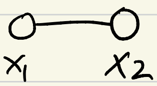
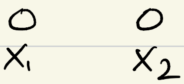
Probabilistic Graphical Model

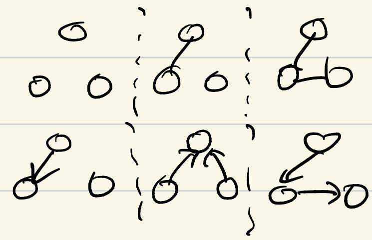
- Bayesian Nework (Directed G.M)
- Markov Random Field (Undirected G.M)
- Factor Graphs.

Graph Separation

Graph $\longrightarrow$ Factorization $\longrightarrow$ Conditional Independence

ex) $X=[X_1, X_2]$

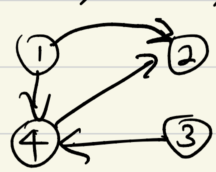
Graph	Factorization	Conditional Independence
	$P(x_1, x_2)$	none
	$P(x_1)P(x_2)$	$x_1 \perp\!\!\!\perp x_2$



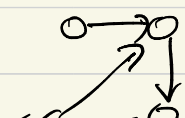
* $\boxed{\text{Directed Graphical Models}}$

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

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



no cycle



cycle

- Parent set of node i : $\pi_i \triangleq \{j \in V \mid (j,i) \in E\}$

- Directed Graphical Model on a DAG $G=(V,E)$ is
  a family of distributions on $X=[X_1, \dots, X_n]$
  that factorizes as

$$P(x_1, \dots, x_n) = \prod_{i=1}^{n} f_i(x_i, x_{\pi_i}) \qquad f_i : \mathcal{X}^{|\pi_i|+1} \to \mathbb{R}_+$$

$$\text{s.t. } \sum_{x_i \in \mathcal{X}} f_i(x_i, x_{\pi_i}) = 1 \qquad x_{\pi_i} \triangleq \{x_j \mid j \in \pi_i\}$$

$P(\cdot \mid x_{\pi_i})$: PMF

Claim: $f_i(x_i, x_{\pi_i}) = P(x_i \mid x_{\pi_i})$, I could of said

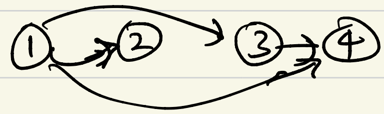
Z_i is a disc. R.V

$\mathbb{P}_{Z_i}(z_i) = \mathbb{P}(Z_i = z_i)$

w.l.o.g.

suppose that $G=(V,E)$ has a topological ordering of $(1,2,\cdots,n)$.

$\triangleq \overrightarrow{(i,j)} \Rightarrow i < j$



(1,2,3,4)

(1,3,2,4)

(1) $P(x_1^n) = \prod_{i=1}^{n} f_i(x_i, x_{\pi_i})$

(2) $P(x_1^{n-1}) = \sum_{x_n \in \mathcal{X}} P(x_1^n) = \underbrace{\left(\sum_{x_n \in \mathcal{X}} f_n(x_n, x_{\pi_n})\right)}_{=1} \times \prod_{i=1}^{n-1} f_i(x_i, x_{\pi_i}) = \prod_{i=1}^{n-1} f_i(x_i, x_{\pi_i})$

Topological ordering

$f_n(x_n, x_{\pi_n}) = \frac{P(x_1^n)}{P(x_1^{n-1})} \triangleq P(x_n \mid x_1^{n-1}) \overset{?}{=} \underline{P(x_n \mid x_{\pi_n})}$

depends only on x_n, x_{π_n}

by induction for rest. $\square$

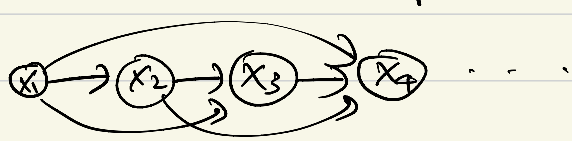
Directed Graphical Models encode Bayes rule (=chain rule)

- chain rule : for any $P(x)$ & any ordering $(1,\cdots,n)$

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

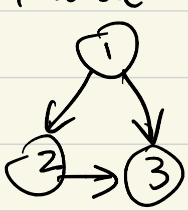
$\Rightarrow$ any $P(x)$ can be represented by a complete DAG

all pairs connected

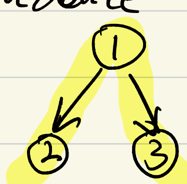


Q. What independence does a DAG (not complete) encode?

ex>

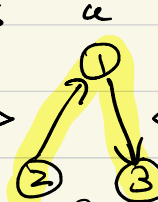


$P(x_1)P(x_2 \mid x_1)P(x_3 \mid x_1, x_2)$



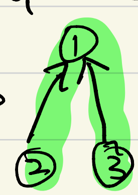
$P(x_1)P(x_2 \mid x_1)P(x_3 \mid x_1)$

$\Leftrightarrow$

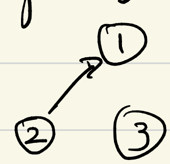


$P(x_2)P(x_1 \mid x_2)P(x_3 \mid x_1)$

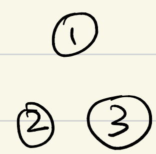
$\not\Leftrightarrow$



$\underline{P(x_2)P(x_3)P(x_1 \mid x_2, x_3)}$



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



$P(x_1)P(x_2)P(x_3)$

dense $\emptyset$

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

$X_2 \perp\!\!\!\perp X_3$

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

all mutually indep.

complex

unstructured $\longleftrightarrow$ simple sparse structured

* Markov Property of Directed Graphical Models.

1. Directed (ordered) Markov Property

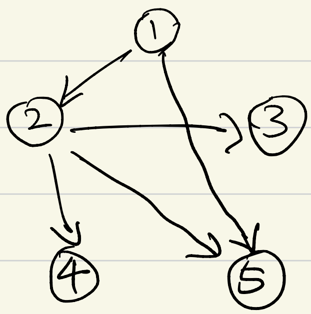
given a DAG $G=(V,E)$, and $(1,2,\cdots,n)$

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

set minus

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

proof> Chainrule □.



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

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

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

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

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

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

⋮

2. Directed (Local) Markov Property

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

, $nd_i \triangleq$ non-descendents
set of nodes not reachable from i by following the direction paths

$X_1 \quad \phi$

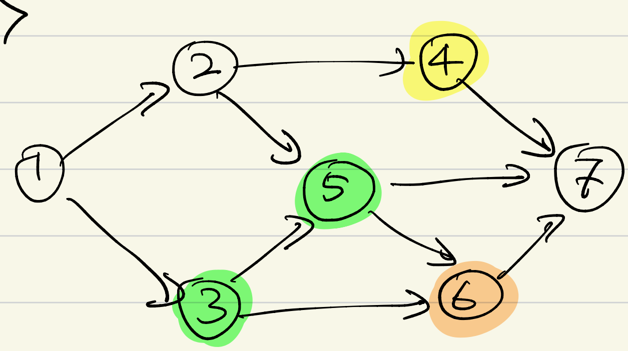
$X_2 \quad \phi$

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

$0,1 \leftarrow$ PROGRAM(A, B, C, G)
Bayes Ball Algorithm

more generally we want to answer if

ex>



$A = \{4\}$

$B = \{3, 5\}$

$C = \{6\}$

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

3. Directed Global Markov Property

for any $A, B, C \subset V$

$X_A \perp\!\!\!\perp X_C \mid X_B$ if A is d-separated from C w.r.t B.

Def. $A \subset V$ is d-separated from C w.r.t B if every path between any node $a \in A$ & $c \in C$ is blocked (by B).

· Path : $a - P_1 - P_2 - \cdots - P_k - c$ that does not repeat any node.
$\quad$ either $(a, P_1) \in E$ or $(P_1, a) \in E$

· Walk : can repeat nodes.

· a path is blocked if somewhere in the path $\exists$ 3 nodes s.t.

Block

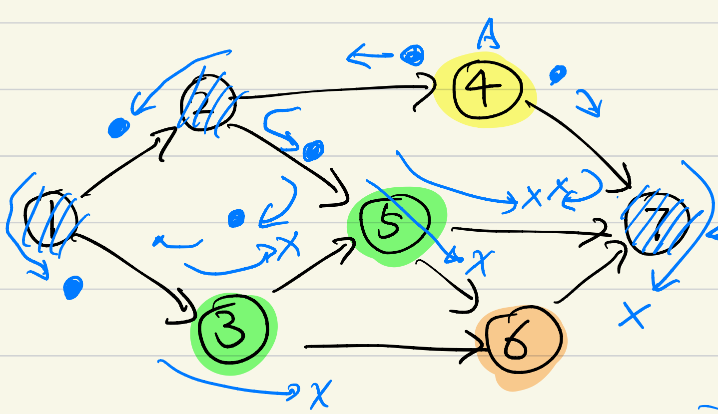
$a \cdots 1 \to 2 \to 3 \cdots c$ & $2 \in B$ $\quad$ Intuition: $X_1 \perp\!\!\!\perp X_3 \mid X_2$

$\cdots 1 \leftarrow 2 \leftarrow 3 \cdots$ & $2 \in B$ $\quad$ $X_1 \perp\!\!\!\perp X_3 \mid X_2$

$\cdots 1 \leftarrow 2 \to 3 \cdots$ & $2 \in B$ $\quad$ $X_1 \perp\!\!\!\perp X_3 \mid X_2$

$\cdots 1 \to 2 \leftarrow 3 \cdots$ & $2 \notin B$ $\quad$ $X_1 \perp\!\!\!\perp X_3$

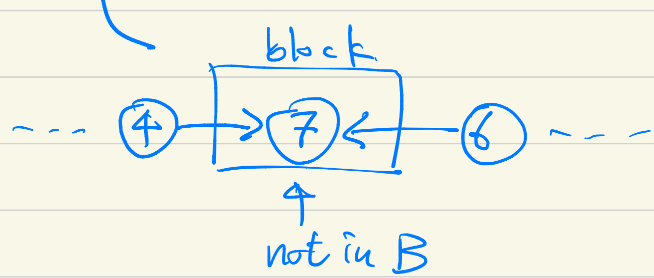
* Bayes ball algorithm identifies conditional independence $X_A \perp\!\!\!\perp X_C \mid X_B$



$A = \{4\}$

$B = \{3, 5\}$

$C = \{6\}$



Theorem. the following are equivalent.

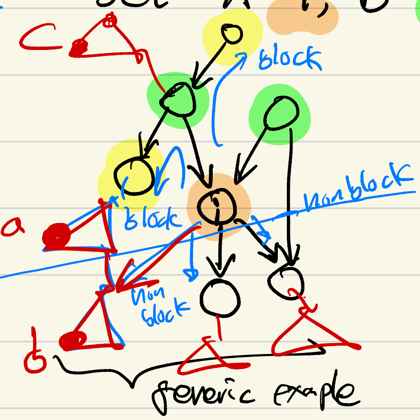
a> $P(x)$ factorizes according to a DAG $G=(V,E)$

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

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

proof. b> $\Rightarrow$ c>

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



$c \perp\!\!\!\perp i \mid \pi_i$

$a \perp\!\!\!\perp i \mid \pi_i$

$b \not\!\perp\!\!\!\perp i \mid \pi_i$

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

directed ordered M.P
only has ordering

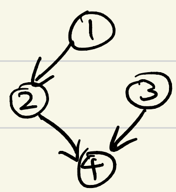
c> $\Rightarrow$ a> any topological ordering. $(1, \cdots, n)$

$$P(x) \underset{\text{chain}}{=} \prod_i P_i(x_i \mid x_1^{i-1}) \underset{\text{local MP}}{=} \prod_i \overbrace{P_i(x_i \mid x_{\pi_i})}^{f_i(x_i, x_{\pi_i})}$$

separate PDF.

$\boxed{\text{a>} \Rightarrow \text{b>}}$ uses induction (separate note)

let $I(G)$ is the set of all C.I implied by $G=(V,E)$



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

let $I(P)$ is the set of all C.I. implied by the factorization of $P(x)$

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

let $X_1 = Z_1$, $X_2 = Z_1$, $X_3 = Z_2$, $X_4 = X_2 \underset{\text{XOR}}{\oplus} X_3$ $\Big\} \Rightarrow P(x_1^4)$

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

R1. $\exists P(X)$ s.t. no DAG G satisfy $I(G) = I(P)$

proof> $$\left.\begin{array}{l} X_1 = X_2 \\ X_3 = X_1 + Z \end{array}\right. \Rightarrow \left.\begin{array}{l} X_1 \perp\!\!\!\perp X_3 \mid X_2 \\ X_2 \perp\!\!\!\perp X_3 \mid X_1 \end{array}\right\} I(P)$$

R2. $\forall G$, $\exists P$ s.t $I(P) = I(G)$.

R3. $\exists G_1 \neq G_2$ s.t. $I(G_1) = I(G_2)$

$$\circ \rightarrow \circ \rightarrow \circ \qquad \circ \leftarrow \circ \rightarrow \circ$$

Def. G is a <u>perfect map</u> for $P(X)$ if $I(P) = I(G)$

Def. G is an <u>I-map</u> for $P(X)$ if $I(G) \subseteq I(P)$

Def. G is a <u>minimal I-map</u> of $P(X)$ if remove any edge in G resules in G no longer beig an I-map

R4. minimal <u>I-map</u> is not unique.

- different ordering of variables
- even withe the same ordering, minimal I-map is not unique.

proof> $X_1 = X_2$
$X_3 = X_1 + Z$

$$1 \rightarrow 2 \rightarrow 3$$

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