CSE 150A-250A AI: Probabilistic Methods

Lecture 3

Fall 2025

Trevor Bonjour
Department of Computer Science and Engineering
University of California, San Diego

Slides adapted from previous versions of the course (Prof. Lawrence, Prof. Alvarado, Prof Berg-Kirkpatrick)

Agenda

Review

Joint Distributions and Inference

Alarm Example

Belief networks

Review

Review

- Types of probabilities:

$P(X, Y)$	joint
$P(Y \mid X)$	conditional
$P(X)$	unconditional (or marginal)

Review

- Types of probabilities:

$P(X, Y)$	joint
$P(Y\|X)$	conditional
$P(X)$	unconditional (or marginal)

- Useful rules:

$$P(A, B, C, \ldots) = P(A)\, P(B|A)\, P(C|A, B) \ldots \quad \text{product rule}$$

$$P(X|Y) = P(Y|X)P(X)/P(Y) \quad \text{Bayes rule}$$

$$P(X) = \sum_y P(X, Y = y) \quad \text{marginalization}$$

Review

- **Types of probabilities:**

$P(X, Y)$	joint
$P(Y \mid X)$	conditional
$P(X)$	unconditional (or marginal)

- **Useful rules:**

$$P(A, B, C, \ldots) = P(A)\, P(B \mid A)\, P(C \mid A, B) \ldots \quad \textbf{product rule}$$

$$P(X \mid Y) = P(Y \mid X) P(X) / P(Y) \quad \textbf{Bayes rule}$$

$$P(X) = \sum_y P(X, Y = y) \quad \textbf{marginalization}$$

- **Conditioning on background evidence E:**

$$P(A, B, C, \ldots \mid E) = P(A \mid E)\, P(B \mid A, E)\, P(C \mid A, B, E) \ldots$$

$$P(X \mid Y, E) = P(Y \mid X, E) P(X \mid E) / P(Y \mid E)$$

$$P(X \mid E) = \sum_y P(X, Y = y \mid E)$$

Marginal and conditional independence

- Marginal independence

$$\left.\begin{aligned} P(X|Y) &= P(X) \\ P(Y|X) &= P(Y) \\ P(X,Y) &= P(X)\,P(Y) \end{aligned}\right\}\ \textit{Each of these implies the other two.}$$

Marginal and conditional independence

- Marginal independence

$$
\left.\begin{aligned}
P(X|Y) &= P(X) \\
P(Y|X) &= P(Y) \\
P(X, Y) &= P(X)\,P(Y)
\end{aligned}\right\}
$$

Each of these implies the other two.

- Conditional independence

$$
\left.\begin{aligned}
P(X|Y, E) &= P(X|E) \\
P(Y|X, E) &= P(Y|E) \\
P(X, Y|E) &= P(X|E)\,P(Y|E)
\end{aligned}\right\}
$$

Each of these also implies the other two.

Example of conditional dependence

- B and E are marginally independent:

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

Example of conditional dependence



- B and E are marginally independent:

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

- But B and E are conditionally dependent given A:

$$
\begin{aligned}
P(B|A) &\neq P(B|E,A) \\
P(E|A) &\neq P(E|B,A) \\
P(B,E|A) &\neq P(B|A)\,P(E|A)
\end{aligned}
$$

Joint Distributions and Inference

Why Joints Matter

The joint distribution $P(X_1, \ldots, X_n)$ is a **complete description of uncertainty**.

Why Joints Matter

The joint distribution $P(X_1, \ldots, X_n)$ is a **complete description of uncertainty**.

From the full joint distribution $P(X_1, \ldots, X_n)$, what kinds of probabilities can we compute?

A. Only marginals, $P(X_i)$

B. Only conditionals, $P(X_i | X_j)$

C. Any marginal or conditional over the variables

D. None of the above

Why Joints Matter

The joint distribution $P(X_1, \ldots, X_n)$ is a **complete description of uncertainty**.

From the full joint distribution $P(X_1, \ldots, X_n)$, what kinds of probabilities can we compute?

A. Only marginals, $P(X_i)$
B. Only conditionals, $P(X_i|X_j)$
C. Any marginal or conditional over the variables
D. None of the above

Inference: Compute the **posterior** distribution of query variables given observed evidence.

Motivation

- Model complexity

  Suppose $X_i \in \{0, 1\}$ are binary random variables.

- **Model complexity**

Suppose $X_i \in \{0, 1\}$ are binary random variables.

How many numbers do we need to specify the joint distribution of $P(X_1 = x_1, \ldots, X_n = x_n)$?

A. $O(n)$

B. $O(2^n)$

C. $O(n^2)$

D. $O(logn)$

$P(x_1 = 1, x_2 = 1)$

$P(x_1 = 0, x_2 = 1)$

$P(x_1 = 1, x_2 = 0)$

$= 0 \quad = 0$

- **Model complexity**

Suppose $X_i \in \{0, 1\}$ are binary random variables.

How many numbers do we need to specify the joint distribution of $P(X_1 = x_1, \ldots, X_n = x_n)$?

A. $O(n)$

B. $O(2^n)$

C. $O(n^2)$

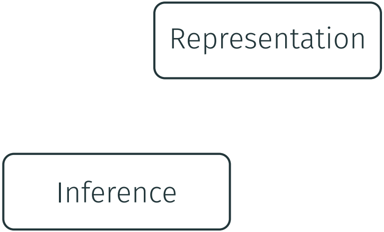
D. $O(logn)$

It requires $O(2^n)$ numbers to specify the joint distribution $P(X_1 = x_1, \ldots, X_n = x_n)$.

Conceptual and Practical Goals

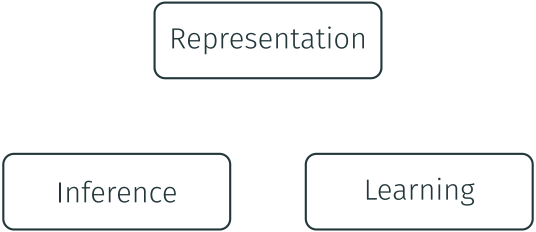
Representation

- **Representation**: compactly encode the joint.



- **Representation**: compactly encode the joint.
- **Inference**: answer queries given evidence.

Conceptual and Practical Goals

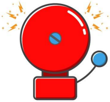


- **Representation**: compactly encode the joint.
- **Inference**: answer queries given evidence.
- **Learning**: estimate structure/parameters from data.

Alarm Example

Alarm example






- Binary random variables

$B \in \{0,1\}$	Was there a burglary?
$E \in \{0,1\}$	Was there an earthquake?
$A \in \{0,1\}$	Was the alarm triggered?
$J \in \{0,1\}$	Did Jamal call?
$M \in \{0,1\}$	Did Maya call?

Joint distribution

- Product rule

  $P(B, E, A, J, M)$

Joint distribution

- Product rule

$$P(B, E, A, J, M)$$

$$= \quad P(B)$$

Joint distribution

- Product rule

$$
\begin{aligned}
& P(B, E, A, J, M) \\
& \quad = \quad P(B)\, P(E|B)
\end{aligned}
$$

Joint distribution

- Product rule

$$P(B, E, A, J, M)$$

$$= \quad P(B)\, P(E|B)\, P(A|B, E)$$

Joint distribution

- Product rule

$$P(B, E, A, J, M)$$

$$= \quad P(B)\, P(E|B)\, P(A|B, E)\, P(J|B, E, A)$$

Joint distribution

- Product rule

$$P(B, E, A, J, M)$$

$$= P(B)\, P(E|B)\, P(A|B, E)\, P(J|B, E, A)\, P(M|B, E, A, J)$$

Joint distribution

- Product rule

$$P(B, E, A, J, M)$$
$$= \quad P(B)\, P(E|B)\, P(A|B, E)\, P(J|B, E, A)\, P(M|B, E, A, J)$$

Note: the above is true no matter what the variables signify.

Joint distribution

- Product rule

$$\begin{aligned} & P(B, E, A, J, M) \\ & \quad = P(B)\, P(E|B)\, P(A|B, E)\, P(J|B, E, A)\, P(M|B, E, A, J) \end{aligned}$$

Note: the above is true no matter what the variables signify.

- Domain-specific assumptions

Joint distribution

- Product rule

$$P(B, E, A, J, M)$$

$$= \quad P(B)\, P(E|B)\, P(A|B, E)\, P(J|B, E, A)\, P(M|B, E, A, J)$$

Note: the above is true no matter what the variables signify.

- Domain-specific assumptions

$$P(E|B) \quad = \quad P(E)$$ marginal independence

Joint distribution

- Product rule

$$
\begin{aligned}
& P(B, E, A, J, M) \\
& \quad = \quad P(B)\, P(E|B)\, P(A|B, E)\, P(J|B, E, A)\, P(M|B, E, A, J)
\end{aligned}
$$

Note: the above is true no matter what the variables signify.

- Domain-specific assumptions

$$
\begin{aligned}
P(E|B) &= P(E) && \text{marginal independence} \\
P(J|B, E, A) &= P(J|A) && \text{conditional independence}
\end{aligned}
$$

Joint distribution

- Product rule

$$P(B, E, A, J, M)$$
$$= P(B)\,P(E|B)\,P(A|B, E)\,P(J|B, E, A)\,P(M|B, E, A, J)$$

Note: the above is true no matter what the variables signify.

- Domain-specific assumptions

$$\begin{aligned} P(E|B) &= P(E) && \text{marginal independence} \\ P(J|B, E, A) &= P(J|A) && \text{conditional independence} \\ P(M|B, E, A, J) &= P(M|A) && \text{conditional independence} \end{aligned}$$

Completing the model

- Joint distribution

$$P(B, E, A, J, M)$$

$$= P(B)\, P(E|B)\, P(A|B, E)\, P(J|B, E, A)\, P(M|B, E, A, J)$$

Completing the model

- Joint distribution

$$
\begin{aligned}
&P(B, E, A, J, M) \\
&\quad = P(B)\, P(E|B)\, P(A|B, E)\, P(J|B, E, A)\, P(M|B, E, A, J) \\
&\quad = P(B)\, P(E)\, P(A|B, E)\, P(J|A)\, P(M|A)
\end{aligned}
$$

Completing the model

- Joint distribution

$$
\begin{aligned}
& P(B, E, A, J, M) \\
&\quad = P(B)\, P(E|B)\, P(A|B, E)\, P(J|B, E, A)\, P(M|B, E, A, J) \\
&\quad = P(B)\, P(E)\, P(A|B, E)\, P(J|A)\, P(M|A)
\end{aligned}
$$

- Conditional probability tables (CPTs)

Completing the model

- Joint distribution

$$
\begin{aligned}
P(B,E,A,J,M) \\
&= P(B)\,P(E|B)\,P(A|B,E)\,P(J|B,E,A)\,P(M|B,E,A,J) \\
&= P(B)\,P(E)\,P(A|B,E)\,P(J|A)\,P(M|A)
\end{aligned}
$$

- Conditional probability tables (CPTs)

$P(B=1) = 0.001$

$P(E=1) = 0.002$

Completing the model

- Joint distribution

$$
\begin{aligned}
P(B, E, A, J, M) \\
&= P(B)\, P(E|B)\, P(A|B, E)\, P(J|B, E, A)\, P(M|B, E, A, J) \\
&= P(B)\, P(E)\, P(A|B, E)\, P(J|A)\, P(M|A)
\end{aligned}
$$

- Conditional probability tables (CPTs)

$P(B=1) = 0.001$

$P(E=1) = 0.002$

B	E	$P(A=1 \mid B, E)$
0	0	0.001
1	0	0.94
0	1	0.29
1	1	0.95

Completing the model

- Joint distribution

$$
\begin{aligned}
& P(B, E, A, J, M) \\
&\quad = P(B)\, P(E|B)\, P(A|B, E)\, P(J|B, E, A)\, P(M|B, E, A, J) \\
&\quad = P(B)\, P(E)\, P(A|B, E)\, P(J|A)\, P(M|A)
\end{aligned}
$$

- Conditional probability tables (CPTs)

$P(B=1) = 0.001$

$P(E=1) = 0.002$

B	E	$P(A=1\|B, E)$
0	0	0.001
1	0	0.94
0	1	0.29
1	1	0.95

A	$P(J=1\|A)$
0	0.05
1	0.9

Completing the model

- Joint distribution

$J \perp B \mid A$

$J \perp E \mid A$

$$
\begin{aligned}
& P(B, E, A, J, M) \\
& \quad = P(B)\, P(E|B)\, P(A|B, E)\, P(J|B, E, A)\, P(M|B, E, A, J) \\
& \quad = P(B)\, P(E)\, P(A|B, E)\, P(J|A)\, P(M|A)
\end{aligned}
$$

- Conditional probability tables (CPTs)

$P(B=1) = 0.001$

$P(E=1) = 0.002$

B	E	$P(A=1\|B, E)$
0	0	0.001
1	0	0.94
0	1	0.29
1	1	0.95

A	$P(J=1\|A)$
0	0.05
1	0.9

A	$P(M=1\|A)$
0	0.01
1	0.7

Inference

- Joint probabilities are easy to compute:

Inference

- Joint probabilities are easy to compute:

$$P(B=1, E=0, A=1, J=1, M=1)$$

Inference

- Joint probabilities are easy to compute:

$$
\begin{aligned}
&P(B=1, E=0, A=1, J=1, M=1) \\
&\quad = \quad P(B=1)\, P(E=0)\, P(A=1|B=1, E=0)\, P(J=1|A=1)\, P(M=1|A=1)
\end{aligned}
$$

Inference

- Joint probabilities are easy to compute:

$$
\begin{aligned}
&P(B=1, E=0, A=1, J=1, M=1) \\
&\quad = P(B=1)\, P(E=0)\, P(A=1|B=1, E=0)\, P(J=1|A=1)\, P(M=1|A=1) \\
&\quad = (0.001)(1-0.002)(0.94)(0.9)(0.7)
\end{aligned}
$$

Inference

- Joint probabilities are easy to compute:

$$
\begin{aligned}
&P(B=1, E=0, A=1, J=1, M=1) \\
&\quad = \quad P(B=1)\, P(E=0)\, P(A=1|B=1, E=0)\, P(J=1|A=1)\, P(M=1|A=1) \\
&\quad = \quad (0.001)(1-0.002)(0.94)(0.9)(0.7)
\end{aligned}
$$

- Any inference can be expressed in terms of joint probabilities:

Inference

- Joint probabilities are easy to compute:

$$
\begin{aligned}
&P(B{=}1, E{=}0, A{=}1, J{=}1, M{=}1) \\
&\quad = P(B{=}1)\, P(E{=}0)\, P(A{=}1|B{=}1, E{=}0)\, P(J{=}1|A{=}1)\, P(M{=}1|A{=}1) \\
&\quad = (0.001)(1 - 0.002)(0.94)(0.9)(0.7)
\end{aligned}
$$

- Any inference can be expressed in terms of joint probabilities:

$$P(B{=}1, E{=}0|M{=}1)$$

- Joint probabilities are easy to compute:

$$\begin{aligned} &P(B=1, E=0, A=1, J=1, M=1) \\ &\quad = P(B=1)\, P(E=0)\, P(A=1|B=1, E=0)\, P(J=1|A=1)\, P(M=1|A=1) \\ &\quad = (0.001)(1-0.002)(0.94)(0.9)(0.7) \end{aligned}$$

- Any inference can be expressed in terms of joint probabilities:

$$\begin{aligned} &P(B=1, E=0|M=1) \\ &\quad = \frac{P(B=1, E=0, M=1)}{P(M=1)} \qquad \boxed{\text{product rule}} \end{aligned}$$

Inference

- Joint probabilities are easy to compute:

$$
\begin{aligned}
& P(B{=}1, E{=}0, A{=}1, J{=}1, M{=}1) \\
&\quad = P(B{=}1)\, P(E{=}0)\, P(A{=}1|B{=}1, E{=}0)\, P(J{=}1|A{=}1)\, P(M{=}1|A{=}1) \\
&\quad = (0.001)(1-0.002)(0.94)(0.9)(0.7)
\end{aligned}
$$

- Any inference can be expressed in terms of joint probabilities:

$$
\begin{aligned}
& P(B{=}1, E{=}0|M{=}1) \\
&\quad = \frac{P(B{=}1, E{=}0, M{=}1)}{P(M{=}1)} \qquad \boxed{\text{product rule}} \\
&\quad = \frac{\sum_{a,j} P(B{=}1, E{=}0, A{=}a, J{=}j, M{=}1)}{\sum_{b',e',a',j'} P(B{=}b', E{=}e', A{=}a', J{=}j', M{=}1)} \qquad \boxed{\text{marginalization}}
\end{aligned}
$$

Inference

- Joint probabilities are easy to compute:

$$
\begin{aligned}
&P(B=1, E=0, A=1, J=1, M=1) \\
&\quad = \quad P(B=1)\, P(E=0)\, P(A=1|B=1, E=0)\, P(J=1|A=1)\, P(M=1|A=1) \\
&\quad = \quad (0.001)(1-0.002)(0.94)(0.9)(0.7)
\end{aligned}
$$

- Any inference can be expressed in terms of joint probabilities:

$$
\begin{aligned}
&P(B=1, E=0|M=1) \\
&\quad = \quad \frac{P(B=1, E=0, M=1)}{P(M=1)} \qquad \boxed{\text{product rule}} \\
&\quad = \quad \frac{\sum_{a,j} P(B=1, E=0, A=a, J=j, M=1)}{\sum_{b',e',a',j'} P(B=b', E=e', A=a', J=j', M=1)} \qquad \boxed{\text{marginalization}}
\end{aligned}
$$

But this approach can be very inefficient!

Efficient inference

How to perform inference most efficiently?

Efficient inference

How to perform inference most efficiently?

1. Visualize models as directed acyclic graphs.
2. Exploit graph structure to organize and simplify calculations.

Efficient inference

How to perform inference most efficiently?

1. Visualize models as directed acyclic graphs.
2. Exploit graph structure to organize and simplify calculations.

We'll spend today on (1) and next lecture on (2).

Visualizing the model

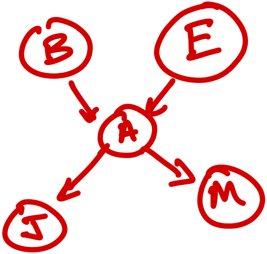
- Joint distribution

$$
\begin{aligned}
P(B, E, A, J, M) \\
&= P(B)\, P(E|B)\, P(A|B, E)\, P(J|B, E, A)\, P(M|B, E, A, J) \\
&= P(B)\, P(E)\, P(A|B, E)\, P(J|A)\, P(M|A)
\end{aligned}
$$

- Joint distribution

$$
\begin{aligned}
P(B, E, A, J, M) \\
&= P(B)\, P(E|B)\, P(A|B, E)\, P(J|B, E, A)\, P(M|B, E, A, J) \\
&= P(B)\, P(E)\, P(A|B, E)\, P(J|A)\, P(M|A)
\end{aligned}
$$

- Directed acyclic graph (DAG)



- Joint distribution

$$\begin{aligned} P(B, E, A, J, M) \\ &= P(B)\,P(E|B)\,P(A|B, E)\,P(J|B, E, A)\,P(M|B, E, A, J) \\ &= P(B)\,P(E)\,P(A|B, E)\,P(J|A)\,P(M|A) \end{aligned}$$

- Directed acyclic graph (DAG)

Visualizing the model

- Joint distribution

$$
\begin{aligned}
P(B, E, A, J, M) & \\
&= P(B)\, P(E|B)\, P(A|B, E)\, P(J|B, E, A)\, P(M|B, E, A, J) \\
&= P(B)\, P(E)\, P(A|B, E)\, P(J|A)\, P(M|A)
\end{aligned}
$$

- Directed acyclic graph (DAG)

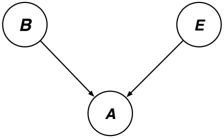
Absent edges encode assumptions of independence.

Visualizing the model

- Joint distribution

$$\begin{aligned} &P(B, E, A, J, M) \\ &\quad = P(B)\, P(E|B)\, P(A|B, E)\, P(J|B, E, A)\, P(M|B, E, A, J) \\ &\quad = P(B)\, P(E)\, P(A|B, E)\, P(J|A)\, P(M|A) \end{aligned}$$

- Directed acyclic graph (DAG)



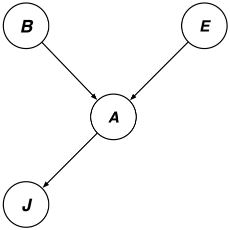
Absent edges encode assumptions of independence.

Visualizing the model

- Joint distribution

$$
\begin{aligned}
P(B, E, A, J, M) \\
&= P(B)\, P(E|B)\, P(A|B, E)\, P(J|B, E, A)\, P(M|B, E, A, J) \\
&= P(B)\, P(E)\, P(A|B, E)\, P(J|A)\, P(M|A)
\end{aligned}
$$

- Directed acyclic graph (DAG)



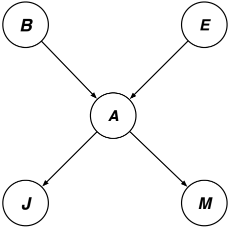
Absent edges encode assumptions of independence.

Visualizing the model

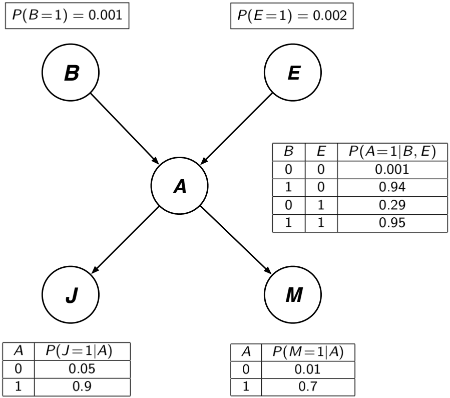
- Joint distribution

$$
\begin{aligned}
P(B,E,A,J,M) & \\
&= P(B)\,P(E|B)\,P(A|B,E)\,P(J|B,E,A)\,P(M|B,E,A,J) \\
&= P(B)\,P(E)\,P(A|B,E)\,P(J|A)\,P(M|A)
\end{aligned}
$$

- Directed acyclic graph (DAG)



Absent edges encode assumptions of independence.



B	E	$P(A=1 \mid B,E)$
0	0	0.001
1	0	0.94
0	1	0.29
1	1	0.95

A	$P(J=1 \mid A)$
0	0.05
1	0.9

A	$P(M=1 \mid A)$
0	0.01
1	0.7

This visual representation of the joint distribution is called a **belief network** (or a **Bayesian network**, or a **probabilistic graphical model**).

Belief networks

Definition

A **belief network** (BN) is a directed acyclic graph (DAG) in which:

Definition

A **belief network** (BN) is a directed acyclic graph (DAG) in which:

1. Nodes represent random variables.

Definition

A **belief network** (BN) is a directed acyclic graph (DAG) in which:

1. Nodes represent random variables.
2. Edges represent dependencies.

Definition

A **belief network** (BN) is a directed acyclic graph (DAG) in which:

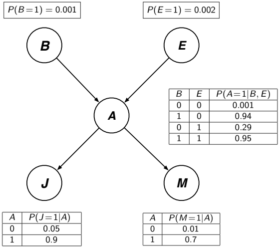
1. Nodes represent random variables.
2. Edges represent dependencies.
3. Conditional probability tables (CPTs) describe how each node depends on its parents.

Definition

A **belief network** (BN) is a directed acyclic graph (DAG) in which:

1. Nodes represent random variables.
2. Edges represent dependencies.
3. Conditional probability tables (CPTs) describe how each node depends on its parents.

BN = DAG + CPTs



From distributions to graphs

From distributions to graphs

- It is always true from the product rule that

$$P(X_1, X_2, \ldots, X_n) \quad = \quad P(X_1)\, P(X_2|X_1) \; \ldots \; P(X_n|X_1, \ldots, X_{n-1})$$

From distributions to graphs

- It is always true from the product rule that

$$\begin{aligned} P(X_1, X_2, \ldots, X_n) &= P(X_1)\, P(X_2|X_1)\ \ldots\ P(X_n|X_1, \ldots, X_{n-1}) \\ &= \prod_{i=1}^{n} P(X_i|X_1, X_2, \ldots, X_{i-1}) \end{aligned}$$

From distributions to graphs

- It is always true from the product rule that

$$\begin{aligned} P(X_1, X_2, \ldots, X_n) &= P(X_1)\, P(X_2|X_1)\, \ldots\, P(X_n|X_1, \ldots, X_{n-1}) \\ &= \prod_{i=1}^{n} P(X_i|X_1, X_2, \ldots, X_{i-1}) \end{aligned}$$

- But suppose in a particular domain that

$$P(X_i|X_1, X_2, \ldots, X_{i-1}) \;=\; P(X_i|\mathrm{parents}(X_i)),$$

  where $\mathrm{parents}(X_i)$ is a subset of $\{X_1, \ldots, X_{i-1}\}$.

From distributions to graphs

- It is always true from the product rule that

$$\begin{aligned} P(X_1, X_2, \ldots, X_n) &= P(X_1)\, P(X_2|X_1)\, \ldots\, P(X_n|X_1, \ldots, X_{n-1}) \\ &= \prod_{i=1}^{n} P(X_i|X_1, X_2, \ldots, X_{i-1}) \end{aligned}$$

- But suppose in a particular domain that

$$P(X_i|X_1, X_2, \ldots, X_{i-1}) = P(X_i|\mathrm{parents}(X_i)),$$

  where $\mathrm{parents}(X_i)$ is a subset of $\{X_1, \ldots, X_{i-1}\}$.

- Big idea: represent conditional dependencies by a DAG.

Constructing a belief network

Three steps:

Constructing a belief network

Three steps:

1. Choose your random variables of interest.

Constructing a belief network

Three steps:

1. Choose your random variables of interest.

2. Choose an ordering of these variables

Constructing a belief network

Three steps:

1. Choose your random variables of interest.
2. Choose an ordering of these variables (e.g., $X_1, X_2, \ldots, X_n$).

Constructing a belief network

Three steps:

1. Choose your random variables of interest.
2. Choose an ordering of these variables (e.g., $X_1, X_2, \ldots, X_n$).
3. While there are variables left:

Constructing a belief network

Three steps:

1. Choose your random variables of interest.
2. Choose an ordering of these variables (e.g., $X_1, X_2, \ldots, X_n$).
3. While there are variables left:
   (a) add the node X_i to the network

Constructing a belief network

Three steps:

1. Choose your random variables of interest.
2. Choose an ordering of these variables (e.g., $X_1, X_2, \ldots, X_n$).
3. While there are variables left:
   (a) add the node X_i to the network
   (b) set the parents of X_i to be the minimal subset satisfying

$$P(X_1, X_2, \ldots, X_n) \;=\; \prod_{i=1}^{n} P(X_i|\mathrm{parents}(X_i)),$$

Constructing a belief network

Three steps:

1. Choose your random variables of interest.
2. Choose an ordering of these variables (e.g., $X_1, X_2, \ldots, X_n$).
3. While there are variables left:
   (a) add the node X_i to the network
   (b) set the parents of X_i to be the minimal subset satisfying
   $$P(X_1, X_2, \ldots, X_n) = \prod_{i=1}^{n} P(X_i | \mathrm{parents}(X_i)),$$
   (c) define the conditional probability table $P(X_i | \mathrm{parents}(X_i))$

Node ordering

Node ordering

- Best ordering:

  Add the "root causes," then the variables they influence, then the next variables that are influenced, etc.

Node ordering

- Best ordering:

  Add the "root causes," then the variables they influence, then the next variables that are influenced, etc.

- Example:

  In the alarm world, a natural ordering is (B, E, A, J, M).

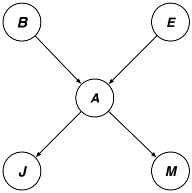
Node ordering

- Best ordering:

  Add the “root causes,” then the variables they influence, then the next variables that are influenced, etc.

- Example:

  In the alarm world, a natural ordering is (B, E, A, J, M).



Node ordering

- What happens if we choose an unnatural ordering?

Node ordering

- What happens if we choose an unnatural ordering?

Ex: (M, J, A, E, B)

- What happens if we choose an unnatural ordering?

Ex: (M, J, A, E, B)

- Adding nodes with this ordering:

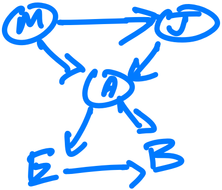
$P(M, J, A, E, B)$ = P(M) P(J|M) P(A|J M) P(E|A J M) P(B|E A J M)

P(J) ~~✗~~ P(J|M)

P(E|A)

P(B|A, E)

P(B|A)



Node ordering

- What happens if we choose an unnatural ordering?

$$\text{Ex: } (M, J, A, E, B)$$

- Adding nodes with this ordering:

$$\begin{aligned} &P(M, J, A, E, B) \\ &\quad = \quad P(M)\, P(J|M)\, P(A|J, M)\, P(E|A, J, M)\, P(B|E, A, J, M) \end{aligned}$$

Node ordering

- What happens if we choose an unnatural ordering?

$$\text{Ex: } (M, J, A, E, B)$$

- Adding nodes with this ordering:

$$\begin{aligned}
&P(M, J, A, E, B) \\
&\quad = \quad P(M)\, P(J|M)\, P(A|J, M)\, P(E|A, J, M)\, P(B|E, A, J, M) \\
&\quad = \quad P(M)
\end{aligned}$$

Node ordering

- What happens if we choose an unnatural ordering?

Ex: (M, J, A, E, B)

- Adding nodes with this ordering:

$$
\begin{aligned}
& P(M, J, A, E, B) \\
& \quad = \quad P(M)\, P(J|M)\, P(A|J, M)\, P(E|A, J, M)\, P(B|E, A, J, M) \\
& \quad = \quad P(M)\, P(J|M)
\end{aligned}
$$

Node ordering

- What happens if we choose an unnatural ordering?

Ex: (M, J, A, E, B)

- Adding nodes with this ordering:

$$
\begin{aligned}
& P(M, J, A, E, B) \\
& \quad = \quad P(M)\, P(J|M)\, P(A|J, M)\, P(E|A, J, M)\, P(B|E, A, J, M) \\
& \quad = \quad P(M)\, P(J|M)\, P(A|J, M)
\end{aligned}
$$

Node ordering

- What happens if we choose an unnatural ordering?

$$\text{Ex: } (M, J, A, E, B)$$

- Adding nodes with this ordering:

$$\begin{aligned} & P(M, J, A, E, B) \\ & \quad = \quad P(M)\, P(J|M)\, P(A|J, M)\, P(E|A, J, M)\, P(B|E, A, J, M) \\ & \quad = \quad P(M)\, P(J|M)\, P(A|J, M)\, P(E|A) \end{aligned}$$

Node ordering

- What happens if we choose an unnatural ordering?

$$\text{Ex: } (M, J, A, E, B)$$

- Adding nodes with this ordering:

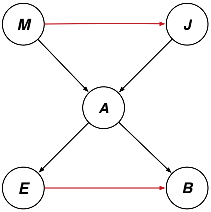
$$\begin{aligned} & P(M, J, A, E, B) \\ & \quad = \quad P(M)\, P(J|M)\, P(A|J, M)\, P(E|A, J, M)\, P(B|E, A, J, M) \\ & \quad = \quad P(M)\, P(J|M)\, P(A|J, M)\, P(E|A)\, P(B|A, E) \end{aligned}$$

- What happens if we choose an unnatural ordering?

$$\text{Ex: } (M, J, A, E, B)$$

- Adding nodes with this ordering:

$$\begin{aligned} & P(M, J, A, E, B) \\ & \quad = \quad P(M)\, P(J|M)\, P(A|J, M)\, P(E|A, J, M)\, P(B|E, A, J, M) \\ & \quad = \quad P(M)\, P(J|M)\, P(A|J, M)\, P(E|A)\, P(B|A, E) \end{aligned}$$



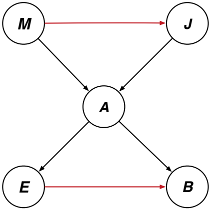
Node ordering

- What happens if we choose an unnatural ordering?

$$\text{Ex: } (M, J, A, E, B)$$

- Adding nodes with this ordering:

$$\begin{aligned} & P(M, J, A, E, B) \\ & \quad = \quad P(M)\, P(J|M)\, P(A|J, M)\, P(E|A, J, M)\, P(B|E, A, J, M) \\ & \quad = \quad P(M)\, P(J|M)\, P(A|J, M)\, P(E|A)\, P(B|A, E) \end{aligned}$$



This belief network has two extra edges.

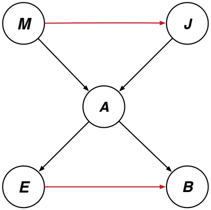
Node ordering

- What happens if we choose an unnatural ordering?

$$\text{Ex: } (M, J, A, E, B)$$

- Adding nodes with this ordering:

$$\begin{aligned} & P(M, J, A, E, B) \\ & \quad = P(M)\, P(J|M)\, P(A|J, M)\, P(E|A, J, M)\, P(B|E, A, J, M) \\ & \quad = P(M)\, P(J|M)\, P(A|J, M)\, P(E|A)\, P(B|A, E) \end{aligned}$$



This belief network has two extra edges.
This DAG does not show $P(B) = P(B|E)$.
This DAG does not show $P(M|A) = P(M|A, J)$.

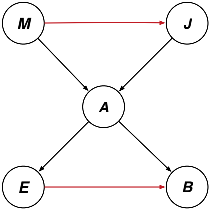
Node ordering

- What happens if we choose an unnatural ordering?

$$\text{Ex: } (M, J, A, E, B)$$

- Adding nodes with this ordering:

$$\begin{aligned} P(M, J, A, E, B) & \\ &= P(M)\, P(J|M)\, P(A|J, M)\, P(E|A, J, M)\, P(B|E, A, J, M) \\ &= P(M)\, P(J|M)\, P(A|J, M)\, P(E|A)\, P(B|A, E) \end{aligned}$$



This belief network has two extra edges.
This DAG does not show $P(B) = P(B|E)$.
This DAG does not show $P(M|A) = P(M|A, J)$.
This belief network has larger CPTs.

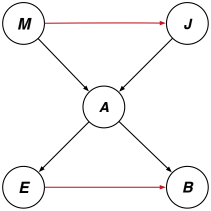
Node ordering

- What happens if we choose an unnatural ordering?

$$\text{Ex: } (M, J, A, E, B)$$

- Adding nodes with this ordering:

$$\begin{aligned} & P(M, J, A, E, B) \\ & \quad = P(M)\, P(J|M)\, P(A|J, M)\, P(E|A, J, M)\, P(B|E, A, J, M) \\ & \quad = P(M)\, P(J|M)\, P(A|J, M)\, P(E|A)\, P(B|A, E) \end{aligned}$$



This belief network has **two extra edges**.
This DAG does not show $P(B) = P(B|E)$.
This DAG does not show $P(M|A) = P(M|A, J)$.
This belief network has **larger CPTs**.
These CPTs may be more difficult to assess.

Advantages of belief networks

Advantages of belief networks

1. Compact representation of complex models

Advantages of belief networks

1. **Compact representation of complex models**

   BNs provide a complete but parsimonious representation of joint probability distributions.

Advantages of belief networks

1. **Compact representation of complex models**

   BNs provide a complete but parsimonious representation of joint probability distributions.

2. **Crisp separation of qualitative vs quantitative knowledge**

Advantages of belief networks

1. **Compact representation of complex models**

   BNs provide a complete but parsimonious representation of joint probability distributions.

2. **Crisp separation of qualitative vs quantitative knowledge**

Qualitative	DAGs encode assumptions of marginal and conditional independence.

Advantages of belief networks

1. **Compact representation of complex models**

   BNs provide a complete but parsimonious representation of joint probability distributions.

2. **Crisp separation of qualitative vs quantitative knowledge**

Qualitative	DAGs encode assumptions of marginal and conditional independence.
Quantitative	CPTs encode numerical influences of some variables on others.

That’s all folks!