Izmir Institute of Technology

CENG 461 – Artificial Intelligence

Representing Uncertainty with Probabilities

2015-16 Fall Term

Introduction

- So far our examples were fully deterministic problems.
  - All actions at any state were known beforehand.
  - State transition results were also deterministic.
- Real life problems often involve uncertainty.
- We will represent this uncertainty by probabilities.
- In particular we will examine **Bayes Networks**, which can represent probabilistic relations of a large number of random variables in a compact manner.

Basic Probability

- Coin flips (H = Heads, T = Tails)
  - Fair Coin: P(H) = 0.5, P(T) = 0.5
  - Biased Coin: P(H) = p, P(T) = 1 – p
- Repeated Independent Coin Flips
  - P(H, H) = P(H) * P(H)
  - P(H, H, T) = P(H) * P(H) * P(T)
  - P(two heads in three flips) =

= P(H, H, T) + P(H, T, H) + P(T, H, H)

= 1 - P(H, H, H) - P(T, T, H) - P(T, H, T) - P(H, T, T) - P(T, T, T)

Joint Probability

- **P(Rain):**

  P(+Rain) and P(-Rain)

- **P(Clouds):**

  P(+Clouds) and P(-Clouds)

- **P(Rain, Clouds):**

  P(+Rain, +Clouds), P(+Rain, -Clouds),

  P(-Rain, +Clouds), P(-Rain, -Clouds)

Independence

- If the joint probability is the product of marginal probabilities (P(X) and P(Y)) then we say the variables are independent and below equation becomes true.

$$P(X,Y) = P(X) * P(Y)$$

- For independent coin flips: $P(H, H, T) = P(H) * P(H) * P(T)$
- Independent random variables do not provide information about each other:

$$P(X \mid Y) = P(X) \quad \text{and} \quad P(Y \mid X) = P(Y)$$

- E.g.

  P(Having a toothache | It is raining) = P(Having a toothache)

- We denote the independence relation as $X \perp Y$

Dependence and Conditional Probability

- Knowing the value of a random variable might affect the probability of the other, we call them *dependent*.
- We write the conditional probability of X given the value of Y as P(X | Y).
- We also know that

$$\Sigma_X P(X \mid Y) = 1$$
$$\Sigma_Y P(Y \mid X) = 1$$

This also tells $P(-X \mid Y) = 1 - P(X \mid Y)$

Exercise: Foreign Language

- A foreign language containing only the letters {a, b} has the following properties:
  - For the first letter of the word: $P(L_1 = a) = 0.4$,
  - If the previous letter is a, then the probability for the following letter is $P(L_N = a \mid L_{N-1} = a) = 0.8$
  - If the previous letter is b, then the probability for the following letter is $P(L_N = a \mid L_{N-1} = b) = 0.3$
- Compute
  - The probability of the words "aba", "aaa"
  - The probability that a three letter word will contain two 'b's
  - The probability that the second letter will be 'a'
  - The probability that the third letter will be 'a'

Joint Probability from Conditionals

- The joint distribution P(X,Y) can be written in terms of the conditionals and the marginal probabilities as

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

- E.g.:

P(Having a toothache | Having a cavity)=0.8

P(Having a cavity)=0.3

P(Having a toothache, Having a cavity)= 0.24

Marginalization and Total Probability

- We can calculate the marginal probabilities from the joint distribution by summation:

$$P(Y) = \Sigma_X P(Y, X)$$

- This can also be written as the total probability of all possible outcomes of one of the variables

$$P(Y) = \Sigma_i P(Y \mid X = i) * P(X = i)$$

Exercise

- Consider that different coin flips are dependent

$P(X_1 = H) = 0.5$, then $P(X_1=T) = 0.5$

$P(X_2 = H \mid X_1 = H) = 0.9$ $P(X_2=T \mid X_1=H) = 0.1$

$P(X_2 = T \mid X_1 = T) = 0.8$ $P(X_2=H \mid X_1=T) = 0.2$

$P(X_2 = H) = ?$

- Solution:

$$P(X_2=H) = P(X_2=H \mid X_1=H) \cdot P(X_1=H) + P(X_2=H \mid X_1=T) \cdot P(X_1=T)$$

$$= 0.9 * 0.5 + 0.2 * 0.5 = 0.55$$

Bayes' Rule

- Remember the joint probability

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

- We can derive the Bayes' Rule from the above equality

$$\underbrace{P(X \mid Y)}_{\text{posterior}} = \frac{\overbrace{P(Y \mid X)}^{\text{likelihood}} * \overbrace{P(X)}^{\text{prior}}}{\underbrace{P(Y)}_{\text{total probability}}} = \frac{P(Y \mid X) * P(X)}{\sum_i P(Y \mid X = i) * P(X = i)}$$

- This is probably the most important equation in probabilistic analysis.

Exercise

- Consider the following cancer diagnostic problem:
  - $P(+Cancer) = 0.01$
  - $P(+Test \mid +Cancer) = 0.9$
  - $P(+Test \mid -Cancer) = 0.2$

$P(-Cancer) = 0.99$

$P(-Test \mid +Cancer) = 0.1$

$P(-Test \mid -Cancer) = 0.8$

- Calculate $P(+Cancer \mid +Test)$
- Solution:

$$P(+Cancer \mid +Test) = P(+Test \mid +Cancer) \cdot P(+Cancer) \,/\, P(+Test)$$

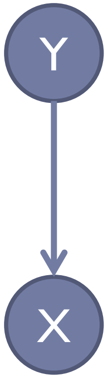
$$= \frac{P(+Test \mid +Cancer) \cdot P(+Cancer)}{P(+Test \mid +Cancer) \cdot P(+Cancer) + P(+Test \mid -Cancer) \cdot P(-Cancer)}$$

$$= \frac{0.9 * 0.01}{0.9 * 0.01 + 0.2 * 0.99} = 0.043$$

Simple Bayes Networks

- We can graphically represent the joint distribution of a set of variables by a directed graph.
- We have a node for each random variable and we have edges for each conditional probability in the factorized joint distribution:

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



Not observable

Observable

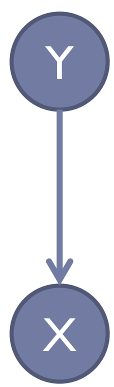
Simple Bayes Networks

- To write the joint distribution for a Bayes Network, we write a factor for each graph node conditioned on its parent nodes.
- Nodes without a parent correspond to marginal probabilities of the corresponding random variable:

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

Not observable

Observable



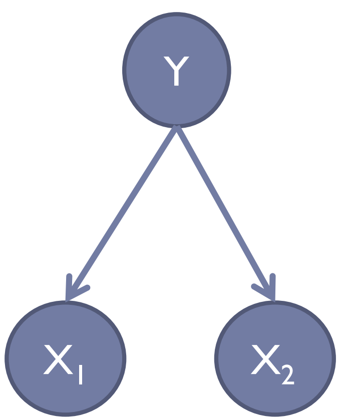
In diagnostic reasoning example:
Y is cancer, we know P(Cancer)

X is test, we know P(+Test|Cancer)

We want to infer P(Cancer|+Test)

Conditional Independence

- Sometimes knowledge of one random variable affects the independence relation between two other random variables.
- In the Bayes Network below, in general, X_1 and X_2 are not independent random variables.
- Given Y, however, X_1 and X_2 are conditionally independent of each other.



$$X_1 \perp X_2 \mid Y$$

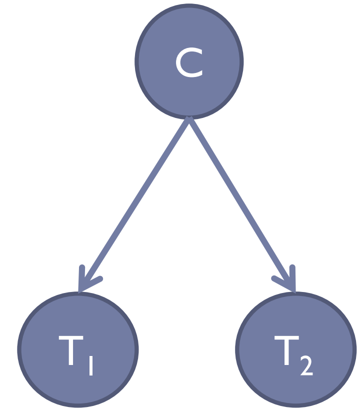
$$P(X_2 | Y, X_1) = P(X_2 | Y)$$
$$P(X_1, X_2, Y) = P(X_1 | Y) * P(X_2 | Y) * P(Y)$$

Exercise

- Consider the following diagnostic problem:
  - $P(+C) = 0.01$
  - $P(+T \mid +C) = 0.9$
  - $P(+T \mid -C) = 0.2$

C: Cancer

T: Test



- What if we apply the test two times?
- Calculate $P(+Cancer \mid +T_1, +T_2)$
- Solution:

Bayes Rule: $P(+C \mid +T_1,+T_2) = P(+T_1,+T_2 \mid +C) \cdot P(+C) \,/\, P(+T_1,+T_2)$

Use conditional indepen. : $P(+T_1,+T_2 \mid +C) = P(+T_1,+C) \cdot P(+T_2,+C)$

Bayes Rule becomes

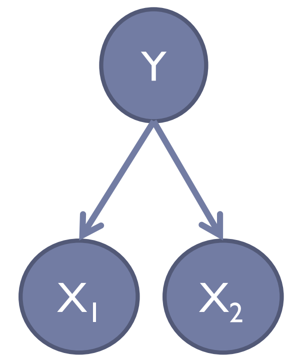
$$= \frac{P(+T_1|+C) \cdot P(+T_2|+C) \cdot P(+C)}{P(+T_1|+C) \cdot P(+T_2|+C) \cdot P(+C) + P(+T_1|-C) \cdot P(+T_2|-C) \cdot P(-C)}$$

$= 0.9*0.9*0.01 \,/\, 0.9*0.9*0.01 + 0.2*0.2*0.99 = 0.1698$

Conditional Independence

- We have seen that given Y, X_1 and X_2 are independent.
- Does that mean X_1 and X_2 are independent even if we do not know Y?

$$\cancel{X_1 \perp X_2}$$

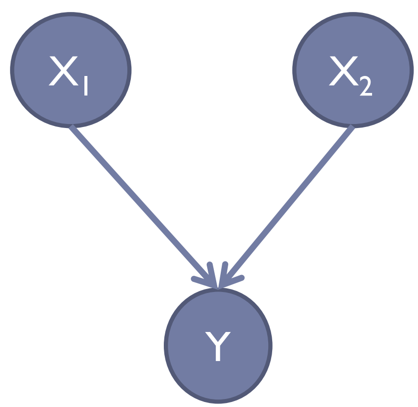


- In cancer diagnosis, intuitively, getting a positive test result increases the prob. of having cancer, and this raises the probability of getting a positive from a second test.
- As an exercise calculate $P(+T_2 \mid +T_1)$

  Hint: Since they are not independent,
  $=P(+T_2 \mid +T_1,+C) \cdot P(+C \mid +T_1) + P(+T_2 \mid +T_1,-C) \cdot P(-C \mid +T_1)$
  $=P(+T_2 \mid +C) \cdot P(+C \mid +T_1) + P(+T_2 \mid -C) \cdot P(-C \mid +T_1)$

Absolute versus Conditional Independence

- We have seen that conditional independence does not imply independence.
- How about the other way around?
- If two random variables are independent, are they always conditionally independent given a third random variable?

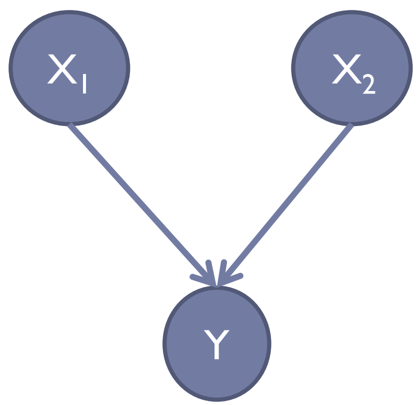


$$X_1 \perp X_2$$

$$X_1 \,?\, X_2 \mid Y$$

Absolute versus Conditional Independence

- In general independence does not imply conditional independence.
- For the Bayes Network below, given Y, X_1 and X_2 are NOT conditionally independent.



$$X_1 \perp X_2$$

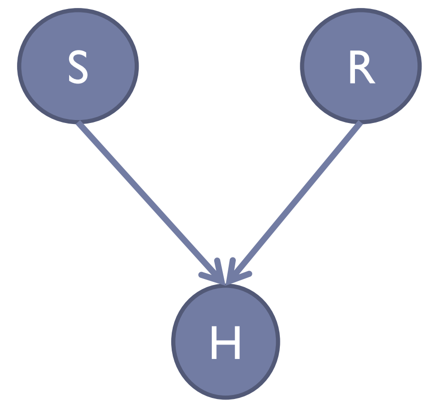
$$X_1 \not\perp X_2 \mid Y$$

$$P(X_1, X_2, Y) = P(Y \mid X_1, X_2) * P(X_1) * P(X_2)$$

Exercise

- Consider the Bayes Network on the right, with H representing the event "Happy", S representing "Sunny weather", and R representing "Raise in salary".
- With the following information
  - $P(+S) = 0.7$
  - $P(+R) = 0.01$
  - $P(+H \mid +S, +R) = 1$
  - $P(+H \mid -S, +R) = 0.9$
  - $P(+H \mid +S, -R) = 0.7$
  - $P(+H \mid -S, -R) = 0.1$
- Calculate $P(+R \mid +S)$

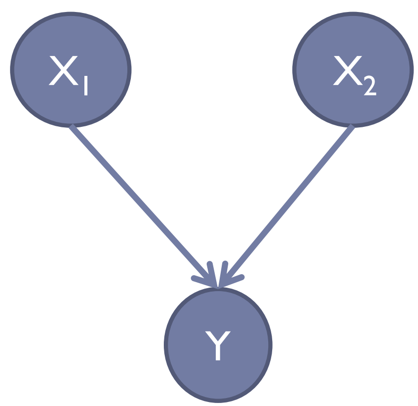
  $= 0.01$



Explaining Away

- Once you know the value of Y, the probabilities of X_1 and X_2 change to explain the value of Y.
- If you also learn the value of X_1, this might already explain the Y observation. Then, X_2 will change since the value of Y is explained away by X_1.

E.g. If happiness is explained away by sunny weather, changes of getting raise decreases.



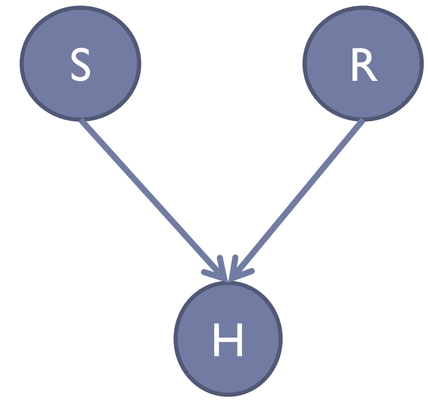
$$X_1 \perp X_2$$

$$X_1 \not\perp X_2 \mid Y$$

$$P(X_1, X_2, Y) = P(Y \mid X_1, X_2) * P(X_1) * P(X_2)$$

Exercise: Explain away

- Consider the Bayes Network on the right, with the following information
  - $P(+S) = 0.7$ $P(+R) = 0.01$
  - $P(+H \mid +S, +R) = 1$
  - $P(+H \mid -S, +R) = 0.9$
  - $P(+H \mid +S, -R) = 0.7$
  - $P(+H \mid -S, -R) = 0.1$



- Calculate $P(+R \mid +H, +S)$

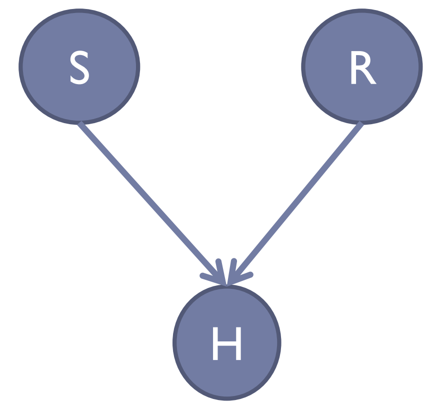
Bayes rule: $P(+R|+H,+S) = P(+H|+R,+S) \cdot P(+R|+S) / P(+H|+S)$

$$= \frac{P(+H|+R,+S) \cdot P(+R)}{P(+H|+R,+S) \cdot P(+R) + P(+H|-R,+S) \cdot P(-R)}$$

$= 0.0142$

Exercise: Explain away

- Same exercise but this time we do not know about the weather.
  - $P(+S) = 0.7$ $P(+R) = 0.01$
  - $P(+H \mid +S, +R) = 1$
  - $P(+H \mid -S, +R) = 0.9$
  - $P(+H \mid +S, -R) = 0.7$
  - $P(+H \mid -S, -R) = 0.1$



- Calculate $P(+R \mid +H)$

  Bayes rule: $P(+R|+H)= P(+H|+R) \cdot P(+R)/ P(+H)$.

  It may or may not be sunny. Long story short:

$$= \frac{P(+H,+R,+S) + P(+H,+R,-S)}{P(+H,+R,+S) + P(+H,+R,-S) + P(+H,-R,+S) + P(+H,-R,-S)}$$

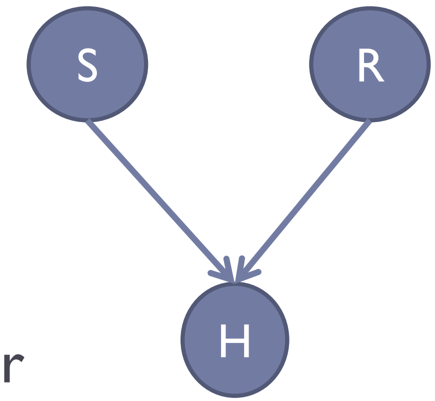
$= 0.0185$

Remember joint prob. $P(+H,+R,+S) = P(+H|+R,+S) \cdot P(+R) \cdot P(+S)$

Exercise: Explain away

- $P(+R) = 0.01$
- $P(+R \mid +S) = P(+R) = 0.01$
- $P(+R \mid +H) = 0.0185$
- $P(+R \mid +H, +S) = 0.0142$
  - Happiness explained away by sunny weather
  - S and R are not conditionally independent, i.e. given H, they become dependent.
- $P(+R \mid +H, -S) = ?$

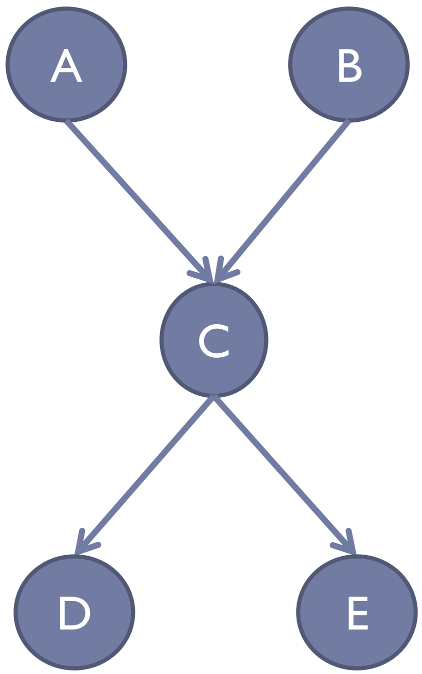
  Do at home.



Bayes Networks

- The joint distribution for the network on the right is

$$
\begin{aligned}
P(A, B, C, D, E) = \ & P(A) * P(B) \\
& * \ P(C|A, B) \\
& * P(D|C) * P(E|C)
\end{aligned}
$$



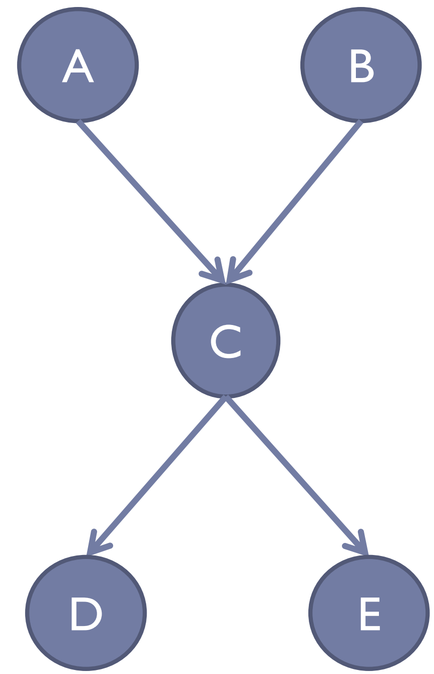
Parameter Counts

- Bayesian networks are more compact than full joint distributions.
- How many parameters (probability values) are required to specify the full joint distribution of 5 binary random variables?

  31

- How many parameters are needed to specify the joint distribution of the Bayes network on the right?

  10



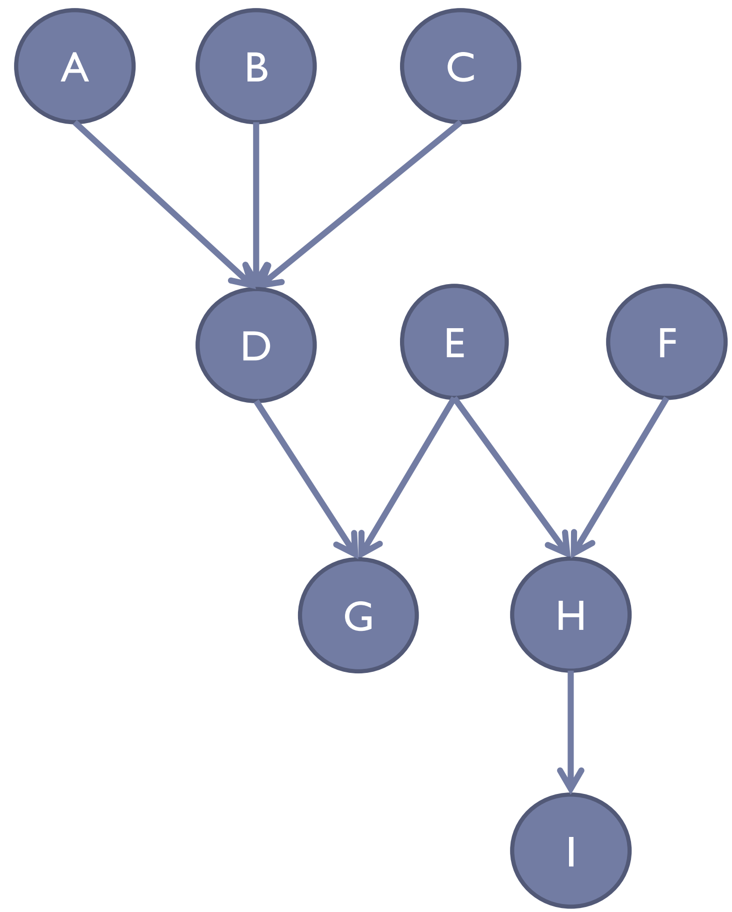
Example

- How many parameters are needed to represent the complete joint distribution of 9 binary random variables?

  → 511

- How many parameters are needed to represent the network on the right?

  → 23



D-Separation

- The concept of D-separation (D stands for dependence) help us to find out if two random variables are dependent or not.
- For the graph on the right,

  is $D \perp E$?

  is $D \perp E \mid C$?

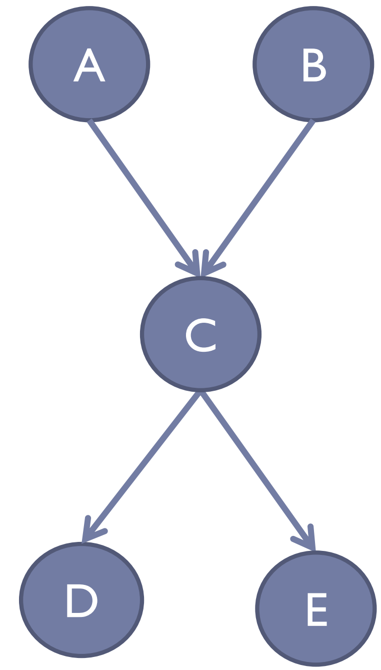
  is $A \perp E$?

  is $A \perp E \mid B$?

  is $A \perp E \mid C$?

  is $A \perp B$?

  is $A \perp B \mid C$?



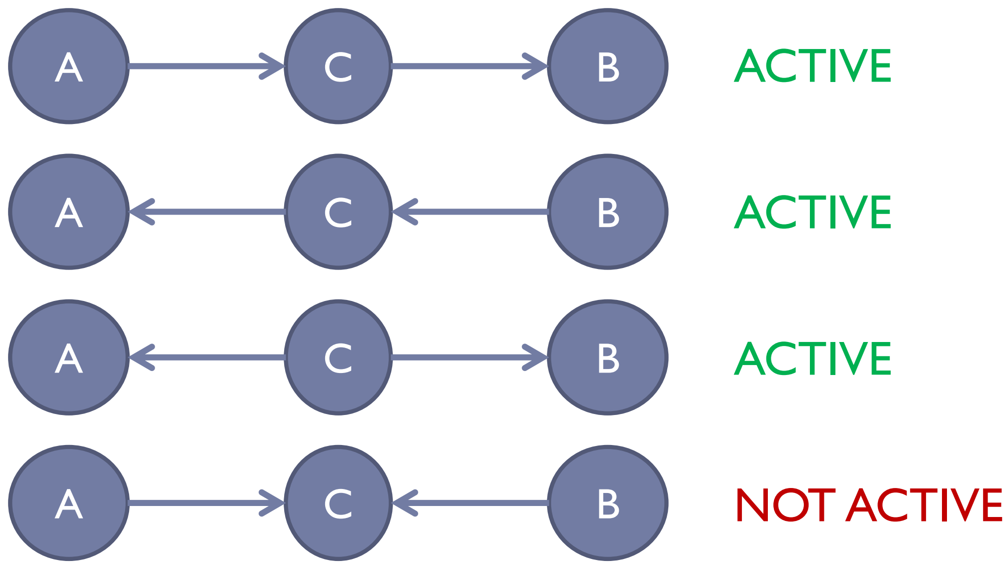
D-Separation

Definitions:

- Given a graph, two random variables are D-connected if any path between them is active.
- An active path carries information, so it causes the connected random variables to be dependent.
- A path is active if every node on it is active.
- We need to decide when a node is active, the rest follows from the previous definitions.

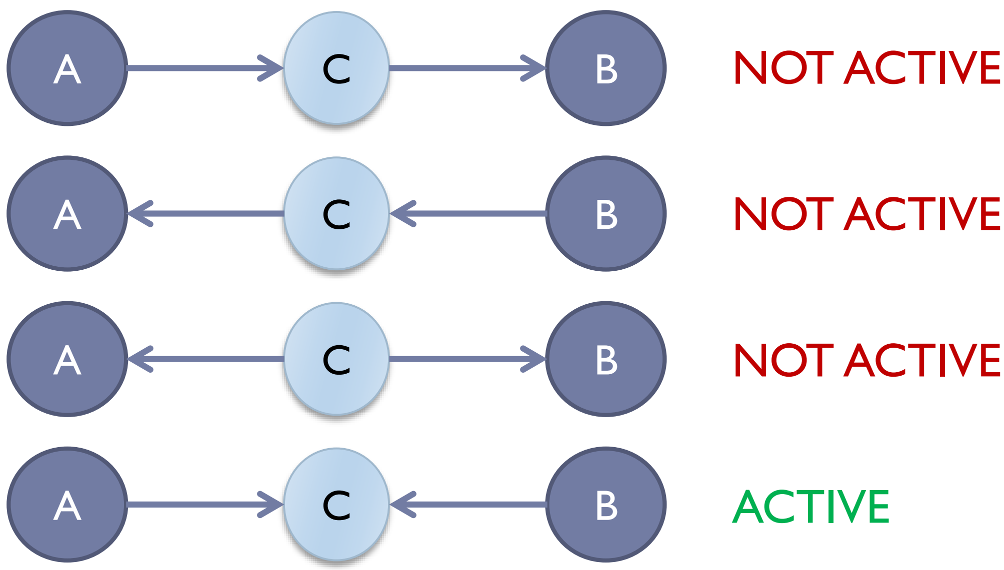
D-Separation

- When none of the nodes are in the conditioned set, i.e. when C is not given, all but the last paths are active.



D-Separation

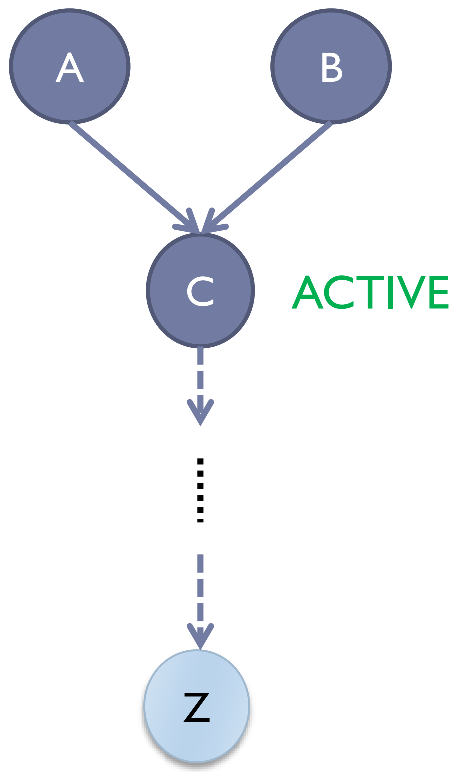
- If node C is in the conditioned set, the path is active only in the last case:



http://www.andrew.cmu.edu/user/scheines/tutor/d-sep.html

D-Separation

- Lastly, if a descendant of node C is in the conditioned set, (e.g. node Z is given here), the node C is active also active as a path element between A and B:



D-Separation

- For the graph on the right

is $C \perp A$? no

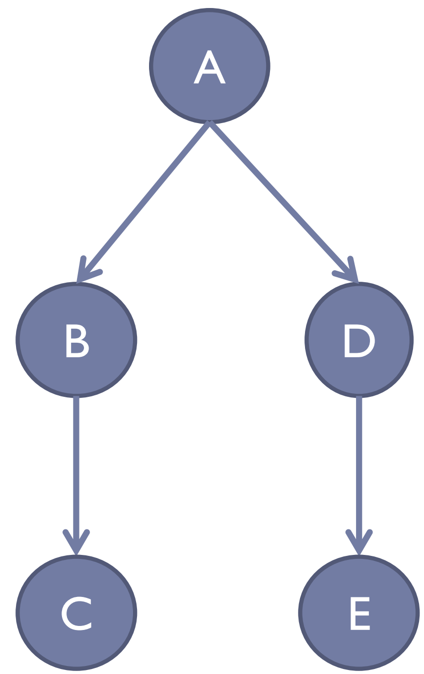
is $C \perp A \mid B$? yes

is $B \perp D$? no

is $C \perp D$? no

is $C \perp D \mid A$? yes

is $C \perp E \mid D$? yes



Exercise

1. $F \perp A$ yes
2. $F \perp A \mid D$ no
3. $F \perp A \mid G$ no
4. $F \perp A \mid H$ yes