_D \mu_{D\to f_3}(D) f_3(C,D) & & & & \\ \mu_{f_3\to C}(0) & = & \mu_{D\to f_3}(0) f_3(0,0) + \mu_{D\to f_3}(1) f_3(0,1) & = & 220\cdot 10 + 1111\cdot 10 & = & 13310 \\ \mu_{f_3\to C}(1) & = & \mu_{D\to f_3}(1) f_3(1,0) + \mu_{D\to f_3}(1) f_3(1,1) & = & 220\cdot 1 + 1111\cdot 1 & = & 1331 \end{array}$$

9. $$\begin{array}{lllll} \mu_{C\to f_2}(C) & = & \mu_{f_3\to C}(C) & & \\ \mu_{C\to f_2}(0) & = & \mu_{f_3\to C}(0) & = & 13310 \\ \mu_{C\to f_2}(1) & = & \mu_{f_3\to C}(1) & = & 1331 \end{array}$$

10. $$\begin{array}{lllll} \mu_{B\to f_2}(B) & = & \mu_{f_1\to B}(B) & & \\ \mu_{B\to f_2}(0) & = & \mu_{f_1\to B}(0) & = & 3 \\ \mu_{B\to f_2}(1) & = & \mu_{f_1\to B}(1) & = & 1 \end{array}$$

11. $$\begin{array}{lll} \mu_{f_2\to A}(A) & = & \sum_{B,C} \mu_{C\to f_2}(C)\mu_{B\to f_2}(B) f_2(A,B,C) \\ \mu_{f_2\to A}(0) & = & \mu_{C\to f_2}(0)\mu_{B\to f_2}(0) f_2(0,0,0) + \\ & & \mu_{C\to f_2}(0)\mu_{B\to f_2}(1) f_2(0,1,0) + \\ & & \mu_{C\to f_2}(1)\mu_{B\to f_2}(0) f_2(0,0,1) + \\ & & \mu_{C\to f_2}(1)\mu_{B\to f_2}(1) f_2(0,1,1) \\ & = & 13310\cdot 3\cdot 5 + \\ & & 13310\cdot 1\cdot 5 + \\ & & 1331\cdot 3\cdot 4 + \\ & & 1331\cdot 1\cdot 8 \\ & = & 1331\cdot 220 \end{array}$$

$$
\begin{aligned}
\mu_{f_2\to A}(1) &= \mu_{C\to f_2}(0)\mu_{B\to f_2}(0)f_2(1,0,0) + \\
&\quad \mu_{C\to f_2}(0)\mu_{B\to f_2}(1)f_2(1,1,0) + \\
&\quad \mu_{C\to f_2}(1)\mu_{B\to f_2}(0)f_2(1,0,1) + \\
&\quad \mu_{C\to f_2}(1)\mu_{B\to f_2}(1)f_2(1,1,1) \\
&= 13310\cdot 3\cdot 1 + \\
&\quad 13310\cdot 1\cdot 1 + \\
&\quad 1331\cdot 3\cdot 2 + \\
&\quad 1331\cdot 1\cdot 3 \\
&= 1331\cdot 49
\end{aligned}
$$

Now Z is computed using this message as

$$
\begin{aligned}
Z &= \sum_{A,B,C,D,E,F} p(A,B,C,D,E,F) \\
&= \sum_{A} \mu_{f_2\to A}(A) \\
&= \mu_{f_2\to A}(0) + \mu_{f_2\to A}(1) \\
&= 1331\cdot 220 + 1331\cdot 49 \\
&= 1331\cdot 269 = 358039
\end{aligned}
$$

and

$$
\begin{aligned}
p(A=0) &= \frac{\mu_{f_2\to A}(0)}{Z} = \frac{1331\cdot 220}{1331\cdot 269} = \frac{220}{269} = 0.8178 \\
p(A=1) &= \frac{\mu_{f_2\to A}(1)}{Z} = \frac{1331\cdot 49}{1331\cdot 269} = \frac{49}{269} = 0.1822
\end{aligned}
$$

Similarly, in order to compute $p(C)$, we need to compute 2 more messages towards C.

12. $$
\begin{aligned}
\mu_{A\to f_2}(0) &= 1 \\
\mu_{A\to f_2}(1) &= 1
\end{aligned}
$$

13. $$
\begin{aligned}
\mu_{f_2\to C}(C) &= \textstyle\sum_{A,B} \mu_{A\to f_2}(A)\mu_{B\to f_2}(B)f_2(A,B,C) \\
\mu_{f_2\to C}(0) &= \mu_{A\to f_2}(0)\mu_{B\to f_2}(0)f_2(0,0,0) + \\
&\quad \mu_{A\to f_2}(0)\mu_{B\to f_2}(1)f_2(0,1,0) + \\
&\quad \mu_{A\to f_2}(1)\mu_{B\to f_2}(0)f_2(1,0,0) + \\
&\quad \mu_{A\to f_2}(1)\mu_{B\to f_2}(1)f_2(1,1,0) \\
&= 1\cdot 3\cdot 5 + 1\cdot 1\cdot 5 + 1\cdot 3\cdot 1 + 1\cdot 1\cdot 1 \\
&= 24 \\
\mu_{f_2\to C}(1) &= \mu_{A\to f_2}(0)\mu_{B\to f_2}(0)f_2(0,0,1) + \\
&\quad \mu_{A\to f_2}(0)\mu_{B\to f_2}(1)f_2(0,1,1) + \\
&\quad \mu_{A\to f_2}(1)\mu_{B\to f_2}(0)f_2(1,0,1) + \\
&\quad \mu_{A\to f_2}(1)\mu_{B\to f_2}(1)f_2(1,1,1) \\
&= 1\cdot 3\cdot 4 + 1\cdot 1\cdot 8 + 1\cdot 3\cdot 2 + 1\cdot 1\cdot 3 \\
&= 29
\end{aligned}
$$

Now, we multiply all incoming messages and normalise

$$
\begin{aligned}
p(C=0) &= \frac{\mu_{f_2\to C}(C=0)\mu_{f_3\to C}(C=0)}{Z} = \frac{24\cdot 13310}{1331\cdot 269} = \frac{240}{269} = 0.8922 \\
p(C=1) &= \frac{\mu_{f_2\to C}(C=1)\mu_{f_3\to C}(C=1)}{Z} = \frac{29\cdot 1331}{1331\cdot 269} = \frac{29}{269} = 0.1078
\end{aligned}
$$

Note that the message $\mu_{f_3 \to C}(C)$ was already computed at step 8 when computing Z.

6. First we factor $p(g)$ as

$$p(g) = \sum_{a,b,c,d,e,f} p(f)p(c|f)p(a|c,d)p(d|f,b)p(b)p(g|d,e)p(e)$$

and perform bucket elimination with the alphabetical ordering

a	$p(a\|c,d)$	
b	$p(b)p(d\|f,b)$	$\gamma_a(c,d) = \sum_a p(a\|c,d) = 1$
c	$p(c\|f)$	$\gamma_b(d,f) = \sum_b p(b)p(d\|f,b)$
d	$p(g\|d,e)$	$\gamma_c(d,f) = \sum_c p(c\|f)\gamma_b(d,f) = \gamma_b(d,f)\sum_c p(c\|f) = \gamma_b(d,f)$
e	$p(e)$	$\gamma_d(e,f,g) = \sum_d p(g\|d,e)\gamma_c(d,f)$
f	$p(f)$	$\gamma_e(f,g) = \sum_e p(e)\gamma_d(e,f,g)$
g		$\gamma_f(g) = \sum_f p(f)\gamma_e(f,g)$

We now have,

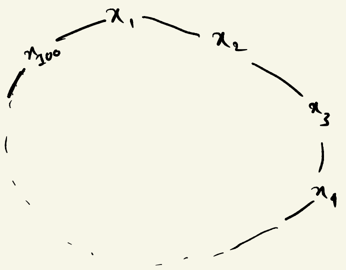
$$\begin{aligned}
p(g) &= \gamma_f(g) \\
&= \sum_f p(f)\gamma_e(f,g) \\
&= \sum_f p(f) \sum_e p(e)\gamma_d(e,f,g) \\
&= \sum_f p(f) \sum_e p(e) \sum_d p(g|d,e)\gamma_c(d,f) \\
&= \sum_f p(f) \sum_e p(e) \sum_d p(g|d,e)\gamma_b(d,f) \\
&= \sum_f p(f) \sum_e p(e) \sum_d p(g|d,e) \sum_b p(b)p(d|f,b)
\end{aligned}$$

① $P(x) = \sum_{i \sim j} \varphi(x_i, x_j)$

$Z = \sum_x P(x) = \sum_x \sum_{i \sim j} \varphi(x_i, x_j)$

Sum-product algorithm

② $P(x) = \varphi(x_1, x_{100}) \prod_{i=1}^{99} \varphi(x_i, x_{i+1})$

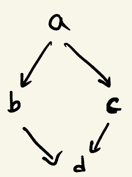


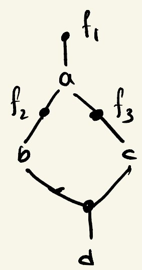
Fix x_1

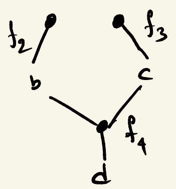
$\max_{x_1} \Big(\max_{x_2} \varphi(x_2, x_1) \max_{x_3} \varphi(x_2, x_3) \max_{x_4} \cdots$

$\cdots \max_{x_{100}} \varphi(x_{100}, x_1) \Big)$

3) $P(a,b,c,d) = P(a) P(b|a) P(c|a) P(d|b,c)$







$P(b,c,d|a) = \dfrac{P(b,c,d,a)}{P(a)}$

$= P(a) P(b|a) P(c|a) P(d|b,c)$

$P(a) \max_{b,c,d} P(b|a) P(c|a) P(d|b,c)$

$\mu_{f_2 \to b} = P(b|a) = f_2$

$\mu_{f_3 \to c} = P(c|a) = f_3$

$\mu_{b \to f_4} = \mu_{f_2 \to b}$

$\mu_{c \to f_4} = \mu_{f_3 \to c}$

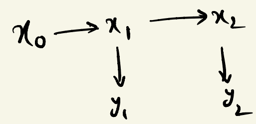
$\mu_{f_4 \to d} = \max_{b,c} f(d,b,c)\, \mu_{f_2 \to b}\, \mu_{f_3 \to c}$

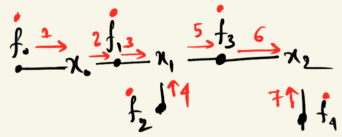
$= \max_{b,c} P(d|b,c) P(b|a) P(c|a)$

$\max P(b,c,d|a) = \max_d \mu_{f_4 \to d}$

d) $P(b^*, c^*, d^* | a = \text{True}) > P(b^*, c^*, d^* | a = \text{False})$ then $a = \text{True}$ else False

④





$P(x_0, x_1, x_2, y_1, y_2)$

$= P(x_0) P(x_1|x_0) P(y_1|x_1) \cdot P(x_2|x_1) P(y_2|x_2)$

$f_0 = P(x_0)$

$f_1 = P(x_1|x_0)$

$f_2 = P(y_1|x_1)$

$f_3 = P(x_2|x_1)$

$f_4 = P(y_2|x_2)$

$P(x_0, x_1, x_2 | y_1 y_2)$

$\max_{x_0, x_1, x_2} P(x_0, x_1, x_2, y_1, y_2)$

① $\mu_{f_0 \to x_0} = f_0 = P(x_0)$

② $\mu_{x_0 \to f_1} = \mu_{f_0 \to x_0} = P(x_0)$

③ $\mu_{f_1 \to x_1} = \max_{x_0} f_1(x_1, x_0) \cdot \mu_{x_0 \to f_1}$

$= \max_{x_0} f_1(x_1, x_0) f_0(x_0)$

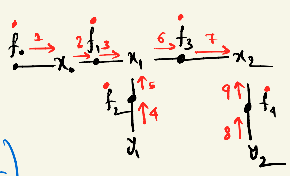
$= \max_{x_0} P(x_1|x_0) P(x_0)$

$\mu_{f_1 \to x_1}(A) = \max_{x_0} P(x_1 = A | x_0) P(x_0)$

$= \max\left(\frac{1}{2} \cdot \frac{1}{4}, \frac{1}{4} \cdot \frac{1}{4}, \frac{1}{2} \cdot \frac{1}{4} \right)$

$= \frac{1}{8}$

$\operatorname{argmax}_{x_0} \mu_{f_1 \to x_1}(A) = A/C$



$\mu_{f_1 \to x_1}(B)$

$= \max\left(\frac{1}{2} \cdot \frac{1}{2}, \frac{1}{4} \cdot \frac{1}{4}, \frac{1}{4} \cdot \frac{1}{4} \right)$

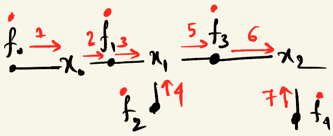
$= \frac{1}{4}$

$\operatorname{argmax}_{x_0} \mu_{f_1 \to x_1}(B) = A$

$\mu_{f_1 \to x_1}(C) = \max\left(\frac{1}{4} \cdot \frac{1}{2}, \frac{1}{2} \cdot \frac{1}{4}, \frac{1}{4} \cdot \frac{1}{4} \right)$

$= \frac{1}{8}$

$\operatorname{argmax}_{x_0} \mu_{f_1 \to x_1}(C) = A/B$



④ $\mu_{f_2 \to x_1}(x_1) = f_2(y_1=0|x_1)$

$\mu_{f_2 \to x_1}(A) = 0.5 = \frac{1}{2}$

$\mu_{f_2 \to x_1}(B) = 0.75 = \frac{3}{4}$

$\mu_{f_2 \to x_1}(C) = 0.75 = \frac{3}{4}$

⑤ $\mu_{x_1 \to f_3}(x_1) = \mu_{f_1 \to x_1}(x_1)\, \mu_{f_2 \to x_1}(x_1)$

$\mu_{x_1 \to f_3}(A) = \frac{1}{8} \times \frac{1}{2} = \frac{1}{16}$

$\mu_{x_1 \to f_3}(B) = \frac{1}{4} \times \frac{3}{4} = \frac{3}{16}$

$\mu_{x_1 \to f_3}(C) = \frac{1}{8} \times \frac{3}{4} = \frac{3}{32}$

⑥ $\mu_{f_3 \to x_2}(x_2) = \max_{x_1} f_3(x_2, x_1) \cdot \mu_{x_1 \to f_3}(x_1)$

$= \max_{x_1} p(x_2|x_1)\, \mu_{x_1 \to f_3}(x_1)$

$\mu_{f_3 \to x_2}(A) = \max\left(\frac{1}{16} \times \frac{1}{4}, \frac{3}{16} \times \frac{1}{4}, \frac{3}{32} \times \frac{1}{2}\right)$

$= \max\left(\frac{1}{64}, \frac{3}{64}, \frac{3}{64}\right)$

$= \frac{3}{64}$

$\arg\max_{x_1} \mu_{f_3 \to x_2}(x_2) = B/C$

$\mu_{f_3 \to x_2}(B) = \max\left(\frac{1}{16} \times \frac{1}{2}, \frac{3}{16} \times \frac{1}{4}, \frac{3}{32} \times \frac{1}{4}\right)$

$= \frac{1}{32}$

$\arg\max_{x_1}(B)$ A

$\mu_{f_3 \to x_2}(C) = \max\left(\frac{1}{16} \times \frac{1}{4}, \frac{3}{16} \times \frac{1}{2}, \frac{3}{32} \times \frac{1}{4}\right)$

$= \frac{3}{32}$

$\arg\max_{x_2}(C)$ B

⑦ $\mu_{f_5 \to x_2}(x_2) = f_5(x_2) = P(1|x_2)$

$\therefore x_2 = \arg\max_{x_2} \mu_{f_5 \to x_2}(x_2) \cdot \mu_{f_3 \to x_2}(x_2)$

$= \arg\max_{x_2} \left(\frac{3}{64} \times \frac{1}{2}, \frac{1}{4} \times \frac{1}{32}, \frac{3}{32} \times \frac{1}{4}\right) = \{A, C\}$

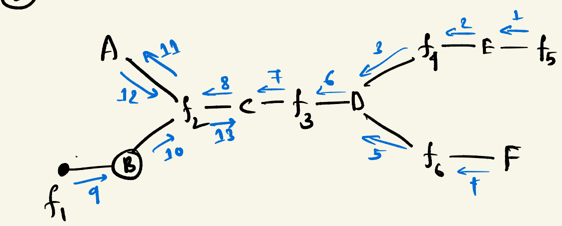
let $x_2^* = C$, $x_1^* = B$, $x_0^* = A$

let $x_2^* = A$, $x_1^* = C$, $x_0^* = B$

let $x_2^* = A$, $x_1^* = B$, $x_0^* = A$

let $x_2^* = A$, $x_1^* = C$, $x_0^* = A$

⑤



① a) $P(A,B,C,D,E,F) =$

$f_1(B)\, f_2(A,B,C)\, f_3(C,D)\, f_4(D,E)\, f_5(E)\, f_6(D,F)$

① $\mu_{f_5 \to E}(E) = f_5(E)$

② $\mu_{E \to f_4}(E) = \mu_{f_5 \to E}(E) = f_5(E)$

$\mu_{E \to f_4}(1) = 1$

$\mu_{E \to f_4}(0) = 10$

③ $\mu_{f_4 \to D}(D) = \sum_E \mu_{E \to f_4}(E) \cdot f_4(D,E)$

$\mu_{f_4 \to D}(0) = \sum f_4(0,E)\, \mu_{E \to f_4}(E)$

$= f_4(0,0)\, \mu_{E \to f_4}(0) + f_4(0,1)\, \mu_{E \to f_4}(1)$

$= 1 \times 10 + 10 \times 1$

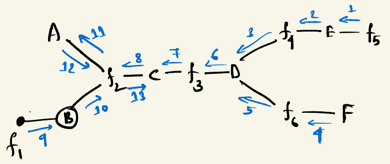
$= 20$

$\mu_{f_4 \to D}(1) = \sum_E f_4(1,E)\, \mu_{E \to f_4}(E)$

$= f_4(1,0)\, \mu_{E \to f_4}(0) + f_4(1,1)\, \mu_{E \to f_4}(1)$

$= 10 \times 10 + 1 \times 1$

$= 101$



④ $\mu_{F\to f_6} = 1$

⑤ $\mu_{f_6\to D} = \sum f_6(D,F)$

$\mu_{f_6\to D}(0) = f_6(0,0) + f_6(0,1)$
$= 1 + 10 = 11$

$\mu_{f_6\to D}(1) = f_6(1,0) + f_6(1,1)$
$= 11$

⑥ $\mu_{D\to f_3}(D) = \mu_{f_4\to D}(D)\cdot\mu_{f_6\to D}(D)$

$\mu_{D\to f_3}(0) = 11\times 20 = 220$

$\mu_{D\to f_3}(1) = 101\times 11 = 1111$

⑦ $\mu_{f_3\to c}(c) = \sum_D f_3(c,D)\cdot\mu_{D\to f_3}(D)$

$\mu_{f_3\to c}(0) = f_3(0,0)\,\mu_{D\to f_3}(0) + f_3(0,1)\,\mu_{D\to f_3}(1)$
$= 10\times 220 + 1111\times 10$
$= 13310$

$\mu_{f_3\to c}(1) = 220\times 1 + 1111\times 1$
$= 220 + 1111 = 1331$

⑧ $\mu_{c\to f_2}(c) = \mu_{f_3\to c}(c)$

$\mu_{c\to f_2}(0) = 13310$

$\mu_{c\to f_2}(1) = 1331$

⑨ $\mu_{f_1\to B}(B) = f_1(B)$

$\mu_{f_1\to B}(0) = 3$

$\mu_{f_1\to B}(1) = 1$

10) $\mu_{B\to f_2}(B) = \mu_{f_1\to B}(B)$

$\mu_{B\to f_2}(0) = 3$

$\mu_{B\to f_2}(1) = 1$

11) $\mu_{f_2\to A}(A)$
$= \sum_{BC} f_2(A,B,C)\,\mu_{B\to f_2}(B)\cdot\mu_{C\to f_2}(C)$

$\mu_{f_2\to A}(0) = \mu_{B\to f_2}(0)\{f_2(0,0,0)\,\mu_{C\to f_2}(0) + f_2(0,0,1)\,\mu_{C\to f_2}(1)\} +$
$= \mu_{B\to f_2}(1)\{f_2(0,1,0)\,\mu_{C\to f_2}(0) + f_2(0,1,1)\,\mu_{C\to f_2}(1)\}$

$= 3(5\times 13310 + 4\times 1331) + (5\times 13310 + 8\times 1331)$

$= 220\times 13310 + 20\times 1331 = 1331\times 220$

$\mu_{f_2\to A}(1) = \mu_{B\to f_2}(0)\{f_2(1,0,0)\,\mu_{C\to f_2}(0) + f_2(1,0,1)\,\mu_{C\to f_2}(1)\} +$
$= \mu_{B\to f_2}(1)\{f_2(1,1,0)\,\mu_{C\to f_2}(0) + f_2(1,1,1)\,\mu_{C\to f_2}(1)\}$

$= 3\times(1\times 13310 + 2\times 1331) + (13310 + 3\times 1331)$

$= 4\times 13310 + 9\times 1331$

$= 49\times 1331$

$Z = \sum_A \mu_{f_1\to A}(A)$

$= 49\times 1331 + 220\times 1331$

$= 269\times 1331$

$P(A=0) = \dfrac{\mu_{f_2\to A}(0)}{Z}$

$= \dfrac{220}{269}$

$P(A=1) = \dfrac{\mu_{f_2\to A}(1)}{Z} = \dfrac{49}{269}$

12

$\mu_{A\to f_2} = 1 \quad \forall A$

13

$\mu_{f_2\to c}(c) = \sum_{A,B} f_2(A,B,c)\,\mu_{A\to f_2}(A)\cdot\mu_{B\to f_2}(B)$

$\mu_{f_2\to c}(0) = \mu_{B\to f_2}(0)\big(f_2(0,0,0) + f_2(1,0,0)\big)$
$+ \mu_{B\to f_2}(1)\big(f_2(0,1,0) + f_2(1,1,0)\big)$

$= 3\cdot(5+1) + 1(5+1)$

$= 18 + 6 = 24$

$\mu_{f_2\to c}(1) = \mu_{B\to f_2}(0)\big(f_2(0,0,1) + f_2(1,0,1)\big)$
$+ \mu_{B\to f_2}(1)\big(f_2(0,1,1) + f_2(1,1,1)\big)$

$= 3\times(4+2) + (8+3)$

$= 3\times 6 + 11 = 29$

$P(c=0) = \dfrac{\mu_{f_2\to c}(0)\;\mu_{f_3\to c}(0)}{Z}$

$= \dfrac{24\times \cancel{13310}^{\,10}}{269\times \cancel{1331}} = \dfrac{240}{269}$

$$P(c=1) = \frac{29 \times \cancel{1331}}{269 \times \cancel{1331}} = \frac{29}{269}$$

⑥

$$P(a,b,c,d,e,f,g) = P(a|c,d) \cdot P(b) \cdot P(f) \cdot P(c|f)\, P(d|f,b)\, P(g|d,e) P(e)$$

a	$P(a\|c,d)$	$r_E(c,d) = \sum_a P(a\|c,d) = 1$
b	$P(b) \cdot P(d\|f,b)$	$r_B(d,f)$, $r_C(d,f)$
c	$P(c\|f)$	
d	$P(g\|d,e)$	$r_D(g,e,f)$
e	$P(e)$	$r_E(g,f)$
f	$P(f)$	$r_f(g) = \sum_f P(f) \cdot r_E(g,f)$
g		

$$P(g) = r_f(g)$$

$$r_B(d,f) = \sum_b P(b) \cdot P(d|f,b)$$

$$r_C(d,f) = \sum_c P(c|f) = 1$$

$$r_D(g,e,f) = \sum_d P(g|d,e)\, r_B(d,f)$$

$$r_E(g,f) = \sum_e P(e) \cdot r_D(g,e,f)$$

$$P(g) = r_f(g)$$

$$= \sum_f P(f) \sum_e P(e) \sum_d P(g|d,e) \sum_b P(b) \cdot P(d|f,b)$$