Probabilistic Graphical Models

Lectures 12

Introduction to Inference
Variable Elimination

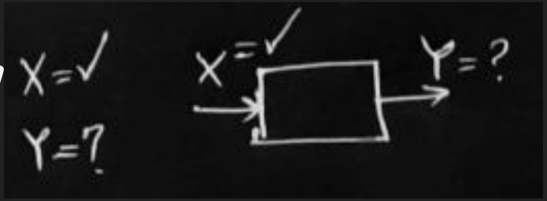
Probabilistic Model Problems

- Representation: $P_\theta(X)$, $P_\theta(X,Y)$, $P_\theta(Y|X)$

Probabilistic Model Problems

- Representation: $P_\theta(X)$, $P_\theta(X,Y)$, $P_\theta(Y|X)$

- Inference (θ know

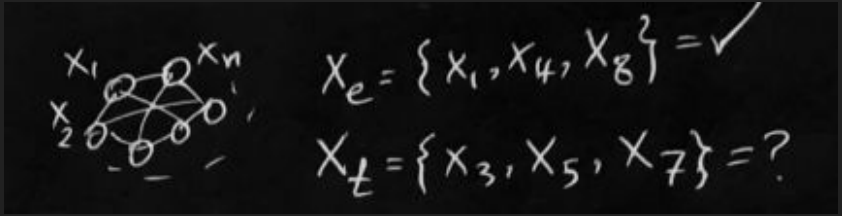


Probabilistic Model Problems

K. N. Toosi
University of Technology

- Representation: $P_\theta(X)$, $P_\theta(X,Y)$, $P_\theta(Y|X)$

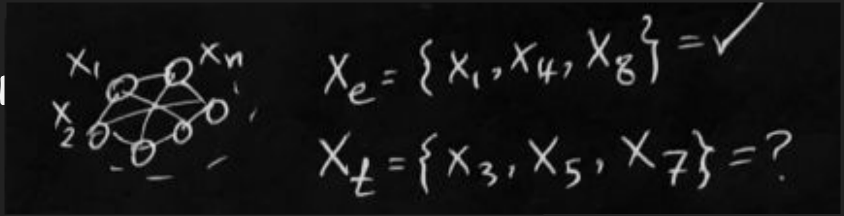
- Inference (θ known)



Probabilistic Model Problems

- Representation: $P_\theta(X)$, $P_\theta(X,Y)$, $P_\theta(Y|X)$

- Inference (θ know

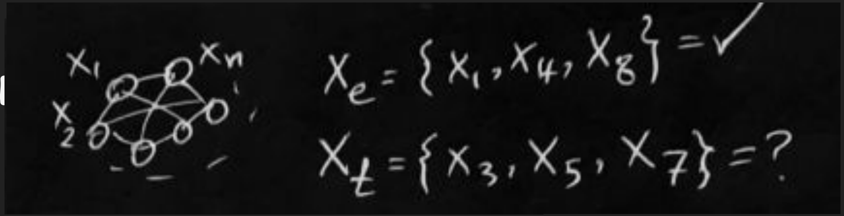


- Sampling (θ known) $P_\theta(X)$ ---> generate data X^1, X^2, …, X^m

Probabilistic Model Problems

K. N. Toosi
University of Technology

- Representation: $P_\theta(X)$, $P_\theta(X,Y)$, $P_\theta(Y|X)$
- Inference (θ know



- Sampling (θ known) $P_\theta(X)$ ---> generate data X^1, X^2, ..., X^m
- Learning (θ unknown) data X^1, X^2, ..., X^m find θ, find structure

Inference

- Given $P(X_1, X_2, \dots, X_N)$ find **target variables** given **observed variables.**

Inference

K. N. Toosi
University of Technology

Inference $P(X) = P(X_1, X_2, \ldots, X_n) = P(X_e, X_t, X_o)$

$X_e \subseteq X$ (evidence, observation, input) known

$X_t \subseteq X$ (target, output) unknown

$X_o = X \setminus (X_e \cup X_t)$ (other variables) we don't care about them

in many cases $X_o = \{\} \Rightarrow X = X_e \cup X_t$

given X_e find X_t

Example

K. N. Toosi
University of Technology

Example

15 (II)

$P(X_1, X_2, X_3, X_4, X_5) = \checkmark$

$X_t = \{X_1, X_3\}$

$X_e = \{X_2, X_5\}$

$X_o = X_4$

given X_2, X_5 find X_1, X_3

Example

Example:

X input

Y output

$p(X, Y)$

or

$p(Y|X)$

$X_e = X$

$X_t = Y$

given X find Y

Example

Given X_e find X_t

What does it mean to find the target variables X_t?

- In probabilistic sense (As a random variable) ⟹ Find $P(X_t)$

  P(Flu) = ?   P(Covid) = ?   P(TB) = ? ....

- In deterministic sense ⟹ Find the most likely X_t (MAP inference)

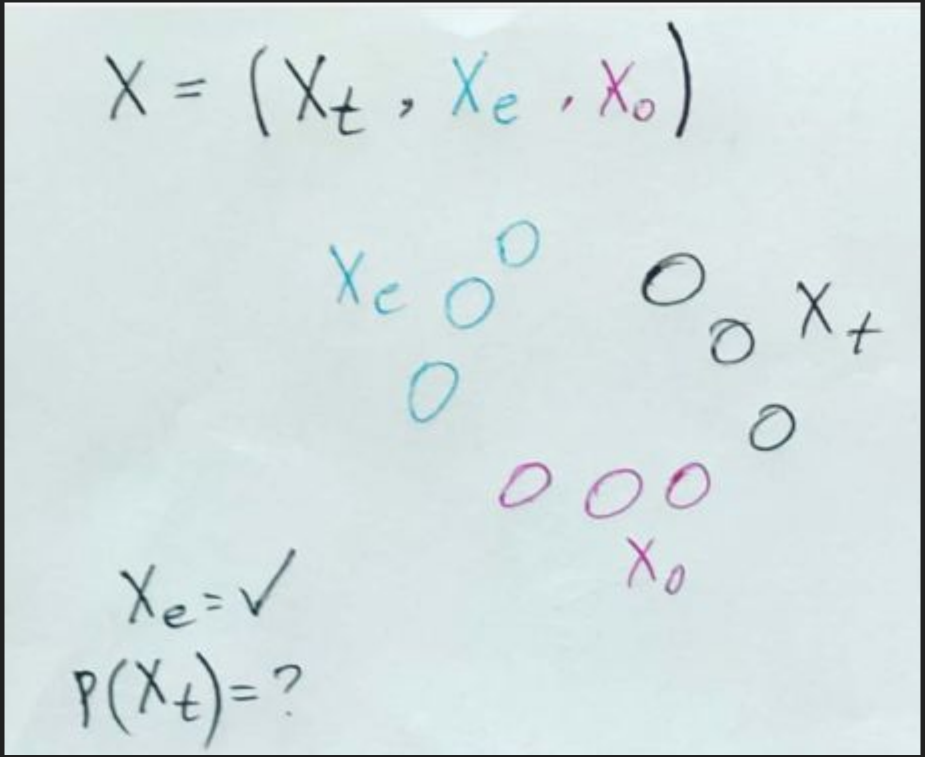
  or Find a reasonable X_t

Distance = 1.2 m , Temperature = 19.3°c

Examples: Semantic Segmentation / Speech Recognition / message Decoding

Messag → [Media] → noisy message

Inference



Simple Case

Simple Case: $X_e = \{\}$ $\quad X = X_t, X_o$

$P(X) = P(X_t, X_o)$

$$P(X_t) = \sum_{X_o} P(X) = \sum_{X_o} P(X_t, X_o)$$

↳ number of summation terms grows exponentially with the size of X_o.

(size of X_o: no. of variables in X_o)

Simple Case

K. N. Toosi
University of Technology

no. of variables in X_0

Eample: $X_0 = \{X_{0,1}, X_{0,2}, \ldots, X_{0,M}\}$ $\quad X_{0,i} \in \{0,1\}$

$$\sum_{X_0} P(X) = \underbrace{\sum_{X_{0,1}=0}^{1} \sum_{X_{0,2}=0}^{1} \sum_{X_{0,3}=0}^{1} \cdots \sum_{X_{0,M}=0}^{1} P(X_t, X_{0,1}, X_{0,2}, \ldots, X_{0,M})}_{2^M \text{ terms}}$$

Intractable for large M

General Case

K. N. Toosi University of Technology

$$P(X) = P(X_t, X_e, X_o) \qquad X_e = x_e \checkmark$$

migh be intractable

$$P(X_t \mid X_e = x_e) = \frac{P(X_t, X_e = x_e)}{P(X_e = x_e)} = \frac{\sum_{X_o} P(X_t, X_o, X_e = x_e)}{\sum_{X_o} \sum_{X_t} P(X_t, X_o, X_e = x_e)}$$

Exploit Factorization

15 (IV)

$$P(X) = P(X_t, X_o)$$

$$P(X_t) = \sum_{X_o} P(X) \Rightarrow \text{might be intractable}$$

$\Rightarrow$ Use the fact that for the PGMs (BN/MRF/CRF) the joint distribution is in the form of product of factors

$$P(X) = \prod_c \phi_c(X_c)$$

X_c comprises a small no. of variable

Exploit Factorization

$P(A, B, C, D, E) \Longrightarrow$ find $P(E)$

$X_e = \{\}$ $X_o = \{A, B, C, D\}$ $X_t = \{E\}$

$$P(E) = \sum_A \sum_B \sum_C \sum_D P(A, B, C, D, E)$$

$2^4 = 16$ summation terms if $A, B, C, D \in \{0, 1\}$

Exploit Factorization

$$\sum_D \sum_C \sum_B \sum_A \phi_0(A)\,\phi_1(A,B)\,\phi_2(B,C)\,\phi_3(C,D)\,\phi_4(D,E)$$

$$\sum_D \sum_C \sum_B \phi_2(B,C)\,\phi_3(C,D)\,\phi_4(D,E) \sum_A \phi_0(A)\,\phi_1(A,B)$$

Exploit Factorization

$$\sum_{D} \sum_{C} \sum_{B} \phi_2(B,C)\, \phi_3(C,D)\, \phi_4(D,E) \sum_{A} \phi_0(A)\, \phi_1(A,B)$$

$\phi_0(A)$

0.2	$\phi_0(A=0)$
0.8	$\phi_0(A=1)$

$\phi_1(A,B)$

	B=0	B=1
A=0	0.7	0.3
A=1	0.5	0.5

Exploit Factorization

$$\sum_D \sum_C \sum_B \phi_2(B,C)\,\phi_3(C,D)\,\phi_4(D,E) \sum_A \phi_0(A)\,\phi_1(A,B)$$

$\phi_0(A)$

0.2	$\phi_0(A=0)$
0.8	$\phi_0(A=1)$

$\phi_1(A,B)$

	B=0	B=1
A=0	0.7	0.3
A=1	0.5	0.5

$$\psi_1(A,B) = \phi_0(A)\,\phi_1(A,B)$$

	B=0	B=1
A=0	0.7x.2	0.3x0.2
A=1	0.5x0.8	0.5x0.8

$= \psi_1(A,B) =$

0.14	0.06
0.4	0.4

Exploit Factorization

$$\sum_D \sum_C \sum_B \phi_2(B,C)\,\phi_3(C,D)\,\phi_4(D,E) \sum_A \phi_0(A)\,\phi_1(A,B)$$

$$\tau_1(B) = \sum_A \psi_1(A,B) = \psi_1(0,B) + \psi_1(1,B)$$

$\psi_1(A,B)$

	B=0	B=1
A=0	0.14	0.06
A=1	0.4	0.4

⇓ +

$\tau_1(B)$ →

B=0	B=1
0.54	0.46

Exploit Factorization

K. N. Toosi
University of Technology

$$\sum_D \sum_C \sum_B \phi_2(B,C)\,\phi_3(C,D)\,\phi_4(D,E) \sum_A \psi_1(A,B)$$

$$\sum_D \sum_C \sum_B \phi_2(B,C)\,\phi_3(C,D)\,\phi_4(D,E)\,\tau_1(B)$$

$$\sum_D \sum_C \phi_3(C,D)\,\phi_4(D,E) \sum_B \underbrace{\phi_2(B,C)\,\tau_1(B)}_{\psi_2(B,C)}$$

$$\sum_D \sum_C \phi_3(C,D)\,\phi_4(D,E)\,\tau_2(C)$$

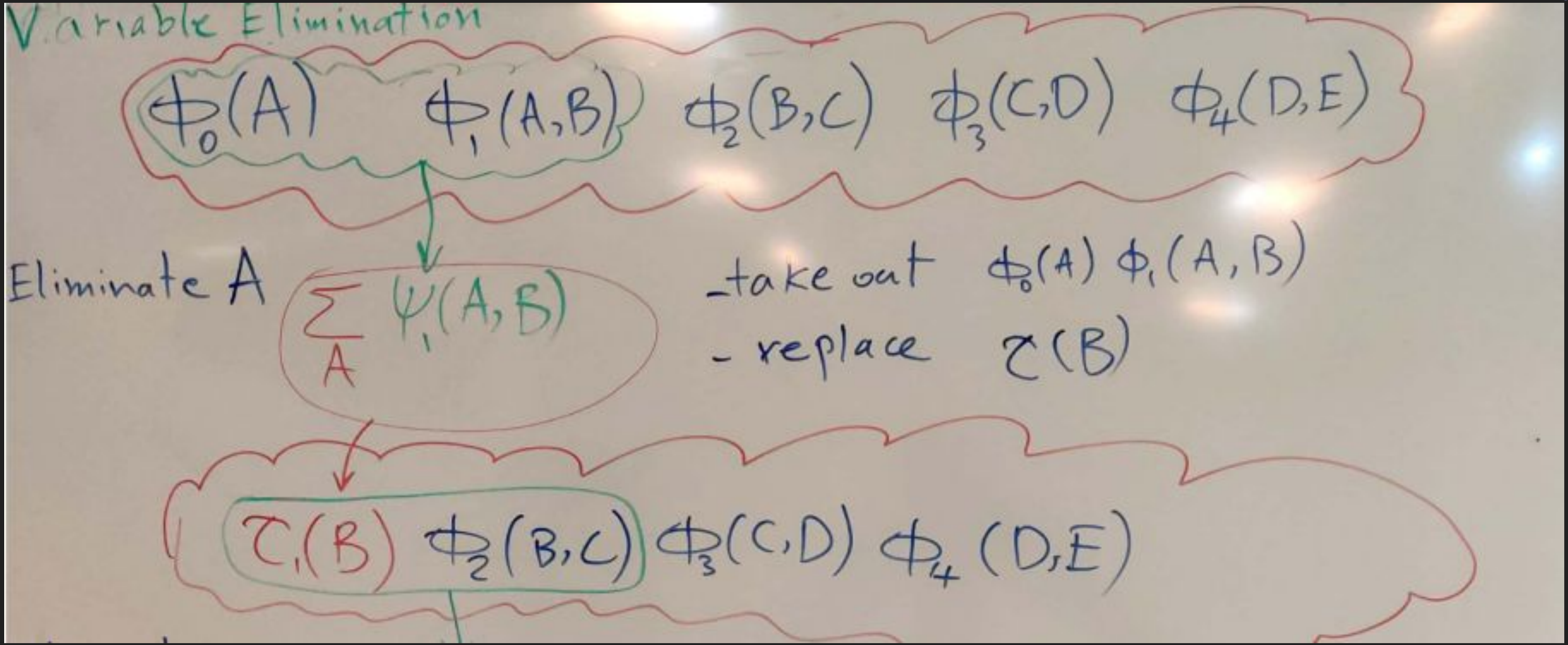
$$\sum_D \phi_4(D,E) \boxed{\sum_C \phi_3(C,D)\,\tau_2(C)}$$

Exploit Factorization

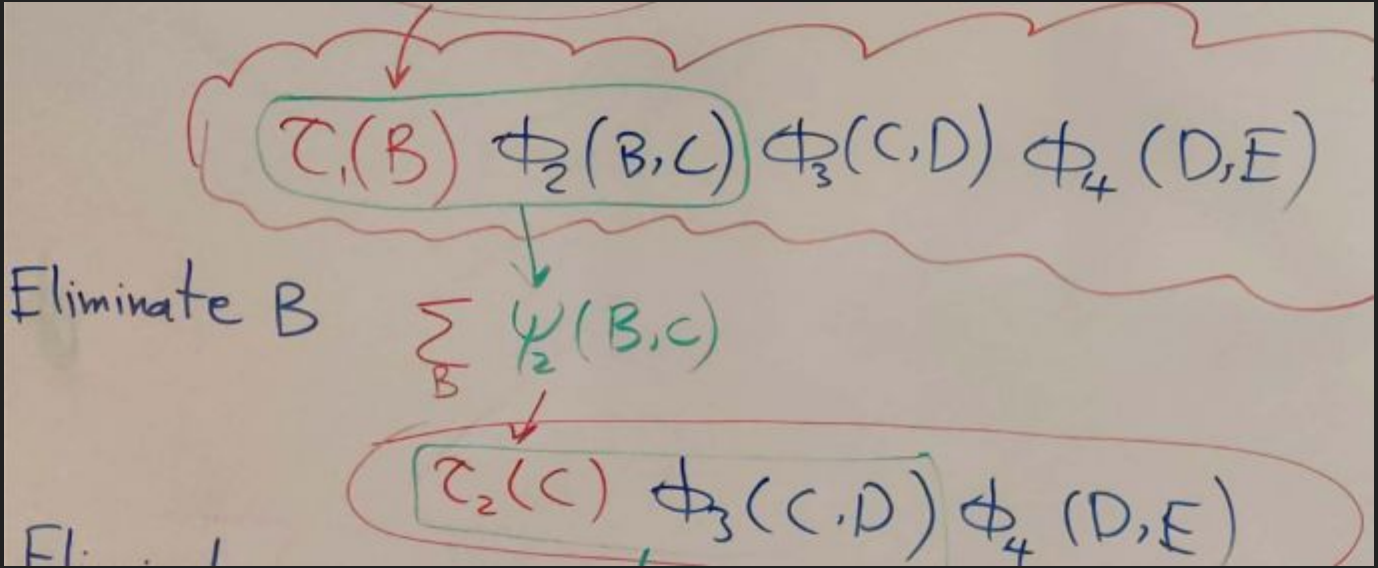
$$\sum_{D} \underbrace{\phi_4(D, E) \quad \tau_3(D)}_{\psi_3(D, E)} = \tau_4(D) = P(D)$$

Variable Elimination

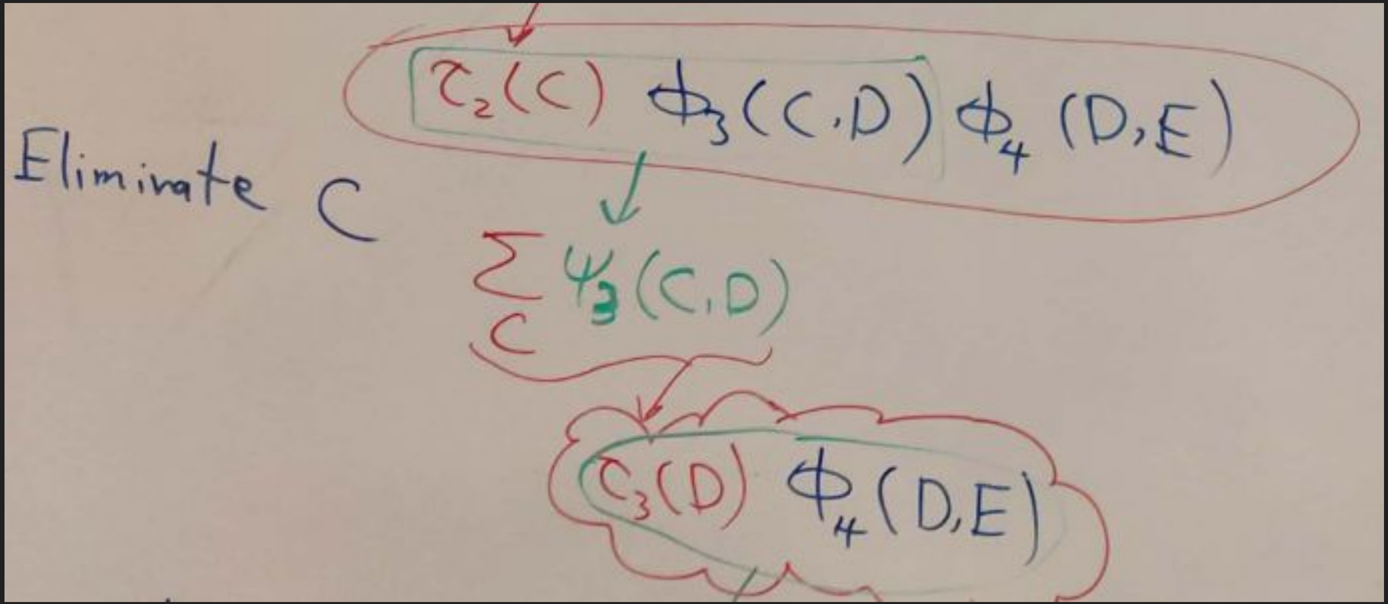
Variable Elimination



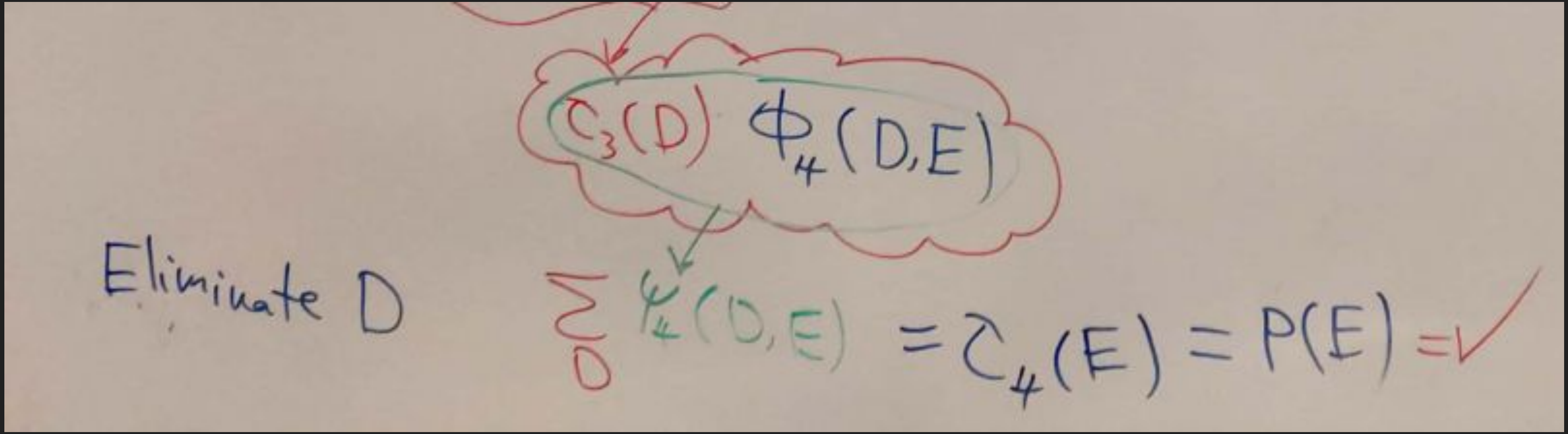
Variable Elimination



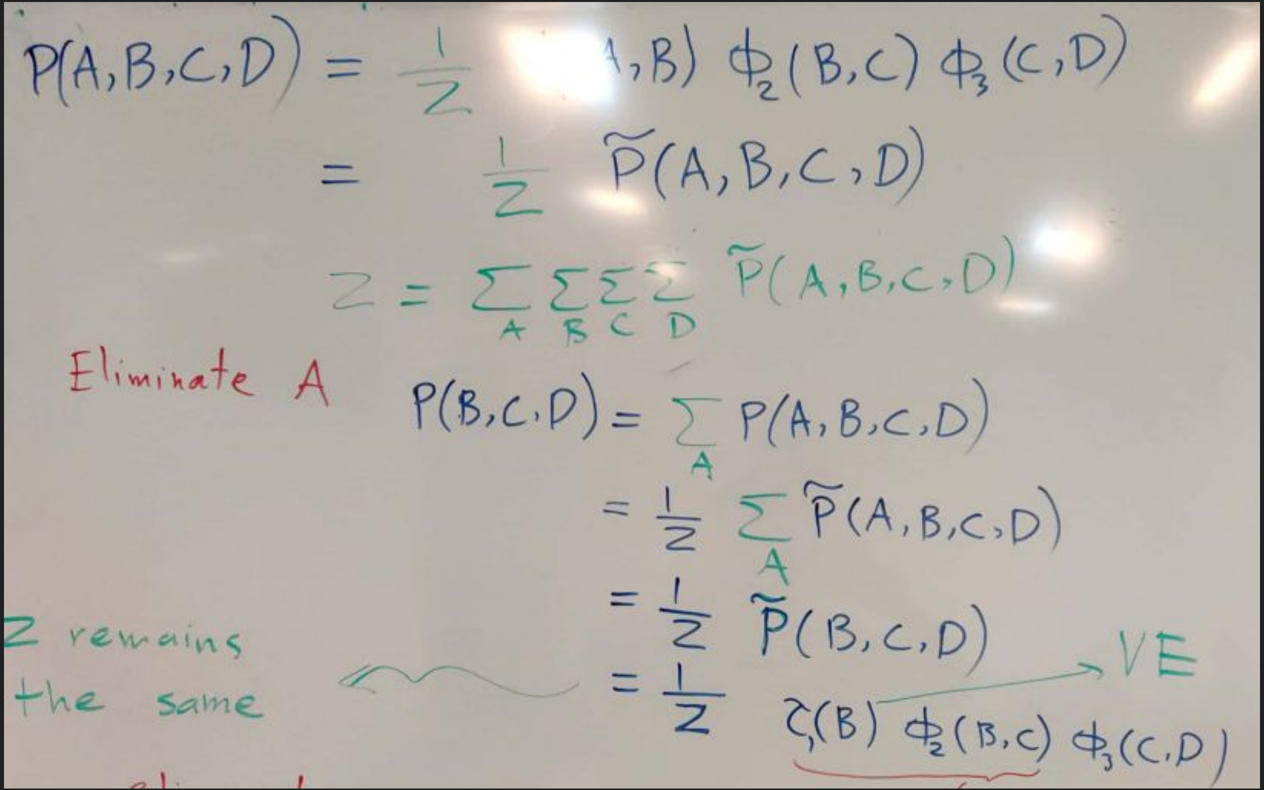
Variable Elimination



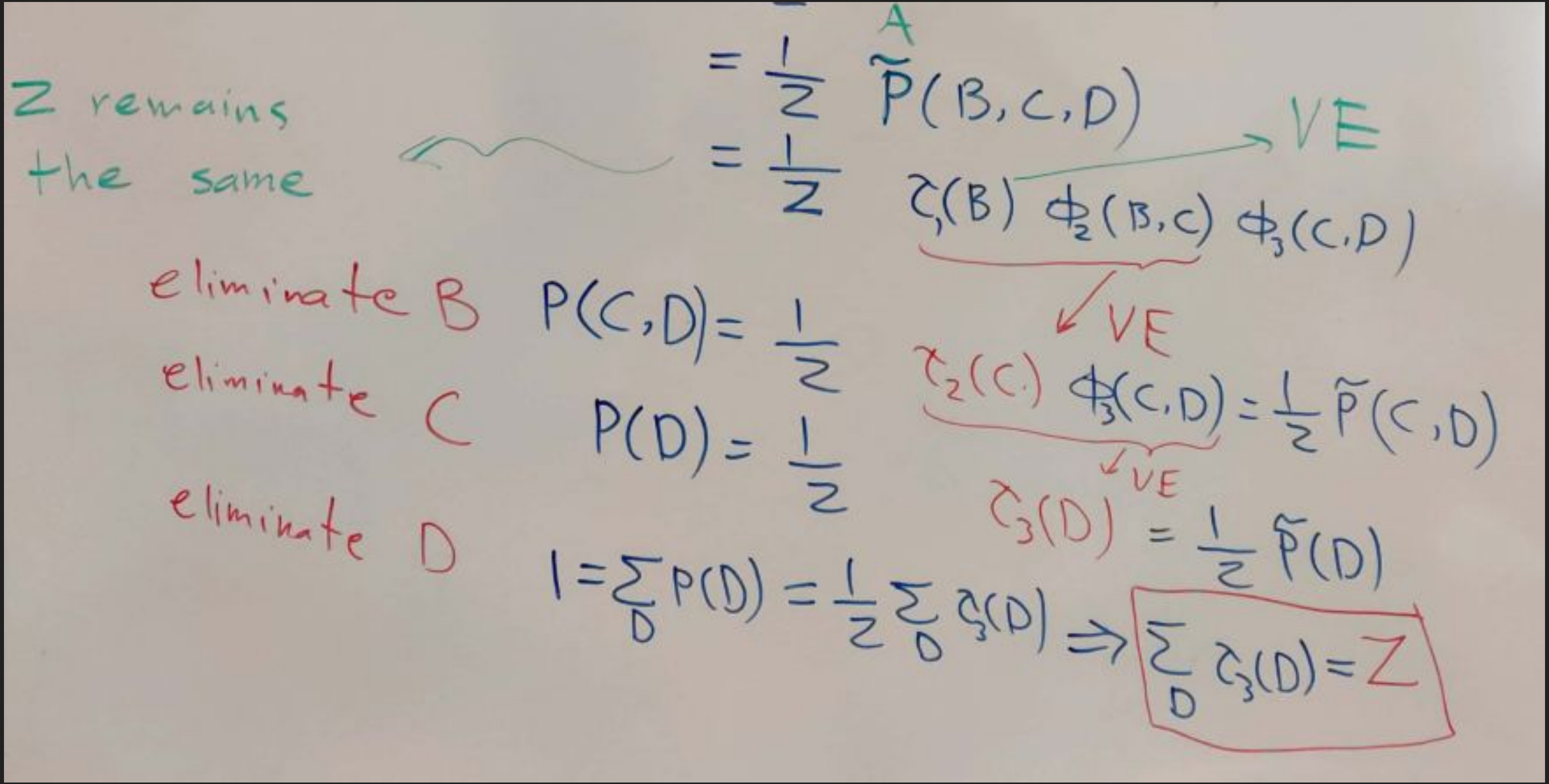
Variable Elimination



Variable Elimination - Unnormalized



Variable Elimination - Unnormalized



Variable Elimination - Unnormalized

K. N. Toosi
University of Technology

$$\sum_{D} \tau_3(D)$$

$$= \sum_{D} \sum_{C} \tau_2(C)\, \phi_3(C,D)$$

$$= \sum_{D} \sum_{C} \sum_{B} \tau_1(B)\, \phi_2(B,C)\, \phi_3(C,D)$$

$$= \sum_{D} \sum_{C} \sum_{B} \sum_{A} \phi_1(A,B)\, \phi_2(B,C)\, \phi_3(C,D)$$

$$= \sum_{A,B,C,D} \tilde{P}(A,B,C,D) = Z$$

Variable Elimination - with evidence

K. N. Toosi
University of Technology

$$P(X) = P(X_e, X_t, X_o) = \underbrace{\frac{1}{Z}\prod_c \Phi_c(X_c)}_{\tilde{P}(X_t, X_e, X_o)}$$

$X_e = x_e$ known

$$P(X_t \mid X_e = x_e) = \frac{\sum_{X_o} P(X_t, x_e, X_o)}{\sum_{X_t}\sum_{X_o} P(X_t, x_e, X_o)} = \frac{\sum_{X_o} \tilde{P}(X_t, x_e, X_o)}{\sum_{X_t}\sum_{X_o} \tilde{P}(X_t, x_e, X_o)}$$

Variable Elimination - with evidence

K. N. Toosi
University of Technology

1- replace $X_e = x_e$

$$\tilde{P}(X_t, x_e, X_o)$$

↓ Variable Elimination
Eliminate all $X_i \in X_o$

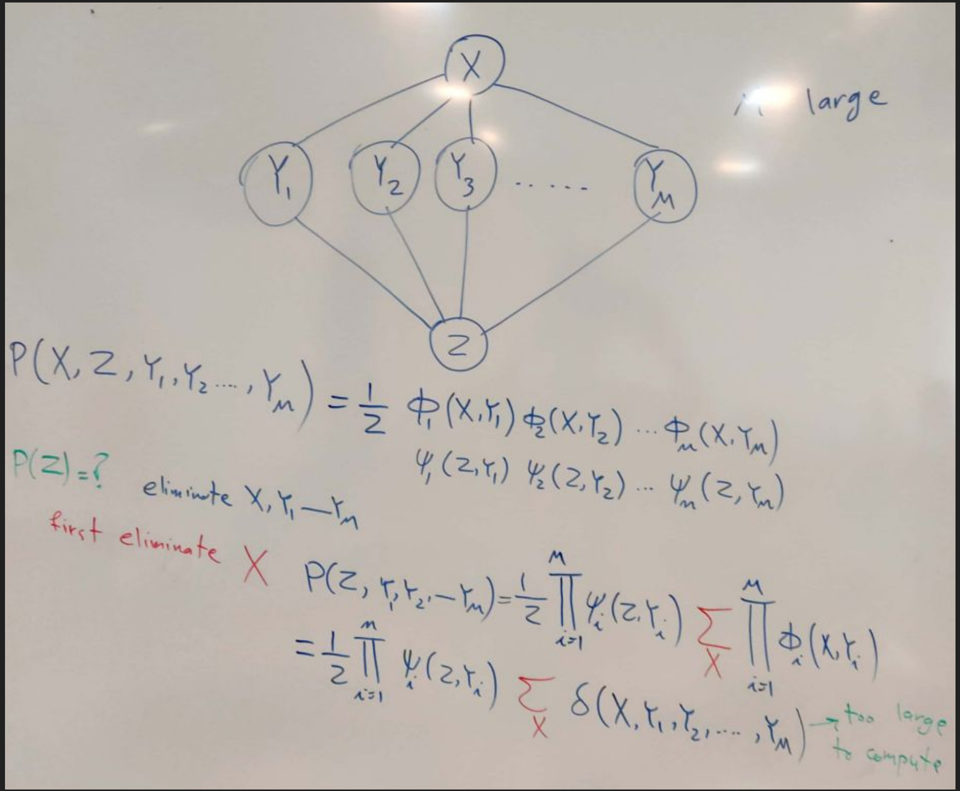
$$\tilde{P}(X_t, x_e)$$

↓ renormalize over X_t

$$P(X_t \mid X_e = x_e) = \frac{\tilde{P}(X_t, x_e)}{\sum_{X_t} \tilde{P}(X_t, x_e)}$$

Elimination order

K. N. Toosi
University of Technology



Elimination order