Probabilistic graphical models

Lecture 3+4 of “Mathematics and AI”

Outline

1. Definition
2. Markov random fields
3. Bayesian networks
4. Inference via message passing

Factorizations of functions

$$p(x,y) = p(x)p(y) \qquad p(x,y) \neq p(x)p(y)$$

$$e^{x+y} = e^x e^y \qquad e^{(x+y)^2} = e^{x^2} e^{y^2} e^{2xy}$$

Definition

Definition

We want to describe complex relationships between an (ordered) set

$$S = \{X_1, X_2, \dots, X_n\}$$

of random variables $X_1, X_2, \dots X_n$. We can do so via the multivariate probability distribution

$$p(S = s) = p(X_1 = x_1, X_2 = x_2, \dots, X_n = x_n)$$

If S can be split into two independent subsets S_1 and S_2, $p(S = s)$ can be factorized:

$$p(S = s) = p(S_1 = s_1)p(S_2 = s_2)$$

Definition

We want to describe complex relationships between an (ordered) set

$$S = \{X_1, X_2, \dots, X_n\}$$

of random variables $X_1, X_2, \dots X_n$. We can do so via the multivariate probability distribution

$$p(s) = p(x_1, x_2, \dots, x_n)$$

If S can be split into two independent subsets S_1 and S_2, $p(s)$ can be factorized:

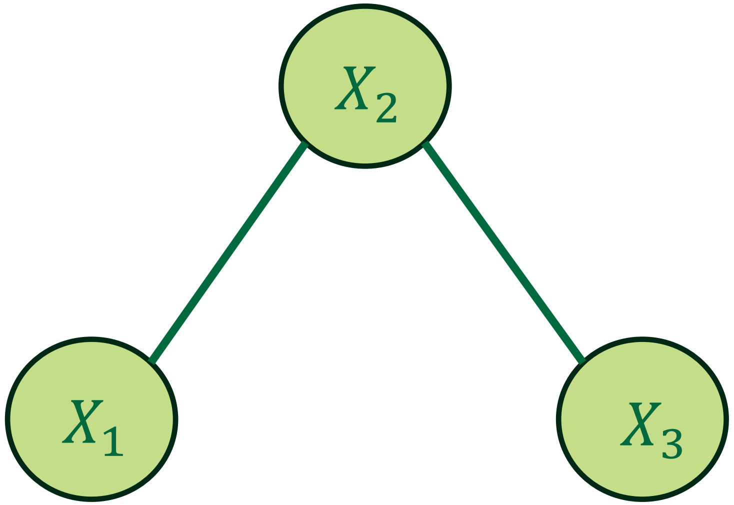
$$p(s) = p(s_1)p(s_2)$$

Definition

A probabilistic graphical model (PGM) includes:

- A multivariate probability distribution
- A graphical representation of its factorization properties

$$p(x_1, x_2, x_3)$$

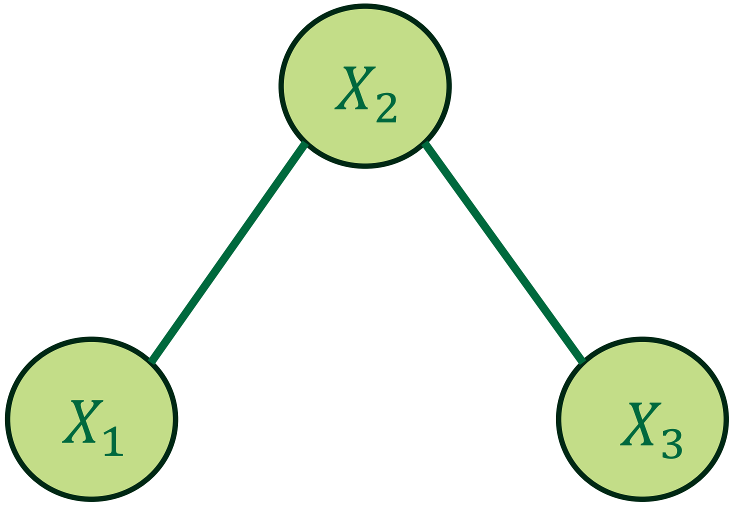


Different variants of PGMs represent factorization properties differently

Markov random fields

Markov random field

$$p(x_1, x_2, x_3) = \frac{1}{Z}\varphi(x_1, x_2)\,\varphi(x_2, x_3)$$

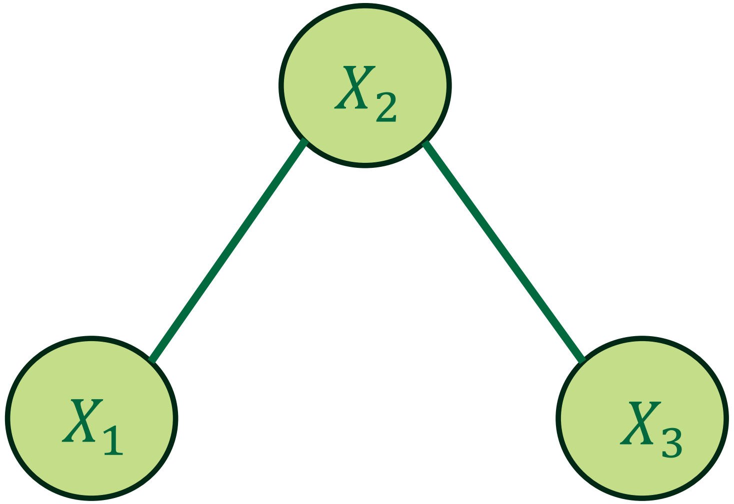


- Undirected graph of statistical dependencies
- Edge = statistical dependence
- Missing edge = independence
- Probability distribution $p(s) = \frac{1}{Z} \prod_{c_i \in \text{Cliques}} \varphi(c_i)$
- Extension: Factor graphs

Non-negative function called "potential" or "factor"

Normalization constant

MRF Example

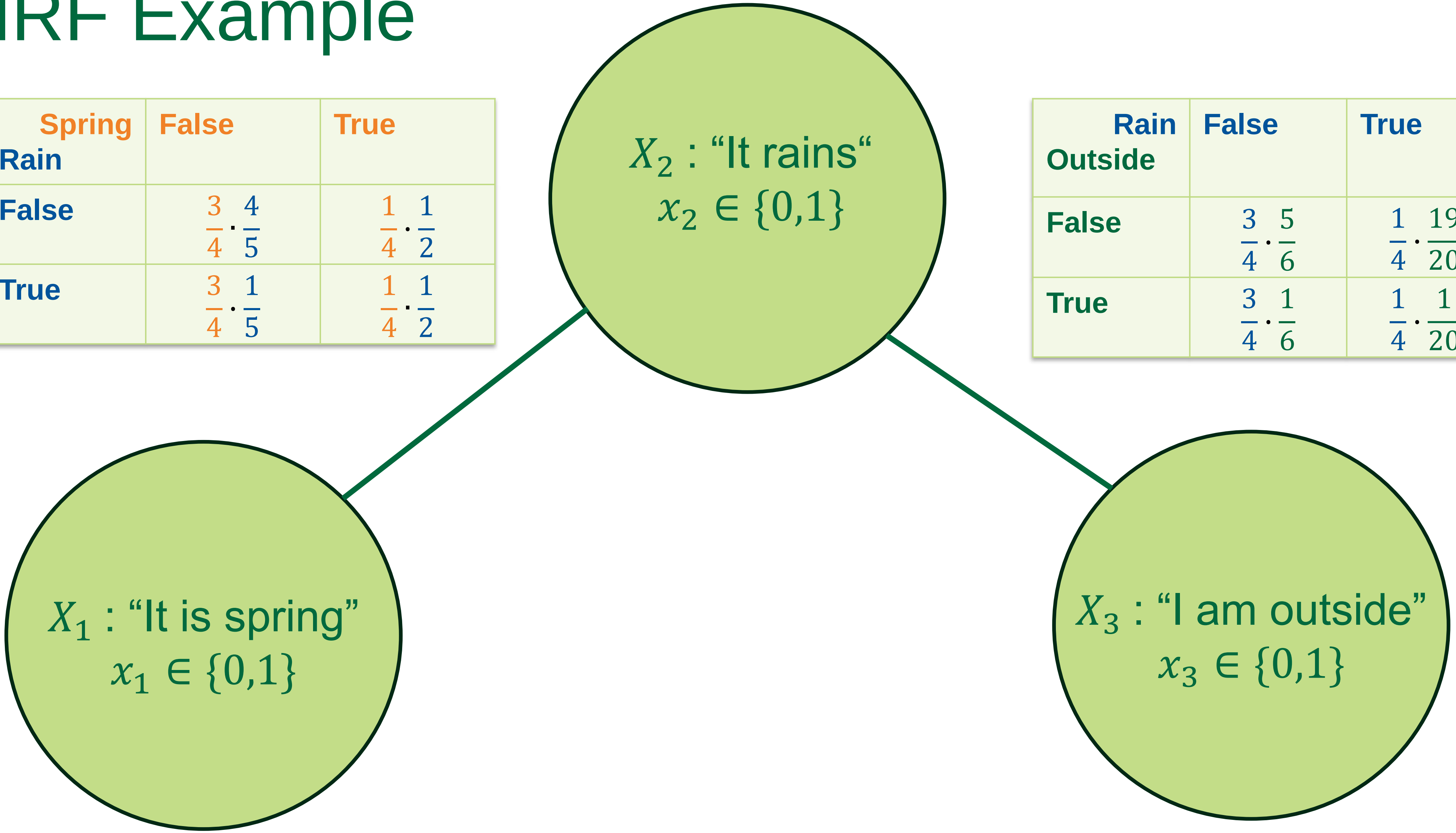


$$p(x_1, x_2, x_3) = \frac{1}{Z} \varphi(x_1, x_2)\, \varphi(x_2, x_3)$$

MRF Example

Spring / Rain	False	True
False	$\frac{3}{4}\cdot\frac{4}{5}$	$\frac{1}{4}\cdot\frac{1}{2}$
True	$\frac{3}{4}\cdot\frac{1}{5}$	$\frac{1}{4}\cdot\frac{1}{2}$

Rain / Outside	False	True
False	$\frac{3}{4}\cdot\frac{5}{6}$	$\frac{1}{4}\cdot\frac{19}{20}$
True	$\frac{3}{4}\cdot\frac{1}{6}$	$\frac{1}{4}\cdot\frac{1}{20}$



MRF Example

Spring / Rain	False	True
False	$\frac{3}{5}$	$\frac{1}{8}$
True	$\frac{3}{20}$	$\frac{1}{8}$

$\varphi(x_1, x_2)$

X_2 : “It rains“
$x_2 \in \{0,1\}$

Rain / Outside	False	True
False	$\frac{15}{24}$	$\frac{19}{80}$
True	$\frac{3}{24}$	$\frac{1}{80}$

$\varphi(x_2, x_3)$

X_1 : “It is spring”
$x_1 \in \{0,1\}$

X_3 : “I am outside”
$x_3 \in \{0,1\}$

Bayesian networks

Bayesian networks

- Directed graph of statistical dependencies
- Edge = statistical dependence
- No edge = independence
- Probability distribution: $$p(s) = \frac{1}{Z} \prod_{X_i} \varphi(x_i, X_i's \text{ parents})$$
- Extension: Directed factor graphs

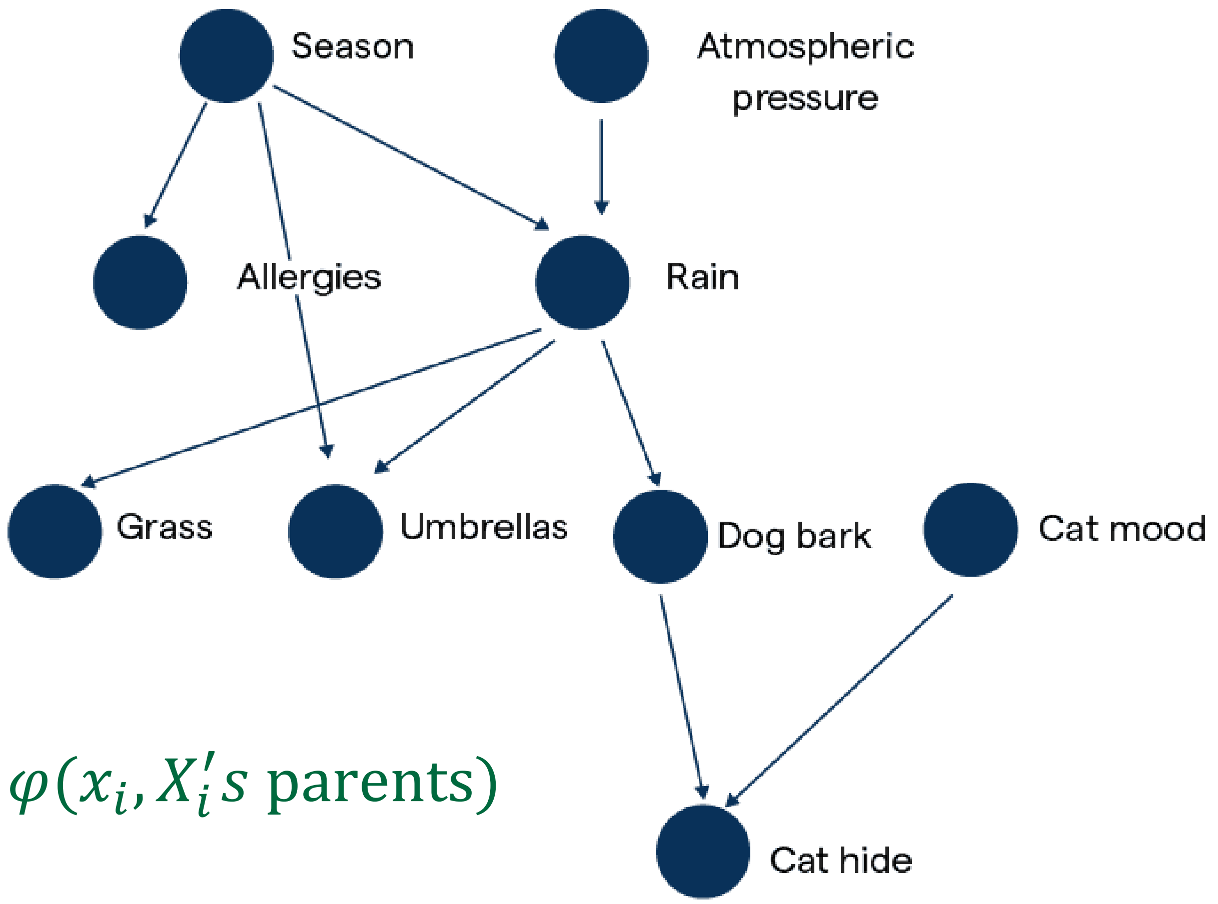
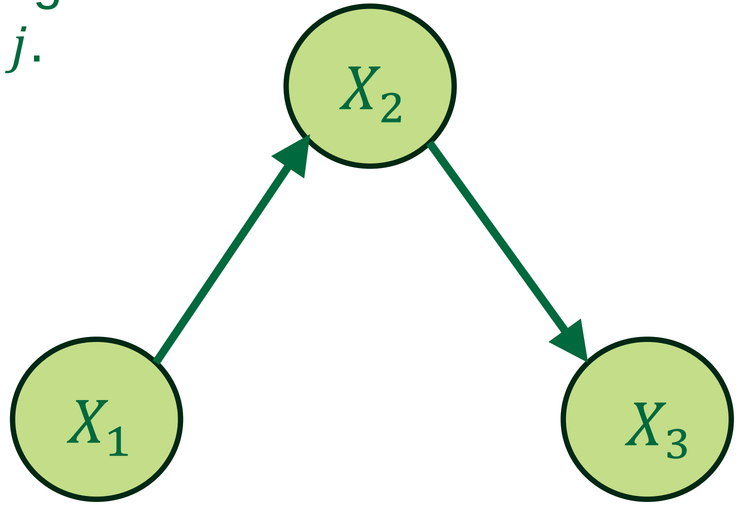


Image source: https://botpenguin.com/glossary/bayesian-networks

DARTMOUTH

BN Example

Assume a tropical node ordering:
If X_i is a parent of X_j, then $i < j$.



$$p(x_1, x_2, x_3) = \frac{1}{Z} \varphi(x_1) \varphi(x_1, x_2)\ \varphi(x_2, x_3)$$
$$= p(x_1) p(x_2 | x_1)\ p(x_3 | x_2)$$

BN Example

X_2 : "It rains"
$x_2 \in \{0,1\}$

$\varphi(x_3, x_2) = p(x_3|x_2)$

Spring / Rain	False	True
False	$\frac{4}{5}$	$\frac{1}{2}$
True	$\frac{1}{5}$	$\frac{1}{2}$

$\varphi(x_2, x_1) = p(x_2|x_1)$

Rain / Outside	False	True
False	$\frac{5}{6}$	$\frac{19}{20}$
True	$\frac{1}{6}$	$\frac{1}{20}$

X_1 : "It is spring"
$x_1 \in \{0,1\}$

Spring	False	True
	$\frac{3}{4}$	$\frac{1}{4}$

$\varphi(x_1) = p(x_1)$

X_3 : "I am outside"
$x_3 \in \{0,1\}$

BN: How does X_1 affect X_3?

- When it rains (i.e. $X_2 = 1$):
  - ➢Look at conditional probabilities $p(X_1, X_3 | X_2 = 1)$

S, R, O	p
F, F, F	
F, F, T	
F, T, F	
F, T, T	
T, F, F	
T, F, T	
T, T, F	
T, T, T	

Spring	False	True
	$\frac{3}{4}$	$\frac{1}{4}$

Spring / Rain	False	True
False	$\frac{4}{5}$	$\frac{1}{2}$
True	$\frac{1}{5}$	$\frac{1}{2}$

Rain / Outside	False	True
False	$\frac{5}{6}$	$\frac{19}{20}$
True	$\frac{1}{6}$	$\frac{1}{20}$

Outside	False	True
	$\frac{19}{20}$	$\frac{1}{20}$

BN: How does X_1 affect X_3?

- When it rains (i.e. $X_2 = 1$):
  - ➢Look at conditional probabilities $p(X_1, X_3 | X_2 = 1)$
- When it does not rain (i.e. $X_2 = 0$):
  - ➢Look at conditional probabilities $p(X_1, X_3 | X_2 = 1)$
- When we make no assumption about X_2
  - ➢Look at marginal probabilities $p(X_1, X_3)$

Spring	False	True
	$\frac{3}{4}$	$\frac{1}{4}$

Rain / Spring	False	True
False	$\frac{4}{5}$	$\frac{1}{5}$
True	$\frac{1}{2}$	$\frac{1}{2}$

Outside / Rain	False	True
False	$\frac{5}{6}$	$\frac{1}{6}$
True	$\frac{19}{20}$	$\frac{1}{20}$

Outside	False	True
	$\frac{5}{6}$	$\frac{1}{6}$

BN: How does X_1 affect X_3?

$$\varphi(x_1) = p(x_1)$$

Spring	False	True
	$\frac{3}{4}$	$\frac{1}{4}$

$$\varphi(x_2, x_1) = p(x_2|x_1)$$

Rain / Spring	False	True
False	$\frac{4}{5}$	$\frac{1}{5}$
True	$\frac{1}{2}$	$\frac{1}{2}$

Outside / Rain	False	True
False	$\frac{5}{6}$	$\frac{1}{6}$
True	$\frac{19}{20}$	$\frac{1}{20}$

$$\varphi(x_3, x_2) = p(x_3|x_2)$$

$$\boldsymbol{p(x_1, x_3) = \sum_{x_2} p(x_1, x_2, x_3)} = \sum_{x_2} p(x_1, x_3|x_2)p(x_2)$$

$$= \sum_{x_2} p(x_3|x_2)p(x_2)p(x_1)$$

$$= \sum_{x_2} p(x_3|x_2)\left[\sum_{x_1} p(x_2|x_1)p(x_1)\right]p(x_1)$$

- When we make no assumption about X_2
  - Look at marginal probabilities $p(X_1, X_3)$

BN: How does X_1 affect X_3 in general?

- X_3 is independent of X_1 when conditioning on X_2
- X_3 depends on X_1 when X_2 is marginalized

More generally:

- Conditioning on all parents of a variable X_i yields a probability distribution in which X_i is independent of all other variables
- Marginalizing a variable X_j yields a probability distribution in which all children of X_j depend on all parents of X_j

Back to MRF: How are X_1 and X_3 related?

- When it rains (i.e. $X_2 = 1$):
  - ➢Look at conditional probabilities $p(X_1, X_3 | X_2 = 1)$
- When it does not rain (i.e. $X_2 = 0$):
  - ➢Look at conditional probabilities $p(X_1, X_3 | X_2 = 1)$
- When we make no assumption about X_2
  - ➢Look at marginal probabilities $p(X_1, X_3)$

Back to MRF: How are X_1 and X_3 related?

- X_1 and X_3 are independent variables when conditioning on X_2
- X_1 and X_3 are dependent variables when X_2 is marginalized

More generally:

- Conditioning on $\overbrace{\text{all neighbors of } X_i}^{\text{“Markov blanket”}}$ yields a probability distribution in which X_i is independent of all other variables and vice versa
- Marginalizing X_j yields a probability distribution in which all neighbors of X_j depend on each other

Inference on probabilistic graph models

Queries for a probabilistic graph model

What is the probability distribution (or most likely value) of variable X_k given variable $X_i = x_i$?

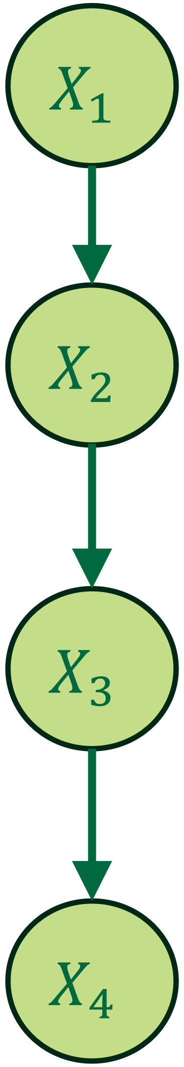
What is the probability distribution (or most likely value) of variable set S_2 given variable set $S_1 = s_1$?

Inference via message passing

- Message passing (a.k.a. belief propagation):
  - Algorithm for finding posterior distributions (and posterior most likely values)
  - Idea: Find solution by passing messages from parents to children

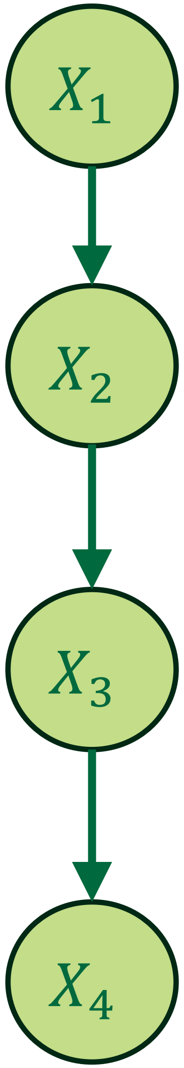
Example: Inference on a directed chain

$$p(x_4) = \sum_{x_1, x_2, x_3} p(x_1, x_2, x_3, x_4)$$



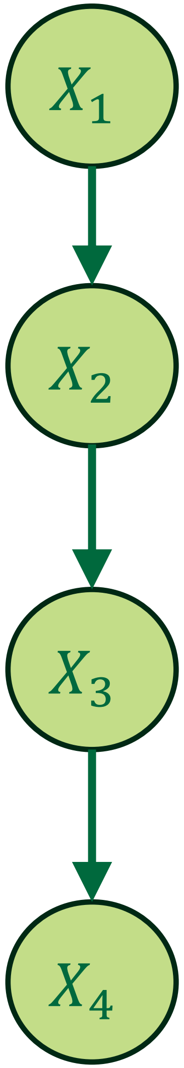
Example: Inference on a directed chain

$$
\begin{aligned}
p(x_4) &= \sum_{x_1,x_2,x_3} p(x_1, x_2, x_3, x_4) \\
&= \sum_{x_1,x_2,x_3} p(x_1)p(x_2|x_1)p(x_3|x_2)p(x_4|x_3)
\end{aligned}
$$



Example: Inference on a directed chain

$$
\begin{aligned}
p(x_4) &= \sum_{x_1,x_2,x_3} p(x_1, x_2, x_3, x_4) \\
&= \sum_{x_1,x_2,x_3} p(x_1)p(x_2|x_1)p(x_3|x_2)p(x_4|x_3) \\
&= \sum_{x_1,x_2,x_3} p(x_4|x_3)p(x_3|x_2)p(x_2|x_1)p(x_1)
\end{aligned}
$$

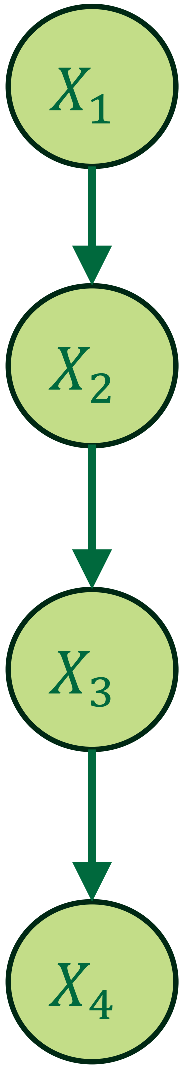


 DARTMOUTH

Example: Inference on a directed chain

$$
\begin{aligned}
p(x_4) &= \sum_{x_1,x_2,x_3} p(x_1,x_2,x_3,x_4) \\
&= \sum_{x_1,x_2,x_3} p(x_1)p(x_2|x_1)p(x_3|x_2)p(x_4|x_3) \\
&= \sum_{x_1,x_2,x_3} p(x_4|x_3)p(x_3|x_2)p(x_2|x_1)p(x_1) \\
&= \sum_{x_3} \underbrace{p(x_4|x_3) \underbrace{\sum_{x_2} p(x_3|x_2) \underbrace{\sum_{x_1} p(x_2|x_1)p(x_1)}_{m_{X_1\to X_2}(x_2)}}_{m_{X_2\to X_3}(x_3)}}_{m_{X_3\to X_4}(x_4)}
\end{aligned}
$$

Messages:



Conclusion on knowledge representation

- Different graphical models for representing
  - Certain structured knowledge
  - Uncertain structured knowledge
- Inferences on semantic networks
  - based on logic
- Inferences on probabilistic graph models
  - based on probability rules
  - Using message-passing algorithm for efficient inference