Introduction to Graphical Models

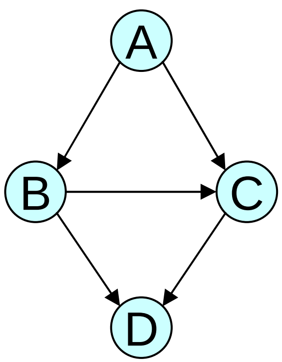
Kenji Fukumizu
The Institute of Statistical Mathematics

Computational Methodology in Statistical Inference II

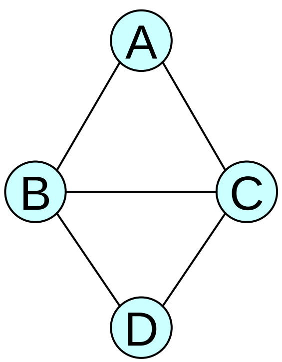
Introduction and Review

Graphical Models – Rough Sketch

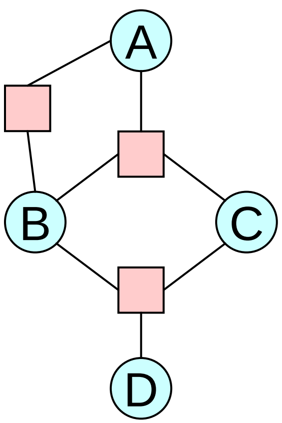
- Graphical models
  - Graph: $G = (V, E)$ V: the set of nodes, E: the set of edges
  - In graphical models,
    - the random variables are represented by the nodes.
    - statistical relationships between the variables are represented by the edges.



Directed graph



Undirected graph



Factor graph

Purpose of using Graphical Models

- Intuitive and visual representation

  A graph is an intuitive way of representing and visualizing the relationships among variables.

- Independence / conditional independence

  A graph represents conditional independence relationships among variables.

  → Causal relationships, decision making, diagnosis system, etc.

- Efficient computation

  With graphs, efficient propagation algorithms can be defined.

  → Belief-propagation, junction tree algorithm

  Which parts of the modeling block efficient computation?

Independence

For simplicity, it is assumed that the distribution of a random variable X has the probability density function $p_X(x)$.

- Independence
  - X and Y are independent ($X \perp\!\!\!\perp Y$)

$$\Leftrightarrow \quad p_{XY}(x, y) = p_X(x) p_Y(y)$$

$X \perp\!\!\!\perp Y$

Dawid's notation

Conditional Independence

- Conditional probability
  - Conditional probability density of Y given X

$$\text{Def.} \quad p_{Y|X}(y \mid x) = \frac{p_{XY}(x,y)}{p_X(x)} = \frac{p_{XY}(x,y)}{\sum_y p_{XY}(x,y)}$$

- Conditional independence
  - X and Y are conditionally independent given Z $\quad (X \perp\!\!\!\perp Y \mid Z)$

$$\underset{\text{def.}}{\Leftrightarrow} \quad p_{XY|Z}(x,y \mid z) = p_{X|Z}(x \mid z)\, p_{Y|Z}(y \mid z)$$

for all z with $p_Z(z) > 0$.

  - $X \perp\!\!\!\perp Y \mid Z \quad \Leftrightarrow \quad p_{X|YZ}(x \mid y,z) = p_{X|Z}(x \mid z)$ for all (y,z) with $p_{YZ}(y,z) > 0$.

If we already know Z, additional information on Y does not increase the knowledge on X.

Conditional Independence - Examples

- Speeding Fine $\not\perp\!\!\!\perp$ Type of Car (perhaps)
- Speeding Fine $\perp\!\!\!\perp$ Type of Car | Speed
- Ability of Team A $\perp\!\!\!\perp$ Ability of Team B
- Ability of Team A $\not\perp\!\!\!\perp$ Ability of Team B | Outcome of Team A and B

Conditional Independence

- Another characterization of cond. independence

<u>Proposition 1</u>

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

$$\Longleftrightarrow$$

there exist functions $f(x,z)$ and $g(y,z)$ such that

$$p_{XYZ}(x,y,z) = f(x,z)g(y,z)$$

for all x, y and z with $p_Z(z) > 0$.

<u>Corollary 2</u>

$$X \perp\!\!\!\perp Y$$

$$\Longleftrightarrow$$

there exist functions $f(x)$ and $g(y)$ such that

$$p_{XY}(x,y) = f(x)g(y) \quad \text{for all } x, y.$$

Conditional Independence

- Proof of Prop.1.

  $\Longrightarrow$ Clear from the definition.

  $\Longleftarrow$ For any x, y, and z with $p_Z(z) > 0$,

$$p_Z(z) = \sum_{x,y} p_{XYZ}(x,y,z) = \sum_{x,y} f(x,z)g(y,z)$$
$$= \left(\sum_x f(x,z)\right)\left(\sum_y g(y,z)\right)$$

We have

$$p_{XY|Z}(x,y|z) = \frac{p_{XYZ}(x,y,z)}{p_Z(z)} = \frac{f(x,z)g(y,z)}{\sum_x f(x,z)\ \sum_y g(y,z)}$$

$$p_{X|Z}(x|z) = \frac{p_{XZ}(x,z)}{p_Z(z)} = \frac{\sum_y p_{XYZ}(x,y,z)}{p_Z(z)} = \frac{f(x,z)\ \cancel{\sum_y g(y,z)}}{\sum_x f(x,z)\ \cancel{\sum_y g(y,z)}}$$

$$p_{Y|Z}(y|z) = \frac{p_{YZ}(y,z)}{p_Z(z)} = \frac{\sum_x p_{XYZ}(x,y,z)}{p_Z(z)} = \frac{\cancel{\sum_x f(x,z)}g(y,z)}{\cancel{\sum_x f(x,z)}\ \sum_y g(y,z)}$$

Thus,

$$p_{XY|Z}(x,y|z) = p_{X|Z}(x|z)p_{Y|Z}(y|z)$$

Undirected Graph and Markov Property

Undirected Graph

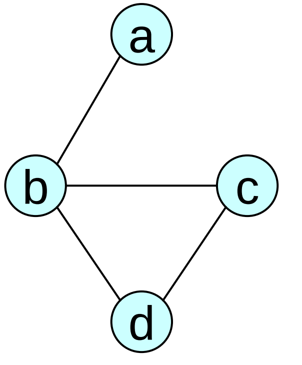
- Undirected Graph
  - $G = (V, E)$: undirected graph

    V: finite set

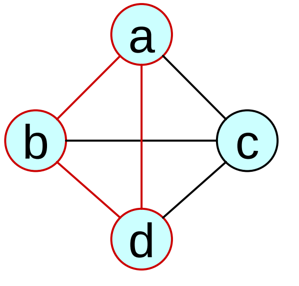
    $E \subset V \times V$, the order is neglected. $(a, b) = (b, a)$

    Example: $V = \{a, b, c, d\}$

    $E = \{(a, b), (b, c), (c, d), (b, d)\}$



- Graph terminology
  - Complete: A subgraph S of V is complete if any a and b ($a \neq b$) in S are connected by an edge.
  - Clique: A clique is a maximal complete subset w.r.t. inclusion.



(a,b,d): complete, but not a clique

Probability and Undirected Graph

- Probability associated with an undirected graph

$G = (V, E)$: undirected graph. $V = \{1, \ldots, n\}$

$X = (X_1, \ldots, X_n)$: random variables indexed by the node set V.

The probability distribution of X is associated with G if there is a non-negative function $\psi_C(X_C)$ for each clique C in G such that

$$p(X) = \prod_{C:\text{clique}} \psi_C(X_C)$$

Notation: for a subset S of V, $X_S = (X_a)_{a \in S}$

- An undirected graph does not specify a single probability, but defines a family of probabilities.
  In other words, it puts restrictions by the conditional independence relations represented by the graph.

Probability and Undirected Graph

$p(X)$ is associated with an undirected graph G if and only if it admits

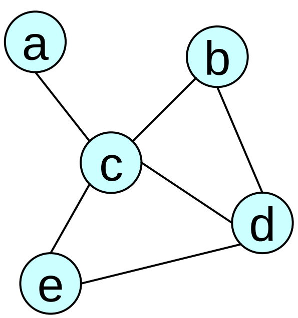
$$p(X) = \frac{1}{Z} \prod_{C:\text{clique}} \psi_C(X_C)$$

Z: normalization constant

'p factorizes w.r.t. G.'

ψ_C: factor (or potential)

Example



$$p(X) = \frac{1}{Z} \psi_1(X_a, X_c) \psi_2(X_b, X_c, X_d) \psi_3(X_c, X_d, X_e)$$

Markov Property

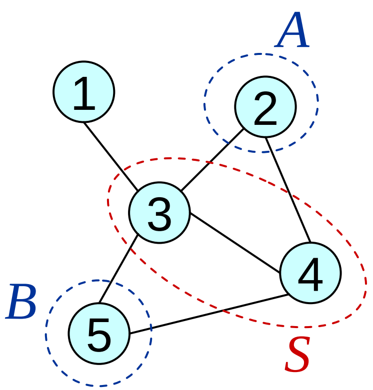
- Undirected graph and Markov property

  Separation:

  $G = (V, E)$: undirected graph.

  A, B, S: disjoint subsets of V.

  S separates A from B if every path between any a in A and b in B intersects with S.



Theorem 3

$G = (V, E)$: undirected graph.

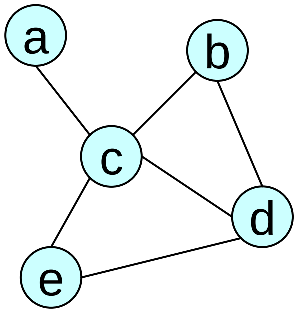
X: random vector with the distribution associated with G.

If S separates A from B, then

$$X_A \perp\!\!\!\perp X_B \mid X_S$$

(Proof: next lecture.)

Markov Property



- Example

$$p(X)=\frac{1}{Z}\psi_1(X_a,X_c)\psi_2(X_b,X_c,X_d)\psi_3(X_c,X_d,X_e)$$

- $\{c, d\}$ separates $\{b\}$ and $\{e\}$ $\Rightarrow$ $X_b \perp\!\!\!\perp X_e \mid X_{\{c,d\}}$

$$p(X_b,X_c,X_d,X_e)=\frac{1}{Z}\sum_{X_a}\psi_1(X_a,X_c)\psi_2(X_b,X_c,X_d)\psi_3(X_c,X_d,X_e)$$

$$=\frac{1}{Z}\boxed{\tilde{\psi}_1(X_c)\psi_2(X_b,X_c,X_d)}\boxed{\psi_3(X_c,X_d,X_e)}$$

$$=\frac{1}{Z}f(X_b,X_c,X_d)g(X_e,X_c,X_d)$$

Use prop.1.

- $\{c\}$ separates $\{a\}$ and $\{b\}$ $\Rightarrow$ $X_a \perp\!\!\!\perp X_b \mid X_c$

$$p(X_a,X_b,X_c)=\frac{1}{Z}\psi_1(X_a,X_c)\sum_{X_d,X_e}\{\psi_2(X_b,X_c,X_d)\psi_3(X_c,X_d,X_e)\}$$

$$=\frac{1}{Z}\boxed{\psi_1(X_a,X_c)}\boxed{g(X_b,X_c)}$$

Use prop.1.

Markov Property

- Global Markov Property

  $G = (V, E)$: undirected graph

  X: random vector indexed by V.

  X satisfies global Markov property relative to G if $X_A \perp\!\!\!\perp X_B \mid X_S$ holds for any triplet (A,B,S) of disjoint subsets of V such that S separates A from B.

  The previous theorem tells
  if the distribution of X factorizes w.r.t. G, then X satisfies global Markov property relative to G.

  Remark: Both of 'factorize' and 'global Markov property' are the properties regarding a relation between the probability $p(X)$ and the undirected graph G.

Markov Property

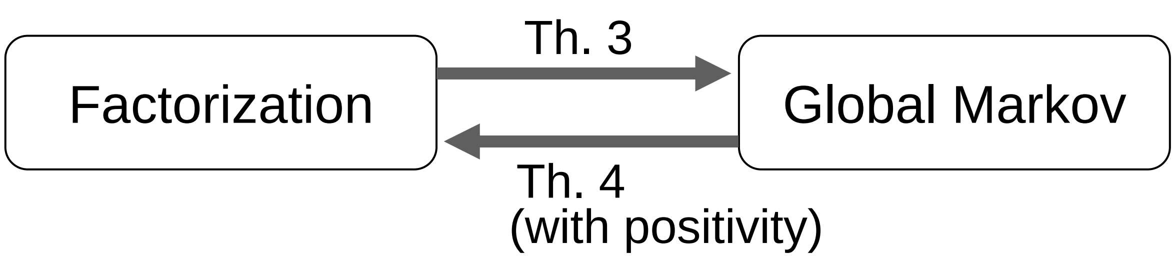
- Hammersley-Clifford theorem (see e.g. Lauritzen. Th.3.9)

Theorem 4

$G = (V, E)$: undirected graph
X: random vector indexed by V.
Assume that the probability density function $p(X)$ of the distribution of X is strictly positive.

If X satisfies global Markov property w.r.t. G, then X factorizes w.r.t. G, *i.e.* $p(X)$ admits the factorization:

$$p(X) = \prod_{C:\text{clique}} \psi_C(X_C).$$



Directed Acyclic Graph and Markov Property

Directed Acyclic Graph

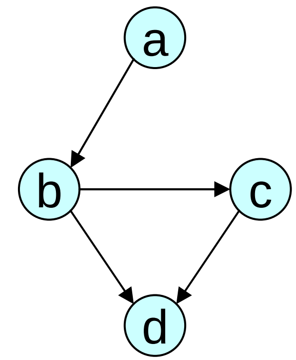
- Directed Graph
  - $G = (V, E)$: directed graph

    V: finite set -- nodes

    $E \subset V \times V$: set of edges

    Example: $V = \{a, b, c, d\}$

    $E = \{(a,b), (b,c), (c,d), (b,d)\}$

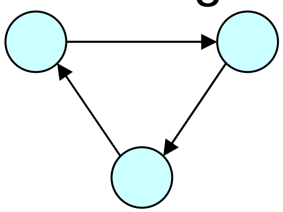


Orient the edge (a,b) by $a \rightarrow b$

  - Directed Acyclic graph (DAG)

    Directed graph with no cycles.

    Cycle: directed path starting and ending at the same node.



DAG and Probability

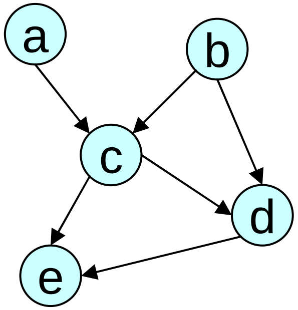
- Probability associated with a DAG
  - A DAG defines a family of probability distributions

$$p(X_1,\ldots,X_n)=\prod_{i=1}^{n} p(X_i \mid X_{pa(i)})$$

$$pa(i)=\{j\in V \mid (i,j)\in E\}$$: parents of node i.

p is said to be associated with DAG G, or p factorizes w.r.t. G.

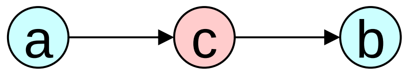
Example:



$$p(X_a,X_b,X_c,X_d,X_e)$$
$$=p(X_a)p(X_b)p(X_c \mid X_a,X_b)p(X_d \mid X_b,X_c)p(X_e \mid X_c,X_d)$$

Conditional Independence with DAG

- Three basic cases

(1) 

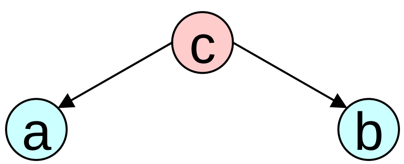
$$p(X_a, X_b, X_c) = p(X_a) p(X_c \mid X_a) p(X_b \mid X_c)$$

$$X_a \perp\!\!\!\perp X_b \mid X_c$$

Note $p(X_a) p(X_c \mid X_a) = p(X_a, X_c) = p(X_c) p(X_a \mid X_c)$

$$\Rightarrow \quad p(X_a, X_b, X_c) = p(X_c) p(X_a \mid X_c) p(X_b \mid X_c)$$

$$p(X_a, X_b \mid X_c) = p(X_a \mid X_c) p(X_b \mid X_c)$$

(2) 

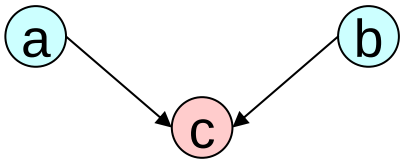
$$p(X_a, X_b, X_c) = p(X_c) p(X_a \mid X_c) p(X_b \mid X_c)$$

$$X_a \perp\!\!\!\perp X_b \mid X_c$$

Note: $p(X_a, X_b, X_c)$ are the same for (1) and (2).

Conditional Independence with DAG

(3)

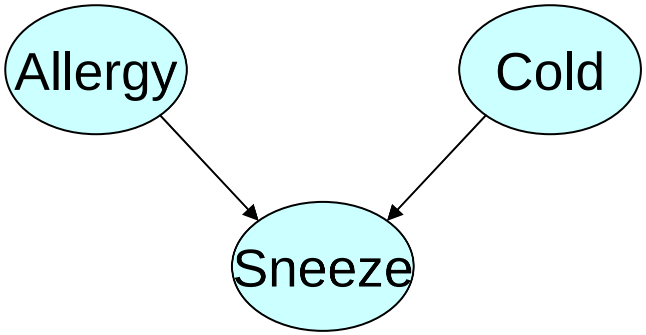


head-to-head
(or v-structure)

$$p(X_a, X_b, X_c) = p(X_a)\,p(X_b)\,p(X_c \mid X_a, X_b)$$

$$X_a \not\perp\!\!\!\perp X_b \mid X_c \qquad X_a \perp\!\!\!\perp X_b$$

Note: $p(X_a, X_b, X_c)$ in (3) are different from (1) and (2).



If you often sneeze, but you do not have cold, then it is more likely you have allergy (hay fever).

D-Separation

- Blocked:

  An undirected path π is said to be blocked by a subset S in V if there exists a node c on the path such that either

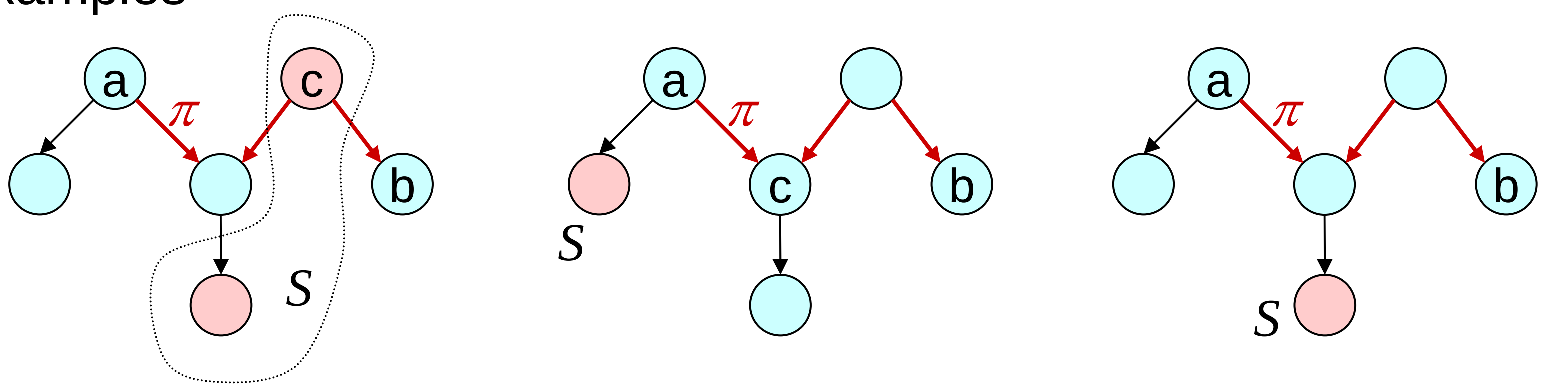
  (i) $c \in S$ and c is not head-to-head in π ($\bigcirc \rightarrow (c) \rightarrow \bigcirc$ or $\bigcirc \leftarrow (c) \rightarrow \bigcirc$),

  (ii) $\bigcirc \rightarrow (c) \leftarrow \bigcirc$ and $(\{c\} \cup de(c)) \cap S = \phi.$

  head-to-head

  Descendent: $de(i) = \{ j \in V \mid \exists \text{directed path from } i \text{ to } j \}$

Examples



π is blocked by S | π is blocked by S | π is NOT blocked by S

D-Separation

- d-separate:

  A, B, S: disjoint subsets of V.

  S d-separates A from B if every undirected path between a in A and b in B is blocked by S.

- d-separation and conditional independence

<u>Theorem 5</u>

X: random vector with the distribution associated with DAG G.

A, B, S: disjoint subsets of V.
If S d-separates A from B, then

$$X_A \perp\!\!\!\perp X_B \mid X_S$$

(Proof not shown in this course. See Lauritzen 1996, 3.23&3.25)

D-Separation

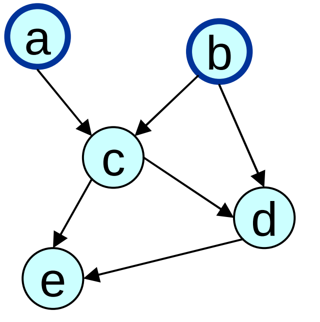
- Example
  - $X_a \perp\!\!\!\perp X_b$

    $S = \phi$.

    a → c ← b is blocked (with c).

    a → c → d ← b is blocked (with d)

    a → c → e ← d ← b is blocked (with e)

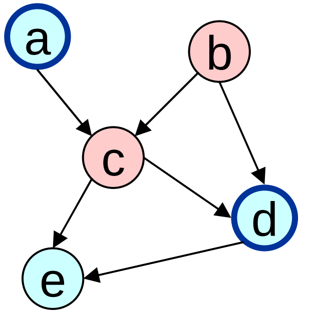


  - $X_a \perp\!\!\!\perp X_d \mid X_{\{b,c\}}$

    a → c → d is blocked (with c).

    a → c ← b ← d is blocked (with b)

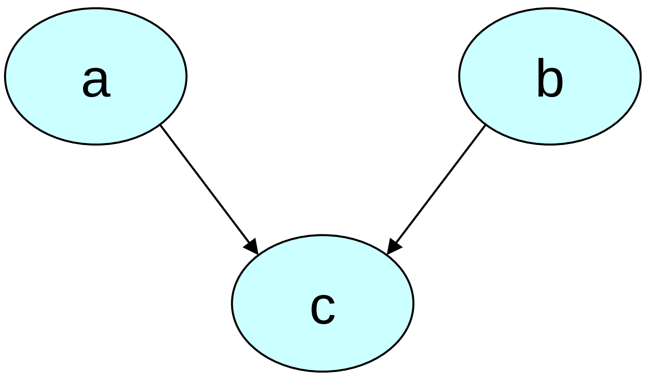
    a → c → e ← d is blocked (with e or c)



Comparison: UDG and DAG

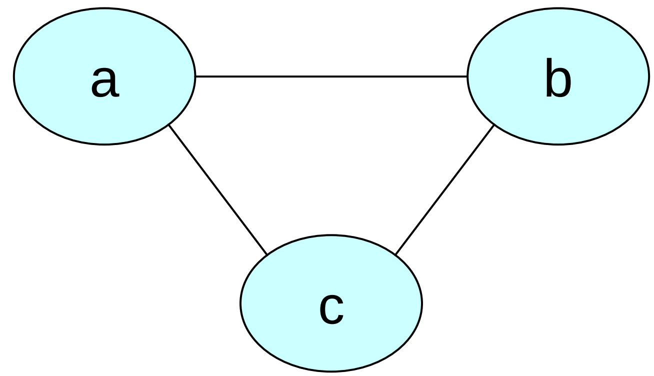
- Limitation of undirected graph

$$p(X_a, X_b, X_c) = p(X_a)\,p(X_b)\,p(X_c \mid X_a, X_b)$$



DAG

$X_a \perp\!\!\!\perp X_b$, $\quad X_a \not\!\perp\!\!\!\perp X_b \mid X_c$



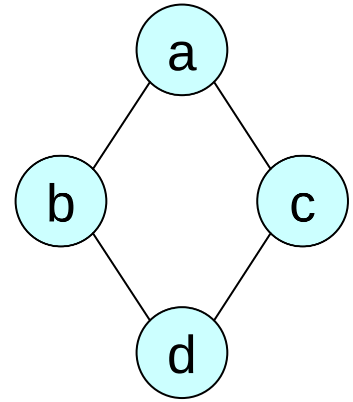
If $X_a \not\!\perp\!\!\!\perp X_c$, $X_b \not\!\perp\!\!\!\perp X_c$, $X_a \not\!\perp\!\!\!\perp X_b \mid X_c$,
any UDG is not able to
express $X_a \perp\!\!\!\perp X_b$.

Comparison: UDG and DAG

- Limitation of DAG

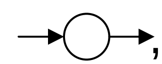
Undirected graph

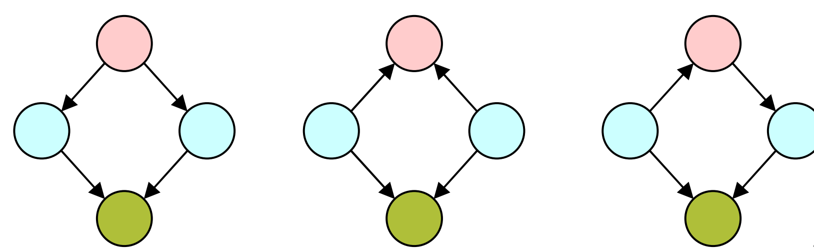
$$p(X_a, X_b, X_c, X_d)$$
$$= p(X_a, X_b)p(X_a, X_c)p(X_b, X_d)p(X_c, X_d)$$



$$X_a \perp\!\!\!\perp X_d \mid X_{\{b,c\}} \qquad X_b \perp\!\!\!\perp X_c \mid X_{\{a,d\}}$$

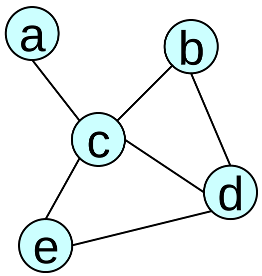
No DAG expresses these conditional independence relationships.

[Sketch of the proof.] If every node had the form , the graph would be a cycle. Thus, there must be a v-structure. Conditional independence of the parents of the v-structure given the other two nodes cannot be expressed by a DAG.



Mini Summary on UDG and DAG

Undirected graph



- Probability associated with G, ($p(X)$ factorizes w.r.t. G)

$$p(X) = \frac{1}{Z} \prod_{C:\text{clique}} \psi_C(X_C)$$

- $p(X)$ factorizes w.r.t. G

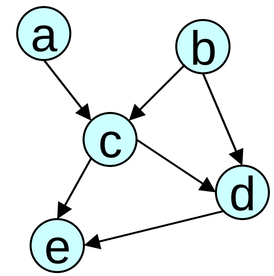
$\Rightarrow$

X is global Markov relative to G.

(*i.e*. if S separates A from B,

then $X_A \perp\!\!\!\perp X_B \mid X_S$.)

Directed acyclic graph (DAG)



- Probability associated with G ($p(X)$ factorizes w.r.t. G)

$$p(X_1, \ldots, X_n) = \prod_{i=1}^{n} p(X_i \mid X_{pa(i)})$$

- $p(X)$ factorizes w.r.t. G

$\Rightarrow$

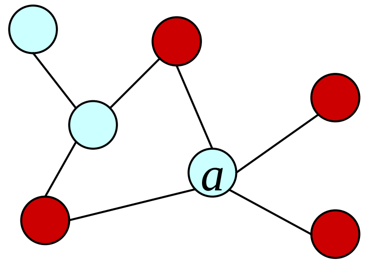
X is d-global Markov relative to G.

(*i.e*. if S d-separates A from B,

then $X_A \perp\!\!\!\perp X_B \mid X_S$.)

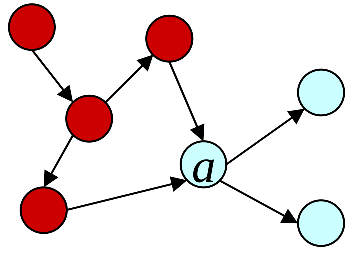
Appendix: Terminology on Graphs

- Undirected graph $G = (V, E)$
  - Adjacent: a and b in V $(a \neq b)$ are adjacent if $(a, b) \in E$.
  - Neighbor: $ne(a) = \{b \in V \mid (a, b) \in E\}.$

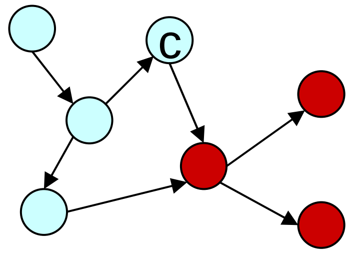


ne(a)

- DAG $G = (V, E)$
  - Parents: $pa(a) = \{b \in V \mid (b, a) \in E\}.$
  - Children: $ch(a) = \{b \in V \mid (a, b) \in E\}.$
  - Ancestors:
    $an(a) = \{b \in V \mid \exists \text{ directed path from } b \text{ to } a\}.$
  - Descendents:
    $de(a) = \{b \in V \mid \exists \text{ directed path from } a \text{ to } b\}.$



an(a)



de(c)

Factor Graph and Markov Property

Factor Graph

- Factor graph $G = (V, E)$

  $V = (I, F)$: two types of nodes

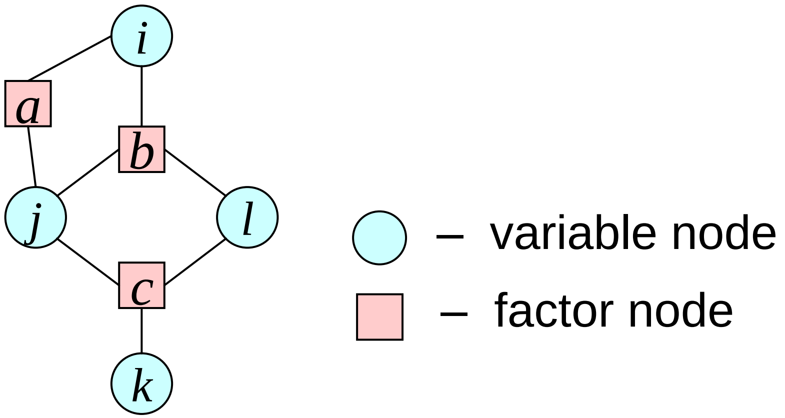
  I: variable nodes

  F: factor nodes

  E: undirected edges

  $$E \subset I \times F \subset V \times V.$$

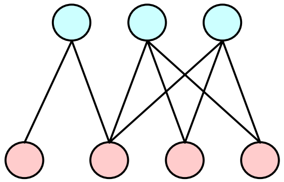
  An edge exists only between a factor node and a variables node.



- A factor graph is in general called bipartite graph.

  A bipartite graph is an undirected graph $G = (V, E)$ such that

  $$V = V_1 \cup V_2,\ V_1 \cap V_2 = \phi,\ E \subset V_1 \times V_2.$$



Probability and Factor graph

- Factor graph to represent factorization
  - $X = (X_i)_{i \in I}$: random vector indexed by a finite set I.
    The density of the distribution of X factorizes as

$$p(X) = \frac{1}{Z} \prod_{a \in F} f_a(X^{(a)})$$

F: finite set.
Z: normalization constant

f_a: non-negative function of a subset of $\{X_1, \ldots, X_n\}$

$X^{(a)} = (X_i)_{i \in I_a}$, where $I_a := \{i \in I \mid (i,a) \in E\}$

  - The factor graph $G = (V, E)$ representing the factorization is given by

$V = (I, F)$

$$E = \{(i,a) \in I \times F \mid i \in I_a\}$$

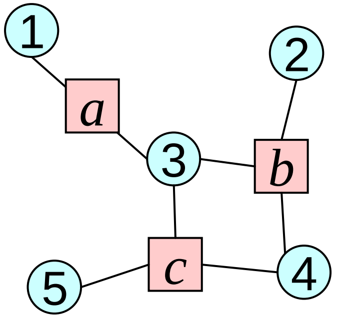
Probability and Factor graph

- Example

$I = \{1,2,3,4,5\}$

$F = \{a,b,c\}$

$$p(X) = \frac{1}{Z} f_a(X_1, X_3) f_b(X_2, X_3, X_4) f_c(X_3, X_4, X_5)$$

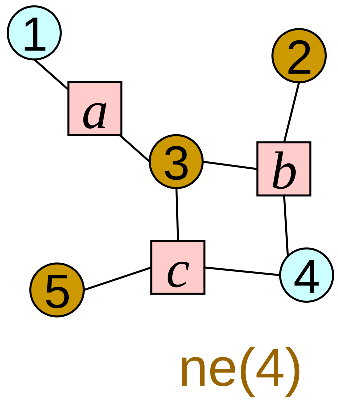


  - A probability is often given by a factorized form, *i.e*., a product of factors with a small number of variables.

Markov Property of Factor Graph

- ne(i): neighbor of a variable node i

$$ne(i) = \{ j \in I \mid \exists a \in F, \{i, j\} \subset I_a \}.$$



- A path in a factor graph is a sequence of variables nodes such that any consecutive two nodes are neighbors.
  e.g. 2 – 3 – 5.

- Factorization → global Markov property

<u>Theorem 6</u>

Assume the probability of X factorizes w.r.t. a factor graph G.

S, A, B: disjoint subsets of the variable nodes I.

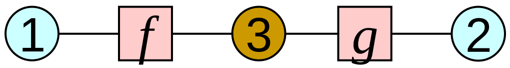
If every path between any a in A and b in B intersects with S, then

$$X_A \perp\!\!\!\perp X_B \mid X_S$$

Markov Property of Factor Graph

- Example 1

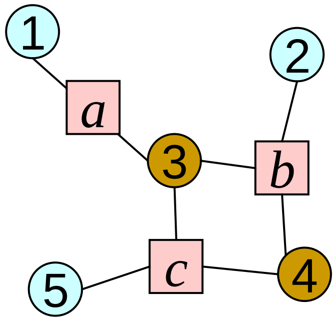
$$p(X) = \frac{1}{Z} f(X_1, X_3) g(X_2, X_3)$$



$$X_1 \perp\!\!\!\perp X_2 \mid X_3$$

- Example 2

$$p(X) = \frac{1}{Z} f_a(X_1, X_3) f_b(X_2, X_3, X_4) f_c(X_3, X_4, X_5)$$



$$X_1 \perp\!\!\!\perp X_5 \mid X_{\{3,4\}}$$

Direct confirmation

$$p(X_1, X_3, X_4, X_5) = \sum_{X_2} p(X) = \frac{1}{Z} f_a(X_1, X_3) \sum_{X_2} f_b(X_2, X_3, X_4) f_c(X_3, X_4, X_5)$$

$$= \frac{1}{Z} \boxed{f_a(X_1, X_3) g(X_3, X_4)} \boxed{f_c(X_3, X_4, X_5)}$$

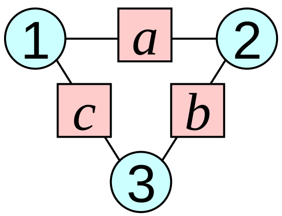
$$= \frac{1}{Z} \varphi(X_1, X_3, X_4) \psi(X_3, X_4, X_5) \quad \text{(Prop.1)}$$

Comparison of Factor Graph and other graphs

- Factor graph and UDG

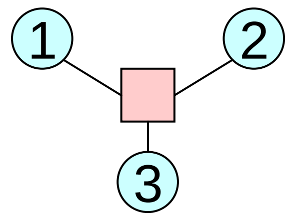
Factor graphs

(i)



$$p(X) = \frac{1}{Z} f_a(X_1, X_2) f_b(X_2, X_3) f_c(X_1, X_3)$$

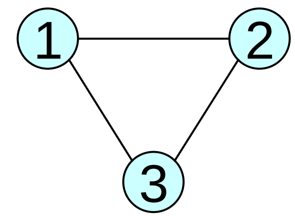
(ii)



$$p(X) = p(X_1, X_2, X_3)$$

Undirected graph

(iii)



$$p(X) = p(X_1, X_2, X_3)$$

UDG cannot distinguish the factorization in (i) and (ii)

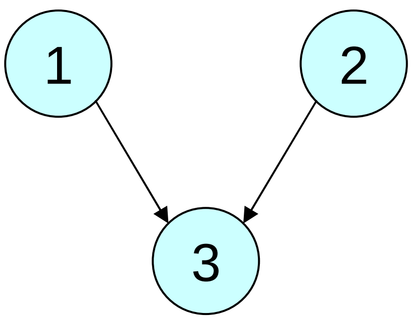
  - All the variable nodes in (i), (ii), and (iii) have the same neighbors, and thus the same conditional independence relationships (no conditional independence).
  - The factor graph representations of (i) and (ii) are different.

Comparison of Factor Graph and other graphs

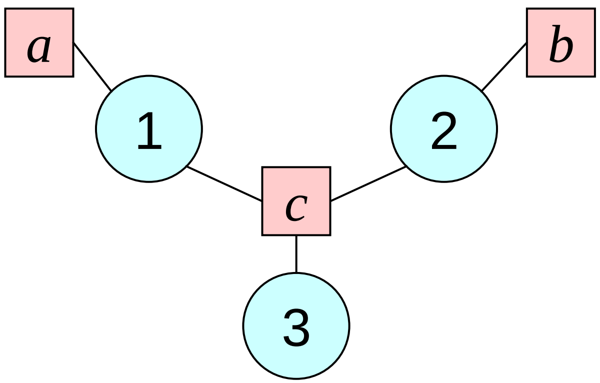
- Factor graph and DAG

$$p(X_1, X_2, X_3) = p(X_1)\,p(X_2)\,p(X_3 \mid X_1, X_2)$$

DAG



Factor graph

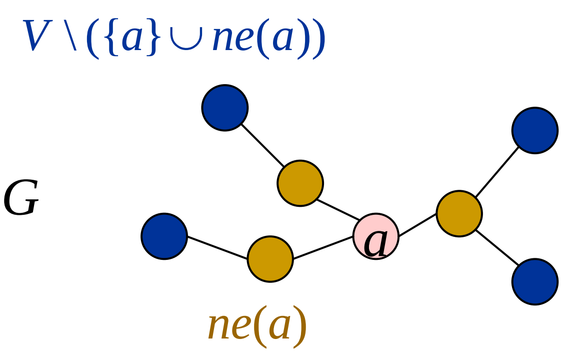


Independence of 1 and 2 cannot be represented.

More on Markov Property

Markov Properties Revisited

- Markov properties for an undirected graph
  $G = (V, E)$: undirected graph.
  X: random vector indexed by V.

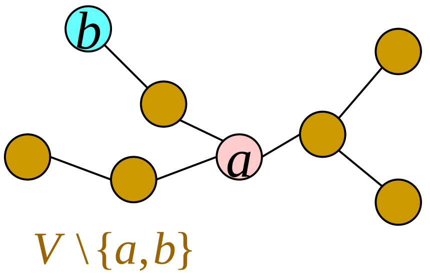


  - Local Markov
    X satisfies local Markov property relative to G if for any node a

$$X_a \perp\!\!\!\perp X_{V \setminus (\{a\} \cup ne(a))} \mid X_{ne(a)}$$

  - Pairwise Markov
    X satisfies pairwise Markov property relative to G if any non-adjacent pair of nodes (a, b) satisfies

$$X_a \perp\!\!\!\perp X_b \mid X_{V \setminus \{a,b\}}$$



Markov Properties Revisited

<u>Theorem 7</u>

Factorization $\Rightarrow$ global Markov $\Rightarrow$ local Markov
$\Rightarrow$ pairwise Markov

proof) factorization $\Rightarrow$ global Markov : Theorem 3.
global Markov $\Rightarrow$ local Markov : easy.
local Markov $\Rightarrow$ pairwise Markov : needs some math (Exercise).

- Hammersley-Clifford asserts that the pairwise Markov property means factorization w.r.t. the graph under positivity of the density. (Theorem 4 assumes 'global Markov', but the assertion holds under 'pairwise Markov' assumptoin.)

- Similar notions are defined for directed and factor graphs.

Proof for Undirected Case

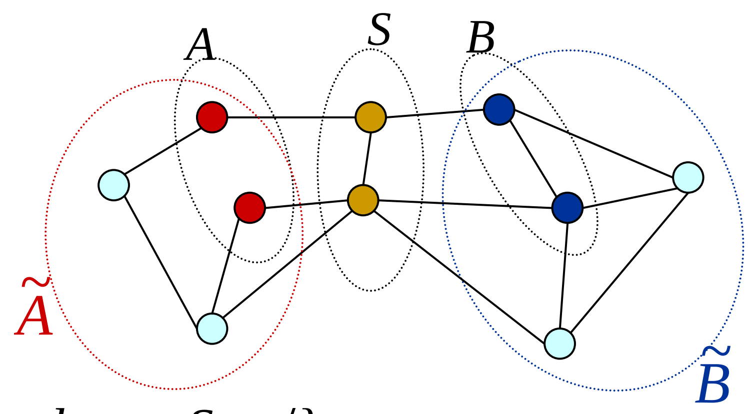
We show a slight generalization of Theorem 3.

Theorem 8

Let $G = (V, E)$ be an undirected graph. If the distribution of X factorizes as

$$p(X) = \frac{1}{Z} \prod_{C:\text{complete}} \psi_C(X_C),$$

then X satisfies global Markov property relative to G, *i.e.*, for a triplet (S, A, B) such that S separates A from B, the conditional independence $X_A \perp\!\!\!\perp X_B \mid X_S$ holds.



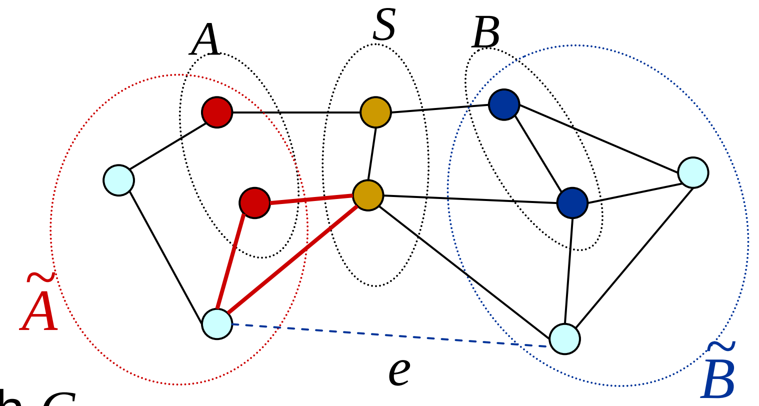
- Proof

Let

$$\tilde{A} = \{d \in V \setminus S \mid \exists a \in A, \exists \pi \text{ path from } a \text{ to } d, \pi \cap S = \phi\},$$

$$\tilde{B} = V \setminus (\tilde{A} \cup S).$$

Proof for Undirected Case



Obviously $A \subset \tilde{A}$,

and since S separates A from B,

$$B \subset \tilde{B}.$$

We can show for any complete subgraph C

$$C \subset S \cup \tilde{A} \text{ or } C \subset S \cup \tilde{B} \text{ holds.}$$

If $C \subset S$, there is nothing to prove.

Assume $C \not\subset S$.

Suppose that the above assertion does not hold, then

$C \cap \tilde{A} \neq \phi$ and $C \cap \tilde{B} \neq \phi$. Let $a \in \tilde{A} \cap C$ and $b \in \tilde{B} \cap C$. Because a and b are in the complete subgraph C, there is an edge e connecting a and b. Since $a \in \tilde{A}$, there is a path π from a to A without intersecting S. Connecting π and e makes a path from b to A without intersecting S, which contradicts with the definition of $\tilde{A}$ and $\tilde{B}$.

Proof for Undirected Case

From this fact,

$$p(X) = \frac{1}{Z} \prod_{C:\text{complete}} \psi_C(X_C) = \frac{1}{Z} \prod_{\substack{C:\text{complete} \\ C \subset S \cup \tilde{A}}} \psi_C(X_C) \prod_{\substack{D:\text{complete} \\ D \subset S \cup \tilde{B}}} \psi_D(X_D)$$

$$= f(X_{\tilde{A}}, X_S) g(X_{\tilde{B}}, X_S)$$

which means

$$X_{\tilde{A}} \perp\!\!\!\perp X_{\tilde{B}} \mid X_S, \qquad \text{(Proposition 1)}$$

and thus

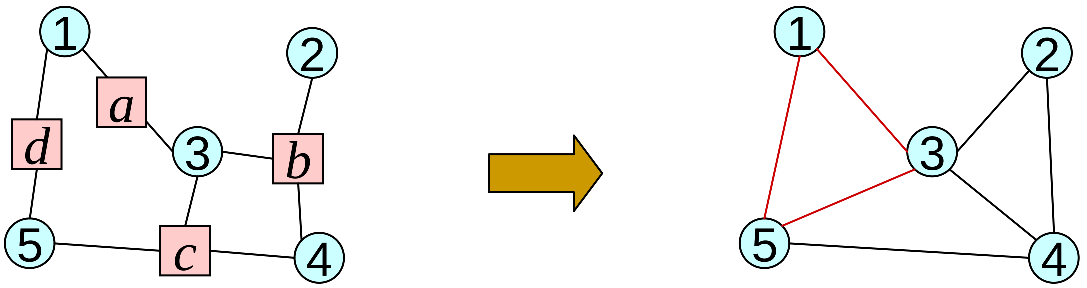
$$X_A \perp\!\!\!\perp X_B \mid X_S.$$

Q.E.D.

Converting Factor Graph to UDG

Neighborhood structure by a factor graph make an undirected graph.

$$p(X) = \frac{1}{Z} f_a(X_1, X_3) f_b(X_2, X_3, X_4) f_c(X_3, X_4, X_5) f_d(X_1, X_5) \quad \cdots \text{(A)}$$



Factor graph G Undirected graph U

Each factor in (A) does not correspond to a clique in U,
but to a complete subgraph in U.

In general, $p(X)$ factorizes as

$$p(X) = \frac{1}{Z} \prod_{C:\text{complete}} \psi_C(X_C),$$

for the converted undirected graph U.

Proof for Factor Graph

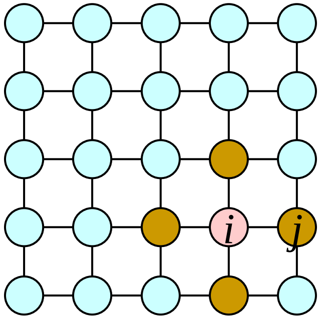
- Proof of Theorem 6 (‘Factorization → Global Markov’ for factor graph)

  From the above observation, the proof is done by Theorem 8.

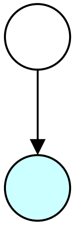
Practical Examples

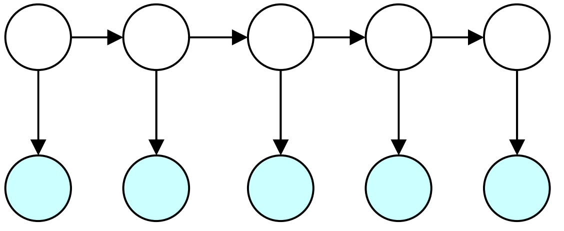
- Markov random field for image analysis

$$p(X) = \frac{1}{Z} \prod_{(i,j)\in E} \exp\left(-U_{ij}(X_i, X_j)\right)$$

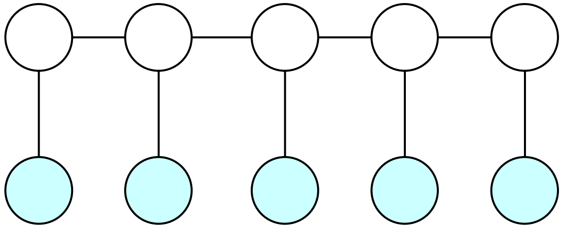


- Mixture model and hidden Markov model





- Conditional random field for sequential data (Lafferty et al. 2001)



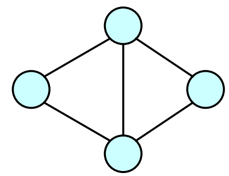
Hidden label sequence

Observation

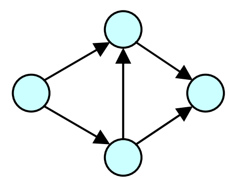
Summary

- A graph represents the conditional independence relationships among random variables.

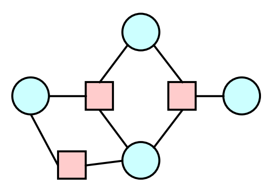
- There are many types of graph to represent probabilities.
  - Undirected graph



  - Directed graph



  - Factor graph



Factorization of the probability distribution w.r.t. a graph means Markov Property of the distribution relative to the graph.