Graphical Models

Pradeep Ravikumar

Co-instructor: Aarti Singh

Machine Learning 10-701
Mar 20-22, 2017

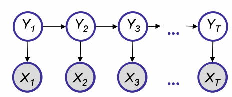
MACHINE LEARNING DEPARTMENT
Carnegie Mellon.
School of Computer Science

Applications of Graphical Models

- Computer vision and graphics
- Natural language processing
- Information retrieval
- Robotic control
- Computational biology
- Medical diagnosis
- Finance and economics
- Error-control codes
- Decision making under uncertainty
- ....

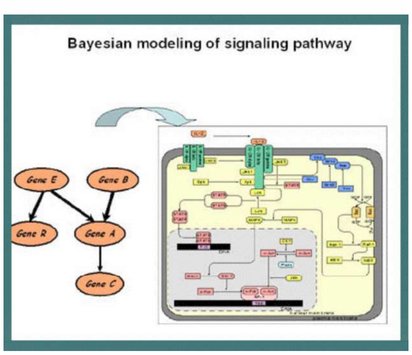
Speech Recognition

Speech Recognition



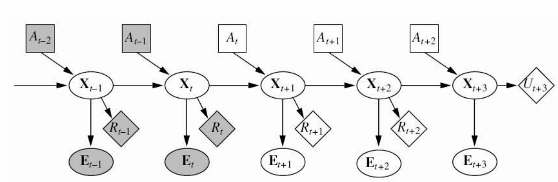
Hidden Markov Model

Bioinformatics

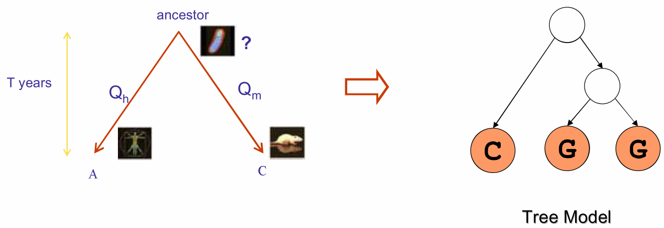


Reinforcement Learning

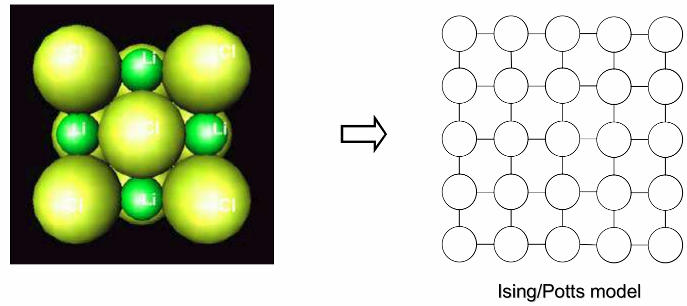
Partially Observed Markov Decision Processes (POMDP)



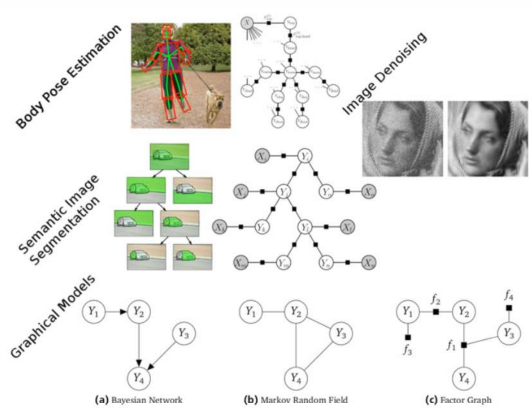
Evolution



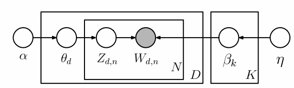
Solid State Physics



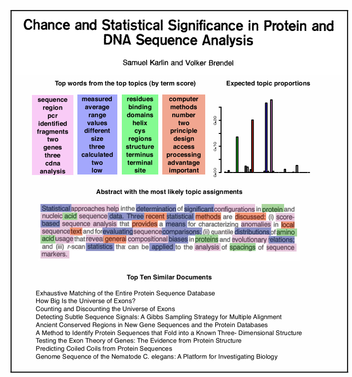
Computer Vision



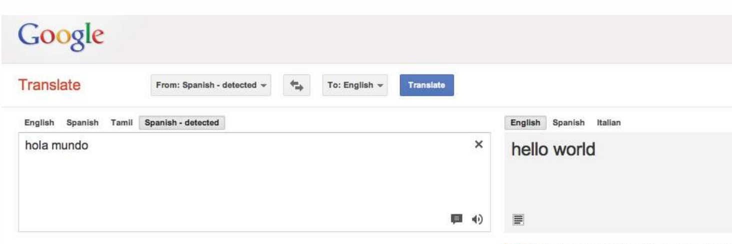
Language Processing

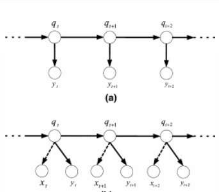


Topic Models



Machine Translation





Graphical Models: The Why

- Compact way of representing distributions among random variables

Graphical Models: The Why

- Compact way of representing distributions among random variables
- Visually appealing way (accessible to domain experts) of representing distributions among random variables

Graphical Models: The Why

- Compact way of representing distributions among random variables
- Visually appealing way (accessible to domain experts) of representing distributions among random variables
- They represent distributions among random variables — probability theory

Graphical Models: The Why

- Compact way of representing distributions among random variables
- Visually appealing way (accessible to domain experts) of representing distributions among random variables
- They represent distributions among random variables — probability theory
- They use graphs to do so — graph theory

Graphical Models: The Why

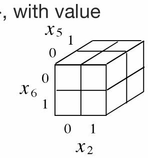
- Compact way of representing distributions among random variables
- Visually appealing way (accessible to domain experts) of representing distributions among random variables
- They represent distributions among random variables — probability theory
- They use graphs to do so — graph theory
  - A marriage of graph and probability theory

Joint Distributions

- Consider a set of random variables $\{X_1, X_2, \ldots, X_n\}$
- Let x_i be a realization of random variable X_i; x_i may be real, discrete, vector in vector space; for now assume variables are discrete.
- Probability Mass Function $p(x_1, \ldots, x_n) := P(X_1 = x_1, \ldots, X_n = x_n)$.
- Shorthand: $X = (X_1, \ldots, X_n)$, $x = (x_1, \ldots, x_n)$, so that $p(x) = P(X = x)$.

Joint Distributions and Tables

- Graphical Models are an answer to the following key question: how do we represent the joint distribution p(x)?
- One way is by using an n-dimensional table
  - a separate cell for each setting of variables {x_1, ... x_n}, with value p(x_1,...,x_n)
  - If each variable takes r values, we need to store r^n values : exponential in n
  - For large values of r,n ; such a representation is out!



Two classes of graphical models:
- Directed Graphical Models
- Undirected Graphical Models

Directed Graphical Models

Directed Graphical Models

Two Definitions

- A Factorization based definition

- Conditional Independence based definition

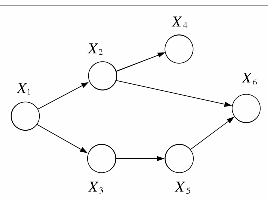
Directed Graphical Models

Two Definitions

- A Factorization based definition

- Conditional Independence based definition

Directed Graphs and Joint Probabilities



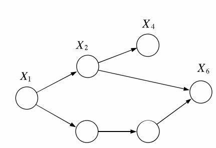
- Directed Graph is a pair $G = (V, E)$, where V is a set of nodes, E is a set of directed (also called oriented) edges. We will assume G is acyclic.

- Each node $i \in V$ is associated with a random variable X_i.

- Letting $V = \{1, 2, \ldots, n\}$, the set of random variables is $\{X_1, X_2, \ldots, X_n\}$.

- We will use node i and the associated random variable X_i interchangeably (though graph nodes and random variables are different formal objects!)

Directed Graphs and Joint Probabilities



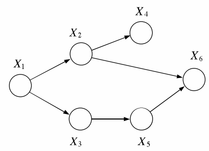
- Each node $i \in V$ has a set of parents π_i
  - $\pi_6 = \{X_2, X_5\}$.

- Let $V = \{1, 2, \ldots, n\}$. Given a set of functions $\{f_i(x_i, x_{\pi_i}) : i \in V\}$, we define a joint probability distribution:

$$p(x_1, \ldots, x_n) := \prod_{i=1}^{n} f_i(x_i, x_{\pi_i}).$$

Only assumptions:
* Non-negative
* Sum to one as a function of x_i

Directed Graphs and Joint Probabilities



- Each node $i \in V$ has a set of parents π_i
  - $\pi_6 = \{X_2, X_5\}$.
- Let $V = \{1, 2, \ldots, n\}$. Given a set of functions $\{f_i(x_i, x_{\pi_i}) : i \in V\}$, we define a joint probability distribution:

$$p(x_1, \ldots, x_n) := \prod_{i=1}^{n} f_i(x_i, x_{\pi_i}).$$

Only assumptions:
* Non-negative
* Sum to one as a function of x_i

- Is it a proper distribution?
  - Is it non-negative?
  - Does it sum to one? (Yes, provided each function $f_i(x_i, x_{\pi_i})$ sums to one as a function of x_i.

Directed Graphs

- Let $V = \{1, 2, \ldots, n\}$. Given a set of non-negative functions $\{f_i(x_i, x_{\pi_i}) : i \in V\}$, each of which sum to one as a function of first argument, we define a joint probability distribution:

$$p(x_1, \ldots, x_n) := \prod_{i=1}^{n} f_i(x_i, x_{\pi_i}).$$

- Exercise: Calculate $p(x_i | x_{\pi_i})$.

Directed Graphs

- Let $V = \{1, 2, \ldots, n\}$. Given a set of non-negative functions $\{f_i(x_i, x_{\pi_i}) : i \in V\}$, each of which sum to one as a function of first argument, we define a joint probability distribution:

$$p(x_1, \ldots, x_n) := \prod_{i=1}^{n} f_i(x_i, x_{\pi_i}).$$

- Exercise: Calculate $p(x_i | x_{\pi_i})$.
- Can be shown that $p(x_i | x_{\pi_i}) = f_i(x_i, x_{\pi_i})$.

Directed Graphs

- Let $V = \{1, 2, \ldots, n\}$. Given a set of non-negative functions $\{f_i(x_i, x_{\pi_i}) : i \in V\}$, each of which sum to one as a function of first argument, we define a joint probability distribution:

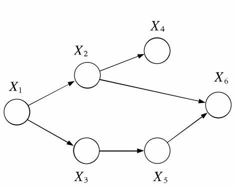
$$p(x_1, \ldots, x_n) := \prod_{i=1}^{n} f_i(x_i, x_{\pi_i}).$$

- Exercise: Calculate $p(x_i | x_{\pi_i})$.
- Can be shown that $p(x_i | x_{\pi_i}) = f_i(x_i, x_{\pi_i})$.

$$p(x_1, x_2, \ldots, x_n) = \prod_{i=1}^{n} p(x_i \,|\, x_{\pi_i}).$$

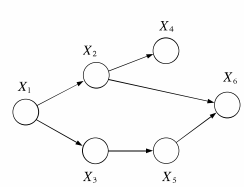
Directed Graphical Models: **set** of distributions that **factorize** according to a given directed acyclic graph G

Directed Graphical Models



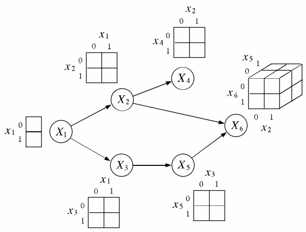
$$p(x_1, x_2, x_3, x_4, x_5, x_6) =$$

Directed Graphical Models



$$p(x_1, x_2, x_3, x_4, x_5, x_6) = p(x_1)p(x_2 \,|\, x_1)p(x_3 \,|\, x_1)p(x_4 \,|\, x_2)p(x_5 \,|\, x_3)p(x_6 \,|\, x_2, x_5)$$

Representational Economy



- Suppose each variable takes r values. Storage cost of a general joint distribution: r^n.
- Storage cost of directed graphical model: for variable i, $r^{|\pi_i|}$
- Exponential dependence scales only with "fan-in" of each node; typically small e.g. each node could have one parent, while number of nodes n could be in the millions.

Representational Economy

- Directed graphical model distributions thus can be represented or stored much more cheaply than a general distribution
  - What's the catch?
  - The graphical model distribution satisfy certain **conditional independence** assumptions

Review: Conditional Independence

X_A and X_B are *independent*, written $X_A \perp\!\!\!\perp X_B$, if:

$$p(x_A, x_B) = p(x_A)p(x_B),$$

X_A and X_C are *conditionally independent given* X_B, written $X_A \perp\!\!\!\perp X_C \,|\, X_B$, if:

$$p(x_A, x_C \,|\, x_B) = p(x_A \,|\, x_B)p(x_C \,|\, x_B),$$

Review: Conditional Independence

X_A and X_B are *independent*, written $X_A \perp\!\!\!\perp X_B$, if:

$$p(x_A, x_B) = p(x_A)p(x_B),$$

X_A and X_C are *conditionally independent given* X_B, written $X_A \perp\!\!\!\perp X_C \,|\, X_B$, if:

$$p(x_A, x_C \,|\, x_B) = p(x_A \,|\, x_B)p(x_C \,|\, x_B),$$

$$\Rightarrow \quad p(x_A \,|\, x_B, x_C) = p(x_A \,|\, x_B).$$

Review: Chain Rule

$$p(x_1, x_2, x_3, x_4, x_5, x_6) = p(x_1)p(x_2 \,|\, x_1)p(x_3 \,|\, x_1, x_2)p(x_4 \,|\, x_1, x_2, x_3)p(x_5 \,|\, x_1, x_2, x_3, x_4)p(x_6 \,|\, x_1, x_2, x_3, x_4, x_5)$$

Chain Rule

$$p(x_1, x_2, x_3, x_4, x_5, x_6) = p(x_1)p(x_2 \,|\, x_1)p(x_3 \,|\, x_1, x_2)p(x_4 \,|\, x_1, x_2, x_3)p(x_5 \,|\, x_1, x_2, x_3, x_4)p(x_6 \,|\, x_1, x_2, x_3, x_4, x_5)$$

In General:

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

Chain Rule

$p(x_1, x_2, x_3, x_4, x_5, x_6)$
$\quad = p(x_1)p(x_2 \,|\, x_1)p(x_3 \,|\, x_1, x_2)p(x_4 \,|\, x_1, x_2, x_3)p(x_5 \,|\, x_1, x_2, x_3, x_4)p(x_6 \,|\, x_1, x_2, x_3, x_4, x_5)$

In General:

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

Compare to directed graphical model distribution:

$$p(x_1, x_2, \ldots, x_n) = \prod_{i=1}^{n} p(x_i \,|\, x_{\pi_i}).$$

Chain Rule

$p(x_1, x_2, x_3, x_4, x_5, x_6)$
$\quad = p(x_1)p(x_2 \,|\, x_1)p(x_3 \,|\, x_1, x_2)p(x_4 \,|\, x_1, x_2, x_3)p(x_5 \,|\, x_1, x_2, x_3, x_4)p(x_6 \,|\, x_1, x_2, x_3, x_4, x_5)$

In General:

$$p(x_1, x_2, \ldots, x_n) = \prod_{i=1}^{n} p(x_i \,|\, x_1, \ldots, x_{i-1}).$$ Holds for all distributions

Compare to:

$$p(x_1, x_2, \ldots, x_n) = \prod_{i=1}^{n} p(x_i \,|\, x_{\pi_i}).$$ Holds for directed graphical model distributions

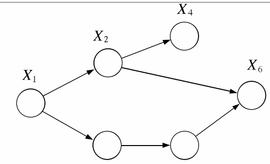
Directed Graphical Models

Two Definitions

- A Factorization based definition

- Conditional Independence based definition

Topological Ordering



Topological Ordering I:
for each node $i \in V$, the nodes in π_i appear before i in the ordering.
e.g. I = (1,2,3,4,5,6)

Let ν_i denote set of nodes that appear earlier than i in the ordering I, *excluding* parent nodes π_i.
e.g. $\nu_5 = \{1, 2, 4\}$.

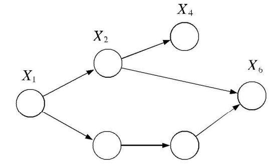
Conditional Independence Assertions

- Given a graph G, and a topological ordering I, we associate to the graph the following set of basic conditional independence statements:

$$\{X_i \perp\!\!\!\perp X_{\nu_i} \,|\, X_{\pi_i}\} \qquad \text{for } i \in \mathcal{V}.$$

Example

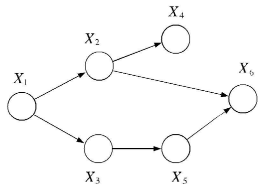
$$\{X_i \perp\!\!\!\perp X_{\nu_i} \,|\, X_{\pi_i}\} \qquad \text{for } i \in \mathcal{V}.$$



Example

$\{X_i \perp\!\!\!\perp X_{\nu_i} \,|\, X_{\pi_i}\}$ for $i \in \mathcal{V}$.

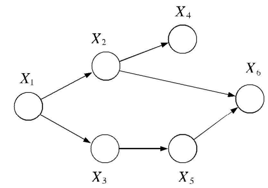
$$X_1 \perp\!\!\!\perp \emptyset \quad | \quad \emptyset$$



Example

$\{X_i \perp\!\!\!\perp X_{\nu_i} \,|\, X_{\pi_i}\}$ for $i \in \mathcal{V}$.

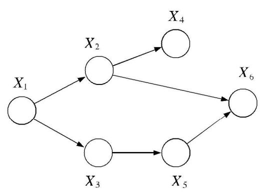
$$\begin{array}{rcl}
X_1 \perp\!\!\!\perp \emptyset & | & \emptyset \\
X_2 \perp\!\!\!\perp \emptyset & | & X_1
\end{array}$$



Example

$\{X_i \perp\!\!\!\perp X_{\nu_i} \,|\, X_{\pi_i}\}$ for $i \in \mathcal{V}$.

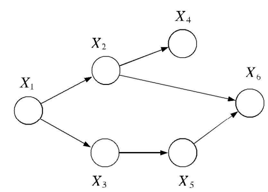
$$\begin{array}{rcl}
X_1 \perp\!\!\!\perp \emptyset & | & \emptyset \\
X_2 \perp\!\!\!\perp \emptyset & | & X_1 \\
X_3 \perp\!\!\!\perp X_2 & | & X_1
\end{array}$$



Example

$\{X_i \perp\!\!\!\perp X_{\nu_i} \,|\, X_{\pi_i}\}$ for $i \in \mathcal{V}$.

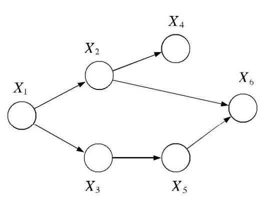
$$\begin{array}{rcl}
X_1 \perp\!\!\!\perp \emptyset & | & \emptyset \\
X_2 \perp\!\!\!\perp \emptyset & | & X_1 \\
X_3 \perp\!\!\!\perp X_2 & | & X_1 \\
X_4 \perp\!\!\!\perp \{X_1, X_3\} & | & X_2
\end{array}$$



Example

$\{X_i \perp\!\!\!\perp X_{\nu_i} \,|\, X_{\pi_i}\}$ for $i \in \mathcal{V}$.

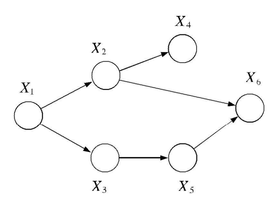
$$\begin{array}{rcl}
X_1 \perp\!\!\!\perp \emptyset & | & \emptyset \\
X_2 \perp\!\!\!\perp \emptyset & | & X_1 \\
X_3 \perp\!\!\!\perp X_2 & | & X_1 \\
X_4 \perp\!\!\!\perp \{X_1, X_3\} & | & X_2 \\
X_5 \perp\!\!\!\perp \{X_1, X_2, X_4\} & | & X_3
\end{array}$$



Example

$\{X_i \perp\!\!\!\perp X_{\nu_i} \,|\, X_{\pi_i}\}$ for $i \in \mathcal{V}$.

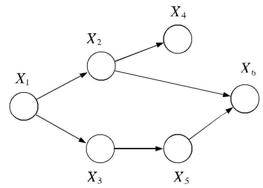
$$\begin{array}{rcl}
X_1 \perp\!\!\!\perp \emptyset & | & \emptyset \\
X_2 \perp\!\!\!\perp \emptyset & | & X_1 \\
X_3 \perp\!\!\!\perp X_2 & | & X_1 \\
X_4 \perp\!\!\!\perp \{X_1, X_3\} & | & X_2 \\
X_5 \perp\!\!\!\perp \{X_1, X_2, X_4\} & | & X_3 \\
X_6 \perp\!\!\!\perp \{X_1, X_3, X_4\} & | & \{X_2, X_5\}
\end{array}$$



Example

$$\{X_i \perp\!\!\!\perp X_{\nu_i} \,|\, X_{\pi_i}\} \qquad \text{for } i \in \mathcal{V}.$$

$$\begin{aligned}
X_1 \perp\!\!\!\perp \emptyset \quad &| \quad \emptyset \\
X_2 \perp\!\!\!\perp \emptyset \quad &| \quad X_1 \\
X_3 \perp\!\!\!\perp X_2 \quad &| \quad X_1 \\
X_4 \perp\!\!\!\perp \{X_1, X_3\} \quad &| \quad X_2 \\
X_5 \perp\!\!\!\perp \{X_1, X_2, X_4\} \quad &| \quad X_3 \\
X_6 \perp\!\!\!\perp \{X_1, X_3, X_4\} \quad &| \quad \{X_2, X_5\}
\end{aligned}$$



Question: do these “asserted” conditional independence statements hold given the joint directed graphical model distribution?

Question: These were “asserted” conditional independence statements; do they actually hold given the joint directed graphical model distribution?

$$p(x_1, x_2, x_3, x_4, x_5, x_6) = p(x_1)p(x_2 \,|\, x_1)p(x_3 \,|\, x_1)p(x_4 \,|\, x_2)p(x_5 \,|\, x_3)p(x_6 \,|\, x_2, x_5)$$

$$\rightarrow X_4 \text{ is independent of } X_1 \text{ and } X_3 \text{ given } X_2 \text{ ?}$$

Question: do these “asserted” conditional independence statements hold given the joint directed graphical model distribution?

$$p(x_1, x_2, x_3, x_4, x_5, x_6) = p(x_1)p(x_2 \,|\, x_1)p(x_3 \,|\, x_1)p(x_4 \,|\, x_2)p(x_5 \,|\, x_3)p(x_6 \,|\, x_2, x_5)$$

$$\rightarrow X_4 \text{ is independent of } X_1 \text{ and } X_3 \text{ given } X_2 \text{ ?}$$

$$\begin{aligned}
p(x_1, x_2, x_3, x_4) &= \sum_{x_5}\sum_{x_6} p(x_1, x_2, x_3, x_4, x_5, x_6) \\
&= \sum_{x_5}\sum_{x_6} p(x_1)p(x_2 \,|\, x_1)p(x_3 \,|\, x_1)p(x_4 \,|\, x_2)p(x_5 \,|\, x_3)p(x_6 \,|\, x_2, x_5) \\
&= p(x_1)p(x_2 \,|\, x_1)p(x_3 \,|\, x_1)p(x_4 \,|\, x_2)\sum_{x_5} p(x_5 \,|\, x_3)\sum_{x_6} p(x_6 \,|\, x_2, x_5) \\
&= p(x_1)p(x_2 \,|\, x_1)p(x_3 \,|\, x_1)p(x_4 \,|\, x_2),
\end{aligned}$$

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

$$p(x_1, x_2, x_3, x_4, x_5, x_6) = p(x_1)p(x_2 \,|\, x_1)p(x_3 \,|\, x_1)p(x_4 \,|\, x_2)p(x_5 \,|\, x_3)p(x_6 \,|\, x_2, x_5)$$

$$\rightarrow X_4 \text{ is independent of } X_1 \text{ and } X_3 \text{ given } X_2 \text{ ?}$$

$$\begin{aligned}
p(x_1, x_2, x_3, x_4) &= \sum_{x_5}\sum_{x_6} p(x_1, x_2, x_3, x_4, x_5, x_6) \\
&= \sum_{x_5}\sum_{x_6} p(x_1)p(x_2 \,|\, x_1)p(x_3 \,|\, x_1)p(x_4 \,|\, x_2)p(x_5 \,|\, x_3)p(x_6 \,|\, x_2, x_5) \\
&= p(x_1)p(x_2 \,|\, x_1)p(x_3 \,|\, x_1)p(x_4 \,|\, x_2)\sum_{x_5} p(x_5 \,|\, x_3)\sum_{x_6} p(x_6 \,|\, x_2, x_5) \\
&= p(x_1)p(x_2 \,|\, x_1)p(x_3 \,|\, x_1)p(x_4 \,|\, x_2),
\end{aligned}$$

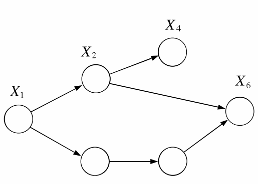
$$\begin{aligned}
p(x_1, x_2, x_3) &= \sum_{x_4} p(x_1)p(x_2 \,|\, x_1)p(x_3 \,|\, x_1)p(x_4 \,|\, x_2) \\
&= p(x_1)p(x_2 \,|\, x_1)p(x_3 \,|\, x_1).
\end{aligned}$$

$$p(x_4 \,|\, x_1, x_2, x_3) = p(x_4 \,|\, x_2) \quad \Rightarrow \quad X_4 \perp\!\!\!\perp \{X_1, X_3\} \,|\, X_2.$$

Question: do these “asserted” conditional independence statements hold given the joint directed graphical model distribution?

Yes.

Question: are there other conditional independence statements that hold given the joint directed graphical model distribution?

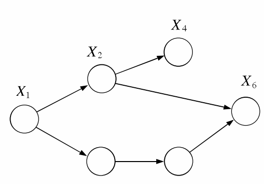


e.g. $X_1 \perp X_6 \,|\, \{X_2, X_3\}$, but is only *implied* by the asserted independence statements earlier

Question: do these “asserted” conditional independence statements hold given the joint directed graphical model distribution?

Yes.

Question: are there other conditional independence statements that hold given the joint directed graphical model distribution?



e.g. $X_1 \perp X_6 \,|\, \{X_2, X_3\}$, but is only *implied* by the asserted independence statements earlier

We can list all the conditional independence statements implied by a graphical model distribution, using “graph separation”

Graphical Models and Conditional Independences

- A Graphical Model is a family of probability distributions. Given a graph, specify any set of conditional distributions of nodes given parents.

Graphical Models and Conditional Independences

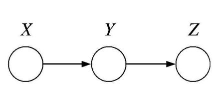
- A Graphical Model is a family of probability distributions. Given a graph, specify any set of conditional distributions of nodes given parents.
- Any such distribution satisfies conditional independences

Graphical Models and Conditional Independences

- A Graphical Model is a family of probability distributions. Given a graph, specify any set of conditional distributions of nodes given parents.
- Any such distribution satisfies conditional independences
  - Some of which we listed earlier; we will now provide an algorithm (Bayes-Ball, d-Separation) to list **all** conditional independence statements satisfied by the given family of graphical model distributions

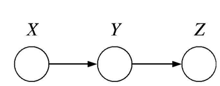
Three Canonical Graphs

Graph I



$\Rightarrow$ $X \perp\!\!\!\perp Z \mid Y.$

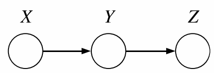
Graph I



$\Rightarrow$ $X \perp\!\!\!\perp Z \mid Y.$

Future (Z) is independent of Past (X) given present (Y)

Graph I

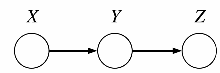


from the graphical model factored form:
$p(x, y, z) = p(x)p(y \,|\, x)p(z \,|\, y),$

which implies:

$$p(z \,|\, x, y) \;=$$

Graph I

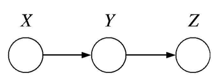


from the graphical model factored form:
$p(x, y, z) = p(x)p(y \,|\, x)p(z \,|\, y),$

which implies:

$$\begin{aligned} p(z \,|\, x, y) &= \frac{p(x,y,z)}{p(x,y)} \\ &= \frac{p(x)p(y \,|\, x)p(z \,|\, y)}{p(x)p(y \,|\, x)} \\ &= p(z \,|\, y), \end{aligned}$$

Graph I

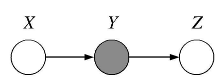


from the graphical model factored form:
$p(x, y, z) = p(x)p(y \,|\, x)p(z \,|\, y),$

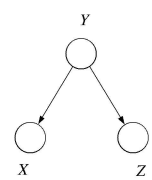
which implies:

$$\begin{aligned} p(z \,|\, x, y) &= \frac{p(x,y,z)}{p(x,y)} \\ &= \frac{p(x)p(y \,|\, x)p(z \,|\, y)}{p(x)p(y \,|\, x)} \\ &= p(z \,|\, y), \end{aligned}$$

$$X \perp\!\!\!\perp Z \,|\, Y.$$

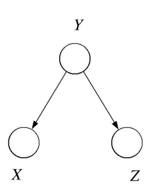


Graph II



$$X \perp\!\!\!\perp Z \,|\, Y.$$

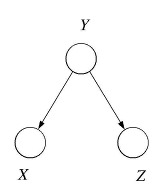
Graph II



$$X \perp\!\!\!\perp Z \,|\, Y.$$

- If X = “shoe-size”, and Z = “gray hair”, then in general these are

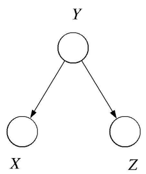
Graph II



$$X \perp\!\!\!\perp Z \,|\, Y.$$

- If X = “shoe-size”, and Z = “gray hair”, then in general these are strongly dependent (children have small shoe-sizes, and no gray hair!)

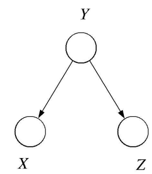
Graph II



$\Rightarrow$ $X \perp\!\!\!\perp Z \,|\, Y.$

- If X = “shoe-size”, and Z = “gray hair”, then in general these are strongly dependent (children have small shoe-sizes, and no gray hair!)
- But if Y = “age”, then given Y, the variables X and Z are

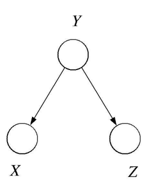
Graph II



$\Rightarrow$ $X \perp\!\!\!\perp Z \,|\, Y.$

- If X = “shoe-size”, and Z = “gray hair”, then in general these are strongly dependent (children have small shoe-sizes, and no gray hair!)
- But if Y = “age”, then given Y, the variables X and Z are no longer dependent.
  - Y “explains” all of the observed dependence between X and Z.

Graph II



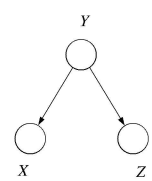
$$p(x, y, z) = p(y)p(x \,|\, y)p(z \,|\, y)$$

which implies:

$$p(x, z \,|\, y) \quad =$$

$\Rightarrow$ $X \perp\!\!\!\perp Z \,|\, Y.$

Graph II



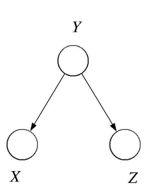
$$p(x, y, z) = p(y)p(x \,|\, y)p(z \,|\, y)$$

which implies:

$$\begin{aligned} p(x, z \,|\, y) &= \frac{p(y)p(x \,|\, y)p(z \,|\, y)}{p(y)} \\ &= p(x \,|\, y)p(z \,|\, y), \end{aligned}$$

$\Rightarrow$ $X \perp\!\!\!\perp Z \,|\, Y.$

Graph II

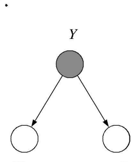


$$p(x, y, z) = p(y)p(x \,|\, y)p(z \,|\, y)$$

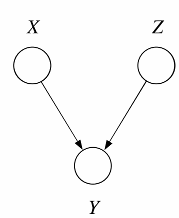
which implies:

$$\begin{aligned} p(x, z \,|\, y) &= \frac{p(y)p(x \,|\, y)p(z \,|\, y)}{p(y)} \\ &= p(x \,|\, y)p(z \,|\, y), \end{aligned}$$

$\Rightarrow$ $X \perp\!\!\!\perp Z \,|\, Y.$

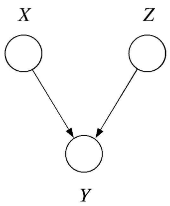


Graph III



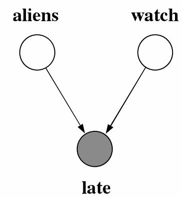
$\Rightarrow$

Graph III

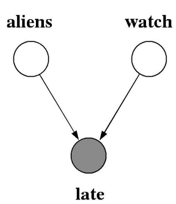


$X \perp\!\!\!\perp Z,$

Graph III



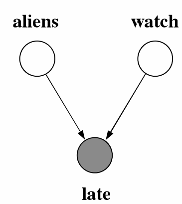
Graph III



$$P(\textbf{aliens} = \text{"yes"} \,|\, \textbf{late} = \text{"yes"}) \neq P(\textbf{aliens} = \text{"yes"} \,|\, \textbf{late} = \text{"yes"}, \textbf{watch} = \text{"no"})$$

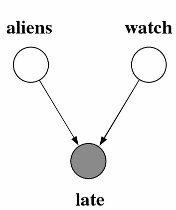
$\textbf{aliens} \not\perp \textbf{watch} \mid \textbf{late}$

Graph III



$\text{aliens} \perp \text{watch}$

Graph III

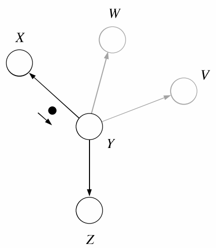


$\text{aliens} \perp \text{watch}$

“Late” can explain away independence between “aliens” and “watch”!

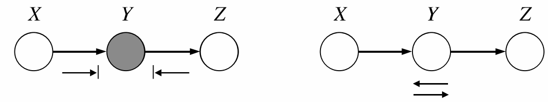
Is there a graph-theoretic way to write down all conditional independence statements given graph?

d-Separation Algorithm

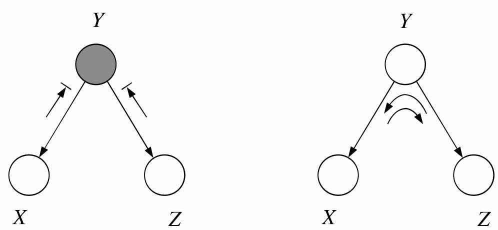


- Is X_A independent of X_B given X_C?
- Drop a ball from any variable in X_A. If it reaches any variable in X_C without being “blocked” by variables in X_B, then answer is no, otherwise yes.

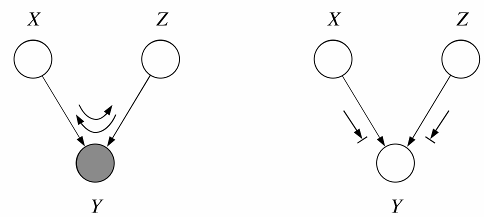
“Blocking” Rules I



“Blocking” Rules II

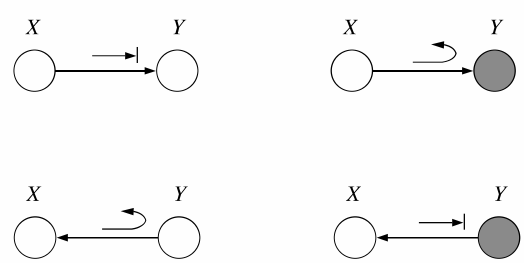


“Blocking” Rules III

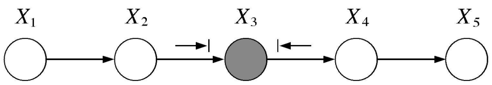


Blocking Rules IV

Blocking Rules IV

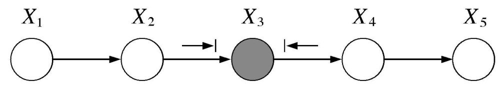


Example I



$X_{i+1} \perp\!\!\!\perp \{X_1, X_2, \ldots, X_{i-1}\} \mid X_i$. **?**

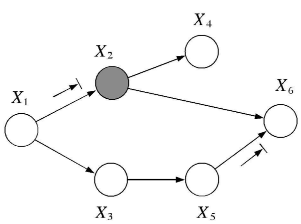
Example I



$X_{i+1} \perp\!\!\!\perp \{X_1, X_2, \ldots, X_{i-1}\} \mid X_i$.

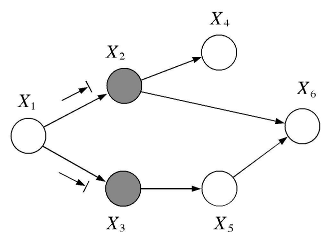
$X_1 \perp\!\!\!\perp X_5 \mid X_4$, $\qquad X_1 \perp\!\!\!\perp X_5 \mid X_2$, $\qquad X_1 \perp\!\!\!\perp X_5 \mid \{X_2, X_4\}$,

Example II



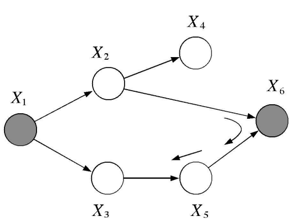
$X_4 \perp\!\!\!\perp \{X_1, X_3\} \mid X_2$ **?**

Example II(b)



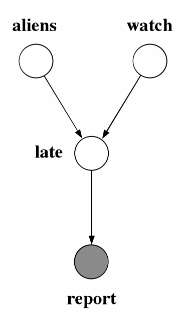
$X_1 \perp\!\!\!\perp X_6 \mid \{X_2, X_3\}$ **?**

Example II(c)

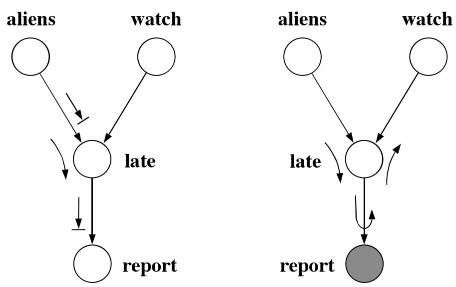


$X_2 \perp\!\!\!\perp X_3 \mid \{X_1, X_6\}$ **?**

Example III



Example III



Dual Characterization of Graphical Models

Characterization I:

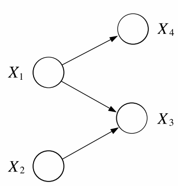
given a graph G, all distributions such that:

$$p(x_1, x_2, \ldots, x_n) \triangleq \prod_{i=1}^{n} p(x_i \,|\, x_{\pi_i}).$$

Dual Characterization of Graphical Models

Characterization II:

given a graph G, all distributions that satisfy all conditional independence statements specified by d-Separation algorithm



$$X_1 \perp\!\!\!\perp X_2$$
$$X_2 \perp\!\!\!\perp X_4$$
$$X_2 \perp\!\!\!\perp X_4 \mid X_1$$
$$X_3 \perp\!\!\!\perp X_4 \mid X_1$$
$$X_2 \perp\!\!\!\perp X_4 \mid \{X_1, X_3\}$$
$$\{X_2, X_3\} \perp\!\!\!\perp X_4 \mid X_1$$

Dual Characterization of Graphical Models

Characterization I:

given a graph G, all distributions such that:

$$p(x_1, x_2, \ldots, x_n) \triangleq \prod_{i=1}^{n} p(x_i \,|\, x_{\pi_i}).$$

Characterization II:

given a graph G, all distributions that satisfy all conditional independence statements specified by Bayes Ball algorithm

Both these are **equivalent!**

Dual Characterization of Graphical Models

Characterization I:

given a graph G, all distributions such that:

$$p(x_1, x_2, \ldots, x_n) \triangleq \prod_{i=1}^{n} p(x_i \,|\, x_{\pi_i}).$$

Characterization II:

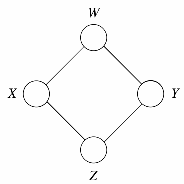
given a graph G, all distributions that satisfy all conditional independence statements specified by d-Separation algorithm

Both these are **equivalent!**

Marriage of Probability Theory (Char. I) and Graph Theory (Char. II)

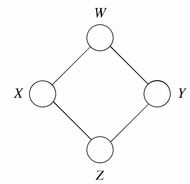
Undirected Graphical Models

Undirected Graphical Models



- Suppose we prefer undirected graphs
- Can we use these to define (sets of) probability distributions?
  - Yes
- "Undirected Graphical Models" or Markov Random Fields

Undirected Graphs

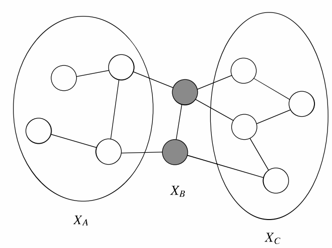


- Undirected Graph is a pair $G = (V, E)$, where V is a set of nodes, E is a set of undirected edges. We do *not* assume G is acyclic.
- As before, each node $i \in V$ is associated with a random variable X_i.
- Letting $V = \{1, 2, \ldots, n\}$, the set of random variables is $\{X_1, X_2, \ldots, X_n\}$.
- We will use node i and the associated random variable X_i interchangeably (though graph nodes and random variables are different formal objects!)

Undirected Graphical Models

- In the directed graphical model case, we started with a **factorized form** before discussing **conditional independence assumptions** because the former was more natural (conditional probabilities vs "Bayes-Ball" algorithm)
- In the undirected graphical model case, the conditional independence assertions are much more natural and intuitive

Undirected Graphs and Cond. Independences



The set X_B separates X_A from X_C (all paths from X_A to X_C pass through X_B) iff $X_A \perp\!\!\!\perp X_C \mid X_B$

Undirected Graphical Models

- Given an undirected graph, we can define the set of conditional independence assertions using graph separation as above

Undirected Graphical Models

- Given an undirected graph, we can define the set of conditional independence assertions using graph separation as above
  - the set of distributions that satisfy all these assertions is the undirected graphical model family given the graph

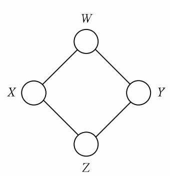
Undirected Graphical Models

- Given an undirected graph, we can define the set of conditional independence assertions using graph separation as above
  - the set of distributions that satisfy all these assertions is the undirected graphical model family given the graph
- Is it possible to reduce directed graphical models to undirected graphical models or vice versa?

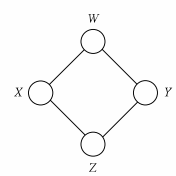
Undirected Graphical Models

- Given an undirected graph, we can define the set of conditional independence assertions using graph separation as above
  - the set of distributions that satisfy all these assertions is the undirected graphical model family given the graph
- Is it possible to reduce directed graphical models to undirected graphical models or vice versa?
  - No

Undirected and Directed Graphical Models



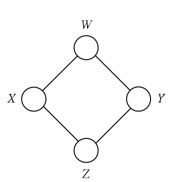
Undirected and Directed Graphical Models



$X \perp\!\!\!\perp Y \,|\, \{W, Z\}$ and $W \perp\!\!\!\perp Z \,|\, \{X, Y\}$

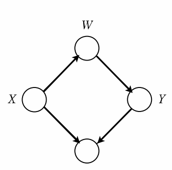
Can we represent the cond. independence assertions using a directed graph?

Undirected and Directed Graphical Models

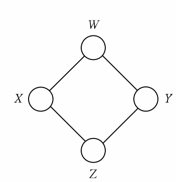


$X \perp\!\!\!\perp Y \,|\, \{W, Z\}$ and $W \perp\!\!\!\perp Z \,|\, \{X, Y\}$

Can we represent the cond. independence assertions using a directed graph?

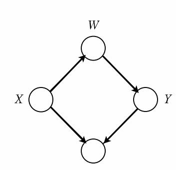


Undirected and Directed Graphical Models



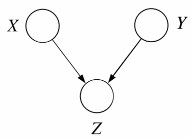
$X \perp\!\!\!\perp Y \,|\, \{W, Z\}$ and $W \perp\!\!\!\perp Z \,|\, \{X, Y\}$

Can we represent the cond. independence assertions using a directed graph?



we have $X \perp\!\!\!\perp Y \,|\, W$, and we do not have $X \perp\!\!\!\perp Y \,|\, \{W, Z\}$

Undirected and Directed Graphical Models



No undirected graph can exactly represent the set of cond. independence assertions given this directed graph

Undirected Graphical Models: Parameterization

Undirected Graphical Models: Parameterization

- Given an undirected graph, we can define the set of conditional independence assertions using graph separation as above

Undirected Graphical Models: Parameterization

- Given an undirected graph, we can define the set of conditional independence assertions using graph separation as above
  - the set of distributions that satisfy all these assertions is the undirected graphical model family given the graph

Undirected Graphical Models: Parameterization

- Given an undirected graph, we can define the set of conditional independence assertions using graph separation as above
  - the set of distributions that satisfy all these assertions is the undirected graphical model family given the graph
- We would like to obtain a "local" parameterization of an undirected graphical model

Undirected Graphical Models: Parameterization

- Given an undirected graph, we can define the set of conditional independence assertions using graph separation as above
  - the set of distributions that satisfy all these assertions is the undirected graphical model family given the graph
- We would like to obtain a "local" parameterization of an undirected graphical model
  - Why? For compact representation of joint distribution!

Undirected Graphical Models: Parameterization

- Given an undirected graph, we can define the set of conditional independence assertions using graph separation as above
  - the set of distributions that satisfy all these assertions is the undirected graphical model family given the graph
- We would like to obtain a “local” parameterization of an undirected graphical model
  - Why? For compact representation of joint distribution!
- For directed graphs, the parameterization was based on local conditional probabilities of nodes given their parents

Undirected Graphical Models: Parameterization

- Given an undirected graph, we can define the set of conditional independence assertions using graph separation as above
  - the set of distributions that satisfy all these assertions is the undirected graphical model family given the graph
- We would like to obtain a “local” parameterization of an undirected graphical model
  - Why? For compact representation of joint distribution!
- For directed graphs, the parameterization was based on local conditional probabilities of nodes given their parents

  - “local” had the interpretation of a set consisting of a node and its parents

Undirected Graphical Models: Parameterization

- We would like to obtain a “local” parameterization of an undirected graphical model
- Conditional probabilities of nodes given neighbors?
  - Hard to ensure consistency of such conditional probabilities across nodes, so won’t be able to choose such functions independently for each node

Undirected Graphical Models: Parameterization

- We would like to obtain a “local” parameterization of an undirected graphical model
- Conditional probabilities of nodes given neighbors?
  - Hard to ensure consistency of such conditional probabilities across nodes, so won’t be able to choose such functions independently for each node

  - Their product need not be a valid joint distribution

Undirected Graphical Models: Parameterization

- How do we use conditional probabilities in the undirected graphical model case (where there are no parents)?
- Conditional probabilities of nodes given neighbors?
  - Hard to ensure consistency of such conditional probabilities across nodes, so won’t be able to choose such functions independently for each node
  - Their product need not be a valid joint distribution
  - Better to not use conditional probabilities, but to use product of some other “local” functions that can be chosen independently

Conditional Independence and Factorization

- If $X_A \perp\!\!\!\perp X_C \mid X_B$, then

$$P(X_A|X_B, X_C) = P(X_A|X_C)$$
$$P(X_A, X_B, X_C) = P(X_A|X_C)\, P(X_B, X_C)$$

- The joint distribution $P(X_A, X_B, X_C) = f(X_A, X_B)\, g(X_B, X_C)$, for some functions f and g.
- Conditional Independence entails factorization!

Conditional Independence and Factorization

- If $X_A \perp\!\!\!\perp X_C \mid X_B$, then

$$P(X_A|X_B, X_C) = P(X_A|X_C)$$
$$P(X_A, X_B, X_C) = P(X_A|X_C)\, P(X_B, X_C)$$

- The joint distribution $P(X_A, X_B, X_C) = f(X_A, X_B)\, g(X_B, X_C)$, for some functions f and g.
- Conditional Independence entails factorization!

- If there is no edge between X_i and X_j, then the joint distribution cannot have a factor such as $\psi(X_i, X_j, X_k)$.
- Factors of the joint distribution can depend only on variables within a fully connected subgraph, also called a clique.
- Without loss of generality, assume factors only over *maximum* cliques

Conditional Independence and Factorization

- If $X_A \perp\!\!\!\perp X_C \mid X_B$, then

$$P(X_A|X_B, X_C) = P(X_A|X_C)$$
$$P(X_A, X_B, X_C) = P(X_A|X_C)\, P(X_B, X_C)$$

- The joint distribution $P(X_A, X_B, X_C) = f(X_A, X_B)\, g(X_B, X_C)$, for some functions f and g.
- Conditional Independence entails factorization!

- If there is no edge between X_i and X_j, then the joint distribution cannot have a factor such as $\psi(X_i, X_j, X_k)$.
- Factors of the joint distribution can depend only on variables within a fully connected subgraph, also called a clique.
- Without loss of generality, assume factors only over *maximum* cliques

Conditional Independence and Factorization

- If $X_A \perp\!\!\!\perp X_C \mid X_B$, then

$$P(X_A|X_B, X_C) = P(X_A|X_C)$$
$$P(X_A, X_B, X_C) = P(X_A|X_C)\, P(X_B, X_C)$$

- The joint distribution $P(X_A, X_B, X_C) = f(X_A, X_B)\, g(X_B, X_C)$, for some functions f and g.
- Conditional Independence entails factorization!

- If there is no edge between X_i and X_j, then the joint distribution cannot have a factor such as $\psi(X_i, X_j, X_k)$.
- Factors of the joint distribution can depend only on variables within a fully connected subgraph, also called a clique.
- Without loss of generality, assume factors only over *maximum* cliques

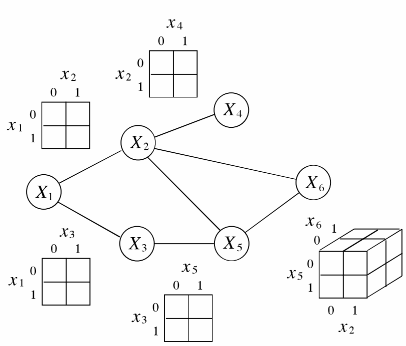
Undirected Graphical Models: Factorization

- Meaning of "local" for undirected graphs should be "maximal clique".
- Let C be indices of a maximal clique in undirected graph G, let $\mathcal{C}$ be set of all such maximal cliques C.
- A "potential function" $\psi_C(x_C)$ is a function that only depends on values x_c of maximal clique variables X_C
- Potential functions assumed to be non-negative, real-valued, but otherwise arbitrary

$$p(x) \triangleq \frac{1}{Z} \prod_{C \in \mathcal{C}} \psi_{X_C}(x_C),$$

Normalization Constant: $Z \triangleq \sum_x \prod_{C \in \mathcal{C}} \psi_{X_C}(x_C)$

Undirected Graphical Models: Factorization

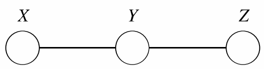


What do potential functions mean?

- We just discussed how local conditional probabilities are not satisfactory as potential functions
- Why not replace potential functions $\psi_C(x_C)$ with marginal probabilities $p(x_C)$?

What do potential functions mean?

- We just discussed how local conditional probabilities are not satisfactory as potential functions
- Why not replace potential functions $\psi_C(x_C)$ with marginal probabilities $p(x_C)$?



$$p(x, y, z) = p(y)p(x \mid y)p(z \mid y).$$

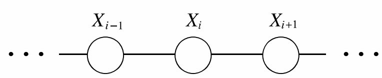
$$p(x, y, z) \neq p(x, y)p(y, z).$$

What do potential functions mean?

- Potential functions are neither conditional nor marginal probabilities, and do not have a local probabilistic interpretation
- But do often have a “pre-probabilistic” interpretation as “agreement,” “constraint” or “energy”
- Potential function favors certain configurations by assigning larger value

What do potential functions mean?

- Potential function favors certain configurations by assigning larger value



$$\begin{array}{cc|c|c|} & & \multicolumn{2}{c}{x_i} \\ & & -1 & 1 \\ \hline x_{i-1} & -1 & 1.5 & 0.2 \\ \hline & 1 & 0.2 & 1.5 \\ \hline \end{array} \qquad \begin{array}{cc|c|c|} & & \multicolumn{2}{c}{x_{i+1}} \\ & & -1 & 1 \\ \hline x_{i} & -1 & 1.5 & 0.2 \\ \hline & 1 & 0.2 & 1.5 \\ \hline \end{array}$$

Dual Characterization of Undirected Graphical Models

Characterization I:

Given a graph G, all distributions expressed as product of potential functions over maximal cliques

Characterization II:

Given a graph G, all distributions that satisfy all conditional independence statements specified by graph separation

Hammersley Clifford Theorem: Characterizations are equivalent!

As in the directed case, a profound link between probability theory and graph theory

Topics in Graphical Models

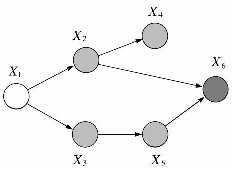
- Representation
  - Which joint probability distributions does a graphical model represent?
- Inference
  - How to answer questions about the joint probability distribution?
    - Marginal distribution of a node variable
    - Most likely assignment of node variables
- Learning
  - How to learn the parameters and structure of a graphical model?

Probabilistic Inference

- Let E and F be disjoint subsets of nodes of a graphical model; with X_E and X_F the disjoint subsets of variables corresponding to E, F
- We want to calculate $P(x_F \mid x_E)$ for arbitrary subsets E and F
  - This is the general **probabilistic inference** problem
  - We will focus initially on a single “query node” F, and directed graphical models

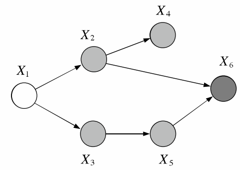
Probabilistic Inference

- We will focus initially on a single “query node” F, and directed graphical models



- Darkly Shaded: $E = \{6\}$
- Unshaded: $F = \{1\}$
- Lightly Shaded: $R = \{2, 3, 4, 5\}$.
- Need: $P(X_F|X_E)$

Probabilistic Inference



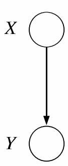
- Darkly Shaded: $E = \{6\}$
- Unshaded: $F = \{1\}$
- Lightly Shaded: $R = \{2, 3, 4, 5\}$.
- Need: $P(X_F|X_E)$

$$p(x_E, x_F) = \sum_{x_R} p(x_E, x_F, x_R),$$ Marginalize out variables in R

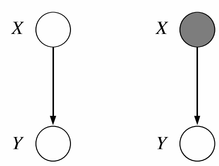
$$p(x_E) = \sum_{x_F} p(x_E, x_F),$$ Marginalize out variables in F

$$p(x_F \,|\, x_E) = \frac{p(x_E, x_F)}{p(x_E)}.$$

Simplest Case

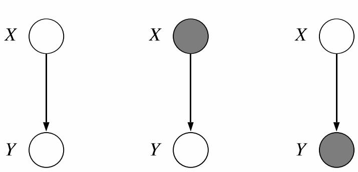


Simplest Case



- P(Y|X) is specified directly by the conditional probability table

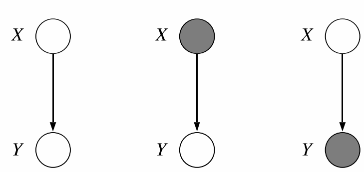
Simplest Case



$$p(x \,|\, y) = \frac{p(y \,|\, x)p(x)}{p(y)}.$$

$$p(y) = \sum_x p(y \,|\, x)p(x).$$

Simplest Case

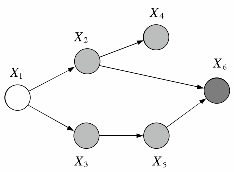


$$p(x \,|\, y) = \frac{p(y \,|\, x)p(x)}{p(y)}.$$

$$p(y) = \sum_x p(y \,|\, x)p(x).$$

(Just) Need extensions to general graphs!

Probabilistic Inference



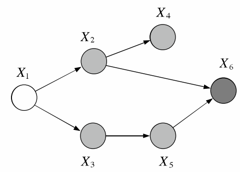
$$p(x_E, x_F) = \sum_{x_R} p(x_E, x_F, x_R),$$

$$p(x_E) = \sum_{x_F} p(x_E, x_F),$$

$$p(x_F \mid x_E) = \frac{p(x_E, x_F)}{p(x_E)}.$$

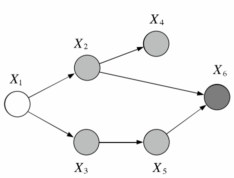
- The summation $\sum_{x_R}$ expands over many variables, one for each indexed by R.
- If each variable takes k values, we have $k^{|R|}$ terms in summation!
- With $|R|$ in the hundreds, naive summation is infeasible

Probabilistic Inference



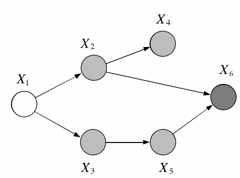
- Consider $p(x_1, \ldots, x_6)$ as a k^6 size probability table

Probabilistic Inference



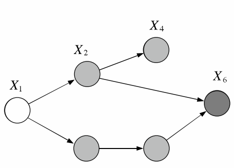
- Consider $p(x_1, \ldots, x_6)$ as a k^6 size probability table
- Summing over even one variable x_6 requires performing the sum for each value of the variables $\{x_1, x_2, \ldots, x_5\}$.

Probabilistic Inference



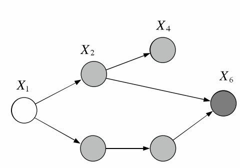
- Consider $p(x_1, \ldots, x_6)$ as a k^6 size probability table
- Summing over even one variable x_6 requires performing the sum for each value of the variables $\{x_1, x_2, \ldots, x_5\}$.
- We must thus perform $O(k^6)$ operations to do a single sum (essentially, we must touch each entry in the table)
- To reduce comp. complexity, we must consider the factorized form of the probability distribution

Variable Elimination



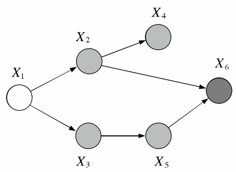
$$p(x_1, x_2, \ldots, x_5) = \sum_{x_6} p(x_1)p(x_2 \mid x_1)p(x_3 \mid x_1)p(x_4 \mid x_2)p(x_5 \mid x_3)p(x_6 \mid x_2, x_5)$$

Variable Elimination



$$\begin{aligned} p(x_1, x_2, \ldots, x_5) &= \sum_{x_6} p(x_1)p(x_2 \mid x_1)p(x_3 \mid x_1)p(x_4 \mid x_2)p(x_5 \mid x_3)p(x_6 \mid x_2, x_5) \\ &= p(x_1)p(x_2 \mid x_1)p(x_3 \mid x_1)p(x_4 \mid x_2)p(x_5 \mid x_3) \sum_{x_6} p(x_6 \mid x_2, x_5). \end{aligned}$$

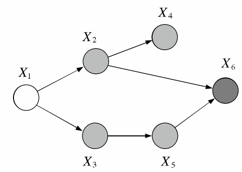
Variable Elimination



$$\begin{aligned} p(x_1, x_2, \ldots, x_5) &= \sum_{x_6} p(x_1)p(x_2 \,|\, x_1)p(x_3 \,|\, x_1)p(x_4 \,|\, x_2)p(x_5 \,|\, x_3)p(x_6 \,|\, x_2, x_5) \\ &= p(x_1)p(x_2 \,|\, x_1)p(x_3 \,|\, x_1)p(x_4 \,|\, x_2)p(x_5 \,|\, x_3) \sum_{x_6} p(x_6 \,|\, x_2, x_5). \end{aligned}$$

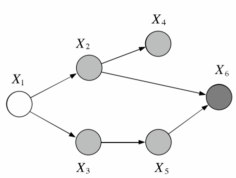
Reduced the count from $O(k^6)$ to $O(k^3)$ (actually we know the sum here is equal to one, but assume we didn't know that)

Variable Elimination



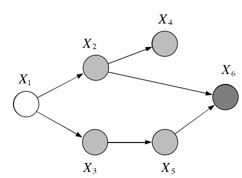
$$p(x_1, \bar{x}_6) = \sum_{x_2}\sum_{x_3}\sum_{x_4}\sum_{x_5} p(x_1)p(x_2 \,|\, x_1)p(x_3 \,|\, x_1)p(x_4 \,|\, x_2)p(x_5 \,|\, x_3)p(\bar{x}_6 \,|\, x_2, x_5)$$

Variable Elimination



$$\begin{aligned} p(x_1, \bar{x}_6) &= \sum_{x_2}\sum_{x_3}\sum_{x_4}\sum_{x_5} p(x_1)p(x_2 \,|\, x_1)p(x_3 \,|\, x_1)p(x_4 \,|\, x_2)p(x_5 \,|\, x_3)p(\bar{x}_6 \,|\, x_2, x_5) \\ &= p(x_1) \sum_{x_2} p(x_2 \,|\, x_1) \sum_{x_3} p(x_3 \,|\, x_1) \sum_{x_4} p(x_4 \,|\, x_2) \sum_{x_5} p(x_5 \,|\, x_3)p(\bar{x}_6 \,|\, x_2, x_5) \end{aligned}$$

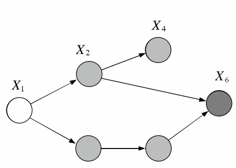
Variable Elimination



$$\begin{aligned} p(x_1, \bar{x}_6) &= \sum_{x_2}\sum_{x_3}\sum_{x_4}\sum_{x_5} p(x_1)p(x_2 \,|\, x_1)p(x_3 \,|\, x_1)p(x_4 \,|\, x_2)p(x_5 \,|\, x_3)p(\bar{x}_6 \,|\, x_2, x_5) \\ &= p(x_1) \sum_{x_2} p(x_2 \,|\, x_1) \sum_{x_3} p(x_3 \,|\, x_1) \sum_{x_4} p(x_4 \,|\, x_2) \sum_{x_5} p(x_5 \,|\, x_3)p(\bar{x}_6 \,|\, x_2, x_5) \\ &= p(x_1) \sum_{x_2} p(x_2 \,|\, x_1) \sum_{x_3} p(x_3 \,|\, x_1) \sum_{x_4} p(x_4 \,|\, x_2) m_5(x_2, x_3) \end{aligned}$$

where we define $m_5(x_2, x_3) \triangleq \sum_{x_5} p(x_5 \,|\, x_3)p(\bar{x}_6 \,|\, x_2, x_5)$.

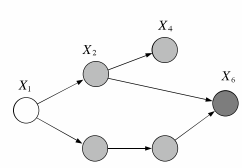
Variable Elimination



$$\begin{aligned} p(x_1, \bar{x}_6) &= p(x_1) \sum_{x_2} p(x_2 \,|\, x_1) \sum_{x_3} p(x_3 \,|\, x_1) m_5(x_2, x_3) \sum_{x_4} p(x_4 \,|\, x_2) \\ &= p(x_1) \sum_{x_2} p(x_2 \,|\, x_1) m_4(x_2) \sum_{x_3} p(x_3 \,|\, x_1) m_5(x_2, x_3). \end{aligned}$$

$m_4(x_2) \triangleq \sum_{x_4} p(x_4 \,|\, x_2)$

Variable Elimination

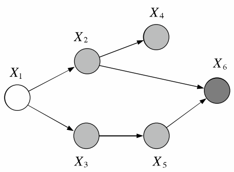


$$\begin{aligned} p(x_1, \bar{x}_6) &= p(x_1) \sum_{x_2} p(x_2 \,|\, x_1) \sum_{x_3} p(x_3 \,|\, x_1) m_5(x_2, x_3) \sum_{x_4} p(x_4 \,|\, x_2) \\ &= p(x_1) \sum_{x_2} p(x_2 \,|\, x_1) m_4(x_2) \sum_{x_3} p(x_3 \,|\, x_1) m_5(x_2, x_3). \end{aligned}$$

$m_4(x_2) \triangleq \sum_{x_4} p(x_4 \,|\, x_2)$

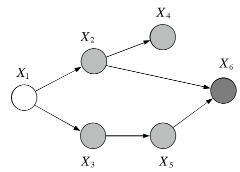
We denote by $m_i(S_i)$ the expression after computing $\sum_{x_i}$ with S_i the index of variables, other than i that appear in the summand

Variable Elimination



$$
\begin{aligned}
p(x_1, \bar{x}_6) &= p(x_1) \sum_{x_2} p(x_2 \,|\, x_1) m_4(x_2) m_3(x_1, x_2) \\
&= p(x_1) m_2(x_1).
\end{aligned}
$$

Variable Elimination

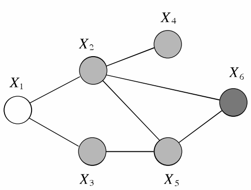


$$
\begin{aligned}
p(x_1, \bar{x}_6) &= p(x_1) \sum_{x_2} p(x_2 \,|\, x_1) m_4(x_2) m_3(x_1, x_2) \\
&= p(x_1) m_2(x_1).
\end{aligned}
$$

$$p(\bar{x}_6) = \sum_{x_1} p(x_1) m_2(x_1).$$

$$p(x_1 \,|\, \bar{x}_6) = \frac{p(x_1) m_2(x_1)}{\sum_{x_1} p(x_1) m_2(x_1)}.$$

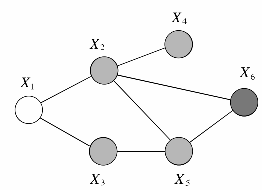
Variable Elimination; undirected graphs



- Potentials $\{\psi_C(x_C)\}$ on the cliques $\{X_1, X_2\}$, $\{X_1, X_3\}$, $\{X_2, X_4\}$, $\{X_3, X_5\}$, and $\{X_2, X_5, X_6\}$.

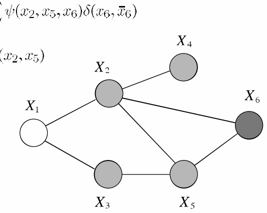
Elimination; undirected graphs

$$
\begin{aligned}
p(x_1, \bar{x}_6) &= \frac{1}{Z} \sum_{x_2} \sum_{x_3} \sum_{x_4} \sum_{x_5} \sum_{x_6} \psi(x_1, x_2) \psi(x_1, x_3) \psi(x_2, x_4) \psi(x_3, x_5) \psi(x_2, x_5, x_6) \delta(x_6, \bar{x}_6) \\
&= \frac{1}{Z} \sum_{x_2} \psi(x_1, x_2) \sum_{x_3} \psi(x_1, x_3) \sum_{x_4} \psi(x_2, x_4) \sum_{x_5} \psi(x_3, x_5) \sum_{x_6} \psi(x_2, x_5, x_6) \delta(x_6, \bar{x}_6)
\end{aligned}
$$



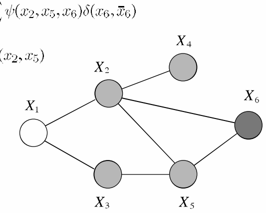
Elimination; undirected graphs

$$
\begin{aligned}
p(x_1, \bar{x}_6) &= \frac{1}{Z} \sum_{x_2} \sum_{x_3} \sum_{x_4} \sum_{x_5} \sum_{x_6} \psi(x_1, x_2) \psi(x_1, x_3) \psi(x_2, x_4) \psi(x_3, x_5) \psi(x_2, x_5, x_6) \delta(x_6, \bar{x}_6) \\
&= \frac{1}{Z} \sum_{x_2} \psi(x_1, x_2) \sum_{x_3} \psi(x_1, x_3) \sum_{x_4} \psi(x_2, x_4) \sum_{x_5} \psi(x_3, x_5) \sum_{x_6} \psi(x_2, x_5, x_6) \delta(x_6, \bar{x}_6) \\
&= \frac{1}{Z} \sum_{x_2} \psi(x_1, x_2) \sum_{x_3} \psi(x_1, x_3) \sum_{x_4} \psi(x_2, x_4) \sum_{x_5} \psi(x_3, x_5) m_6(x_2, x_5)
\end{aligned}
$$



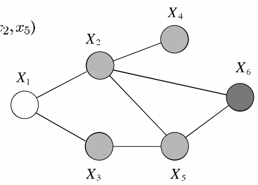
Elimination; undirected graphs

$$
\begin{aligned}
p(x_1, \bar{x}_6) &= \frac{1}{Z} \sum_{x_2} \sum_{x_3} \sum_{x_4} \sum_{x_5} \sum_{x_6} \psi(x_1, x_2) \psi(x_1, x_3) \psi(x_2, x_4) \psi(x_3, x_5) \psi(x_2, x_5, x_6) \delta(x_6, \bar{x}_6) \\
&= \frac{1}{Z} \sum_{x_2} \psi(x_1, x_2) \sum_{x_3} \psi(x_1, x_3) \sum_{x_4} \psi(x_2, x_4) \sum_{x_5} \psi(x_3, x_5) \sum_{x_6} \psi(x_2, x_5, x_6) \delta(x_6, \bar{x}_6) \\
&= \frac{1}{Z} \sum_{x_2} \psi(x_1, x_2) \sum_{x_3} \psi(x_1, x_3) \sum_{x_4} \psi(x_2, x_4) \sum_{x_5} \psi(x_3, x_5) m_6(x_2, x_5) \\
&= \frac{1}{Z} \sum_{x_2} \psi(x_1, x_2) \sum_{x_3} \psi(x_1, x_3) m_5(x_2, x_3) \sum_{x_4} \psi(x_2, x_4)
\end{aligned}
$$



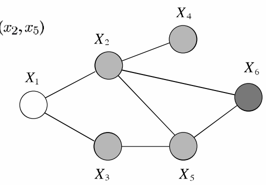
Elimination; undirected graphs

$$
\begin{aligned}
p(x_1, \bar{x}_6) &= \frac{1}{Z}\sum_{x_2}\sum_{x_3}\sum_{x_4}\sum_{x_5}\sum_{x_6}\psi(x_1,x_2)\psi(x_1,x_3)\psi(x_2,x_4)\psi(x_3,x_5)\psi(x_2,x_5,x_6)\delta(x_6,\bar{x}_6) \\
&= \frac{1}{Z}\sum_{x_2}\psi(x_1,x_2)\sum_{x_3}\psi(x_1,x_3)\sum_{x_4}\psi(x_2,x_4)\sum_{x_5}\psi(x_3,x_5)\sum_{x_6}\psi(x_2,x_5,x_6)\delta(x_6,\bar{x}_6) \\
&= \frac{1}{Z}\sum_{x_2}\psi(x_1,x_2)\sum_{x_3}\psi(x_1,x_3)\sum_{x_4}\psi(x_2,x_4)\sum_{x_5}\psi(x_3,x_5)m_6(x_2,x_5) \\
&= \frac{1}{Z}\sum_{x_2}\psi(x_1,x_2)\sum_{x_3}\psi(x_1,x_3)m_5(x_2,x_3)\sum_{x_4}\psi(x_2,x_4) \\
&= \frac{1}{Z}\sum_{x_2}\psi(x_1,x_2)m_4(x_2)\sum_{x_3}\psi(x_1,x_3)m_5(x_2,x_3)
\end{aligned}
$$



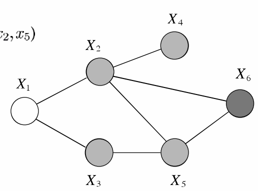
Elimination; undirected graphs

$$
\begin{aligned}
p(x_1, \bar{x}_6) &= \frac{1}{Z}\sum_{x_2}\sum_{x_3}\sum_{x_4}\sum_{x_5}\sum_{x_6}\psi(x_1,x_2)\psi(x_1,x_3)\psi(x_2,x_4)\psi(x_3,x_5)\psi(x_2,x_5,x_6)\delta(x_6,\bar{x}_6) \\
&= \frac{1}{Z}\sum_{x_2}\psi(x_1,x_2)\sum_{x_3}\psi(x_1,x_3)\sum_{x_4}\psi(x_2,x_4)\sum_{x_5}\psi(x_3,x_5)\sum_{x_6}\psi(x_2,x_5,x_6)\delta(x_6,\bar{x}_6) \\
&= \frac{1}{Z}\sum_{x_2}\psi(x_1,x_2)\sum_{x_3}\psi(x_1,x_3)\sum_{x_4}\psi(x_2,x_4)\sum_{x_5}\psi(x_3,x_5)m_6(x_2,x_5) \\
&= \frac{1}{Z}\sum_{x_2}\psi(x_1,x_2)\sum_{x_3}\psi(x_1,x_3)m_5(x_2,x_3)\sum_{x_4}\psi(x_2,x_4) \\
&= \frac{1}{Z}\sum_{x_2}\psi(x_1,x_2)m_4(x_2)\sum_{x_3}\psi(x_1,x_3)m_5(x_2,x_3) \\
&= \frac{1}{Z}\sum_{x_2}\psi(x_1,x_2)m_4(x_2)m_3(x_1,x_2)
\end{aligned}
$$



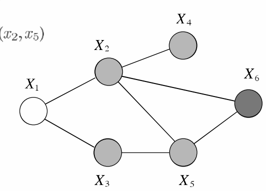
Elimination; undirected graphs

$$
\begin{aligned}
p(x_1, \bar{x}_6) &= \frac{1}{Z}\sum_{x_2}\sum_{x_3}\sum_{x_4}\sum_{x_5}\sum_{x_6}\psi(x_1,x_2)\psi(x_1,x_3)\psi(x_2,x_4)\psi(x_3,x_5)\psi(x_2,x_5,x_6)\delta(x_6,\bar{x}_6) \\
&= \frac{1}{Z}\sum_{x_2}\psi(x_1,x_2)\sum_{x_3}\psi(x_1,x_3)\sum_{x_4}\psi(x_2,x_4)\sum_{x_5}\psi(x_3,x_5)\sum_{x_6}\psi(x_2,x_5,x_6)\delta(x_6,\bar{x}_6) \\
&= \frac{1}{Z}\sum_{x_2}\psi(x_1,x_2)\sum_{x_3}\psi(x_1,x_3)\sum_{x_4}\psi(x_2,x_4)\sum_{x_5}\psi(x_3,x_5)m_6(x_2,x_5) \\
&= \frac{1}{Z}\sum_{x_2}\psi(x_1,x_2)\sum_{x_3}\psi(x_1,x_3)m_5(x_2,x_3)\sum_{x_4}\psi(x_2,x_4) \\
&= \frac{1}{Z}\sum_{x_2}\psi(x_1,x_2)m_4(x_2)\sum_{x_3}\psi(x_1,x_3)m_5(x_2,x_3) \\
&= \frac{1}{Z}\sum_{x_2}\psi(x_1,x_2)m_4(x_2)m_3(x_1,x_2) \\
&= \frac{1}{Z}m_2(x_1).
\end{aligned}
$$



Elimination; undirected graphs

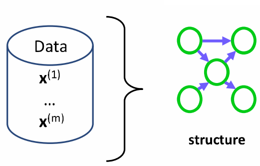
$$
\begin{aligned}
p(x_1, \bar{x}_6) &= \frac{1}{Z}\sum_{x_2}\sum_{x_3}\sum_{x_4}\sum_{x_5}\sum_{x_6}\psi(x_1,x_2)\psi(x_1,x_3)\psi(x_2,x_4)\psi(x_3,x_5)\psi(x_2,x_5,x_6)\delta(x_6,\bar{x}_6) \\
&= \frac{1}{Z}\sum_{x_2}\psi(x_1,x_2)\sum_{x_3}\psi(x_1,x_3)\sum_{x_4}\psi(x_2,x_4)\sum_{x_5}\psi(x_3,x_5)\sum_{x_6}\psi(x_2,x_5,x_6)\delta(x_6,\bar{x}_6) \\
&= \frac{1}{Z}\sum_{x_2}\psi(x_1,x_2)\sum_{x_3}\psi(x_1,x_3)\sum_{x_4}\psi(x_2,x_4)\sum_{x_5}\psi(x_3,x_5)m_6(x_2,x_5) \\
&= \frac{1}{Z}\sum_{x_2}\psi(x_1,x_2)\sum_{x_3}\psi(x_1,x_3)m_5(x_2,x_3)\sum_{x_4}\psi(x_2,x_4) \\
&= \frac{1}{Z}\sum_{x_2}\psi(x_1,x_2)m_4(x_2)\sum_{x_3}\psi(x_1,x_3)m_5(x_2,x_3) \\
&= \frac{1}{Z}\sum_{x_2}\psi(x_1,x_2)m_4(x_2)m_3(x_1,x_2) \\
&= \frac{1}{Z}m_2(x_1).
\end{aligned}
$$



$$p(\bar{x}_6) = \frac{1}{Z}\sum_{x_1} m_2(x_1),$$

$$p(x_1 \mid \bar{x}_6) = \frac{m_2(x_1)}{\sum_{x_1} m_2(x_1)},$$ <--- No Z!

Learning



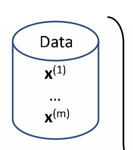
\+ CPTs – $P(X_i \mid \mathbf{Pa}_{X_i})$

parameters

Given set of m independent samples (assignments of random variables),

find the best (most likely?) Bayes Net (graph Structure + CPTs)

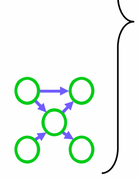
Learning the CPTs (given structure)



For each discrete variable X_k

Compute MLE or MAP estimates for

$$p(x_k|\text{pa}_k)$$



Recall

MLE: $P(X_i = x_i \mid X_j = x_j) = \dfrac{\text{Count}(X_i = x_i, X_j = x_j)}{\text{Count}(X_j = x_j)}$

MAP: Add psuedocounts