グラフィカルモデルの基礎

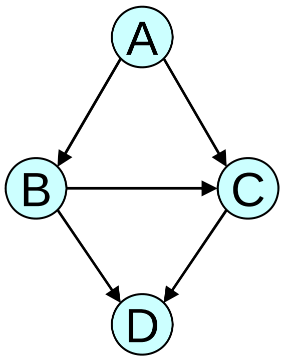
Kenji Fukumizu
The Institute of Statistical Mathematics

計算推論科学概論 II （2010年度, 後期）

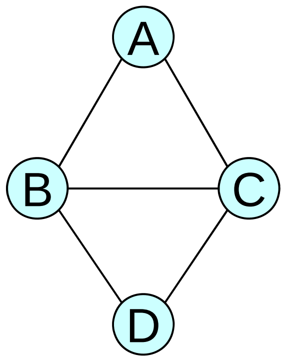
Introduction and Review

Graphical Model – Rough Sketch

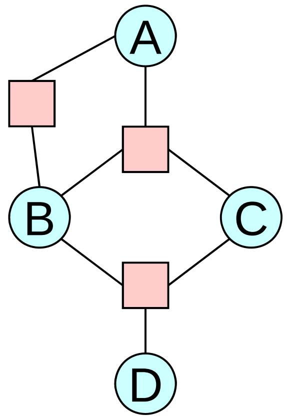
- Graphical model
  - Graph: $G = (V, E)$ V: the set of nodes, E: the set of edges
  - In graphical models,
    - the random variables are represented by the nodes.
    - statistical relationships between the variables are represented by the edges.



Directed graph



Undirected graph



Factor graph*

* Factor graphは本講義では議論しない

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

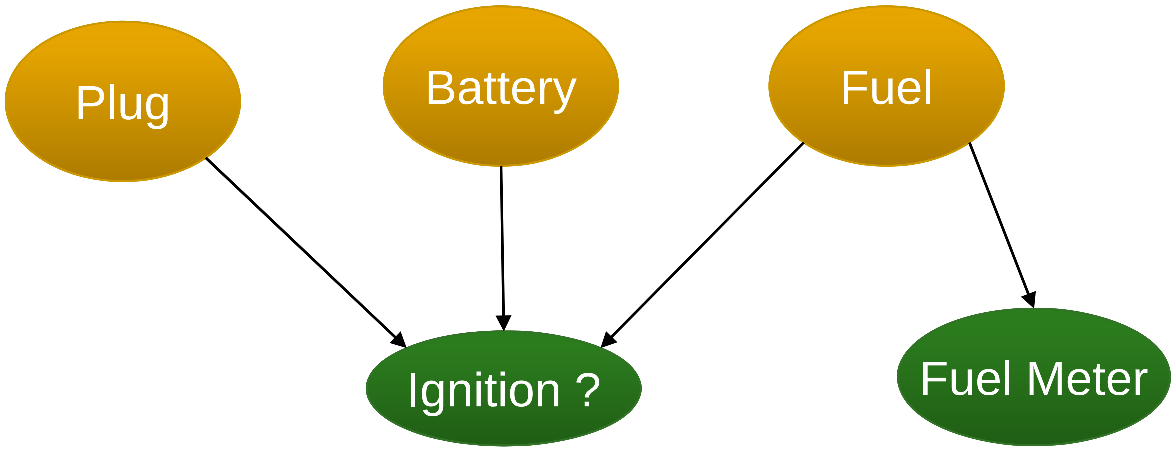
  With graphs, efficient propagation algorithms can be defined.

  → Belief-propagation, junction tree algorithm

  Which parts of the modeling block efficient computation?

Example: Diagnosis

Car start problem



If Fuel Meter indicates “full” and Plug is checked to be clean, it is more likely that the battery is dead.

Review: 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

Review: Conditional Probability

- Conditional probability density of Y given X

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

Review: Conditional Independence

David's notation

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

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

    "conditional " independendnence

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

    $\Leftrightarrow \quad p_{X|YZ}(x \mid y, z) = p_{X|Z}(x \mid z)$ 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

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

Conditional Independence

- Proof of Prop.1.

  $\Rightarrow$ Clear from the definition.

  $\Leftarrow$ 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

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

$$p_{Y|Z}(y\,|\,z) = \sum_x p_{XY|Z}(x,y\,|\,z) = \frac{\cancel{\sum_x f(x,z)}g(y,z)}{\cancel{\sum_{\tilde{x}} f(\tilde{x},z)}\sum_{\tilde{y}} g(\tilde{y},z)}$$

$$p_{X|Z}(x\,|\,z) = \sum_y p_{XY|Z}(x,y\,|\,z) = \frac{f(x,z)\cancel{\sum_y g(y,z)}}{\sum_{\tilde{x}} f(\tilde{x},z)\cancel{\sum_{\tilde{y}} g(\tilde{y},z)}}$$

Thus,

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

Undirected Graph and Markov Property

Undirected Graph

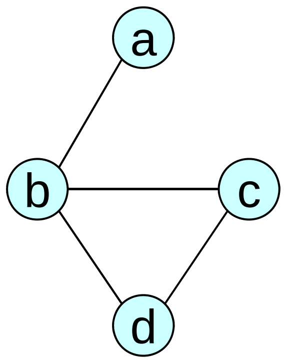
- Undirected Graph
  - $G = (V, E)$: undirected graph （無向グラフ）

    V: finite set

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

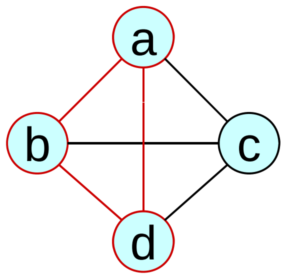
    Example: $V = \{a, b, c, d\}$

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



- Graph terminology
  - Complete: A subgraph S of V is complete（完全）

    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

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

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

The probability density function $p(X)$ factorizes with respect to G if there is a non-negative function $\psi_C(X_C)$ for each clique C in G such that

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

(Z: normalization constant)

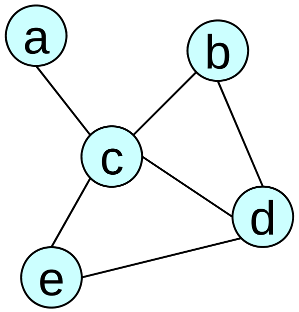
Notation: for a subset S of V, $X_S = (X_a)_{a\in S}$

- A.k.a. Markov random field

- An undirected graph does not specify a single probability, but defines a family of probabilities, *i.e*. it defines a set of conditional independence relations.

Probability and Undirected Graph

Example

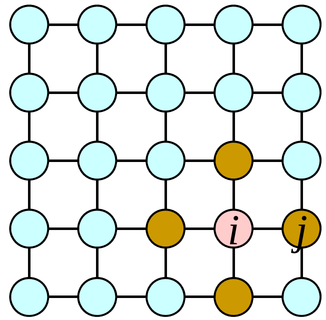


$$p(X) = \frac{1}{Z}\psi_1(X_a, X_c)\psi_2(X_b, X_c, X_d)\psi_3(X_c, X_d, X_e)$$

Practical Example

- Markov random field for image

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



X_i in $\{+1,-1\}$ (binary image)

e.g. $U_{ij}(X_i, X_j) = \beta X_i X_j$ (neighbors tend to take the same value)

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

c.f. Ising model in statistical physics

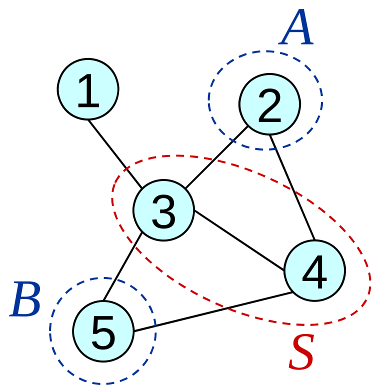
Markov Property

Def. 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 such that the p.d.f. factorizes w.r.t. G.

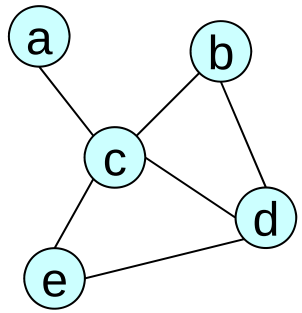
If S separates A from B, then

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

(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 $\{e\}$ $\Rightarrow$ $X_b \perp\!\!\!\perp X_e \mid X_{\{c,d\}}$

$$\begin{aligned} p(X_b, X_c, X_d, X_e) &= \frac{1}{Z}\sum_{X_a}\psi_1(X_a, X_c)\psi_2(X_b, X_c, X_d)\psi_3(X_c, X_d, X_e) \\ &= \frac{1}{Z}\boxed{\tilde{\psi}_1(X_c)\psi_2(X_b, X_c, X_d)}\boxed{\psi_3(X_c, X_d, X_e)} \\ &= \frac{1}{Z} f(X_b, X_c, X_d) g(X_e, X_c, X_d) \end{aligned}$$

Use prop.1.

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

$$\begin{aligned} p(X_a, X_b, X_c) &= \frac{1}{Z}\psi_1(X_a, X_c)\sum_{X_d, X_e}\{\psi_2(X_b, X_c, X_d)\psi_3(X_c, X_d, X_e)\} \\ &= \frac{1}{Z}\boxed{\psi_1(X_a, X_c)}\boxed{g(X_b, X_c)} \end{aligned}$$

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. (Graph tells the conditional independence)

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

Markov Property

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

<u>Theorem 4</u>

$G = (V, E)$: 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:

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

$$\begin{tikzcd} \boxed{\text{Factorization}} \arrow[r, shift left, "\text{Th. 3}"] & \boxed{\text{Global Markov}} \arrow[l, shift left, "\text{Th. 4 (with positivity)}"] \end{tikzcd}$$

Directed Acyclic Graph and Markov Property

Directed Acyclic Graph

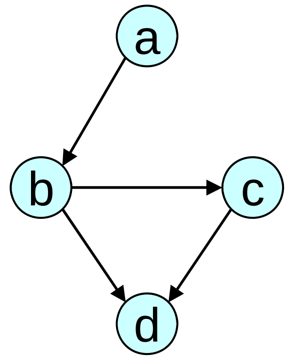
- Directed Graph
  - $G = (V, E)$: directed graph (有向グラフ)

    V: finite set -- nodes

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

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

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

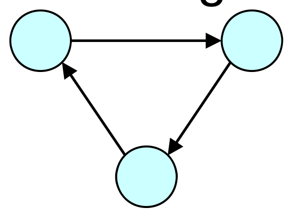


Orient the edge (a,b) by $a \rightarrow b$

  - Directed Acyclic graph (DAG, 非巡回有向グラフ)

    Directed graph with no cycles.

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



DAG and Probability

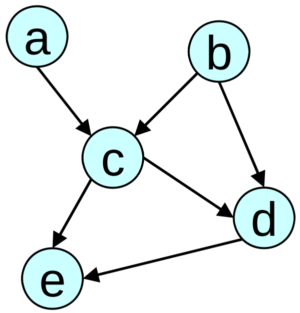
- Probability associated with a DAG
  - A DAG defines a family of probability distributions

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

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

p factorizes wr.t. DAG G.

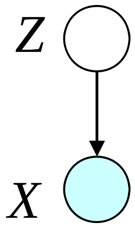
Example:



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

Practical Examples

- Finite mixture model

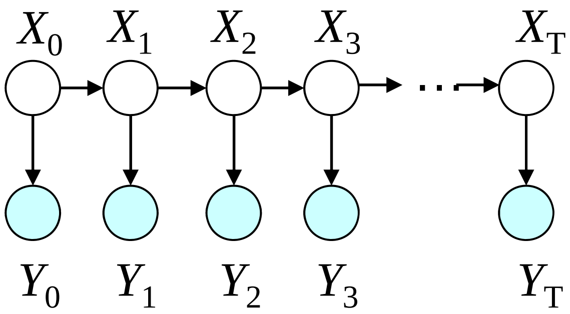


$Z \in$ Finite set

$$p(X,Z) = p(Z)p(X \mid Z)$$

$$p(X) = \sum_{Z} p(Z)p(X \mid Z)$$

- Hidden Markov model



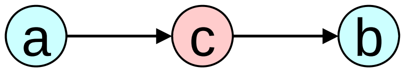
$$p(X,Y) = p(X_0)\prod_{t=1}^{T} p(X_t \mid X_{t-1})p(Y_t \mid X_t)$$

- Mixture model and HMM are discussed later.

Conditional Independence with DAG

- Three basic cases

(1)



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

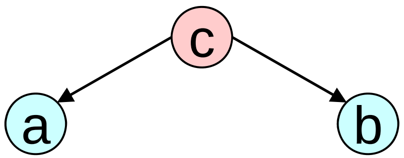
$$X_a \perp\!\!\!\perp X_b \mid X_c$$

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

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

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

(2)



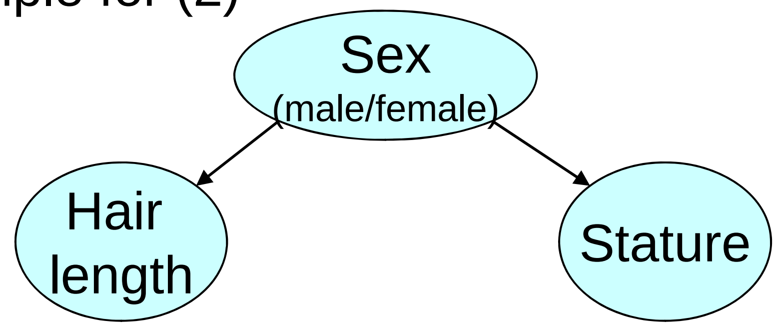
$$p(X_a, X_b, X_c) = p(X_c)\,p(X_a \mid X_c)\,p(X_b \mid X_c)$$

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

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

Conditional Independence with DAG

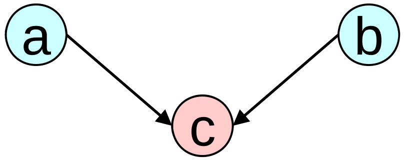
- Example for (2)



Sex (male/female) does influence both of hair length and stature, but given male (or female), hair length and stature are independent.

Conditional Independence with DAG

(3)

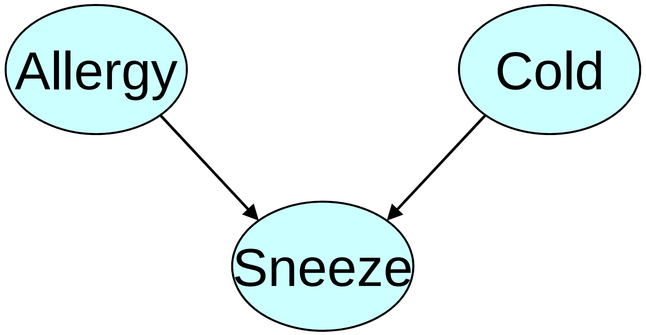


head-to-head
(or v-structure)

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

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

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

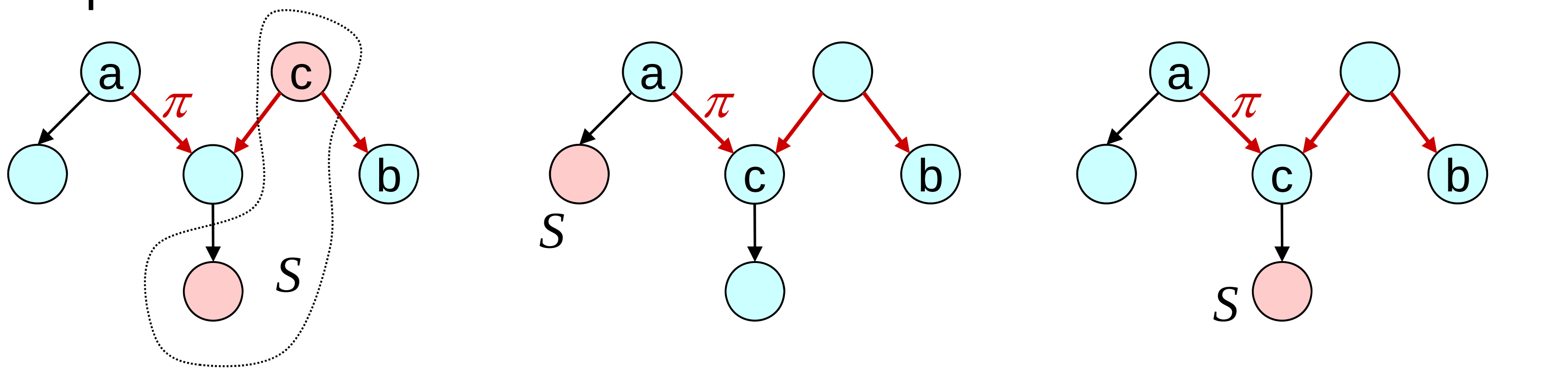
  (i) $c \in S$ and c is not head-to-head in π ($\bigcirc \rightarrow c \rightarrow \bigcirc$ or $\bigcirc \leftarrow c \rightarrow \bigcirc$),

  (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

Theorem 5

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

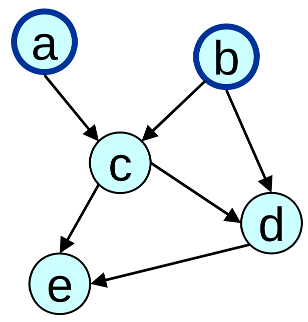
- Example
  - $X_a \perp\!\!\!\perp X_b$

    $S = \phi.$

    a → c ← b is blocked (with c).

    a → c → d ← b is blocked (with d)

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

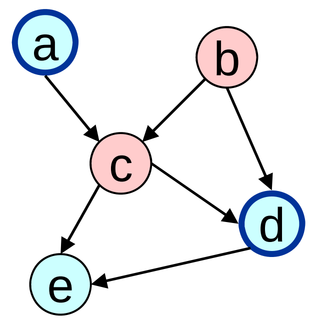


  - $X_a \perp\!\!\!\perp X_d \mid X_{\{b,c\}}$

    a → c → d is blocked (with c).

    a → c ← b ← d is blocked (with b)

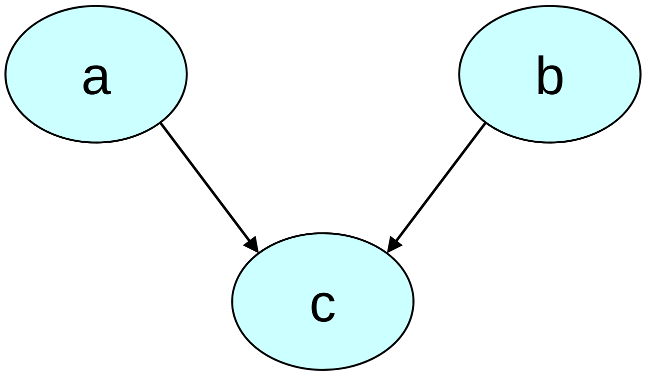
    a → c → e ← d is blocked (with e or c)

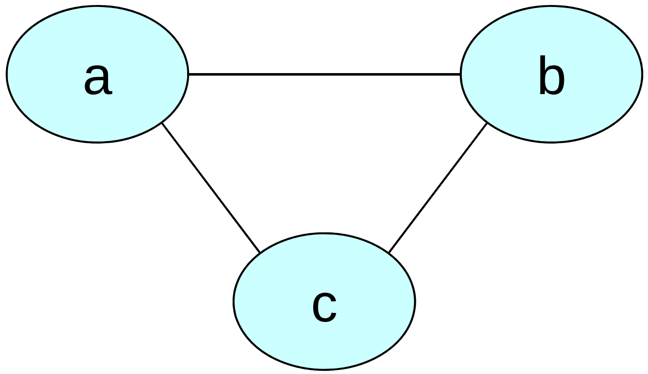


Comparison: UDG and DAG

- 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

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

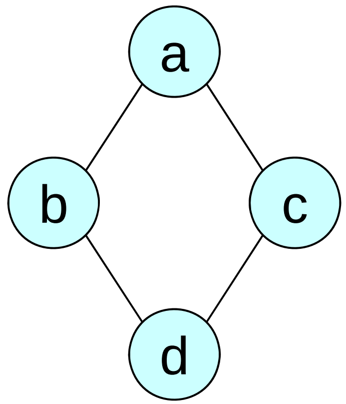
If $X_a \not\perp\!\!\!\perp X_c$, $X_b \not\perp\!\!\!\perp X_c$, $X_a \not\perp\!\!\!\perp X_b \mid X_c$, any UDG is not able to express $X_a \perp\!\!\!\perp X_b$.

Comparison: UDG and DAG

- Limitation of DAG

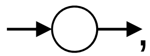
Undirected graph

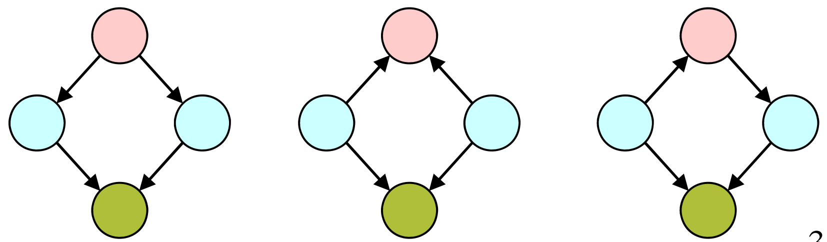
$$p(X_a, X_b, X_c, X_d)$$
$$= p(X_a, X_b)p(X_a, X_c)p(X_b, X_d)p(X_c, X_d)$$



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

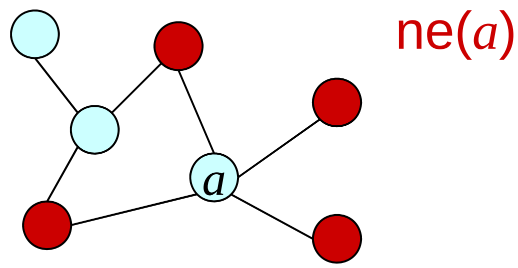
No DAG expresses these conditional independence relationships.

[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.

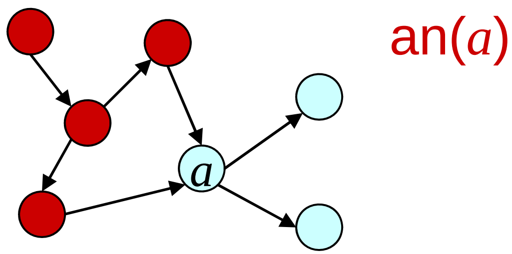


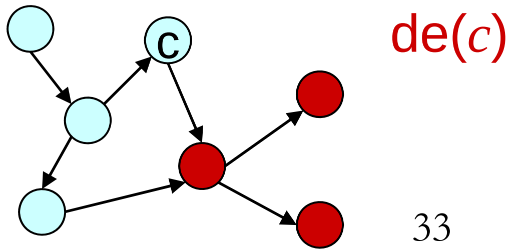
Appendix: Terminology on Graphs

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



- 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(先祖): $an(a) = \{b \in V \mid \exists \text{directed path from } b \text{ to } a\}$.
  - Descendents(子孫): $de(a) = \{b \in V \mid \exists \text{directed path from } a \text{ to } b\}$.



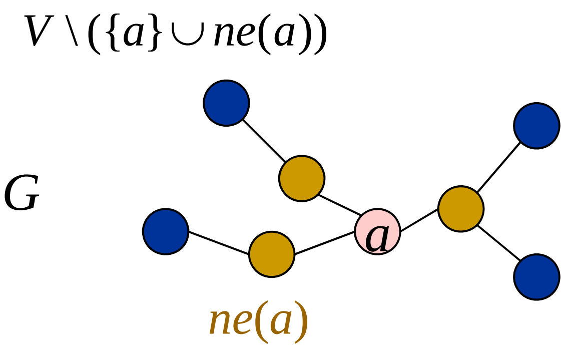


More on Markov Property

Markov Properties Revisited

- Markov properties for an undirected graph

  $G = (V, E)$: undirected graph.
  X: random vector indexed by V.

  - Local Markov

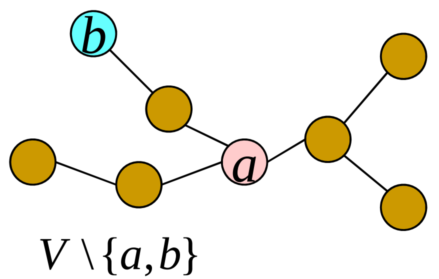
    X satisfies local Markov property relative to G if for any node a

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

  - 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

Theorem 7

Factorization $\Longrightarrow$ global Markov $\Longrightarrow$ local Markov $\Longrightarrow$ pairwise Markov

proof) factorization $\Longrightarrow$ global Markov : Theorem 3.

global Markov $\Longrightarrow$ local Markov : easy.

local Markov $\Longrightarrow$ 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

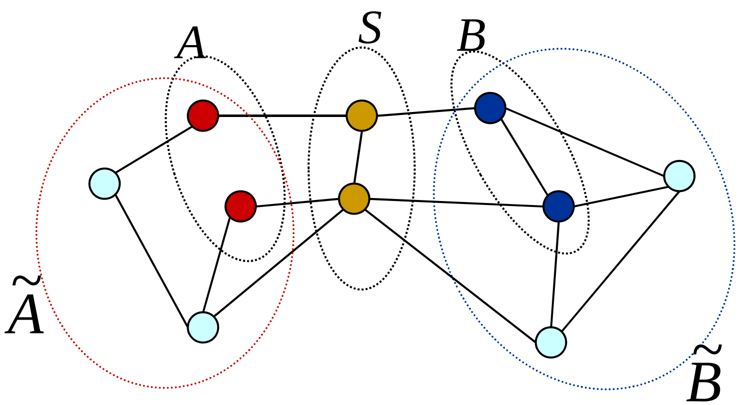
We show a slight generalization of Theorem 3.

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.

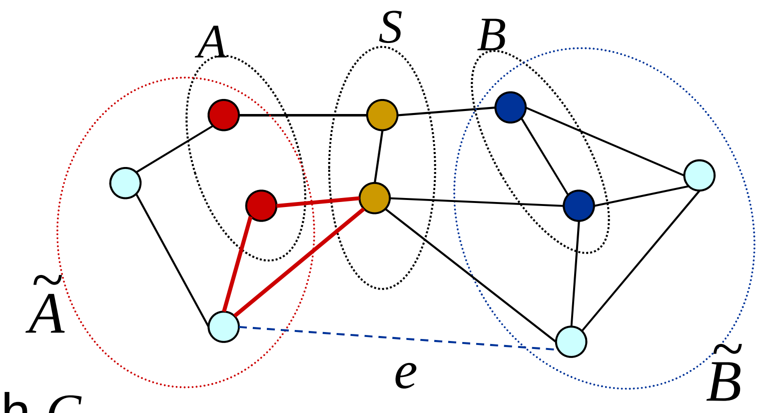


- Proof

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

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

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

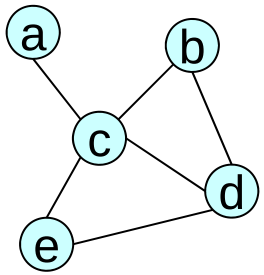
and thus

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

Q.E.D.

Mini Summary

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

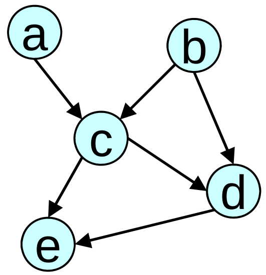
$\Rightarrow$

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