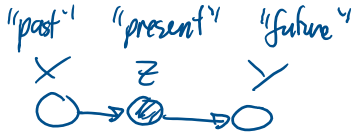
today: DGM (cont.)
UGM

conditional independence in DGM

basic 3 node graphs

1) Markov chain



here $\underset{\text{future}}{Y} \perp\!\!\!\perp \underset{\text{past}}{X} \mid \underset{\text{present}}{Z}$

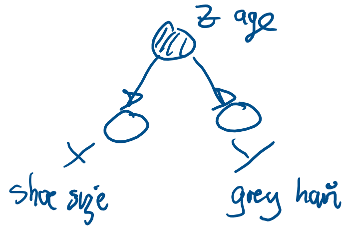
$p(x,y,z) = p(x)\,p(z|x)\,p(y|z)$

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

but $X \not\perp\!\!\!\perp Y$ (for some $p \in \mathcal{L}(G)$)

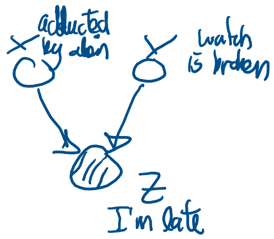
2) latent cause

(hidden variable)
↓
Z is not observed



statements same as before

3) explaining away
competing effect
(v structure)



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

but $X \not\perp\!\!\!\perp Y \mid Z$ (for some $p \in \mathcal{L}(G)$)

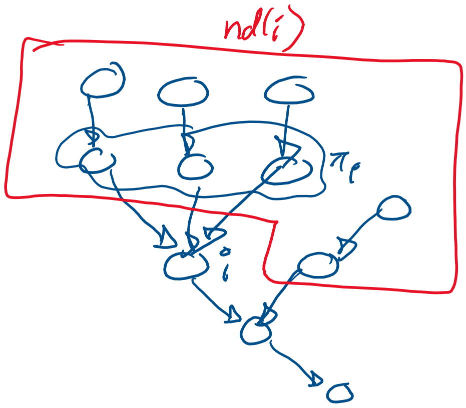
non-monotonic property of conditioning

p(alien) tiny

p(alien | late) > p(alien)

p(alien | late, watch broken) < p(alien | late)

more conditional indep. statements in a DGM



let $nd(i) \triangleq \{ j \neq i : \text{no path } i \rightsquigarrow j \}$

"non-descendants" of i

prop: $\boxed{p \in \mathcal{L}(G) \Longleftrightarrow X_i \perp\!\!\!\perp X_{nd(i)} \mid X_{\pi_i} \quad \forall i}$

(by decomposition: $\Rightarrow X_i \perp\!\!\!\perp X_j \mid X_{\pi_i} \;\; \forall j \in nd(i)$)

proof:

$\Rightarrow$) let i be fixed

key pt.: then $\exists$ top. ordering of G s.t. $nd(i)$ are exactly before i

i.e. $(nd(i), i, \underbrace{\text{descendents}(i)}_{\text{pluck all these?}})$

$$p(x_i, x_{nd(i)}) = p(x_i | x_{\pi_i}) \prod_{j \in nd(i)} p(x_j | x_{\pi_j})$$

$$p(x_i | x_{nd(i)}) = \frac{p(x_i, x_{nd(i)})}{p(x_{nd(i)})} = \frac{p(x_i|x_{\pi_i}) \prod_{j \in nd(i)} \cancel{p(x_j|x_{\pi_j})}}{\underbrace{\left(\sum_{x_i'} p(x_i'|x_{\pi_i})\right)}_{1} \underbrace{\prod_j \cancel{\;}}_{\text{no } x_i' \text{ here}}} = p(x_i | x_{\pi_i}) \quad //$$

$\pi_i \subseteq nd(i)$

$\Leftarrow$) (suppose p satisfies all these cond. indep. statements)

let $1{:}n$ be (WLOG) a top. sort of G

then have $1{:}i-1 \subseteq nd(i)$ $\forall i$ [by top. sort property]

$\Rightarrow X_i \perp\!\!\!\perp X_{1:i-1} | X_{\pi_i}$ by decomposition $\quad \left[X_i \perp\!\!\!\perp (X_{1:i-1 \setminus \pi_i}, X_{\pi_i}) | X_{\pi_i} \right.$

$$p(x_V) = \prod_{i=1}^{n} p(x_i | x_{1:i-1}) \quad \text{(by chain rule)}$$

$$= \prod_{i=1}^{n} p(x_i | x_{\pi_i}) \quad \text{(by cond. indep.)}$$

$\Rightarrow p \in \mathcal{L}(G)$ //

other cond. indep. statements?

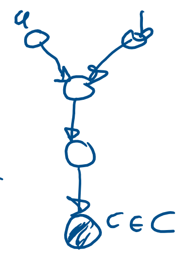
chain from a to b: just any (undirected) path in graph from a to b

d-separation:

def: set A & B are said to be d-separated by C (in G) $\leftarrow$ set of R.V. condition on

iff all chains from $a \in A$ to $b \in B$ are "blocked" given C

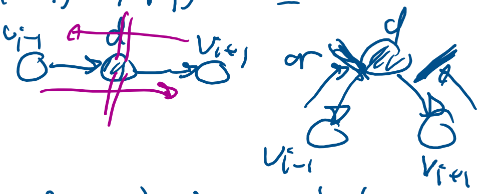
(intuition: do not want to have this situation



where a chain from a to b is "blocked" given C at a node d

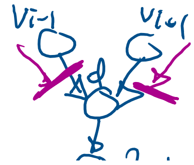
if a) either $d \in C$ and (v_{i-1}, d, v_{i+1}) is not a v-structure $\leftarrow$ subpath of chain

(standard separation) (e.g.

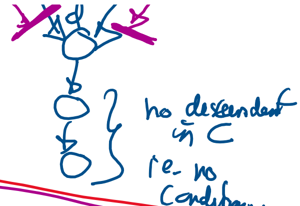


b) $d \notin C$ and (v_{i-1}, d, v_{i+1}) is a v-structure

and no descendent of d is in C



and no descendent of d is in C



prop: $p \in \mathcal{L}(G) \iff X_A \perp\!\!\!\perp X_B \mid X_C \quad \forall A,B,C \subseteq V$ s.t. A & B are d-separated given C

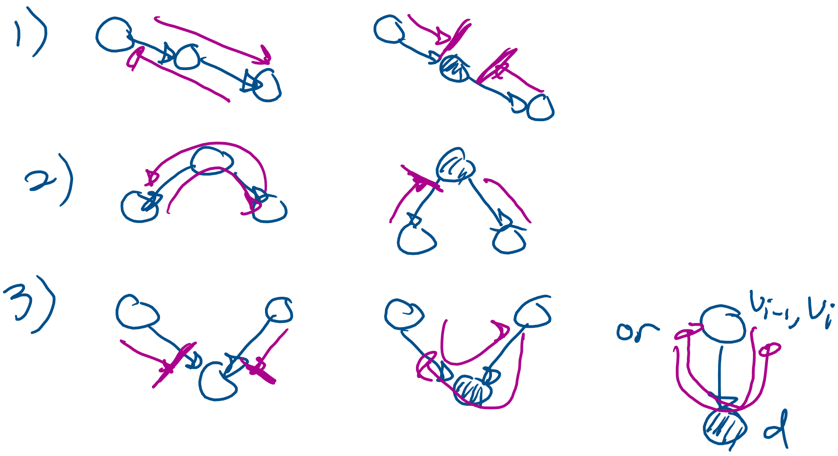
{ this covers all cond. indep. statements which are true $\forall p \in \mathcal{L}(G)$ }

"Bayes ball" alg. "intuitive" alg to check d-separation

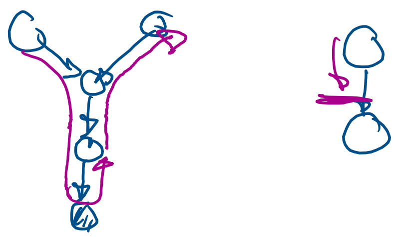
rules balls/chains being blocked

a) in d-separation

b)



see Alg. 3.1 in K&F book



15h30

more properties of DGM

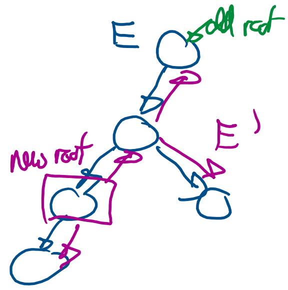
inclusion: $E \subseteq E' \Rightarrow \mathcal{L}(G) \subseteq \mathcal{L}(G')$

reversal: if G is a directed tree (or a forest) (there is no v-structure)

$\hookrightarrow |\pi_i| \leq 1 \ \forall i$

let E' be another directed tree (with same undirected edges) by choosing a different root

$\Rightarrow \mathcal{L}(G) = \mathcal{L}(G')$

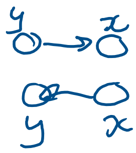


rephrasing: all directed trees from an undirected tree give same DGM

(*) why direction of edges are not causal in DGM

undirected tree give same DGM

the not causal in DGM



marginalization :

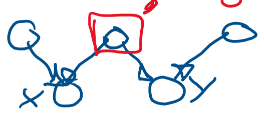
- marginalize a leaf node n gives a smaller DGM

  let $S = \{ q$ is a dist. on $x_{1:n-1}$ s.t. $q(x_{1:n-1}) = \sum_{x_n} p(x_{1:n})$ for some $p \in \mathcal{L}(G) \}$

  then $S = \mathcal{L}(G')$ where G' is G with leaf n plucked (removed)

- not true for all marginalizations

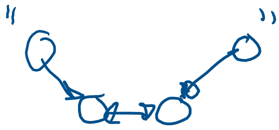
  e.g.



marginalize this node

↳ get a set of dist. $\neq$ any $\mathcal{L}(G')$ for any G'

i.e. marginalization is not a "closed" operation on DGMs



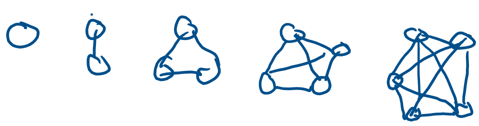
undirected GM (UGM) (aka. Markov random field or Markov network)

let $G = (V, E)$ be an undirected graph

let $\mathcal{C}$ be the set of all cliques of G

clique = set of nodes which are fully connected

$(\text{if } C \in \mathcal{C} \Leftrightarrow \forall i, j \in C, i \neq j, \{i, j\} \in E)$



UGM associated with G

$$\mathcal{L}(G) \triangleq \{ p \text{ is a dist. over } X_V : p(x_V) = \frac{1}{Z} \prod_{C \in \mathcal{C}} \psi_C(x_C) \}$$

for some "potentials" $\psi_C : \Omega_{X_C} \to \mathbb{R}^+$

and $Z \triangleq \sum_{x_V} \left(\prod_{C \in \mathcal{C}} \psi_C(x_C) \right)$

normalization constant

"partition fct."

notes: → where could think of C to (i, π_i) : $\psi_C(x_C) = q(x_i | x_{\pi_i})$

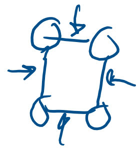
notes:

- unlike a DGM → where could think of C to $\{i, \pi_i\}$; $\psi_C(x_C) = p(x_i | x_{\pi_i})$

  $\psi_C(x_C)$ is **not directly related** $p(x_C)$

- can rescale any potential without changing joint p i.e. $\psi_C^{new}(x_C) \triangleq \psi_C^{old}(x_C) \cdot A$ for some $A > 0$

  (unlike in DGM, where f_i's were uniquely defined by p)

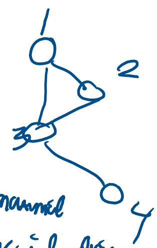


- it is sufficient to consider $\mathcal{C}_{max}$, the set of maximal cliques

  e.g. $C' \subseteq C$

  redefine $\psi_C^{new}(x_C) \triangleq \psi_C^{old}(x_C) \cdot \psi_{C'}^{old}(x_{C'})$

  (if C' belongs to more than one C, just pick one)



$\{1,2\}$ is not maximal

$\subset$

$\{1,2,3\}$ is maximal clique

[we'll see later, sometimes convenient to consider]

"over-parameterization" of trees using $\psi_{ij}(x_i, x_j)$ $\psi_i(x_i)$

properties of UGM

- as before, $E \subseteq E' \Rightarrow \mathcal{L}(G) \subseteq \mathcal{L}(G')$

  $E = \emptyset \Rightarrow \mathcal{L}(G) =$ set of fully factorized dist.

  $E =$ all pairs (i.e. V one big clique in G) $\Rightarrow \mathcal{L}(G) =$ all of dist. on x_V

- if $\psi_C(x_C) > 0 \quad \forall x_C, \forall C$

  then can write $p(x_V) = \exp\Big(\sum_{C \in \mathcal{C}} \overbrace{\log \psi_C(x_C)}^{\langle \theta_C, T_C(x_C) \rangle} - \log Z\Big)$

  physics ~~link~~ negative energy fct

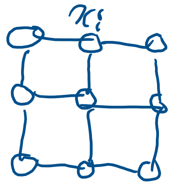
  $T_C(x_C) = \begin{pmatrix} 0 \\ 1 \\ 0 \\ 0 \end{pmatrix} \leftarrow$ "x_C"

  $\theta_C(x_C) \triangleq \log \psi_C(x_C)$

  $\dim T_C = |\Omega_{x_C}|$

e.g. Ising model in physics

e.g. Ising model in physics



$x_i \in \{0,1\}$

node potential $E_i = -\log \psi_i(x_i = 1)$

edge potential $E_{ij} = -\log \psi_{ij}(x_i = 1, x_j = 1)$