Introduction to Artificial Intelligence

Introduction to Bayesian Networks

Andres Mendez-Vazquez

February 11, 2020

Outline

Cinvestav

Outline

Cinvestav

History

History

- '60s The first expert systems. IF-THEN rules.
- 1968 Attempts to use probabilities in expert systems (Gorry & Barnett).
- 1973 Gave up - to heavy calculations! (Gorry).
- 1976 MYCIN: Medical predicate logic expert system with certainty factors (Shortliffe).
- 1976 PROSPECTOR: Predicts the likely location of mineral deposits. Uses Bayes' rule. (Duda et al.).

History

History

- '60s The first expert systems. IF-THEN rules.
- 1968 Attempts to use probabilities in expert systems (Gorry & Barnett).
- 1973 Gave up - to heavy calculations! (Gorry).
- 1976 MYCIN: Medical predicate logic expert system with certainty factors (Shortliffe).
- 1976 PROSPECTOR: Predicts the likely location of mineral deposits. Uses Bayes' rule. (Duda et al.).

Summary until mid '80s

- "Pure logic will solve the AI problems!"
- "Probability theory is intractable to use and too complicated for complex models."

History

History

- '60s The first expert systems. IF-THEN rules.
- 1968 Attempts to use probabilities in expert systems (Gorry & Barnett).
- 1973 Gave up - to heavy calculations! (Gorry).
- 1976 MYCIN: Medical predicate logic expert system with certainty factors (Shortliffe).
- 1976 PROSPECTOR: Predicts the likely location of mineral deposits. Uses Bayes' rule. (Duda et al.).

Summary until mid '80s

- "Pure logic will solve the AI problems!"
- "Probability theory is intractable to use and too complicated for complex models."

History

History

- ‘60s The first expert systems. IF-THEN rules.
- 1968 Attempts to use probabilities in expert systems (Gorry & Barnett).
- 1973 Gave up - to heavy calculations! (Gorry).
- 1976 MYCIN: Medical predicate logic expert system with certainty factors (Shortliffe).
- 1976 PROSPECTOR: Predicts the likely location of mineral deposits. Uses Bayes' rule. (Duda et al.).

Summary until mid '80s

- "Pure logic will solve the AI problems!"
- "Probability theory is intractable to use and too complicated for complex models."

History

History

- ‘60s The first expert systems. IF-THEN rules.
- 1968 Attempts to use probabilities in expert systems (Gorry & Barnett).
- 1973 Gave up - to heavy calculations! (Gorry).
- 1976 MYCIN: Medical predicate logic expert system with certainty factors (Shortliffe).
- 1976 PROSPECTOR: Predicts the likely location of mineral deposits. Uses Bayes’ rule. (Duda et al.).

Summary until mid ’80s

- “Pure logic will solve the AI problems!”
- “Probability theory is intractable to use and too complicated for complex models.”

History

History

- '60s The first expert systems. IF-THEN rules.
- 1968 Attempts to use probabilities in expert systems (Gorry & Barnett).
- 1973 Gave up - to heavy calculations! (Gorry).
- 1976 MYCIN: Medical predicate logic expert system with certainty factors (Shortliffe).
- 1976 PROSPECTOR: Predicts the likely location of mineral deposits. Uses Bayes' rule. (Duda et al.).

Summary until mid '80s

- "Pure logic will solve the AI problems!"
- "Probability theory is intractable to use and too complicated for complex models."

History

History

- ‘60s The first expert systems. IF-THEN rules.
- 1968 Attempts to use probabilities in expert systems (Gorry & Barnett).
- 1973 Gave up - to heavy calculations! (Gorry).
- 1976 MYCIN: Medical predicate logic expert system with certainty factors (Shortliffe).
- 1976 PROSPECTOR: Predicts the likely location of mineral deposits. Uses Bayes’ rule. (Duda et al.).

Summary until mid ’80s

- “Pure logic will solve the AI problems!”
- “Probability theory is intractable to use and too complicated for complex models.”

But...

More History

- 1986 Bayesian networks were revived and reintroduced to expert systems (Pearl).
- 1988 Breakthrough for efficient calculation algorithms (Lauritzen & Spiegelhalter) tractable calculations on Bayesian Networkss.
- 1995 In Windows95™ for printer-trouble shooting and Office assistance ("the paper clip").
- 1999 Bayesian Networks are getting more and more used. Ex. Gene expression analysis, Business strategy etc.
- 2000 Widely used - A Bayesian Network tool will be shipped with every Windows™ Commercial Server.

But...

More History

- 1986 Bayesian networks were revived and reintroduced to expert systems (Pearl).
- 1988 Breakthrough for efficient calculation algorithms (Lauritzen & Spiegelhalter) tractable calculations on Bayesian Networkss.
- 1995 In Windows95™ for printer-trouble shooting and Office assistance ("the paper clip").
- 1999 Bayesian Networks are getting more and more used. Ex. Gene expression analysis, Business strategy etc.
- 2000 Widely used - A Bayesian Network tool will be shipped with every Windows™ Commercial Server.

But...

More History

- 1986 Bayesian networks were revived and reintroduced to expert systems (Pearl).
- 1988 Breakthrough for efficient calculation algorithms (Lauritzen & Spiegelhalter) tractable calculations on Bayesian Networkss.
- 1995 In Windows95™ for printer-trouble shooting and Office assistance (“the paper clip”).
- 1999 Bayesian Networks are getting more and more used. Ex. Gene expression analysis, Business strategy etc.
- 2000 Widely used - A Bayesian Network tool will be shipped with every Windows™ Commercial Server.

But...

More History

- 1986 Bayesian networks were revived and reintroduced to expert systems (Pearl).
- 1988 Breakthrough for efficient calculation algorithms (Lauritzen & Spiegelhalter) tractable calculations on Bayesian Networkss.
- 1995 In Windows95™ for printer-trouble shooting and Office assistance ("the paper clip").
- 1999 Bayesian Networks are getting more and more used. Ex. Gene expression analysis, Business strategy etc.
- 2000 Widely used - A Bayesian Network tool will be shipped with every Windows™ Commercial Server.

But...

More History

- 1986 Bayesian networks were revived and reintroduced to expert systems (Pearl).
- 1988 Breakthrough for efficient calculation algorithms (Lauritzen & Spiegelhalter) tractable calculations on Bayesian Networkss.
- 1995 In Windows95™ for printer-trouble shooting and Office assistance ("the paper clip").
- 1999 Bayesian Networks are getting more and more used. Ex. Gene expression analysis, Business strategy etc.
- 2000 Widely used - A Bayesian Network tool will be shipped with every Windows™ Commercial Server.

Furtheron 2000-2015

Bayesian Networks are use in

- Spam Detection.
- Gene Discovery.
- Signal Processing.
- Ranking.
- Forecasting.
- etc.

Something Notable

We are interested more and more on building automatically Bayesian Networks using data!!!

Cinvestav

Furtheron 2000-2015

Bayesian Networks are use in

- Spam Detection.
- Gene Discovery.
- Signal Processing.
- Ranking.
- Forecasting.
- etc.

Something Notable

We are interested more and more on building automatically Bayesian Networks using data!!!

Bayesian Network Advantages

Many of Them

1. Since in a Bayesian network encodes all variables, missing data entries can be handled successfully.
2. When used for learning casual relationships, they help better understand a problem domain as well as forecast consequences.
3. it is ideal to use a Bayesian network for representing prior data and knowledge.
4. Over-fitting of data can be avoidable when using Bayesian networks and Bayesian statistical methods.

Cinvestav

Bayesian Network Advantages

Many of Them

1. Since in a Bayesian network encodes all variables, missing data entries can be handled successfully.
2. When used for learning casual relationships, they help better understand a problem domain as well as forecast consequences.
3. it is ideal to use a Bayesian network for representing prior data and knowledge.
4. Over-fitting of data can be avoidable when using Bayesian networks and Bayesian statistical methods.

Bayesian Network Advantages

Many of Them

1. Since in a Bayesian network encodes all variables, missing data entries can be handled successfully.
2. When used for learning casual relationships, they help better understand a problem domain as well as forecast consequences.
3. it is ideal to use a Bayesian network for representing prior data and knowledge.
4. Over-fitting of data can be avoidable when using Bayesian networks and Bayesian statistical methods.

Cinvestav

Bayesian Network Advantages

Many of Them

1. Since in a Bayesian network encodes all variables, missing data entries can be handled successfully.
2. When used for learning casual relationships, they help better understand a problem domain as well as forecast consequences.
3. it is ideal to use a Bayesian network for representing prior data and knowledge.
4. Over-fitting of data can be avoidable when using Bayesian networks and Bayesian statistical methods.

Outline

Outline

Cinvestav

Bayes Theorem

One Version

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

- $P(A|B)$ is the conditional probability of A, given B. It is also called the posterior probability because it is derived from or depends upon the specified value of B.
- $P(B|A)$ is the conditional probability of B given A. It is also called the likelihood.
- $P(B)$ is the prior or marginal probability of B, and acts as a normalizing constant.

Bayes Theorem

One Version

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

Where

- $P(A)$ is the **prior probability** or marginal probability of A. It is "prior" in the sense that it does not take into account any information about B.
- $P(A|B)$ is the **conditional probability** of A, given B. It is also called the posterior probability because it is derived from or depends upon the specified value of B.
- $P(B|A)$ is the **conditional probability** of B given A. It is also called the likelihood.
- $P(B)$ is the **prior or marginal probability** of B, and acts as a normalizing constant.

Bayes Theorem

One Version

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

Where

- $P(A)$ is the **prior probability** or marginal probability of A. It is "prior" in the sense that it does not take into account any information about B.
- $P(A|B)$ is the **conditional probability** of A, given B. It is also called the posterior probability because it is derived from or depends upon the specified value of B.
- $P(B|A)$ is the **conditional probability** of B given A. It is also called the likelihood.
- $P(B)$ is the **prior or marginal probability** of B, and acts as a normalizing constant.

Bayes Theorem

One Version

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

Where

- $P(A)$ is the **prior probability** or marginal probability of A. It is "prior" in the sense that it does not take into account any information about B.
- $P(A|B)$ is the **conditional probability** of A, given B. It is also called the posterior probability because it is derived from or depends upon the specified value of B.
- $P(B|A)$ is the **conditional probability** of B given A. It is also called the likelihood.
- $P(B)$ is the **prior or marginal probability** of B, and acts as a normalizing constant.

Bayes Theorem

One Version

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

Where

- $P(A)$ is the **prior probability** or marginal probability of A. It is "prior" in the sense that it does not take into account any information about B.
- $P(A|B)$ is the **conditional probability** of A, given B. It is also called the posterior probability because it is derived from or depends upon the specified value of B.
- $P(B|A)$ is the **conditional probability** of B given A. It is also called the likelihood.
- $P(B)$ is the **prior or marginal probability** of B, and acts as a normalizing constant.

A Simple Example

Consider two related variables:

1. Drug (D) with values y or n
2. Test (T) with values $+$ve or $-$ve

A Simple Example

Consider two related variables:

1. Drug (D) with values y or n
2. Test (T) with values $+$ve or $-$ve

Initial Probabilities

- $P(D = y) = 0.001$
- $P(T = +ve | D = y) = 0.8$
- $P(T = +ve | D = n) = 0.01$

A Simple Example

Consider two related variables:

1. Drug (D) with values y or n
2. Test (T) with values $+$ve or $-$ve

Initial Probabilities

- $P(D = y) = 0.001$
- $P(T = +ve | D = y) = 0.8$
- $P(T = +ve | D = n) = 0.01$

A Simple Example

Consider two related variables:

1. Drug (D) with values y or n
2. Test (T) with values $+$ve or $-$ve

Initial Probabilities

- $P(D = y) = 0.001$
- $P(T = +ve|D = y) = 0.8$
- $P(T = +ve|D = n) = 0.01$

A Simple Example

Consider two related variables:

1. Drug (D) with values y or n
2. Test (T) with values $+$ve or –ve

Initial Probabilities

- $P(D = y) = 0.001$
- $P(T = +ve|D = y) = 0.8$
- $P(T = +ve|D = n) = 0.01$

A Simple Example

What is the probability that a person has taken the drug?

$$P\left(D=y|T=+ve\right)=\frac{P\left(T=+ve|D=y\right)P\left(D{=}\text{y}\right)}{P\left(T=+ve|D=y\right)P\left(D{=}\text{y}\right)+P\left(T=+ve|D=n\right)P\left(D{=}\text{n}\right)}$$

Let me develop the equation

Using simply

$$P\left(A,B\right)=P\left(A|B\right)P\left(B\right)\ \text{(Chain Rule)} \tag{1}$$

A Simple Example

What is the probability that a person has taken the drug?

$$P(D = y|T = +ve) = \frac{P(T = +ve|D = y)\,P(D=\text{y})}{P(T = +ve|D = y)\,P(D=\text{y}) + P(T = +ve|D = n)\,P(D=\text{n})}$$

Let me develop the equation

Using simply

$$P(A, B) = P(A|B)\,P(B) \quad \text{(Chain Rule)} \tag{1}$$

Cinvestav

Outline

Cinvestav

A More Complex Case

Increase Complexity

- Suppose now that there is a similar link between Lung Cancer (L) and a chest X-ray (X) and that we also have the following relationships:
  - History of smoking (S) has a direct influence on bronchitis (B) and lung cancer (L);
  - L and B have a direct influence on fatigue (F).

Cinvestav

A More Complex Case

Increase Complexity

- Suppose now that there is a similar link between Lung Cancer (L) and a chest X-ray (X) and that we also have the following relationships:
  - History of smoking (S) has a direct influence on bronchitis (B) and lung cancer (L);
  - L and B have a direct influence on fatigue (F).

Question

- What is the probability that someone has bronchitis given that they smoke, have fatigue and have received a positive X-ray result?

A More Complex Case

Increase Complexity

- Suppose now that there is a similar link between Lung Cancer (L) and a chest X-ray (X) and that we also have the following relationships:
  - History of smoking (S) has a direct influence on bronchitis (B) and lung cancer (L);
  - L and B have a direct influence on fatigue (F).

Question

- What is the probability that someone has bronchitis given that they smoke, have fatigue and have received a positive X-ray result?

A More Complex Case

Increase Complexity

- Suppose now that there is a similar link between Lung Cancer (L) and a chest X-ray (X) and that we also have the following relationships:
  - History of smoking (S) has a direct influence on bronchitis (B) and lung cancer (L);
  - L and B have a direct influence on fatigue (F).

Question

- What is the probability that someone has bronchitis given that they smoke, have fatigue and have received a positive X-ray result?

A More Complex Case

Short Hand

$$P\left(b_1|s_1,f_1,x_1\right)=\frac{P\left(b_1,s_1,f_1,x_1\right)}{P\left(s_1,f_1,x_1\right)}=\frac{\sum_l P\left(b_1,s_1,f_1,x_1,l\right)}{\sum_{b,l} P\left(b,s_1,f_1,x_1,l\right)}$$

Values for the Complex Case

Table

Feature	Value	When the Feature Takes this Value
H	h_1	There is a history of smoking
	h_2	There is no history of smoking
B	b_1	Bronchitis is present
	b_2	Bronchitis is absent
L	l_1	Lung cancer is present
	l_2	Lung cancer is absent
F	f_1	Fatigue is present
	f_2	Fatigue is absent
C	c_1	Chest X-ray is positive
	c_2	Chest X-ray is negative

Problem with Large Instances

The joint probability distribution $P(H, B, L, F, C)$

- For five binary variables there are $2^5 = 32$ values in the joint distribution (for 100 variables there are over 2^{100} values)
- How are these values to be obtained?

Problem with Large Instances

The joint probability distribution $P(H, B, L, F, C)$

- For five binary variables there are $2^5 = 32$ values in the joint distribution (for 100 variables there are over 2^{100} values)
- **How are these values to be obtained?**

We can try to do inference

- To obtain posterior distributions once some evidence is available requires summation over an exponential number of terms!!!

Cinvestav

Problem with Large Instances

The joint probability distribution $P(H, B, L, F, C)$

- For five binary variables there are $2^5 = 32$ values in the joint distribution (for 100 variables there are over 2^{100} values)
- **How are these values to be obtained?**

We can try to do inference

- To obtain posterior distributions once some evidence is available requires summation over an exponential number of terms!!!

Ok!!!

- We need something BETTER!!!

Problem with Large Instances

The joint probability distribution $P(H, B, L, F, C)$

- For five binary variables there are $2^5 = 32$ values in the joint distribution (for 100 variables there are over 2^{100} values)
- **How are these values to be obtained?**

We can try to do inference

- To obtain posterior distributions once some evidence is available requires summation over an exponential number of terms!!!

Ok!!!

- **We need something BETTER!!!**

Cinvestav

Outline

Cinvestav

Bayesian Networks

Definition

A Bayesian network consists of

- A Graph
  - Nodes represent the random variables.
  - Directed edges (arrows) between pairs of nodes.
  - it must be a Directed Acyclic Graph (DAG) – no directed cycles.
  - The graph represents independence relationships between variables.

Bayesian Networks

Definition

A Bayesian network consists of

- A Graph
  - Nodes represent the random variables.
  - Directed edges (arrows) between pairs of nodes.
  - it must be a Directed Acyclic Graph (DAG) – no directed cycles.
  - The graph represents independence relationships between variables.

This allows to define

- Conditional Probability Specifications:
  - The conditional probability of each variable given its parents in the DAG.

Bayesian Networks

Definition

A Bayesian network consists of

- A Graph
  - Nodes represent the random variables.
  - Directed edges (arrows) between pairs of nodes.
  - it must be a Directed Acyclic Graph (DAG) – no directed cycles.
  - The graph represents independence relationships between variables.

This allows to define

- Conditional Probability Specifications:
  - The conditional probability of each variable given its parents in the DAG.

Bayesian Networks

Definition

A Bayesian network consists of

- A Graph
  - Nodes represent the random variables.
  - Directed edges (arrows) between pairs of nodes.
  - it must be a Directed Acyclic Graph (DAG) – no directed cycles.
  - The graph represents independence relationships between variables.

This allows to define

- Conditional Probability Specifications:
  - The conditional probability of each variable given its parents in the DAG.

Cinvestav

Bayesian Networks

Definition

A Bayesian network consists of

- A Graph
  - Nodes represent the random variables.
  - Directed edges (arrows) between pairs of nodes.
  - it must be a Directed Acyclic Graph (DAG) – no directed cycles.
  - The graph represents independence relationships between variables.

This allows to define

- Conditional Probability Specifications:
  - The conditional probability of each variable given its parents in the DAG.

Bayesian Networks

Definition

A Bayesian network consists of

- A Graph
  - Nodes represent the random variables.
  - Directed edges (arrows) between pairs of nodes.
  - it must be a Directed Acyclic Graph (DAG) – no directed cycles.
  - The graph represents independence relationships between variables.

This allows to define

- Conditional Probability Specifications:
  - The conditional probability of each variable given its parents in the DAG.

Bayesian Networks

Definition

A Bayesian network consists of

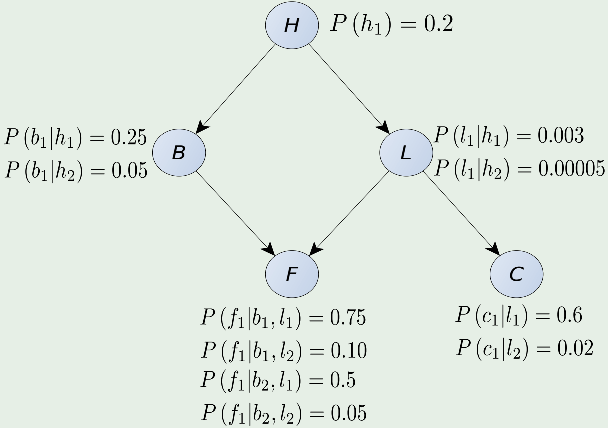
- A Graph
  - Nodes represent the random variables.
  - Directed edges (arrows) between pairs of nodes.
  - it must be a Directed Acyclic Graph (DAG) – no directed cycles.
  - The graph represents independence relationships between variables.

This allows to define

- Conditional Probability Specifications:
  - The conditional probability of each variable given its parents in the DAG.

Example

DAG for the previous Lung Cancer Problem



Outline

Markov Condition

Definition

- Suppose we have a joint probability distribution P of the random variables in some set V and a DAG $G = (V, E)$.
  - We say that (G, P) satisfies **the Markov condition** if for each variable $X \in V$, $\{X\}$ is conditionally independent of the set of all its non-descendents given the set of all its parents.

Cinvestav

Markov Condition

Definition

- Suppose we have a joint probability distribution P of the random variables in some set V and a DAG $G = (V, E)$.
  - We say that (G, P) satisfies **the Markov condition** if for each variable $X \in V$, $\{X\}$ is conditionally independent of the set of all its non-descendents given the set of all its parents.

Notation

- $\text{PA}_X =$ set of parents of X.
- $\text{ND}_X =$ set of non-descendants of X.

Markov Condition

Definition

- Suppose we have a joint probability distribution P of the random variables in some set V and a DAG $G = (V, E)$.
  - We say that (G, P) satisfies **the Markov condition** if for each variable $X \in V$, $\{X\}$ is conditionally independent of the set of all its non-descendents given the set of all its parents.

Notation

- $\mathrm{PA}_X =$ set of parents of X.
- $\mathrm{ND}_X =$ set of non-descendants of X.

We use the following the notation

$$I_P(\{X\}, ND_X | PA_X)$$

Markov Condition

Definition

- Suppose we have a joint probability distribution P of the random variables in some set V and a DAG $G = (V, E)$.
  - We say that (G, P) satisfies **the Markov condition** if for each variable $X \in V$, $\{X\}$ is conditionally independent of the set of all its non-descendents given the set of all its parents.

Notation

- $\text{PA}_X =$ set of parents of X.
- $\text{ND}_X =$ set of non-descendants of X.

We use the following the notation

$$I_P(\{X\}, ND_X | PA_X)$$

Markov Condition

Definition

- Suppose we have a joint probability distribution P of the random variables in some set V and a DAG $G = (V, E)$.
  - We say that (G, P) satisfies **the Markov condition** if for each variable $X \in V$, $\{X\}$ is conditionally independent of the set of all its non-descendents given the set of all its parents.

Notation

- $\text{PA}_X =$ set of parents of X.
- $\text{ND}_X =$ set of non-descendants of X.

We use the following the notation

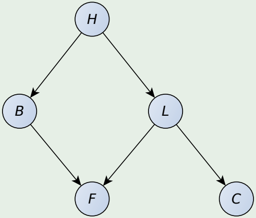
$$I_P\left(\{X\}, ND_X | PA_X\right)$$

Outline

Cinvestav

Example

We have that

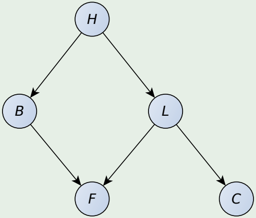


Given the previous DAG we have

Node	PA	Conditional Independence
C	$\{L\}$	$I_P(\{C\}, \{H, B, F\} \mid \{L\})$
B	$\{H\}$	$I_P(\{B\}, \{L, C\} \mid \{H\})$
F	$\{B, L\}$	$I_P(\{F\}, \{H, C\} \mid \{B, L\})$
L	$\{H\}$	$I_P(\{L\}, \{B\} \mid \{H\})$

Example

We have that



Given the previous DAG we have

Node	PA	Conditional Independence
C	$\{L\}$	$I_P(\{C\}, \{H, B, F\} \mid \{L\})$
B	$\{H\}$	$I_P(\{B\}, \{L, C\} \mid \{H\})$
F	$\{B, L\}$	$I_P(\{F\}, \{H, C\} \mid \{B, L\})$
L	$\{H\}$	$I_P(\{L\}, \{B\} \mid \{H\})$

Outline

Cinvestav

Using the Markov Condition

First Decompose a Joint Distribution using the Chain Rule

$$P(c, f, l, b, h) = P(c|b, s, l, f)\, P(f|b, h, l)\, P(l|b, h)\, P(b|h)\, P(h) \quad (2)$$

Using the Markov condition in the following DAG

We have the following equivalences

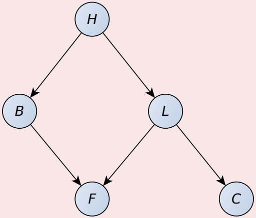
- $P(c|b, h, l, f) = P(c|l)$
- $P(f|b, h, l) = P(f|b, l)$
- $P(l|b, h) = P(l|h)$

Using the Markov Condition

First Decompose a Joint Distribution using the Chain Rule

$$P(c, f, l, b, h) = P(c|b, s, l, f)\, P(f|b, h, l)\, P(l|b, h)\, P(b|h)\, P(h) \qquad (2)$$

Using the Markov condition in the following DAG



We have the following equivalences

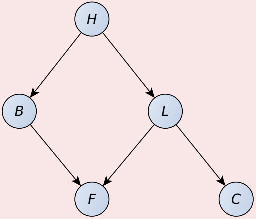
- $P(c|b, h, l, f) = P(c|l)$
- $P(f|b, h, l) = P(f|b, l)$
- $P(l|b, h) = P(l|h)$

Using the Markov Condition

First Decompose a Joint Distribution using the Chain Rule

$$P(c, f, l, b, h) = P(c|b, s, l, f)\, P(f|b, h, l)\, P(l|b, h)\, P(b|h)\, P(h) \tag{2}$$

Using the Markov condition in the following DAG



We have the following equivalences

- $P(c|b, h, l, f) = P(c|l)$
- $P(f|b, h, l) = P(f|b, l)$
- $P(l|b, h) = P(l|h)$

Using the Markov Condition

Finally

$$P(c, f, l, b, h) = P(c|l)\, P(f|b, l)\, P(l|h)\, P(b|h)\, P(h) \quad (3)$$

Outline

Cinvestav

Representing the Joint Distribution

Theorem (Product of Conditional Probabilities of the Parents)

If (G, P) satisfies the Markov condition, then P is equal to the product of its conditional distributions of all nodes given values of their parents, whenever these conditional distributions exist.

General Representation

- In general, for a network with nodes $X_1, X_2, ..., X_n \Rightarrow$

$$P(x_1, x_2, ..., x_n) = \prod_{i=1}^{n} P(x_i | PA(x_i))$$

Representing the Joint Distribution

Theorem (Product of Conditional Probabilities of the Parents)

If (G, P) satisfies the Markov condition, then P is equal to the product of its conditional distributions of all nodes given values of their parents, whenever these conditional distributions exist.

General Representation

- In general, for a network with nodes $X_1, X_2, ..., X_n \Rightarrow$

$$P(x_1, x_2, ..., x_n) = \prod_{i=1}^{n} P(x_i | PA(x_i))$$

Proof

We prove the case where P is discrete

Order the nodes so that if Y is a descendant of Z, then Y follows Z in the ordering.

- Topological Sorting.

Proof

We prove the case where P is discrete

Order the nodes so that if Y is a descendant of Z, then Y follows Z in the ordering.

- Topological Sorting.

This is called

Ancestral ordering.

Proof

We prove the case where P is discrete

Order the nodes so that if Y is a descendant of Z, then Y follows Z in the ordering.

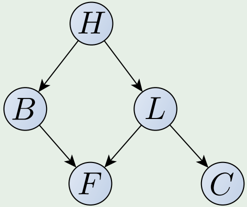
- Topological Sorting.

This is called

Ancestral ordering.

Proof

For example

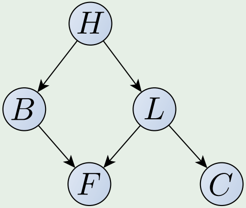


The ancestral ordering are

$$[H, L, B, C, F] \text{ and } [H, B, L, F, C] \tag{4}$$

Proof

For example



The ancestral ordering are

$$[H, L, B, C, F] \text{ and } [H, B, L, F, C] \tag{4}$$

Proof

Now

Let $X_1, X_2, ..., X_n$ be the resultant ordering.

For a given set of values of $x_1, x_2, ..., x_n$

Let pa_i be the subsets of these values containing the values of $X_i's$ parents

Thus, we need to prove that whenever $P(\text{pa}_i) \neq 0$ for $1 \leq i \leq n$

$$P(x_n, x_{n-1}, ..., x_1) = P(x_n|\text{pa}_n)\,P(x_{n-1}|\text{pa}_{n-1})\ldots P(x_1|\text{pa}_1) \quad (5)$$

Proof

Now

Let $X_1, X_2, ..., X_n$ be the resultant ordering.

For a given set of values of $x_1, x_2, ..., x_n$

Let pa_i be the subsets of these values containing the values of $X_i's$ parents

Thus, we need to prove that whenever $P(\text{pa}_i) \neq 0$ for $1 \leq i \leq n$

$$P(x_n, x_{n-1}, ..., x_1) = P(x_n|\text{pa}_n)\, P(x_{n-1}|\text{pa}_{n-1}) \ldots P(x_1|\text{pa}_1) \quad (5)$$

Proof

Now

Let $X_1, X_2, ..., X_n$ be the resultant ordering.

For a given set of values of $x_1, x_2, ..., x_n$

Let pa_i be the subsets of these values containing the values of $X_i's$ parents

Thus, we need to prove that whenever $P(\text{pa}_i) \neq 0$ for $1 \leq i \leq n$

$$P(x_n, x_{n-1}, ..., x_1) = P(x_n|\text{pa}_n) P(x_{n-1}|\text{pa}_{n-1}) ... P(x_1|\text{pa}_1) \quad (5)$$

Cinvestav

Proof

Something Notable

We show this using induction on the number of variables in the network.

- Assume then $P(\text{pa}_i) \neq 0$ for $1 \leq i \leq n$ for a combination of $x_i's$ values.

Proof

Something Notable

We show this using induction on the number of variables in the network.

- Assume then $P(\mathrm{pa}_i) \neq 0$ for $1 \leq i \leq n$ for a combination of $x_i's$ values.

Base Case of Induction

Since pa_1 is empty, then

$$P(x_1) = P(x_1|\mathrm{pa}_1) \quad (6)$$

Proof

Something Notable

We show this using induction on the number of variables in the network.

- Assume then $P\left(\text{pa}_i\right) \neq 0$ for $1 \leq i \leq n$ for a combination of $x_i's$ values.

Base Case of Induction

Since pa_1 is empty, then

$$P\left(x_1\right) = P\left(x_1|\text{pa}_1\right) \tag{6}$$

Inductive Hypothesis

Suppose for this combination of values of the x_i's that

$$P\left(x_i, x_{i-1}, ..., x_1\right) = P\left(x_i|\text{pa}_i\right) P\left(x_{i-1}|\text{pa}_{i-1}\right) ... P\left(x_1|\text{pa}_1\right) \tag{7}$$

Proof

Something Notable

We show this using induction on the number of variables in the network.

- Assume then $P(\mathsf{pa}_i) \neq 0$ for $1 \leq i \leq n$ for a combination of $x_i's$ values.

Base Case of Induction

Since pa_1 is empty, then

$$P(x_1) = P(x_1|\mathsf{pa}_1) \tag{6}$$

Inductive Hypothesis

Suppose for this combination of values of the x_i's that

$$P(x_i, x_{i-1}, ..., x_1) = P(x_i|\mathsf{pa}_i)\, P(x_{i-1}|\mathsf{pa}_{i-1}) \ldots P(x_1|\mathsf{pa}_1) \tag{7}$$

Proof

Something Notable

We show this using induction on the number of variables in the network.

- Assume then $P(\mathsf{pa}_i) \neq 0$ for $1 \leq i \leq n$ for a combination of $x_i's$ values.

Base Case of Induction

Since pa_1 is empty, then

$$P(x_1) = P(x_1|\mathsf{pa}_1) \tag{6}$$

Inductive Hypothesis

Suppose for this combination of values of the x_i's that

$$P(x_i, x_{i-1}, \dots, x_1) = P(x_i|\mathsf{pa}_i) P(x_{i-1}|\mathsf{pa}_{i-1}) \dots P(x_1|\mathsf{pa}_1) \tag{7}$$

Proof

Something Notable

We show this using induction on the number of variables in the network.

- Assume then $P\left(\mathsf{pa}_i\right) \neq 0$ for $1 \leq i \leq n$ for a combination of $x_i's$ values.

Base Case of Induction

Since pa_1 is empty, then

$$P\left(x_1\right) = P\left(x_1|\mathsf{pa}_1\right) \tag{6}$$

Inductive Hypothesis

Suppose for this combination of values of the x_i's that

$$P\left(x_i, x_{i-1}, ..., x_1\right) = P\left(x_i|\mathsf{pa}_i\right) P\left(x_{i-1}|\mathsf{pa}_{i-1}\right) ...P\left(x_1|\mathsf{pa}_1\right) \tag{7}$$

Proof

Inductive Step

We need show for this combination of values of the x_i's that

$$P\left(x_{i+1}, x_i, ..., x_1\right) = P\left(x_{i+1}|\mathsf{pa}_{i+1}\right) P\left(x_i|\mathsf{pa}_i\right) ... P\left(x_1|\mathsf{pa}_1\right) \tag{8}$$

Case 1

For this combination of values:

$$P\left(x_i, x_{i-1}, ..., x_1\right) = 0 \tag{9}$$

By Conditional Probability, we have

$$P\left(x_{i+1}, x_i, ..., x_1\right) = P\left(x_{i+1}|x_i, ..., x_1\right) P\left(x_i, ..., x_1\right) = 0 \tag{10}$$

Cinvestav

Proof

Inductive Step

We need show for this combination of values of the x_i's that

$$P\left(x_{i+1}, x_i, ..., x_1\right) = P\left(x_{i+1} | \mathsf{pa}_{i+1}\right) P\left(x_i | \mathsf{pa}_i\right) ... P\left(x_1 | \mathsf{pa}_1\right) \tag{8}$$

Case 1

For this combination of values:

$$P\left(x_i, x_{i-1}, ..., x_1\right) = 0 \tag{9}$$

By Conditional Probability, we have

$$P\left(x_{i+1}, x_i, ..., x_1\right) = P\left(x_{i+1} | x_i, ..., x_1\right) P\left(x_i, ..., x_1\right) = 0 \tag{10}$$

Proof

Inductive Step

We need show for this combination of values of the x_i's that

$$P(x_{i+1}, x_i, ..., x_1) = P(x_{i+1}|\mathsf{pa}_{i+1})\, P(x_i|\mathsf{pa}_i) ... P(x_1|\mathsf{pa}_1) \quad (8)$$

Case 1

For this combination of values:

$$P(x_i, x_{i-1}, ..., x_1) = 0 \quad (9)$$

By Conditional Probability, we have

$$P(x_{i+1}, x_i, ..., x_1) = P(x_{i+1}|x_i, ..., x_1)\, P(x_i, ..., x_1) = 0 \quad (10)$$

Proof

Due to the previous equalities and the inductive hypothesis

There is some k, $1 \leq k \leq i$ such that $P\left(x_k|\text{pa}_k\right) = 0$ because after all

$$P\left(x_i|\text{pa}_i\right) P\left(x_{i-1}|\text{pa}_{i-1}\right) ... P\left(x_1|\text{pa}_1\right) = 0 \quad (11)$$

Thus, the equality holds

Now for the Case 2

Case 2

For this combination of values $P\left(x_i, x_{i-1}, ..., x_1\right) \neq 0$

Proof

Due to the previous equalities and the inductive hypothesis

There is some k, $1 \leq k \leq i$ such that $P\left(x_k|\mathsf{pa}_k\right) = 0$ because after all

$$P\left(x_i|\mathsf{pa}_i\right) P\left(x_{i-1}|\mathsf{pa}_{i-1}\right) ... P\left(x_1|\mathsf{pa}_1\right) = 0 \tag{11}$$

Thus, the equality holds

Now for the Case 2

Case 2

For this combination of values $P\left(x_i, x_{i-1}, ..., x_1\right) \neq 0$

Proof

Due to the previous equalities and the inductive hypothesis

There is some k, $1 \leq k \leq i$ such that $P\left(x_k|\text{pa}_k\right) = 0$ because after all

$$P\left(x_i|\text{pa}_i\right) P\left(x_{i-1}|\text{pa}_{i-1}\right) ...P\left(x_1|\text{pa}_1\right) = 0 \tag{11}$$

Thus, the equality holds

Now for the Case 2

Case 2

For this combination of values $P\left(x_i, x_{i-1}, ..., x_1\right) \neq 0$

Proof

Thus by the Rule of Conditional Probability

$$P\left(x_{i+1}, x_i, ..., x_1\right) = P\left(x_{i+1} | x_i, ..., x_1\right) P\left(x_i, ..., x_1\right)$$

Proof

Thus by the Rule of Conditional Probability

$$P\left(x_{i+1}, x_i, ..., x_1\right) = P\left(x_{i+1} | x_i, ..., x_1\right) P\left(x_i, ..., x_1\right)$$

Definition Markov Condition (Remember!!!)

- Suppose we have a joint probability distribution P of the random variables in some set V and a DAG $G = (V, E)$.
  - We say that (G, P) satisfies **the Markov condition** if for each variable $X \in V$, $\{X\}$ is conditionally independent of the set of all its non-descendents given the set of all its parents.

Proof

Thus by the Rule of Conditional Probability

$$P\left(x_{i+1}, x_i, ..., x_1\right) = P\left(x_{i+1}|x_i, ..., x_1\right) P\left(x_i, ..., x_1\right)$$

Definition Markov Condition (Remember!!!)

- Suppose we have a joint probability distribution P of the random variables in some set V and a DAG $G = (V, E)$.
  - We say that (G, P) satisfies **the Markov condition** if for each variable $X \in V$, $\{X\}$ is conditionally independent of the set of all its non-descendents given the set of all its parents.

Cinvestav

Proof

Given this Markov Condition and the fact that $X_1, ..., X_i$ are all non-descendants of X_{i+1}

We have that

$$P\left(x_{i+1}, x_i, ..., x_1\right) = P\left(x_{i+1}|\text{pa}_{i+1}\right) P\left(x_i, ..., x_1\right)$$
$$= P\left(x_{i+1}|\text{pa}_{i+1}\right) P\left(x_i|\text{pa}_i\right) \cdots P\left(x_1|\text{pa}_1\right) \text{ (IH)}$$

Q.E.D.

Proof

Given this Markov Condition and the fact that $X_1, ..., X_i$ are all non-descendants of X_{i+1}

We have that

$$\begin{aligned} P\left(x_{i+1}, x_i, ..., x_1\right) &= P\left(x_{i+1}|\mathsf{pa}_{i+1}\right) P\left(x_i, ..., x_1\right) \\ &= P\left(x_{i+1}|\mathsf{pa}_{i+1}\right) P\left(x_i|\mathsf{pa}_i\right) \cdots P\left(x_1|\mathsf{pa}_1\right) \ \text{(IH)} \end{aligned}$$

Q.E.D.

Outline

Cinvestav

Example

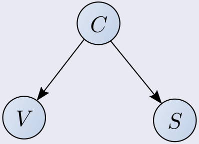
Imagine the following Random Variables

Variable	Value	Outcome
V	v_1	All objects containing a '1'
	v_2	All objects containing a '2'
S	s_1	All square objects
	s_2	All round objects
C	c_1	All black objects
	c_2	All white objects

Cinvestav

Example

Using the following graph

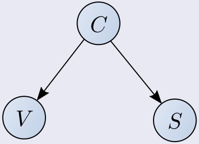


Using the chain rule

$$
\begin{aligned}
P(v, s, c) &= P(v|s, c)\, P(s|c)\, P(c) \\
&= P(v|c)\, P(s|c)\, P(c)
\end{aligned}
$$

Example

Using the following graph

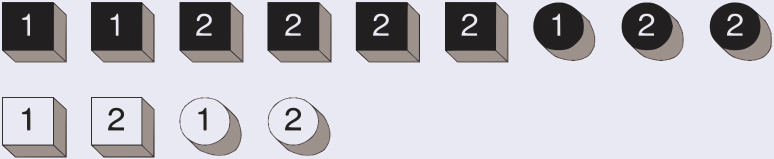


Using the chain rule

$$
\begin{aligned}
P(v,s,c) &= P(v|s,c)\,P(s|c)\,P(c)\\
&= P(v|c)\,P(s|c)\,P(c)
\end{aligned}
$$

Then, using the following probabilities

We have



c	s	v	$P(v\|s,c)$	$P(v\|c)$
c_1	s_1	v_1	1/3	1/3
c_1	s_1	v_2	2/3	2/3
c_1	s_2	v_1	1/3	1/3
c_1	s_2	v_2	2/3	2/3
c_2	s_1	v_1	1/2	1/2
c_2	s_1	v_2	1/2	1/2
c_2	s_2	v_1	1/2	1/2
c_2	s_2	v_2	1/2	1/2

Therefore

We have the following probabilities

$$P\left(c_1, s_1, v_1\right) = \frac{2}{13}$$

With the following using the Markov Condition

$$P\left(v_1|c_1\right) P\left(s_1|c_1\right) P\left(c_1\right) = P\left(One|Black\right) P\left(Square|Black\right) P\left(Black\right)$$
$$= \frac{1}{3} \times \frac{2}{3} \times \frac{9}{13} = \frac{2}{13}$$

Therefore

We have the following probabilities

$$P(c_1, s_1, v_1) = \frac{2}{13}$$

With the following using the Markov Condition

$$P(v_1|c_1)\,P(s_1|c_1)\,P(c_1) = P(One|Black)\,P(Square|Black)\,P(Black)$$
$$= \frac{1}{3} \times \frac{2}{3} \times \frac{9}{13} = \frac{2}{13}$$

Outline

Cinvestav

Now

OBSERVATIONS

- There are good savings in the Number of Values

Brute Force Approach

- on n binary variables requires m^n, if $m = \max\{|v_i| \, |V\}_{i=1}^{n}$.

For a Bayesian Network with n binary variables and each node has at most k parents

- Then, less than $m^k n$ values are required!!!

Now

OBSERVATIONS

- There are good savings in the Number of Values

Brute Force Approach

- on n binary variables requires m^n, if $m = \max\left\{|v_i|\,|V\right\}_{i=1}^{n}$.

For a Bayesian Network with n binary variables and each node has at most k parents

- Then, less than $m^k n$ values are required!!!

Now

OBSERVATIONS

- There are good savings in the Number of Values

Brute Force Approach

- on n binary variables requires m^n, if $m = \max\left\{|v_i|\,|V\right\}_{i=1}^{n}$.

For a Bayesian Network with n binary variables and each node has at most k parents

- Then, less than $m^k n$ values are required!!!

Outline

Cinvestav

It is more!!!

Theorem (Markov Condition on a DAG)

- Let a DAG G be given in which each node is a random variable, and let a discrete conditional probability distribution of each node given values of its parents in G be specified.
- Then, the product of these conditional distributions yields a joint probability distribution P of the variables, and (G, P) satisfies the Markov condition.

Cinvestav

It is more!!!

Theorem (Markov Condition on a DAG)

- Let a DAG G be given in which each node is a random variable, and let a discrete conditional probability distribution of each node given values of its parents in G be specified.
- Then, the product of these conditional distributions yields a joint probability distribution P of the variables, and (G, P) satisfies the Markov condition.

Note

- Notice that the theorem requires that specified conditional distributions be discrete.
- Often in the case of continuous distributions it still holds.

Cinvestav

It is more!!!

Theorem (Markov Condition on a DAG)

- Let a DAG G be given in which each node is a random variable, and let a discrete conditional probability distribution of each node given values of its parents in G be specified.
- Then, the product of these conditional distributions yields a joint probability distribution P of the variables, and (G, P) satisfies the Markov condition.

Note

- Notice that the theorem requires that specified conditional distributions be discrete.
- Often in the case of continuous distributions it still holds.

Cinvestav

It is more!!!

Theorem (Markov Condition on a DAG)

- Let a DAG G be given in which each node is a random variable, and let a discrete conditional probability distribution of each node given values of its parents in G be specified.
- Then, the product of these conditional distributions yields a joint probability distribution P of the variables, and (G, P) satisfies the Markov condition.

Note

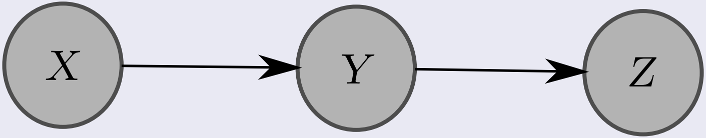
- Notice that the theorem requires that specified conditional distributions be discrete.
- Often in the case of continuous distributions it still holds.

Outline

Cinvestav

Example

We have the following DAG and probabilities



$$P(x_1) = 0.3 \qquad P(y_1|x_1) = 0.6 \qquad P(z_1|y_1) = 0.2$$
$$P(x_2) = 0.7 \qquad P(y_2|x_1) = 0.4 \qquad P(z_2|y_1) = 0.8$$

$$P(y_1|x_2) = 0.0 \qquad P(z_1|y_2) = 0.5$$
$$P(y_2|x_2) = 1.0 \qquad P(z_2|y_2) = 0.5$$

Cinvestav

Then

We have the according to a Markov Condition on a DAG

$$P(x, y, z) = P(z|y)\, P(y|x)\, P(x)$$

Which, we have that

- It satisfies the Markov Condition.

Then

We have the according to a Markov Condition on a DAG

$$P(x, y, z) = P(z|y)\, P(y|x)\, P(x)$$

Which, we have that

- It satisfies the Markov Condition.

Outline

Cinvestav

Causality in Bayesian Networks

Definition of a Cause

The one, such as a person, an event, or a condition, that is responsible for an action or a result.

Causality in Bayesian Networks

Definition of a Cause

The one, such as a person, an event, or a condition, that is responsible for an action or a result.

However

- Although useful, this simple definition is certainly not the last word on the concept of causation.
  - Actually Philosophers are still wrangling the issue!!!

Causality in Bayesian Networks

Definition of a Cause

The one, such as a person, an event, or a condition, that is responsible for an action or a result.

However

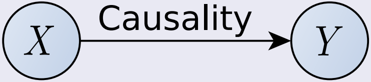
- Although useful, this simple definition is certainly not the last word on the concept of causation.
  - Actually Philosophers are still wrangling the issue!!!

Cinvestav

Causality in Bayesian Networks

Nevertheless, It sheds light in the issue

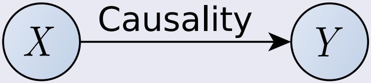
- If the action of making variable X take some value sometimes changes the value taken by a variable Y.



Causality in Bayesian Networks

Nevertheless, It sheds light in the issue

- If the action of making variable X take some value sometimes changes the value taken by a variable Y.



Here, we assume X is responsible for sometimes changing Y 's value

- Thus, we conclude X is a cause of Y.

Furthermore

Formally

We say we **manipulate** X when we force X to take some value.

- We say X **causes** Y if there is some manipulation of X that leads to a change in the probability distribution of Y.

Cinvestav

Furthermore

Formally

We say we **manipulate** X when we force X to take some value.

- We say X **causes** Y if there is some manipulation of X that leads to a change in the probability distribution of Y.

Thus

We assume causes and their effects are statistically correlated.

Furthermore

Formally

We say we **manipulate** X when we force X to take some value.

- We say X **causes** Y if there is some manipulation of X that leads to a change in the probability distribution of Y.

Thus

We assume causes and their effects are statistically correlated.

However

Variables can be correlated without one causing the other.

Furthermore

Formally

We say we **manipulate** X when we force X to take some value.

- We say X **causes** Y if there is some manipulation of X that leads to a change in the probability distribution of Y.

Thus

We assume causes and their effects are statistically correlated.

However

Variables can be correlated without one causing the other.

Outline

Cinvestav

Precautionary Tale: Causality and Bayesian Networks

Important

Not every Bayesian Networks describes causal relationships between the variables.

Precautionary Tale: Causality and Bayesian Networks

Important

Not every Bayesian Networks describes causal relationships between the variables.

Consider

- Consider the dependence between Lung Cancer, L, and the X-ray test, X.
- By focusing on just these variables we might be tempted to represent them by the following Bayesian Networks.

Precautionary Tale: Causality and Bayesian Networks

Important

Not every Bayesian Networks describes causal relationships between the variables.

Consider

- Consider the dependence between Lung Cancer, L, and the X-ray test, X.
- By focusing on just these variables we might be tempted to represent them by the following Bayesian Networks.

Precautionary Tale: Causality and Bayesian Networks

Important

Not every Bayesian Networks describes causal relationships between the variables.

Consider

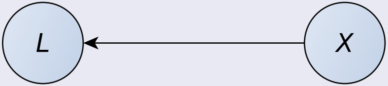
- Consider the dependence between Lung Cancer, L, and the X-ray test, X.
- By focusing on just these variables we might be tempted to represent them by the following Bayesian Networks.

$$P(l_1) = 0.001 \qquad L \longrightarrow X \qquad \begin{array}{l} P(x_1|l_1) = 0.6 \\ P(x_1|l_2) = 0.02 \end{array}$$

Precautionary Tale: Causality and Bayesian Networks

However, we can try the same

$$P(l_1|x_1) = 0.02915$$
$$P(l_1|x_2) = 0.00041$$



$$P(x_1) = 0.02058$$

Remark

Be Careful

- It is tempting to think that Bayesian Networks can be created by creating a DAG where the edges represent direct causal relationships between the variables.

Outline

Cinvestav

However

Causal DAG

- Given a set of variables V, if for every $X, Y \in V$ we draw an edge from X to Y $\iff$ X is a direct cause of Y relative to V, we call the resultant DAG a **causal DAG**.

However

Causal DAG

- Given a set of variables V, if for every $X, Y \in V$ we draw an edge from X to Y $\iff$ X is a direct cause of Y relative to V, we call the resultant DAG a **causal DAG**.

We want

- If we create a causal DAG $G = (V, E)$ and assume the probability distribution of the variables in V satisfies the Markov condition with G:
  - we say we are making the causal **Markov assumption**.

However

Causal DAG

- Given a set of variables V, if for every $X, Y \in V$ we draw an edge from X to Y $\iff$ X is a direct cause of Y relative to V, we call the resultant DAG a **causal DAG**.

We want

- If we create a causal DAG $G = (V, E)$ and assume the probability distribution of the variables in V satisfies the Markov condition with G:
  - we say we are making the causal **Markov assumption**.

In General

- The Markov condition holds for a causal DAG.

Cinvestav

However

Causal DAG

- Given a set of variables V, if for every $X, Y \in V$ we draw an edge from X to Y $\iff$ X is a direct cause of Y relative to V, we call the resultant DAG a **causal DAG**.

We want

- If we create a causal DAG $G = (V, E)$ and assume the probability distribution of the variables in V satisfies the Markov condition with G:
  - we say we are making the causal **Markov assumption**.

In General

- The Markov condition holds for a causal DAG.

However, we still want to know if the Markov Condition Holds

Remark

There are several thing that the DAG needs to have in order to have the Markov Condition.

However, we still want to know if the Markov Condition Holds

Remark

There are several thing that the DAG needs to have in order to have the Markov Condition.

Examples of those

- Common Causes
- Common Effects

However, we still want to know if the Markov Condition Holds

Remark

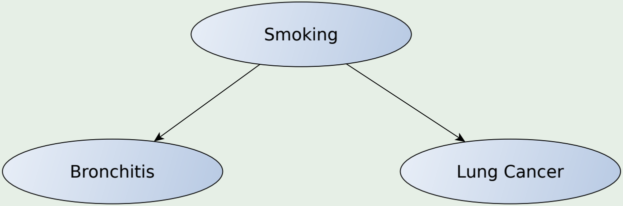
There are several thing that the DAG needs to have in order to have the Markov Condition.

Examples of those

- Common Causes
- Common Effects

How to have a Markov Assumption : Common Causes

Consider

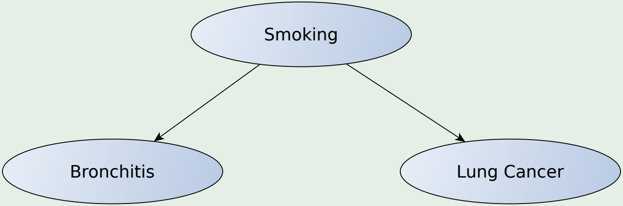


Markov condition

$$I_p(\{B\},\{L\}\mid\{S\}) \Rightarrow P(B|L,S)=P(B|S) \quad (12)$$

How to have a Markov Assumption : Common Causes

Consider



Markov condition

$$I_p\left(\{B\},\{L\} \mid \{S\}\right) \Rightarrow P(B|L,S) = P(B|S) \tag{12}$$

How to have a Markov Assumption : Common Causes

If we know the causal relationships

$$S \rightarrow B \text{ and } S \rightarrow L \tag{13}$$

Now!!!

- If we know that you smoke...

How to have a Markov Assumption : Common Causes

If we know the causal relationships

$$S \to B \text{ and } S \to L \tag{13}$$

Now!!!

- If we know that you smoke...

How to have a Markov Assumption : Common Causes

Then, because of the blocking of information from smoking

- Finding out that Bronchitis will not give us any more information about the probability of having Lung Cancer.

Markov condition

- It is satisfied!!!

How to have a Markov Assumption : Common Causes

Then, because of the blocking of information from smoking

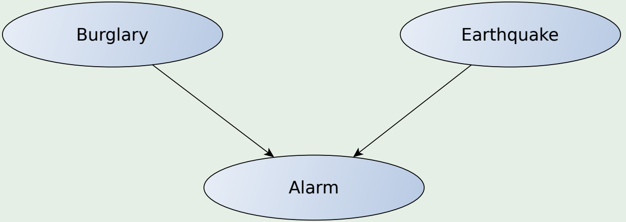
- Finding out that Bronchitis will not give us any more information about the probability of having Lung Cancer.

Markov condition

- It is satisfied!!!

How to have a Markov Assumption : Common Effects

Consider



Markov Condition

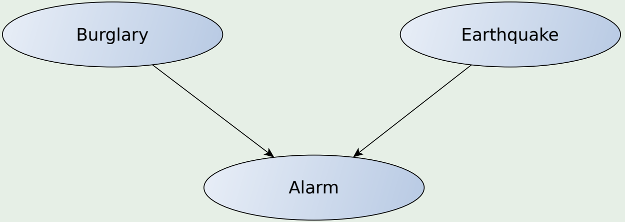
$$I_p(B,W) \Rightarrow P(B|E) = P(B) \quad (14)$$

Thus

We would expect Raining and Ballgame to be independent of each other which is in agreement with the Markov condition.

How to have a Markov Assumption : Common Effects

Consider



Markov Condition

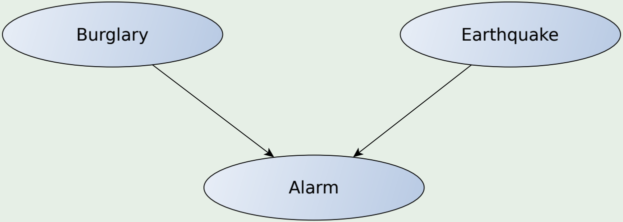
$$l_p\left(B, W\right) \Rightarrow P(B|E) = P(B) \tag{14}$$

Thus

We would expect Raining and Ballgame to be independent of each other which is in agreement with the Markov condition.

How to have a Markov Assumption : Common Effects

Consider



Markov Condition

$$l_p\left(B, W\right) \Rightarrow P(B|E) = P(B) \tag{14}$$

Thus

We would expect Raining and Ballgame to be independent of each other which is in agreement with the Markov condition.

How to have a Markov Assumption : Common Effects

However

We would, however expect them to be conditionally dependent given Alarm.

Thus

If the alarm has gone off, news that there had been an earthquake would 'explain away' the idea that a burglary had taken place.

Then

Again in agreement with the Markov condition.

How to have a Markov Assumption : Common Effects

However

We would, however expect them to be conditionally dependent given Alarm.

Thus

If the alarm has gone off, news that there had been an earthquake would 'explain away' the idea that a burglary had taken place.

Then

Again in agreement with the Markov condition.

Cinvestav

How to have a Markov Assumption : Common Effects

However

We would, however expect them to be conditionally dependent given Alarm.

Thus

If the alarm has gone off, news that there had been an earthquake would ‘explain away’ the idea that a burglary had taken place.

Then

Again in agreement with the Markov condition.

Outline

Cinvestav

The Causal Markov Condition

What do we want?

The basic idea is that the Markov condition holds for a causal DAG.

Rules to construct A Causal Graph

Conditions

1. There must be no hidden common causes.
2. There must not be selection bias.
3. There must be no feedback loops.

Rules to construct A Causal Graph

Conditions

1. There must be no hidden common causes.
2. There must not be selection bias.
3. There must be no feedback loops.

Observations

- Even with these there is a lot of controversy as to its validity.
- It seems to be false in quantum mechanical.

Rules to construct A Causal Graph

Conditions

1. There must be no hidden common causes.
2. There must not be selection bias.
3. There must be no feedback loops.

Observations

- Even with these there is a lot of controversy as to its validity.
- It seems to be false in quantum mechanical.

Rules to construct A Causal Graph

Conditions

1. There must be no hidden common causes.
2. There must not be selection bias.
3. There must be no feedback loops.

Observations

- Even with these there is a lot of controversy as to its validity.
- It seems to be false in quantum mechanical.

Cinvestav

Rules to construct A Causal Graph

Conditions

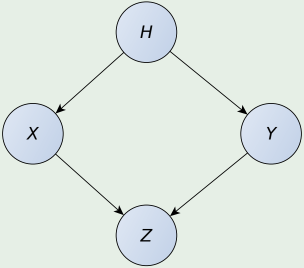
1. There must be no hidden common causes.
2. There must not be selection bias.
3. There must be no feedback loops.

Observations

- Even with these there is a lot of controversy as to its validity.
- It seems to be false in quantum mechanical.

Hidden Common Causes?

Consider the following DAG

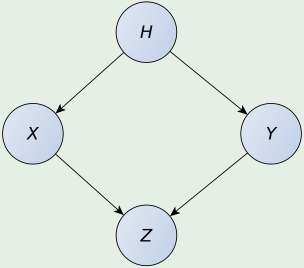


Something Notable

- If a DAG is created on the basis of causal relationships between the variables under consideration,
  - Then X and Y would be marginally independent according to the Markov condition,
  - If Information is given to $H = h_i$

Hidden Common Causes?

Consider the following DAG

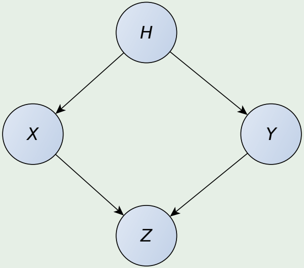


Something Notable

- If a DAG is created on the basis of causal relationships between the variables under consideration,
  - Then X and Y would be marginally independent according to the Markov condition.
  - If Information is given to $H = h_i$

Hidden Common Causes?

Consider the following DAG

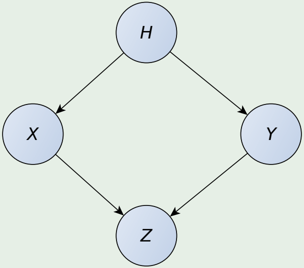


However

- If H is hidden, they will normally be dependent.

Hidden Common Causes?

Consider the following DAG



However

- If H is hidden, they will normally be dependent.

Outline

Cinvestav

Inference in Bayesian Networks

What do we want from Bayesian Networks?

The main point of Bayesian Networkss is to enable probabilistic inference to be performed.

Inference in Bayesian Networks

What do we want from Bayesian Networks?

The main point of Bayesian Networkss is to enable probabilistic inference to be performed.

Two different types of inferences

1. Belief Updating.
2. Abduction Inference.

Inference in Bayesian Networks

What do we want from Bayesian Networks?

The main point of Bayesian Networkss is to enable probabilistic inference to be performed.

Two different types of inferences

1. Belief Updating.
2. Abduction Inference.

Inference in Bayesian Networks

Belief updating

It is used to obtain the posterior probability of one or more variables given evidence concerning the values of other variables.

Abductive inference

It finds the most probable configuration of a set of variables (hypothesis) given certain evidence.

Inference in Bayesian Networks

Belief updating

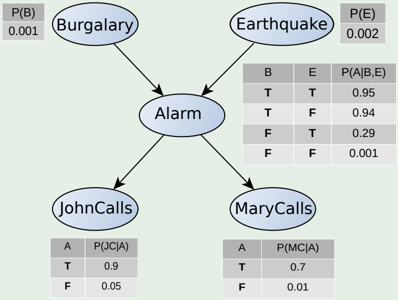
It is used to obtain the posterior probability of one or more variables given evidence concerning the values of other variables.

Abductive inference

It finds the most probable configuration of a set of variables (hypothesis) given certain evidence.

Using the Structure I

Consider the following Bayesian Networks



P(B)
0.001

P(E)
0.002

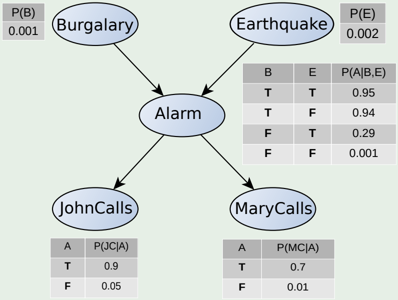
B	E	P(A|B,E)
T	**T**	0.95
T	**F**	0.94
F	**T**	0.29
F	**F**	0.001

A	P(JC|A)
T	0.9
F	0.05

A	P(MC|A)
T	0.7
F	0.01

Using the Structure I

Consider the following Bayesian Networks



P(B)
0.001

P(E)
0.002

B	E	P(A|B,E)
T	**T**	0.95
T	**F**	0.94
F	**T**	0.29
F	**F**	0.001

A	P(JC|A)
T	0.9
F	0.05

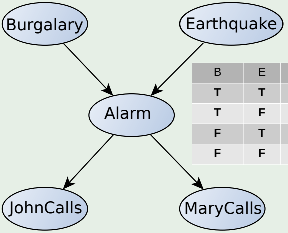
A	P(MC|A)
T	0.7
F	0.01

Consider answering a query in a Bayesian Network

- $Q=$ set of query variables
- $e=$ evidence (set of instantiated variable-value pairs)
- Inference = computation of conditional distribution $P(Q|e)$

Using the Structure I

Consider the following Bayesian Networks



P(B)
0.001

P(E)
0.002

B	E	P(A|B,E)
T	**T**	0.95
T	**F**	0.94
F	**T**	0.29
F	**F**	0.001

A	P(JC|A)
T	0.9
F	0.05

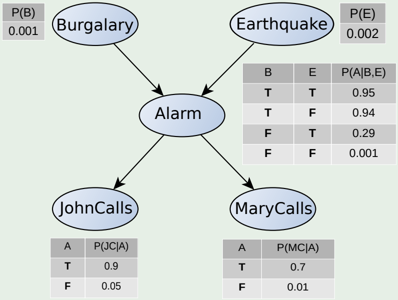
A	P(MC|A)
T	0.7
F	0.01

Consider answering a query in a Bayesian Network

- $Q=$ set of query variables
- $e=$ evidence (set of instantiated variable-value pairs)
- Inference = computation of conditional distribution $P(Q|e)$

Using the Structure I

Consider the following Bayesian Networks



P(B)
0.001

P(E)
0.002

B	E	P(A|B,E)
T	**T**	0.95
T	**F**	0.94
F	**T**	0.29
F	**F**	0.001

A	P(JC|A)
T	0.9
F	0.05

A	P(MC|A)
T	0.7
F	0.01

Consider answering a query in a Bayesian Network

- $Q=$ set of query variables
- $e=$ evidence (set of instantiated variable-value pairs)
- Inference $=$ computation of conditional distribution $P(Q|e)$

Using the Structure II

Examples

- $P(burglary|alarm)$
- $P(earthquake|JCalls, MCalls)$
- $P(JCalls, MCalls|burglary, earthquake)$

Using the Structure II

Examples

- $P(burglary|alarm)$
- $P(earthquake|JCalls, MCalls)$
- $P(JCalls, MCalls|burglary, earthquake)$

So

Can we use the structure of the Bayesian Network to answer such queries efficiently?

- Note: Generally speaking, complexity is inversely proportional to sparsity of graph

Cinvestav

Using the Structure II

Examples

- $P(burglary|alarm)$
- $P(earthquake|JCalls, MCalls)$
- $P(JCalls, MCalls|burglary, earthquake)$

So

Can we use the structure of the Bayesian Network to answer such queries efficiently?

Answer

YES

- Note: Generally speaking, complexity is inversely proportional to sparsity of graph

Using the Structure II

Examples

- $P(burglary|alarm)$
- $P(earthquake|JCalls, MCalls)$
- $P(JCalls, MCalls|burglary, earthquake)$

So

Can we use the structure of the Bayesian Network to answer such queries efficiently?

Answer

YES

- Note: Generally speaking, complexity is inversely proportional to sparsity of graph

Using the Structure II

Examples

- $P(burglary|alarm)$
- $P(earthquake|JCalls, MCalls)$
- $P(JCalls, MCalls|burglary, earthquake)$

So

Can we use the structure of the Bayesian Network to answer such queries efficiently?

Answer

YES

- Note: Generally speaking, complexity is inversely proportional to sparsity of graph

Using the Structure II

Examples

- $P(burglary|alarm)$
- $P(earthquake|JCalls, MCalls)$
- $P(JCalls, MCalls|burglary, earthquake)$

So

Can we use the structure of the Bayesian Network to answer such queries efficiently?

Answer

YES

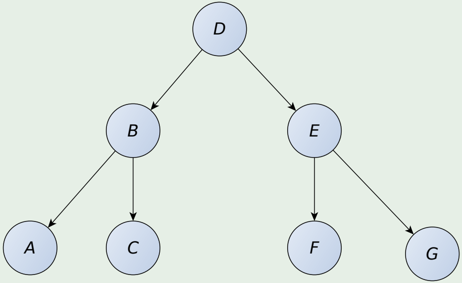
- Note: Generally speaking, complexity is inversely proportional to sparsity of graph

Outline

Cinvestav

Example

DAG



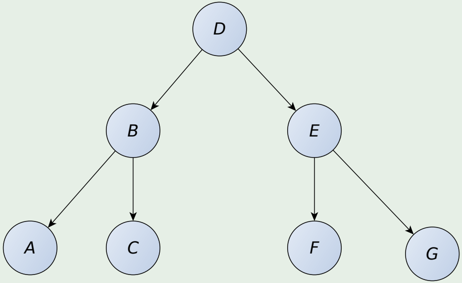
We have the following model

$p(a,b,c,d,e,f,g)$ is modeled by

$$p(a,b,c,d,e,f,g) = p(a|b)\,p(c|b)\,p(f|e)\,p(g|e)\,p(b|d)\,p(e|d)\,p(d)$$

Example

DAG



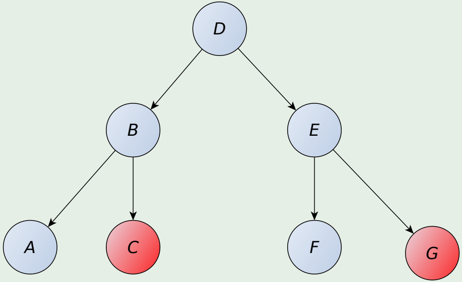
We have the following model

$p(a,b,c,d,e,f,g)$ is modeled by

$$p(a,b,c,d,e,f,g) = p(a|b)\,p(c|b)\,p(f|e)\,p(g|e)\,p(b|d)\,p(e|d)\,p(d)$$

Example

Given values in $C = c$ and $G = g$

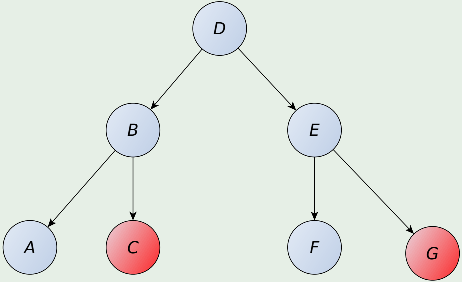


We want to calculate the following

$$p(a|c,g)$$

Example

Given values in $C = c$ and $G = g$

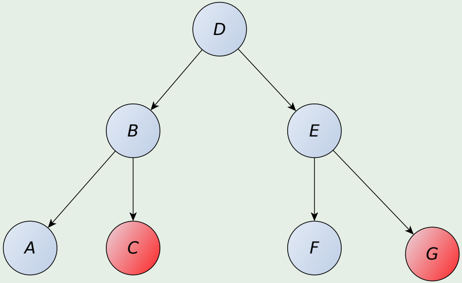


We want to calculate the following

$$p\left(a|c, g\right)$$

Example

Then, if you have brute force approach



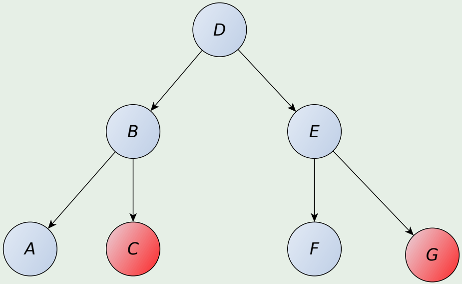
However, a direct calculation leads to use a demarginalization

$$p(a|c,g) = \sum_{b,d,e,f} p(a,b,d,e,f|c,g)$$

- This will require that if we fix the value of a, c and g to have a complexity of $O(m^4)$ with $m = \max\{|B|, |D|, |E|, |F|\}$

Example

Then, if you have brute force approach



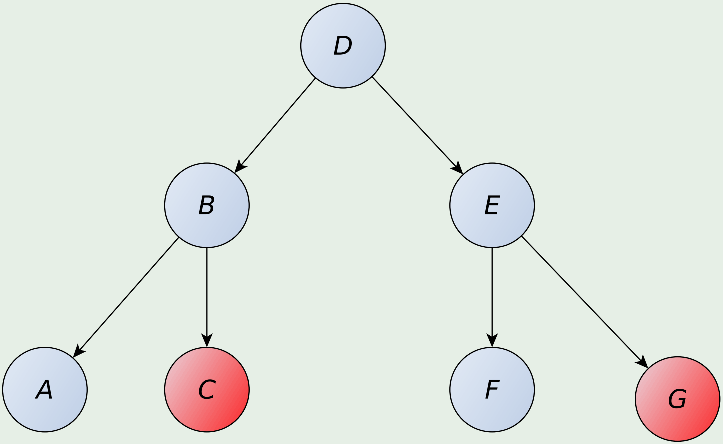
However, a direct calculation leads to use a demarginalization

$$p\left(a|c,g\right)=\sum_{b,d,e,f}p\left(a,b,d,e,f|c,g\right)$$

- This will require that if we fix the value of a, c and g to have a complexity of $O\left(m^{4}\right)$ with $m=\max\left\{\left|B\right|,\left|D\right|,\left|E\right|,\left|F\right|\right\}$

Example

We get some information about $(a = a_i, c = c_i, g = g_i)$



Thus, we have by the Markov Condition

First, we use the chain representation

$$
\begin{aligned}
p\left(a=a_i, b, d, e, f \mid c=c_i, g=g_i\right) = & p\left(a=a_i \mid b, d, e, f, c=c_i, g=g_i\right) \times \cdots \\
& \ldots p\left(b \mid d, e, f, c=c_i, g=g_i\right) \times \cdots \\
& \ldots p\left(d \mid e, f, c=c_i, g=g_i\right) \times \cdots \\
& \ldots \left(e \mid f, c=c_i, g=g_i\right) \times \cdots \\
& \ldots p\left(f \mid c=c_i, g=g_i\right) \times \cdots \\
& \ldots p\left(c=c_i \mid g=g_i\right) \times p\left(g=g_i\right)
\end{aligned}
$$

Thus, we have by the Markov Condition

First, we use the chain representation

$$
\begin{aligned}
p\left(a=a_i, b, d, e, f \mid c=c_i, g=g_i\right) = & \; p\left(a=a_i \mid b, d, e, f, c=c_i, g=g_i\right) \times \cdots \\
& \ldots p\left(b \mid d, e, f, c=c_i, g=g_i\right) \times \cdots \\
& \ldots p\left(d \mid e, f, c=c_i, g=g_i\right) \times \cdots \\
& \ldots \left(e \mid f, c=c_i, g=g_i\right) \times \cdots \\
& \ldots p\left(f \mid c=c_i, g=g_i\right) \times \cdots \\
& \ldots p\left(c=c_i \mid g=g_i\right) \times p\left(g=g_i\right)
\end{aligned}
$$

Thus, we have by the Markov Condition

First, we use the chain representation

$$
\begin{aligned}
p\left(a=a_i, b, d, e, f | c=c_i, g=g_i\right) = & p\left(a=a_i | b, d, e, f, c=c_i, g=g_i\right) \times \cdots \\
& ...p\left(b | d, e, f, c=c_i, g=g_i\right) \times \cdots \\
& ...p\left(d | e, f, c=c_i, g=g_i\right) \times \cdots \\
& ...\left(e | f, c=c_i, g=g_i\right) \times \cdots \\
& ...p\left(f | c=c_i, g=g_i\right) \times \cdots \\
& ...p\left(c=c_i | g=g_i\right) \times p\left(g=g_i\right)
\end{aligned}
$$

Thus, we have by the Markov Condition

First, we use the chain representation

$$
\begin{aligned}
p\left(a=a_i, b, d, e, f | c=c_i, g=g_i\right) = & \; p\left(a=a_i | b, d, e, f, c=c_i, g=g_i\right) \times \cdots \\
& ...p\left(b | d, e, f, c=c_i, g=g_i\right) \times \cdots \\
& ...p\left(d | e, f, c=c_i, g=g_i\right) \times \cdots \\
& ...\left(e | f, c=c_i, g=g_i\right) \times \cdots \\
& ...p\left(f | c=c_i, g=g_i\right) \times \cdots \\
& ...p\left(c=c_i | g=g_i\right) \times p\left(g=g_i\right)
\end{aligned}
$$

Thus, we have by the Markov Condition

First, we use the chain representation

$$
\begin{aligned}
p\left(a = a_i, b, d, e, f | c = c_i, g = g_i\right) = & p\left(a = a_i | b, d, e, f, c = c_i, g = g_i\right) \times \cdots \\
& ...p\left(b | d, e, f, c = c_i, g = g_i\right) \times \cdots \\
& ...p\left(d | e, f, c = c_i, g = g_i\right) \times \cdots \\
& ...\left(e | f, c = c_i, g = g_i\right) \times \cdots \\
& ...p\left(f | c = c_i, g = g_i\right) \times \cdots \\
& ...p\left(c = c_i | g = g_i\right) \times p\left(g = g_i\right)
\end{aligned}
$$

Thus, we have by the Markov Condition

First, we use the chain representation

$$
\begin{aligned}
p\left(a=a_i,b,d,e,f|c=c_i,g=g_i\right)=&p\left(a=a_i|b,d,e,f,c=c_i,g=g_i\right)\times\cdots\\
&...p\left(b|d,e,f,c=c_i,g=g_i\right)\times\cdots\\
&...p\left(d|e,f,c=c_i,g=g_i\right)\times\cdots\\
&...\left(e|f,c=c_i,g=g_i\right)\times\cdots\\
&...p\left(f|c=c_i,g=g_i\right)\times\cdots\\
&...p\left(c=c_i|g=g_i\right)\times p\left(g=g_i\right)
\end{aligned}
$$

Then, we have that

Using the DAG structure

$$p\left(a=a_i,b,d,e,f|c=c_i,g=g_i\right) = p\left(a=a_i|b\right)p\left(b|d,c=c_i\right)\times\cdots$$
$$\ldots p\left(d|e\right)p\left(e,f|g=g_i\right)$$

Then given the original sum at the de-margenalization

$$p\left(a=a_i,b,d,e,f|c=c_i,g=g_i\right) = \sum_b p\left(a=a_i|b\right)\sum_d p\left(b|d,c=c_i\right)\times\cdots$$
$$\ldots\sum_e p\left(d|e\right)\sum_f p\left(e,f|g=g_i\right)$$

Then, we have that

Using the DAG structure

$$
\begin{aligned}
p\left(a=a_i,b,d,e,f|c=c_i,g=g_i\right)=&p\left(a=a_i|b\right)p\left(b|d,c=c_i\right)\times\cdots\\
&...p\left(d|e\right)p\left(e,f|g=g_i\right)
\end{aligned}
$$

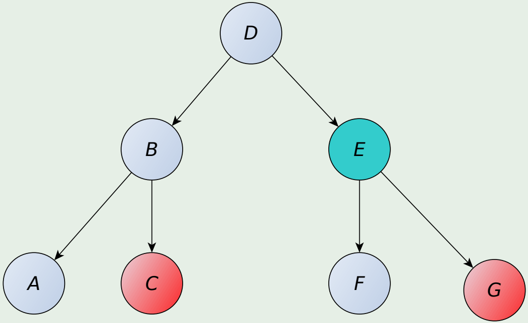
Then given the original sum at the de-margenalization

$$
\begin{aligned}
p\left(a=a_i,b,d,e,f|c=c_i,g=g_i\right)=&\sum_b p\left(a=a_i|b\right)\sum_d p\left(b|d,c=c_i\right)\times\cdots\\
&...\sum_e p\left(d|e\right)\sum_f p\left(e,f|g=g_i\right)
\end{aligned}
$$

Cinvestav

Now, we can concentrate $\sum_f p\left(e, f|g = g_i\right)$

Now, using the relation with respect to E

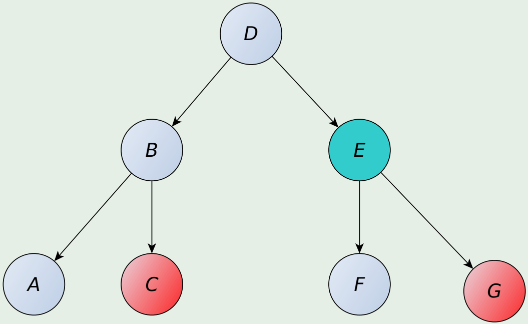


Using this information, we can reduce one of the sums by marginalization

$$\sum_f p\left(e, f|g = g_i\right) = p\left(e|g = g_i\right)$$

Now, we can concentrate $\sum_f p\left(e, f | g = g_i\right)$

Now, using the relation with respect to E



Using this information, we can reduce one of the sums by marginalization

$$\sum_f p\left(e, f | g = g_i\right) = p\left(e | g = g_i\right)$$

How?

Remember that

$$\sum_f p\left(e, f|g=g_i\right) = \sum_f p\left(e|f, g=g_i\right) p\left(f|g=g_i\right)$$

$$= \sum_f pp\left(e|f, g=g_i\right)$$

$$= p\left(e|g=g_i\right)$$

How?

Remember that

$$
\begin{aligned}
\sum_f p\left(e, f | g = g_i\right) &= \sum_f p\left(e | f, g = g_i\right) p\left(f | g = g_i\right) \\
&= \sum_f pp\left(e | f, g = g_i\right) \\
&= p\left(e | g = g_i\right)
\end{aligned}
$$

How?

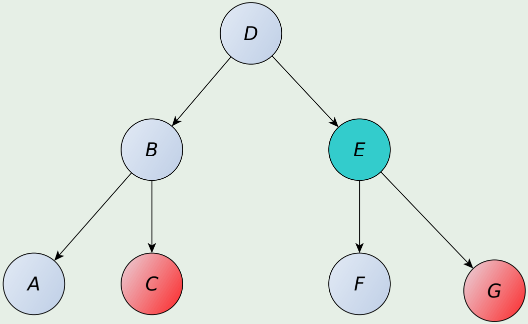
Remember that

$$\begin{aligned}\sum_f p\left(e,f|g=g_i\right) =& \sum_f p\left(e|f,g=g_i\right)p\left(f|g=g_i\right)\\ =& \sum_f pp\left(e|f,g=g_i\right)\\ =& p\left(e|g=g_i\right)\end{aligned}$$

Cinvestav

Then, we have that

DAG

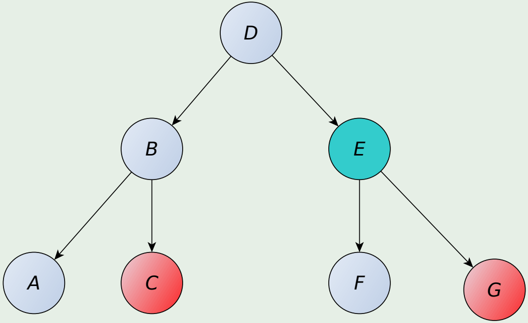


Thus, we can reduce the size of our sum

$$\sum_{b} p\left(a = a_i | b\right) \sum_{d} p\left(b | d, c = c_i\right) \sum_{e} p\left(d | e\right) p\left(e | g = g_i\right)$$

Then, we have that

DAG

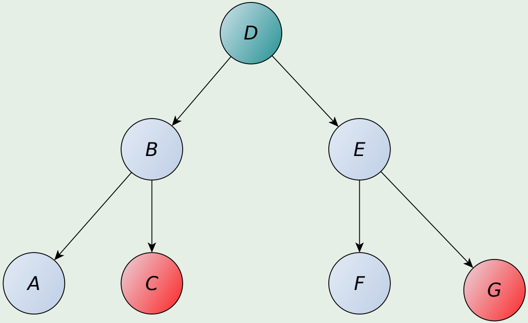


Thus, we can reduce the size of our sum

$$\sum_{b} p\left(a=a_i|b\right) \sum_{d} p\left(b|d, c=c_i\right) \sum_{e} p\left(d|e\right) p\left(e|g=g_i\right)$$

Then, we can use the realtion with respect to D

Given the following DAG

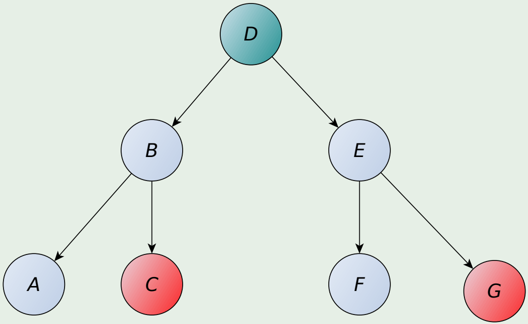


Now, we can calculate the probability of D by using the chain rule

$$p(d|e)\,p(e|g=g_i) = p(d|e,g=g_i)\,p(e|g=g_i) = p(d,e|g=g_i)$$

Then, we can use the realtion with respect to D

Given the following DAG

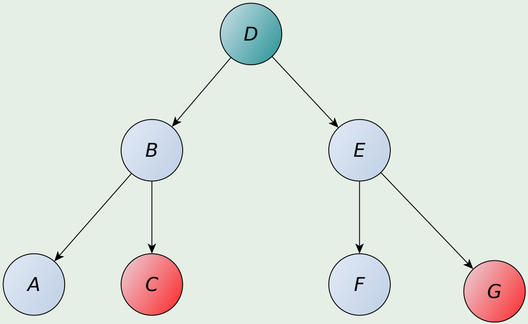


Now, we can calculate the probability of D by using the chain rule

$$p\left(d|e\right)p\left(e|g=g_i\right)=p\left(d|e,g=g_i\right)p\left(e|g=g_i\right)=p\left(d,e|g=g_i\right)$$

Example

DAG

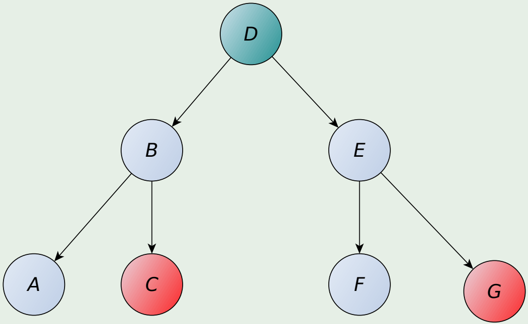


Now, we can calculate the probability of D by using the chain rule

$$\sum_{b} p\left(a = a_i | b\right) \sum_{d} p\left(b | d, c = c_i\right) \sum_{e} p\left(d, e | g = g_i\right)$$

Example

DAG

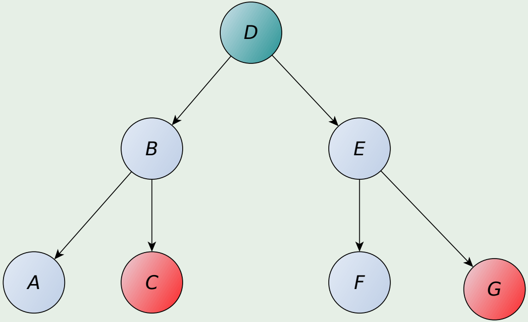


Now, we can calculate the probability of D by using the chain rule

$$\sum_b p\left(a = a_i | b\right) \sum_d p\left(b | d, c = c_i\right) \sum_e p\left(d, e | g = g_i\right)$$

Example

DAG

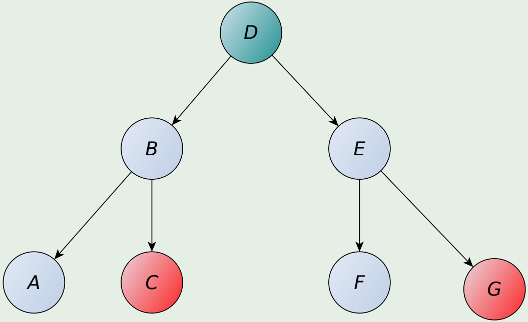


Now, we sum over all possible values of E

$$\sum_{e} p\left(d, e | g = g_i\right) = p\left(d | g = g_i\right)$$

Example

DAG

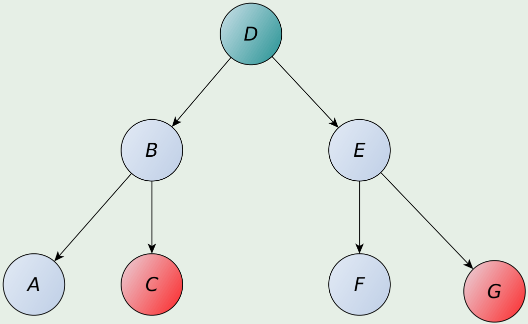


Now, we sum over all possible values of E

$$\sum_{e} p\left(d, e|g = g_i\right) = p\left(d|g = g_i\right)$$

Example

DAG

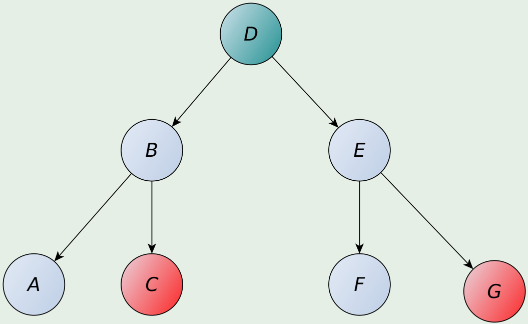


We get the following

$$\sum_b p(a = a_i|b) \sum_d p(b|d, c = c_i)\, p(d|g = g_i)$$

Example

DAG

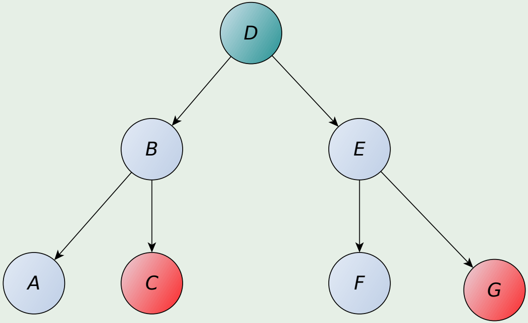


We get the following

$$\sum_{b} p\left(a = a_i | b\right) \sum_{d} p\left(b | d, c = c_i\right) p\left(d | g = g_i\right)$$

Example

DAG

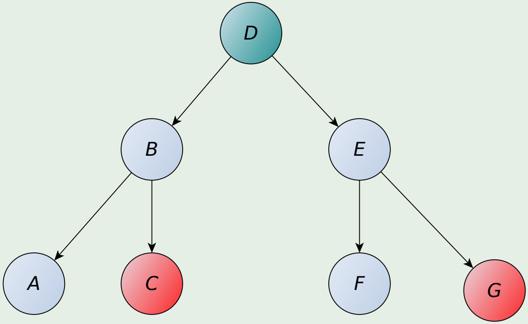


Again the chain rule for D

$$p\left(b|d, c=c_i\right) p\left(d|g=g_i\right) = p\left(b|d, c=c_i, g=g_i\right) p\left(d|c=c_i, g=g_i\right)$$
$$= p\left(b, d|c=c_i, g=g_i\right)$$

Example

DAG

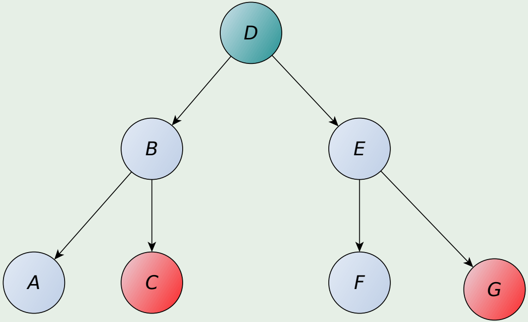


Again the chain rule for D

$$p(b|d, c = c_i)\, p(d|g = g_i) = p(b|d, c = c_i, g = g_i)\, p(d|c = c_i, g = g_i)$$
$$= p(b, d|c = c_i, g = g_i)$$

Example

DAG

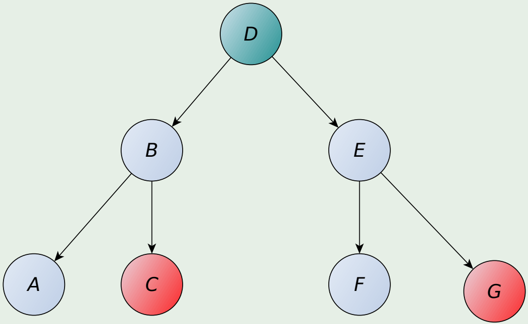


Now, we sum over all possible values of D

$$\sum_{b} p\left(a = a_i | b\right) p\left(b | c = c_1, g = g_1\right)$$

Example

DAG

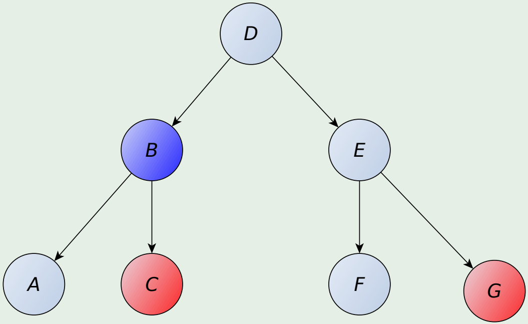


Now, we sum over all possible values of D

$$\sum_b p\left(a = a_i | b\right) p\left(b | c = c_i, g = g_i\right)$$

Example

DAG

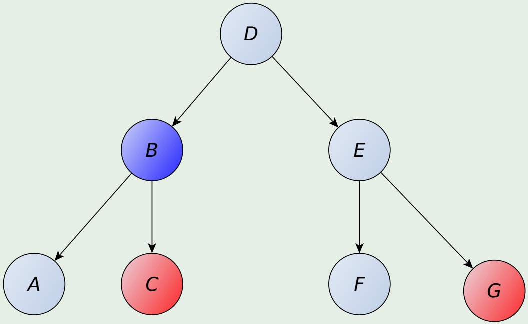


Now, we use the chain rule for reducing again

$$p(a=a_i|b)\,p(b|c=c_i,g=g_i) = p(a=a_i,b|c=c_i,g=g_i)$$

Example

DAG



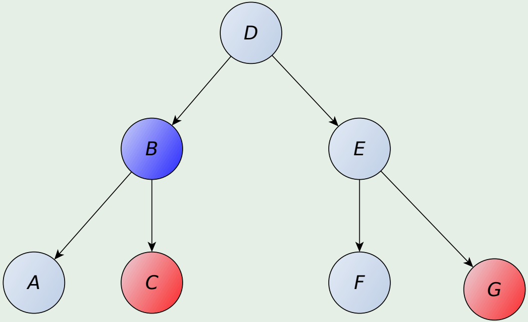
Now, we use the chain rule for reducing again

$$p\left(a=a_i|b\right)p\left(b|c=c_i,g=g_i\right)=p\left(a=a_i,b|c=c_i,g=g_i\right)$$

Cinvestav

Example

DAG

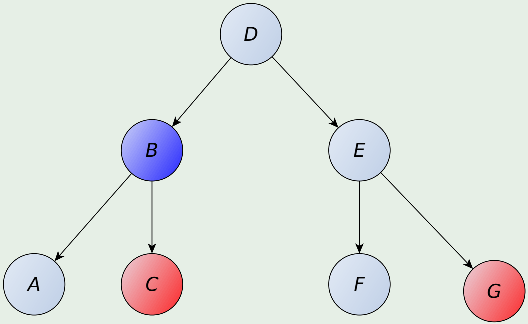


Now, we use the chain rule for reducing again

$$\sum_{b} p\left(a = a_i, b | c = c_i, g = g_i\right) = p\left(a = a_i | c = c_i, g = g_i\right)$$

Example

DAG



Now, we use the chain rule for reducing again

$$\sum_{b} p\left(a = a_i, b | c = c_i, g = g_i\right) = p\left(a = a_i | c = c_i, g = g_i\right)$$

Complexity

Because this can be computed using a sequence of four for loops

The complexity simply becomes $O(m)$ when compared with $O(m^4)$

Outline

Cinvestav

General Strategy for Inference

Query

- Want to compute $P(q|e)$!!!

Step 1

- $P(q|e) = \frac{P(q,e)}{P(e)} = \alpha P(q,e)$, since $\alpha = P(e)$ is constant wrt Q.

Step 2

- $P(q,e) = \sum_{a..z} P(q,e,a,b,\ldots,z)$, by the law of total probability.

General Strategy for Inference

Query

- Want to compute $P(q|e)$!!!

Step 1

- $P(q|e) = \frac{P(q,e)}{P(e)} = aP(q,e)$, since $a = P(e)$ is constant wrt Q.

Step 2

- $P(q,e) = \sum_{a..z} P(q,e,a,b,\ldots,z)$, by the law of total probability.

General Strategy for Inference

Query

- Want to compute $P(q|e)$!!!

Step 1

- $P(q|e) = \frac{P(q,e)}{P(e)} = aP(q,e)$, since $a = P(e)$ is constant wrt Q.

Step 2

- $P(q,e) = \sum_{a..z} P(q,e,a,b,\ldots.z)$, by the law of total probability.

General Strategy for inference

Step 3

- $\sum_{a..z} P(q, e, a, b, \ldots .z) = \sum_{a..z} \prod P(\text{variable } i \mid \text{parents } i)$ (using Bayesian network factoring)

Step 4

- Distribute summations across product terms for efficient computation.

General Strategy for inference

Step 3

- $\sum_{a..z} P(q, e, a, b, \ldots .z) = \sum_{a..z} \prod P(\text{variable } i \mid \text{parents } i)$ (using Bayesian network factoring)

Step 4

- Distribute summations across product terms for efficient computation.

Outline

Inference – An Overview

Case 1

- Trees and singly connected networks – only one path between any two nodes:
  - Message passing (Pearl, 1988)

Inference – An Overview

Case 1

- Trees and singly connected networks – only one path between any two nodes:
  - Message passing (Pearl, 1988)

Case 2

- Multiply connected networks:
  - A range of algorithms including cut-set conditioning (Pearl, 1988), junction tree propagation (Lauritzen and Spiegelhalter, 1988), bucket elimination (Dechter, 1996) to mention a few.
  - A range of algorithms for approximate inference.

- Here the focus will be on message passing and junction tree propagation for discrete variables.

Inference – An Overview

Case 1

- Trees and singly connected networks – only one path between any two nodes:
  - Message passing (Pearl, 1988)

Case 2

- Multiply connected networks:
  - A range of algorithms including cut-set conditioning (Pearl, 1988), junction tree propagation (Lauritzen and Spiegelhalter, 1988), bucket elimination (Dechter, 1996) to mention a few.
  - A range of algorithms for approximate inference.

Notes

- Both exact and approximate inference are NP-hard in the worst case.
- Here the focus will be on message passing and junction tree propagation for discrete variables.

Inference – An Overview

Case 1

- Trees and singly connected networks – only one path between any two nodes:
  - Message passing (Pearl, 1988)

Case 2

- Multiply connected networks:
  - A range of algorithms including cut-set conditioning (Pearl, 1988), junction tree propagation (Lauritzen and Spiegelhalter, 1988), bucket elimination (Dechter, 1996) to mention a few.
  - A range of algorithms for approximate inference.

Notes

- Both exact and approximate inference are NP-hard in the worst case.
- Here the focus will be on message passing and junction tree propagation for discrete variables.

Inference – An Overview

Case 1

- Trees and singly connected networks – only one path between any two nodes:
  - Message passing (Pearl, 1988)

Case 2

- Multiply connected networks:
  - A range of algorithms including cut-set conditioning (Pearl, 1988), junction tree propagation (Lauritzen and Spiegelhalter, 1988), bucket elimination (Dechter, 1996) to mention a few.
  - A range of algorithms for approximate inference.

Notes

- Both exact and approximate inference are NP-hard in the worst case.
- Here the focus will be on message passing and junction tree propagation for discrete variables.

Inference – An Overview

Case 1

- Trees and singly connected networks – only one path between any two nodes:
  - Message passing (Pearl, 1988)

Case 2

- Multiply connected networks:
  - A range of algorithms including cut-set conditioning (Pearl, 1988), junction tree propagation (Lauritzen and Spiegelhalter, 1988), bucket elimination (Dechter, 1996) to mention a few.
  - A range of algorithms for approximate inference.

Notes

- Both exact and approximate inference are NP-hard in the worst case.
- Here the focus will be on message passing and junction tree propagation for discrete variables.

Inference – An Overview

Case 1

- Trees and singly connected networks – only one path between any two nodes:
  - Message passing (Pearl, 1988)

Case 2

- Multiply connected networks:
  - A range of algorithms including cut-set conditioning (Pearl, 1988), junction tree propagation (Lauritzen and Spiegelhalter, 1988), bucket elimination (Dechter, 1996) to mention a few.
  - A range of algorithms for approximate inference.

Notes

- Both exact and approximate inference are NP-hard in the worst case.
- Here the focus will be on message passing and junction tree propagation for discrete variables.