Tufts

Lecture 13: Independence & Inference in Bayes Nets

Artificial Intelligence (CS 131)

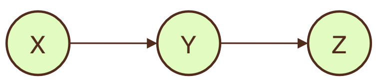
1

Independence in Bayes Nets

- The descendants of a node in a BN are its children + children's children + …
  - Anything else (upstream or down) is a non-descendant
- A node is conditionally independent of all of its *non*-descendants, *given* its parents
  - **Note**: *not* saying *absolutely* independent, only conditionally
- Question: given some particular evidence, what are all the actual independence relationships?

2

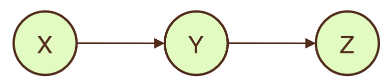
Simplest Case: Indirect Connection



- When we look at a BN, if there is no arrow connecting two things directly, then there is no *direct* influence
  - However, *indirect* connections can exist
  - When they do, independence depends upon *evidence* we have
- Given the basic definition, is it true that Z is independent of X, given Y?

3

Independence in Indirect Connection



- Given basic definition, is Z independent of X, given Y?
- The graph structure tells us that:

$$P(X,Y,Z) = P(X)\,P(Y \mid X)\,P(Z \mid Y)$$

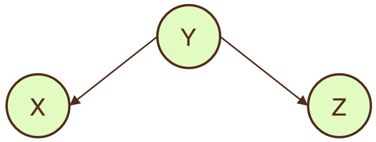
- Therefore, we can show:

$$P(Z \mid X,Y) = \frac{P(X,Y,Z)}{P(X,Y)} = \frac{P(X)\,P(Y \mid X)\,P(Z \mid Y)}{P(Y \mid X)P(X)} = P(Z \mid Y)$$

- If we *know* the value of Y, then Z and X *are* independent

4

Common Cause and Independence

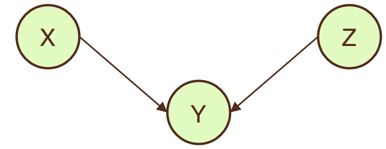


- Here, both X and Z depend on Y (like symptoms of the same disease), but there is no direct connection between them
- Again, if we *know* Y, then Z and X *are* independent

$$P(Z \mid X, Y) = \frac{P(X, Y, Z)}{P(X, Y)} = \frac{P(Y)\, P(X \mid Y)\, P(Z \mid Y)}{P(X \mid Y) P(Y)} = P(Z \mid Y)$$

- Note that if we *do not* know Y, then Z and X *may have* some dependency upon one another

Shared Effects and Independence



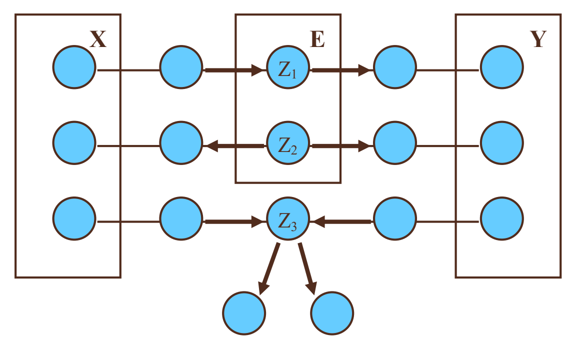
- Here, we have no arrow between X and Z, which tells us that they are marginally independent (if we have *no evidence*)
- However, this is very different from before: if we *do* know Y, then X and Z *may not* be independent any more
- Perhaps X and Z are two separate coin flips, and Y is a variable that is true if and only if both coins come up the same
  - Then X and Z *are* independent flips *by themselves*
  - If we know Y is true, however, X and Z are completely *dependent*

d-separation in Bayes Nets

- A basic way to define and discover independence in a BN
- Definition: If X, Y and E are *distinct* sets of nodes in a BN, then E d-separates X and Y if it blocks every undirected path from X to Y
- A path is blocked by E if there exists node z such that:

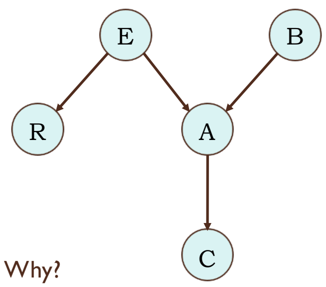
1. z is in E and z has one incoming & one outgoing arrow
2. z is in E and z has two outgoing arrows
3. z has two incoming arrows and *neither* z *nor any of its descendants* are in E

Three types of path-blocking



d-separation Implies Independence

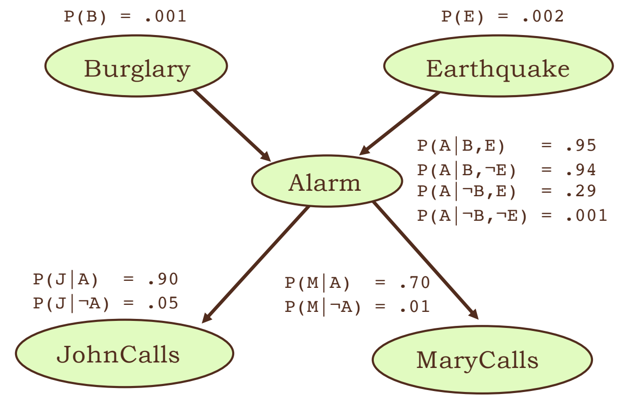
- If sets of variables X and Y *are* d-separated by E, *then* they are conditionally independent given E
- If they are *not* d-separated, then we *cannot assume* that they are independent



- Consider the example:

1. Is R independent of C given E? Why?
2. Is R independent of C given A? Why?
3. Is E independent of B given A? Why?
4. Is E independent of B given C? Why?
5. Is E independent of B given *no evidence at all*? Why?

Basic Inference in Bayes Nets: Earthquakes & Burglaries



Calculating Joint Probabilities

- Remember: given a joint distribution over all variables, we can answer *any* probabilistic query
  - A Bayes Net represents a joint distribution compactly
  - We can use BN to compute things like joint likelihood
- For instance, say we want to know how likely it is that both John and Mary call while a false alarm is occurring (i.e., neither a burglary nor an earthquake)

$$P(J, M, A, \neg B, \neg E) = ?$$

Calculating Joint Probabilities

- We want to know how likely it is that both John and Mary call while a false alarm is occurring…
- By the chain rule, we can write this as:

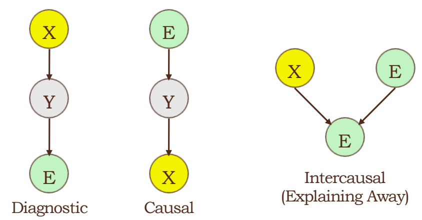
$$P(J, M, A, \neg B, \neg E) = P(J \mid M, A, \neg B, \neg E)\, P(M \mid A, \neg B, \neg E)\, P(A \mid \neg B, \neg E)\, P(\neg B \mid \neg E)\, P(\neg E)$$

- Now we can *simplify* the equation, *eliminating* any of the conditional terms that are *not direct parents*, and then solve since we have all necessary remaining numbers in the CPTs:

$$P(J, M, A, \neg B, \neg E) = P(J \mid A)\, P(M \mid A)\, P(A \mid \neg B, \neg E)\, P(\neg B)\, P(\neg E)$$
$$= 0.9 \times 0.7 \times 0.001 \times 0.999 \times 0.998 = \mathbf{0.0067}$$

General Inference in Bayes Nets

- BN shows probabilistic relationships among variables
- When we make a query (do inference), we divide the variables into 3 separate parts:
  1. Query (X): variables whose probability we want to know
  2. Evidence (E): known values of some variables
  3. Remainder (Y): other variables left over



Examples of Different Types of Queries

- Posterior: the probability of X, given evidence E:

$$P(X \mid E) = \frac{P(X,E)}{P(E)} = \alpha P(X,E) = \alpha \sum_{Y} P(X,E,Y)$$

- Most probable explanation: combination of *all* the remaining variables Y with highest probability given evidence E:

$$MPE(E) = \text{argmax}_y \; P(y,e)$$

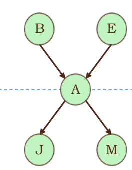
- Maximum a posteriori (MAP): combination of *some* variables V with highest probability given evidence E:

$$MAP(V \mid E) = \text{argmax}_v \; P(v,e)$$

Examples of MAP Queries

- Speech recognition: take the sound signal as the evidence, and calculate the most likely sequence of words that might have produced it
- Face identification: take a set of labeled examples of different people (assigning a name to each picture), plus a new picture as evidence, and calculate the most likely name of the person in the new picture
- Web search: take past user linking and page content, plus the new words typed into the search box as evidence, and generate a list of pages, ordered from most likely down to less likely

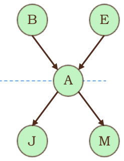
Simple Enumerative Inference



- The BN framework allows us to represent our joint probability distributions in an intelligent way
- We would like to be able to query in a smart way, too, without adding up too many terms
- Consider the query:

$$\begin{aligned} P(B \mid j,m) &= P(B,j,m)/P(j,m) \\ &= \alpha\, P(B,j,m) \\ &= \alpha \sum_{e} \sum_{a} P(B,e,a,j,m) \end{aligned}$$

Simple Enumerative Inference



- We can rewrite the full joint using CPT entries:

$$P(B\,|\,j,m) = \alpha \sum_e \sum_a P(B)\,P(e)\,P(a\,|\,B,e)\,P(j\,|\,a)\,P(m\,|\,a)$$
$$= \alpha\, P(B) \sum_e P(e) \sum_a P(a\,|\,B,e)\,P(j\,|\,a)\,P(m\,|\,a)$$

- We can calculate the normalization constant in the same way:

$$\alpha = \frac{1}{P(j,m)}$$
$$= 1/\sum_b \sum_e \sum_a P(b,e,a,j,m)$$
$$= 1/\sum_b \sum_e \sum_a P(b)\,P(e)\,P(a\,|\,b,e)\,P(j\,|\,a)\,P(m\,|\,a)$$
$$= 1/\sum_b P(b) \sum_e P(e) \sum_a P(a\,|\,b,e)\,P(j\,|\,a)\,P(m\,|\,a)$$

17

An Enumerative Inference Algorithm

function ENUMERATION-ASK(X, **e**, bn) **returns** a distribution over X
  inputs: X, the query variable
    e, observed values for variables **E**
    bn, a Bayes net with variables $vars$

  $\mathbf{Q}(X) \leftarrow$ a distribution over X, initially empty
  for each value x_i of X **do**
    $\mathbf{Q}(x_i) \leftarrow$ ENUMERATE-ALL($vars$, $\mathbf{e}_{x_i}$)
      where $\mathbf{e}_{x_i}$ is **e** extended with $X = x_i$
  return NORMALIZE($\mathbf{Q}(X)$)

function ENUMERATE-ALL($vars$, **e**) **returns** a real number
  if EMPTY?($vars$) **then return** 1.0
  $V \leftarrow$ FIRST($vars$)
  if V is an evidence variable with value v in **e**
    then return $P(v\,|\,parents(V)) \times$ ENUMERATE-ALL(REST($vars$), **e**)
    else return $\sum_v P(v\,|\,parents(V)) \times$ ENUMERATE-ALL(REST($vars$), $\mathbf{e}_v$)
      where $\mathbf{e}_v$ is **e** extended with $V = v$

For each value $x_i \in X$, want to know

$$P(x_i|e) = \frac{P(x_i,\, e)}{P(e)}$$

ENUMERATE-ALL calculates $Q(x_i) = P(x_i,\, e)$.

NORMALIZE($\mathbf{Q}(X)$): multiplies $\mathbf{Q}(X)$ by $1/P(\mathbf{e})$.

Calculate $P(\mathbf{e})$ by running: ENUMERATE-ALL(bn.VARS, **e**).

18

An Enumerative Inference Algorithm

function ENUMERATION-ASK(X, **e**, bn) **returns** a distribution over X
  inputs: X, the query variable
    e, observed values for variables **E**
    bn, a Bayes net with variables $vars$

  $\mathbf{Q}(X) \leftarrow$ a distribution over X, initially empty
  for each value x_i of X **do**
    $\mathbf{Q}(x_i) \leftarrow$ ENUMERATE-ALL($vars$, $\mathbf{e}_{x_i}$)
      where $\mathbf{e}_{x_i}$ is **e** extended with $X = x_i$
  return NORMALIZE($\mathbf{Q}(X)$)

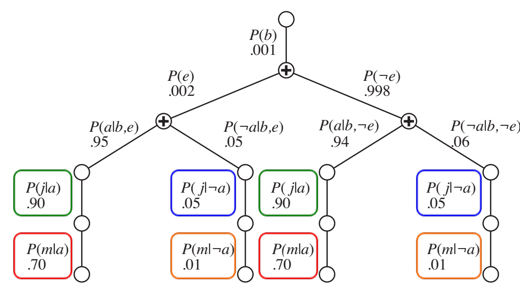
function ENUMERATE-ALL($vars$, **e**) **returns** a real number
  if EMPTY?($vars$) **then return** 1.0
  $V \leftarrow$ FIRST($vars$)
  if V is an evidence variable with value v in **e**
    then return $P(v\,|\,parents(V)) \times$ ENUMERATE-ALL(REST($vars$), **e**)
    else return $\sum_v P(v\,|\,parents(V)) \times$ ENUMERATE-ALL(REST($vars$), $\mathbf{e}_v$)
      where $\mathbf{e}_v$ is **e** extended with $V = v$

For every variable V in $vars$, calculation depends upon whether we have evidence ($v \in e$) or not

To be able to properly access probabilities from CPTs, must always select V such that $parents(V) \in e$

19

The Evaluation Tree



- Although suitable for reasonably sized BNs, this is an inefficient algorithm, due to repetition
  - Runs in time $O(d^n)$ for n variables with domains of size d
  - Caching results to avoid re-computation can help somewhat to make this more manageable

20

Answering Queries Intelligently

- There are many equivalent ways of calculating probabilities in a BN
  - If we aren't smart, however, we can end up wasting effort, adding up the same numbers over and over again
  - Algorithms are designed to try to rearrange problems to avoid this problem
- Variable Elimination: re-orders and re-writes probability calculations to reduce the amount of work

Variable Elimination in Chains

- Consider BN in form of a ***chain***:

$$A \rightarrow B \rightarrow C \rightarrow D \rightarrow E$$

- Get probability of value e by marginalizing

$$P(e) = \sum_d \sum_c \sum_b \sum_a P(a,b,c,d,e)$$

- So, by chain rule & basic facts about BNs:

$$P(e) = \sum_d \sum_c \sum_b \sum_a P(a)\,P(b\,|\,a)\,P(c\,|\,a,b)\,P(d\,|\,a,b,c)\,P(e\,|\,a,b,c,d)$$

$$= \sum_d \sum_c \sum_b \sum_a P(a)\,P(b\,|\,a)\,P(c\,|\,b)\,P(d\,|\,c)\,P(e\,|\,d)$$

- **Note**: no sum over e (*want to know* that value)

Variable Elimination in Chains

$$A \rightarrow B \rightarrow C \rightarrow D \rightarrow E$$

- Now, rearrange the terms to save effort
- Since variable a does not appear in last three $P(\ldots)$ terms, move them ***outside*** of that sum

$$P(e) = \sum_d \sum_c \sum_b \sum_a P(a)\,P(b\,|\,a)\,P(c\,|\,b)\,P(d\,|\,c)\,P(e\,|\,d)$$

$$= \sum_d \sum_c \sum_b P(c\,|\,b)\,P(d\,|\,c)\,P(e\,|\,d) \sum_a P(a)\,P(b\,|\,a)$$

Variable Elimination in Chains

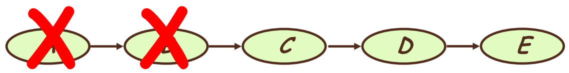
$$\cancel{A} \rightarrow B \rightarrow C \rightarrow D \rightarrow E$$

- Now we can ***eliminate*** a, by doing inner summation:

$$P(e) = \sum_d \sum_c \sum_b P(c\,|\,b)\,P(d\,|\,c)\,P(e\,|\,d) \underbrace{\sum_a P(a)\,P(b\,|\,a)}$$

$$= \sum_d \sum_c \sum_b P(c\,|\,b)\,P(d\,|\,c)\,P(e\,|\,d) P(b)$$

Variable Elimination in Chains



- Next, we can eliminate b, rearranging again to move terms without b outside, and summing:

$$P(e) = \sum_d \sum_c \sum_b P(c \mid b)\, P(d \mid c)\, P(e \mid d)\, P(b)$$

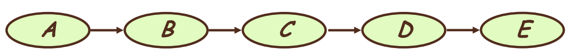
$$= \sum_d \sum_c P(d \mid c)\, P(e \mid d) \sum_b P(c \mid b)\, P(b)$$

$$= \sum_d \sum_c P(d \mid c)\, P(e \mid d)\, P(c)$$

What About Evidence?

- These basic chain elimination steps can continue, eliminating c and then d the same way, until we have $P(e)$
- This is the elementary (prior) probability of e, given *no knowledge* other than the basic BN
- However, we can also reason about the probability of *other* variables, *given evidence*
- What if we *already know* the value of evidence e? How do we calculate the probability of its ancestor a (which may be its *actual* cause)?

Chain Elimination with Evidence



- Work backwards, ***from*** evidence (e) **to** ancestor (a)
- Again, chain rule and basic properties of BNs gives:

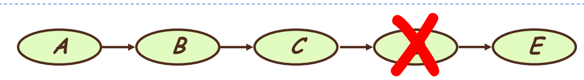
$$P(a,e) = \sum_b \sum_c \sum_d P(a,b,c,d,e)$$

$$= \sum_b \sum_c \sum_d P(a)\, P(b \mid a)\, P(c \mid a,b)\, P(d \mid a,b,c)\, P(e \mid a,b,c,d)$$

$$= \sum_b \sum_c \sum_d P(a)\, P(b \mid a)\, P(c \mid b)\, P(d \mid c)\, P(e \mid d)$$

- ***Note***: no sum over a ***or*** e (want former, know latter)

Chain Elimination with Evidence



- First, we can rearrange and eliminate d:

$$P(a,e) = \sum_b \sum_c \sum_d P(a)\, P(b \mid a)\, P(c \mid b)\, P(d \mid c)\, P(e \mid d)$$

$$= \sum_b \sum_c P(a)\, P(b \mid a)\, P(c \mid b) \sum_d P(d \mid c)\, P(e \mid d)$$

$$= \sum_b \sum_c P(a)\, P(b \mid a)\, P(c \mid b)\, P(e \mid c)$$

Chain Elimination with Evidence

- Then, we do the same to eliminate c:

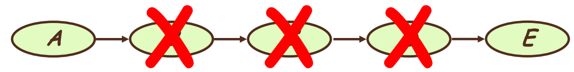
$$P(a,e) = \sum_b \sum_c P(a)\,P(b \mid a)\,P(c \mid b)\,P(e \mid c)$$

$$= \sum_b P(a)\,P(b \mid a) \sum_c P(c \mid b)\,P(e \mid c)$$

$$= \sum_b P(a)\,P(b \mid a)\,P(e \mid b)$$

29

Chain Elimination with Evidence



- Finally, eliminate b:

$$P(a,e) = \sum_b P(a)\,P(b \mid a)\,P(e \mid b)$$

$$= P(a) \sum_b P(b \mid a)\,P(e \mid b)$$

$$= P(a)\,P(e \mid a)$$

- Normalization, to get $P(e)$, will now tell us the answer, $P(a \mid e)$, since

$$P(a \mid e) = \frac{P(a,e)}{P(e)} = \frac{P(a)\,P(e \mid a)}{P(e)}$$

30