Machine Intelligence

5. Reasoning Under Uncertainty, Part II: Bayesian Networks

How to Organize Your Knowledge

Álvaro Torralba

AALBORG UNIVERSITET

Fall 2022

Thanks to Thomas D. Nielsen and Jörg Hoffmann for slide sources

Agenda

Our Agenda for This Topic

→ Our treatment of the topic "Probabilistic Reasoning" consists of Chapters 4-6.

- **Chapter 4**: All the basic machinery at use in Bayesian networks.

  → Sets up the framework and basic operations.

- **This Chapter**: Bayesian networks: What they are and how to build them.

  → The most wide-spread and successful practical framework for probabilistic reasoning.

- **Chapter 6**: Bayesian networks: how to use them.

  → How to use Bayesian Networks to answer our questions.

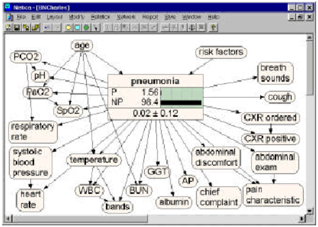
Some Applications

→ A ubiquitous problem: Observe "symptoms", need to infer "causes".

Some Applications

→ A ubiquitous problem: Observe "symptoms", need to infer "causes".

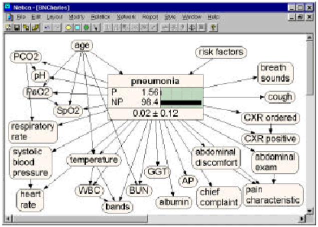
Medical Diagnosis



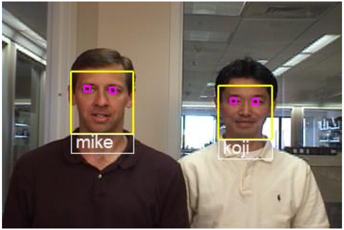
Some Applications

→ A ubiquituous problem: Observe "symptoms", need to infer "causes".

Medical Diagnosis



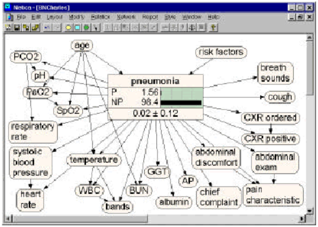
Face Recognition



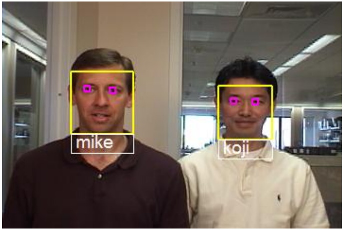
Some Applications

→ A ubiquituous problem: Observe "symptoms", need to infer "causes".

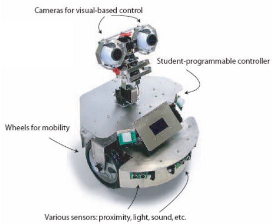
Medical Diagnosis



Face Recognition



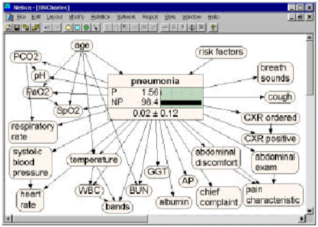
Self-Localization



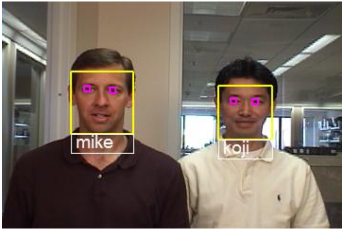
Some Applications

→ A ubiquituous problem: Observe "symptoms", need to infer "causes".

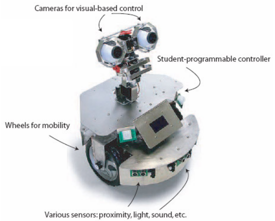
Medical Diagnosis



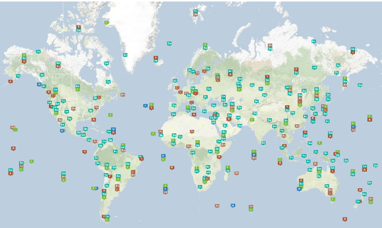
Face Recognition



Self-Localization



Nuclear Test Ban



Our Agenda for This Chapter

- **Recap: Conditional Independence**

  → A brief recap on the main notion exploited by Bayesian Networks.

Our Agenda for This Chapter

- **Recap: Conditional Independence**

  → A brief recap on the main notion exploited by Bayesian Networks.

- **What is a Bayesian Network?** What is the syntax?

  → Tells you what Bayesian networks look like.

Our Agenda for This Chapter

- **Recap: Conditional Independence**

  $\rightarrow$ A brief recap on the main notion exploited by Bayesian Networks.

- **What is a Bayesian Network?** What is the syntax?

  $\rightarrow$ Tells you what Bayesian networks look like.

- **What is the Meaning of a Bayesian Network?** What is the semantics?

  $\rightarrow$ Makes the intuitive meaning precise.

Our Agenda for This Chapter

- **Recap: Conditional Independence**

  → A brief recap on the main notion exploited by Bayesian Networks.

- **What is a Bayesian Network?** What is the syntax?

  → Tells you what Bayesian networks look like.

- **What is the Meaning of a Bayesian Network?** What is the semantics?

  → Makes the intuitive meaning precise.

- **D-Separation**: How evidence is transmitted along the Bayesian Network?

  → Some intuition about how BNs work.

Our Agenda for This Chapter

- **Recap: Conditional Independence**

  → A brief recap on the main notion exploited by Bayesian Networks.

- **What is a Bayesian Network?** What is the syntax?

  → Tells you what Bayesian networks look like.

- **What is the Meaning of a Bayesian Network?** What is the semantics?

  → Makes the intuitive meaning precise.

- **D-Separation**: How evidence is transmitted along the Bayesian Network?

  → Some intuition about how BNs work.

- **Constructing Bayesian Networks:** How do we design these networks? What effect do our choices have on their size?

  → Before you can start doing inference, you need to model your domain.

Our Agenda for This Chapter

- **Recap: Conditional Independence**

  → A brief recap on the main notion exploited by Bayesian Networks.

- **What is a Bayesian Network?** What is the syntax?

  → Tells you what Bayesian networks look like.

- **What is the Meaning of a Bayesian Network?** What is the semantics?

  → Makes the intuitive meaning precise.

- **D-Separation**: How evidence is transmitted along the Bayesian Network?

  → Some intuition about how BNs work.

- **Constructing Bayesian Networks:** How do we design these networks? What effect do our choices have on their size?

  → Before you can start doing inference, you need to model your domain.

- **Mediating Variables:** How to introduce mediating variables to make the network precise and compact?

  → More advance notions on how to construct Bayesian Networks.

Our Agenda for This Chapter

- **Recap: Conditional Independence**

  → A brief recap on the main notion exploited by Bayesian Networks.

- **What is a Bayesian Network?** What is the syntax?

  → Tells you what Bayesian networks look like.

- **What is the Meaning of a Bayesian Network?** What is the semantics?

  → Makes the intuitive meaning precise.

- **D-Separation**: How evidence is transmitted along the Bayesian Network?

  → Some intuition about how BNs work.

- **Constructing Bayesian Networks:** How do we design these networks? What effect do our choices have on their size?

  → Before you can start doing inference, you need to model your domain.

- **Mediating Variables:** How to introduce mediating variables to make the network precise and compact?

  → More advance notions on how to construct Bayesian Networks.

- **Inference in Bayesian Networks:** Next Chapter

Agenda

Compact Specifications by Independence

The variables $A_1, \dots, A_k$ and $B_1, \dots, B_m$ are independent if

$$P(A_1, \dots, A_k \mid B_1, \dots, B_m) = P(A_1, \dots, A_k)$$

This is equivalent to:

$$P(A_1, \dots, A_k, B_1, \dots, B_m) = P(A_1, \dots, A_k) \cdot P(B_1, \dots, B_m)$$

Compact Specifications by Independence

The variables $A_1, \dots, A_k$ and $B_1, \dots, B_m$ are independent if

$$P(A_1, \dots, A_k \mid B_1, \dots, B_m) = P(A_1, \dots, A_k)$$

This is equivalent to:

$$P(A_1, \dots, A_k, B_1, \dots, B_m) = P(A_1, \dots, A_k) \cdot P(B_1, \dots, B_m)$$

→ Independence can be exploited to represent the full joint probability distribution more compactly.

$M =$	$F =$ W	D	L	$P(M)$
W				.5625
D	M and F are independent			.25
L				.1875
$P(F)$	.3281	.2656	.4062	

The probability for each possible world then is defined, e.g.

$$P(M = D, F = L) = 0.25 \cdot 0.4062 = 0.10155$$

→ Usually, random variables are independent only under particular conditions: conditional independence, see next section.

Questionnaire

$$\begin{array}{ll} \textit{Hair length}: & D_{\textit{Hair length}} = \{\textit{long}, \textit{short}\} \\ \textit{Height}: & D_{\textit{Height}} = \{\textit{tall}, \textit{medium}\} \end{array}$$

Question!

Are Hair length and Height independent?

(A): Yes

(B): No

Questionnaire

$$\begin{array}{ll} \textit{Hair length}: & D_{\textit{Hair length}} = \{\textit{long}, \textit{short}\} \\ \textit{Height}: & D_{\textit{Height}} = \{\textit{tall}, \textit{medium}\} \\ \textit{Sex}: & D_{\textit{Sex}} = \{\textit{male}, \textit{female}\} \end{array}$$

Question!

Are Hair length and Height independent?

(A): Yes (B): No

Joint Distribution:

	Sex			
	male		*female*	
	Hair length		*Hair length*	
Height	*long*	*short*	*long*	*short*
tall	0.06	0.24	0.07	0.03
medium	0.04	0.16	0.28	0.12

Questionnaire

$$\begin{array}{ll} \textit{Hair length}: & D_{\textit{Hair length}} = \{\textit{long}, \textit{short}\} \\ \textit{Height}: & D_{\textit{Height}} = \{\textit{tall}, \textit{medium}\} \\ \textit{Sex}: & D_{\textit{Sex}} = \{\textit{male}, \textit{female}\} \end{array}$$

Question!

Are Hair length and Height independent?

(A): Yes (B): No

Joint Distribution:

	Sex			
	male		*female*	
	Hair length		*Hair length*	
Height	*long*	*short*	*long*	*short*
tall	0.06	0.24	0.07	0.03
medium	0.04	0.16	0.28	0.12

$P(\textit{Hair length}, \textit{Height})$ $P(\textit{Height})$, $P(\textit{Height} \mid \textit{Hair length})$:

	Hair length		
Height	*long*	*short*	
tall	0.13	0.27	0.4
	0.289	0.49	
medium	0.32	0.28	0.6
	0.711	0.51	

$\leadsto$ No, *Hair length* and *Height* are not independent. Tall people have shorter hair!
Does that sound right to you?

Conditional Independence

Definition. *Given sets of random variables* $\mathbf{Z}_1$, $\mathbf{Z}_2$, $\mathbf{Z}$, *we say that* $\mathbf{Z}_1$ *and* $\mathbf{Z}_2$ *are conditionally independent given* $\mathbf{Z}$ *if:*

$$\mathbf{P}(\mathbf{Z}_1, \mathbf{Z}_2 \mid \mathbf{Z}) = \mathbf{P}(\mathbf{Z}_1 \mid \mathbf{Z})\mathbf{P}(\mathbf{Z}_2 \mid \mathbf{Z})$$

We alternatively say that $\mathbf{Z}_1$ *is conditionally independent of* $\mathbf{Z}_2$ *given* $\mathbf{Z}$*.*

Conditional Independence

Definition. *Given sets of random variables* $\mathbf{Z}_1$, $\mathbf{Z}_2$, $\mathbf{Z}$, *we say that* $\mathbf{Z}_1$ *and* $\mathbf{Z}_2$ *are conditionally independent given* $\mathbf{Z}$ *if:*

$$\mathbf{P}(\mathbf{Z}_1, \mathbf{Z}_2 \mid \mathbf{Z}) = \mathbf{P}(\mathbf{Z}_1 \mid \mathbf{Z})\mathbf{P}(\mathbf{Z}_2 \mid \mathbf{Z})$$

We alternatively say that $\mathbf{Z}_1$ *is conditionally independent of* $\mathbf{Z}_2$ *given* $\mathbf{Z}$.

$P(\textit{Hair length}, \textit{Height} \mid \textit{Sex} = \textit{female})$, $P(\textit{Height} \mid \textit{Sex} = \textit{female})$,
$P(\textit{Height} \mid \textit{Hair length}, \textit{Sex} = \textit{female})$:

	Hair length		
Height	*long*	*short*	
tall	0.14	0.06	0.2
	0.2	0.2	
medium	0.56	0.24	0.8
	0.8	0.8	

$\rightsquigarrow$ *Hair length* and *Height* are independent given *Sex*=*female*.

Also given *Sex*=*male*, so *Hair length* and *Height* are independent given *Sex*.

Conditional Independence

Definition. *Given sets of random variables* $\mathbf{Z}_1$, $\mathbf{Z}_2$, $\mathbf{Z}$, *we say that* $\mathbf{Z}_1$ *and* $\mathbf{Z}_2$ *are conditionally independent given* $\mathbf{Z}$ *if:*

$$\mathbf{P}(\mathbf{Z}_1, \mathbf{Z}_2 \mid \mathbf{Z}) = \mathbf{P}(\mathbf{Z}_1 \mid \mathbf{Z})\mathbf{P}(\mathbf{Z}_2 \mid \mathbf{Z})$$

We alternatively say that $\mathbf{Z}_1$ *is conditionally independent of* $\mathbf{Z}_2$ *given* $\mathbf{Z}$.

$P(\textit{Hair length}, \textit{Height} \mid \textit{Sex} = \textit{female})$, $P(\textit{Height} \mid \textit{Sex} = \textit{female})$,
$P(\textit{Height} \mid \textit{Hair length}, \textit{Sex} = \textit{female})$:

	Hair length		
Height	*long*	*short*	
tall	0.14	0.06	0.2
	0.2	0.2	
medium	0.56	0.24	0.8
	0.8	0.8	

$\rightsquigarrow$ *Hair length* and *Height* are independent given *Sex=female*.

Also given *Sex=male*, so *Hair length* and *Height* are independent given *Sex*.

Proposition. *If* $\mathbf{Z}_1$ *and* $\mathbf{Z}_2$ *are conditionally independent given* $\mathbf{Z}$, *then* $\mathbf{P}(\mathbf{Z}_1 \mid \mathbf{Z}_2, \mathbf{Z}) = \mathbf{P}(\mathbf{Z}_1 \mid \mathbf{Z})$.

Conditional Independence

Definition. *Given sets of random variables* $\mathbf{Z}_1$, $\mathbf{Z}_2$, $\mathbf{Z}$, *we say that* $\mathbf{Z}_1$ *and* $\mathbf{Z}_2$ *are conditionally independent given* $\mathbf{Z}$ *if:*

$$\mathbf{P}(\mathbf{Z}_1, \mathbf{Z}_2 \mid \mathbf{Z}) = \mathbf{P}(\mathbf{Z}_1 \mid \mathbf{Z})\mathbf{P}(\mathbf{Z}_2 \mid \mathbf{Z})$$

We alternatively say that $\mathbf{Z}_1$ *is conditionally independent of* $\mathbf{Z}_2$ *given* $\mathbf{Z}$.

$P(\mathit{Hair\ length}, \mathit{Height} \mid \mathit{Sex} = \mathit{female})$, $P(\mathit{Height} \mid \mathit{Sex} = \mathit{female})$,
$P(\mathit{Height} \mid \mathit{Hair\ length}, \mathit{Sex} = \mathit{female})$:

	Hair length		
Height	*long*	*short*	
tall	0.14	0.06	0.2
	0.2	0.2	
medium	0.56	0.24	0.8
	0.8	0.8	

$\rightsquigarrow$ *Hair length* and *Height* are independent given *Sex*=*female*.

Also given *Sex*=*male*, so *Hair length* and *Height* are independent given *Sex*.

Proposition. *If* $\mathbf{Z}_1$ *and* $\mathbf{Z}_2$ *are conditionally independent given* $\mathbf{Z}$, *then* $\mathbf{P}(\mathbf{Z}_1 \mid \mathbf{Z}_2, \mathbf{Z}) = \mathbf{P}(\mathbf{Z}_1 \mid \mathbf{Z})$.

If we already know the Sex of a person, knowing the hair length does not tell us anything about their height!

Agenda

John, Mary, and My Brand-New Alarm

Example

I got very valuable stuff at home. So I bought an alarm.

John, Mary, and My Brand-New Alarm

Example

I got very valuable stuff at home. So I bought an alarm. Unfortunately, the alarm just rings at home, doesn't call me on my mobile.

John, Mary, and My Brand-New Alarm

Example

I got very valuable stuff at home. So I bought an alarm. Unfortunately, the alarm just rings at home, doesn't call me on my mobile. I've got two neighbors, Mary and John, who'll call me if they hear the alarm.

John, Mary, and My Brand-New Alarm

Example

I got very valuable stuff at home. So I bought an alarm. Unfortunately, the alarm just rings at home, doesn't call me on my mobile. I've got two neighbors, Mary and John, who'll call me if they hear the alarm. The problem is that, sometimes, the alarm is caused by an earthquake.

John, Mary, and My Brand-New Alarm

Example

I got very valuable stuff at home. So I bought an alarm. Unfortunately, the alarm just rings at home, doesn't call me on my mobile. I've got two neighbors, Mary and John, who'll call me if they hear the alarm. The problem is that, sometimes, the alarm is caused by an earthquake. Also, John might confuse the alarm with his telephone, and Maria might miss the alarm altogether because she typically listens to loud music.

John, Mary, and My Brand-New Alarm

Example

I got very valuable stuff at home. So I bought an alarm. Unfortunately, the alarm just rings at home, doesn't call me on my mobile. I've got two neighbors, Mary and John, who'll call me if they hear the alarm. The problem is that, sometimes, the alarm is caused by an earthquake. Also, John might confuse the alarm with his telephone, and Maria might miss the alarm altogether because she typically listens to loud music.

$\rightarrow$ Note: This example is by Russel and Norvig.

John, Mary, and My Brand-New Alarm

Example

I got very valuable stuff at home. So I bought an alarm. Unfortunately, the alarm just rings at home, doesn't call me on my mobile. I've got two neighbors, Mary and John, who'll call me if they hear the alarm. The problem is that, sometimes, the alarm is caused by an earthquake. Also, John might confuse the alarm with his telephone, and Maria might miss the alarm altogether because she typically listens to loud music.

→ Note: This example is by Russel and Norvig. If it was by me, (a) there would be no valuable stuff at home, (b) the neighbors would be called "Juan" and "Maria", and (c) there would be no earthquakes.

John, Mary, and My Brand-New Alarm

Example

I got very valuable stuff at home. So I bought an alarm. Unfortunately, the alarm just rings at home, doesn't call me on my mobile. I've got two neighbors, Mary and John, who'll call me if they hear the alarm. The problem is that, sometimes, the alarm is caused by an earthquake. Also, John might confuse the alarm with his telephone, and Maria might miss the alarm altogether because she typically listens to loud music.

→ Note: This example is by Russel and Norvig. If it was by me, (a) there would be no valuable stuff at home, (b) the neighbors would be called “Juan” and “Maria”, and (c) there would be no earthquakes.

Random variables: (All Boolean)

$$Burglary, Earthquake, Alarm, JohnCalls, MaryCalls$$

John, Mary, and My Brand-New Alarm

Example

I got very valuable stuff at home. So I bought an alarm. Unfortunately, the alarm just rings at home, doesn't call me on my mobile. I've got two neighbors, Mary and John, who'll call me if they hear the alarm. The problem is that, sometimes, the alarm is caused by an earthquake. Also, John might confuse the alarm with his telephone, and Maria might miss the alarm altogether because she typically listens to loud music.

$\rightarrow$ Note: This example is by Russel and Norvig. If it was by me, (a) there would be no valuable stuff at home, (b) the neighbors would be called "Juan" and "Maria", and (c) there would be no earthquakes.

Random variables: (All Boolean)

Burglary, *Earthquake*, *Alarm*, *JohnCalls*, *MaryCalls*

Question: We want to compute the probability of atomic events (e.g. $P(Burglary, Earthquake, Alarm, JohnCalls, MaryCalls)$).
Do we need to store a table with 2^5 combinations?

Example continued

Chain rule

$$
\begin{aligned}
&P(\mathit{Burglary}, \mathit{Earthquake}, \mathit{Alarm}, \mathit{JohnCalls}, \mathit{MaryCalls}) = \\
&\quad P(\mathit{Burglary}) \cdot P(\mathit{Earthquake} \mid \mathit{Burglary}) \cdot P(\mathit{Alarm} \mid \mathit{Burglary}, \mathit{Earthquake}) \cdot \\
&\quad P(\mathit{JohnCalls} \mid \mathit{Burglary}, \mathit{Earthquake}, \mathit{Alarm}) \cdot \\
&\quad P(\mathit{MaryCalls} \mid \mathit{Burglary}, \mathit{Earthquake}, \mathit{Alarm}, \mathit{JohnCalls})
\end{aligned}
$$

Example continued

Chain rule

$$\begin{aligned} P(\mathit{Burglary}, \mathit{Earthquake}, \mathit{Alarm}, \mathit{JohnCalls}, \mathit{MaryCalls}) = \\ & P(\mathit{Burglary}) \cdot P(\mathit{Earthquake} \mid \mathit{Burglary}) \cdot P(\mathit{Alarm} \mid \mathit{Burglary}, \mathit{Earthquake}) \cdot \\ & P(\mathit{JohnCalls} \mid \mathit{Burglary}, \mathit{Earthquake}, \mathit{Alarm}) \cdot \\ & P(\mathit{MaryCalls} \mid \mathit{Burglary}, \mathit{Earthquake}, \mathit{Alarm}, \mathit{JohnCalls}) \end{aligned}$$

Conditional independence assumptions

$$P(\mathit{Earthquake} \mid \mathit{Burglary}) \ =$$

Example continued

Chain rule

$$\begin{aligned}&P(\textit{Burglary}, \textit{Earthquake}, \textit{Alarm}, \textit{JohnCalls}, \textit{MaryCalls}) = \\ &\quad P(\textit{Burglary}) \cdot P(\textit{Earthquake} \mid \textit{Burglary}) \cdot P(\textit{Alarm} \mid \textit{Burglary}, \textit{Earthquake}) \cdot \\ &\quad P(\textit{JohnCalls} \mid \textit{Burglary}, \textit{Earthquake}, \textit{Alarm}) \cdot \\ &\quad P(\textit{MaryCalls} \mid \textit{Burglary}, \textit{Earthquake}, \textit{Alarm}, \textit{JohnCalls})\end{aligned}$$

Conditional independence assumptions

$P(\textit{Earthquake} \mid \textit{Burglary}) \ = \ P(\textit{Earthquake})$
Burglaries and earthquakes are independent (this is actually debatable $\rightarrow$ design decision!)

$P(\textit{Alarm} \mid \textit{Burglary}, \textit{Earthquake}) \ =$

Example continued

Chain rule

$$P(\mathit{Burglary}, \mathit{Earthquake}, \mathit{Alarm}, \mathit{JohnCalls}, \mathit{MaryCalls}) = \\ P(\mathit{Burglary}) \cdot P(\mathit{Earthquake} \mid \mathit{Burglary}) \cdot P(\mathit{Alarm} \mid \mathit{Burglary}, \mathit{Earthquake}) \cdot \\ P(\mathit{JohnCalls} \mid \mathit{Burglary}, \mathit{Earthquake}, \mathit{Alarm}) \cdot \\ P(\mathit{MaryCalls} \mid \mathit{Burglary}, \mathit{Earthquake}, \mathit{Alarm}, \mathit{JohnCalls})$$

Conditional independence assumptions

$P(\mathit{Earthquake} \mid \mathit{Burglary}) = P(\mathit{Earthquake})$
Burglaries and earthquakes are independent (this is actually debatable $\rightarrow$ design decision!)

$P(\mathit{Alarm} \mid \mathit{Burglary}, \mathit{Earthquake}) = P(\mathit{Alarm} \mid \mathit{Burglary}, \mathit{Earthquake})$

John and Mary call if and only if they hear the alarm (they don't care about earthquakes)
$P(\mathit{JohnCalls} \mid \mathit{Burglary}, \mathit{Earthquake}, \mathit{Alarm}) =$

Example continued

Chain rule

$$
\begin{aligned}
&P(\mathit{Burglary}, \mathit{Earthquake}, \mathit{Alarm}, \mathit{JohnCalls}, \mathit{MaryCalls}) = \\
&\quad P(\mathit{Burglary}) \cdot P(\mathit{Earthquake} \mid \mathit{Burglary}) \cdot P(\mathit{Alarm} \mid \mathit{Burglary}, \mathit{Earthquake}) \cdot \\
&\quad P(\mathit{JohnCalls} \mid \mathit{Burglary}, \mathit{Earthquake}, \mathit{Alarm}) \cdot \\
&\quad P(\mathit{MaryCalls} \mid \mathit{Burglary}, \mathit{Earthquake}, \mathit{Alarm}, \mathit{JohnCalls})
\end{aligned}
$$

Conditional independence assumptions

$P(\mathit{Earthquake} \mid \mathit{Burglary}) = P(\mathit{Earthquake})$
Burglaries and earthquakes are independent (this is actually debatable $\rightarrow$ design decision!)

$P(\mathit{Alarm} \mid \mathit{Burglary}, \mathit{Earthquake}) = P(\mathit{Alarm} \mid \mathit{Burglary}, \mathit{Earthquake})$

John and Mary call if and only if they hear the alarm (they don't care about earthquakes)
$P(\mathit{JohnCalls} \mid \mathit{Burglary}, \mathit{Earthquake}, \mathit{Alarm}) = P(\mathit{JohnCalls} \mid \mathit{Alarm})$

$P(\mathit{MaryCalls} \mid \mathit{Burglary}, \mathit{Earthquake}, \mathit{Alarm}, \mathit{JohnCalls}) =$

Example continued

Chain rule

$$\begin{aligned}P(\mathit{Burglary}, \mathit{Earthquake}, \mathit{Alarm}, \mathit{JohnCalls}, \mathit{MaryCalls}) = \\ P(\mathit{Burglary}) \cdot P(\mathit{Earthquake} \mid \mathit{Burglary}) \cdot P(\mathit{Alarm} \mid \mathit{Burglary}, \mathit{Earthquake}) \cdot \\ P(\mathit{JohnCalls} \mid \mathit{Burglary}, \mathit{Earthquake}, \mathit{Alarm}) \cdot \\ P(\mathit{MaryCalls} \mid \mathit{Burglary}, \mathit{Earthquake}, \mathit{Alarm}, \mathit{JohnCalls})\end{aligned}$$

Conditional independence assumptions

$P(\mathit{Earthquake} \mid \mathit{Burglary}) = P(\mathit{Earthquake})$
Burglaries and earthquakes are independent (this is actually debatable $\rightarrow$ design decision!)

$P(\mathit{Alarm} \mid \mathit{Burglary}, \mathit{Earthquake}) = P(\mathit{Alarm} \mid \mathit{Burglary}, \mathit{Earthquake})$

John and Mary call if and only if they hear the alarm (they don't care about earthquakes)
$P(\mathit{JohnCalls} \mid \mathit{Burglary}, \mathit{Earthquake}, \mathit{Alarm}) = P(\mathit{JohnCalls} \mid \mathit{Alarm})$

$P(\mathit{MaryCalls} \mid \mathit{Burglary}, \mathit{Earthquake}, \mathit{Alarm}, \mathit{JohnCalls}) = P(\mathit{MaryCalls} \mid \mathit{Alarm})$

Example continued

Chain rule

$$P(\mathit{Burglary}, \mathit{Earthquake}, \mathit{Alarm}, \mathit{JohnCalls}, \mathit{MaryCalls}) = \\ P(\mathit{Burglary}) \cdot P(\mathit{Earthquake} \mid \mathit{Burglary}) \cdot P(\mathit{Alarm} \mid \mathit{Burglary}, \mathit{Earthquake}) \cdot \\ P(\mathit{JohnCalls} \mid \mathit{Burglary}, \mathit{Earthquake}, \mathit{Alarm}) \cdot \\ P(\mathit{MaryCalls} \mid \mathit{Burglary}, \mathit{Earthquake}, \mathit{Alarm}, \mathit{JohnCalls})$$

Conditional independence assumptions

$$P(\mathit{Earthquake} \mid \mathit{Burglary}) = P(\mathit{Earthquake})$$

Burglaries and earthquakes are independent (this is actually debatable → design decision!)

$$P(\mathit{Alarm} \mid \mathit{Burglary}, \mathit{Earthquake}) = P(\mathit{Alarm} \mid \mathit{Burglary}, \mathit{Earthquake})$$

John and Mary call if and only if they hear the alarm (they don't care about earthquakes)

$$P(\mathit{JohnCalls} \mid \mathit{Burglary}, \mathit{Earthquake}, \mathit{Alarm}) = P(\mathit{JohnCalls} \mid \mathit{Alarm})$$

$$P(\mathit{MaryCalls} \mid \mathit{Burglary}, \mathit{Earthquake}, \mathit{Alarm}, \mathit{JohnCalls}) = P(\mathit{MaryCalls} \mid \mathit{Alarm})$$

Simplified representation of joint distribution

$$P(\mathit{Burglary}, \mathit{Earthquake}, \mathit{Alarm}, \mathit{JohnCalls}, \mathit{MaryCalls}) = P(\mathit{Burglary}) \cdot P(\mathit{Earthquake}) \cdot \\ P(\mathit{Alarm} \mid \mathit{Burglary}, \mathit{Earthquake}) \cdot P(\mathit{JohnCalls} \mid \mathit{Alarm}) \cdot P(\mathit{MaryCalls} \mid \mathit{Alarm})$$

John, Mary, and My Alarm: Designing the BN

Cooking Recipe: (1) Design the random variables $X_1, \dots, X_n$; (2) Identify their dependencies; (3) Insert the conditional probability tables $\mathbf{P}(X_i \mid Parents(X_i))$.

John, Mary, and My Alarm: Designing the BN

Cooking Recipe: (1) Design the random variables $X_1, \dots, X_n$; (2) Identify their dependencies; (3) Insert the conditional probability tables $\mathbf{P}(X_i \mid Parents(X_i))$.

Example: Let's cook!

1. Random variables: $Burglary$, $Earthquake$, $Alarm$, $JohnCalls$, $MaryCalls$. (All Boolean)

John, Mary, and My Alarm: Designing the BN

Cooking Recipe: (1) Design the random variables $X_1, \dots, X_n$; (2) Identify their dependencies; (3) Insert the conditional probability tables $\mathbf{P}(X_i \mid Parents(X_i))$.

Example: Let's cook!

1. Random variables: $Burglary$, $Earthquake$, $Alarm$, $JohnCalls$, $MaryCalls$. (All Boolean)
2. Dependencies:

John, Mary, and My Alarm: Designing the BN

Cooking Recipe: (1) Design the random variables $X_1, \dots, X_n$; (2) Identify their dependencies; (3) Insert the conditional probability tables $\mathbf{P}(X_i \mid Parents(X_i))$.

Example: Let's cook!

1. Random variables: $Burglary$, $Earthquake$, $Alarm$, $JohnCalls$, $MaryCalls$. (All Boolean)
2. Dependencies: Burglaries and earthquakes are independent (this is actually debatable $\rightarrow$ design decision!); the alarm might be activated by either. John and Mary call if and only if they hear the alarm (they don't care about earthquakes).

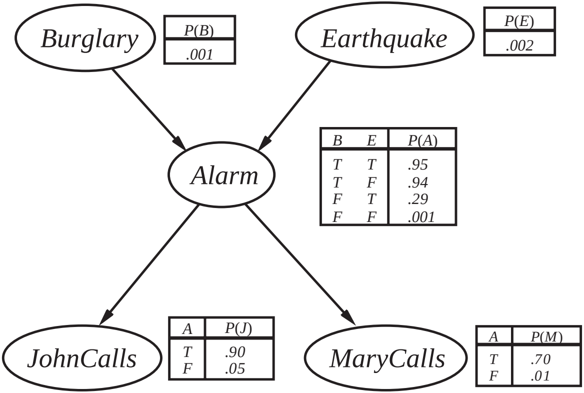
John, Mary, and My Alarm: Designing the BN

Cooking Recipe: (1) Design the random variables $X_1, \dots, X_n$; (2) Identify their dependencies; (3) Insert the conditional probability tables $\mathbf{P}(X_i \mid Parents(X_i))$.

Example: Let's cook!

1. Random variables: $Burglary$, $Earthquake$, $Alarm$, $JohnCalls$, $MaryCalls$. (All Boolean)
2. Dependencies: Burglaries and earthquakes are independent (this is actually debatable $\rightarrow$ design decision!); the alarm might be activated by either. John and Mary call if and only if they hear the alarm (they don't care about earthquakes).
3. Conditional probability tables: Assess the probabilities, see next slide.

John, Mary, and My Alarm: The BN



P(B)
.001

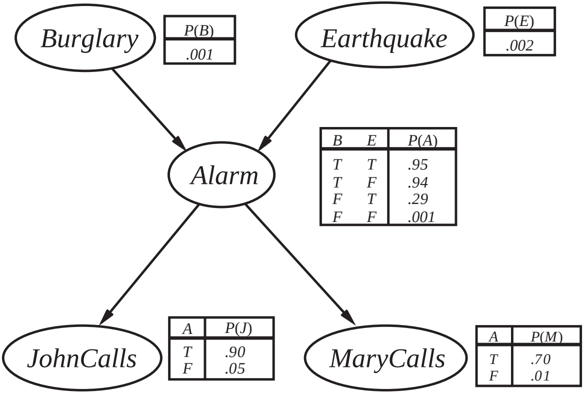
P(E)
.002

B	E	P(A)
T	T	.95
T	F	.94
F	T	.29
F	F	.001

A	P(J)
T	.90
F	.05

A	P(M)
T	.70
F	.01

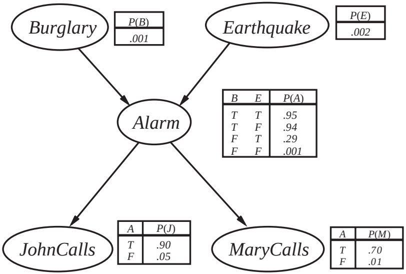
John, Mary, and My Alarm: The BN



Note: In each $\mathbf{P}(X_i \mid Parents(X_i))$, we show only $\mathbf{P}(X_i = true \mid Parents(X_i))$. We don't show $\mathbf{P}(X_i = false \mid Parents(X_i))$ which is $= 1 - \mathbf{P}(X_i = true \mid Parents(X_i))$.

The Syntax of Bayesian Networks

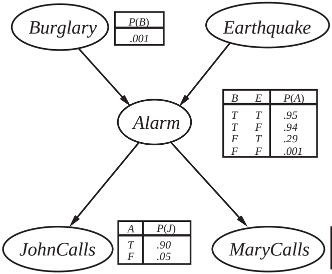
Representation of conditional dependencies in a **directed** and **acyclic** graph:



Definition (Bayesian Network). *Given random variables* $X_1, \dots, X_n$ *with finite domains* $D_1, \dots, D_n$*, a Bayesian network is an acyclic directed graph* $BN = (\{X_1, \dots, X_n\}, E)$. *We denote* $Parents(X_i) := \{X_j \mid (X_j, X_i) \in E\}$*.*
Each X_i *is associated with a function* $CPT(X_i) : D_i \times (\bigtimes_{X_j \in Parents(X_i)} D_j) \mapsto [0,1]$*.*
$CPT(X_i)$ is a **conditional probability table** specifying the conditional distribution $P(X_i \mid \textit{parents}(A_i))$.

The Syntax of Bayesian Networks

Representation of conditional dependencies in a **directed** and **acyclic** graph:



P(B)
.001

P(E)
.002

B	E	P(A)
T	T	.95
T	F	.94
F	T	.29
F	F	.001

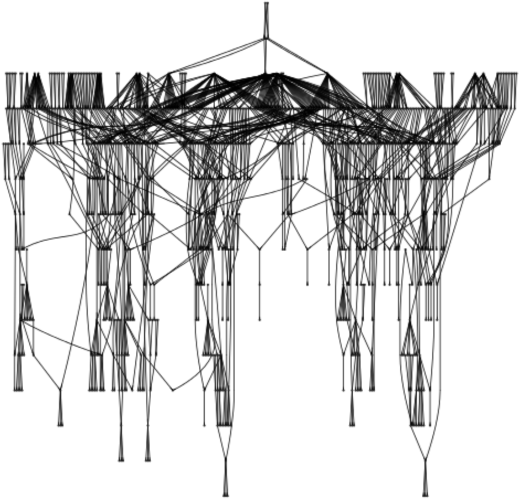
A	P(J)
T	.90
F	.05

A	P(M)
T	.70
F	.01

Definition (Bayesian Network). *Given random variables $X_1, \dots, X_n$ with finite domains $D_1, \dots, D_n$, a Bayesian network is an acyclic directed graph $BN = (\{X_1, \dots, X_n\}, E)$. We denote $Parents(X_i) := \{X_j \mid (X_j, X_i) \in E\}$. Each X_i is associated with a function $CPT(X_i) : D_i \times (\times_{X_j \in Parents(X_i)} D_j) \mapsto [0, 1]$.*
$CPT(X_i)$ is a **conditional probability table** specifying the conditional distribution $P(X_i \mid \textit{parents}(A_i))$.

[$\rightarrow$ Why "acyclic"? Slide 19 (*) $\mathbf{P}(X_1, \dots, X_n) = \Pi_{i=1}^{n} \mathbf{P}(X_i \mid Parents(X_i))$. By (*), acyclic BN suffice to represent any full joint probability distribution. But for cyclic BN, (*) does NOT hold, indeed cyclic BNs may be self-contradictory.]

The Munin network



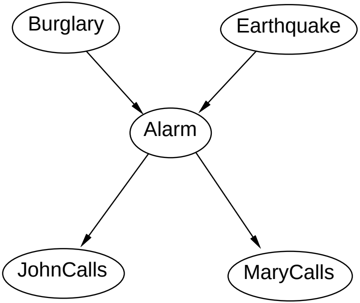
Characteristics:

- Approximately 1100 variables.
- Each variable has between 2 and 20 values.
- 10^{600} possible state configurations!

A system for diagnosing neuro-muscular diseases.

Agenda

The Semantics of BNs: Example

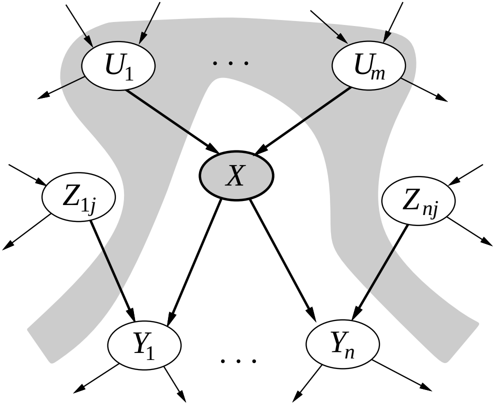


- $Alarm$ depends on $Burglary$ and $Earthquake$.
- $MaryCalls$ only depends on $Alarm$.
  $P(MaryCalls \mid Alarm, Burglary) = P(MaryCalls \mid Alarm)$

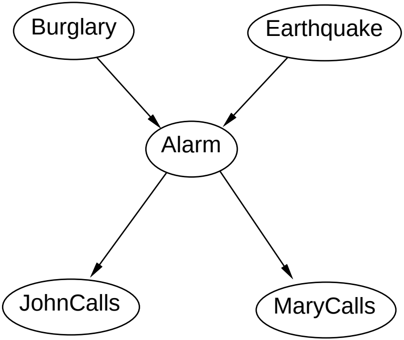
→ Bayesian networks represent sets of independence assumptions.

The Semantics of BNs: General Case

→ Each node X in a BN is conditionally independent of its non-descendants given its parents $Parents(X)$.

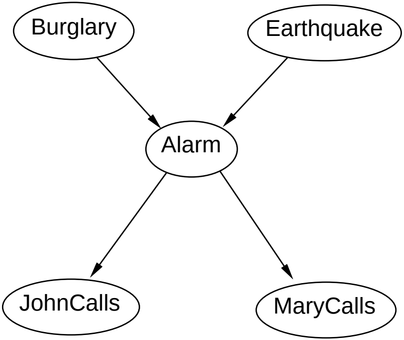


The Semantics of BNs: Example, ctd.



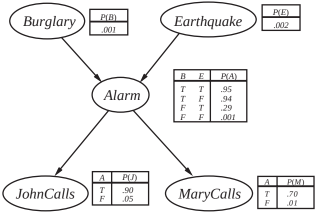
→ Given the value of $Alarm$, $MaryCalls$ is independent of?

The Semantics of BNs: Example, ctd.



→ Given the value of $Alarm$, $MaryCalls$ is independent of?
$\{Burglary, Earthquake, JohnCalls\}$.

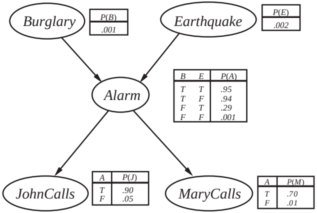
The Semantics of BNs: Formal



Definition. *Given a Bayesian network $BN = (\{X_1, \dots, X_n\}, E)$, we identify BN with the following two assumptions:*

(A) *For $1 \leq i \leq n$, X_i is conditionally independent of $NonDescendants(X_i)$ given $Parents(X_i)$, where $NonDescendants(X_i) := \{X_j \mid (X_i, X_j) \notin E^*\} \setminus Parents(X_i)$ with E^* denoting the transitive closure of E.*

The Semantics of BNs: Formal



Definition. *Given a Bayesian network $BN = (\{X_1, \dots, X_n\}, E)$, we identify BN with the following two assumptions:*

(A) *For $1 \leq i \leq n$, X_i is conditionally independent of $NonDescendants(X_i)$ given $Parents(X_i)$, where $NonDescendants(X_i) := \{X_j \mid (X_i, X_j) \notin E^*\} \setminus Parents(X_i)$ with E^* denoting the transitive closure of E.*

(B) *For $1 \leq i \leq n$, all values x_i of X_i, and all value combinations $parents(X_i)$ of $Parents(X_i)$, we have $P(x_i \mid parents(X_i)) = CPT(x_i, parents(X_i))$.*

Recovering the Full Joint Probability Distribution

"A Bayesian network is a methodology for representing the full joint probability distribution."

Recovering the Full Joint Probability Distribution

"A Bayesian network is a methodology for representing the full joint probability distribution."

$\rightarrow$ How to recover the full joint probability distribution $\mathbf{P}(X_1, \dots, X_n)$ from $BN = (\{X_1, \dots, X_n\}, E)$?

Recovering the Full Joint Probability Distribution

"A Bayesian network is a methodology for representing the full joint probability distribution."

→ How to recover the full joint probability distribution $\mathbf{P}(X_1, \dots, X_n)$ from $BN = (\{X_1, \dots, X_n\}, E)$?

Chain rule: For any ordering $X_1, \dots, X_n$, we have:

$$\mathbf{P}(X_1, \dots, X_n) = \mathbf{P}(X_n \mid X_{n-1}, \dots, X_1)\mathbf{P}(X_{n-1} \mid X_{n-2}, \dots, X_1) \dots \mathbf{P}(X_1)$$

Choose $X_1, \dots, X_n$ consistent with BN: $X_j \in Parents(X_i) \Longrightarrow j < i$.

Recovering the Full Joint Probability Distribution

"A Bayesian network is a methodology for representing the full joint probability distribution."

$\rightarrow$ How to recover the full joint probability distribution $\mathbf{P}(X_1, \dots, X_n)$ from $BN = (\{X_1, \dots, X_n\}, E)$?

Chain rule: For any ordering $X_1, \dots, X_n$, we have:

$$\mathbf{P}(X_1, \dots, X_n) = \mathbf{P}(X_n \mid X_{n-1}, \dots, X_1)\mathbf{P}(X_{n-1} \mid X_{n-2}, \dots, X_1) \dots \mathbf{P}(X_1)$$

Choose $X_1, \dots, X_n$ consistent with BN: $X_j \in Parents(X_i) \implies j < i$.

Exploit conditional independence: With BN assumption (A), instead of $\mathbf{P}(X_i \mid X_{i-1} \dots, X_1)$ we can use $\mathbf{P}(X_i \mid Parents(X_i))$:

$$\mathbf{P}(X_1, \dots, X_n) = \Pi_{i=1}^{n} \mathbf{P}(X_i \mid Parents(X_i))$$

The distributions $\mathbf{P}(X_i \mid Parents(X_i))$ are given by BN assumption (B).

Recovering the Full Joint Probability Distribution

"A Bayesian network is a methodology for representing the full joint probability distribution."

→ How to recover the full joint probability distribution $\mathbf{P}(X_1, \dots, X_n)$ from $BN = (\{X_1, \dots, X_n\}, E)$?

Chain rule: For any ordering $X_1, \dots, X_n$, we have:

$$\mathbf{P}(X_1, \dots, X_n) = \mathbf{P}(X_n \mid X_{n-1}, \dots, X_1)\mathbf{P}(X_{n-1} \mid X_{n-2}, \dots, X_1) \dots \mathbf{P}(X_1)$$

Choose $X_1, \dots, X_n$ consistent with BN: $X_j \in Parents(X_i) \implies j < i$.

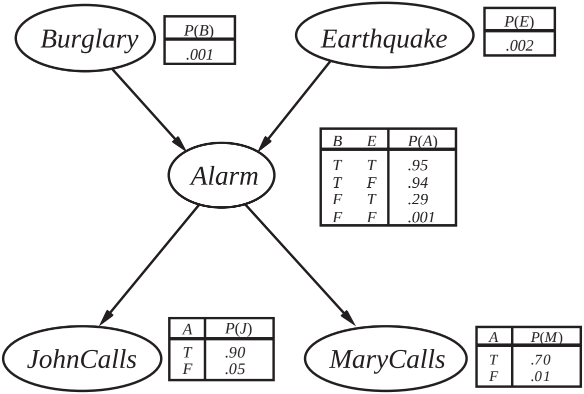
Exploit conditional independence: With BN assumption (A), instead of $\mathbf{P}(X_i \mid X_{i-1} \dots, X_1)$ we can use $\mathbf{P}(X_i \mid Parents(X_i))$:

$$\mathbf{P}(X_1, \dots, X_n) = \Pi_{i=1}^{n} \mathbf{P}(X_i \mid Parents(X_i))$$

The distributions $\mathbf{P}(X_i \mid Parents(X_i))$ are given by BN assumption (B).

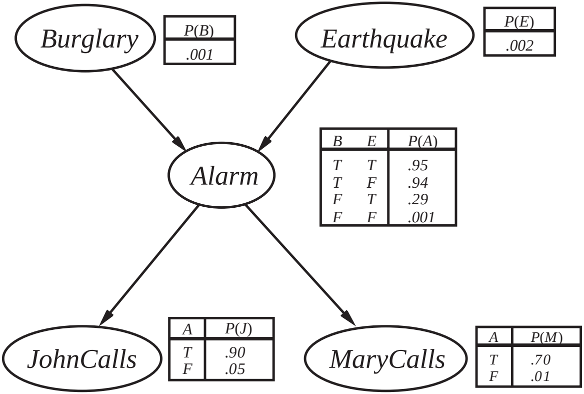
→ Same for atomic events $P(x_1, \dots, x_n)$.

Recovering a Probability for John, Mary, and the Alarm



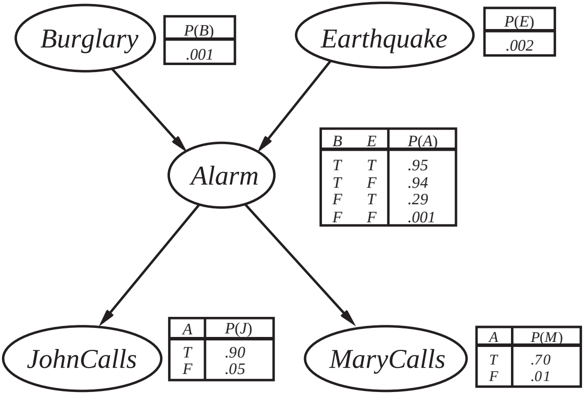
$P(j, m, a, \neg b, \neg e) =$

Recovering a Probability for John, Mary, and the Alarm



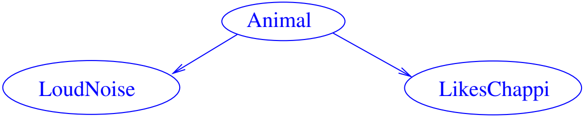
$$
\begin{aligned}
P(j, m, a, \neg b, \neg e) &= P(j \mid a) \cdot P(m \mid a) \cdot P(a \mid \neg b, \neg e) \cdot P(\neg b) \cdot P(\neg e) \\
&=
\end{aligned}
$$

Recovering a Probability for John, Mary, and the Alarm



$$
\begin{aligned}
{\color{red}P(j, m, a, \neg b, \neg e)} &= P(j \mid a) \cdot P(m \mid a) \cdot P(a \mid \neg b, \neg e) \cdot P(\neg b) \cdot P(\neg e) \\
&= 0.9 * 0.7 * 0.001 * 0.999 * 0.998 \\
&= 0.000628
\end{aligned}
$$

Questionnaire

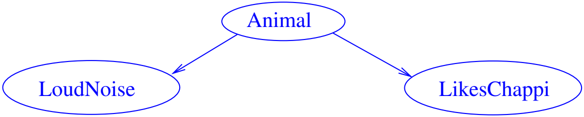


Question!

Say BN is the Bayesian network above. Which statements are correct?

- (A): $Animal$ is independent of $LikesChappi$.
- (B): $LoudNoise$ is independent of $LikesChappi$.
- (C): $Animal$ is conditionally independent of $LikesChappi$ given $LoudNoise$.
- (D): $LikesChappi$ is conditionally independent of $LoudNoise$ given $Animal$.

Questionnaire



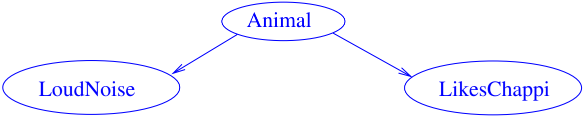
Question!

Say BN is the Bayesian network above. Which statements are correct?

- (A): $Animal$ is independent of $LikesChappi$.
- (B): $LoudNoise$ is independent of $LikesChappi$.
- (C): $Animal$ is conditionally independent of $LikesChappi$ given $LoudNoise$.
- (D): $LikesChappi$ is conditionally independent of $LoudNoise$ given $Animal$.

(A) No: $likeschappi$ indicates dog.

Questionnaire



Question!

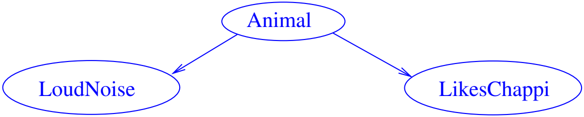
Say BN is the Bayesian network above. Which statements are correct?

- (A): $Animal$ is independent of $LikesChappi$.
- (B): $LoudNoise$ is independent of $LikesChappi$.
- (C): $Animal$ is conditionally independent of $LikesChappi$ given $LoudNoise$.
- (D): $LikesChappi$ is conditionally independent of $LoudNoise$ given $Animal$.

(A) No: $likeschappi$ indicates dog.

(B) No: Not knowing what animal it is, $likeschappi$ is an indication for dog which indicates $loudnoise$.

Questionnaire



Question!

Say BN is the Bayesian network above. Which statements are correct?

(A): $Animal$ is independent of $LikesChappi$.

(B): $LoudNoise$ is independent of $LikesChappi$.

(C): $Animal$ is conditionally independent of $LikesChappi$ given $LoudNoise$.

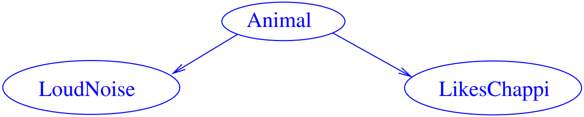
(D): $LikesChappi$ is conditionally independent of $LoudNoise$ given $Animal$.

(A) No: $likeschappi$ indicates dog.

(B) No: Not knowing what animal it is, $likeschappi$ is an indication for dog which indicates $loudnoise$.

(C) No: For example, even if we know $loudnoise$, knowing in addition that $likeschappi$ gives us a stronger indication of $Animal = dog$.

Questionnaire



Question!

Say BN is the Bayesian network above. Which statements are correct?

- (A): $Animal$ is independent of $LikesChappi$.
- (B): $LoudNoise$ is independent of $LikesChappi$.
- (C): $Animal$ is conditionally independent of $LikesChappi$ given $LoudNoise$.
- (D): $LikesChappi$ is conditionally independent of $LoudNoise$ given $Animal$.

(A) No: $likeschappi$ indicates dog.

(B) No: Not knowing what animal it is, $likeschappi$ is an indication for dog which indicates $loudnoise$.

(C) No: For example, even if we know $loudnoise$, knowing in addition that $likeschappi$ gives us a stronger indication of $Animal = dog$.

(D) Yes: $X_i = LikesChappi$ is conditionally independent of $NonDescendants(X_i) = \{LoudNoise\}$ given $Parents(X_i) = \{Animal\}$.

Agenda

D-separation

Given a Causal Network that relates causes to effects, what information is relevant to predict the value of a variable?

D-separation

Given a Causal Network that relates causes to effects, what information is relevant to predict the value of a variable?

In the context of Bayesian Netowrks:
Let X, Y, and E be disjoint sets of variables.

Is X conditionally independent from Y given E?

D-separation

Given a Causal Network that relates causes to effects, what information is relevant to predict the value of a variable?

In the context of Bayesian Netowrks:
Let X, Y, and E be disjoint sets of variables.

Is X conditionally independent from Y given E?

$\rightarrow$We introduce d-separation as a general way of answering that question

Bayesian Network Example 1

Bayesian Network Example 1

- If there has been a flooding does that tell me something about the amount of rain that has fallen?

Bayesian Network Example 1

- If there has been a flooding does that tell me something about the amount of rain that has fallen?

  Yes, Rainfall and Flooding are not independent!

- The water level is high: If there has been a flooding does that tell me anything new about the amount of rain that has fallen?

Bayesian Network Example 1

- If there has been a flooding does that tell me something about the amount of rain that has fallen?

  Yes, Rainfall and Flooding are not independent!

- The water level is high: If there has been a flooding does that tell me anything new about the amount of rain that has fallen?

  No, Rainfall and Flooding are conditionally independent given WaterLevel!

Bayesian Network Example 1

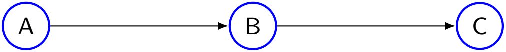
- If there has been a flooding does that tell me something about the amount of rain that has fallen?

  Yes, Rainfall and Flooding are not independent!

- The water level is high: If there has been a flooding does that tell me anything new about the amount of rain that has fallen?

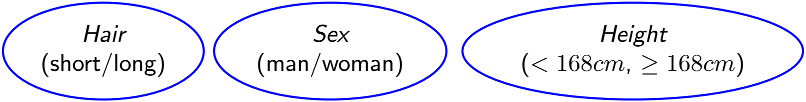
  No, Rainfall and Flooding are conditionally independent given WaterLevel!

Serial Connection (Indirect cause)

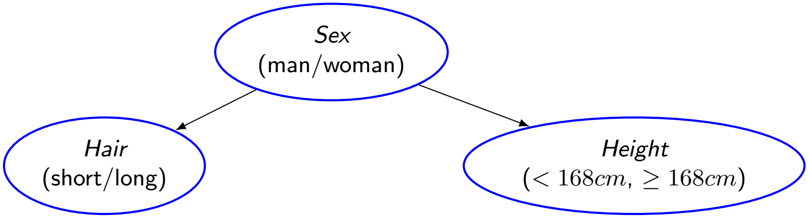


Information can be transmitted between A and C through B if B is not observed. If B is observed, A and C are independent.

Bayesian Network Example 2

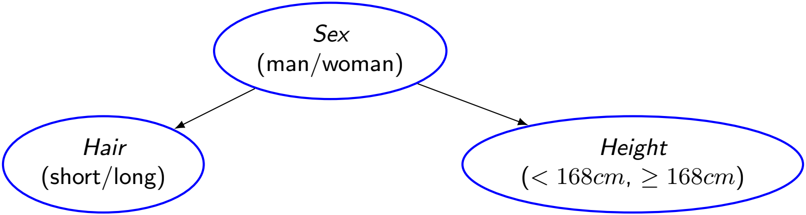


Bayesian Network Example 2



- If a person has long hair does that say something about his/her stature?

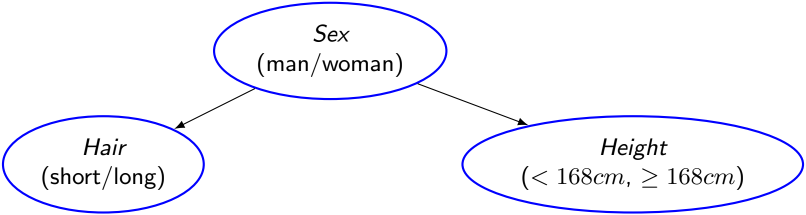
Bayesian Network Example 2



- If a person has long hair does that say something about his/her stature?

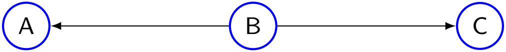
  Yes, Hair and Height are not independent!
- It is a woman: If she has long hair does that say something about her stature?

Bayesian Network Example 2



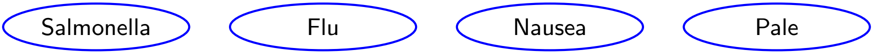
- If a person has long hair does that say something about his/her stature?
  Yes, Hair and Height are not independent!
- It is a woman: If she has long hair does that say something about her stature?
  No, Hair and Height are conditionally independent given Sex!

Diverging Connection (common cause)

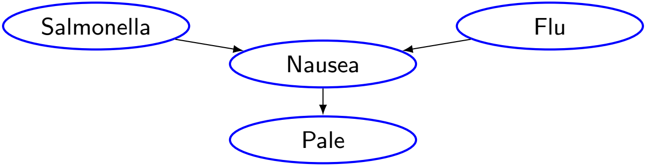


Information can be transmitted between A and C through B if B is not observed. If B is observed, A and C are independent.

Bayesian Network Example 3

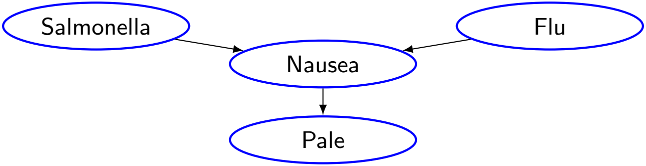


Bayesian Network Example 3



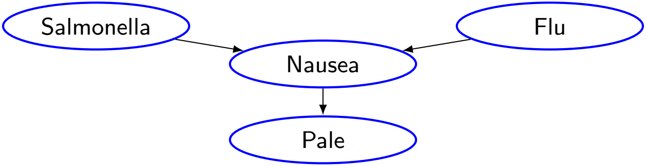
- Does salmonella have an impact on Flu?

Bayesian Network Example 3



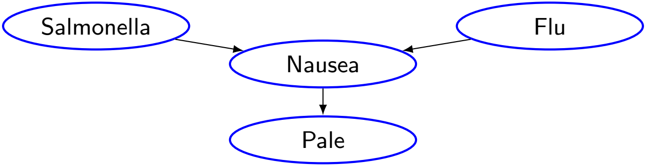
- Does salmonella have an impact on Flu?
  No, salmonella is independent of Flu

Bayesian Network Example 3



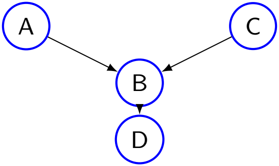
- Does salmonella have an impact on Flu?
  No, salmonella is independent of Flu
- If a person is Pale, does salmonella then have an impact on Flu?

Bayesian Network Example 3



- Does salmonella have an impact on Flu?
  No, salmonella is independent of Flu
- If a person is Pale, does salmonella then have an impact on Flu?
  Yes, salmonella can explain why the person is pale!

Converging Connection (common effect)

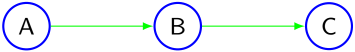


A and C are independent unless we have information about B or its descendants.
If B or a descendant (D) are observed, we can transmit information between A and C.

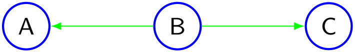
D-Separation Rules

Does evidence from A affect our knowledge about C?

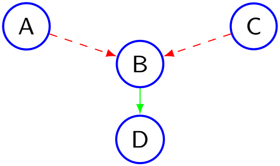
- Evidence may be transmitted through a serial or diverging connection unless it is instantiated.
- Evidence may be transmitted through a converging connection only if either the variable in the connection or one of its descendants has received evidence.



Serial



Diverging

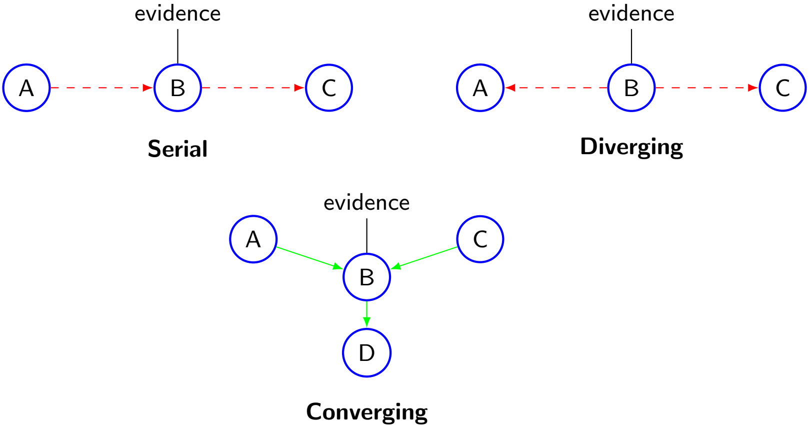


Converging

D-Separation Rules

Does evidence from A affect our knowledge about C?

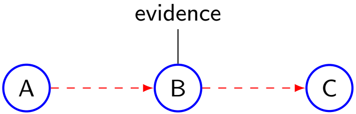
- Evidence may be transmitted through a serial or diverging connection unless it is instantiated.
- Evidence may be transmitted through a converging connection only if either the variable in the connection or one of its descendants has received evidence.



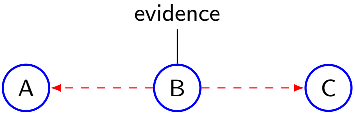
D-Separation Rules

Does evidence from A affect our knowledge about C?

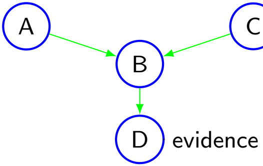
- Evidence may be transmitted through a serial or diverging connection unless it is instantiated.
- Evidence may be transmitted through a converging connection only if either the variable in the connection or one of its descendants has received evidence.



Serial



Diverging



Converging

The d-separation theorem

Theorem

For all pairwise disjoint sets $\mathbf{A}, \mathbf{B}, \mathbf{C}$ of nodes in a Bayesian network:

If $\mathbf{C}$ d-separates $\mathbf{A}$ from $\mathbf{B}$, then $P(\mathbf{A} \mid \mathbf{B}, \mathbf{C}) = P(\mathbf{A} \mid \mathbf{C})$.

There are no more general graphical conditions than d-separation for which such a result holds.

The d-separation theorem

Theorem

For all pairwise disjoint sets $\mathbf{A}, \mathbf{B}, \mathbf{C}$ of nodes in a Bayesian network:

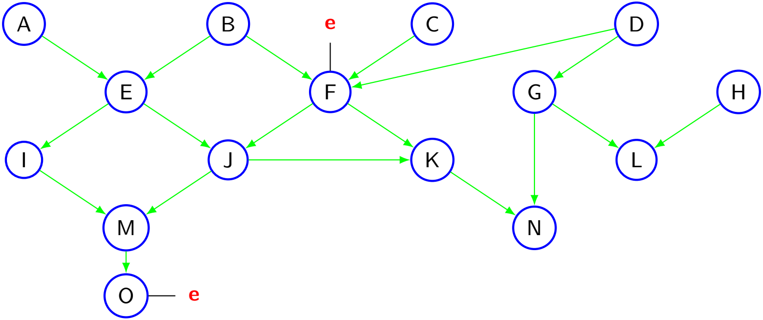
If $\mathbf{C}$ *d-separates* $\mathbf{A}$ *from* $\mathbf{B}$*, then* $P(\mathbf{A} \mid \mathbf{B}, \mathbf{C}) = P(\mathbf{A} \mid \mathbf{C})$.

There are no more general graphical conditions than d-separation for which such a result holds.

Why is d-separation important?

- Gaining insight: given a (correct) Bayesian network model, can derive insight into the dependencies among the variables
- Debugging a model: given a Bayesian network model, check whether entailed independence relations are plausible
- Correctness of algorithms: certain computational procedures depend on validity of special independence relations

Transmission of Evidence: Questionnaire



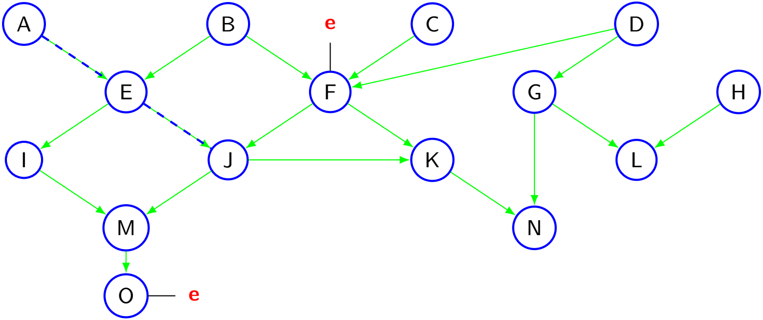
Question!

Given evidence e, can knowledge of A have an impact on our knowledge of _?

(A): J (B): G

(C): H (D): B

Transmission of Evidence: Questionnaire



Question!

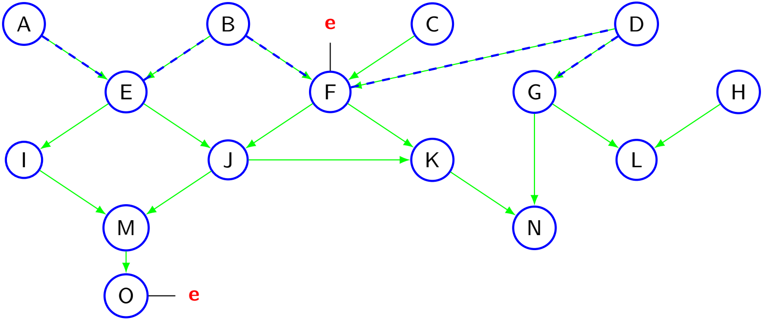
Given evidence e, can knowledge of A have an impact on our knowledge of _?

(A): J (B): G

(C): H (D): B

(A) yes

Transmission of Evidence: Questionnaire



Question!

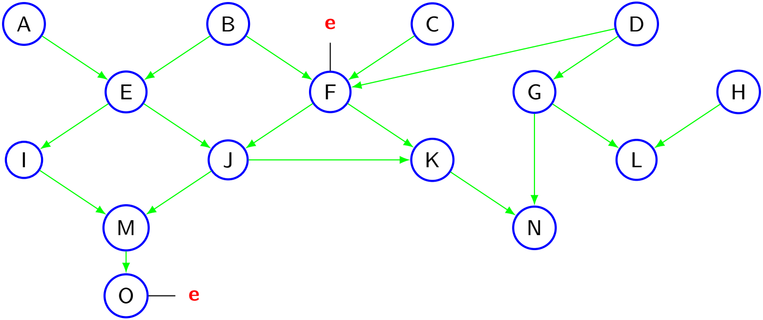
Given evidence e, can knowledge of A have an impact on our knowledge of _?

(A): J (B): G

(C): H (D): B

(A) yes (B) yes

Transmission of Evidence: Questionnaire



Question!

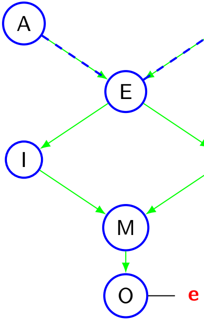
Given evidence e, can knowledge of A have an impact on our knowledge of _?

(A): J (B): G

(C): H (D): B

(A) yes (B) yes (C) no

Transmission of Evidence: Questionnaire



Question!

Given evidence e, can knowledge of A have an impact on our knowledge of _?

(A): J (B): G

(C): H (D): B

(A) yes (B) yes (C) no (D) yes (note O is a descendant of E)

D-separation Algorithm

Algorithm for testing whether A and B are d-separated given evidence E

1. Construct the subgraph consisting of variables A, B and E plus all of their ancestors (for all nodes added, include also their parents).
2. Add undirected edges between all parents of a common child.
3. Replace directed edges by undirected edges
4. Remove all evidence E nodes from the graph (along with their connections).
5. If A and B are d-separated if there is no path between them in the resulting graph.

Agenda

Constructing a Bayesian Network

Construction via chain rule

1. put the random variables in some order
2. write the joint distribution using chain rule
3. simplify conditional probability factors by conditional independence assumptions. That determines the *parents* of each node, i.e. the graph structure
4. specify the conditional probability tables

Note: the structure of the resulting network strongly depends on the chosen order of the variables.

Construction via causality

- Draw and edge from variable A to variable B if A has a direct causal influence on A.

Note: this may not always be possible:

- *Inflation* $\rightarrow$ *salaries* or *salaries* $\rightarrow$ *Inflation* ?
- *Rain* doesn't cause *Sun*, and *Sun* doesn't cause *Rain*, but they are not independent either!

Constructing Bayesian Networks

BN construction algorithm:

1. Initialize $BN := (\{X_1, \dots, X_n\}, E)$ where $E = \emptyset$.

Constructing Bayesian Networks

BN construction algorithm:

1. Initialize $BN := (\{X_1, \ldots, X_n\}, E)$ where $E = \emptyset$.
2. Fix any order of the variables, $X_1, \ldots, X_n$.

Constructing Bayesian Networks

BN construction algorithm:

1. Initialize $BN := (\{X_1, \dots, X_n\}, E)$ where $E = \emptyset$.
2. Fix any order of the variables, $X_1, \dots, X_n$.
3. **for** $i := 1, \dots, n$ **do**

Constructing Bayesian Networks

BN construction algorithm:

1. Initialize $BN := (\{X_1, \dots, X_n\}, E)$ where $E = \emptyset$.
2. Fix any order of the variables, $X_1, \dots, X_n$.
3. **for** $i := 1, \dots, n$ **do**
   a. Choose a minimal set $Parents(X_i) \subseteq \{X_1, \dots, X_{i-1}\}$ so that $\mathbf{P}(X_i \mid X_{i-1} \dots, X_1) = \mathbf{P}(X_i \mid Parents(X_i))$.

Constructing Bayesian Networks

BN construction algorithm:

1. Initialize $BN := (\{X_1, \dots, X_n\}, E)$ where $E = \emptyset$.
2. Fix any order of the variables, $X_1, \dots, X_n$.
3. **for** $i := 1, \dots, n$ **do**
   a. Choose a minimal set $Parents(X_i) \subseteq \{X_1, \dots, X_{i-1}\}$ so that $\mathbf{P}(X_i \mid X_{i-1} \dots, X_1) = \mathbf{P}(X_i \mid Parents(X_i))$.
   b. For each $X_j \in Parents(X_i)$, insert (X_j, X_i) into E.

Constructing Bayesian Networks

BN construction algorithm:

1. Initialize $BN := (\{X_1, \dots, X_n\}, E)$ where $E = \emptyset$.
2. Fix any order of the variables, $X_1, \dots, X_n$.
3. **for** $i := 1, \dots, n$ **do**
   a. Choose a minimal set $Parents(X_i) \subseteq \{X_1, \dots, X_{i-1}\}$ so that $\mathbf{P}(X_i \mid X_{i-1} \dots, X_1) = \mathbf{P}(X_i \mid Parents(X_i))$.
   b. For each $X_j \in Parents(X_i)$, insert (X_j, X_i) into E.
   c. Associate X_i with $CPT(X_i)$ corresponding to $\mathbf{P}(X_i \mid Parents(X_i))$.

Constructing Bayesian Networks

BN construction algorithm:

1. Initialize $BN := (\{X_1, \dots, X_n\}, E)$ where $E = \emptyset$.
2. Fix any order of the variables, $X_1, \dots, X_n$.
3. **for** $i := 1, \dots, n$ **do**
   a. Choose a minimal set $Parents(X_i) \subseteq \{X_1, \dots, X_{i-1}\}$ so that $\mathbf{P}(X_i \mid X_{i-1} \dots, X_1) = \mathbf{P}(X_i \mid Parents(X_i))$.
   b. For each $X_j \in Parents(X_i)$, insert (X_j, X_i) into E.
   c. Associate X_i with $CPT(X_i)$ corresponding to $\mathbf{P}(X_i \mid Parents(X_i))$.

Attention! *Which* variables we need to include into $Parents(X_i)$ depends on what "$\{X_1, \dots, X_{i-1}\}$" is . . . !

Constructing Bayesian Networks

BN construction algorithm:

1. Initialize $BN := (\{X_1, \dots, X_n\}, E)$ where $E = \emptyset$.
2. Fix any order of the variables, $X_1, \dots, X_n$.
3. **for** $i := 1, \dots, n$ **do**
   a. Choose a minimal set $Parents(X_i) \subseteq \{X_1, \dots, X_{i-1}\}$ so that $\mathbf{P}(X_i \mid X_{i-1} \dots, X_1) = \mathbf{P}(X_i \mid Parents(X_i))$.
   b. For each $X_j \in Parents(X_i)$, insert (X_j, X_i) into E.
   c. Associate X_i with $CPT(X_i)$ corresponding to $\mathbf{P}(X_i \mid Parents(X_i))$.

Attention! *Which* variables we need to include into $Parents(X_i)$ depends on what "$\{X_1, \dots, X_{i-1}\}$" is . . . !

$\rightarrow$ The size of the resulting BN depends on the chosen order $X_1, \dots, X_n$.

Constructing Bayesian Networks

BN construction algorithm:

1. Initialize $BN := (\{X_1, \dots, X_n\}, E)$ where $E = \emptyset$.
2. Fix any order of the variables, $X_1, \dots, X_n$.
3. **for** $i := 1, \dots, n$ **do**
   a. Choose a minimal set $Parents(X_i) \subseteq \{X_1, \dots, X_{i-1}\}$ so that $\mathbf{P}(X_i \mid X_{i-1} \dots, X_1) = \mathbf{P}(X_i \mid Parents(X_i))$.
   b. For each $X_j \in Parents(X_i)$, insert (X_j, X_i) into E.
   c. Associate X_i with $CPT(X_i)$ corresponding to $\mathbf{P}(X_i \mid Parents(X_i))$.

Attention! *Which* variables we need to include into $Parents(X_i)$ depends on what "$\{X_1, \dots, X_{i-1}\}$" is ... !

$\rightarrow$ The size of the resulting BN depends on the chosen order $X_1, \dots, X_n$.

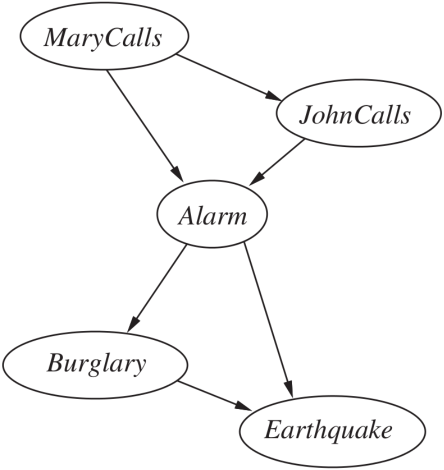
$\rightarrow$ The size of a Bayesian network is *not* a fixed property of the domain. It depends on the skill of the designer.

John and Mary Depend on the Variable Order!

Example: $MaryCalls, JohnCalls, Alarm, Burglary, Earthquake.$

John and Mary Depend on the Variable Order!

Example: $MaryCalls, JohnCalls, Alarm, Burglary, Earthquake.$

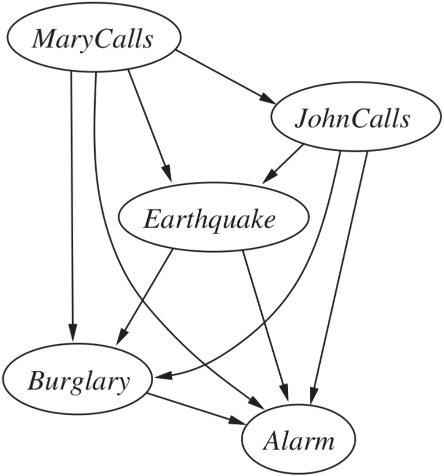


John and Mary Depend on the Variable Order! Ctd.

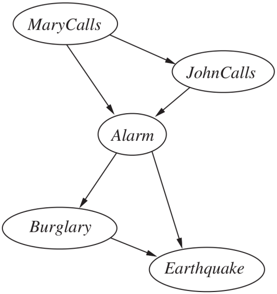
Example: $MaryCalls, JohnCalls, Earthquake, Burglary, Alarm.$

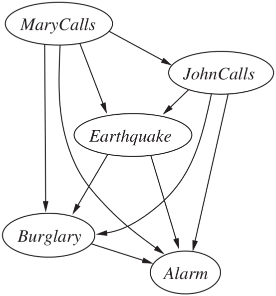
John and Mary Depend on the Variable Order! Ctd.

Example: $MaryCalls$, $JohnCalls$, $Earthquake$, $Burglary$, $Alarm$.

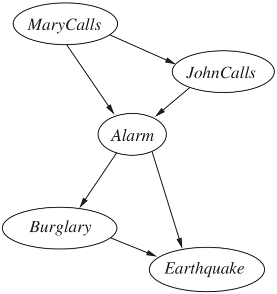


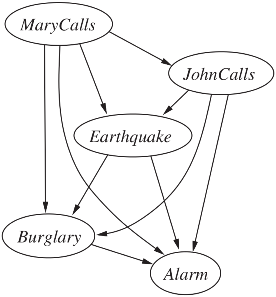
John and Mary, What Went Wrong?





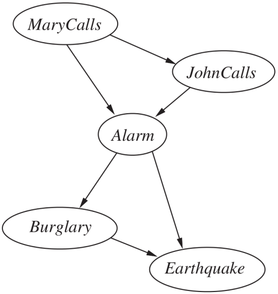
John and Mary, What Went Wrong?

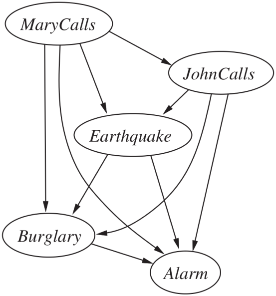




$\rightarrow$ These BNs link from symptoms to causes! $(\mathbf{P}(Cavity \mid Toothache))$

John and Mary, What Went Wrong?

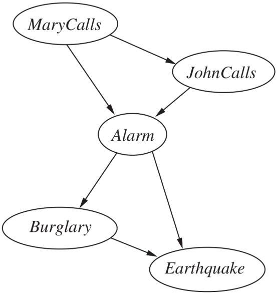


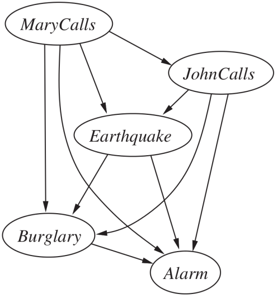


→ These BNs link from symptoms to causes! $(\mathbf{P}(Cavity \mid Toothache))$

- We fail to identify many conditional independence relations (e.g., get dependencies between conditionally independent symptoms).

John and Mary, What Went Wrong?





→ These BNs link from symptoms to causes! $(\mathbf{P}(Cavity \mid Toothache))$

- We fail to identify many conditional independence relations (e.g., get dependencies between conditionally independent symptoms).
- Also recall: Conditional probabilities $\mathbf{P}(Symptom \mid Cause)$ are more robust and often easier to assess than $\mathbf{P}(Cause \mid Symptom)$.

→ We should order causes before symptoms.

The Size of Bayesian Networks

Definition. *Given random variables* $X_1, \dots, X_n$ *with finite domains* $D_1, \dots, D_n$*, the size of* $BN = (\{X_1, \dots, X_n\}, E)$ *is defined as* $size(BN) := \sum_{i=1}^{n} |D_i| * \Pi_{X_j \in Parents(X_i)} |D_j|$. (= The total number of entries in the CPTs.)

The Size of Bayesian Networks

Definition. *Given random variables* $X_1, \dots, X_n$ *with finite domains* $D_1, \dots, D_n$*, the size of* $BN = (\{X_1, \dots, X_n\}, E)$ *is defined as* $size(BN) := \sum_{i=1}^{n} |D_i| * \Pi_{X_j \in Parents(X_i)} |D_j|$. (= The total number of entries in the CPTs.)

→ Smaller BN $\Longrightarrow$ assess less probabilities, more efficient inference.

- Explicit full joint probability distribution has size $\Pi_{i=1}^{n} |D_i|$.

The Size of Bayesian Networks

Definition. *Given random variables* $X_1, \dots, X_n$ *with finite domains* $D_1, \dots, D_n$*, the size of* $BN = (\{X_1, \dots, X_n\}, E)$ *is defined as* $size(BN) := \sum_{i=1}^{n} |D_i| * \Pi_{X_j \in Parents(X_i)} |D_j|$. (= The total number of entries in the CPTs.)

→ Smaller BN ⟹ assess less probabilities, more efficient inference.

- Explicit full joint probability distribution has size $\Pi_{i=1}^{n} |D_i|$.
- If $|Parents(X_i)| \leq k$ for every X_i, and $D_{\max}$ is the largest variable domain, then $size(BN) \leq n * |D_{\max}|^{k+1}$.

The Size of Bayesian Networks

Definition. *Given random variables* $X_1, \dots, X_n$ *with finite domains* $D_1, \dots, D_n$*, the size of* $BN = (\{X_1, \dots, X_n\}, E)$ *is defined as* $size(BN) := \sum_{i=1}^{n} |D_i| * \Pi_{X_j \in Parents(X_i)} |D_j|$. (= The total number of entries in the CPTs.)

→ Smaller BN $\Longrightarrow$ assess less probabilities, more efficient inference.

- Explicit full joint probability distribution has size $\Pi_{i=1}^{n} |D_i|$.
- If $|Parents(X_i)| \leq k$ for every X_i, and $D_{\max}$ is the largest variable domain, then $size(BN) \leq n * |D_{\max}|^{k+1}$.
  → For $|D_{\max}| = 2$, $n = 20$, $k = 4$ we have $2^{20} = 1048576$ probabilities, but a Bayesian network of size $\leq 20 * 2^5 = 640 \dots$!

The Size of Bayesian Networks

Definition. *Given random variables $X_1, \dots, X_n$ with finite domains $D_1, \dots, D_n$, the size of $BN = (\{X_1, \dots, X_n\}, E)$ is defined as $size(BN) :=$ $\sum_{i=1}^{n} |D_i| * \Pi_{X_j \in Parents(X_i)} |D_j|$.* (= The total number of entries in the CPTs.)

$\rightarrow$ Smaller BN $\Longrightarrow$ assess less probabilities, more efficient inference.

- Explicit full joint probability distribution has size $\Pi_{i=1}^{n} |D_i|$.
- If $|Parents(X_i)| \leq k$ for every X_i, and $D_{\max}$ is the largest variable domain, then $size(BN) \leq n * |D_{\max}|^{k+1}$.
  $\rightarrow$ For $|D_{\max}| = 2$, $n = 20$, $k = 4$ we have $2^{20} = 1048576$ probabilities, but a Bayesian network of size $\leq 20 * 2^5 = 640$ ...!
- In the *worst case*, $size(BN) = \sum_{i=1}^{n} \Pi_{j=1}^{i} |D_i|$, namely if

The Size of Bayesian Networks

Definition. *Given random variables* $X_1, \dots, X_n$ *with finite domains* $D_1, \dots, D_n$*, the size of* $BN = (\{X_1, \dots, X_n\}, E)$ *is defined as* $size(BN) := \sum_{i=1}^{n} |D_i| * \Pi_{X_j \in Parents(X_i)} |D_j|$. (= The total number of entries in the CPTs.)

→ Smaller BN $\Longrightarrow$ assess less probabilities, more efficient inference.

- Explicit full joint probability distribution has size $\Pi_{i=1}^{n} |D_i|$.
- If $|Parents(X_i)| \leq k$ for every X_i, and $D_{\max}$ is the largest variable domain, then $size(BN) \leq n * |D_{\max}|^{k+1}$.
  → For $|D_{\max}| = 2$, $n = 20$, $k = 4$ we have $2^{20} = 1048576$ probabilities, but a Bayesian network of size $\leq 20 * 2^5 = 640 \dots$!
- In the *worst case*, $size(BN) = \sum_{i=1}^{n} \Pi_{j=1}^{i} |D_i|$, namely if *every variable depends on all its predecessors in the chosen order*.

→ BNs are compact if each variable is directly influenced only by few of its predecessor variables.

Questionnaire

Question!

What is the Bayesian network we get by constructing according to the ordering $X_1 = LoudNoise, X2 = Animal, X3 = LikesChappi$?

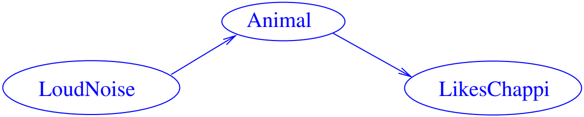
Question!

What is the Bayesian network we get by constructing according to the ordering $X_1 = LoudNoise, X2 = LikesChappi, X3 = Animal$?

Questionnaire

Question!

What is the Bayesian network we get by constructing according to the ordering $X_1 = LoudNoise, X2 = Animal, X3 = LikesChappi$?



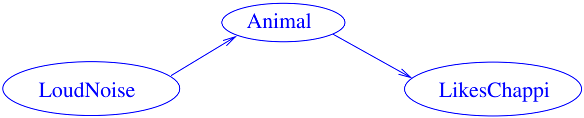
Question!

What is the Bayesian network we get by constructing according to the ordering $X_1 = LoudNoise, X2 = LikesChappi, X3 = Animal$?

Questionnaire

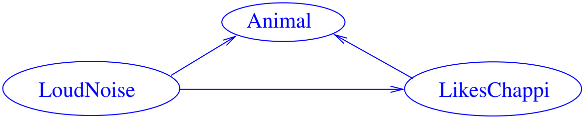
Question!

What is the Bayesian network we get by constructing according to the ordering $X_1 = LoudNoise, X2 = Animal, X3 = LikesChappi$?



Question!

What is the Bayesian network we get by constructing according to the ordering $X_1 = LoudNoise, X2 = LikesChappi, X3 = Animal$?



Building models

Example

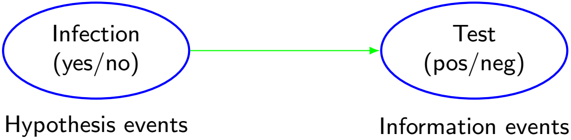
Milk from a cow may be infected. To detect whether or not the milk is infected, you can apply a test which may either give a positive or a negative test result. The test is not perfect: It may give false positives as well as false negatives.

Building models

Example

Milk from a cow may be infected. To detect whether or not the milk is infected, you can apply a test which may either give a positive or a negative test result. The test is not perfect: It may give false positives as well as false negatives.

We set links from hypothesis to test (causal relationship: having the disease is what causes the test to be positive and not viceversa):

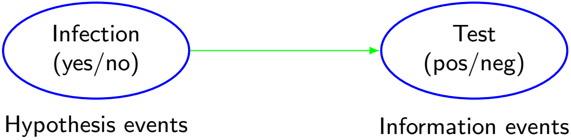


Building models

Example

Milk from a cow may be infected. To detect whether or not the milk is infected, you can apply a test which may either give a positive or a negative test result. The test is not perfect: It may give false positives as well as false negatives.

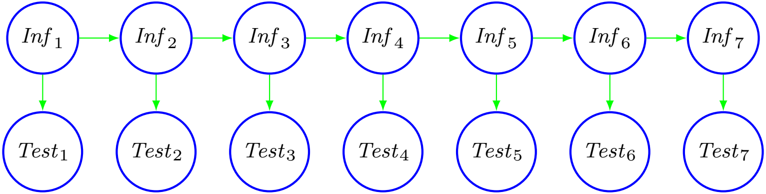
We set links from hypothesis to test (causal relationship: having the disease is what causes the test to be positive and not viceversa):



What if we repeat the test multiple days?

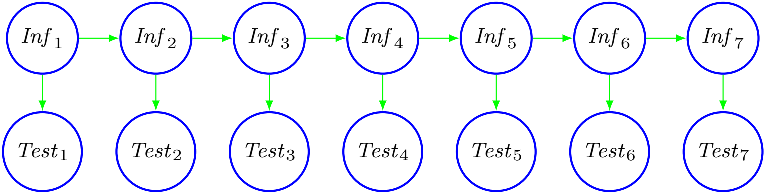
7-day model I

Infections develop over time:



7-day model I

Infections develop over time:



Assumption

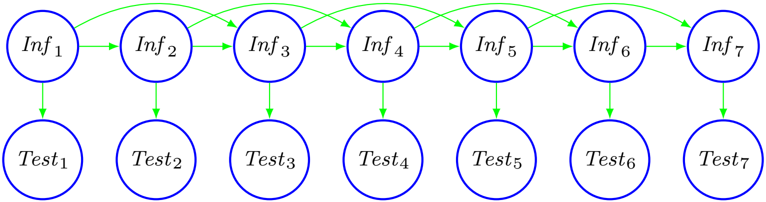
The Markov property: If I know the present, then the past has no influence on the future:

$$\text{Inf}_{i-1} \text{ is d-separated from } \text{Inf}_{i+1} \text{ given } \text{Inf}_i.$$

But what if yesterday's Inf-state has an impact on tomorrow's Inf-state?

7-day model II

Non-Markov relations

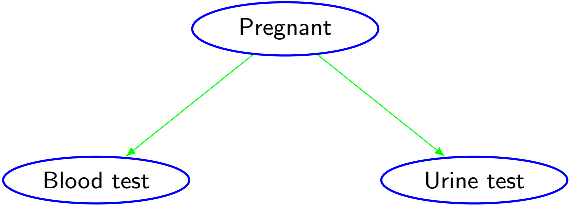


Yesterday's Inf-state has an impact on tomorrow's Inf-state.

Agenda

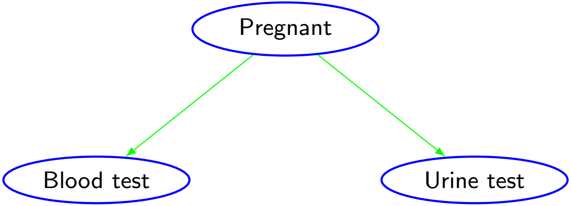
Insemination of a cow

Six weeks after the insemination of a cow, there are two tests: a Blood test and a Urine test.



Insemination of a cow

Six weeks after the insemination of a cow, there are two tests: a Blood test and a Urine test.



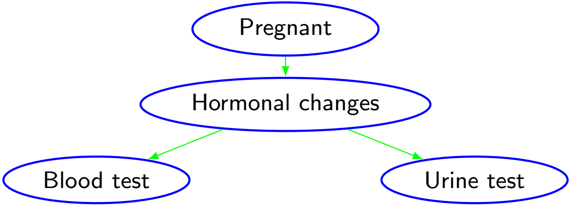
Check the conditional independences

If we know that the cow is pregnant, will a negative blood test then change our expectation for the urine test?

If it will, then the model does not reflect reality!

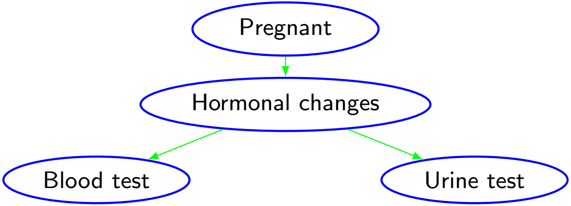
Insemination of a cow: A more correct model

We introduce a mediating variable: Hormonal changes



Insemination of a cow: A more correct model

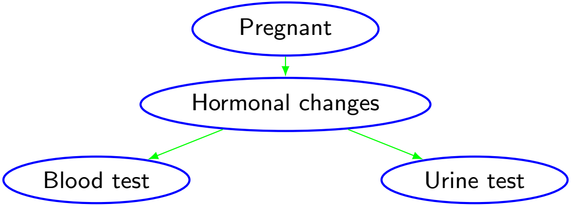
We introduce a mediating variable: Hormonal changes



But does this actually make a difference?

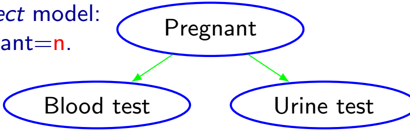
Insemination of a cow: A more correct model

We introduce a mediating variable: Hormonal changes



But does this actually make a difference?

Assume that both tests are negative in the *incorrect* model:
This will overestimate the probability for Pregnant=n.

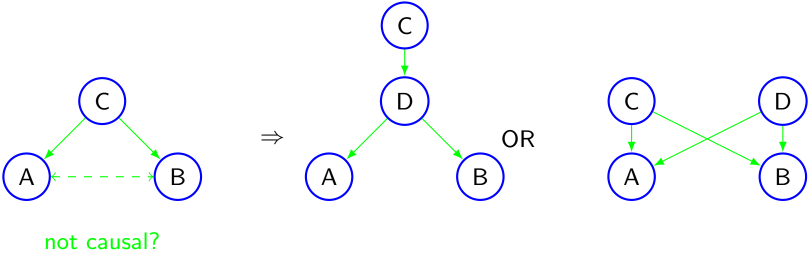


Why mediating variables?

Why do we introduce mediating variables:

- Necessary to catch the correct conditional independences.
- Can ease the specification of the probabilities in the model.

For example: If you find that there is a dependence between two variables A and B, but cannot determine a causal relation: Try with a mediating variable!



A simplified poker game

The game consists of:

- Two players.
- Three cards to each player.
- Two rounds of changing cards (max two cards in the second round)

What kind of hand does my opponent have?

A simplified poker game

The game consists of:

- Two players.
- Three cards to each player.
- Two rounds of changing cards (max two cards in the second round)

What kind of hand does my opponent have?

Hypothesis variable:

OH - {no, 1a, 2v, fl, st, 3v, sf}

Information variables:

FC - {0, 1, 2, 3} and SC - {0, 1, 2}

A simplified poker game

The game consists of:

- Two players.
- Three cards to each player.
- Two rounds of changing cards (max two cards in the second round)

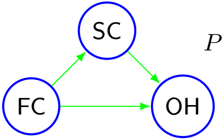
What kind of hand does my opponent have?

Hypothesis variable:

OH - {no, 1a, 2v, fl, st, 3v, sf}

Information variables:

FC - {0, 1, 2, 3} and SC - {0, 1, 2}

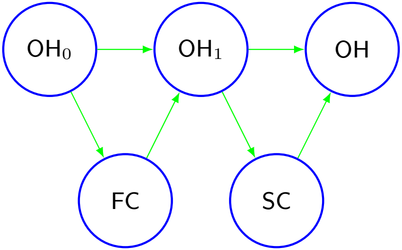


But how do we find:
$P(\text{FC})$, $P(\text{SC}|\text{FC})$ and $P(\text{OH}|\text{SC},\text{FC})$??

A simplified poker game: Mediating variables

Introduce mediating variables:

- The opponent's initial hand, OH_0.
- The opponent's hand after the first change of cards, OH_1.



Note that:

- The states of OH_0 and OH_1 are different from OH.
- We can estimate $P(\mathsf{FC} \mid \mathsf{OH}_0)$, $P(\mathsf{OH}_1 \mid \mathsf{FC}, \mathsf{OH}_0)$, etc. more easily. For example, $P(\mathsf{OH}_1 \mid \mathsf{FC}, \mathsf{OH}_0)$ can be computed by considering the possible cards that could be drawn.

Agenda

Summary

- Bayesian networks (BN) are a wide-spread tool to model uncertainty, and to reason about it. A BN represents conditional independence relations between random variables. It consists of a graph encoding the variable dependencies, and of conditional probability tables (CPTs).
- Given a BN we can recover the full joint probability distribution using the chain rule.
- We can use d-separation to determine whether two variables are independent given some evidence.
- Given a variable order, the BN is small if every variable depends on only a few of its predecessors.
- When designing Bayesian Networks, we need to consider what variables should be added and what are their dependencies. Introducing mediating variables can help to simplify the model.