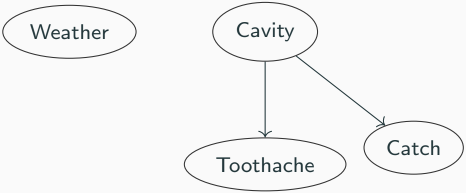
Lecture 22: Probabilistic Reasoning with Bayesian Networks

Last Lecture Takeaway

Conditional Independence reduces the number of probabilities required to specify the Joint Probability Distribution

Representing Conditional Dependencies: Bayesian Networks

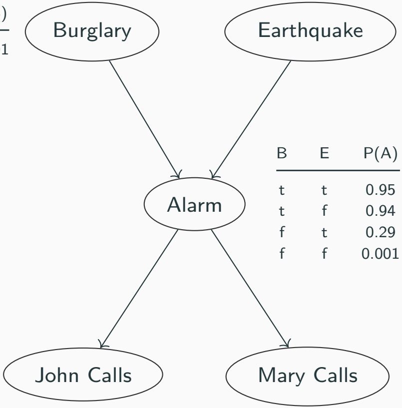
CPDs: $P(X_i|Parents(X_i))$



Arcs and Independence

Each variable is conditionally independent of its **non-descendants**, given its parents

$$\text{If } X \notin Parents^*(Y), \text{ then}$$
$$P(X|Parents(X), Y) = P(X|Parents(X))$$



P(B)
0.001

P(E)
0.002

B	E	P(A)
t	t	0.95
t	f	0.94
f	t	0.29
f	f	0.001

A	P(J)
t	0.7
f	0.01

A	P(M)
t	0.90
f	0.05

...but what is a Bayesian Network? Why?

- **Numeric view**: a Bayesian Network is a compact representation of the Joint Probability Distribution
- **Topological view**: a Bayesian Network is a collection of conditional independence statements.

...but what is a Bayesian Network? Why?

- **Numeric view**: a Bayesian Network is a compact representation of the Joint Probability Distribution
- **Topological view**: a Bayesian Network is a collection of conditional independence statements.

Numeric view

$$P(X_1 = x_1 \wedge \cdots \wedge X_n = x_n)$$

Numeric view

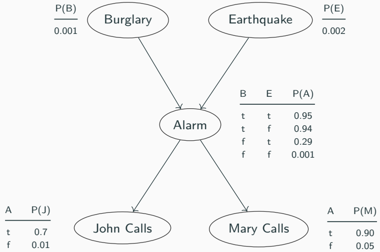
$$P(x_1, \ldots, x_n)$$

Numeric view

$$P(x_1, \ldots, x_n)$$
$$= \prod_{i=1}^{n} P(x_i | Parents(X_i))$$

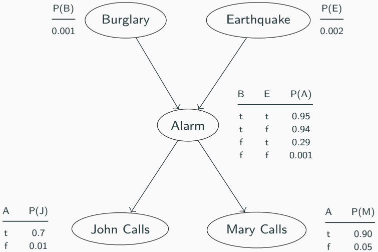
Numeric view

$P(j \wedge m \wedge a \wedge \neg b \wedge \neg e)$



Numeric view

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

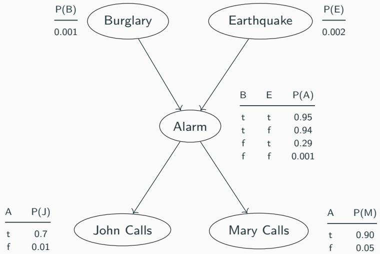


Numeric view

$P(j \wedge m \wedge a \wedge \neg b \wedge \neg e)$

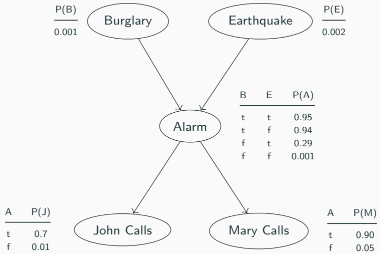
$= P(j|a)P(m|a)P(a|\neg b \wedge \neg e)P(\neg b)P(\neg e)$

$= 0.9 \times 0.7 \times 0.001 \times 0.999 \times 0.998$



Numeric view

$$P(j \wedge m \wedge a \wedge \neg b \wedge \neg e)$$
$$= P(j|a)P(m|a)P(a|\neg b \wedge \neg e)P(\neg b)P(\neg e)$$
$$= 0.9 \times 0.7 \times 0.001 \times 0.999 \times 0.998$$
$$= 0.00062$$



Constructing a BN

$$P(x_1, \ldots, x_n)$$
$$= P(x_n | x_{n-1}, \ldots, x_1) P(x_{n-1}, \ldots, x_1)$$
$$= P(x_n | x_{n-1}, \ldots, x_1) P(x_{n-1} | x_{n-2}, \ldots, x_1) \ldots P(x_2 | x_1) P(x_1)$$
$$= \prod_{i=1}^{n} P(x_i | x_{i-1}, \ldots, x_1)$$

Constructing a BN

$$P(x_1, \ldots, x_n)$$
$$= P(x_n|x_{n-1}, \ldots, x_1)P(x_{n-1}, \ldots, x_1)$$
$$= P(x_n|x_{n-1}, \ldots, x_1)P(x_{n-1}|x_{n-2}, \ldots, x_1) \ldots P(x_2|x_1)P(x_1)$$
$$= \prod_{i=1}^{n} P(x_i|x_{i-1}, \ldots, x_1)$$

Constructing a BN

$$P(x_1, \ldots, x_n)$$
$$= P(x_n|x_{n-1}, \ldots, x_1)P(x_{n-1}, \ldots, x_1)$$
$$= P(x_n|x_{n-1}, \ldots, x_1)P(x_{n-1}|x_{n-2}, \ldots, x_1)\ldots P(x_2|x_1)P(x_1)$$
$$= \prod_{i=1}^{n} P(x_i|x_{i-1}, \ldots, x_1)$$

Constructing a BN

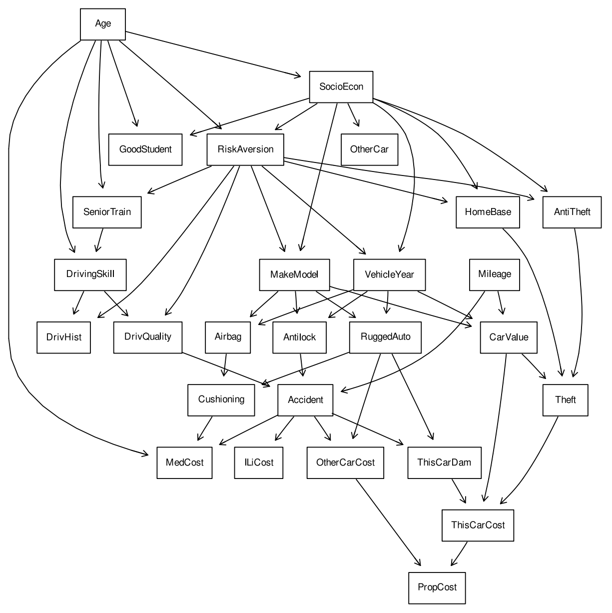
$$
\begin{aligned}
&P(x_1, \ldots, x_n) \\
&= P(x_n | x_{n-1}, \ldots, x_1) P(x_{n-1}, \ldots, x_1) \\
&= P(x_n | x_{n-1}, \ldots, x_1) P(x_{n-1} | x_{n-2}, \ldots, x_1) \ldots P(x_2 | x_1) P(x_1) \\
&= \prod_{i=1}^{n} P(x_i | x_{i-1}, \ldots, x_1)
\end{aligned}
$$

Constructing a BN

$$
\begin{aligned}
&P(x_1, \ldots, x_n) \\
&= P(x_n | x_{n-1}, \ldots, x_1) P(x_{n-1}, \ldots, x_1) \\
&= P(x_n | x_{n-1}, \ldots, x_1) P(x_{n-1} | x_{n-2}, \ldots, x_1) \ldots P(x_2 | x_1) P(x_1) \\
&= \prod_{i=1}^{n} P(x_i | x_{i-1}, \ldots, x_1)
\end{aligned}
$$

$$P(MaryCalls | JohnCalls, Alarm, Earthquake, Burglary) = P(MaryCalls | Alarm)$$

Compactness



Topological Semantics

$$P(x_1, \ldots, x_n) = \prod_{i=1}^{n} P(x_i | Parents(X_i))$$

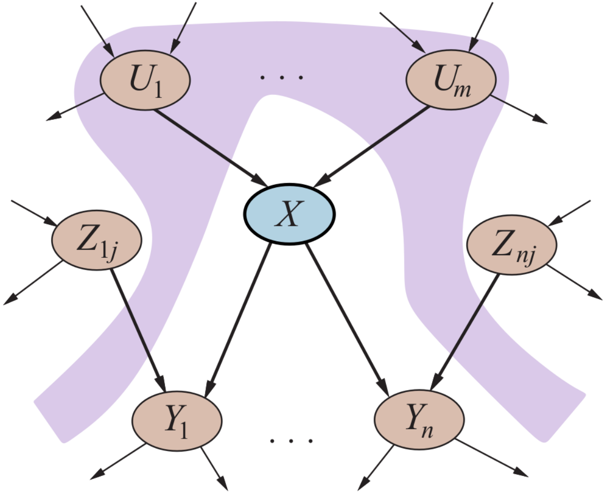
Topological Semantics

$$P(x_1, \ldots, x_n) = \prod_{i=1}^{n} P(x_i | Parents(X_i))$$

1. A node is conditionally independent of its **non-descendants** given its parents
2. A node is conditionally independent of all other nodes, given its parents, children and children's parents (i.e., its **Markov blanket**)

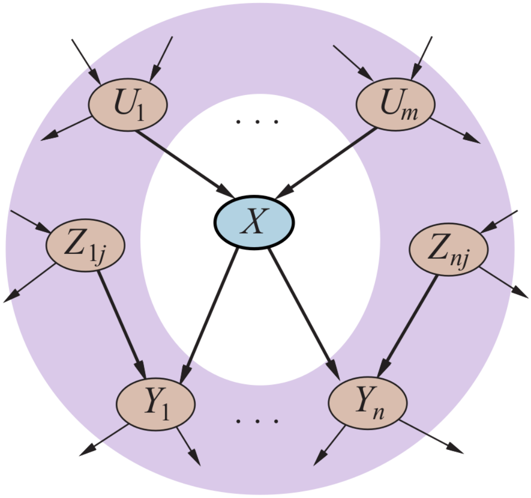
Topological Semantics

A node is conditionally independent of its **non-descendants** given its parents



Topological Semantics

A node is conditionally independent of all other nodes, given its parents, children and children's parents (i.e., its **Markov blanket**)



Toplogical Semantics: Bonus Question

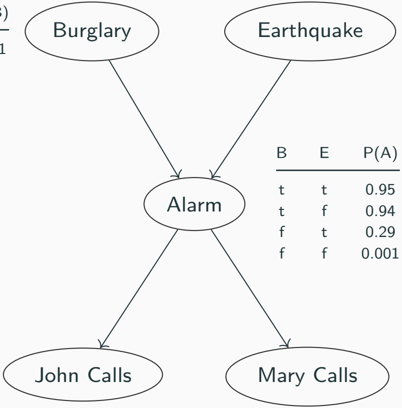
Q: Given evidence of variables $\{Z_1 \ldots Z_i\}$, are variables $\{X_1 \ldots X_j\}$ conditionally independent of variables $\{Y_1 \ldots Y_k\}$?

Toplogical Semantics: Bonus Question

Q: Given evidence of variables $\{Z_1 \ldots Z_i\}$, are variables $\{X_1 \ldots X_j\}$ conditionally independent of variables $\{Y_1 \ldots Y_k\}$?

Prove it?

Context-Specific Independence



P(B)
0.01

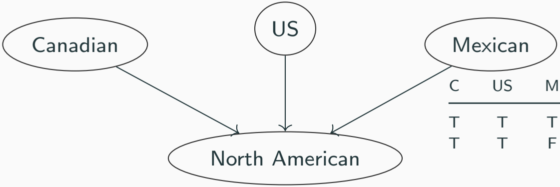
P(E)
0.02

B	E	P(A)
t	t	0.95
t	f	0.94
f	t	0.29
f	f	0.001

A	P(J)
t	0.7
f	0.01

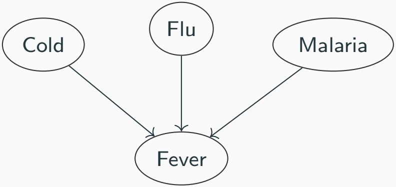
A	P(M)
t	0.90
f	0.05

Context Specific Independence



C	US	M	P(NA)
T	T	T	1
T	T	F	1
		...	
F	F	F	0

Noisy-OR



$$P(\textit{effect}|\textit{cause}) < 1$$

$$P(\neg \textit{fever}|\textit{cold}, \neg \textit{flu}, \neg \textit{malaria}) = 0.6$$
$$P(\neg \textit{fever}|\neg \textit{cold}, \textit{flu}, \neg \textit{malaria}) = 0.2$$
$$P(\neg \textit{fever}|\neg \textit{cold}, \neg \textit{flu}, \textit{malaria}) = 0.1$$

Noisy-OR

$$P(\neg fever | cold, flu, \neg malaria) =$$
$$P(\neg fever | cold, \neg flu, \neg malaria) P(\neg fever | \neg cold, flu, \neg malaria)$$

Noisy-OR

Cold	*Flu*	*Malaria*	$P(fever)$	$P(\neg fever)$
F	F	F		
F	F	T	0.9	**0.1**
F	T	F	0.8	**0.2**
F	T	T		
T	F	F	0.4	**0.6**
T	F	T		
T	T	F		
T	T	T		

$$P(\neg fever | cold, \neg flu, \neg malaria) = 0.6$$
$$P(\neg fever | \neg cold, flu, \neg malaria) = 0.2$$
$$P(\neg fever | \neg cold, \neg flu, malaria) = 0.1$$

Noisy-OR

Cold	*Flu*	*Malaria*	$P(fever)$	$P(\neg fever)$
F	F	F	0.0	1.0
F	F	T	0.9	**0.1**
F	T	F	0.8	**0.2**
F	T	T		
T	F	F	0.4	**0.6**
T	F	T		
T	T	F		
T	T	T		

$$P(\neg fever|cold, \neg flu, \neg malaria) = 0.6$$
$$P(\neg fever|\neg cold, flu, \neg malaria) = 0.2$$
$$P(\neg fever|\neg cold, \neg flu, malaria) = 0.1$$

Noisy-OR

Cold	*Flu*	*Malaria*	$P(fever)$	$P(\neg fever)$
F	F	F	0.0	1.0
F	F	T	0.9	**0.1**
F	T	F	0.8	**0.2**
F	T	T	0.98	$0.02 = 0.2 \times 0.1$
T	F	F	0.4	**0.6**
T	F	T		
T	T	F		
T	T	T		

$$P(\neg fever | cold, \neg flu, \neg malaria) = 0.6$$
$$P(\neg fever | \neg cold, flu, \neg malaria) = 0.2$$
$$P(\neg fever | \neg cold, \neg flu, malaria) = 0.1$$

Noisy-OR

Cold	*Flu*	*Malaria*	$P(fever)$	$P(\neg fever)$
F	F	F	0.0	1.0
F	F	T	0.9	**0.1**
F	T	F	0.8	**0.2**
F	T	T	0.98	$0.02 = 0.2 \times 0.1$
T	F	F	0.4	**0.6**
T	F	T	0.94	$0.06 = 0.6 \times 0.1$
T	T	F		
T	T	T		

$$P(\neg fever|cold, \neg flu, \neg malaria) = 0.6$$
$$P(\neg fever|\neg cold, flu, \neg malaria) = 0.2$$
$$P(\neg fever|\neg cold, \neg flu, malaria) = 0.1$$

Noisy-OR

Cold	*Flu*	*Malaria*	$P(fever)$	$P(\neg fever)$
F	F	F	0.0	1.0
F	F	T	0.9	**0.1**
F	T	F	0.8	**0.2**
F	T	T	0.98	$0.02 = 0.2 \times 0.1$
T	F	F	0.4	**0.6**
T	F	T	0.94	$0.06 = 0.6 \times 0.1$
T	T	F	0.88	$0.12 = 0.6 \times 0.2$
T	T	T		

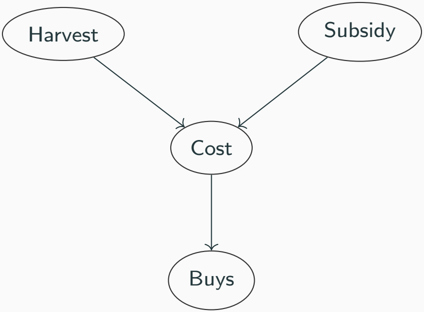
$$P(\neg fever | cold, \neg flu, \neg malaria) = 0.6$$
$$P(\neg fever | \neg cold, flu, \neg malaria) = 0.2$$
$$P(\neg fever | \neg cold, \neg flu, malaria) = 0.1$$

Noisy-OR

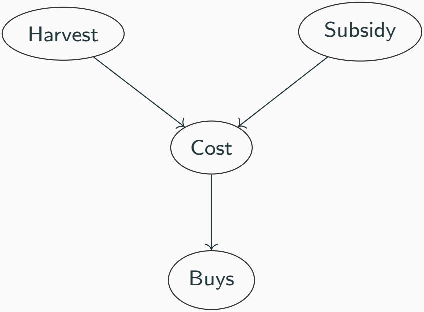
Cold	*Flu*	*Malaria*	$P(fever)$	$P(\neg fever)$
F	F	F	0.0	1.0
F	F	T	0.9	**0.1**
F	T	F	0.8	**0.2**
F	T	T	0.98	$0.02 = 0.2 \times 0.1$
T	F	F	0.4	**0.6**
T	F	T	0.94	$0.06 = 0.6 \times 0.1$
T	T	F	0.88	$0.12 = 0.6 \times 0.2$
T	T	T	0.988	$0.012 = 0.6 \times 0.2 \times 0.1$

$$P(\neg fever | cold, \neg flu, \neg malaria) = 0.6$$
$$P(\neg fever | \neg cold, flu, \neg malaria) = 0.2$$
$$P(\neg fever | \neg cold, \neg flu, malaria) = 0.1$$

BNs with Continuous Variables



BNs with Continuous Variables



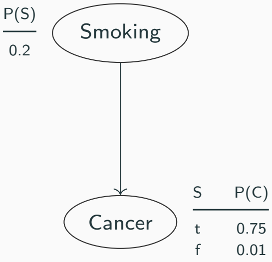
$$P(c|h, subsidy) = \mathcal{N}(c; a_t h + b_t, \sigma_t^2) = \frac{1}{\sigma_t\sqrt{2\pi}} \exp -\frac{1}{2}\left(\frac{c-(a_t h + b_t)}{\sigma_t}\right)^2$$

Bonus: What Bayesian Networks are Not

Survey of 10000 people over 65

- Smokers: 2000
- Non Smokers: 8000

- Non Smokers with Cancer: 80
- Smokers with Cancer: 1500

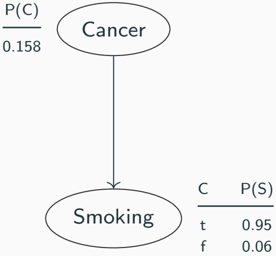


Bonus: What Bayesian Networks are Not

Survey of 10000 people over 65

- Smokers: 2000
- Non Smokers: 8000

- Non Smokers with Cancer: 80
- Smokers with Cancer: 1500

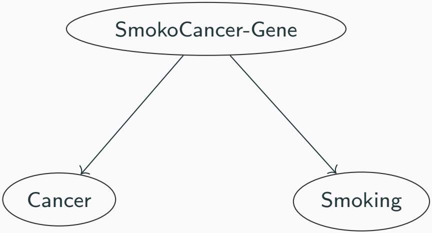


Bonus: What Bayesian Networks are Not

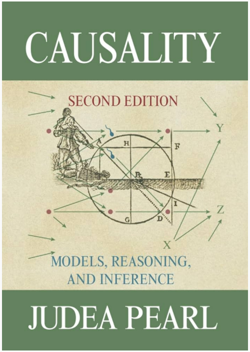
Survey of 10000 people over 65

- Smokers: 2000
- Non Smokers: 8000

- Non Smokers with Cancer: 80
- Smokers with Cancer: 1500



- Level 11 Course: Methods for Causal Inference (INFR11207)



Summary

- Bayesian Networks: Topological and Numerical semantics
- Context Specific Independence
- Continuous Variables and Hybrid Networks