## Exercise 1. Conversion to factor graphs

(a) Draw an undirected graph and an undirected factor graph for $p\left(x_{1}, x_{2}, x_{3}\right)=p\left(x_{1}\right) p\left(x_{2}\right) p\left(x_{3} \mid x_{1}, x_{2}\right)$

## Solution.


(b) Draw an undirected factor graph for the directed graphical model defined by the graph below.


Solution. The graph specifies probabilistic models that factorise as

$$
p\left(x_{1}, \ldots, x_{4}, y_{1}, \ldots, y_{4}\right)=p\left(x_{1}\right) p\left(y_{1} \mid x_{1}\right) \prod_{i=2}^{4} p\left(y_{i} \mid x_{i}\right) p\left(x_{i} \mid x_{i-1}\right)
$$

It is the graph for a Hidden Markov model. The corresponding factor graph is shown below.

(c) Draw the moralised graph and an undirected factor graph for directed graphical models defined by the graph below (this kind of graph is called a polytree: there are no loops but a node may have more than one parent).


Solution. The moral graph is obtained by connecting the parents of the collider node $x_{4}$. See the graph on the left in the figure below.
For the factor graph, we note that the directed graph defines the following class of probabilistic models

$$
p\left(x_{1}, \ldots x_{6}\right)=p\left(x_{1}\right) p\left(x_{2}\right) p\left(x_{3} \mid x_{1}\right) p\left(x_{4} \mid x_{1}, x_{2}\right) p\left(x_{5} \mid x_{4}\right) p\left(x_{6} \mid x_{4}\right)
$$

This gives the factor graph on right in the figure below.



Note:

- The moral graph contains a loop while the factor graph does not. The factor graph is still a polytree. This can be exploited for inference.
- One may choose to group some factors together in order to obtain a factor graph with a particular structure (see factor graph below)



## Exercise 2. Sum-product message passing

We here re-consider the factor tree from the lecture on exact inference.


Let all variables be binary, $x_{i} \in\{0,1\}$, and the factors be defined as follows:

(a) Mark the graph with arrows indicating all messages that need to be computed for the computation of $p\left(x_{1}\right)$.

## Solution.


(b) Compute the messages that you have identified.

Assuming that the computation of the messages is scheduled according to a common clock, group the messages together so that all messages in the same group can be computed in parallel during a clock cycle.

Solution. Since the variables are binary, each message can be represented as a twodimensional vector. We use the convention that the first element of the vector corresponds to the message for $x_{i}=0$ and the second element to the message for $x_{i}=1$. For example,

$$
\begin{equation*}
\mu_{\phi_{A} \rightarrow x_{1}}=\binom{2}{4} \tag{S.1}
\end{equation*}
$$

means that the message $\mu_{\phi_{A} \rightarrow x_{1}}\left(x_{1}\right)$ equals 2 for $x_{1}=0$, i.e. $\mu_{\phi_{A} \rightarrow x_{1}}(0)=2$.
The following figure shows a grouping (scheduling) of the computation of the messages.


## Clock cycle 1:

$$
\begin{equation*}
\mu_{\phi_{A} \rightarrow x_{1}}=\binom{2}{4} \quad \mu_{\phi_{B} \rightarrow x_{2}}=\binom{4}{4} \quad \mu_{x_{4} \rightarrow \phi_{D}}=\binom{1}{1} \quad \mu_{\phi_{F} \rightarrow x_{5}}=\binom{1}{8} \tag{S.2}
\end{equation*}
$$

## Clock cycle 2:

$$
\begin{equation*}
\mu_{x_{2} \rightarrow \phi_{C}}=\mu_{\phi_{B} \rightarrow x_{2}}=\binom{4}{4} \quad \mu_{x_{5} \rightarrow \phi_{E}}=\mu_{\phi_{F} \rightarrow x_{5}}=\binom{1}{8} \tag{S.3}
\end{equation*}
$$

Message $\mu_{\phi_{D} \rightarrow x_{3}}$ is defined as

$$
\begin{equation*}
\mu_{\phi_{D} \rightarrow x_{3}}\left(x_{3}\right)=\sum_{x_{4}} \phi_{D}\left(x_{3}, x_{4}\right) \mu_{x_{4} \rightarrow \phi_{D}}\left(x_{4}\right) \tag{S.4}
\end{equation*}
$$

so that

$$
\begin{align*}
\mu_{\phi_{D} \rightarrow x_{3}}(0) & =\sum_{x_{4}=0}^{1} \phi_{D}\left(0, x_{4}\right) \mu_{x_{4} \rightarrow \phi_{D}}\left(x_{4}\right)  \tag{S.5}\\
& =\phi_{D}(0,0) \mu_{x_{4} \rightarrow \phi_{D}}(0)+\phi_{D}(0,1) \mu_{x_{4} \rightarrow \phi_{D}}(1)  \tag{S.6}\\
& =8 \cdot 1+2 \cdot 1  \tag{S.7}\\
& =10  \tag{S.8}\\
\mu_{\phi_{D} \rightarrow x_{3}}(1) & =\sum_{x_{4}=0}^{1} \phi_{D}\left(1, x_{4}\right) \mu_{x_{4} \rightarrow \phi_{D}}\left(x_{4}\right)  \tag{S.9}\\
& =\phi_{D}(1,0) \mu_{x_{4} \rightarrow \phi_{D}}(0)+\phi_{D}(1,1) \mu_{x_{4} \rightarrow \phi_{D}}(1)  \tag{S.10}\\
& =2 \cdot 1+6 \cdot 1  \tag{S.11}\\
& =8 \tag{S.12}
\end{align*}
$$

and thus

$$
\begin{equation*}
\mu_{\phi_{D} \rightarrow x_{3}}=\binom{10}{8} \tag{S.13}
\end{equation*}
$$

The above computations can be written more compactly in matrix notation. Let $\boldsymbol{\phi}_{\boldsymbol{D}}$ be the matrix that contains the outputs of $\phi_{D}\left(x_{3}, x_{4}\right)$

$$
\phi_{\boldsymbol{D}}=\left(\begin{array}{ll}
\phi_{D}\left(x_{3}=0, x_{4}=0\right) & \phi_{D}\left(x_{3}=0, x_{4}=1\right)  \tag{S.14}\\
\phi_{D}\left(x_{3}=1, x_{4}=0\right) & \phi_{D}\left(x_{3}=1, x_{4}=1\right)
\end{array}\right)=\left(\begin{array}{ll}
8 & 2 \\
2 & 6
\end{array}\right) .
$$

We can then write $\mu_{\phi_{D} \rightarrow x_{3}}$ in terms of a matrix vector product,

$$
\begin{equation*}
\mu_{\phi_{D} \rightarrow x_{3}}=\phi_{D} \mu_{x_{4} \rightarrow \phi_{D}} \tag{S.15}
\end{equation*}
$$

## Clock cycle 3:

Representing the factor $\phi_{E}$ as matrix $\phi_{E}$,

$$
\phi_{E}=\left(\begin{array}{ll}
\phi_{E}\left(x_{3}=0, x_{5}=0\right) & \phi_{E}\left(x_{3}=0, x_{5}=1\right)  \tag{S.16}\\
\phi_{E}\left(x_{3}=1, x_{5}=0\right) & \phi_{E}\left(x_{3}=1, x_{5}=1\right)
\end{array}\right)=\left(\begin{array}{ll}
3 & 6 \\
6 & 3
\end{array}\right),
$$

we can write

$$
\begin{equation*}
\mu_{\phi_{E} \rightarrow x_{3}}\left(x_{3}\right)=\sum_{x_{5}} \phi_{E}\left(x_{3}, x_{5}\right) \mu_{x_{5} \rightarrow \phi_{E}}\left(x_{5}\right) \tag{S.17}
\end{equation*}
$$

as a matrix vector product,

$$
\begin{align*}
\mu_{\phi_{E} \rightarrow x_{3}} & =\phi_{E} \mu_{x_{5} \rightarrow \phi_{E}}  \tag{S.18}\\
& =\left(\begin{array}{ll}
3 & 6 \\
6 & 3
\end{array}\right)\binom{1}{8}  \tag{S.19}\\
& =\binom{51}{30} . \tag{S.20}
\end{align*}
$$

## Clock cycle 4:

Variable node $x_{3}$ has received all incoming messages, and can thus output $\mu_{x_{3} \rightarrow \phi_{C}}$,

$$
\begin{equation*}
\mu_{x_{3} \rightarrow \phi_{C}}\left(x_{3}\right)=\mu_{\phi_{D} \rightarrow x_{3}}\left(x_{3}\right) \mu_{\phi_{E} \rightarrow x_{3}}\left(x_{3}\right) . \tag{S.21}
\end{equation*}
$$

Using $\odot$ to denote element-wise multiplication of two vectors, we have

$$
\begin{align*}
\boldsymbol{\mu}_{x_{3} \rightarrow \phi_{C}} & =\mu_{\phi_{D} \rightarrow x_{3}} \odot \mu_{\phi_{E} \rightarrow x_{3}}  \tag{S.22}\\
& =\binom{10}{8} \odot\binom{51}{30}  \tag{S.23}\\
& =\binom{510}{240} . \tag{S.24}
\end{align*}
$$

## Clock cycle 5:

Factor node $\phi_{C}$ has received all incoming messages, and can thus output $\mu_{\phi_{C} \rightarrow x_{1}}$,

$$
\begin{equation*}
\mu_{\phi_{C} \rightarrow x_{1}}\left(x_{1}\right)=\sum_{x_{2}, x_{3}} \phi_{C}\left(x_{1}, x_{2}, x_{3}\right) \mu_{x_{2} \rightarrow \phi_{C}}\left(x_{2}\right) \mu_{x_{3} \rightarrow \phi_{C}}\left(x_{3}\right) . \tag{S.25}
\end{equation*}
$$

Writing out the sum for $x_{1}=0$ and $x_{1}=1$ gives

$$
\begin{align*}
\mu_{\phi_{C} \rightarrow x_{1}}(0)= & \sum_{x_{2}, x_{3}} \phi_{C}\left(0, x_{2}, x_{3}\right) \mu_{x_{2} \rightarrow \phi_{C}}\left(x_{2}\right) \mu_{x_{3} \rightarrow \phi_{C}}\left(x_{3}\right)  \tag{S.26}\\
= & \left.\phi_{C}\left(0, x_{2}, x_{3}\right) \mu_{x_{2} \rightarrow \phi_{C}}\left(x_{2}\right) \mu_{x_{3} \rightarrow \phi_{C}}\left(x_{3}\right)\right|_{\left(x_{2}, x_{3}\right)=(0,0)}+  \tag{S.27}\\
& \left.\phi_{C}\left(0, x_{2}, x_{3}\right) \mu_{x_{2} \rightarrow \phi_{C}}\left(x_{2}\right) \mu_{x_{3} \rightarrow \phi_{C}}\left(x_{3}\right)\right|_{\left(x_{2}, x_{3}\right)=(1,0)}+  \tag{S.28}\\
& \left.\phi_{C}\left(0, x_{2}, x_{3}\right) \mu_{x_{2} \rightarrow \phi_{C}}\left(x_{2}\right) \mu_{x_{3} \rightarrow \phi_{C}}\left(x_{3}\right)\right|_{\left(x_{2}, x_{3}\right)=(0,1)}+  \tag{S.29}\\
& \left.\phi_{C}\left(0, x_{2}, x_{3}\right) \mu_{x_{2} \rightarrow \phi_{C}}\left(x_{2}\right) \mu_{x_{3} \rightarrow \phi_{C}}\left(x_{3}\right)\right|_{\left(x_{2}, x_{3}\right)=(1,1)}  \tag{S.30}\\
= & 4 \cdot 4 \cdot 510+  \tag{S.31}\\
& 2 \cdot 4 \cdot 510+  \tag{S.32}\\
& 2 \cdot 4 \cdot 240+  \tag{S.33}\\
& 6 \cdot 4 \cdot 240  \tag{S.34}\\
= & 19920  \tag{S.35}\\
\mu_{\phi_{C} \rightarrow x_{1}}(1)= & \sum_{x_{2}, x_{3}} \phi_{C}\left(1, x_{2}, x_{3}\right) \mu_{x_{2} \rightarrow \phi_{C}}\left(x_{2}\right) \mu_{x_{3} \rightarrow \phi_{C}}\left(x_{3}\right)  \tag{S.36}\\
= & \left.\phi_{C}\left(1, x_{2}, x_{3}\right) \mu_{x_{2} \rightarrow \phi_{C}}\left(x_{2}\right) \mu_{x_{3} \rightarrow \phi_{C}}\left(x_{3}\right)\right|_{\left(x_{2}, x_{3}\right)=(0,0)}+  \tag{S.37}\\
& \left.\phi_{C}\left(1, x_{2}, x_{3}\right) \mu_{x_{2} \rightarrow \phi_{C}}\left(x_{2}\right) \mu_{x_{3} \rightarrow \phi_{C}}\left(x_{3}\right)\right|_{\left(x_{2}, x_{3}\right)=(1,0)}+  \tag{S.38}\\
& \left.\phi_{C}\left(1, x_{2}, x_{3}\right) \mu_{x_{2} \rightarrow \phi_{C}}\left(x_{2}\right) \mu_{x_{3} \rightarrow \phi_{C}}\left(x_{3}\right)\right|_{\left(x_{2}, x_{3}\right)=(0,1)}+  \tag{S.39}\\
& \left.\phi_{C}\left(1, x_{2}, x_{3}\right) \mu_{x_{2} \rightarrow \phi_{C}}\left(x_{2}\right) \mu_{x_{3} \rightarrow \phi_{C}}\left(x_{3}\right)\right|_{\left(x_{2}, x_{3}\right)=(1,1)}=  \tag{S.40}\\
= & 2 \cdot 4 \cdot 510+  \tag{S.41}\\
& 6 \cdot 4 \cdot 510+  \tag{S.42}\\
& 6 \cdot 4 \cdot 240+  \tag{S.43}\\
& 4 \cdot 4 \cdot 240 \tag{S.44}
\end{align*}
$$

and hence

$$
\begin{equation*}
\mu_{\phi_{C} \rightarrow x_{1}}=\binom{19920}{25920} \tag{S.46}
\end{equation*}
$$

After step 5 , variable node $x_{1}$ has received all incoming messages and the marginal can be computed.
In addition to the messages needed for computation of $p\left(x_{1}\right)$ one can compute all messages in the graph in five clock cycles, see Figure 1. This means that all marginals, as well as the joints of those variables sharing a factor node, are available after five clock cycles.
(c) What is $p\left(x_{1}=1\right)$ ?

Solution. We compute the marginal $p\left(x_{1}\right)$ as

$$
\begin{equation*}
p\left(x_{1}\right) \propto \mu_{\phi_{A} \rightarrow x_{1}}\left(x_{1}\right) \mu_{\phi_{C} \rightarrow x_{1}}\left(x_{1}\right) \tag{S.47}
\end{equation*}
$$



Figure 1: Answer to Exercise 2 Question (b): Computing all messages in five clock cycles. If we also computed the messages toward the leaf factor nodes, we needed six cycles, but they are not necessary for computation of the marginals so they are omitted.
which is in vector notation

$$
\begin{align*}
\binom{p\left(x_{1}=0\right)}{p\left(x_{1}=1\right)} & \propto \boldsymbol{\mu}_{\phi_{\boldsymbol{A}} \rightarrow x_{\mathbf{1}}} \odot \boldsymbol{\mu}_{\phi_{\boldsymbol{C}} \rightarrow x_{\mathbf{1}}}  \tag{S.48}\\
& \propto\binom{2}{4} \odot\binom{19920}{25920}  \tag{S.49}\\
& \propto\binom{39840}{103680} \tag{S.50}
\end{align*}
$$

Normalisation gives

$$
\begin{align*}
\binom{p\left(x_{1}=0\right)}{p\left(x_{1}=1\right)} & =\frac{1}{39840+103680}\binom{39840}{103680}  \tag{S.51}\\
& =\binom{0.2776}{0.7224} \tag{S.52}
\end{align*}
$$

so that $p\left(x_{1}=1\right)=0.7224$.
Note the relatively large numbers in the messages that we computed. In other cases, one may obtain very small ones depending on the scale of the factors. This can cause numerical issues that can be addressed by working in the logarithmic domain (see Barber's paragraph on log messages, p86 in his book)
(d) Draw the factor graph corresponding to $p\left(x_{1}, x_{3}, x_{4}, x_{5} \mid x_{2}=1\right)$ and provide the numerical values for all factors.

Solution. The pmf represented by the original factor graph is

$$
p\left(x_{1}, \ldots, x_{5}\right) \propto \phi_{A}\left(x_{1}\right) \phi_{B}\left(x_{2}\right) \phi_{C}\left(x_{1}, x_{2}, x_{3}\right) \phi_{D}\left(x_{3}, x_{4}\right) \phi_{E}\left(x_{3}, x_{5}\right) \phi_{F}\left(x_{5}\right)
$$

The conditional $p\left(x_{1}, x_{3}, x_{4}, x_{5} \mid x_{2}=1\right)$ is proportional to $p\left(x_{1}, \ldots, x_{5}\right)$ with $x_{2}$ fixed to $x_{2}=1$, i.e.

$$
\begin{align*}
p\left(x_{1}, x_{3}, x_{4}, x_{5} \mid x_{2}=1\right) & \propto p\left(x_{1}, x_{2}=1, x_{3}, x_{4}, x_{5}\right)  \tag{S.53}\\
& \propto \phi_{A}\left(x_{1}\right) \phi_{B}\left(x_{2}=1\right) \phi_{C}\left(x_{1}, x_{2}=1, x_{3}\right) \phi_{D}\left(x_{3}, x_{4}\right) \phi_{E}\left(x_{3}, x_{5}\right) \phi_{F}\left(x_{5}\right) \tag{S.54}
\end{align*}
$$

$$
\begin{equation*}
\propto \phi_{A}\left(x_{1}\right) \phi_{C}^{x_{2}}\left(x_{1}, x_{3}\right) \phi_{D}\left(x_{3}, x_{4}\right) \phi_{E}\left(x_{3}, x_{5}\right) \phi_{F}\left(x_{5}\right) \tag{S.55}
\end{equation*}
$$

where $\phi_{C}^{x_{2}}\left(x_{1}, x_{3}\right)=\phi_{C}\left(x_{1}, x_{2}=1, x_{3}\right)$. The numerical values of $\phi_{C}^{x_{2}}\left(x_{1}, x_{3}\right)$ can be read from the table defining $\phi_{C}\left(x_{1}, x_{2}, x_{3}\right)$, extracting those rows where $x_{2}=1$,

|  | $x_{1}$ | $x_{2}$ | $x_{3}$ | $\phi_{C}$ |
| :--- | :--- | :--- | :--- | :--- |
|  | 0 | 0 | 0 | 4 |
|  | 1 | 0 | 0 | 2 |
| $\rightarrow$ | 0 | 1 | 0 | 2 |
| $\rightarrow$ | 1 | 1 | 0 | 6 |
|  | 0 | 0 | 1 | 2 |
|  | 1 | 0 | 1 | 6 |
| $\rightarrow$ | 0 | 1 | 1 | 6 |
| $\rightarrow$ | 1 | 1 | 1 | 4 |


| $x_{1}$ | $x_{3}$ | $\phi_{C}^{x_{2}}$ |
| :--- | :--- | :--- |
| 0 | 0 | 2 |
| 1 | 0 | 6 |
| 0 | 1 | 6 |
| 1 | 1 | 4 |

The factor graph for $p\left(x_{1}, x_{3}, x_{4}, x_{5} \mid x_{2}=1\right)$ is shown below. Factor $\phi_{B}$ has disappeared since it only depended on $x_{2}$ and thus became a constant. Factor $\phi_{C}$ is replaced by $\phi_{C}^{x_{2}}$ defined above. The remaining factors are the same as in the original factor graph.

(e) Compute $p\left(x_{1}=1 \mid x_{2}=1\right)$, re-using messages that you have already computed for the evaluation of $p\left(x_{1}=1\right)$.

Solution. The message $\mu_{\phi_{A} \rightarrow x_{1}}$ is the same as in the original factor graph and $\mu_{x_{3} \rightarrow \phi_{C}^{x_{2}}}=$ $\mu_{x_{3} \rightarrow \phi_{C}}$. This is because the outgoing message from $x_{3}$ corresponds to the effective factor obtained by summing out all variables in the sub-trees attached to $x_{3}$ (without the $\phi_{C}^{x_{2}}$ branch), and these sub-trees do not depend on $x_{2}$.
The message $\mu_{\phi_{C}^{x_{2}} \rightarrow x_{1}}$ needs to be newly computed. We have

$$
\begin{equation*}
\mu_{\phi_{C}^{x_{2}} \rightarrow x_{1}}\left(x_{1}\right)=\sum_{x_{3}} \phi_{C}^{x_{2}}\left(x_{1}, x_{3}\right) \mu_{x_{3} \rightarrow \phi_{C}^{x_{2}}} \tag{S.56}
\end{equation*}
$$

or in vector notation

$$
\begin{align*}
\boldsymbol{\mu}_{\boldsymbol{\phi}_{C}^{x_{2}} \rightarrow x_{1}} & =\boldsymbol{\phi}_{C}^{\boldsymbol{x}_{2}} \boldsymbol{\mu}_{x_{3} \rightarrow \phi_{C}^{x_{2}}}  \tag{S.57}\\
& =\left(\begin{array}{ll}
\phi_{C}^{x_{2}}\left(x_{1}=0, x_{3}=0\right) & \phi_{C}^{x_{2}}\left(x_{1}=0, x_{3}=1\right) \\
\phi_{C}^{x_{2}}\left(x_{1}=1, x_{3}=0\right) & \phi_{C}^{x_{2}}\left(x_{1}=1, x_{3}=1\right)
\end{array}\right) \boldsymbol{\mu}_{x_{3} \rightarrow \phi_{C}^{x_{2}}}  \tag{S.58}\\
& =\left(\begin{array}{ll}
2 & 6 \\
6 & 4
\end{array}\right)\binom{510}{240}  \tag{S.59}\\
& =\binom{2460}{4020} \tag{S.60}
\end{align*}
$$

We thus obtain for the marginal posterior of $x_{1}$ given $x_{2}=1$ :

$$
\begin{align*}
\binom{p\left(x_{1}=0 \mid x_{2}=1\right)}{p\left(x_{1}=1 \mid x_{2}=1\right)} & \propto \boldsymbol{\mu}_{\phi_{\boldsymbol{A}} \rightarrow x_{\mathbf{1}}} \odot \boldsymbol{\mu}_{\boldsymbol{\phi}_{\boldsymbol{C}}^{\boldsymbol{x}_{2}} \rightarrow \boldsymbol{x}_{\mathbf{1}}}  \tag{S.61}\\
& \propto\binom{2}{4} \odot\binom{2460}{4020}  \tag{S.62}\\
& \propto\binom{4920}{16080} . \tag{S.63}
\end{align*}
$$

Normalisation gives

$$
\begin{equation*}
\binom{p\left(x_{1}=0 \mid x_{2}=1\right)}{p\left(x_{1}=1 \mid x_{2}=1\right)}=\binom{0.2343}{0.7657} \tag{S.64}
\end{equation*}
$$

and thus $p\left(x_{1}=1 \mid x_{2}=1\right)=0.7657$. The posterior probability is slightly larger than the prior probability, $p\left(x_{1}=1\right)=0.7224$.

