CS 331: Artificial Intelligence
Bayesian Networks (Inference)

1

Inference

- Suppose you are given a Bayesian network with the graph structure and the parameters all figured out

- Now you would like to use it to do inference

- You need inference to make predictions or classifications with a Bayes net

2

Inference Example

- Suppose you are using a movie recommendation service which looks at the following attributes of a movie:
  - *Genre: Action, Comedy, or Romance*
  - *Rating: G, PG-13, or R*
  - *HasRobotsTakingOverWorld*: *true* or *false*
- The class label is *Recommendation* which values *true* or *false*.
- Need to predict:

  P(*Recommendation = true | Genre = Action, Rating = R, HasRobotsTakingOverWorld = true)*

3

Another Example

- You are very sick and you visit your doctor.
- The doctor is able to get the following information from you:
  - *HasFever = true*
  - *HasCough = true*
  - *HasBreathingProblems = true*
  - *AteBaconRecently = true*
- What's the probability you have *SwineFlu* given the above?

4

Another Example

- Need to compute
  P(*SwineFlu = true | HasFever = true,
  HasCough = true, HasBreathingProblems =
  true, AteBaconRecently = true*)
- Suppose you pass out before you say a word to
  the doctor. The doctor is only able to
  determine you have a fever. What is
  P(*SwineFlu = true | HasFever = true*)?

5

Queries

*P(Recommendation=true | Genre=Action, Rating = R,
HasRobotsTakingOverWorld=true*)

P(*SwineFlu = true | HasFever = true, HasCough = true,
HasBreathingProblems = true, AteBaconRecently = true*)

P(*SwineFlu = true | HasFever = true)*

These are all called queries

6

3

Query Example

P(*SwineFlu = true | HasFever = true)*

Query Variable

Evidence Variable

Unobserved variables: *HasCough*, *HasBreathingProblems*, *AteBaconRecently*

Queries Formalized

We will use the following notation:
- X = query variable
- $E = \{E_1, \ldots, E_m\}$ is the set of evidence variables
- e = observed event
- $Y = \{Y_1, \ldots, Y_l\}$ are the non-evidence (or hidden) variables
- The complete set of variables $\boldsymbol{X} = \{X\} \cup \boldsymbol{E} \cup \boldsymbol{Y}$

Need to calculate the query $\boldsymbol{P}(X \mid \boldsymbol{e})$

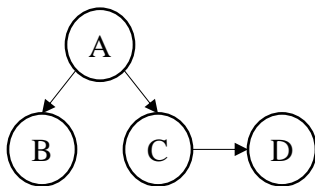# Inference by Enumeration

- Recall that:

$$P(X \mid e) = \alpha P(X, e) = \alpha \sum_{y} P(X, e, y)$$

$$P(x_1, ..., x_n) = \prod_{i=1}^{n} P(x_i \mid parents(X_i))$$

This means you can answer queries by computing sums of products of conditional probabilities from the network
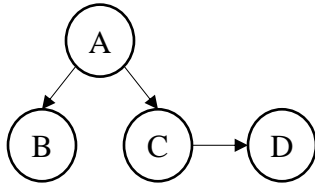
9

Example #1

Query: P(B=true | C=true)

How do you solve this? 2 steps:

1. Express it in terms of the joint probability distribution P(A, B, C,D)

2. Express the joint probability distribution in terms of the entries in the CPTs of the Bayes net
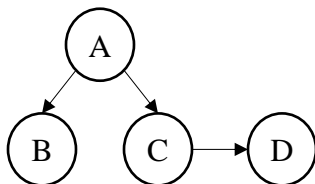
10

5

Example #1

Whenever you see a conditional like P(B=true | C=true), use the Chain Rule:

P(B | C) = P(B, C) / P(C)

$$P(B = true \mid C = true)$$

$$= \frac{P(B = true, C = true)}{P(C = true)}$$

Example #1

Whenever you need to get a subset of the variables e.g. P(B,C) from the full joint distribution P(A,B,C,D), use marginalization:
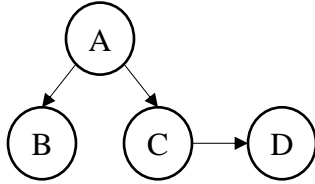
$$P(X) = \sum_{y} P(X, Y = y)$$

$$P(B = true \mid C = true)$$

$$= \frac{P(B = true, C = true)}{P(C = true)}$$

$$= \frac{\sum_{a} \sum_{d} P(A = a, B = true, C = true, D = d)}{\sum_{a} \sum_{b} \sum_{d} P(A = a, B = b, C = true, D = d)}$$

Example #1

To express the joint probability distribution as the entries in the CPTs, use:

$$P(X_1,...,X_N)$$
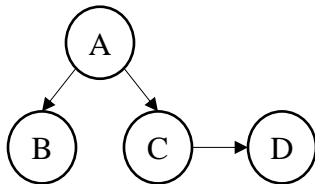
$$= \prod_{i=1}^{N} P(X_i \mid Parents(X_i))$$

$$= \frac{\sum_a \sum_d P(A = a, B = true, C = true, D = d)}{\sum_a \sum_b \sum_d P(A = a, B = b, C = true, D = d)}$$

$$= \frac{\sum_a \sum_d P(A = a)P(B = true \mid A = a)P(C = true \mid A = a)P(D = d \mid C = true)}{\sum_a \sum_b \sum_d P(A = a)P(B = b \mid A = a)P(C = true \mid A = a)P(D = d \mid C = true)}$$
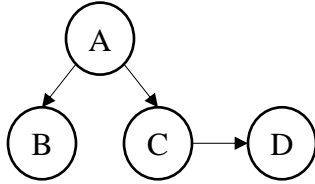
13

Example #1

Take the probabilities that don't depend on the terms in the summation and move them outside the summation

$$= \frac{\sum_a \sum_d P(A = a)P(B = true \mid A = a)P(C = true \mid A = a)P(D = d \mid C = true)}{\sum_a \sum_b \sum_d P(A = a)P(B = b \mid A = a)P(C = true \mid A = a)P(D = d \mid C = true)}$$

$$= \frac{\sum_a P(A = a)P(B = true \mid A = a)P(C = true \mid A = a)\sum_d P(D = d \mid C = true)}{\sum_a P(A = a)\sum_b P(B = b \mid A = a)P(C = true \mid A = a)\sum_d P(D = d \mid C = true)}$$

Example #1
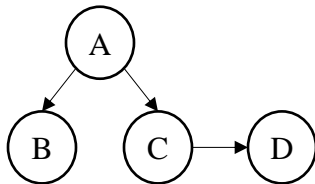

Take the probabilities that don't depend on the terms in the summation and move them outside the summation

Sums to 1

$$= \frac{\sum_a \sum_d P(A=a)P(B=true \mid A=a)P(C=true \mid A=a)P(D=d \mid C=true)}{\sum_a \sum_b \sum_d P(A=a)P(B=b \mid A=a)P(C=true \mid A=a)P(D=d \mid C=true)}$$

$$= \frac{\sum_a P(A=a)P(B=true \mid A=a)P(C=true \mid A=a)\sum_d P(D=d \mid C=true)}{\sum_a P(A=a)\sum_b P(B=b \mid A=a)P(C=true \mid A=a)\sum_d P(D=d \mid C=true)}$$

Sums to 1

Example #1
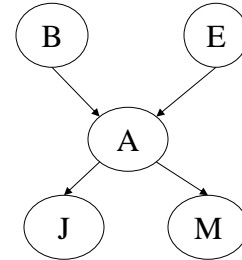

Take the probabilities that don't depend on the terms in the summation and move them outside the summation

$$= \frac{\sum_a P(A=a)P(B=true \mid A=a)P(C=true \mid A=a)}{\sum_a P(A=a)\sum_b P(B=b \mid A=a)\underline{P(C=true \mid A=a)}}$$ ← Doesn't depend on b. Can move to the left

$$= \frac{\sum_a P(A=a)P(B=true \mid A=a)P(C=true \mid A=a)}{\sum_a P(A=a)P(C=true \mid A=a)\underline{\sum_b P(B=b \mid A=a)}}$$ ← Sums to 1

$$= \frac{\sum_a P(A=a)P(B=true \mid A=a)P(C=true \mid A=a)}{\sum_a P(A=a)P(C=true \mid A=a)}$$

Example #2

$P(B = true \mid J = true, M = true)$

$= \dfrac{P(B = true, J = true, M = true)}{P(J = true, M = true)}$

$= \dfrac{\sum_e \sum_a P(B = true, E = e, A = a, J = true, M = true)}{\sum_b \sum_e \sum_a P(B = b, E = e, A = a, J = true, M = true)}$

$= \dfrac{\sum_e \sum_a \left(\begin{array}{l} P(B = true)P(E = e)P(A = a \mid B = true, E = e)* \\ P(J = true \mid A = a)P(M = true \mid A = a) \end{array} \right)}{\sum_b \sum_e \sum_a \left(\begin{array}{l} P(B = b)P(E = e)P(A = a \mid B = b, E = e)* \\ P(J = true \mid A = a)P(M = true \mid A = a) \end{array} \right)}$

$= \dfrac{P(B = true)\sum_e P(E = e)\sum_a \left(\begin{array}{l} P(A = a \mid B = true, E = e)* \\ P(J = true \mid A = a)P(M = true \mid A = a) \end{array} \right)}{\sum_b P(B = b)\sum_e P(E = e)\sum_a \left(\begin{array}{l} P(A = a \mid B = b, E = e)* \\ P(J = true \mid A = a)P(M = true \mid A = a) \end{array} \right)}$

B   E

A

J   M

17

Complexity of Exact Inference

Burglary   Earthquake

Alarm

JohnCalls   MaryCalls
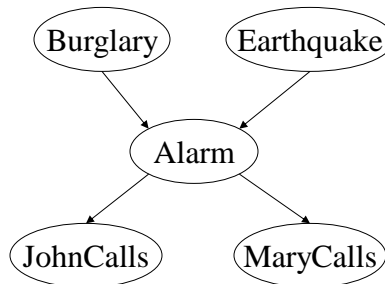
- The Burglary/Earthquake Bayesian network is an example of a polytree
- Singly connected networks (aka polytrees) have at most one undirected path between any two nodes in the network

18

9

Complexity of Exact Inference

- Polytrees have a nice property:

  The time and space complexity of exact inference in polytrees is linear in the number of variables

- What about multiply connected networks?

```
        Cloudy
       /      \
      /        \
 Sprinkler    Rain
      \        /
       \      /
        Wet
        Grass
```
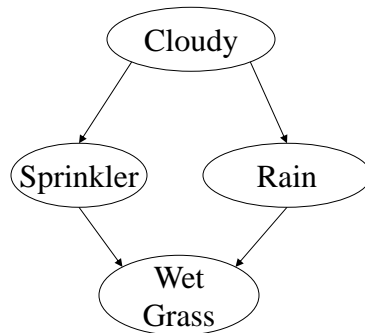
Complexity of Exact Inference

- What about for multiply connected networks?
- Exponential time and space complexity in the number of variables in the worst case
- Bad news: Inference in Bayesian networks is NP-hard
- Even worse news: inference is #P-hard (strictly harder than NP-complete problems)

20

The Good News

- Although exact inference is NP-hard, approximate inference is tractable
  - Lots of promising methods like sampling, MCMC, variational methods, etc.
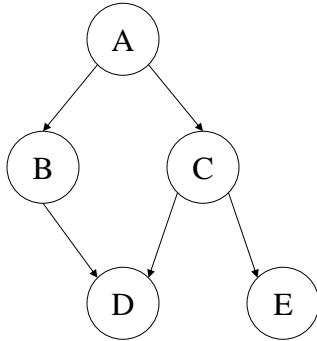- Approximate inference is a current research topic in Machine Learning

21

What You Should Know

- How to do exact inference in probabilistic queries of Bayes nets
- The complexity of inference for polytrees and multiply connected networks
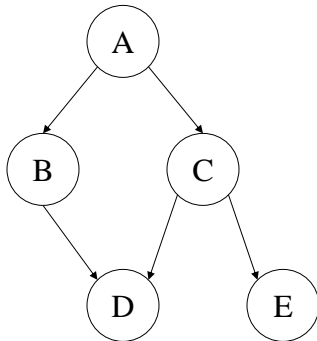
22

In-Class Exercise

A

B   C

D   E

Write out the equations for the following probabilities using probabilities you can obtain from the Bayesian network. You will have to leave it in symbolic form because the CPTs are not shown.

1. P(A=true, B=true, C=true, D=true, E=true)

23

In-Class Exercise

A

B   C

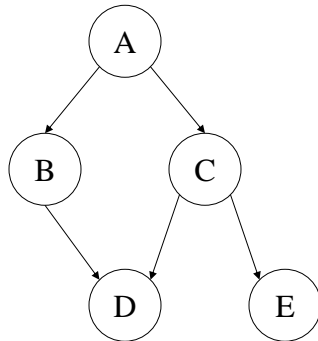D   E

2. P(B=true | D=true)

24

12

In-Class Exercise

2. P(A=true, D=true, E=true | B=true, C=true)
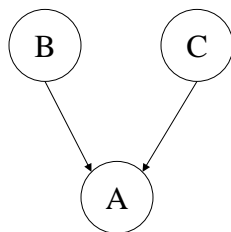
In-Class Exercise

B	P(B)
false	0.25
true	0.75

C	P(C)
false	0.1
true	0.9

C	B	A	P(A|B,C)
false	false	false	0.1
false	false	true	0.9
false	true	false	0.2
false	true	true	0.8
true	false	false	0.3
true	false	true	0.7
true	true	false	0.4
true	true	true	0.6

Can you come up with another Bayes net structure (using only the 3 nodes above) that represents the same joint probability distribution?
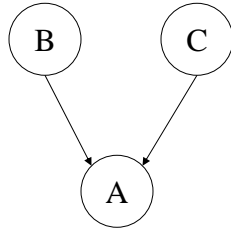
In-Class Exercise

B	P(B)
false	0.25
true	0.75

C	P(C)
false	0.1
true	0.9

C	B	A	P(A|B,C)
false	false	false	0.1
false	false	true	0.9
false	true	false	0.2
false	true	true	0.8
true	false	false	0.3
true	false	true	0.7
true	true	false	0.4
true	true	true	0.6

What is P(B=false,C=false)? Can you think of a quick way
to calculate this?

27