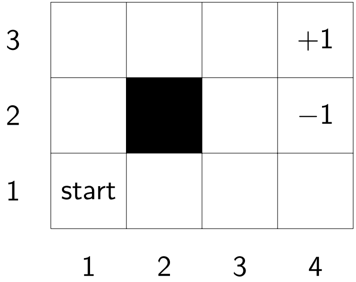
CS-541: Artificial Intelligence
Lecture 8

Abdul Rafae Khan
Department of Computer Science
Stevens Institute of Technology
akhan4@stevens.edu

April 18, 2022

Recap

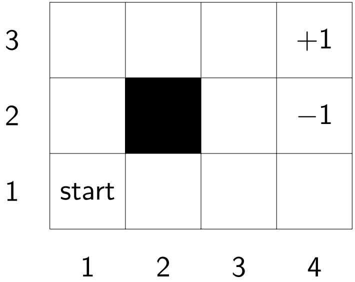
Grid world is 3×4 grid
Start state is at (0,0)
Reward $+1$ at (4,3)
Reward -1 at (4,2)



Recap

For any state, three possible moves

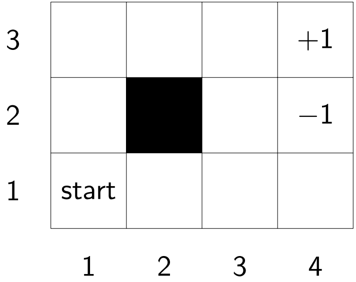
- up: 0.8
- left: 0.1
- right: 0.1



Recap

For any state, three possible moves

- up: 0.8
- left: 0.1
- right: 0.1



How to find the optimal utility?

Recap

For any state, three possible moves

- up: ~~0.8~~ ?
- left: ~~0.1~~ ?
- right: ~~0.1~~ ?

3				$+1$
2				-1
1	start			
	1	2	3	4

Unknown transition probability. How to find the optimal utility?

Recap

Q-Learning:
Start with some Q-values
Iteratively update the Q-values using the following equation

$$\text{for each } (s, a, r, s')$$
$$\hat{Q}_{opt}(s, a) \leftarrow (1 - \eta)\hat{Q}_{opt}(s, a) + \eta\Big(r + \gamma \max_{a' \in Actions(s')} \hat{Q}_{opt}(s', a')\Big)$$

Joint Probability

2 random variables: sunshine $S \in \{0, 1\}$ and rain $R \in \{0, 1\}$
The table represents the probability when each variable has the corresponding value
The total probabilities sum to 1

s	r	$\mathbb{P}(S, R)$
0	0	0.20
0	1	0.08
1	0	0.70
1	1	0.02

Uppercase letters denote the random variable and lowercase letters denote the values

Marginal Probability

s	$\mathbb{P}(S = s)$
0	0.28
1	0.72

Aggregate all the probabilities for the specific value of the random variable
$\mathbb{P}(S = 0) = \mathbb{P}(S = 0, R = 0) + \mathbb{P}(S = 0, R = 1) = 0.20 + 0.08$
The total marginal probabilities should sum to 1

Conditional Probability

s	$\mathbb{P}(S = s \mid R = 1)$
0	0.8
1	0.2

Select only the probabilities for the specific value of the random variable and normalize them

The total normalized conditional probabilities should sum to 1

Probabilistic Inference

Variables: sunshine (S), rain (R), wind (W), humidity (H)
Joint distribution: $\mathbb{P}(S, R, W, H)$
Probabilistic Inference:
Condition on evidence (wind,humidity): $W = 1,\ H = 1$
Interested in query (rain?): R

$$\mathbb{P}(\underbrace{R}_{query} \mid \underbrace{W = 1, H = 1}_{condition})$$

S is marginalized out

Probabilistic Inference

$$\mathbb{P}(\underbrace{R}_{\textit{query}} \mid \underbrace{W = 1, H = 1}_{\textit{condition}})$$

1) we observe some variables (**evidence**)
2) we are interested another set of variables which we didn't observe (**query**)
3) The process of answering this query is called **probabilistic inference**

Overview

How to specify a joint distribution $\mathbb{P}(X_1, \cdots, X_n)$ **compactly**?
- Bayesian network

How to compute queries $\mathbb{P}(R | T = 1, A = 1)$ **efficiently**?
- Variable elimination, Gibbs sampling etc.

Alarm Network

Earthquakes and **Burglary** are independent random variables
An alarm can go off for either of the two

Suppose you get an alarm
If the news reports there was an **earthquake**,
will the probability of **burglary** increase, decrease or remain same?

Bayesian Network (Alarm)

Suppose you get an alarm
If the news reports there was an **earthquake**,
will it change the probability of **burglary**?

Joint probability: $\mathbb{P}(E, B, A)$

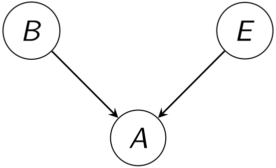
Questions:

$$\mathbb{P}(B = 1 | A = 1) \qquad ? \qquad \mathbb{P}(B = 1 | A = 1, E = 1)$$

Bayesian Network (Alarm)

b	p(B=b)
1	ϵ
0	$1-\epsilon$

e	p(E=e)
1	ϵ
0	$1-\epsilon$



Bayesian Network (Alarm)

b	e	a	$p(a \mid b, e)$
0	0	0	1
0	0	1	0
0	1	0	0
0	1	1	1
1	0	0	0
1	0	1	1
1	1	0	0
1	1	1	0

Bayesian Network (Alarm)

$$
\begin{aligned}
p(b) &= \epsilon \cdot [b = 1] + (1 - \epsilon)[b = 0] \\
p(e) &= \epsilon \cdot [e = 1] + (1 - \epsilon)[e = 0] \\
p(a|b, e) &= [a = (b \vee e)]
\end{aligned}
$$

$$
\mathbb{P}(B = b, E = e, A = a) \stackrel{\text{def}}{=\!=} p(b)p(e)p(a|b, e)
$$

Note:

$\mathbb{P}$ is reserved for the joint distribution over random variables

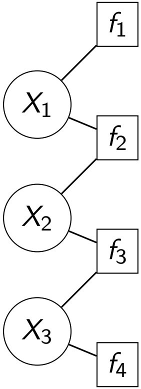
p denotes the local conditional distribution (factor)

Factor Graph

A bipartite graphs which can represent the factorization of a function

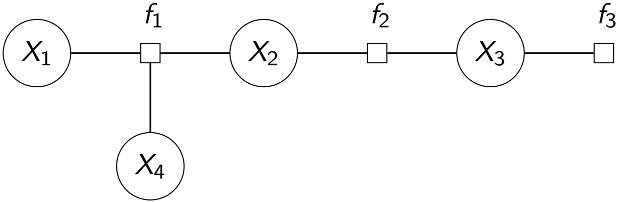
$$\mathbb{P}(X_1, X_2, \cdots, X_n) = \prod_{i=1}^{n} f_i(S_i)$$

$$S \subseteq \{X_1, X_2, \cdots X_4\}$$



• Joint probability written as a product of functions f
• Each function f depends on the subset of the random variables
• Simplify the probability distributions
• Used to visualize and describe independence relations

Factor Graph

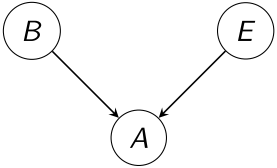


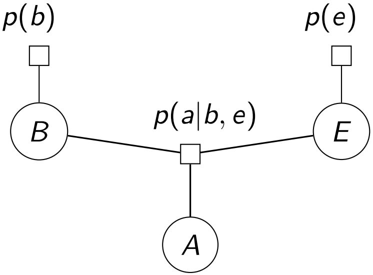
X_1, X_2 and X_4 depend on function f_1, and so on

Joint probability can be written as a product of these functions

$$\mathbb{P}(X_1, X_2, X_3, X_4) = f_1(X_1, X_2, X_4) f_2(X_2, X_3) f_3(X_3)$$

Bayesian Network (Alarm)





$$\mathbb{P}(B = b, E = e, A = a) = p(b)p(e)p(a|b, e)$$

Bayesian networks are special cases of factor graphs

A single factor connects all parents

Probabilistic Inference (Alarm)

Probability of B

b	p(B=b)
1	ϵ
0	$1-\epsilon$

Probability of E

e	p(E=e)
1	ϵ
0	$1-\epsilon$

Joint Probability of B, E, A

b	e	a	$\mathbb{P}(B, E, A)$
0	0	0	$(1-\epsilon)^2$
0	0	1	0
0	1	0	0
0	1	1	$(1-\epsilon)\epsilon$
1	0	0	0
1	0	1	$\epsilon(1-\epsilon)$
1	1	0	0
1	1	1	ϵ^2

Probabilistic Inference (Alarm)

b	e	a	$\mathbb{P}(B, E, A)$
0	0	0	$(1-\epsilon)^2$
0	0	1	0
0	1	0	0
0	1	1	$(1-\epsilon)\epsilon$
1	0	0	0
1	0	1	$\epsilon(1-\epsilon)$
1	1	0	0
1	1	1	ϵ^2

Questions:

$$\mathbb{P}(B=1) = \epsilon(1-\epsilon) + \epsilon^2 = \epsilon$$

$$\mathbb{P}(B=1|A=1) = \frac{\epsilon(1-\epsilon)+\epsilon^2}{\epsilon(1-\epsilon)+\epsilon^2+(1-\epsilon)\epsilon} = \frac{1}{2-\epsilon}$$

$$\mathbb{P}(B=1|A=1,E=1) = \frac{\epsilon^2}{\epsilon^2+(1-\epsilon)\epsilon} = \epsilon$$

Probabilistic Inference (Alarm)

b	e	a	$\mathbb{P}(B, E, A)$
0	0	0	$(1-\epsilon)^2$
0	0	1	0
0	1	0	0
0	1	1	$(1-\epsilon)\epsilon$
1	0	0	0
1	0	1	$\epsilon(1-\epsilon)$
1	1	0	0
1	1	1	ϵ^2

Questions:

$$\mathbb{P}(B=1) = \epsilon(1-\epsilon) + \epsilon^2 = \epsilon$$

$$\mathbb{P}(B=1|A=1) = \frac{\epsilon(1-\epsilon)+\epsilon^2}{\epsilon(1-\epsilon)+\epsilon^2+(1-\epsilon)\epsilon} = \frac{1}{2-\epsilon}$$

$$\mathbb{P}(B=1|A=1, E=1) = \frac{\epsilon^2}{\epsilon^2+(1-\epsilon)\epsilon} = \epsilon$$

Earthquake decrease the probability of burglary!*

* This is not a causal statement!

Probabilistic Inference (Alarm)

b	e	a	$\mathbb{P}(B, E, A)$
0	0	0	$(1-\epsilon)^2$
0	0	1	0
0	1	0	0
0	1	1	$(1-\epsilon)\epsilon$
1	0	0	0
1	0	1	$\epsilon(1-\epsilon)$
1	1	0	0
1	1	1	ϵ^2

Questions:

$$\mathbb{P}(B=1) = \epsilon(1-\epsilon) + \epsilon^2 = \epsilon$$

$$\mathbb{P}(B=1|A=1) = \frac{\epsilon(1-\epsilon)+\epsilon^2}{\epsilon(1-\epsilon)+\epsilon^2+(1-\epsilon)\epsilon} = \frac{1}{2-\epsilon}$$

$$\mathbb{P}(B=1|A=1,E=1) = \frac{\epsilon^2}{\epsilon^2+(1-\epsilon)\epsilon} = \epsilon$$

Earthquake decrease the probability of burglary!*

* This is not a causal statement!

Probabilistic Inference (Alarm)

b	e	a	$\mathbb{P}(B, E, A)$
0	0	0	$(1-\epsilon)^2$
0	0	1	0
0	1	0	0
0	1	1	$(1-\epsilon)\epsilon$
1	0	0	0
1	0	1	$\epsilon(1-\epsilon)$
1	1	0	0
1	1	1	ϵ^2

Questions:

$$\mathbb{P}(B=1) = \epsilon(1-\epsilon) + \epsilon^2 = \epsilon$$

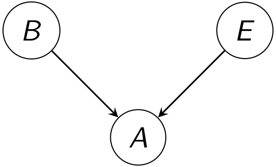
$$\mathbb{P}(B=1|A=1) = \frac{\epsilon(1-\epsilon)+\epsilon^2}{\epsilon(1-\epsilon)+\epsilon^2+(1-\epsilon)\epsilon} = \frac{1}{2-\epsilon}$$

$$\mathbb{P}(B=1|A=1, E=1) = \frac{\epsilon^2}{\epsilon^2+(1-\epsilon)\epsilon} = \epsilon$$

Earthquake decrease the probability of burglary!*

* This is not a causal statement!

Bayesian Network

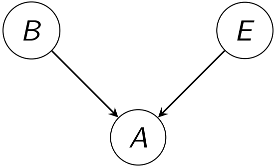


If two causes positively influence an effect.
If we condition on an effect and further condition on one causes
This reduces the probability of the second cause

$$\mathbb{P}(B = 1 | A = 1, E = 1) < \mathbb{P}(B = 1 | A = 1)$$

Note: Only if causes are independent

Conditional Independence



If two causes positively influence an effect.
If we condition on an effect and further condition on one causes
This reduces the probability of the second cause

$$\mathbb{P}(B = 1|A = 1, E = 1) < \mathbb{P}(B = 1|A = 1)$$

Note: Only if causes are independent

Bayesian Network (definition)

Let $X = (X_1, \cdots, X_n)$ are random variables
Bayesian Network is a directed acyclic graph which specifies joint distribution over X as a product of local conditional distributions

$$\mathbb{P}(X_1 = x_1, \cdots, X_n = x_n) \stackrel{\text{def}}{=\!=} \prod_{i=1}^{n} p(x_i | x_{parents(i)})$$

Special Properties

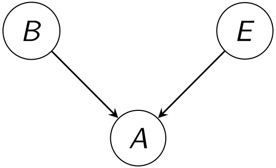
All local distributions satisfy:

$$\mathbb{P}(B = b, E = e) \stackrel{\text{def}}{=\!=} \sum_{x_i} p(x_i | x_{parents(i)}) = 1 \text{ if for each } x_{parents(i)}$$

Implications:

- Consistency of sub-Bayesian networks
- Consistency of conditional distributions

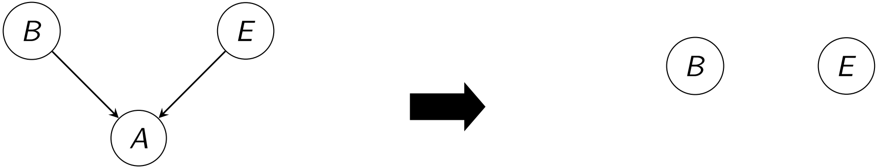
Consistency of sub-Bayesian Networks



$$
\begin{aligned}
\mathbb{P}(B=b, E=e) &\stackrel{\text{def}}{=\!=} \sum_a \mathbb{P}(B=b, E=e, A=a) \\
&\stackrel{\text{def}}{=\!=} \sum_a p(b)p(e)p(a|b,e) \\
&= p(b)p(e)\sum_a p(a|b,e) \\
&= p(b)p(e)
\end{aligned}
$$

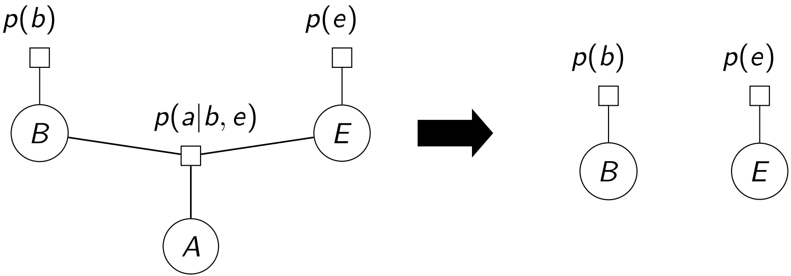
Consistency of sub-Bayesian Networks

Marginalization of a leaf node yields a Bayesian network without the node



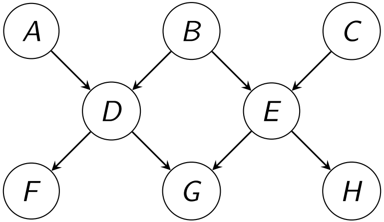
Consistency of sub-Bayesian Networks

Marginalization of a leaf node yields a Bayesian network without the node



Consistency of local conditionals

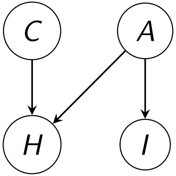
Local conditional distributions are the true conditional distributions.



$$\underbrace{\mathbb{P}(D = d | A = a, B = b)}_{\text{from probabilistic inference}} = \underbrace{p(d|a, b)}_{\text{by definition}}$$

Medical Diagnosis

Cold or Allergies?
If you have cough and itchy eyes, do you have a cold?

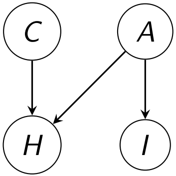


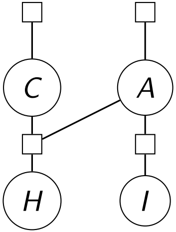
Random variables:
Cold (C), allergy (A), cough (H), and itchy eyes (I)

Medical Diagnosis

Cold or Allergies?

If you have cough and itchy eyes, do you have a cold?

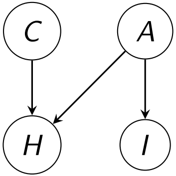




Medical Diagnosis

Cold or Allergies?

If you have cough and itchy eyes, do you have a cold?



Random variables:

Cold (C), allergy (A), cough (H), and itchy eyes (I)

$$\mathbb{P}(C = c, A = a, H = h, I = i) = p(c)p(a)p(h|c,a)p(i|a)$$

$$\mathbb{P}(C = 1|H = 1, I = 1) = ?$$

$$\mathbb{P}(C = 1|H = 1) = ?$$

Medical Diagnosis

Cold or Allergies?

If you have cough and itchy eyes, do you have a cold?

c	$p(c)$
1	0.1
0	0.9

Table: Cold

a	$p(a)$
1	0.2
0	0.8

Table: Allergy

a	i	$p(i\|a)$
1	1	0.9
1	0	0.1
0	0	0.9
0	1	0.1

Table: Itchy eyes

c	a	h	$p(h\|c,a)$
1	0	1	0.9
1	0	0	0.1
0	1	1	0.9
0	1	0	0.1
1	1	1	0.9
1	1	0	0.1
0	0	1	0.1
0	0	0	0.9

Table: Cough

Medical Diagnosis

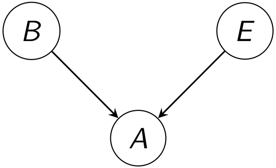
$$\mathbb{P}(C = c, A = a, H = h, I = i) = p(c)p(a)p(h|c, a)p(i|a)$$

$$\begin{aligned}
\mathbb{P}(C = 1|H = 1, I = 1) &= \sum_{a \in A} \mathbb{P}(C = c, A = a, H = h, I = i) \\
&= \sum_{a \in A} p(c)p(a)p(h|c, a)p(i|a)
\end{aligned}$$

Whiteboard

Probabilistic Programming

Probabilistic program: A randomized program that sets the random variable



$$B \sim \mathit{Bernoulli}(\epsilon)$$
$$E \sim \mathit{Bernoulli}(\epsilon)$$
$$A = B \vee E$$

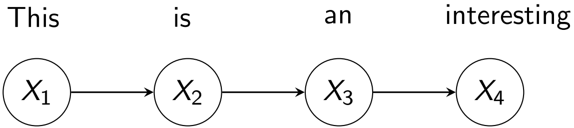
```
def bernoulli(epsilon):
    return random.random()<epsilon

def alarm(epsilon):
    B = bernoulli(epsilon)
    E = bernoulli(epsilon)
    A = (B or E)
    return A
```

Probabilistic Programming: Language Modeling

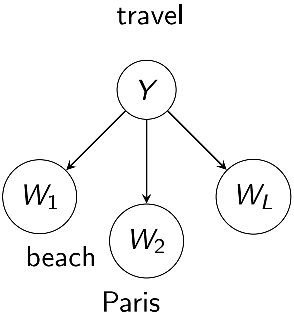
Probabilistic program: A randomized program that sets the random variable

for each position $i = 1, 2, \cdots, n$
    generate words $X_i \sim p(X_i | X_{i-1})$



- Given the previous word, what is the next word?

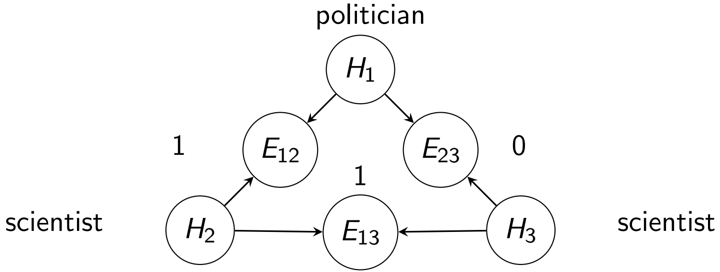
Probabilistic Programming: Topic Modeling




```
generate label Y ~ p(Y)
    for each position i = 1, ..., L:
        generate word W_i ~ p(W_i|Y)
```

- Given a text document, what is the topic?

Probabilistic Programming: Social Network Analysis



for each person $i = 1, \cdots, n$:

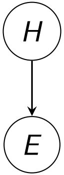
  generate person type $H_i \sim p(H_i)$

for each pair of people $i \neq j$:

  generate connectedness $E_{ij} \sim p(E_{ij} | H_i, H_j)$

- Given a social network (graph over n people), what types of people are there?

Summary so far



- Many different types of models
- Come up with scenarios of how the data (input) was generated through quantities of interest (output)
- Opposite of how we normally do classification!

Probabilistic Inference (definition)

Input

Bayesian network: $\mathbb{P}(X_1, \cdots, X_n)$

Evidence: $E = e$ where $E \subseteq X$ is subset of variables

Query: $Q \subseteq X$ is subset of variables

$$\Downarrow$$

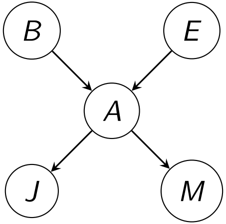
Output

$$\mathbb{P}(Q = q | E = e) \text{ for all values of } q$$

For example: If coughing but no itchy eyes, do you have a cold?

$$\mathbb{P}(C | H = 1, I = 0)$$

Enumeration



$$\mathbb{P}(B|J=j, M=m) = \sum_{a,e} \mathbb{P}(B, J=j, M=m, A, E)$$

$$\sum_{a,e} p(B)p(e)p(a|B,e)p(j|a)p(m|a)$$

So inference in Bayes nets means computing sums of products of numbers

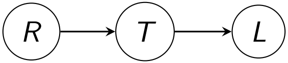
Enumeration

Random Variables: Rain (R), Traffic (T) and Late (L)

$$R \longrightarrow T \longrightarrow L$$

Enumeration

Random Variables: Rain (R), Traffic (T) and Late (L)



$$\mathbb{P}(R=r, T=t, L=l) = \mathbb{P}(R=r)\mathbb{P}(T=t|R=r)\mathbb{P}(L=l|T=t)$$
$$\stackrel{\text{def}}{=\!=} p(r)p(t|r)p(l|t)$$

$P(R)$	
+r	0.1
-r	0.9

$P(T\|R)$		
+r	+t	0.8
+r	-t	0.2
-r	+t	0.1
-r	-t	0.9

$P(L\|T)$		
+t	+l	0.3
+t	-l	0.7
-t	+l	0.1
-t	-l	0.9

Enumeration

Random Variables: Rain (R), Traffic (T) and Late (L)

$$R \rightarrow T \rightarrow L$$

$$\begin{aligned}\mathbb{P}(R = r, T = t, L = l) &= \mathbb{P}(R = r)\mathbb{P}(T = t|R = r)\mathbb{P}(L = l|T = t) \\ &\stackrel{\text{def}}{=\!=} p(r)p(t|r)p(l|t)\end{aligned}$$

$$\mathbb{P}(L = l) = ?$$

Enumeration

Random Variables: Rain (R), Traffic (T) and Late (L)

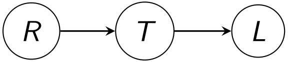
$$R \longrightarrow T \longrightarrow L$$

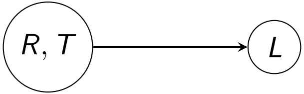
$$\begin{aligned}\mathbb{P}(R = r, T = t, L = l) &= \mathbb{P}(R = r)\mathbb{P}(T = t|R = r)\mathbb{P}(L = l|T = t) \\ &\stackrel{\text{def}}{=\!=} p(r)p(t|r)p(l|t)\end{aligned}$$

$$\mathbb{P}(L = l) = ?$$

Enumeration

Step1: Join factors (similar to database join)





$P(R)$	
+r	0.1
-r	0.9

$\times$

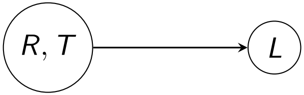
$P(T \mid R)$		
+r	+t	0.8
+r	-t	0.2
-r	+t	0.1
-r	-t	0.9

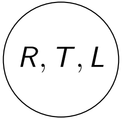
$\rightarrow$

$P(R, T)$		
+r	+t	0.08
+r	-t	0.02
-r	+t	0.09
-r	-t	0.81

Enumeration

Step1: Join factors





$P(T \mid R)$		
+r	+t	0.8
+r	-t	0.2
-r	+t	0.1
-r	-t	0.9

$\times$

$P(L \mid T)$		
+t	+l	0.3
+t	-l	0.7
-t	+l	0.1
-t	-l	0.9

$\rightarrow$

$P(R, T, L)$			
+r	+t	+l	0.024
+r	+t	-l	0.056
+r	-t	+l	0.002
+r	-t	-l	0.018
-r	+t	+l	0.027
-r	+t	-l	0.063
-r	-t	+l	0.081
-r	-t	-l	0.729

Enumeration

Step2: Eliminate factors (remember marginalization)

R, T, L $\qquad\qquad$ T, L

$P(R, T, L)$			
+r	+t	+l	0.024
+r	+t	-l	0.056
+r	-t	+l	0.002
+r	-t	-l	0.018
-r	+t	+l	0.027
-r	+t	-l	0.063
-r	-t	+l	0.081
-r	-t	-l	0.729

$\xrightarrow{\text{Sum out } R}$

$P(L\|T)$		
+t	+l	0.051
+t	-l	0.119
-t	+l	0.083
-t	-l	0.747

Enumeration

Step2: Eliminate factors (remember marginalization)

T, L $\qquad\qquad$ L

$P(L \vert T)$		
+t	+l	0.051
+t	-l	0.119
-t	+l	0.083
-t	-l	0.747

$\xrightarrow{\text{Sum out } T}$

$P(L)$	
+l	0.134
-l	0.866

Variable Elimination

$$X_1 \to X_2 \to X_3 \to X_4$$

Query: $\mathbb{P}(X_3 = x_3 | X_2 = 5)$ for all x_3

$$
\begin{aligned}
\textbf{Tedious way!} &= \frac{\mathbb{P}(X_3 = x_3, X_2 = 5)}{\mathbb{P}(X_2 = 5)} \\
&\propto \mathbb{P}(X_3 = x_3, X_2 = 5) \\
&\propto \sum_{x_1, x_4} \mathbb{P}(X_1 = x_1, X_2 = 5, X_3 = x_3, X_4 = x_4) \\
&\propto \sum_{x_1, x_4} p(x_1) p(x_2 = 5 | x_1) p(x_3 | x_2 = 5) p(x_4 | x_3) \\
&\propto \left(\sum_{x_1} p(x_1) p(x_2 = 5 | x_1) \right) p(x_3 | x_2 = 5) \sum_{x_4} p(x_4 | x_3) \\
&\propto p(x_3 | x_2 = 5)
\end{aligned}
$$

Variable Elimination

$$X_1 \rightarrow X_2 \rightarrow X_3 \rightarrow X_4$$

Query: $\mathbb{P}(X_3 = x_3 | X_2 = 5)$ for all x_3

Faster way!

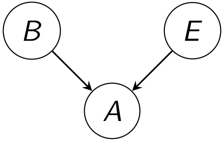
Variable Elimination

Query: $\mathbb{P}(Q|E = e)$ for all x_3

General Probabilistic Inference Strategy

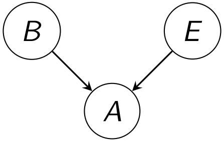
1. Marginalize (remove) non-ancestors of Q or E
2. Convert Bayesian network to factor graph
3. Condition on evidence ($E = e$)
4. Marginalize disconnected
5. Run probabilistic inference algorithm (manual, variable elimination etc.)

Variable Elimination: Alarm



Query: $\mathbb{P}(B)$

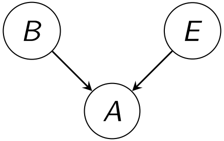
Variable Elimination: Alarm



Query: $\mathbb{P}(B)$

- Marginalize out A and E

Variable Elimination: Alarm

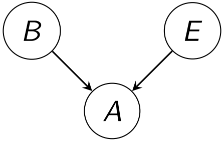


Query: $\mathbb{P}(B)$

- Marginalize out A and E

Query: $\mathbb{P}(B|A=1)$

Variable Elimination: Alarm



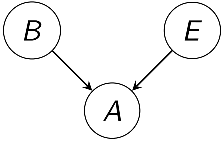
Query: $\mathbb{P}(B)$

- Marginalize out A and E

Query: $\mathbb{P}(B|A=1)$

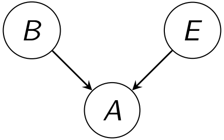
- Condition on $A = 1$

Variable Elimination: Alarm



Query: $\mathbb{P}(B|A=1) = ?$

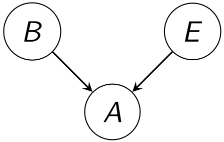
Variable Elimination: Alarm



Query: $\mathbb{P}(B|A = 1) =?$

$$\mathbb{P}(B = b|A = 1) \propto p(b)f(b)$$
$$f(b) = \sum_{e} p(e)p(a = 1|b, e)$$

Variable Elimination: Alarm



Query: $\mathbb{P}(B|A=1) = ?$

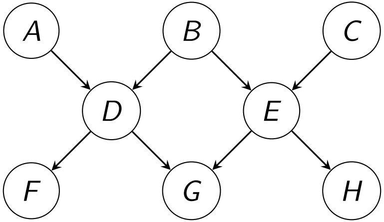
$$\mathbb{P}(B=b|A=1) \propto p(b)f(b)$$

$$\mathbb{P}(B=1|A=1) = \epsilon(\epsilon + (1-\epsilon)) = \epsilon$$

$$\mathbb{P}(B=0|A=1) = (1-\epsilon)(\epsilon + 0) = \epsilon(1-\epsilon)$$

$$\implies \mathbb{P}(B=b|A=1) = \frac{\epsilon}{\epsilon + \epsilon(1-\epsilon)} = \frac{1}{2-\epsilon}$$

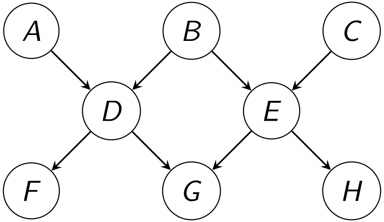
Variable Elimination



Query: $\mathbb{P}(C|B = b)$

Query: $\mathbb{P}(C, H|E = e)$

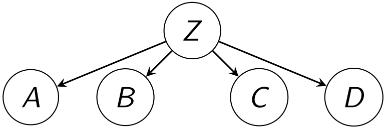
Variable Elimination



Query: $\mathbb{P}(C|B=b)$
- Marginalize out everything else, note $C \perp\!\!\!\perp B$

Query: $\mathbb{P}(C, H|E=e)$
- Marginalize out A, D, F, G, note $C \perp\!\!\!\perp H|E$

Order Matters

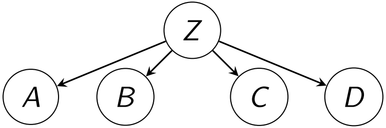


Order the terms Z, A, B, C, D

$$
\begin{aligned}
P(D) &= \sum_{z,a,b,c} p(z)p(a|z)p(b|z)p(c|z)p(D|z) \\
&= \sum_{z} p(z) \sum_{a} p(a|z) \sum_{b} p(b|z) \sum_{c} p(c|z)p(D|z)
\end{aligned}
$$

- Largest factor has 2 variables (D, Z)

Order Matters



Order the terms A, B, C, D, Z

$$\begin{aligned} P(D) &= \sum_{z,a,b,c} p(a|z)p(b|z)p(c|z)p(D|z)p(z) \\ &= \sum_a \sum_b \sum_c \sum_z p(a|z)p(b|z)p(c|z)p(D|z)p(z) \end{aligned}$$

- Largest factor has 4 variables (A, B, C, D)
- In general, with n leaves, factor of size 2^n

Recap

Bayesian Networks
Probabilistic programming
Probabilistic Inference
Enumeration Method
Variable Elimination Method
Variable Elimination Method with factor graph

References

Stuart Russell and Xiaodong Song (2021)
CS 188 — Introduction to Artificial Intelligence
University of California, Berkeley

Chelsea Finn and Nima Anari (2021)
CS221 — Artificial Intelligence: Principles and Techniques
Stanford University

The End