ECE521 Winter 2017: Assignment 4

FuYuan Tee, (999295837) * Chee Loong Soon, (999295793) †

April 9th, 2017

Contents

*Equal Contribution (50%), fuyuan.tee@mail.utoronto.ca

†Equal Contribution (50%), cheeloong.soon@mail.utoronto.ca

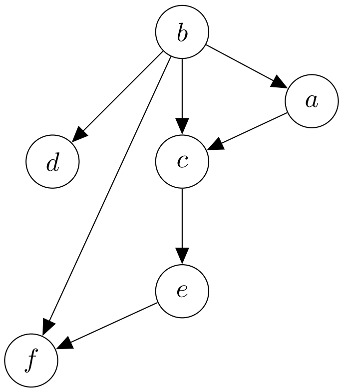
1 Graphical Models

1.1 Graphical models from factorization

$$
\begin{aligned}
P(a, b, c, d, e, f,) &= P(a|b)P(b)P(c|a, b)P(d|b)P(e|c)P(f|b, e) \\
&= P(f|b, e)P(e|c)P(c|a, b)P(a|b)P(d|b)P(b)
\end{aligned}
\tag{1}
$$

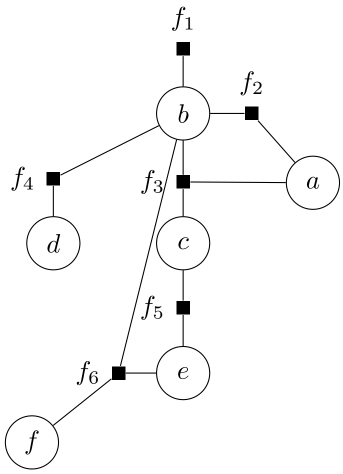
1.1.1 Sketch Bayesian Networks

Bayesian Network corresponding to joint distribution 1.



1.1.2 Sketch Factor Graph

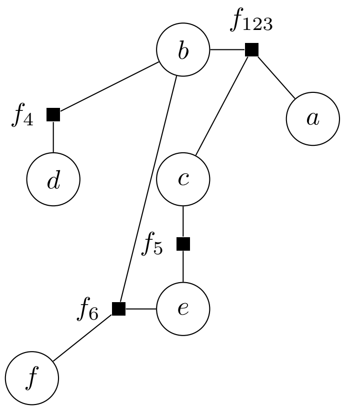
Factor Graph corresponding to joint distribution 1.



The factors with corresponding distribution are shown in 2.

$$\begin{aligned}
f_1(b) &= P(b) \\
f_2(a, b) &= P(a|b) \\
f_3(a, b, c) &= P(c|a, b) \\
f_4(b, d) &= P(d|b) \\
f_5(c, e) &= P(e|c) \\
f_6(b, e, f) &= P(f|b, e)
\end{aligned} \tag{2}$$

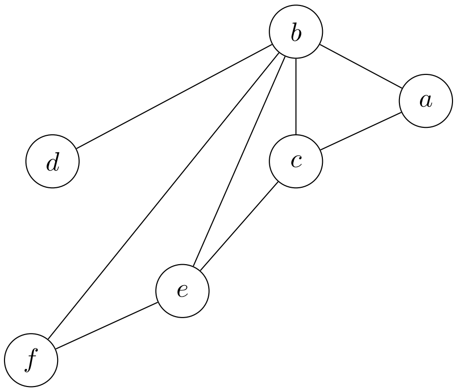
According to ECE521 Winter 2017 Tutorial 8 page 33, this can be simplified to



The factors with corresponding distribution corresponding to the simplified Factor Graph shown in equation 3.

$$
\begin{aligned}
f_{123}(a,b,c) &= P(c|a,b)P(a|b)P(b) \\
f_4(b,d) &= P(d|b) \\
f_5(c,e) &= P(e|c) \\
f_6(e,f) &= P(f|b,e)
\end{aligned}
\tag{3}
$$

1.1.3 Sketch Markov Random Field



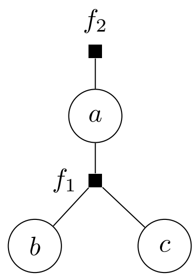
The equations are shown in 4. The parameters to each maximum cliques are the vertex to each maximal clique.

$$
\begin{aligned}
\psi_1(a,b,c) &= P(c|a,b)P(a|b)P(b) \\
\psi_2(b,c,e) &= P(e|c) \\
\psi_3(b,e,f) &= P(f|b,e) \\
\psi_4(b,d) &= P(d|b)
\end{aligned}
\tag{4}
$$

1.2 Conversion between graphical models

1.2.1 Factor Graph

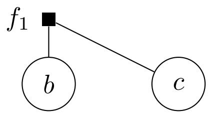
Factor Graph (a)



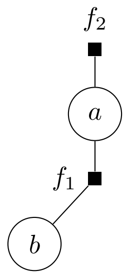
The conditional independence properties for Factor Graph (a) are

$$
\begin{aligned}
&b \not\!\perp\!\!\!\perp c | a \\
&a \not\!\perp\!\!\!\perp b | c \\
&a \not\!\perp\!\!\!\perp c | b
\end{aligned}
\tag{5}
$$

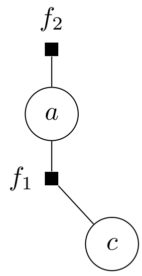
$b \not\!\perp\!\!\!\perp c | a$



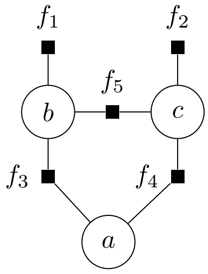
$a \not\!\perp\!\!\!\perp b | c$



$a \not\!\perp\!\!\!\perp c | b$



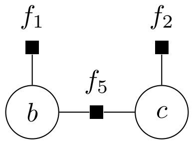
Factor Graph (b)



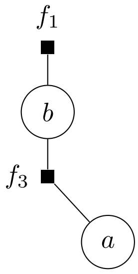
The conditional independence properties for Factor Graph b) are

$$
\begin{aligned}
b &\not\perp\!\!\!\perp c \mid a \\
a &\not\perp\!\!\!\perp b \mid c \\
a &\not\perp\!\!\!\perp c \mid b
\end{aligned}
\tag{6}
$$

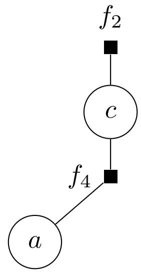
$b \not\perp\!\!\!\perp c | a$



$a \not\perp\!\!\!\perp b | c$



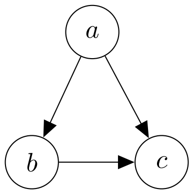
$a \not\perp\!\!\!\perp c | b$



1.2.1.1 Factor Graph to BN

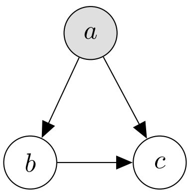
1.2.1.1.1 Factor Graph (a) to BN Converting to Bayesian Networks while maintaining conditional independence in 5 has two different solutions according to conversion rules given in class.

Solution 1:



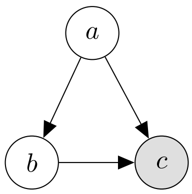
$$\begin{aligned} P(a,b,c) = P(c|a,b)P(b|a)P(a) &= \frac{1}{Z} f_1(a,b,c) f_2(a) \\ Z &= \sum_{a,b,c} f_1(a,b,c) f_2(a) \end{aligned} \tag{7}$$

$b \not\perp\!\!\!\perp c|a$



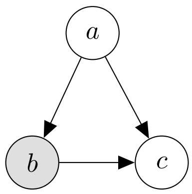
Cascade from b to c

$a \not\perp\!\!\!\perp b|c$



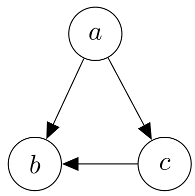
Cascade from a to b. V-structure from c to a and b.

$a \not\perp\!\!\!\perp c|b$



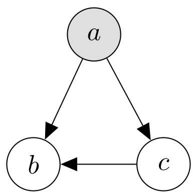
Cascade from a to c

Solution 2:



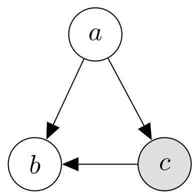
$$
\begin{aligned}
P(a,b,c) = P(b|a,c)P(c|a)P(a) &= \frac{1}{Z} f_1(a,b,c) f_2(a) \\
Z &= \sum_{a,b,c} f_1(a,b,c) f_2(a)
\end{aligned} \tag{8}
$$

$b \not\perp\!\!\!\perp c|a$



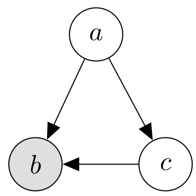
Cascade from c to b

$a \not\perp\!\!\!\perp b|c$



Cascade from a to b

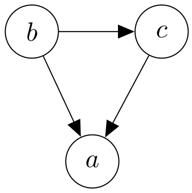
$a \not\perp\!\!\!\perp c|b$



Cascade from a to c. V-structure from b to a and c.

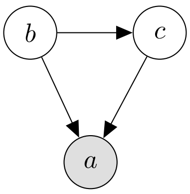
1.2.1.1.2 Factor Graph (b) to BN Converting to Bayesian Networks while maintaining conditional independence in 6 has two different solutions.

Solution 1:



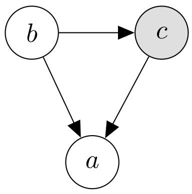
$$
\begin{aligned}
P(a,b,c) = P(a|b,c)P(c|b)P(b) &= \frac{1}{Z} f_1(b) f_2(c) f_3(a,b) f_4(a,c) f_5(b,c) \\
Z &= \sum_{a,b,c} f_1(b) f_2(c) f_3(a,b) f_4(a,c) f_5(b,c)
\end{aligned} \tag{9}
$$

$b \not\perp\!\!\!\perp c | a$



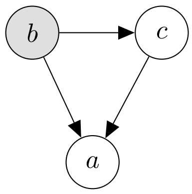
Cascade from b to c, V-structure from a to b and c

$a \not\perp\!\!\!\perp b | c$



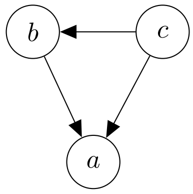
Cascade from b to a

$a \not\perp\!\!\!\perp c | b$



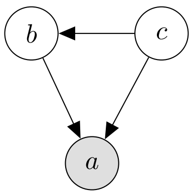
Cascade from c to a

Solution 2:



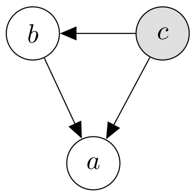
$$
\begin{aligned}
P(a,b,c) = P(a|b,c)P(b|c)P(c) &= \frac{1}{Z} f_1(b) f_2(c) f_3(a,b) f_4(a,c) f_5(b,c) \\
Z &= \sum_{a,b,c} f_1(b) f_2(c) f_3(a,b) f_4(a,c) f_5(b,c)
\end{aligned} \tag{10}
$$

$b \not\perp\!\!\!\perp c | a$



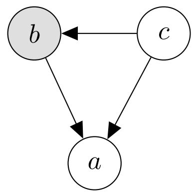
Cascade from c to b, V-structure from a to b and c

$a \not\perp\!\!\!\perp b | c$



Cascade from b to a

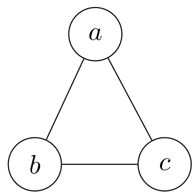
$a \not\perp\!\!\!\perp c | b$



Cascade from c to a

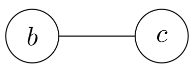
1.2.1.2 Factor Graph to MRF

1.2.1.2.1 Factor Graph (a) to MRF Converting to Markov Random Field while maintaining conditional independence in 5.

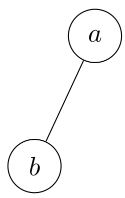


$$\psi_1(a, b, c) = f_1(a, b, c) f_2(a) \tag{11}$$

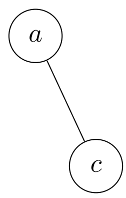
$b \not\perp\!\!\!\perp c | a$



$a \not\perp\!\!\!\perp b | c$

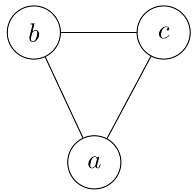


$a \not\perp\!\!\!\perp c | b$



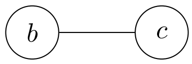
1.2.1.2.2 Factor Graph (b) to MRF

Converting to Markov Random Field while maintaining conditional independence in 6.

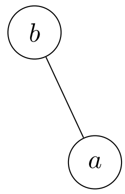


$$\psi_1(a,b,c) = f_1(b)f_2(c)f_3(a,b)f_4(a,c)f_5(b,c) \tag{12}$$

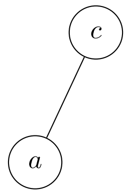
$b \not\perp\!\!\!\perp c|a$



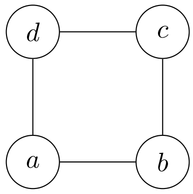
$a \not\perp\!\!\!\perp b|c$



$a \not\perp\!\!\!\perp c|b$



1.2.2 Markov Random Field

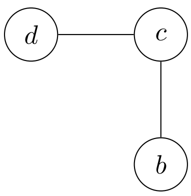


The conditional independence properties for Markov Random Field are shown in equations 14 (conditioned upon 1 variable) and 15 (conditioned upon 2 variables). Equation 13 shows the marginal independence properties. However, we have assumed in this questions that the marginal independence properties do not have to be maintained. Instead, we seek for generality by showing we can convert to Factor Graph while maintaining the marginal independences in 13 whereas we show we cannot convert to Bayesian Networks regardless of whether or not we maintain marginal independence properties in 13.

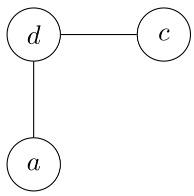
$$
\begin{aligned}
a \not\perp\!\!\!\perp b|\emptyset &= a \not\perp\!\!\!\perp b \\
a \not\perp\!\!\!\perp c|\emptyset &= a \not\perp\!\!\!\perp c \\
a \not\perp\!\!\!\perp d|\emptyset &= a \not\perp\!\!\!\perp d \\
b \not\perp\!\!\!\perp c|\emptyset &= b \not\perp\!\!\!\perp c \\
b \not\perp\!\!\!\perp d|\emptyset &= b \not\perp\!\!\!\perp d \\
c \not\perp\!\!\!\perp d|\emptyset &= c \not\perp\!\!\!\perp d
\end{aligned}
\tag{13}
$$

$$
\begin{gathered}
b \not\perp\!\!\!\perp c \mid a \\
b \not\perp\!\!\!\perp d \mid a \\
c \not\perp\!\!\!\perp d \mid a \\
a \not\perp\!\!\!\perp c \mid b \\
a \not\perp\!\!\!\perp d \mid b \\
c \not\perp\!\!\!\perp d \mid b \\
a \not\perp\!\!\!\perp b \mid c \\
a \not\perp\!\!\!\perp d \mid c \\
b \not\perp\!\!\!\perp d \mid c \\
a \not\perp\!\!\!\perp b \mid d \\
a \not\perp\!\!\!\perp c \mid d \\
b \not\perp\!\!\!\perp c \mid d
\end{gathered} \tag{14}
$$

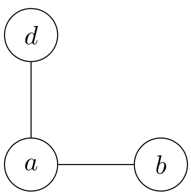
$b \not\perp\!\!\!\perp c|a$ and $b \not\perp\!\!\!\perp d|a$ and $c \not\perp\!\!\!\perp d|a$



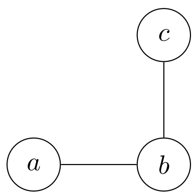
$a \not\perp\!\!\!\perp c|b$ and $a \not\perp\!\!\!\perp d|b$ and $c \not\perp\!\!\!\perp d|b$



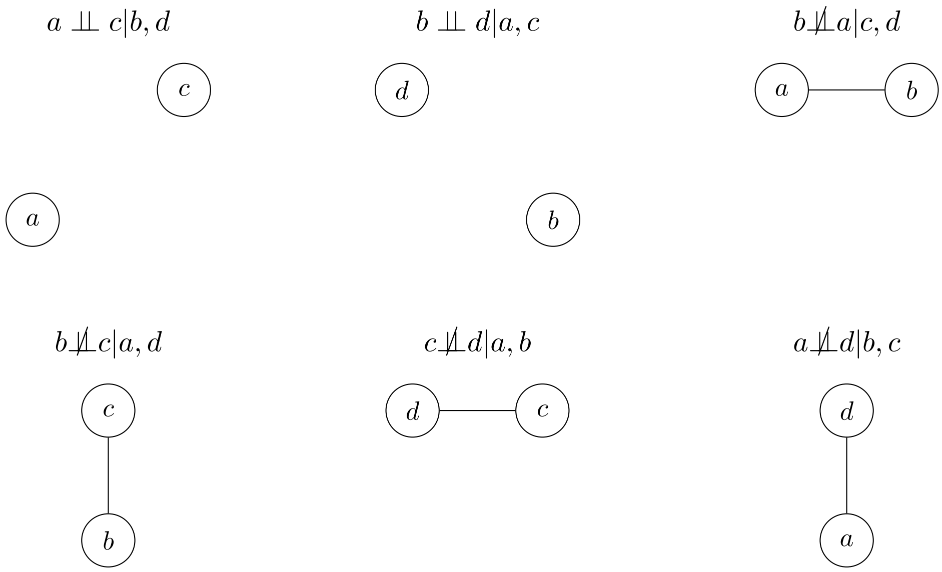
$a \not\perp\!\!\!\perp b|c$ and $a \not\perp\!\!\!\perp d|c$ and $b \not\perp\!\!\!\perp d|c$



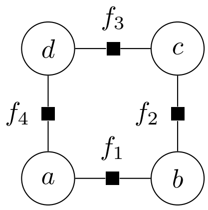
$a \not\perp\!\!\!\perp b|d$ and $a \not\perp\!\!\!\perp c|d$ and $b \not\perp\!\!\!\perp c|d$



$$
\begin{aligned}
a \perp\!\!\!\perp c|b,d \\
b \perp\!\!\!\perp d|a,c \\
b \not\perp\!\!\!\perp a|c,d \\
b \not\perp\!\!\!\perp c|a,d \\
c \not\perp\!\!\!\perp d|a,b \\
a \not\perp\!\!\!\perp d|b,c
\end{aligned}
\tag{15}
$$

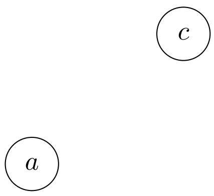


1.2.2.1 MRF to Factor Graph

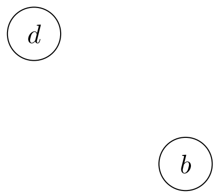


The proof for conditional independence have a similar looking graph as the Markov Random Field itself. Also, according to slide 9 of Lecture 18 ECE521 Winter 2017, it is shown that Markov Random Field is a subset of Factor Graph, meaning that all Markov Random Field can be represented by Factor Graphs. This means that it must always be possible to convert any Markov Random Field to a Factor Graph.

$a \perp\!\!\!\perp c | b, d$



$b \perp\!\!\!\perp d | a, c$



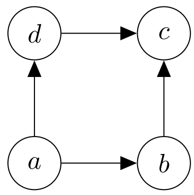
1.2.2.2 MRF to BN

Conversion from Markov Random Field (MRF) above to Bayesian Networks (BN) does not exist. Bayesian Networks are acyclic directed graphs.

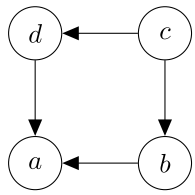
Therefore, below are the all possible factorizations of the MRF to BN which are acyclic and follows the conversion rule given in lectures (every edge that exist in MRF must also exist in the BN to be able to convert it back into an equivalent MRF), each with a different possible acyclic pair of root to leaf paths of the BN.

Since there must be an edge, each edge can be in 1 of 2 directions. There are 4 edges, so there are $2^4 = 16$ possible factorizations. 2 of the factorizations results in directed cycles, which is not a BN by definition. Below we show the remaining 14 ($16 - 2 = 14$) acyclic possible factorizations.

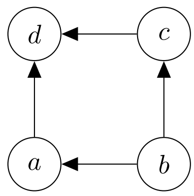
Root a, Leaf c



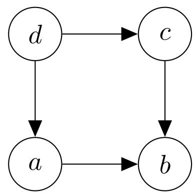
Root c, Leaf a



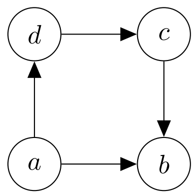
Root b, Leaf d



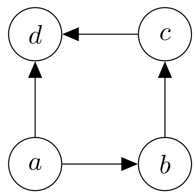
Root d, Leaf b



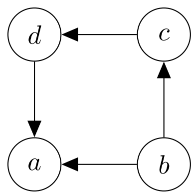
Root a, Leaf b



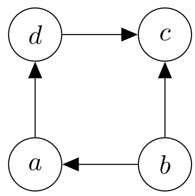
Root a, Leaf d



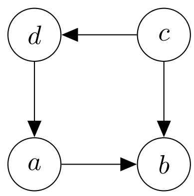
Root b, Leaf a



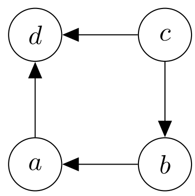
Root b, Leaf c



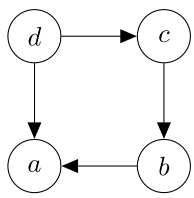
Root c, Leaf b



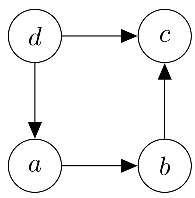
Root c, Leaf d



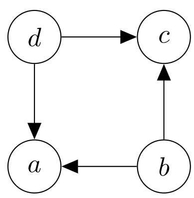
Root d, Leaf a



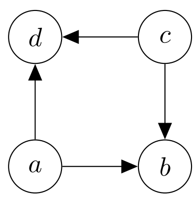
Root d, Leaf c



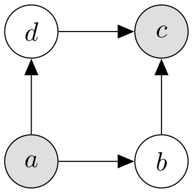
Root b and d, Leaf a and c



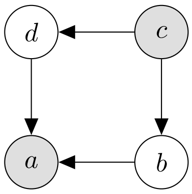
Root a and c, Leaf b and d



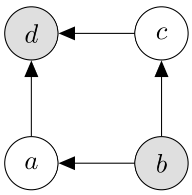
We can Proof by Counter Example that each of the possible factorization above does not satisfy at least one of the conditional independence properties given in equation 15.

Root a, Leaf c: 

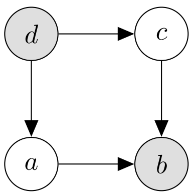
but $b \not\perp\!\!\!\perp d | a, c$ due to V-structure from c to b and d.

Root c, Leaf a: 

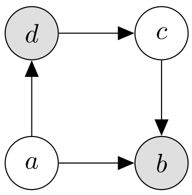
but $b \not\perp\!\!\!\perp d | a, c$ due to V-structure from a to b and d.

Root b, Leaf d: 

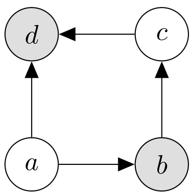
but $a \not\perp\!\!\!\perp c | b, d$ due to V-structure from d to a and c.

Root d, Leaf b: 

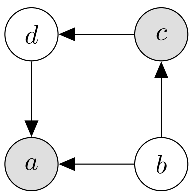
but $a \not\perp\!\!\!\perp c | b, d$ due to V-structure from b to a and c.

Root a, Leaf b: 

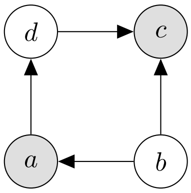
but $a \not\perp\!\!\!\perp c | b, d$ due to V-structure from b to a and c.

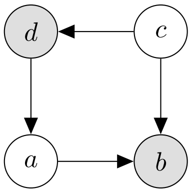
Root a, Leaf d: 

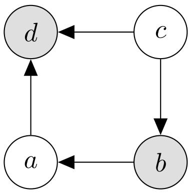
but $a \not\perp\!\!\!\perp c | b, d$ due to V-structure from d to a and c.

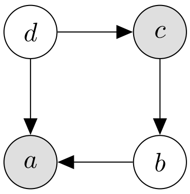
Root b, Leaf a: 

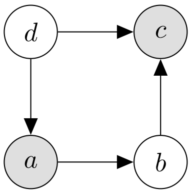
but $b \not\perp\!\!\!\perp d | a, c$ due to V-structure from a to b and d.

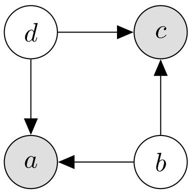
Root b, Leaf c:  but $b \not\perp\!\!\!\perp d | a, c$ due to V-structure from c to b and d.

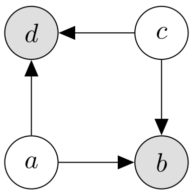
Root c, Leaf b:  but $a \not\perp\!\!\!\perp c | b, d$ due to V-structure from b to a and c.

Root c, Leaf d:  but $a \not\perp\!\!\!\perp c | b, d$ due to V-structure from d to a and c.

Root d, Leaf a:  but $b \not\perp\!\!\!\perp d | a, c$ due to V-structure from a to b and d.

Root d, Leaf c:  but $b \not\perp\!\!\!\perp d | a, c$ due to V-structure from c to b and d.

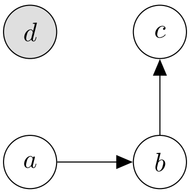
Root b and d, Leaf a and c:  but $b \not\perp\!\!\!\perp d | a, c$ due to V-structure from a and c to b and d.

Root a and c, Leaf b and d:  but $a \not\perp\!\!\!\perp c | b, d$ due to V-structure from b and d to a and c.

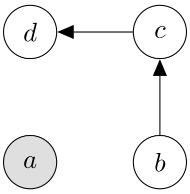
Thus, we have explained that there is no equivalent Bayesian Networks that implies the same conditional independence properties as the Markov Random Field.

This can be concisely explained and proven below.

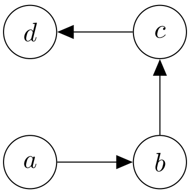
To satisfy $a \not\!\perp\!\!\!\perp c|d$ in equation 14, there needs to be a path from a to c (path from c to a would just be a symmetric argument). The path must go through b.



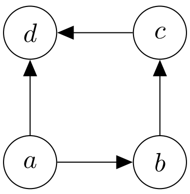
Now, in order to satisfy $b \not\!\perp\!\!\!\perp d|a$ in equation 14, there needs to be a path from b to d. (Path from d to b would just be a symmetric argument). The only possible path is through c.



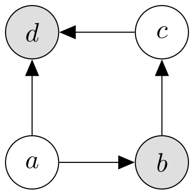
Combining both, we get.



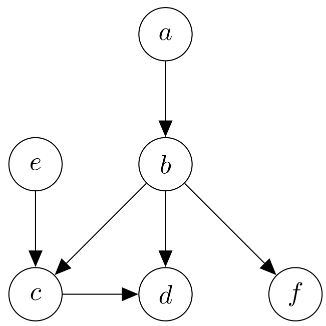
This currently does not satisfy $a \not\!\perp\!\!\!\perp d|b, c$ in equation 15. Therefore, we need to try adding an edge between a and d. The only possible assignment is from a to d as d to a results in an acyclic directed graph which is not a Bayesian Network by definition. Our final factorization results in



However, this does not satisfy $a \perp\!\!\!\perp c|b, d$ as shown below due to V-structure from d to a and c.



Therefore, there is no possible BN factorization for the given MRF as we have shown it is impossible to satisfy all the conditional independence properties.



1.3 Conditional Independence in Bayesian Networks

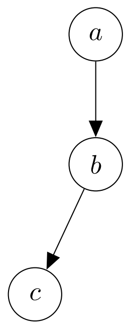
1.3.1 Express Joint Probability of Bayesian Networks

$$P(a, b, c, d, e, f) = P(d|b, c)P(c|b, e)P(e)P(f|b)P(b|a)P(a) \tag{16}$$

1.3.2 Determine TRUE or FALSE

1.3.2.1 $a \perp\!\!\!\perp c$

FALSE

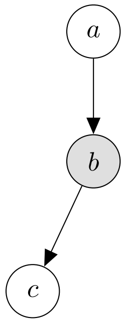


Cascade from a to b to c.

This shows that $a \not\perp\!\!\!\perp c$.

1.3.2.2 $a \perp\!\!\!\perp c \mid b$

TRUE

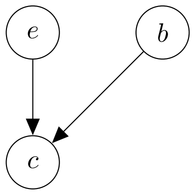


Cascade from a to c is blocked given b.

This shows that $a \perp\!\!\!\perp c \mid b$.

1.3.2.3 $e \perp\!\!\!\perp b$

TRUE

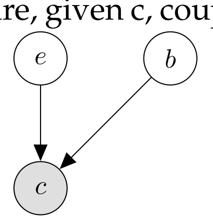


V-structure from b and e to c are blocked if c is not given.

1.3.2.4 $e \perp\!\!\!\perp b \mid c$

FALSE

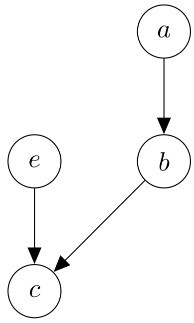
V-structure, given c, couples b and e. This is because b can explain away e with respect to c.



This shows that $e \not\perp\!\!\!\perp b \mid c$.

1.3.2.5 $a \perp\!\!\!\perp e$

TRUE

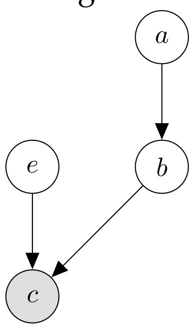


V-structure from a through b and e to c are blocked if c is not given.

1.3.2.6 $a \perp\!\!\!\perp e \mid c$

FALSE

V-structure, given c, couples a through b and e. This is because a through b can explain away e with respect to c.



This shows that $a \not\perp\!\!\!\perp e \mid c$.

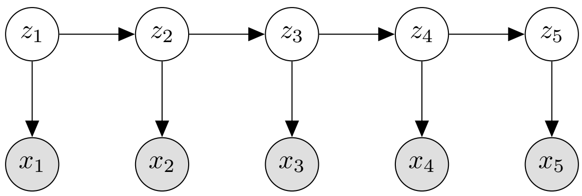
2 Message Passing [practice]

No solution written as practice. Included only so that Table of Contents matches assignment handout.

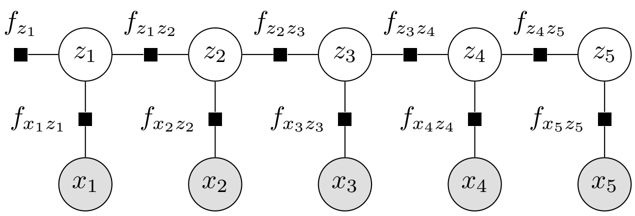
3 Hidden Markov Models

3.1 Factor graph representation

3.1.1 Sketch Bayesian Networks



3.1.2 Sketch Factor Graph



Annotate factors are shown in equations 17, 18, 19.

$$f_{z_1} = P(z_1) \tag{17}$$

$$f_{x_t z_t} = P(x_t|z_t) \qquad t \in \{1,2,3,4,5\} \tag{18}$$

$$f_{z_{t-1} z_t} = P(z_t|z_{t-1}) \quad t \in \{2,3,4,5\} \tag{19}$$

3.2 Inference by passing messages

3.2.1 Computing the message $\mu_{z_4 \to f_{z_3 z_4}}(z_4)$

Applying the variable-to-factor rule,

$$\begin{aligned}\mu_{z_4 \to f_{f_{z_3 z_4}}}(z_4) &= \prod_{f_i \in Ne(z_4) \setminus f_{z_3 z_4}} \mu_{f_i \to z_4}(z_4) \\ &= \mu_{f_{x_4 z_4} \to z_4}(z_4) \cdot \mu_{f_{z_4 z_5} \to z_4}(z_4)\end{aligned} \tag{20}$$

3.3 Message-passing as bi-direction RNNs

The dimension of the matrices and vectors are shown in equation 21. Observed variables can take on M discrete values. Latent variable can take on K latent states.

$$
\begin{aligned}
x_1 &\in \mathbb{R}^{M\times 1} \\
x_2 &\in \mathbb{R}^{M\times 1} \\
x_3 &\in \mathbb{R}^{M\times 1} \\
x_4 &\in \mathbb{R}^{M\times 1} \\
x_5 &\in \mathbb{R}^{M\times 1} \\
W &\in \mathbb{R}^{M\times K} \\
T &\in \mathbb{R}^{K\times K} \\
\pi &\in \mathbb{R}^{K\times 1}
\end{aligned}
\tag{21}
$$

The Emission and Transition matrix are shown in equation 22.

$$
\begin{aligned}
W_{mk} &= P(x_t = m \mid z_t = k) \\
T_{ij} &= P(z_t = i \mid z_{t-1} = j)
\end{aligned}
\tag{22}
$$

3.3.1 Computing vectorized message $\mu_{f_{z_2 z_3} \to z_3}(z_3)$

This is a forward propagation. Hence, we multiply by the Transition Matrix, T.

The messages for priors and observed variables were calculated:

$$\mu_{f_{z_1} \to z_1}(z_1) = \pi \qquad \in \mathbb{R}^{K \times 1} \tag{23}$$

$$\mu_{x_1 \to f_{x_1 z_1}}(x_1) = x_1 \qquad \in \mathbb{R}^{M \times 1} \tag{24}$$

$$\mu_{x_2 \to f_{x_2 z_2}}(x_2) = x_2 \qquad \in \mathbb{R}^{M \times 1} \tag{25}$$

$$f_{x_t z_t}(x_t, z_t) = W^T \qquad \in \mathbb{R}^{K \times M} \qquad t \in \{1, 2, 3, 4, 5\} \tag{26}$$

$$f_{z_{t-1} z_t}(z_{t-1}, z_t) = T \qquad \in \mathbb{R}^{K \times K} \qquad t \in \{2, 3, 4, 5\} \tag{27}$$

The intermediate messages were then calculated:

$$\mu_{f_{x_1 z_1} \to z_1}(z_1) = \sum_{x_1} f_{x_1 z_1}(x_1, z_1) \cdot \mu_{x_1 \to f_{x_1 z_1}}(x_1) = W^T x_1 \in \mathbb{R}^{K \times 1} \tag{28}$$

$$\mu_{f_{x_2 z_2} \to z_2}(z_2) = \sum_{x_2} f_{x_2 z_2}(x_2, z_2) \cdot \mu_{x_2 \to f_{x_2 z_2}}(x_2) = W^T x_2 \in \mathbb{R}^{K \times 1} \tag{29}$$

$$\mu_{z_1 \to f_{z_1 z_2}}(z_1) = \mu_{f_{z_1} \to z_1}(z_1) \cdot \mu_{f_{x_1 z_1} \to z_1}(z_1) \qquad = (\pi) \circ (W^T x_1) \in \mathbb{R}^{K \times 1} \tag{30}$$

$$\mu_{f_{z_1 z_2} \to z_2}(z_2) = \sum_{z_1} f_{z_1 z_2}(z_1, z_2) \cdot \mu_{z_1 \to f_{z_1 z_2}}(z_1) \quad = T\left((\pi) \circ (W^T x_1)\right) \in \mathbb{R}^{K \times 1} \tag{31}$$

$$\mu_{z_2 \to f_{z_2 z_3}}(z_2) = \mu_{f_{z_1 z_2} \to z_2}(z_2) \cdot \mu_{f_{x_2 z_2} \to z_2}(z_2) \qquad = \left[T\left((\pi) \circ (W^T x_1)\right)\right] \circ \left(W^T x_2\right) \in \mathbb{R}^{K \times 1} \tag{32}$$

Finally, the vectorized message was obtained:

$$\mu_{f_{z_2 z_3} \to z_3}(z_3) = \sum_{z_2} f_{z_2 z_3}(z_2, z_3) \cdot \mu_{z_2 \to f_{z_2 z_3}}(z_2) = T\left\{\left[T\left((\pi) \circ (W^T x_1)\right)\right] \circ \left(W^T x_2\right)\right\} \in \mathbb{R}^{K \times 1} \tag{33}$$

where $\circ$ is the Hadamard product. Everything else is matrix multiplication.

3.3.2 Computing vectorized message $\mu_{z_3 \to f_{z_2 z_3}}(z_3)$

This is a backward propagation. Hence, we multiply by the transpose of the Transition Matrix, T^T.

Each observed variable-to-factor, and their corresponding factor-to-latent variable messages were calculated:

$$\mu_{x_t \to f_{x_t z_t}}(x_t) = x_t \qquad \in \mathbb{R}^{M \times 1} \tag{34}$$

$$f_{x_t z_t}(x_t, z_t) = W^T \qquad \in \mathbb{R}^{K \times M} \tag{35}$$

$$f_{z_{t-1} z_t}(z_{t-1}, z_t) = T \qquad \in \mathbb{R}^{K \times K} \tag{36}$$

$$\mu_{f_{x_t z_t} \to z_t}(z_t) = \sum_{x_t} f_{x_t z_t}(x_t, z_t) \cdot \mu_{x_t \to f_{x_t z_t}}(x_t) = W^T x_t \qquad \in \mathbb{R}^{K \times 1} \tag{37}$$

where $t = 3, 4, 5$.

Next, the intermediate messages were calculated:

$$w\mu_{z_5 \to f_{z_4 z_5}}(z_5) = \mu_{f_{x_5 z_5} \to z_5}(z_5) \qquad = (W^T x_5) \in \mathbb{R}^{K \times 1} \tag{38}$$

$$\mu_{f_{z_4 z_5} \to z_4}(z_4) = \sum_{z_5} f_{z_4 z_5}(z_4, z_5) \cdot \mu_{z_5 \to f_{z_4 z_5}}(z_5) = (T^T(W^T x_5)) \in \mathbb{R}^{K \times 1} \tag{39}$$

$$\mu_{z_4 \to f_{z_3 z_4}}(z_4) = \mu_{f_{x_4 z_4} \to z_4}(z_4) \cdot \mu_{f_{z_4 z_5} \to z_4}(z_4) \qquad = \left(W^T x_4\right) \circ \left(T^T(W^T x_5)\right) \in \mathbb{R}^{K \times 1} \tag{40}$$

$$\mu_{f_{z_3 z_4} \to z_3}(z_3) = \sum_{z_4} f_{z_3 z_4}(z_3, z_4) \cdot \mu_{z_4 \to f_{z_3 z_4}}(z_4) = T^T \left[\left(W^T x_4\right) \circ \left(T^T(W^T x_5)\right)\right] \in \mathbb{R}^{K \times 1} \tag{41}$$

Finally, the vectorized message was obtained:

$$\mu_{z_3 \to f_{z_2 z_3}}(z_3) = \mu_{f_{x_3 z_3} \to z_3}(z_3) \cdot \mu_{f_{z_3 z_4} \to z_3}(z_3) = \left(W^T x_3\right) \circ \left\{T^T \left[\left(W^T x_4\right) \circ \left(T^T(W^T x_5)\right)\right]\right\} \in \mathbb{R}^{K \times 1} \tag{42}$$

where $\circ$ is the Hadamard product. Everything else is matrix multiplication.