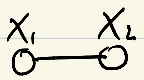
Probabilistic Graphical Model

- Bayesian Networks (Directed Graphical Models)
- Markov Random Fields (Undirected Graphical Models)
- Factor Graphs

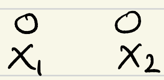
Graph Separation

Graph $\rightarrow$ Factorization $\rightarrow$ (Conditional) Independence

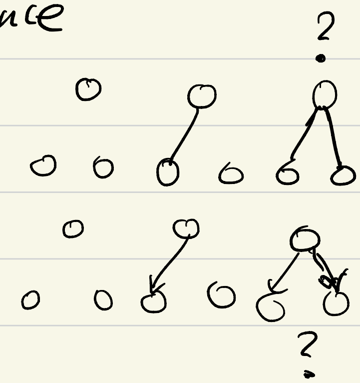
ex) $X=[x_1, x_2]$



$P(x_1, x_2)$



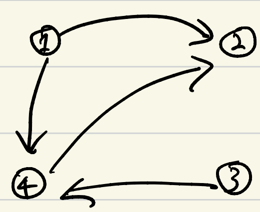
$P(x_1)P(x_2)$ $\quad x_1 \perp\!\!\!\perp x_2$



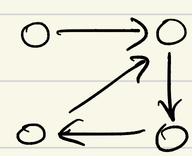
* Directed Graphical Models

Def. · Directed Acyclic Graph (DAG) $\quad G=(V,E)$

- set of nodes $V=\{1, \cdots, n\}$
- set of directed edges $E=\{(1,2), (4,2), (3,4), (1,4)\}$
  $(1,2) \neq (2,1)$



no cycle



cycle

- Parent set of i : $\pi_i = \{ j \in V \mid (j,i) \in E \}$
- Directed Graphical Model on $G=(V,E)$ is
  a family of probability distributions on $X=[x_1, x_2, \cdots, x_n]$
  that factorizes as

$$P(x_1, \cdots, x_n) = \prod_{i=1}^{n} f_i(x_i, \underset{\{x_j \mid j \in \pi_i\}}{\underset{\shortparallel}{x_{\pi_i}}})$$

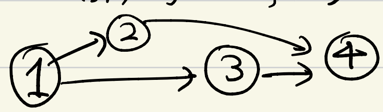
$f_i : \mathcal{X}^{|\pi_i|+1} \rightarrow \mathbb{R}_+$

s.t $\sum_{x_i} f_i(x_i, x_{\pi_i}) = 1.$

Claim: $f_i(X_i, X_{\pi_i}) = P(X_i \mid X_{\pi_i})$

without loss of generality, suppose $\boxed{\text{topological ordering}}$ of DAG $G=(V,E)$ which is $(1, 2, \cdots, n)$

ordering of nodes s.t. $(\vec{i,j}) \Rightarrow i<j$



$$P(X_1^n) = \prod_{i=1}^{n} f_i(X_i, X_{\pi_i})$$

$X_1^i \triangleq [X_1 \cdots X_i]$

$$P(X_1^{n-1}) = \sum_{X_n} P(X_1^n) \underset{\substack{\uparrow \\ \text{Topological ordering}}}{=} \prod_{i=1}^{n-1} f_i(X_i, X_{\pi_i}) \times \underbrace{\sum_{X_n} f_n(X_n, X_{\pi_n})}_{=1} = \prod_{i=1}^{n-1} f_i(X_i, X_{\pi_i})$$

$$f_n(X_n, X_{\pi_n}) = \frac{P(X_1^n)}{P(X_1^{n-1})} = P(X_n \mid X_1^{n-1}) = P(X_n \mid X_1^{n-1})$$

L.H.S is a function of only (X_n, X_{π_n})

by induction, follows for all $i \in [n] \triangleq \{1, 2, \cdots, n\}$

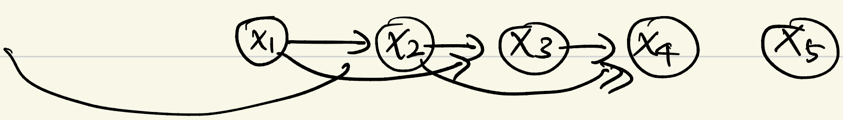
Directed Graphical Models encode Bayes rule (chain rule).

*Chain rule: for any $P(x)$ and any ordering $(1, 2, \cdots, n)$

$$P(x) = P(x_1) P(x_2 \mid x_1) P(x_3 \mid x_1, x_2) \cdots P(x_n \mid x_1^{n-1})$$

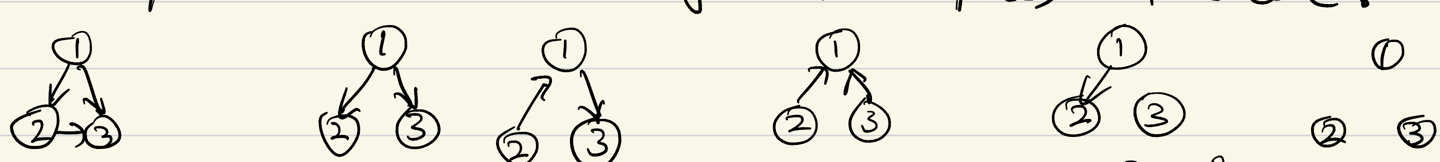
$\Rightarrow$ can be represented by a complete DAG.

all pairs connected



Q. what independence structure does general (incomplete) DAG encode?

ex>



$P(x_2)P(x_1 \mid x_2)P(x_3)$ $\quad$ $P(x_1)P(x_2 \mid x_1)P(x_3)$

$P(x_1)P(x_2 \mid x_1)P(x_3 \mid x_1, x_2)$ $\quad$ $P(x_1)P(x_2 \mid x_1)P(x_3 \mid x_1)$ $\quad$ $P(x_2)P(x_3)P(x_1 \mid x_2, x_3)$ $\quad$ $P(x_1)P(x_2)P(x_3)$

$(X_1, X_2) \perp\!\!\!\perp X_3$

ϕ $\quad$ $X_2 \perp\!\!\!\perp X_3 \mid X_1$ $\quad$ $X_2 \perp\!\!\!\perp X_3$ $\quad$ X_1, X_2, X_3 all independent of one another

complex & unstructured $\longleftrightarrow$ simple & structured

* Markov Property of Directed Graphical Models.

1. Directed ordered Markov Property

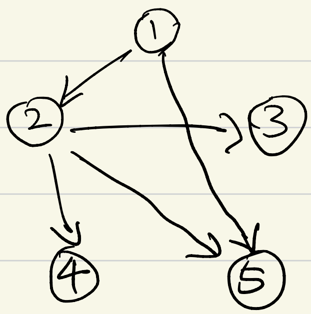
given a DAG $G=(V,E)$ with a topological ordering $(1,2,\cdots,n)$

$$X_i \perp\!\!\!\perp X_{pr_i \setminus \pi_i} \mid X_{\pi_i}$$

$\uparrow$ set minus

$Pr_i \triangleq$ nodes before i in the ordering.

proof> chain rule □



$(1,2,3,4,5)$ ordering

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

$(1,2,4,5,3)$ ordering

$X_3 \perp\!\!\!\perp (X_1, X_4, X_5) \mid X_2$

2. Directed Local Markov Property

$$X_i \perp\!\!\!\perp X_{nd_i \setminus \pi_i} \mid X_{\pi_i}$$

X_1 no independence.

X_2 ״

$X_3 \perp\!\!\!\perp (X_1, X_4, X_5) \mid X_2$

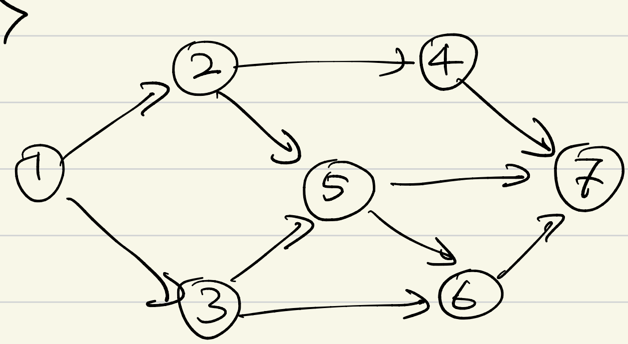
$X_4 \perp\!\!\!\perp (X_1, X_3, X_5) \mid X_2$

$X_5 \perp\!\!\!\perp (X_3, X_4) \mid (X_1, X_2)$

$nd_i \triangleq$ non-descendent

set of nodes not reachable from i by following directed paths.

ex>



more generally we want to answer if

$A = \{4\}$

$B = \{3, 5\}$

$C = \{6\}$

$X_A \perp\!\!\!\perp X_C \mid X_B$

3. Directed (Global) Markov Property.

$X_A \perp\!\!\!\perp X_C \mid X_B$ for any $A, B, C \subset V$

if A is d-separated from C w.r.t. B.

Def. $A \subseteq V$ is d-separated from $C \subseteq V$ w.r.t $B \subseteq V$ if every path between any nodes $a \in A$ and $c \in C$ is blocked

- Path : $a - p_1 - p_2 - \cdots - c$ that does not repeat nodes

vs. walk : sequence of connected nodes that can repeat.

- a path is blocked if somewhere along the path

Intuition : consider $V = \{1, 2, 3\}$

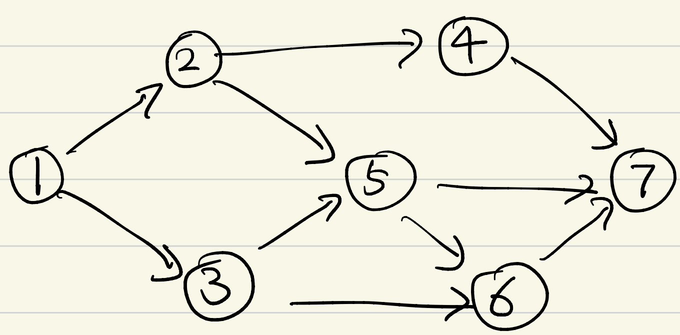
$\cdots\ ① \rightarrow ② \rightarrow ③ \cdots$ & $2 \in B$ $\qquad 1 \perp\!\!\!\perp 3 \mid 2$

$\cdots\ ① \leftarrow ② \leftarrow ③ \cdots$ & $2 \in B$ $\qquad 1 \perp\!\!\!\perp 3 \mid 2$

$\cdots\ ① \leftarrow ② \rightarrow ③ \cdots$ & $2 \in B$ $\qquad 1 \perp\!\!\!\perp 3 \mid 2$

$\cdots\ ① \rightarrow ② \leftarrow ③$ & $2 \notin B$ $\qquad 1 \perp\!\!\!\perp 3$ but $1 \not\perp\!\!\!\perp 3 \mid 2$

* Bayes ball algorithm identifies conditional independence $x_A \perp\!\!\!\perp x_C \mid x_B$



$A = \{4\}$

$B = \{3, 5\}$

$C = \{6\}$

Theorem. the following are equivalent.

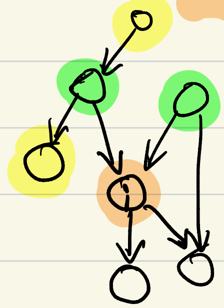
a> $P(x)$ factorizes according to a DAG $G=(V,E)$

b> $P(x)$ satisfies the directed global Markov property.

c> $P(x)$ satisfies the directed local Markov property

proof. b> $\Rightarrow$ c>

set $A = i$, $B = \pi_i$, $C = nd_i \setminus \pi_i$



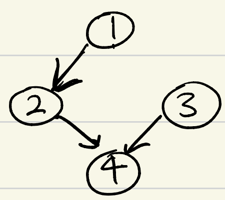
c> $\Rightarrow$ a>

Consider any topological ordering, say $(1, 2, \cdots, n)$

$$P(x) \underset{\substack{\uparrow \\ \text{chain rule}}}{=} \prod_i P_i(x_i \mid x_1^{i-1}) \underset{\substack{\uparrow \\ \text{local Markov}}}{=} \prod_i \underbrace{P_i(x_i \mid x_{\pi_i})}_{f_i(x_i,\, x_{\pi_i})}$$

a> $\Rightarrow$ b> use induction (separate notes).

let $I(G)$ denote the set of C.I. relations implied by G.

ex>  $I(G) = \{ X_3 \perp\!\!\!\perp (X_1, X_2),\ X_1 \perp\!\!\!\perp (X_3, X_4) \mid X_2 \}$

let $I(P)$ denote the set of C.I. relations implied by $P(x)$

ex> Consider indep Z_1, Z_2, $\sim \text{Bern}(1/3)$

let $X_1 = Z_1$, $X_3 = Z_3$, $X_2 = X_1$, $X_4 = X_2 \oplus X_3$

then $I(P) = \{ X_3 \perp\!\!\!\perp (X_1, X_2),\ X_1 \perp\!\!\!\perp (X_3, X_4) \mid X_2 \}$

remark 1. $\exists P$ s.t. there is no DAG G satisfying $I(P) = I(G)$

proof) $P = \begin{cases} X_1 = X_2 \\ X_3 = X_1 + Z \end{cases}$ $\quad I(P) = \begin{cases} X_1 \perp\!\!\!\perp X_3 \mid X_2 \\ X_2 \perp\!\!\!\perp X_3 \mid X_1 \end{cases}$

remark 2. $\forall G$, $\exists P$ s.t. $I(G) = I(P)$.

remark 3. $\exists G_1 \neq G_2$ s.t. $I(G_1) = I(G_2)$

$\circ \to \circ \to \circ \qquad \circ \leftarrow \circ \to \circ$.

Def. DAG $G = (V, E)$ is a <u>perfect map</u> of P if $I(P) = I(G)$.

Def. DAG $G = (V, E)$ is an <u>I-map</u> of P if $I(G) \subseteq I(P)$

Def DAG $G = (V, E)$ is a minimal I-map of P if removing a single edge from G results in G no longer being an I-map.

Remark 4. minimal I-map is not unique.

- different ordering gives different I-maps.
- even with the same ordering, I-map is not unique

$$\begin{tikzcd} 1 \arrow[r] & 2 \arrow[r] & 3 \end{tikzcd}$$

$$\begin{tikzcd} 1 \arrow[r] \arrow[rr, bend left] & 2 & 3 \end{tikzcd}$$

P: $X_1 = X_2$

$X_3 = X_1 + Z$

$I(P) = \{ X_3 \perp\!\!\!\perp X_1 \mid X_2,\ X_3 \perp\!\!\!\perp X_2 \mid X_1 \}$