Belief Propagation in Bayesian Networks

Céline Comte

NOKIA Bell Labs

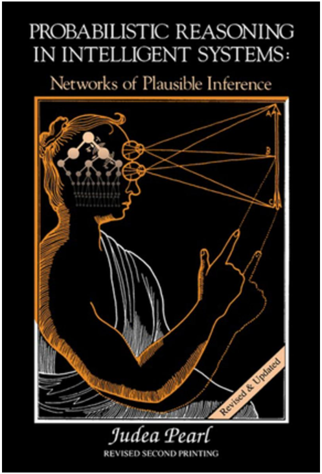
TELECOM ParisTech

Reading Group "Network Theory"
November 5, 2018

Introduction

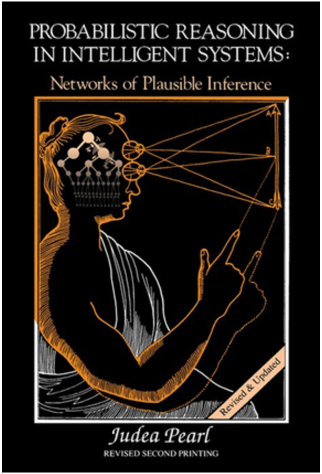
(First-order) logic
Represent causal relations between variables by a directed acyclic graph

Probabilities
Weight these causal relations by probabilities that implicitly account for non-represented variables



Introduction

"Belief networks are directed acyclic graphs in which the nodes represent propositions (or variables), the arcs signify direct dependencies between the linked propositions, and the strengths of these dependencies are quantified by conditional probabilities" (Pearl, 1986)



Bayesian networks are also ...

- A memory-efficient way of storing a PMF
- Based on simple probability rules
  (more details in a few slides)
- Inspired by human causal reasoning (Pearl, 1986, 1988)
- Used for decision taking if a utility function is provided
- Applied in many fields: medecine diagnoses, turbo-codes, (programming) language detection, ...
- Related to other models: Markov random fields, Markov chains, hidden Markov models, ...

References: Pearl's articles and book

- J. Pearl (1982). "Reverend Bayes on Inference Engines: A Distributed Hierarchical Approach". In: AAAI'82
  - → Belief propagation in causal trees
- J. Pearl (1986). "Fusion, propagation, and structuring in belief networks". In: Artificial Intelligence
  - → Belief propagation in causal trees and polytrees
- J. Pearl (1988). Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference. Morgan Kaufmann
  - → A complete reference
    (thanks Achille for providing me with this book)

References: Textbooks

- T. D. Nielsen and F. V. Jensen (2007). Bayesian Networks and Decision Graphs. Springer-Verlag
  → A lot of examples in Chapters 2 and 3
- D. Koller, N. Friedman, and F. Bach (2009). Probabilistic Graphical Models: Principles and Techniques. MIT Press
- M. Jordan (Last modified in 2015). An Introduction to Probabilistic Graphical Models.
  → Definition and belief probagation
  (thanks Nathan for pointing this reference)

Outline

Reminders on probability theory

Bayesian networks

Belief propagation in trees

Belief propagation in polytrees

Outline

Reminders on probability theory

Bayesian networks

Belief propagation in trees

Belief propagation in polytrees

Independence and conditional independence

Remark: We work exclusively with discrete random variables

Independence and conditional independence

Remark: We work exclusively with discrete random variables

- A and B are **marginally independent** (written $A \perp\!\!\!\perp B$) if one of these three equivalent conditions is satisfied:
  - $P(A, B) = P(A)P(B)$
  - $P(A \mid B) = P(A)$
  - $P(B \mid A) = P(B)$

Independence and conditional independence

Remark: We work exclusively with discrete random variables

- A and B are **marginally independent** (written $A \perp\!\!\!\perp B$) if one of these three equivalent conditions is satisfied:
  - $P(A, B) = P(A)P(B)$
  - $P(A \mid B) = P(A)$
  - $P(B \mid A) = P(B)$

- A and B are **conditonally independent** given C (written $A \perp\!\!\!\perp B \mid C$) if one of these three equivalent conditions is satisfied:
  - $P(A, B \mid C) = P(A \mid C)P(B \mid C)$
  - $P(A \mid B, C) = P(A \mid C)$
  - $P(B \mid A, C) = P(B \mid C)$

Useful rules

- **The chain rule of probabilities**
  If $A_1, \dots, A_n$ are random variables, we have

$$P(A_1, \dots, A_n) = P(A_1) \times P(A_2 \mid A_1) \times P(A_3 \mid A_1, A_2) \times \dots \times P(A_n \mid A_1, \dots, A_{n-1})$$

Useful rules

- **The chain rule of probabilities**
  If $A_1, \ldots, A_n$ are random variables, we have

$$P(A_1, \ldots, A_n) = P(A_1) \times P(A_2 \mid A_1) \times P(A_3 \mid A_1, A_2) \times \cdots \times P(A_n \mid A_1, \ldots, A_{n-1})$$

- **Law of total probability**
  If A and B are two random variables,

$$P(B) = \sum_A P(B \mid A) P(A)$$

Useful rules

- **Bayes' rule**
  If A and B are two random variables,

$$P(B \mid A) = \frac{P(A \mid B)P(B)}{P(A)}$$

We can see $P(A)$ as a **normalizing constant**: we can first compute $P(B \mid A) \propto P(A \mid B)P(B)$ for each value of B and then normalize to obtain $P(B \mid A)$ without computing $P(A)$

Glossary

- **Belief** in a random variable (**conviction** in french)
  Marginal distribution of this random variable
  (given the value of some observed variables)

Glossary

- **Belief** in a random variable (**conviction** in french)
  Marginal distribution of this random variable
  (given the value of some observed variables)
- **Observe** a random variable

Glossary

- **Belief** in a random variable (**conviction** in french)
  Marginal distribution of this random variable
  (given the value of some observed variables)
- **Observe** a random variable
- **Evidence** (piece of evidence)
  The set of random variables that have been observed

Outline

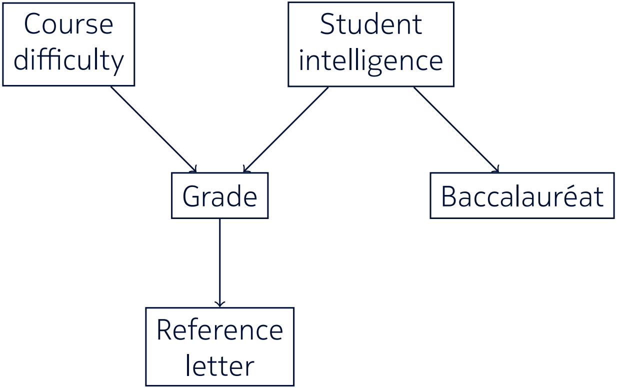
Reminders on probability theory

Bayesian networks

Belief propagation in trees

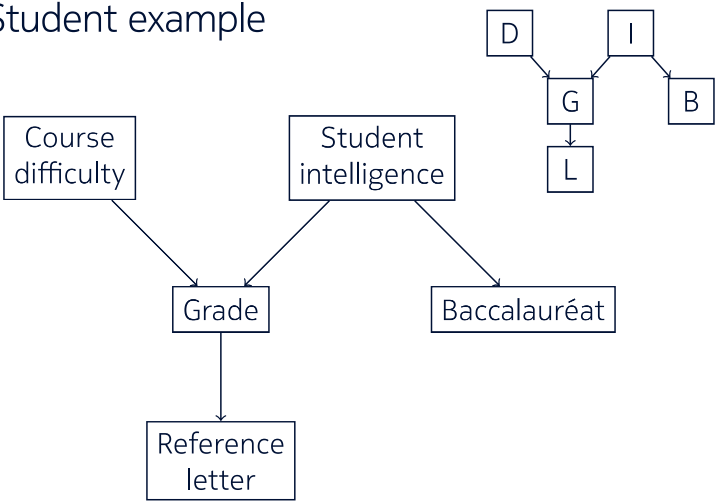
Belief propagation in polytrees

The Student example



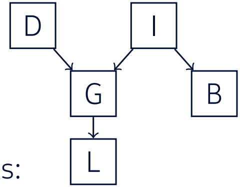
Borrowed from (Koller, Friedman, and Bach, 2009)

The Student example



Borrowed from (Koller, Friedman, and Bach, 2009)

The Student example

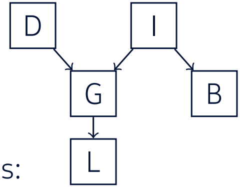


- **Local Markov property**
  Each node is conditionally independent of its non-descendants given its parents:

$$D \perp\!\!\!\perp \{I, B\}, \quad I \perp\!\!\!\perp D, \quad G \perp\!\!\!\perp B \mid \{D, I\},$$

$$B \perp\!\!\!\perp \{D, G, L\} \mid I, \quad L \perp\!\!\!\perp \{D, I, B\} \mid G$$

The Student example



- **Local Markov property**
  Each node is conditionally independent of its non-descendants given its parents:

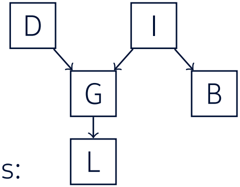
$$D \perp\!\!\!\perp \{I, B\}, \quad I \perp\!\!\!\perp D, \quad G \perp\!\!\!\perp B \mid \{D, I\},$$

$$B \perp\!\!\!\perp \{D, G, L\} \mid I, \quad L \perp\!\!\!\perp \{D, I, B\} \mid G$$

- **Chain rule of Bayesian networks**
  By the chain rule of probabilities:

$$\begin{aligned} P(D, I, G, B, L) &= P(D)P(I \mid D)P(G \mid D, I)P(B \mid D, I, G)P(L \mid D, I, G, B), \\ &= P(D)P(I)P(G \mid D, I)P(B \mid I)P(L \mid G) \end{aligned}$$

The Student example



- **Local Markov property**
  Each node is conditionally independent of its non-descendants given its parents:

$$D \perp\!\!\!\perp \{I, B\}, \quad I \perp\!\!\!\perp D, \quad G \perp\!\!\!\perp B \mid \{D, I\},$$

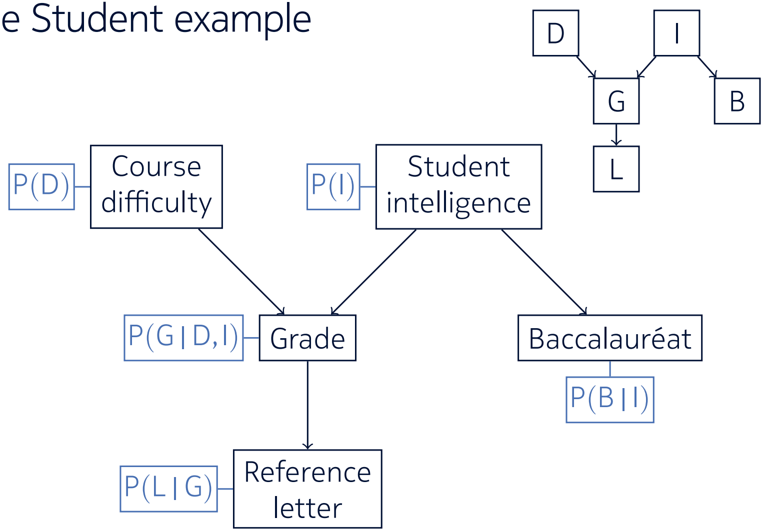
$$B \perp\!\!\!\perp \{D, G, L\} \mid I, \quad L \perp\!\!\!\perp \{D, I, B\} \mid G$$

- **Chain rule of Bayesian networks**
  By the chain rule of probabilities:

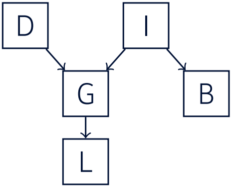
$$\begin{aligned} P(D, I, G, B, L) &= P(D)P(I \mid D)P(G \mid D, I)P(B \mid D, I, G)P(L \mid D, I, G, B), \\ &= P(D)P(I)P(G \mid D, I)P(B \mid I)P(L \mid G) \end{aligned}$$

These two definitions are equivalent

The Student example



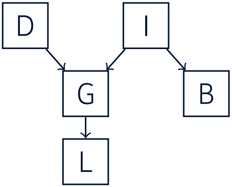
Bayesian networks in general



Described by

- A **directed acyclic graph**
  - Nodes ~ (discrete) random variables $X_1,\dots,X_n$
  - Arrows ~ conditional (in)dependencies
- Local **conditional probability tables (CPT)**
  - $P(X_i \mid parents(X_i))$ for each node X_i

Bayesian networks in general



Two equivalent definitions

- **Local Markov property**
  Each node is conditionally independent of its non-descendants given its parents
- **Chain rule of Bayesian networks**

$$P(X_1, \ldots, X_n) = \prod_{i=1}^{n} P(X_i \mid \text{parents}(X_i))$$

Proof of the equivalence: Corollary 4 p.20 of (Pearl, 1988)

Base case: serial connection

$$X \perp\!\!\!\perp Z \mid Y \qquad P(X,Y,Z) = P(X)P(Y \mid X)P(Z \mid Y)$$

Base case: serial connection

$X \perp\!\!\!\perp Z \mid Y \qquad P(X,Y,Z) = P(X)P(Y \mid X)P(Z \mid Y)$

- Interpretation: chain of causality
  X “causes” Y that “causes” Z

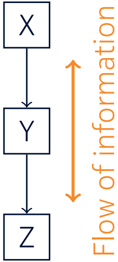
Base case: serial connection

$$\begin{tikzcd} \boxed{X} \arrow[d] \\ \boxed{Y} \arrow[d] \\ \boxed{Z} \end{tikzcd}$$

$$X \perp\!\!\!\perp Z \mid Y \qquad P(X,Y,Z) = P(X)P(Y \mid X)P(Z \mid Y)$$

- Interpretation: chain of causality
  X “causes” Y that “causes” Z
- Information can flow between X and Z through Y (that is, observing X changes our belief in Z and vice versa), unless Y is observed

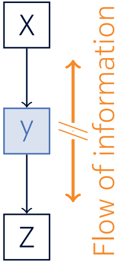
Base case: serial connection



$X \perp\!\!\!\perp Z \mid Y \qquad P(X,Y,Z) = P(X)P(Y \mid X)P(Z \mid Y)$

- Interpretation: chain of causality
  X “causes” Y that “causes” Z
- Information can flow between X and Z through Y (that is, observing X changes our belief in Z and vice versa), unless Y is observed

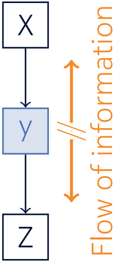
Base case: serial connection



$X \perp\!\!\!\perp Z \mid Y \qquad P(X,Y,Z) = P(X)P(Y \mid X)P(Z \mid Y)$

- Interpretation: chain of causality
  X "causes" Y that "causes" Z
- Information can flow between X and Z through Y (that is, observing X changes our belief in Z and vice versa), unless Y is observed

Base case: serial connection

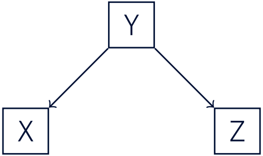


$X \perp\!\!\!\perp Z \mid Y$ $\qquad P(X,Y,Z) = P(X)P(Y \mid X)P(Z \mid Y)$

- Interpretation: chain of causality
  X "causes" Y that "causes" Z
- Information can flow between X and Z through Y (that is, observing X changes our belief in Z and vice versa), unless Y is observed
- Example: Markov chains

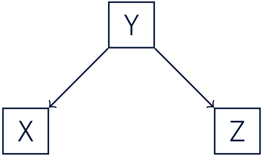
  Public

Base case: diverging connection



$X \perp\!\!\!\perp Z \mid Y \qquad P(X,Y,Z) = P(X \mid Y)P(Y)P(Z \mid Y)$

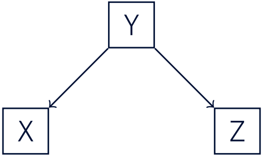
Base case: diverging connection



$$X \perp\!\!\!\perp Z \mid Y \qquad P(X,Y,Z) = P(X \mid Y)P(Y)P(Z \mid Y)$$

- Interpretation: a single root cause Y with two observable consequences X and Z

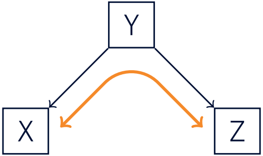
Base case: diverging connection



$$X \perp\!\!\!\perp Z \mid Y \qquad P(X,Y,Z) = P(X \mid Y)P(Y)P(Z \mid Y)$$

- Interpretation: a single root cause Y with two observable consequences X and Z
- Information can flow between X and Z through Y, unless Y is observed

Base case: diverging connection

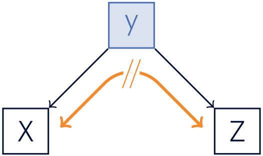


$X \perp\!\!\!\perp Z \mid Y \qquad P(X,Y,Z) = P(X \mid Y)P(Y)P(Z \mid Y)$

- Interpretation: a single root cause Y with two observable consequences X and Z
- Information can flow between X and Z through Y, unless Y is observed

  Public

Base case: diverging connection



$X \perp\!\!\!\perp Z \mid Y \qquad P(X,Y,Z) = P(X \mid Y)P(Y)P(Z \mid Y)$

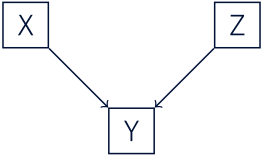
- Interpretation: a single root cause Y with two observable consequences X and Z
- Information can flow between X and Z through Y, unless Y is observed

Base case: converging connection

$$\begin{tikzcd} \boxed{X} \arrow[dr] & & \boxed{Z} \arrow[dl] \\ & \boxed{Y} & \end{tikzcd}$$

$$X \perp\!\!\!\perp Z \qquad P(X,Y,Z) = P(X)P(Y \mid X,Z)P(Z)$$

Base case: converging connection

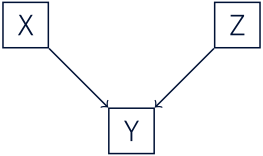


$$X \perp\!\!\!\perp Z \qquad P(X,Y,Z) = P(X)P(Y \mid X,Z)P(Z)$$

- Interpretation: two possible explanations X and Z for an observed consequence Y

  Public NOKIA Bell Labs TELECOM ParisTech

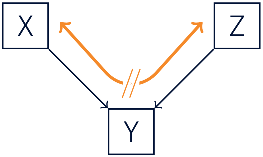
Base case: converging connection



$$X \perp\!\!\!\perp Z \qquad P(X,Y,Z) = P(X)P(Y \mid X,Z)P(Z)$$

- Interpretation: two possible explanations X and Z for an observed consequence Y
- “Explaining away” effect: information cannot flow between X and Z, unless Y is observed

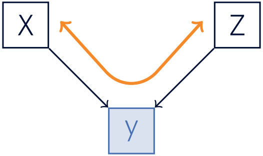
Base case: converging connection



$$X \perp\!\!\!\perp Z \qquad P(X,Y,Z) = P(X)P(Y \mid X,Z)P(Z)$$

- Interpretation: two possible explanations X and Z for an observed consequence Y
- "Explaining away" effect: information cannot flow between X and Z, unless Y is observed

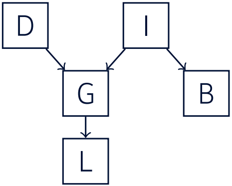
Base case: converging connection



$X \perp\!\!\!\perp Z \qquad P(X,Y,Z) = P(X)P(Y \mid X,Z)P(Z)$

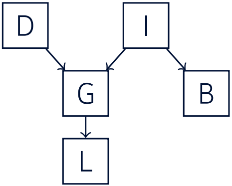
- Interpretation: two possible explanations X and Z for an observed consequence Y
- "Explaining away" effect: information cannot flow between X and Z, unless Y is observed

Implied independencies



Similar to the “Strong Markov property” of Markov chains

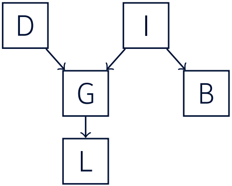
Implied independencies



Similar to the “Strong Markov property” of Markov chains

Which are correct?

Implied independencies

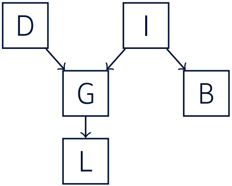


Similar to the “Strong Markov property” of Markov chains

Which are correct?

1. $G \perp\!\!\!\perp B$?

Implied independencies

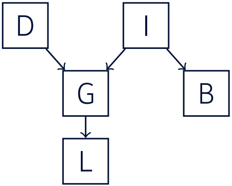


Similar to the “Strong Markov property” of Markov chains

Which are correct?

1. $G \perp\!\!\!\perp B$? No

Implied independencies

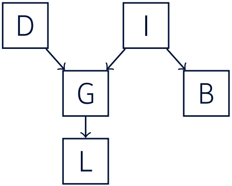


Similar to the “Strong Markov property” of Markov chains

Which are correct?

1. $G \perp\!\!\!\perp B$? No
2. $B \perp\!\!\!\perp L$?

Implied independencies

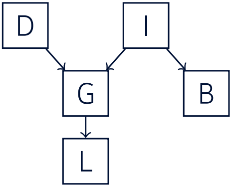


Similar to the “Strong Markov property” of Markov chains

Which are correct?

1. $G \perp\!\!\!\perp B$? No
2. $B \perp\!\!\!\perp L$? No

Implied independencies

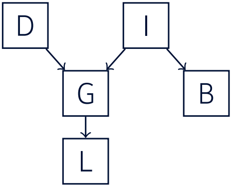


Similar to the “Strong Markov property” of Markov chains

Which are correct?

1. $G \perp\!\!\!\perp B$? No
2. $B \perp\!\!\!\perp L$? No
3. $D \perp\!\!\!\perp L$?

Implied independencies

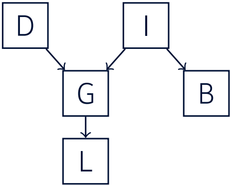


Similar to the “Strong Markov property” of Markov chains

Which are correct?

1. $G \perp\!\!\!\perp B$? No
2. $B \perp\!\!\!\perp L$? No
3. $D \perp\!\!\!\perp L$? No

Implied independencies



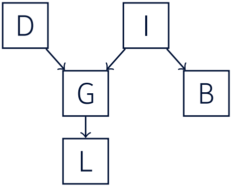
Similar to the "Strong Markov property" of Markov chains

Which are correct?

1. $G \perp\!\!\!\perp B$? No
2. $B \perp\!\!\!\perp L$? No
3. $D \perp\!\!\!\perp L$? No
4. $D \perp\!\!\!\perp B$?

  Public

Implied independencies

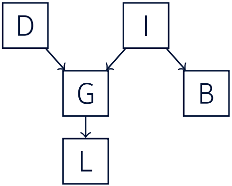


Similar to the "Strong Markov property" of Markov chains

Which are correct?

1. $G \perp\!\!\!\perp B$? No
2. $B \perp\!\!\!\perp L$? No
3. $D \perp\!\!\!\perp L$? No
4. $D \perp\!\!\!\perp B$? Yes (by the local Markov property applied to D)

Implied independencies

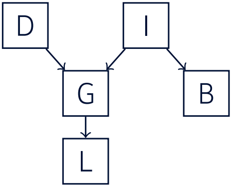


Similar to the “Strong Markov property” of Markov chains

Which are correct?

1. $G \perp\!\!\!\perp B$? No
2. $B \perp\!\!\!\perp L$? No
3. $D \perp\!\!\!\perp L$? No
4. $D \perp\!\!\!\perp B$? Yes (by the local Markov property applied to D)
5. $D \perp\!\!\!\perp B \mid G$?

Implied independencies

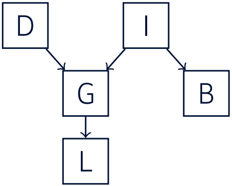


Similar to the “Strong Markov property” of Markov chains

Which are correct?

1. $G \perp\!\!\!\perp B$? No
2. $B \perp\!\!\!\perp L$? No
3. $D \perp\!\!\!\perp L$? No
4. $D \perp\!\!\!\perp B$? Yes (by the local Markov property applied to D)
5. $D \perp\!\!\!\perp B \mid G$? No (“explaining away” effect)

Implied independencies

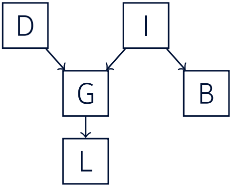


Similar to the “Strong Markov property” of Markov chains

Which are correct?

1. $G \perp\!\!\!\perp B$? No
2. $B \perp\!\!\!\perp L$? No
3. $D \perp\!\!\!\perp L$? No
4. $D \perp\!\!\!\perp B$? Yes (by the local Markov property applied to D)
5. $D \perp\!\!\!\perp B \mid G$? No (“explaining away” effect)
6. $D \perp\!\!\!\perp B \mid \{I, G\}$?

Implied independencies

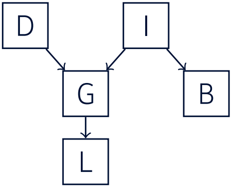


Similar to the "Strong Markov property" of Markov chains

Which are correct?

1. $G \perp\!\!\!\perp B$? No
2. $B \perp\!\!\!\perp L$? No
3. $D \perp\!\!\!\perp L$? No
4. $D \perp\!\!\!\perp B$? Yes (by the local Markov property applied to D)
5. $D \perp\!\!\!\perp B \mid G$? No ("explaining away" effect)
6. $D \perp\!\!\!\perp B \mid \{I, G\}$? Yes

Implied independencies



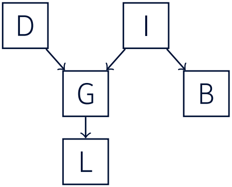
Proof of 6 $D \perp\!\!\!\perp B \mid \{I, G\}$

$$P(D, B \mid I, G)$$

  Public

NOKIA Bell Labs TELECOM ParisTech

Implied independencies

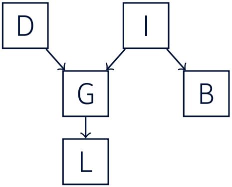


Proof of 6 $D \perp\!\!\!\perp B \mid \{I, G\}$

$$P(D, B \mid I, G)$$

$$= P(D \mid G, I) P(B \mid I, G) \qquad \square$$

Implied independencies

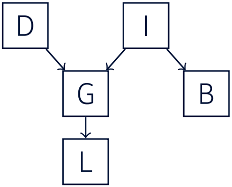


Proof of 6 $D \perp\!\!\!\perp B \mid \{I, G\}$

$$P(D, B \mid I, G) = \frac{P(G \mid D, I, B) P(D, B \mid I)}{P(G \mid I)} \quad \left(\begin{matrix}\text{Bayes'} \\ \text{rule}\end{matrix}\right)$$

$$= P(D \mid G, I) P(B \mid I, G) \qquad \square$$

Implied independencies



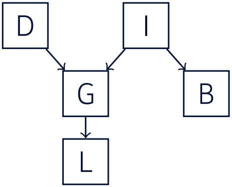
Proof of 6 $D \perp\!\!\!\perp B \mid \{I, G\}$

$$P(D,B \mid I,G) = \frac{P(G \mid D,I,B)P(D,B \mid I)}{P(G \mid I)} \quad \left(\begin{array}{c}\text{Bayes'}\\ \text{rule}\end{array}\right)$$

$$= \frac{P(G \mid D,I)P(D,B \mid I)}{P(G \mid I)} \quad \left(\begin{array}{c}\text{local Markov property}\\ \text{applied to G}\end{array}\right)$$

$$= P(D \mid G,I)P(B \mid I,G) \qquad \square$$

Implied independencies

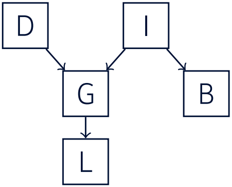


Proof of 6 $D \perp\!\!\!\perp B \mid \{I, G\}$

$$
\begin{aligned}
P(D,B \mid I,G) &= \frac{P(G \mid D,I,B)P(D,B \mid I)}{P(G \mid I)} && \begin{pmatrix}\text{Bayes'} \\ \text{rule}\end{pmatrix} \\
&= \frac{P(G \mid D,I)P(D,B \mid I)}{P(G \mid I)} && \begin{pmatrix}\text{local Markov property} \\ \text{applied to G}\end{pmatrix} \\
&= \frac{P(G \mid D,I)P(D \mid I)P(B \mid D,I)}{P(G \mid I)} && \begin{pmatrix}\text{definition of condi-} \\ \text{tional probabilities}\end{pmatrix}
\end{aligned}
$$

$$= P(D \mid G,I)P(B \mid I,G) \qquad \square$$

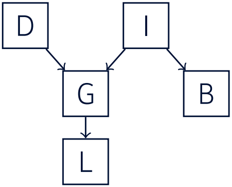
Implied independencies



Proof of (6) $D \perp\!\!\!\perp B \mid \{I, G\}$

$$
\begin{aligned}
P(D,B \mid I,G) &= \frac{P(G \mid D,I,B)P(D,B \mid I)}{P(G \mid I)} && \text{(Bayes' rule)} \\
&= \frac{P(G \mid D,I)P(D,B \mid I)}{P(G \mid I)} && \text{(local Markov property applied to G)} \\
&= \frac{P(G \mid D,I)P(D \mid I)P(B \mid D,I)}{P(G \mid I)} && \text{(definition of conditional probabilities)} \\
&= \frac{P(G \mid D,I)P(D \mid I)}{P(G \mid I)} P(B \mid D,I) \\
&= P(D \mid G,I)P(B \mid I,G) && \square
\end{aligned}
$$

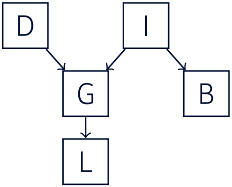
Implied independencies



Proof of 6 $D \perp\!\!\!\perp B \mid \{I, G\}$

$$
\begin{aligned}
P(D,B \mid I,G) &= \frac{P(G \mid D,I,B)P(D,B \mid I)}{P(G \mid I)} && \begin{pmatrix}\text{Bayes'}\\ \text{rule}\end{pmatrix} \\
&= \frac{P(G \mid D,I)P(D,B \mid I)}{P(G \mid I)} && \begin{pmatrix}\text{local Markov property}\\ \text{applied to G}\end{pmatrix} \\
&= \frac{P(G \mid D,I)P(D \mid I)P(B \mid D,I)}{P(G \mid I)} && \begin{pmatrix}\text{definition of condi-}\\ \text{tional probabilities}\end{pmatrix} \\
&= \frac{P(G \mid D,I)P(D \mid I)}{P(G \mid I)} P(B \mid D,I) \\
&= P(D \mid G,I)P(B \mid D,I) && \begin{pmatrix}\text{Bayes'}\\ \text{rule}\end{pmatrix} \\
&= P(D \mid G,I)P(B \mid I,G) && \square
\end{aligned}
$$

Implied independencies



Proof of 6 $D \perp\!\!\!\perp B \mid \{I, G\}$

$$
\begin{aligned}
P(D, B \mid I, G) &= \frac{P(G \mid D, I, B)P(D, B \mid I)}{P(G \mid I)} && \begin{pmatrix}\text{Bayes'}\\ \text{rule}\end{pmatrix}\\
&= \frac{P(G \mid D, I)P(D, B \mid I)}{P(G \mid I)} && \begin{pmatrix}\text{local Markov property}\\ \text{applied to G}\end{pmatrix}\\
&= \frac{P(G \mid D, I)P(D \mid I)P(B \mid D, I)}{P(G \mid I)} && \begin{pmatrix}\text{definition of condi-}\\ \text{tional probabilities}\end{pmatrix}\\
&= \frac{P(G \mid D, I)P(D \mid I)}{P(G \mid I)} P(B \mid D, I)\\
&= P(D \mid G, I)P(B \mid D, I) && \begin{pmatrix}\text{Bayes'}\\ \text{rule}\end{pmatrix}\\
&= P(D \mid G, I)P(B \mid I, G) && \begin{pmatrix}\text{local Markov property}\\ \text{applied to B}\end{pmatrix} \quad \square
\end{aligned}
$$

Memory and time complexity

Parameters

Memory and time complexity

Parameters

- n number of random variables
  (typically, $n \sim$ hundreds or thousands)

Memory and time complexity

Parameters

- n number of random variables (typically, $n \sim$ hundreds or thousands)
- r number of values each variable can take

Memory and time complexity

Parameters

- n number of random variables (typically, $n \sim$ hundreds or thousands)
- r number of values each variable can take
- $d^{\uparrow}$ maximum number of parents of a node

Memory and time complexity

Parameters

- n number of random variables (typically, $n \sim$ hundreds or thousands)
- r number of values each variable can take
- $d^{\uparrow}$ maximum number of parents of a node

Memory complexity

Memory and time complexity

Parameters

- n number of random variables (typically, $n \sim$ hundreds or thousands)
- r number of values each variable can take
- $d^{\uparrow}$ maximum number of parents of a node

Memory complexity

- If we store the probability distribution: $O(r^n)$ entries

Memory and time complexity

Parameters

- n number of random variables (typically, $n \sim$ hundreds or thousands)
- r number of values each variable can take
- $d^{\uparrow}$ maximum number of parents of a node

Memory complexity

- If we store the probability distribution: $O(r^n)$ entries
- If we store the node parents and the conditional probability tables: $O(n(r + r^{d^{\uparrow}})) = O(nr^{d^{\uparrow}})$ entries

Memory and time complexity

Parameters

- n number of random variables (typically, $n \sim$ hundreds or thousands)
- r number of values each variable can take
- $d^{\uparrow}$ maximum number of parents of a node

Memory complexity

- If we store the probability distribution: $O(r^n)$ entries
- If we store the node parents and the conditional probability tables: $O(n(r + r^{d^{\uparrow}})) = O(nr^{d^{\uparrow}})$ entries

What about the time complexity?

Inference

“A guess that you make or an opinion that you form based on the information that you have” (Cambridge dictionary)

Inference

"A guess that you make or an opinion that you form based on the information that you have" (Cambridge dictionary)
→ Bayesian networks: compute or update the belief in each variable given some evidence

Inference

"A guess that you make or an opinion that you form based on the information that you have" (Cambridge dictionary)
→ Bayesian networks: compute or update the belief in each variable given some evidence

Belief propagation, a.k.a. **sum-product message passing**: Propagate the information through the network, starting from the evidence node(s)

Inference

“A guess that you make or an opinion that you form based on the information that you have” (Cambridge dictionary)
→ Bayesian networks: compute or update the belief in each variable given some evidence

Belief propagation, a.k.a. **sum-product message passing**: Propagate the information through the network, starting from the evidence node(s)

- Each variable is a “separate processor” (a neuron?) that knows its own CPT and the messages received from its direct neighbors (Pearl, 1982)

Inference

"A guess that you make or an opinion that you form based on the information that you have" (Cambridge dictionary)
→ Bayesian networks: compute or update the belief in each variable given some evidence

Belief propagation, a.k.a. **sum-product message passing**: Propagate the information through the network, starting from the evidence node(s)

- Each variable is a "separate processor" (a neuron?) that knows its own CPT and the messages received from its direct neighbors (Pearl, 1982)
- Dynamic programming

Outline

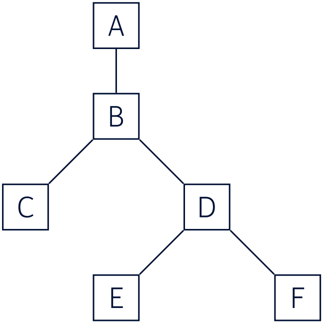
Reminders on probability theory

Bayesian networks

Belief propagation in trees

Belief propagation in polytrees

Tree Bayesian network

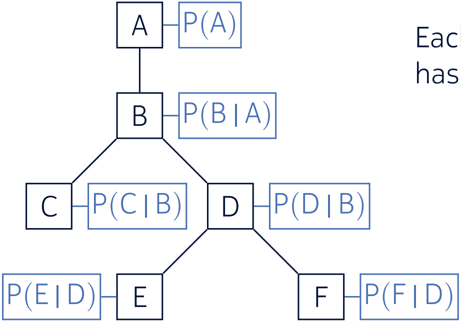


Each node (except the root) has at most one parent

NOKIA Bell Labs

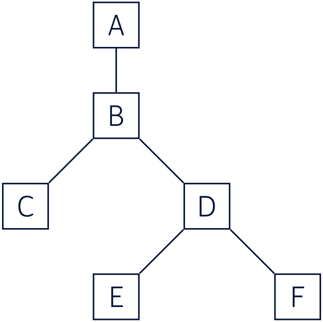
TELECOM ParisTech

Tree Bayesian network



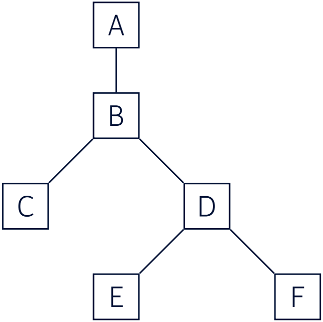
Each node (except the root) has at most one parent

Tree Bayesian network



Each node (except the root) has at most one parent

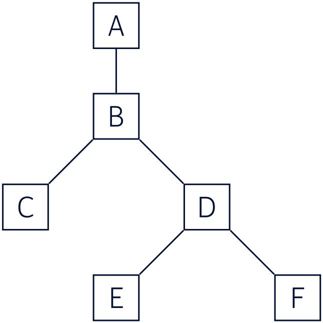
Tree Bayesian network



Each node (except the root) has at most one parent

Each node **separates** the tree: its non-descendants and the subtrees rooted at each of its children are conditionally independent given this node

Tree Bayesian network

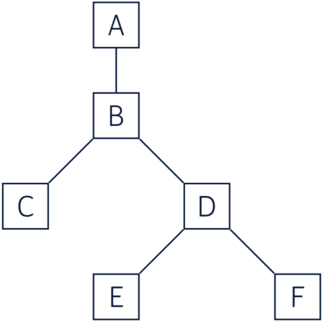


Each node (except the root) has at most one parent

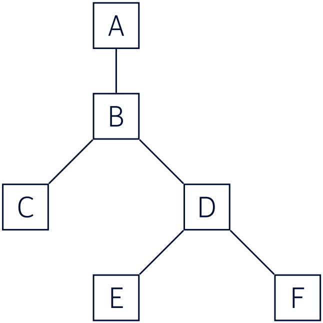
Each node **separates** the tree: its non-descendants and the subtrees rooted at each of its children are conditionally independent given this node

Remark: We will explain the propagation algorithm on this toy example borrowed from (Pearl, 1988)

No evidence

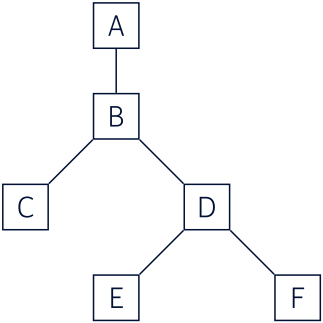


No evidence



- P(A): parameter

No evidence

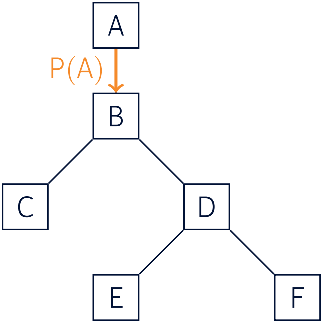


- $P(A)$: parameter
- $P(B) = \sum_{A} P(B \mid A) P(A)$

NOKIA Bell Labs

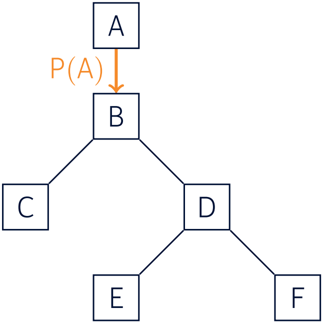
TELECOM ParisTech

No evidence



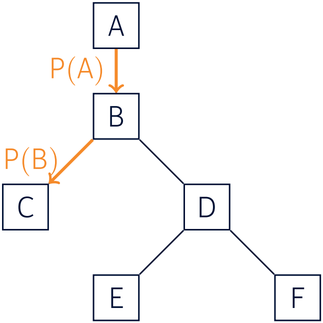
- P(A): parameter
- $P(B) = \sum_A P(B \mid A) P(A)$

No evidence



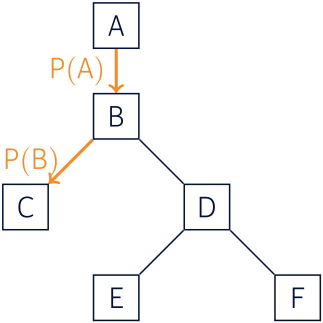
- $P(A)$: parameter
- $P(B) = \sum_{A} P(B \mid A) P(A)$
- $P(C) = \sum_{B} P(C \mid B) P(B)$

No evidence



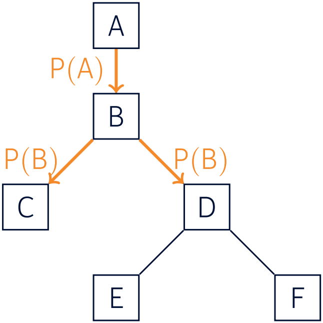
- P(A): parameter
- $P(B) = \sum_A P(B \mid A) P(A)$
- $P(C) = \sum_B P(C \mid B) P(B)$

No evidence



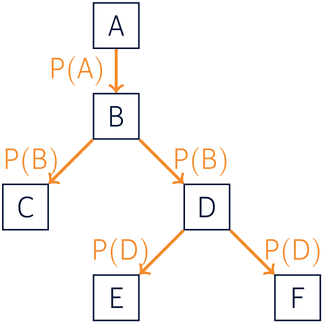
- P(A): parameter
- $P(B) = \sum_A P(B \mid A) P(A)$
- $P(C) = \sum_B P(C \mid B) P(B)$
- $P(D) = \sum_B P(D \mid B) P(B)$

No evidence



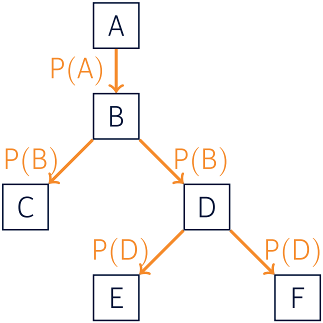
- $P(A)$: parameter
- $P(B) = \sum_{A} P(B \mid A) P(A)$
- $P(C) = \sum_{B} P(C \mid B) P(B)$
- $P(D) = \sum_{B} P(D \mid B) P(B)$

No evidence



- $P(A)$: parameter
- $P(B) = \sum_{A} P(B \mid A) P(A)$
- $P(C) = \sum_{B} P(C \mid B) P(B)$
- $P(D) = \sum_{B} P(D \mid B) P(B)$

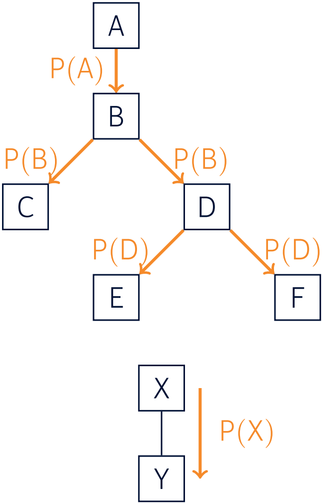
No evidence



- $P(A)$: parameter
- $P(B) = \sum_{A} P(B \mid A) P(A)$
- $P(C) = \sum_{B} P(C \mid B) P(B)$
- $P(D) = \sum_{B} P(D \mid B) P(B)$

Top-down propagation
Complexity $O(nr^2)$

No evidence



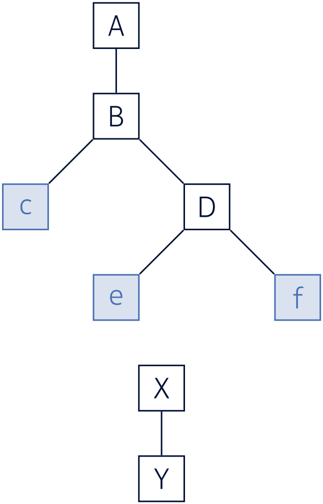
- $P(A)$: parameter
- $P(B) = \sum_A P(B \mid A) P(A)$
- $P(C) = \sum_B P(C \mid B) P(B)$
- $P(D) = \sum_B P(D \mid B) P(B)$

Top-down propagation
Complexity $O(nr^2)$

  Public

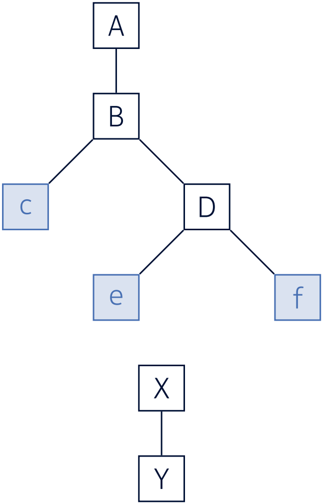
NOKIA Bell Labs TELECOM ParisTech

Three pieces of evidence



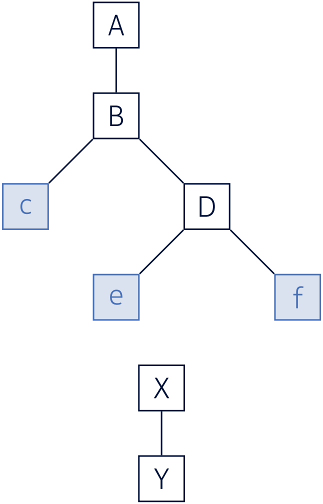
- **Evidence:** We observe that $C = c$, $E = e$, and $F = f$

Three pieces of evidence



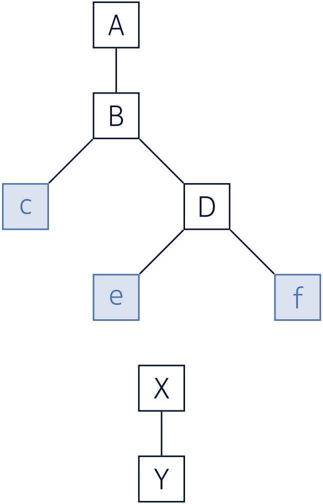
- **Evidence:** We observe that $C = c$, $E = e$, and $F = f$
- **Objective:** Compute the belief $BEL(X) = P(X \mid c, e, f)$ of each node X

Three pieces of evidence



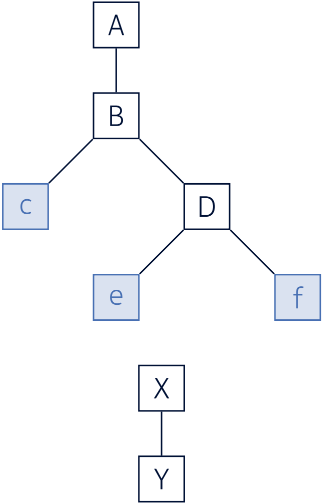
- **Evidence:** We observe that $C = c$, $E = e$, and $F = f$
- **Objective:** Compute the belief $BEL(X) = P(X \mid c, e, f)$ of each node X
- **Principle:** Propagate the information through the network, starting from the evidence nodes

Causal and diagnostic support



By Bayes' rule:

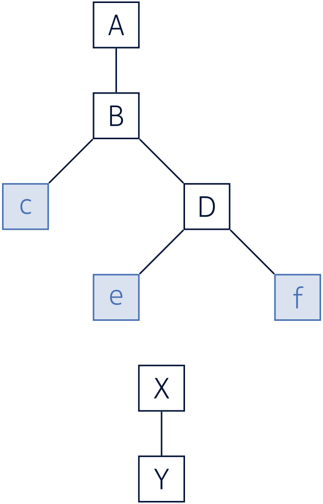
Causal and diagnostic support



By Bayes' rule:

- $P(D \mid c, e, f) = \dfrac{P(e, f \mid D, c) P(D \mid c)}{P(e, f \mid c)}$

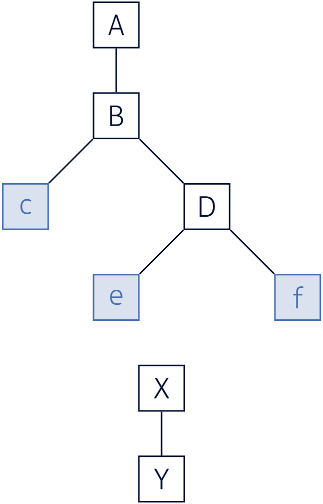
Causal and diagnostic support



By Bayes' rule:

- $$P(D \mid c, e, f) = \frac{P(e, f \mid D, c)P(D \mid c)}{P(e, f \mid c)}$$
  $$= \frac{P(e, f \mid D)P(D \mid c)}{P(e, f \mid c)}$$

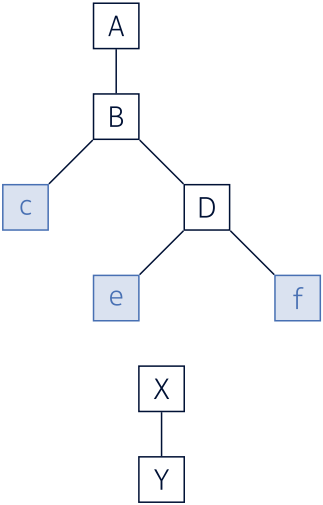
Causal and diagnostic support



By Bayes' rule:

- $$P(D \mid c,e,f) = \frac{P(e,f \mid D,c)P(D \mid c)}{P(e,f \mid c)} = \frac{P(e,f \mid D)P(D \mid c)}{P(e,f \mid c)} \propto P(e,f \mid D)P(D \mid c)$$

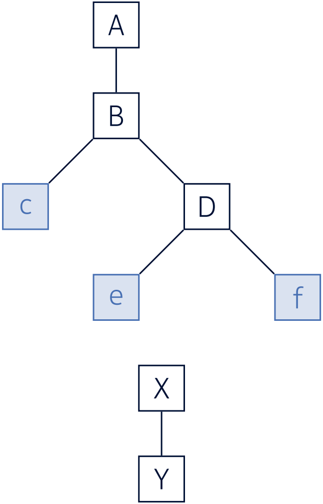
Causal and diagnostic support



By Bayes' rule:

- $$P(B \mid c,e,f) = \frac{P(c,e,f \mid B)P(B)}{P(c,e,f)}$$
- $$P(D \mid c,e,f) = \frac{P(e,f \mid D,c)P(D \mid c)}{P(e,f \mid c)} = \frac{P(e,f \mid D)P(D \mid c)}{P(e,f \mid c)} \propto P(e,f \mid D)P(D \mid c)$$

Causal and diagnostic support

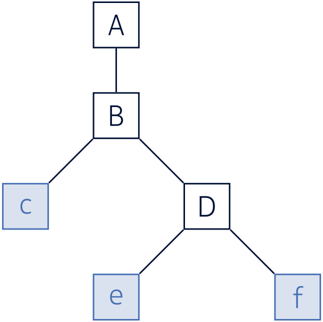


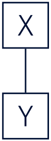
By Bayes' rule:

- $$P(B \mid c,e,f) = \frac{P(c,e,f \mid B)P(B)}{P(c,e,f)} \propto P(c,e,f \mid B)P(B)$$
- $$P(D \mid c,e,f) = \frac{P(e,f \mid D,c)P(D \mid c)}{P(e,f \mid c)} = \frac{P(e,f \mid D)P(D \mid c)}{P(e,f \mid c)} \propto P(e,f \mid D)P(D \mid c)$$

NOKIA Bell Labs TELECOM ParisTech

Causal and diagnostic support

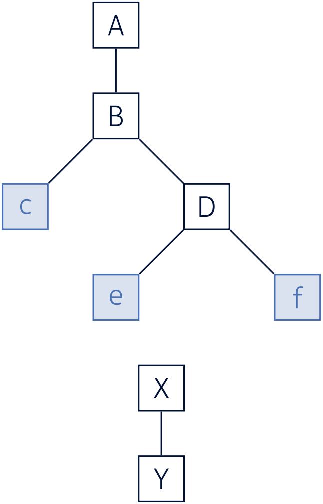




By Bayes' rule:

- $$P(A \mid c,e,f) = \frac{P(c,e,f \mid A)P(A)}{P(c,e,f)}$$
- $$P(B \mid c,e,f) = \frac{P(c,e,f \mid B)P(B)}{P(c,e,f)} \propto P(c,e,f \mid B)P(B)$$
- $$P(D \mid c,e,f) = \frac{P(e,f \mid D,c)P(D \mid c)}{P(e,f \mid c)} = \frac{P(e,f \mid D)P(D \mid c)}{P(e,f \mid c)} \propto P(e,f \mid D)P(D \mid c)$$

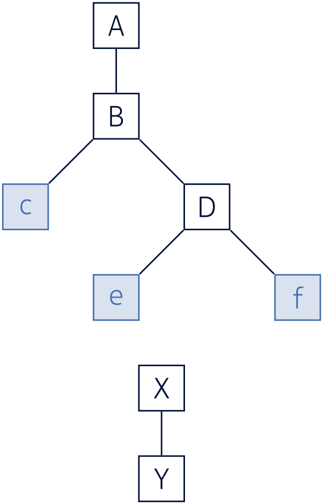
Causal and diagnostic support



By Bayes' rule:

- $$P(A \mid c,e,f) = \frac{P(c,e,f \mid A)P(A)}{P(c,e,f)} \propto P(c,e,f \mid A)P(A)$$
- $$P(B \mid c,e,f) = \frac{P(c,e,f \mid B)P(B)}{P(c,e,f)} \propto P(c,e,f \mid B)P(B)$$
- $$P(D \mid c,e,f) = \frac{P(e,f \mid D,c)P(D \mid c)}{P(e,f \mid c)} = \frac{P(e,f \mid D)P(D \mid c)}{P(e,f \mid c)} \propto P(e,f \mid D)P(D \mid c)$$

Causal and diagnostic support

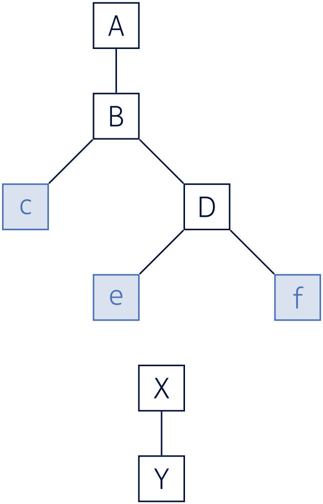


For each X, we compute

- Diagnostic support $P\left(\begin{matrix}\text{evidence}\\ \text{below X}\end{matrix} \mid X\right)$
  Bottom-up propagation
- Causal support $P\left(X \mid \begin{matrix}\text{evidence}\\ \text{above X}\end{matrix}\right)$
  Top-down propagation

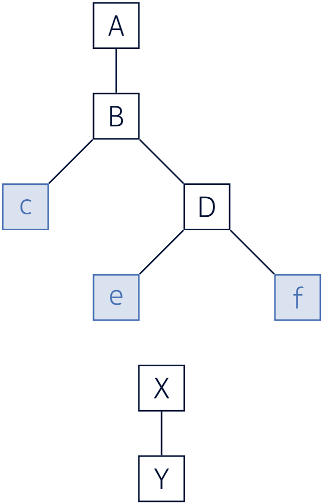
$$\mathrm{BEL}(X) \propto P\left(\begin{matrix}\text{evidence}\\ \text{below X}\end{matrix} \mid X\right) \times P\left(X \mid \begin{matrix}\text{evidence}\\ \text{above X}\end{matrix}\right)$$

Diagnostic support $P\left(\begin{matrix}\text{evidence}\\ \text{below } X\end{matrix} \mid X\right)$



Bottom-up propagation

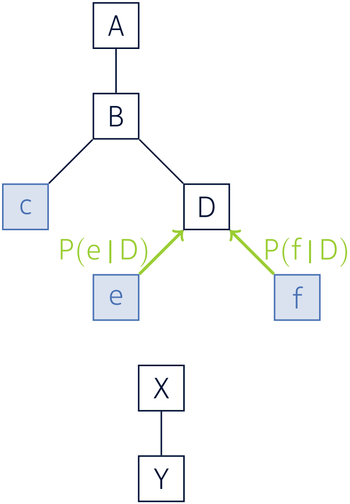
Diagnostic support $P\left(\begin{matrix}\text{evidence}\\ \text{below X}\end{matrix} \mid X\right)$



Bottom-up propagation

- $P(e, f \mid D) = P(e \mid D) P(f \mid D)$

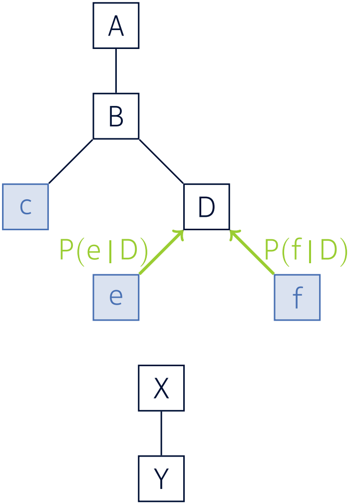
Diagnostic support $P\left(\begin{matrix}\text{evidence} \\ \text{below } X\end{matrix} \,\middle|\, X\right)$



Bottom-up propagation

- $P(e, f \mid D) = P(e \mid D) P(f \mid D)$

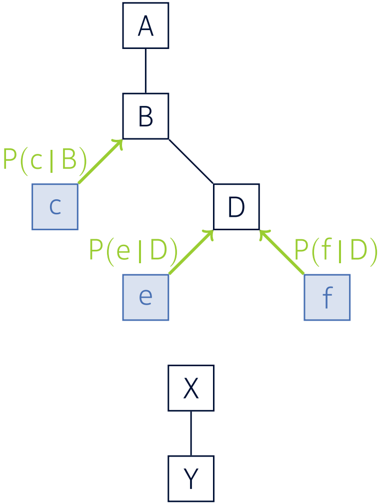
Diagnostic support $P\left(\begin{matrix}\text{evidence}\\ \text{below X}\end{matrix} \mid X\right)$



Bottom-up propagation

- $P(e, f \mid D) = P(e \mid D) P(f \mid D)$
- $P(c, e, f \mid B) = P(c \mid B) P(e, f \mid B)$

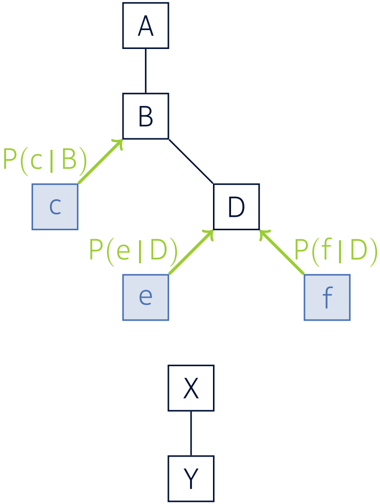
Diagnostic support $P\left(\begin{smallmatrix}\text{evidence}\\ \text{below } X\end{smallmatrix} \mid X\right)$



Bottom-up propagation

- $P(e,f \mid D) = P(e \mid D)P(f \mid D)$
- $P(c,e,f \mid B) = P(c \mid B)P(e,f \mid B)$

Diagnostic support $P\left(\begin{matrix}\text{evidence}\\ \text{below } X\end{matrix} \mid X\right)$



Bottom-up propagation

- $P(e,f \mid D) = P(e \mid D)P(f \mid D)$
- $P(c,e,f \mid B) = P(c \mid B)P(e,f \mid B)$

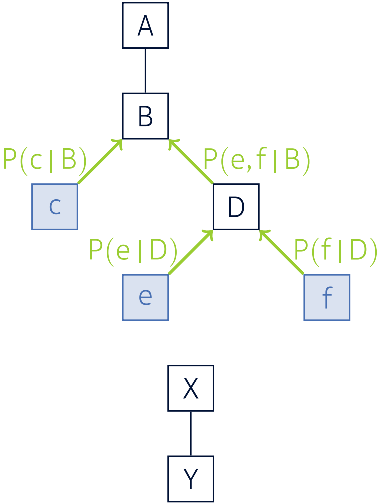
Compute $P(e,f \mid B)$:

$$P(e,f \mid B) = \sum_D P(e,f \mid B,D)P(D \mid B)$$
$$= \sum_D P(e,f \mid D)P(D \mid B)$$

  Public

NOKIA Bell Labs TELECOM ParisTech

Diagnostic support $P\left(\begin{matrix}\text{evidence}\\ \text{below } X\end{matrix} \mid X\right)$



Bottom-up propagation

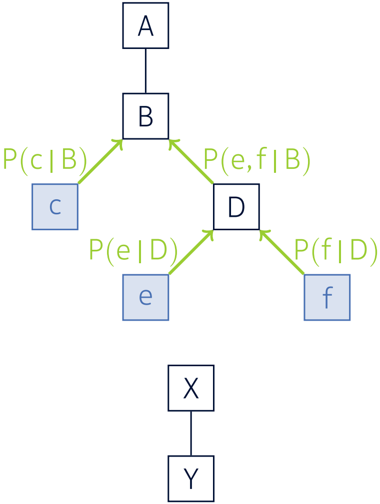
- $P(e,f \mid D) = P(e \mid D)P(f \mid D)$
- $P(c,e,f \mid B) = P(c \mid B)P(e,f \mid B)$

Compute $P(e,f \mid B)$:

$$P(e,f \mid B) = \sum_D P(e,f \mid B,D)P(D \mid B)$$
$$= \sum_D P(e,f \mid D)P(D \mid B)$$

  Public NOKIA Bell Labs TELECOM ParisTech

Diagnostic support $P\left(\begin{matrix}\text{evidence}\\ \text{below X}\end{matrix} \,\middle|\, X\right)$



- $P(c,e,f \mid A) = P(c,e,f \mid A)$

  Compute $P(c,e,f \mid A)$:

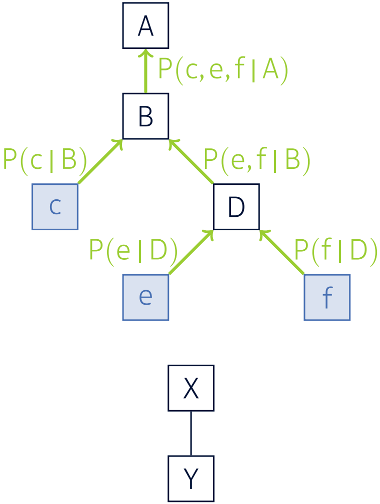
  $$\begin{aligned} &P(c,e,f \mid A) \\ &= \sum_B P(c,e,f \mid A,B)P(B \mid A) \\ &= \sum_B P(c,e,f \mid B)P(B \mid A) \end{aligned}$$

  Public

NOKIA Bell Labs

TELECOM ParisTech

Diagnostic support $P\left(\begin{matrix}\text{evidence}\\\text{below } X\end{matrix} \mid X\right)$

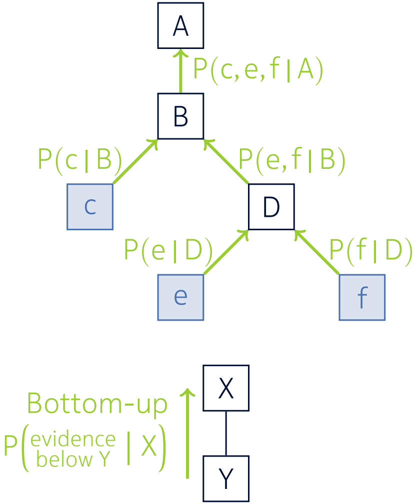


- $P(c,e,f \mid A) = P(c,e,f \mid A)$

  Compute $P(c,e,f \mid A)$:

  $$P(c,e,f \mid A) = \sum_B P(c,e,f \mid A,B)P(B \mid A) = \sum_B P(c,e,f \mid B)P(B \mid A)$$

Diagnostic support $P\left(\begin{matrix}\text{evidence}\\ \text{below } X\end{matrix} \mid X\right)$



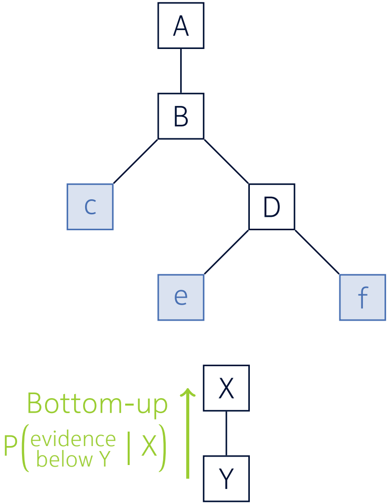
- $P(c,e,f \mid A) = P(c,e,f \mid A)$

  Compute $P(c,e,f \mid A)$:

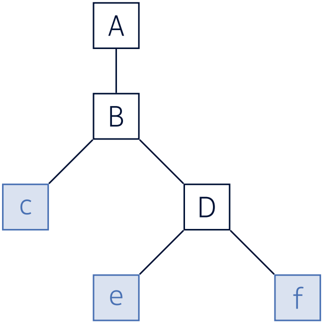
  $$P(c,e,f \mid A)$$
  $$= \sum_B P(c,e,f \mid A,B)P(B \mid A)$$
  $$= \sum_B P(c,e,f \mid B)P(B \mid A)$$

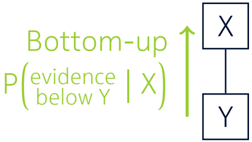
NOKIA Bell Labs
TELECOM ParisTech

Causal support $P\left(X \mid \substack{\text{evidence} \\ \text{above } X}\right)$



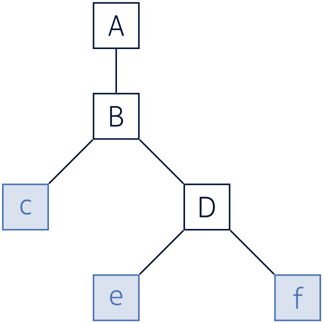
Causal support $P\left(X \mid \begin{array}{c}\text{evidence}\\ \text{above } X\end{array}\right)$





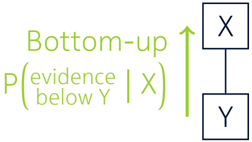
- P(A): parameter

Causal support $P\left(X \mid \begin{matrix}\text{evidence}\\ \text{above } X\end{matrix}\right)$

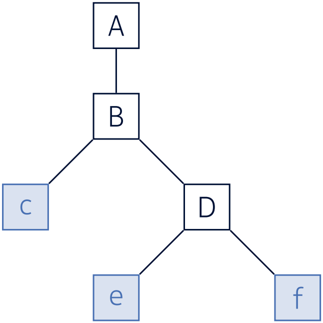


- $P(A)$: parameter

  $\rightarrow \mathrm{BEL}(A) \propto P(c,e,f \mid A)P(A)$



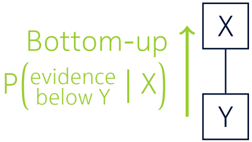
Causal support $P\left(X \mid \begin{array}{c}\text{evidence}\\ \text{above } X\end{array}\right)$



- $P(A)$: parameter

  $\rightarrow BEL(A) \propto P(c,e,f \mid A)P(A)$

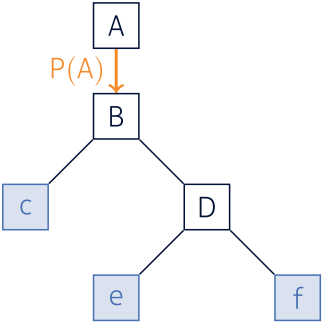
- $P(B) = \sum_A P(B \mid A)P(A)$



  Public

NOKIA Bell Labs TELECOM ParisTech

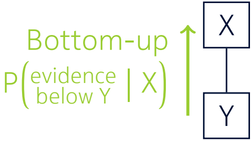
Causal support $P\left(X \mid \begin{matrix}\text{evidence}\\ \text{above } X\end{matrix}\right)$



- $P(A)$: parameter

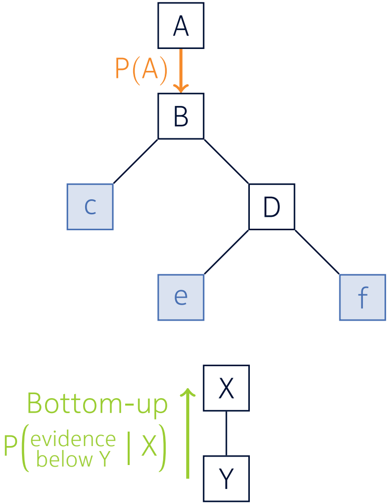
  $\rightarrow \mathrm{BEL}(A) \propto P(c, e, f \mid A)P(A)$

- $P(B) = \sum_{A} P(B \mid A)P(A)$



   NOKIA Bell Labs TELECOM ParisTech

Causal support $P\left(X \mid \begin{matrix}\text{evidence} \\ \text{above } X\end{matrix}\right)$



- $P(A)$: parameter

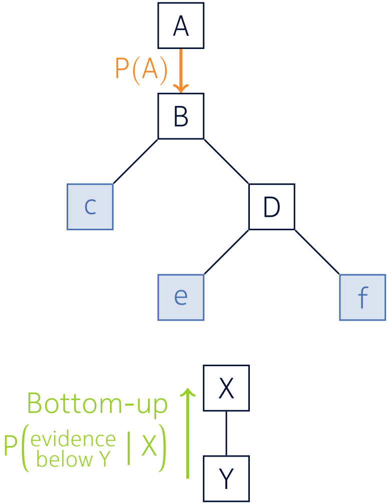
  $\rightarrow BEL(A) \propto P(c,e,f \mid A)P(A)$

- $P(B) = \sum_A P(B \mid A)P(A)$

  $\rightarrow BEL(B) \propto P(c,e,f \mid B)P(B)$

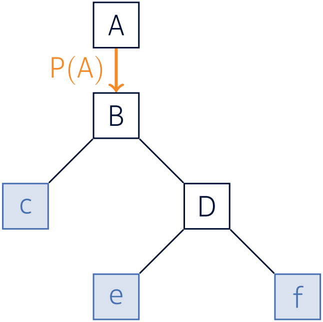
NOKIA Bell Labs TELECOM ParisTech

Causal support $P\left(X \mid \begin{matrix}\text{evidence} \\ \text{above } X\end{matrix}\right)$

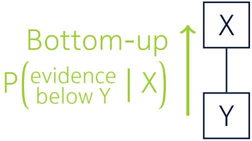


  Public NOKIA Bell Labs TELECOM ParisTech

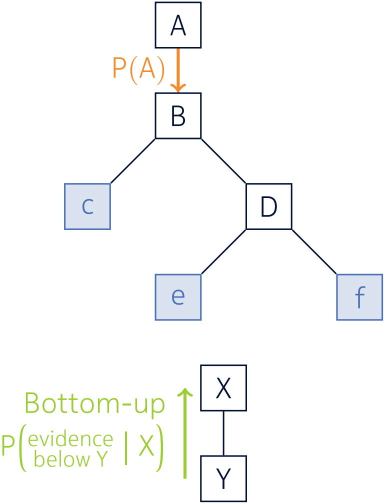
Causal support $P\left(X \mid \begin{matrix}\text{evidence}\\ \text{above } X\end{matrix}\right)$



- $P(D \mid c) = \sum_B P(D \mid B, c) P(B \mid c)$

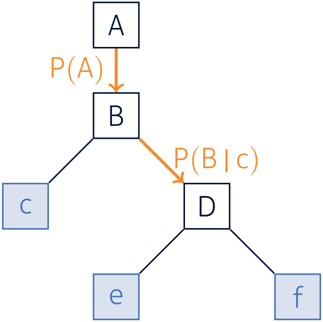


Causal support $P\left(X \mid \begin{smallmatrix}\text{evidence}\\ \text{above } X\end{smallmatrix}\right)$

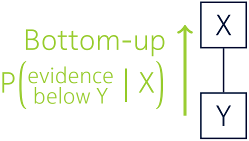


- $$P(D \mid c) = \sum_B P(D \mid B, c) P(B \mid c)$$
$$= \sum_B P(D \mid B) P(B \mid c)$$

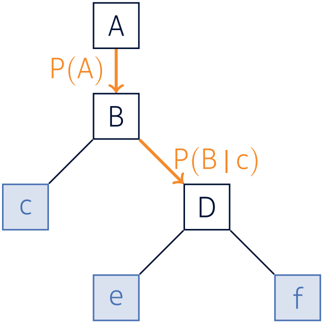
Causal support $P\left(X \mid \begin{matrix}\text{evidence}\\ \text{above } X\end{matrix}\right)$



- $P(D \mid c) = \sum_B P(D \mid B, c) P(B \mid c)$
  $= \sum_B P(D \mid B) P(B \mid c)$



Causal support $P\left(X \mid \begin{smallmatrix}\text{evidence}\\\text{above } X\end{smallmatrix}\right)$



- $P(D \mid c) = \sum_B P(D \mid B, c) P(B \mid c)$

  $= \sum_B P(D \mid B) P(B \mid c)$

Compute $P(B \mid c)$:

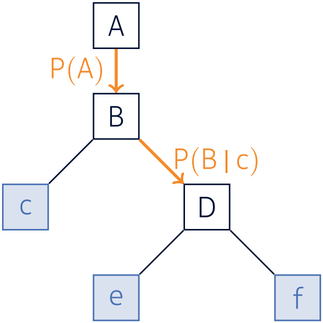
$$P(B \mid c, e, f) = \frac{P(e, f \mid B, c) P(B \mid c)}{P(e, f \mid c)}$$

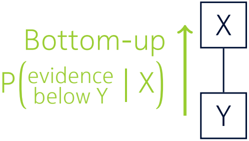
i.e. $P(B \mid c) \propto \dfrac{BEL(B)}{P(e, f \mid B)}$

  Public

NOKIA Bell Labs TELECOM ParisTech

Causal support $P\left(X \,\middle|\, \substack{\text{evidence} \\ \text{above } X}\right)$





- $$P(D \mid c) = \sum_B P(D \mid B, c) P(B \mid c)$$
  $$= \sum_B P(D \mid B) P(B \mid c)$$

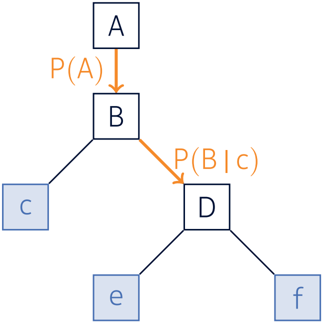
Compute $P(B \mid c)$:

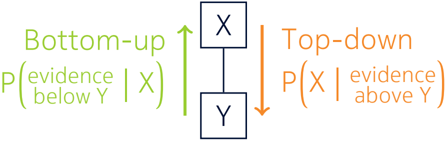
$$P(B \mid c, e, f) = \frac{P(e, f \mid B, c) P(B \mid c)}{P(e, f \mid c)}$$

i.e. $$P(B \mid c) \propto \frac{BEL(B)}{P(e, f \mid B)}$$

$\rightarrow BEL(D) \propto P(e, f \mid D) P(D \mid c)$

Causal support $P\left(X \mid \begin{smallmatrix}\text{evidence}\\ \text{above } X\end{smallmatrix}\right)$





- $$P(D \mid c) = \sum_B P(D \mid B, c) P(B \mid c)$$
$$= \sum_B P(D \mid B) P(B \mid c)$$

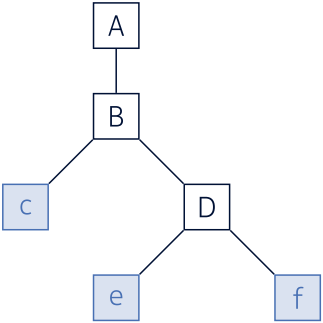
Compute $P(B \mid c)$:

$$P(B \mid c, e, f) = \frac{P(e, f \mid B, c) P(B \mid c)}{P(e, f \mid c)}$$

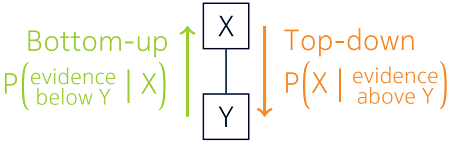
i.e. $P(B \mid c) \propto \dfrac{BEL(B)}{P(e, f \mid B)}$

$\rightarrow BEL(D) \propto P(e, f \mid D) P(D \mid c)$

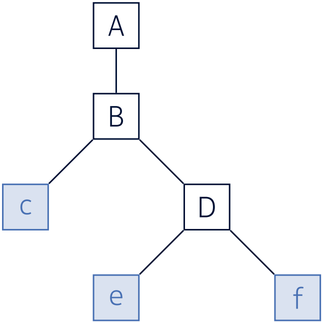
Summary

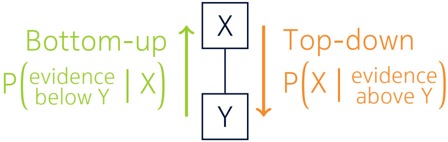


Algorithm



Summary

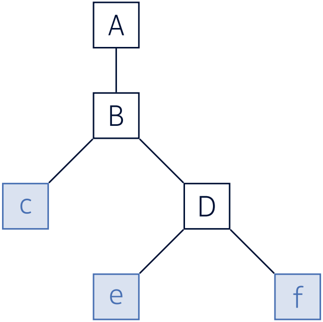




Algorithm

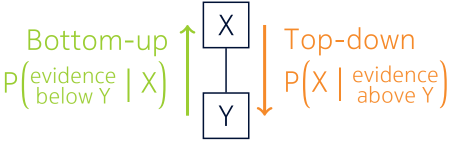
- Diagnostic support $P\left(\begin{smallmatrix}\text{evidence}\\ \text{below } X\end{smallmatrix} \mid X\right)$
  Bottom-up propagation

Summary



Algorithm

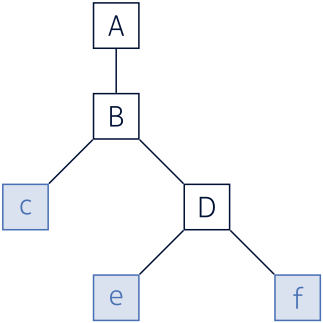
- Diagnostic support $P\left(\begin{matrix}\text{evidence}\\ \text{below } X\end{matrix} \,\middle|\, X\right)$
  Bottom-up propagation
- Causal support $P\left(X \,\middle|\, \begin{matrix}\text{evidence}\\ \text{above } X\end{matrix}\right)$
  Top-down propagation



NOKIA Bell Labs

TELECOM ParisTech

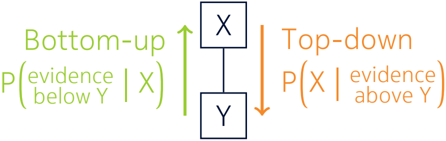
Summary



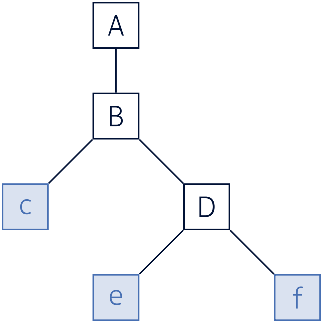
Algorithm

- Diagnostic support $P\left(\begin{smallmatrix}\text{evidence}\\ \text{below } X\end{smallmatrix} \mid X\right)$
  Bottom-up propagation
- Causal support $P\left(X \mid \begin{smallmatrix}\text{evidence}\\ \text{above } X\end{smallmatrix}\right)$
  Top-down propagation

In general



Summary

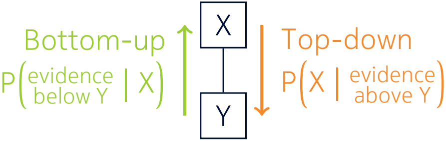


Algorithm

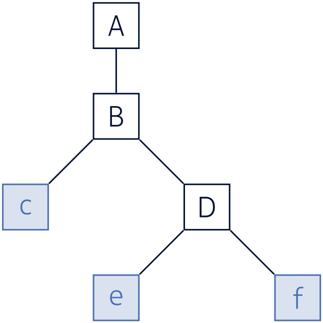
- Diagnostic support $P\left(\begin{smallmatrix}\text{evidence}\\ \text{below } X\end{smallmatrix} \mid X\right)$
  Bottom-up propagation
- Causal support $P\left(X \mid \begin{smallmatrix}\text{evidence}\\ \text{above } X\end{smallmatrix}\right)$
  Top-down propagation

In general

- Use a topological ordering



Summary

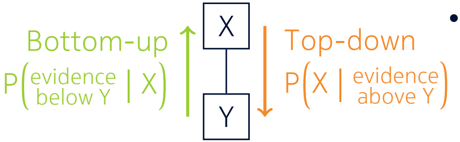


Algorithm

- Diagnostic support $P\left(\begin{smallmatrix}\text{evidence}\\ \text{below } X\end{smallmatrix} \mid X\right)$
  Bottom-up propagation
- Causal support $P\left(X \mid \begin{smallmatrix}\text{evidence}\\ \text{above } X\end{smallmatrix}\right)$
  Top-down propagation

In general

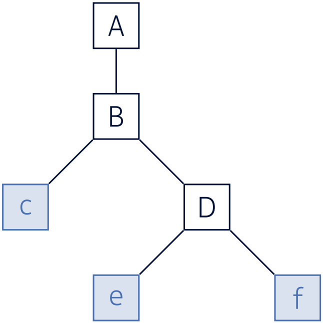
- Use a topological ordering
- Complexity: $O(rd^{\downarrow} + r^2 + r)$

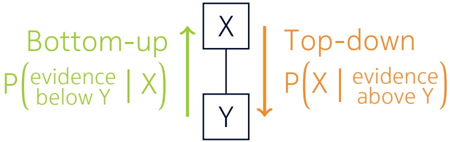


  Public

NOKIA Bell Labs TELECOM ParisTech

Additional remarks

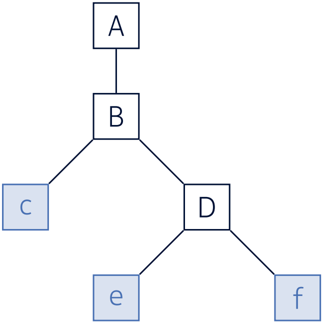




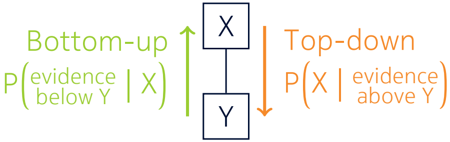
- If the evidence node is not a leaf: add a phantom node

  Public NOKIA Bell Labs TELECOM ParisTech

Additional remarks

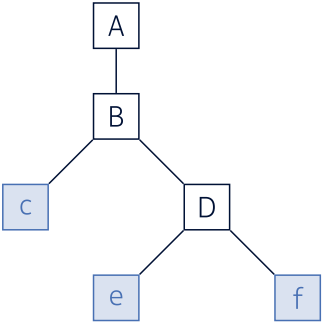


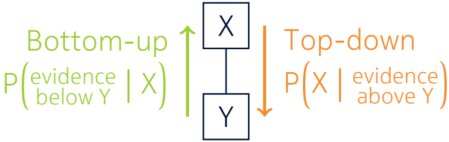
- If the evidence node is not a leaf: add a phantom node
- The calculations can be written as matrix products
  - Belief, causal and diagnostic supports, messages ~ Vectors
  - CPT ~ Matrix



NOKIA Bell Labs TELECOM ParisTech

Additional remarks





- If the evidence node is not a leaf: add a phantom node
- The calculations can be written as matrix products
  - Belief, causal and diagnostic supports, messages ~ Vectors
  - CPT ~ Matrix
- Asynchronous / parallel updates: acknowledgements (Pearl, 1982, 1986, 1988)

Outline

Reminders on probability theory

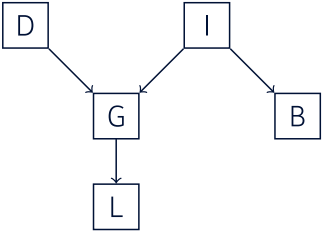
Bayesian networks

Belief propagation in trees

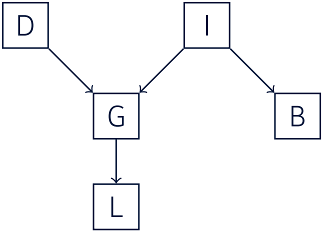
Belief propagation in polytrees

Polytree (or singly-connected) Bayesian network

The underlying undirected graph is a tree



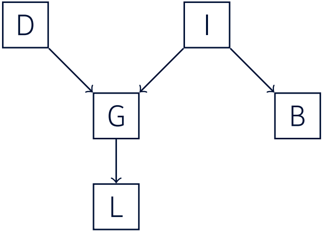
Polytree (or singly-connected) Bayesian network



The underlying undirected graph is a tree

Separation properties

Polytree (or singly-connected) Bayesian network

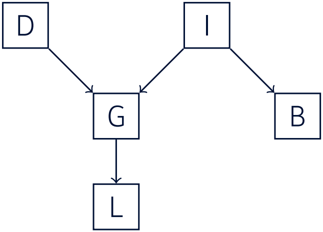


The underlying undirected graph is a tree

Separation properties

- Given a node, the non-descendants and the subtrees rooted at each child are independent

Polytree (or singly-connected) Bayesian network

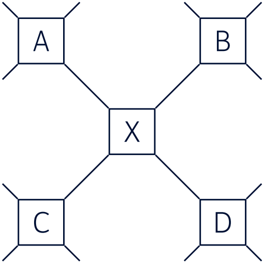


The underlying undirected graph is a tree

Separation properties

- Given a node, the non-descendants and the subtrees rooted at each child are independent
- If we don't condition on a node nor any of its descendants, the inversed subtrees rooted at its ancestors are independent

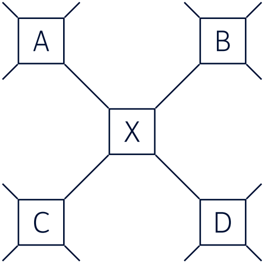
Belief propagation



  Public

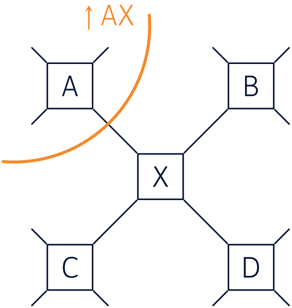
Belief propagation

$$P(X \mid \text{evidence}) \propto P\left(\begin{matrix}\text{evidence}\\ \text{below } X\end{matrix} \,\middle|\, X\right) \times P\left(X \,\middle|\, \begin{matrix}\text{evidence}\\ \text{above } X\end{matrix}\right)$$



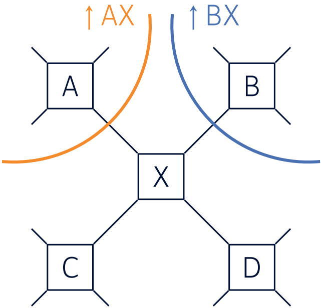
Belief propagation

$$P(X \mid \text{evidence}) \propto P\left(\begin{matrix}\text{evidence}\\ \text{below } X\end{matrix} \,\middle|\, X\right) \times P\left(X \,\middle|\, \begin{matrix}\text{evidence}\\ \text{above } X\end{matrix}\right)$$



Belief propagation

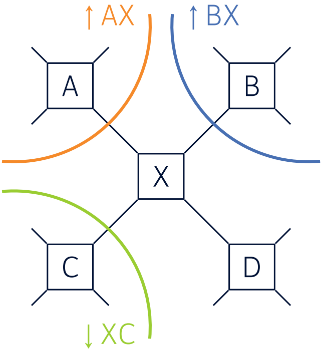
$$P(X \mid \text{evidence}) \propto P\left(\begin{smallmatrix}\text{evidence} \\ \text{below } X\end{smallmatrix} \mid X\right) \times P\left(X \mid \begin{smallmatrix}\text{evidence} \\ \text{above } X\end{smallmatrix}\right)$$



  Public

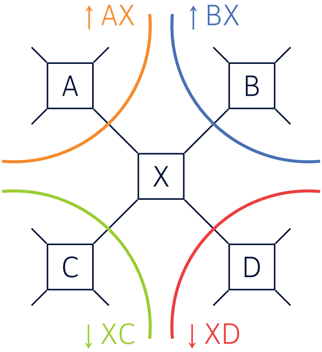
Belief propagation

$$P(X \mid \text{evidence}) \propto P\left(\begin{matrix}\text{evidence}\\ \text{below } X\end{matrix} \,\middle|\, X\right) \times P\left(X \,\middle|\, \begin{matrix}\text{evidence}\\ \text{above } X\end{matrix}\right)$$



Belief propagation

$$P(X \mid \text{evidence}) \propto P\left(\begin{matrix}\text{evidence}\\ \text{below } X\end{matrix} \,\middle|\, X\right) \times P\left(X \,\middle|\, \begin{matrix}\text{evidence}\\ \text{above } X\end{matrix}\right)$$



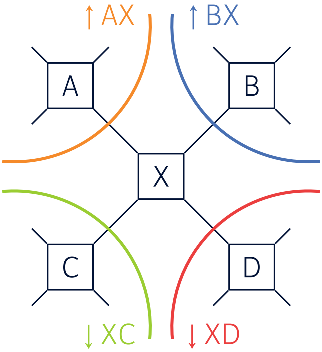
  Public

NOKIA Bell Labs

TELECOM ParisTech

Belief propagation

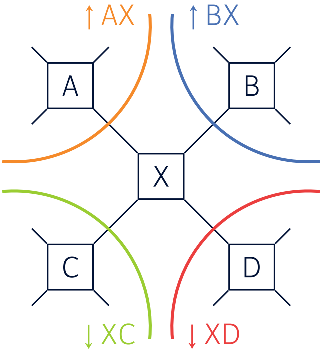
$$P(X \mid \text{evidence}) \propto P\left(\begin{matrix}\text{evidence} \\ \text{below X}\end{matrix} \mid X\right) \times P\left(X \mid \begin{matrix}\text{evidence} \\ \text{above X}\end{matrix}\right)$$



- Diagnostic support $P\left(\begin{matrix}\text{evidence} \\ \text{below X}\end{matrix} \mid X\right)$
  Bottom-up propagation
  ↓ XC and ↓ XD are independent given X

Belief propagation

$$P(X \mid \text{evidence}) \propto P\left(\begin{matrix}\text{evidence}\\ \text{below } X\end{matrix} \,\middle|\, X\right) \times P\left(X \,\middle|\, \begin{matrix}\text{evidence}\\ \text{above } X\end{matrix}\right)$$



- Diagnostic support $P\left(\begin{matrix}\text{evidence}\\ \text{below } X\end{matrix} \,\middle|\, X\right)$
  Bottom-up propagation
  ↓ XC and ↓ XD are independent given X
- Causal support $P\left(X \,\middle|\, \begin{matrix}\text{evidence}\\ \text{above } X\end{matrix}\right)$
  Top-down propagation
  ↑ AX and ↑ BX are independent

Conclusion

Conclusion

- **Bayesian networks**
  A memory-efficient way of storing a PMF by leveraging conditional independencies between variables

Conclusion

- **Bayesian networks**
  A memory-efficient way of storing a PMF by leveraging conditional independencies between variables

- **Belief propagation**
  A time-efficient algorithm for computing the belief
  - Asynchronous, parallelizable
  - Exact in (poly)trees
  - In general, extended to the junction tree algorithm and to other (approximate) algorithms

