Probabilistic Graph Models:
Factor Graphs Inference Using Belief Propagation

ECE/CS 498 DS U/G

Lecture 19

Ravi K. Iyer

Dept. of Electrical and Computer Engineering

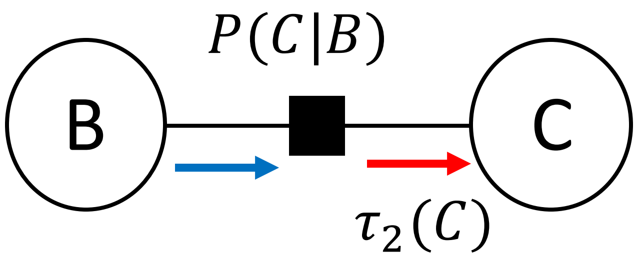
University of Illinois at Urbana Champaign

Announcements

- MP3 Checkpoint 1 (Task 0 and Task 1) due on Wednesday, April 10
- MP3 Task 2 will be released on Wednesday, April 10
- In Class Activity 5 on Wednesday, April 10

Belief Propagation – Message from factor to variable

Recall from previous example:

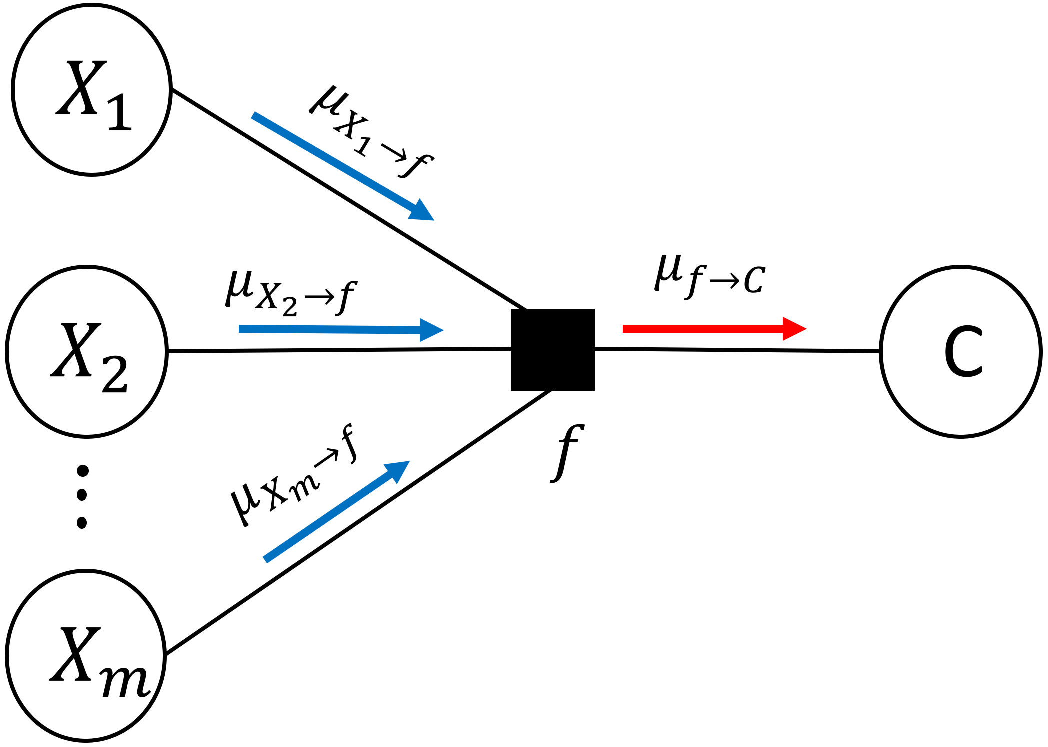


$$\tau_2(C) = \sum_B P(C|B)\tau_1(B)$$

To get the general expression, denote by:

$$f(B,C) = P(C|B)$$
$$\mu_{f\to C}(C) = \tau_2(C)$$
$$\mu_{B\to f}(B) = \tau_1(B)$$

In general:

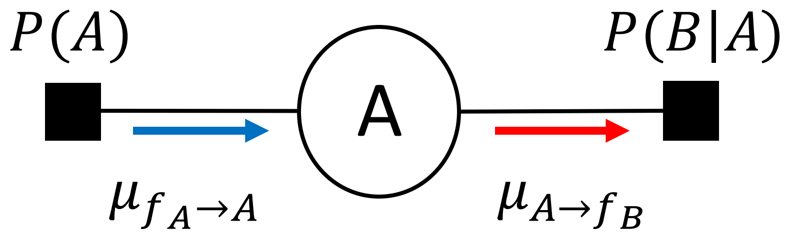


$$\mu_{f\to C}(C) = \sum_{X_1,X_2,\ldots,X_m} f(C, X_1, \ldots, X_m) \prod_{i=1}^{m} \mu_{X_i\to f}(X_i)$$

Message from factor to variable: Product of all incoming messages and factor, sum out previous variables

Belief Propagation – Message variable to factor

Recall from previous example:

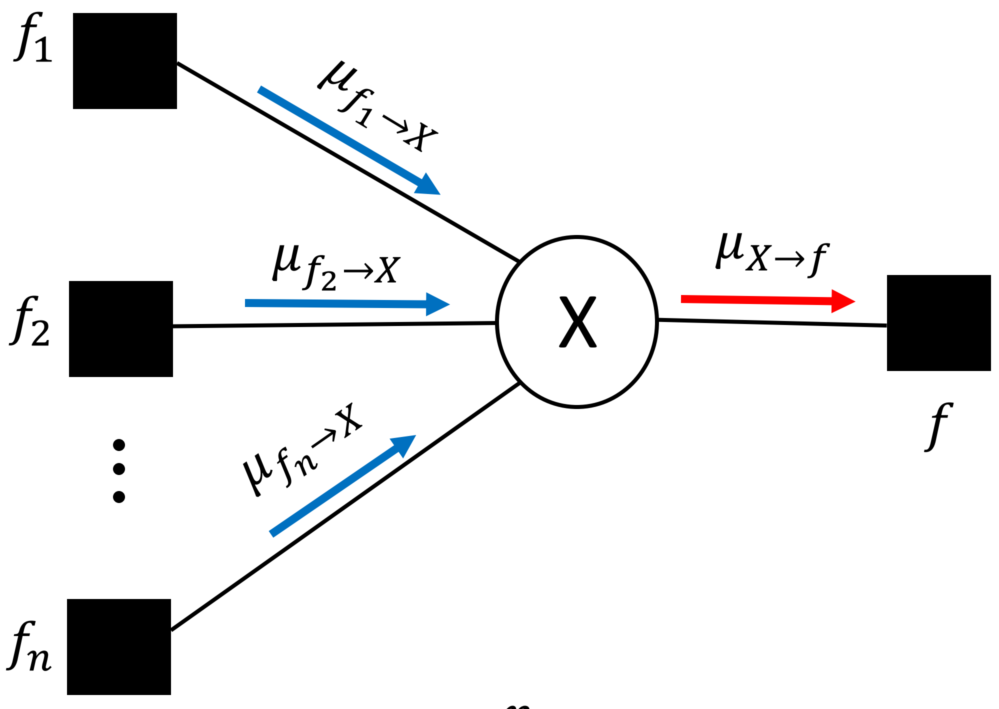


$$\mu_{A \to f_B}(A) = \mu_{f_A \to A}(A) = P(A)$$

Where,

$$f_A(A) = P(A)$$
$$f_B(A, B) = P(B|A)$$

In general:



$$\mu_{X \to f}(X) = \prod_{i=1}^{n} \mu_{f_i \to X}(X)$$

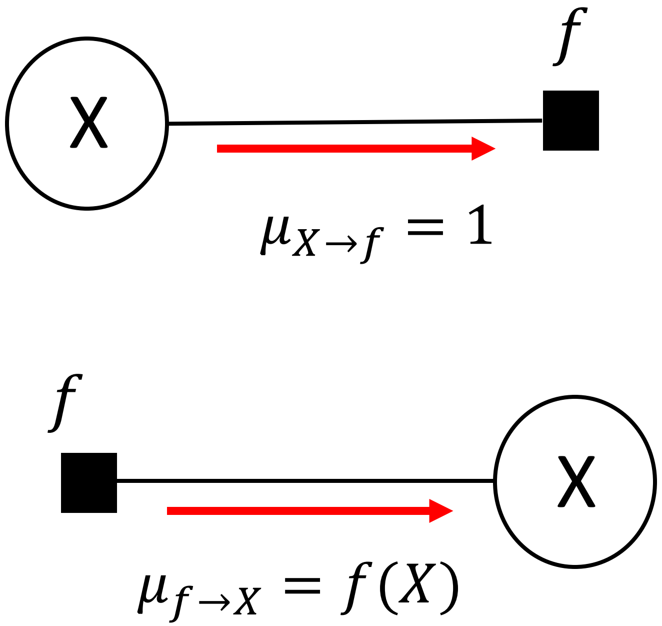
Message from factor to variable: Product of all incoming messages

Belief Propagation: General Algorithm

- How to start the algorithm
  - Choose a node in the factor graph as root node
  - Compute all the leaf-to-root messages
  - Compute all the root-to-leaf messages

- Initial Conditions
  - Starting from a factor leaf/root node, the initial factor-to-variable message is the factor itself
  - Starting from a variable leaf/root node, the initial variable-to-factor message is a vector of ones

- Computing marginals
  - Marginal is given by the product of all incoming messages



Initialization

Factor Graph application in Natural Language Processing

Background: A fake news is a lie, i.e., a purported fact, that is intended to mislead a target audience.

Motivation: Fake news are prevalent on the Internet.
Post 2016 election, Facebook users engaged with Fake news site roughly 70 million times per month [REF].
Fake news are being used to sway votes in India election, 2019. [REF] [REF]

Problem Statement: How to distinguish facts from fake news?

Input: Natural Language Documents

"Barack Obama was born in Hawaii"
"Barack Obama is from Kenya"

Output: Fact database

Facts:
Born_in(Obama, Hawaii) = TRUE
Born_in(Obama, Kenya) = FALSE

Example Factor Graph for fact-checking Obama's place of birth

INPUT: NATURAL LANGUAGE DOCUMENTS

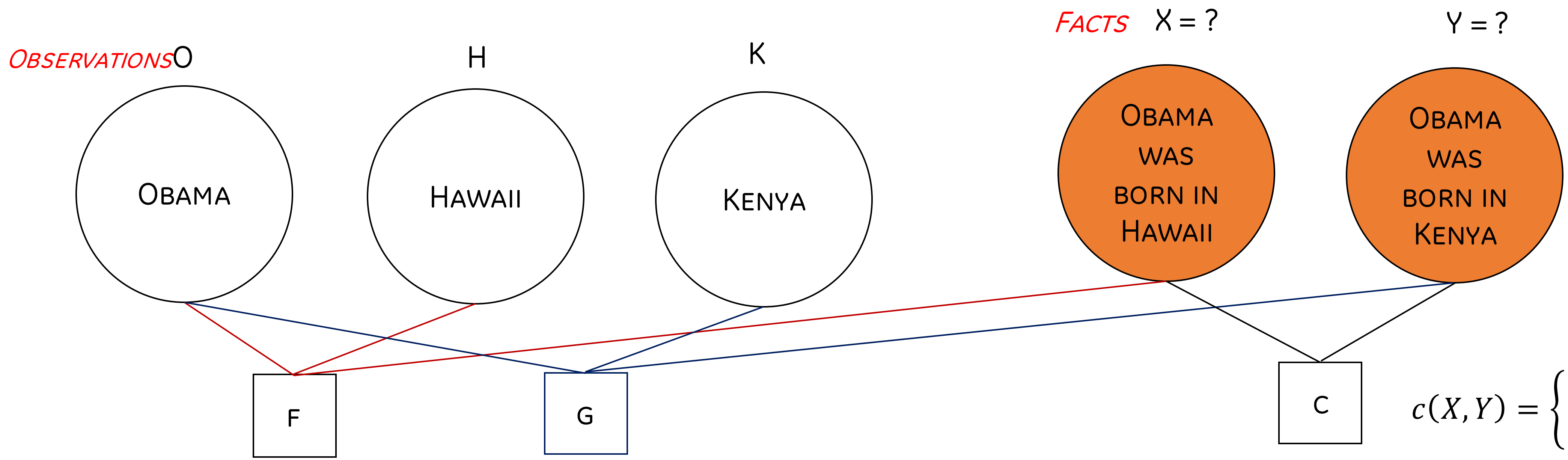
"Barack Obama was born in Hawaii"
"Barack Obama is from Kenya"

OUTPUT: FACT DATABASE

Facts:
Born_in(Obama, Hawaii) = TRUE
Born_in(Obama, Kenya) = FALSE

$n_{OH} = 10000, n_{OK} = 5000$

X, Y IN {0,1}



$$c(X,Y) = \begin{cases} -\inf \ if \ X = Y = 1 \\ 1 \ \ if \ X \mathrel{!}= Y \end{cases}$$

$$P(O,H,K,X,Y) = \frac{1}{Z} f(O,H,X) g(O,K,Y) c(X,Y)$$

$$f(O,H,X) = \begin{cases} \log(1 + n_{OH}), X = 1 \\ 1, \quad X = 0 \end{cases}$$

$$g(O,K,Y) = \begin{cases} \log(1 + n_{OK}), Y = 1 \\ 1, \quad Y = 0 \end{cases}$$

$$f(O,H,X = 1) = \log(10000 + 1) = 4$$
$$f(O,H,X = 0) = 1$$

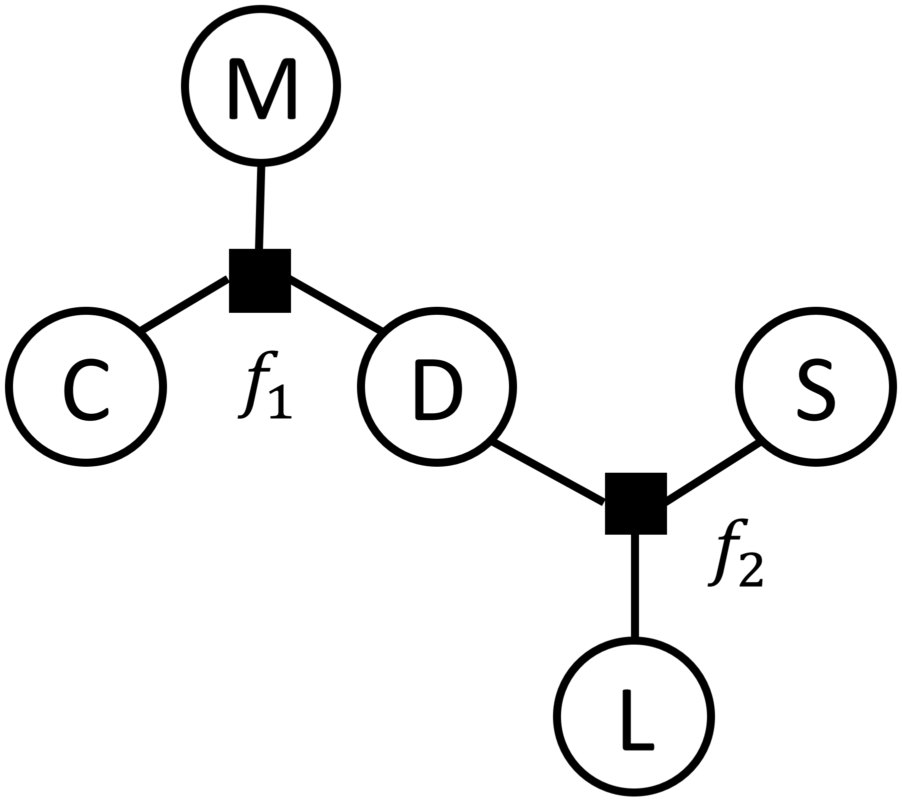
$$g(O,K,Y = 1) = \log(5000 + 1) = 3.7$$
$$g(O,K,Y = 0) = 1$$

Example:

$$P(O,H,K,X = 1,Y = 0) = \frac{1}{Z} f(O,H,X) g(O,K,Y) c(X,Y)$$
$$P(O,H,K,X = 0,Y = 0) = \frac{1}{Z} f(O,H,X) g(O,K,Y) c(X,Y)$$
$$P(O,H,K,X = 0,Y = 1) = \frac{1}{Z} f(O,H,X) g(O,K,Y) c(X,Y)$$

Lung Cancer Example

Researcher has shown that a mutation in a particular gene and the presence of vascular disease has a correlation with defect in a particular protein's structure. Another study showed that defect in the particular protein's structure and smoking are correlated with the presence of lung cancer. Using this information, an health provider came up with the factor graph shown alongside.



M: Mutation $\{m^0, m^1\}$

C: Vascular disease $\{c^0, c^1\}$

S: Smoking $\{s^0, s^1\}$

D: Defective protein $\{d^0, d^1\}$

L: Lung Cancer $\{l^0, l^1\}$

The health provider is interested in evaluating the probability that a patient will contract lung cancer given that he has the mutation i.e., $P(L|M = m1)$.

Lung Cancer Example

Factor functions are given as follows

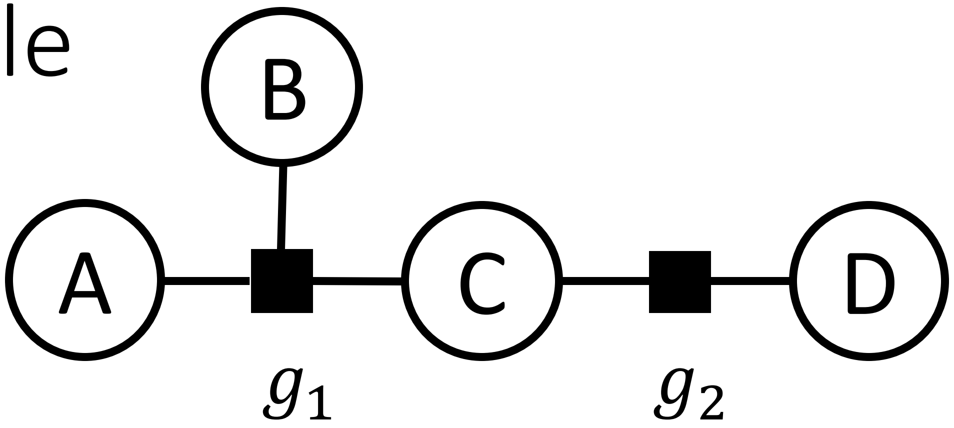
C	M	D	f_1
c^0	m^0	d^0	10
c^0	m^0	d^1	1
c^0	m^1	d^0	5
c^0	m^1	d^1	10
c^1	m^0	d^0	8
c^1	m^0	d^1	2
c^1	m^1	d^0	10
c^1	m^1	d^1	15

D	S	L	f_2
d^0	s^0	l^0	20
d^0	s^0	l^1	1
d^0	s^1	l^0	5
d^0	s^1	l^1	10
d^1	s^0	l^0	10
d^1	s^0	l^1	5
d^1	s^1	l^0	1
d^1	s^1	l^1	15

Inference is linked to conditioning

- Bayesian Networks and HMM: Some variables were observed, other must be estimated
- Inference question: P(variables of interest | observations)
  - We have seen algorithms to solve these for BN and HMM
- Belief propagation can be used to compute marginals of single variables in Factor Graphs
- FGs are a generalization of BN and HMMs
  - This implies, that belief propagation algorithm that we applied to FGs can be applied to other PGMs
- Key question with FGs: What if you observe certain variables e.g., Mutation (M). Given the observation, what is the distribution of a variable of interest e.g., Lung Cancer (L)
- Solution: Take the marginal values and use them to compute the conditional values. E.g., $P(L|M) = P(L,M)/P(M)$

Conditioning on a variable Example



Consider the factor graph with binary variable A, B, C, D. Calculate $P(A|B = b^0)$.

Bayes rule: $P(A|b^0) = \frac{P(A,b^0)}{P(b^0)}$

$$P(A, b^0) = \sum_{C,D} P(A, b^0, C, D)$$

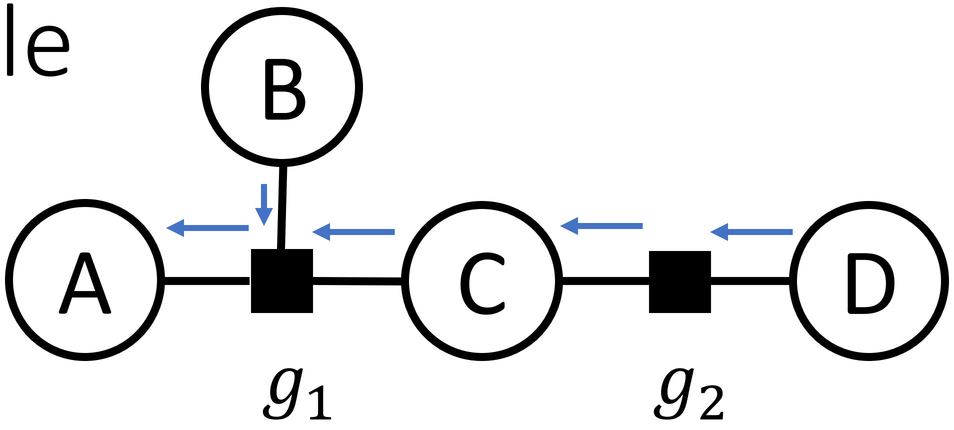
$$= \frac{1}{Z} \sum_{C,D} g_1(A, b^0, C)\, g_2(C, D)$$

$$P(b^0) = \sum_{A,C,D} P(A, b^0, C, D)$$

B	**A**	**C**	$\boldsymbol{g_1}$
b^0	a^0	c^0	7
b^0	a^0	c^1	2
b^0	a^1	c^0	4
b^0	a^1	c^1	15
b^1	a^0	c^0	0
b^1	a^0	c^1	11
b^1	a^1	c^0	10
b^1	a^1	c^1	100

D	**C**	$\boldsymbol{g_2}$
d^0	c^0	0.4
d^0	c^1	6
d^1	c^0	10
d^1	c^1	3

Conditioning on a variable Example



Consider the factor graph with binary variable A, B, C, D. Calculate $P(A|B = b^0)$.

Bayes rule: $P(A|b^0) = \frac{P(A,b^0)}{P(b^0)}$

$$P(A, b^0) = \sum_{C,D} P(A, b^0, C, D)$$

$$= \frac{1}{Z} \sum_{C,D} g_1(A, b^0, C)\ g_2(C, D)$$

$$P(b^0) = \sum_{A,C,D} P(A, b^0, C, D)$$

Belief Propagation:

$$P(A, b^0) \propto \mu_{g_1 \to A}(A, b^0)$$

$$\mu_{g_1 \to A}(A, B) = \sum_{C} g_1(A, B, C)\ \mu_{C \to g_1}(C) \mu_{B \to g_1}(B)$$

Set $B = b^0$ in the above to get $\mu_{g_1 \to A}(A, b^0)$

Compute above marginals using Belief Propagation[1].

[1]See https://www.psi.toronto.edu/~jimmy/ece521/Tut10.pdf for details.

Alternate method

$$P(A, b^0) = \sum_{C,D} P(A, b^0, C, D)$$

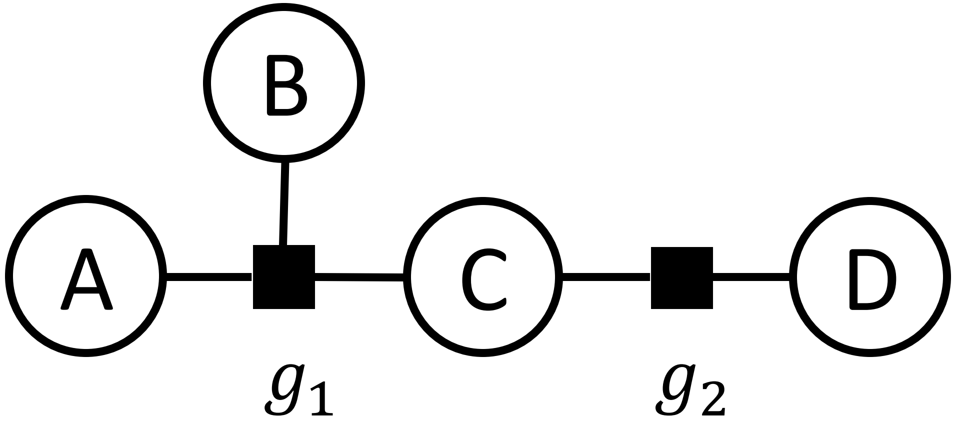
$$= \frac{1}{Z} \sum_{C,D} \boxed{g_1(A, b^0, C)}\, g_2(C, D)$$

$$= \frac{1}{Z} \sum_{C,D} g_1'(A, C)\, g_2(C, D)$$

$$P(b^0) = \sum_{A,C,D} P(A, b^0, C, D)$$

$$= \frac{1}{Z} \sum_{A,C,D} g_1'(A, C)\, g_2(C, D)$$

Fixing one variable is equivalent to modifying the factor function



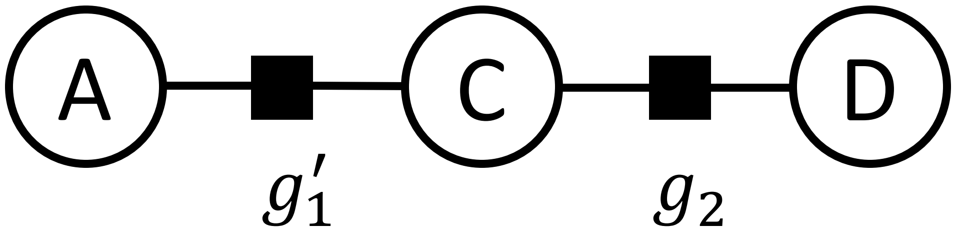
B	A	C	g_1
b^0	a^0	c^0	7
b^0	a^0	c^1	2
b^0	a^1	c^0	4
b^0	a^1	c^1	15
b^1	a^0	c^0	0
b^1	a^0	c^1	11
b^1	a^1	c^0	10
b^1	a^1	c^1	100

A	C	g_1'
a^0	c^0	7
a^0	c^1	2
a^1	c^0	4
a^1	c^1	15

D	C	g_2
d^0	c^0	0.4
d^0	c^1	6
d^1	c^0	10
d^1	c^1	3

Alternate method

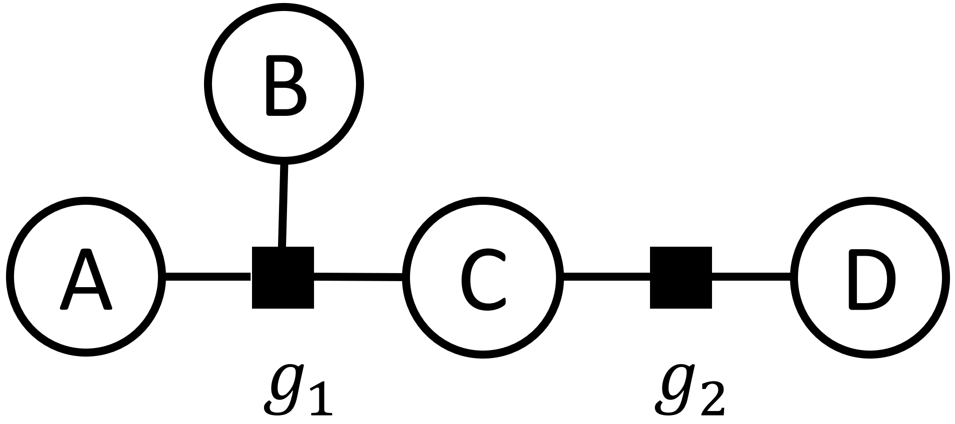
- Fixing one variable is equivalent to modifying the factor function
- Equivalent to defining a factor graph



A	**C**	$\boldsymbol{g_1'}$
a^0	c^0	7
a^0	c^1	2
a^1	c^0	4
a^1	c^1	15

D	**C**	$\boldsymbol{g_2}$
d^0	c^0	0.4
d^0	c^1	6
d^1	c^0	10
d^1	c^1	3

Factor graph after observing B=b^0



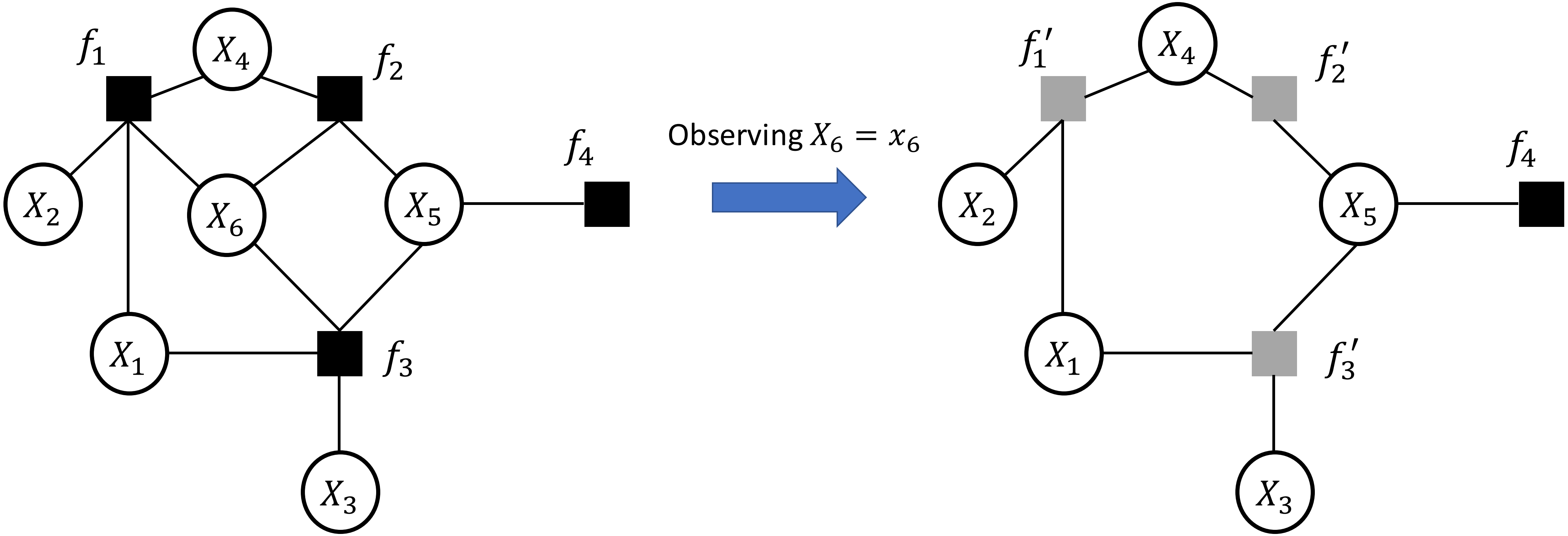
B	**A**	**C**	$\boldsymbol{g_1}$
b^0	a^0	c^0	7
b^0	a^0	c^1	2
b^0	a^1	c^0	4
b^0	a^1	c^1	15
b^1	a^0	c^0	0
b^1	a^0	c^1	11
b^1	a^1	c^0	10
b^1	a^1	c^1	100

D	**C**	$\boldsymbol{g_2}$
d^0	c^0	0.4
d^0	c^1	6
d^1	c^0	10
d^1	c^1	3

Original factor graph

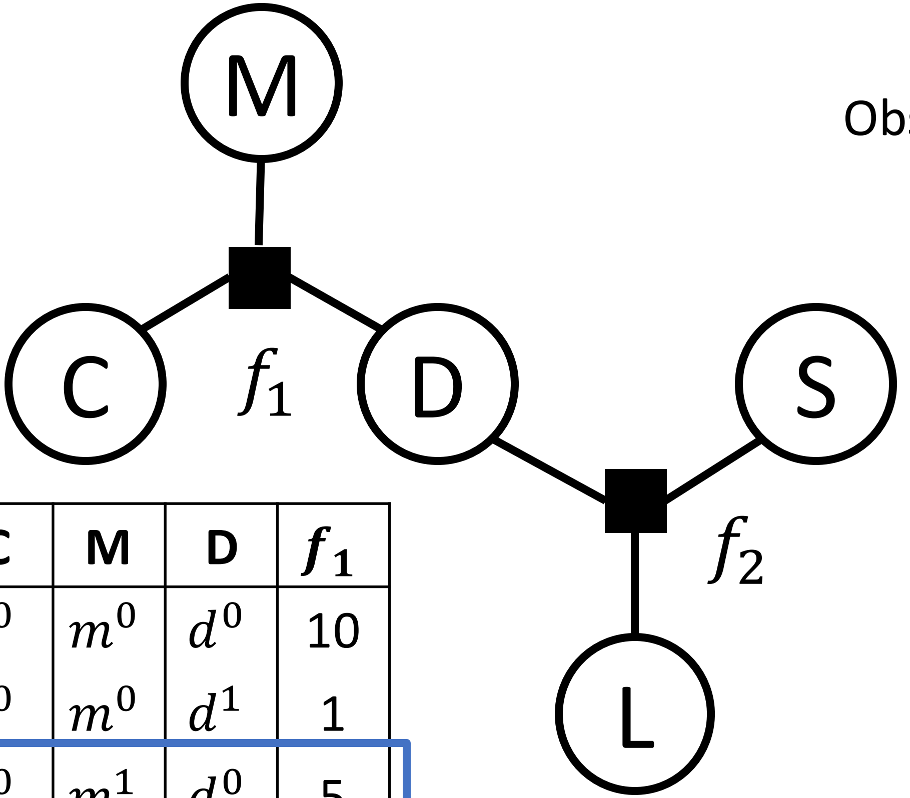
Eliminating Observed Variables

- If a variable X is **observed**, i.e., its value is given, then it is a ***constant*** in all functions that include X
- We can **eliminate** X from the graph by removing the corresponding node and modifying all neighboring factor functions to treat it as a constant



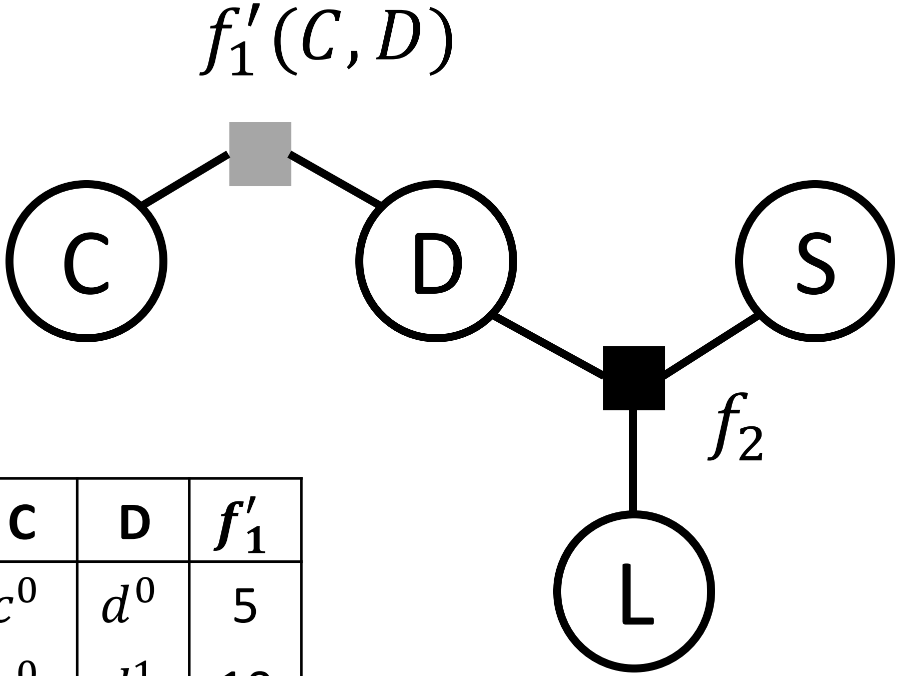
After conditioning on $X_6 = x_6$, we get $P(X_1, \ldots, X_5, X_6 = x_6) = f_1'(X_1, X_2, X_4) f_2'(X_4, X_5) f_3'(X_1, X_3, X_5) f_4\ (X_5)$. The functions $f_1'(.), f_2'(.), f_3'(.)$ (in grey) are distinct from $f_1(.), f_2(.), f_3(.)$

Lung Cancer Example: Modify graph



C	M	D	f_1
c^0	m^0	d^0	10
c^0	m^0	d^1	1
c^0	m^1	d^0	5
c^0	m^1	d^1	10
c^1	m^0	d^0	8
c^1	m^0	d^1	2
c^1	m^1	d^0	10
c^1	m^1	d^1	15

Observing $M = m^1$

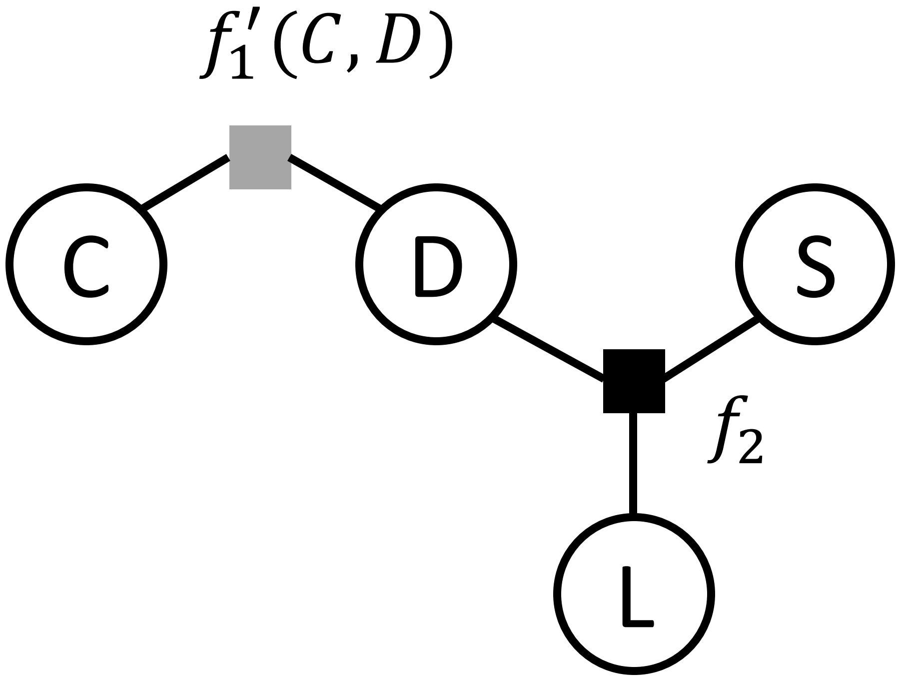


C	D	f_1'
c^0	d^0	5
c^0	d^1	10
c^1	d^0	10
c^1	d^1	15

Lung Cancer Example: Modified graph

C	M	D	f'_1
c^0	m^1	d^0	5
c^0	m^1	d^1	10
c^1	m^1	d^0	10
c^1	m^1	d^1	15

D	S	L	f_2
d^0	s^0	l^0	20
d^0	s^0	l^1	1
d^0	s^1	l^0	5
d^0	s^1	l^1	10
d^1	s^0	l^0	10
d^1	s^0	l^1	5
d^1	s^1	l^0	1
d^1	s^1	l^1	15



- Apply belief propagation on the modified graph with L as the root node
- Marginal of L in the above FG is equal to $P(L|M = m^1)$ in the original FG

Lung Cancer Example: Belief Propagation

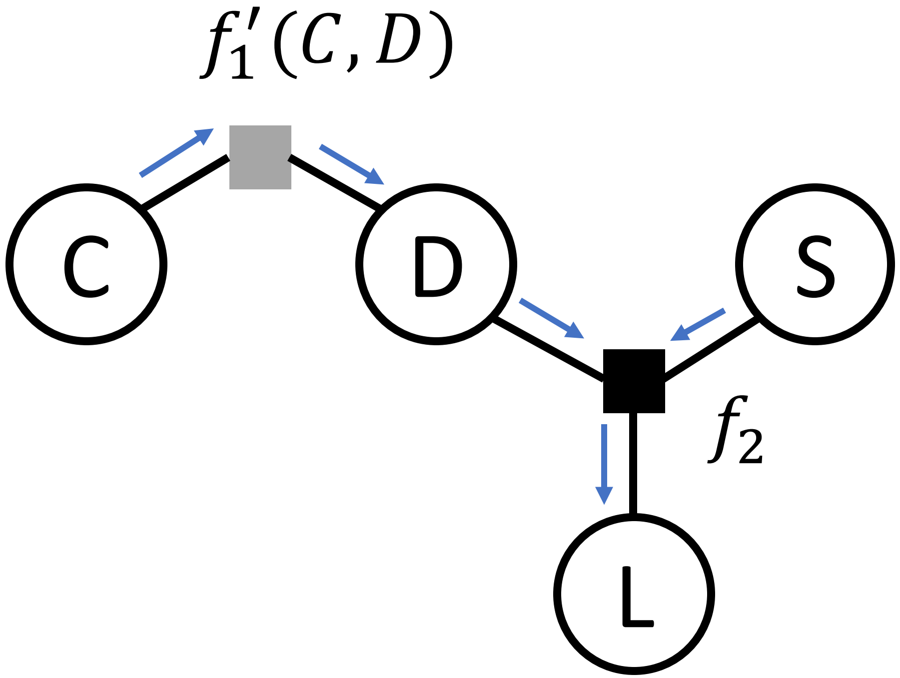
$$\mu_{C\to f_1'}(C) = 1$$

$$\mu_{f_1^1\to D}(D) = \sum_C f_1'(C,D)\mu_{C\to f_1'}(C)$$

$$\mu_{f_1^1\to D}(d^0) = \sum_C f_1'(C,d^0)\mu_{C\to f_1'}(C)$$
$$= f_1'(c^0,d^0)\mu_{C\to f_1'}(c^0) + f_1'(c^1,d^0)\mu_{C\to f_1'}(c^1)$$
$$= 5\times 1 + 10\times 1 = 15$$

$$\mu_{f_1^1\to D}(d^1) = \sum_C f_1'(C,d^1)\mu_{C\to f_1'}(C)$$
$$= f_1'(c^0,d^1)\mu_{C\to f_1'}(c^0) + f_1'(c^1,d^1)\mu_{C\to f_1'}(c^1)$$
$$= 10\times 1 + 15\times 1 = 25$$

$$\mu_{D\to f_2}(D) = \mu_{f_1'\to D}(D)$$



Lung Cancer Example: Belief Propagation

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

$$\mu_{f_2\to L}(L) = \sum_{S,D} f_2(D,S,L)\mu_{D\to f_2}(D)\mu_{S\to f_2}(S)$$

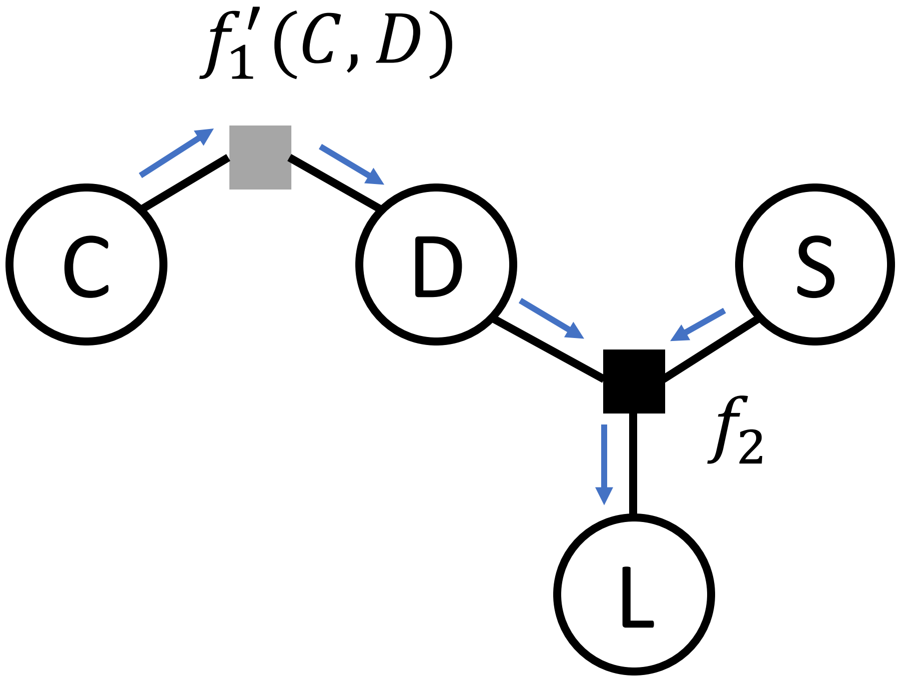
$$\mu_{f_2\to L}(l^0) = \sum_{S,D} f_2(D,S,l^0)\mu_{D\to f_2}(D)\mu_{S\to f_2}(S)$$

$$= f_2(s^0,d^0,l^0)\mu_{D\to f_2}(d^0) + f_2(s^0,d^1,l^0)\mu_{D\to f_2}(d^1)$$

$$+f_2(s^1,d^0,l^0)\mu_{D\to f_2}(d^0) + f_2(s^1,d^1,l^0)\mu_{D\to f_2}(d^1)$$

$$= 20\times 15 + 10\times 25 + 5\times 15 + 1\times 25 = 650$$

$$\mu_{f_2\to L}(l^1) = 665$$



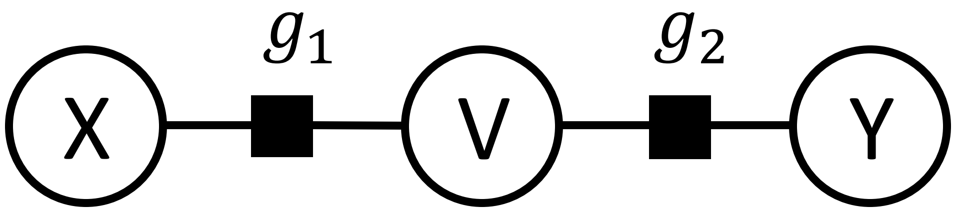
$$P(L|m^1) \propto \mu_{f_2\to L}(L)$$

Recall that the FG was modified after observing $M = m^1$, so all marginal distributions on the modified graph will be conditioned on $M = m^1$

Appropriate normalization gives: $P(L|m^1) = \frac{1}{665+650}\begin{bmatrix}650\\665\end{bmatrix}$

Conditional independence in FG

- Looked at factorization of the joint distribution and inference in BN, HMM and FG
- Conditional independence between variables can be derived in BN and HMM. What about FG?
- **Example**: Is $X \perp\!\!\!\perp Y | V$?
- **Solution**: We want to show conditional independence:



$$X \perp\!\!\!\perp Y | V \Leftrightarrow P(X|Y,V) = P(X|V)$$

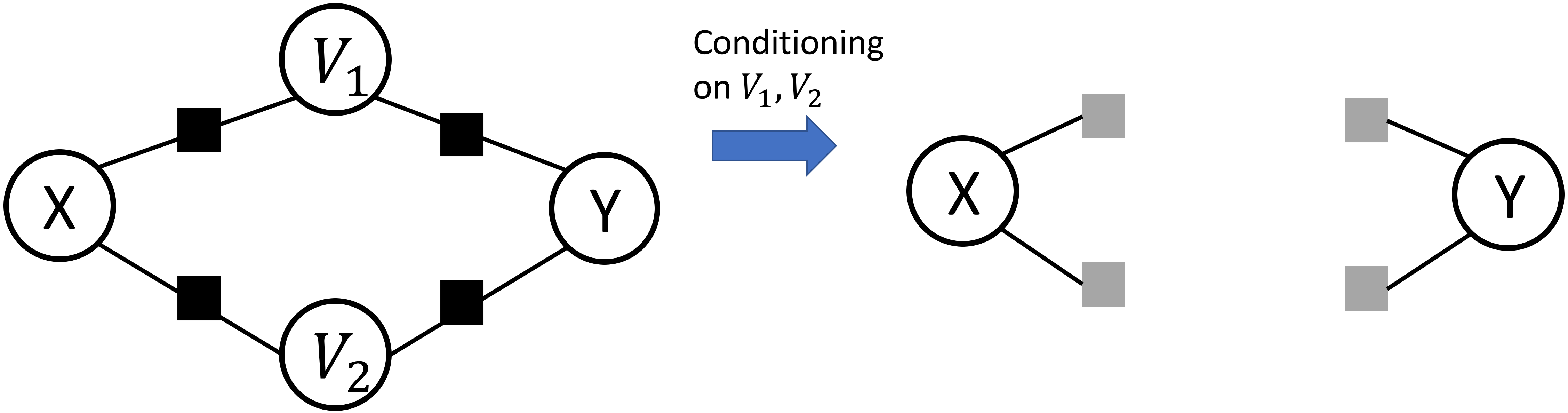
$$P(X,Y,V) = \frac{1}{Z} g_1(X,V) g_2(Y,V) \qquad P(Y,V) = \sum_X \frac{1}{Z} g_1(X,V) g_2(Y,V)$$

$$P(X|Y,V) = \frac{P(X,Y,V)}{P(Y,V)} = \frac{\frac{1}{Z} g_1(X,V) g_2(Y,V)}{\sum_X \frac{1}{Z} g_1(X,V) g_2(Y,V)} = \frac{g_1(X,V)}{\sum_X g_1(X,V)}$$

Since the RHS does not depend on Y, it follows that X is independent of Y given V

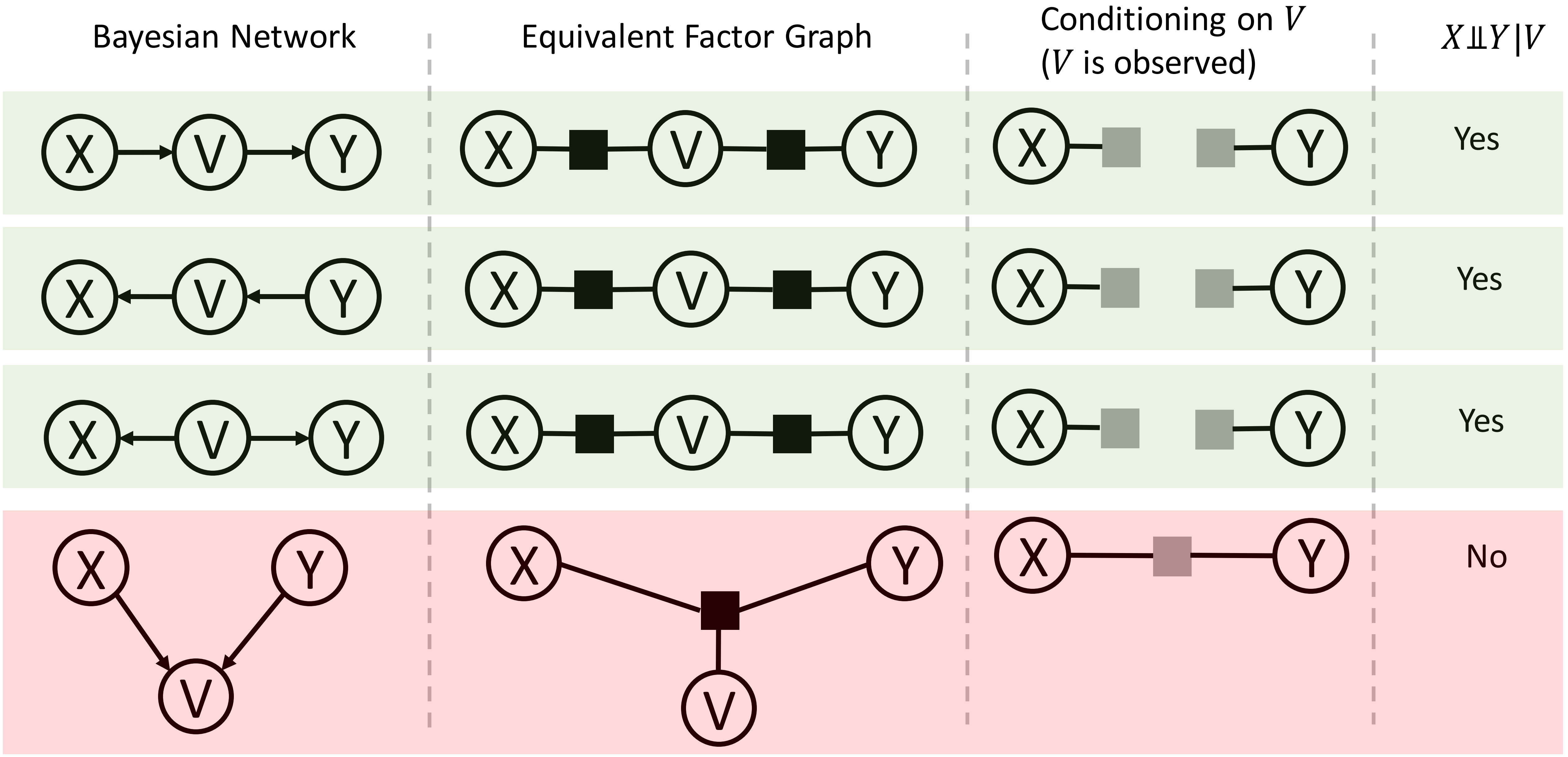
Conditional independence in FG

- A path is a sequence of neighboring nodes
- $X \perp\!\!\!\perp Y | \mathcal{V}$ if **every path** between X and Y contains some node $V \in \mathcal{V}$
- Corollary: Given the neighbors of X, the variable X is conditionally independent of all other variables



$X \perp\!\!\!\perp Y | \mathcal{V}$ where $\mathcal{V} = \{V_1, V_2\}$

Conditional Independence in BN using equivalent FG



References

- https://web.stanford.edu/~montanar/RESEARCH/BOOK/partC.pdf
- https://www.cs.toronto.edu/~urtasun/courses/GraphicalModels/lecture10.pdf
- http://mlg.eng.cam.ac.uk/zoubin/course04/factorprop.pdf
- http://mlg.eng.cam.ac.uk/teaching/4f13/1011/lect04.pdf
- https://www.psi.toronto.edu/~jimmy/ece521/Tut10.pdf