

today: directed graphical model

note: for simplicity, consider discrete R.V. first

conditional distribution concept is tricky to formalize

see Borel-Kolmogorov paradox

https://en.wikipedia.org/wiki/Borel%E2%80%93Kolmogorov_paradox

Directed graphical model

Let $G=(V,E)$ be a DAG

a directed graphical model (DGM) (associated with G) (aka Bayesian network)

is a family of distributions over X_V

here, $\mathcal{L}(G) \triangleq \left\{ p \text{ is a dist. over } X_V : \begin{array}{l} \text{(not necessarily unique)} \\ \exists \text{ local factors } f_i \text{'s s.t.} \\ p(x_V) = \prod_{i=1}^n f_i(x_i | x_{\pi_i}) \end{array} \right\}$

for potentials f_i s.t. 1) $f_i \geq 0$

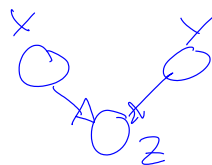
$f_i: \Omega_{x_i} \times \Omega_{x_{\pi_i}} \rightarrow [0,1]$

2) $\forall i, \sum_{x_i} f_i(x_i | x_{\pi_i}) = 1 \quad \forall x_{\pi_i}$

f_i is like a CPT
(conditional probability table)

terminology is: if can write $p(x_V) = \prod_{i=1}^n f_i(x_i | x_{\pi_i})$ this is where G comes into play

then we say that " p factorizes according to G "

one example:  $p \in \mathcal{S}(G) \Leftrightarrow \exists f_x, f_y, f_z$
 $p(x, y, z) = f_x(x) f_y(y) f_z(z|x, y)$

"leaf plucking" property fundamental property of DGM

let n be a leaf in G (i.e. n is not a parent of anything)

then $p(x_{1:n-1}) \in \mathcal{S}(G - \{n\})$
remove n from G

$$\text{if } p(x_{1:n}) = \prod_{i=1}^n f_i(x_i | x_{\pi_i}) \Rightarrow p(x_{1:n-1}) = \prod_{i=1}^{n-1} f_i(x_i | x_{\pi_i})$$

proof:

$$p(x_n, x_{1:n-1}) = f_n(x_n | x_{\pi_n}) \prod_{j \neq n} f_j(x_j | x_{\pi_j})$$

no n in any π_j

$$p(x_{1:n-1}) = \sum_{x_n} p(x_n, x_{1:n-1})$$

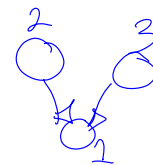
$$= \left(\sum_{x_n} f_n(x_n | x_{\pi_n}) \right) \left(\prod_{j \neq n} f_j(x_j | x_{\pi_j}) \right)$$

1 by definition

proposition: if $p \in \mathcal{S}(G)$;

let $\{f_j\}$ be a factorization w.r. to G for p

$$\text{then } \forall i \quad p(x_i | x_{\pi_i}) = f_i(x_i | x_{\pi_i})$$



proof: w.r. to n not Σ_1 n is a leaf node

2 3 1
~~2~~ ~~3~~ ~~1~~
 1 2 3

proof: wlog, let $\{1, \dots, n\}$ be a top sort

let i be fixed; we want to prove $p(x_i | x_{\pi_i}) = f_i(x_i | x_{\pi_i})$

$$n \text{ is a leaf} \rightarrow \text{pluck it to get } p(x_{1:n-1}) = \prod_{j=1}^{n-1} f_j(x_j | x_{\pi_j})$$

factorization for a DAG where x_{n-1} is now a leaf

pluck $n-1$ as well?

keep doing this [in reverse order $n-1, n-2, \dots, i+1$] to get

$$p(x_{1:i}) = f_i(x_i | x_{\pi_i}) \prod_{j < i} f_j(x_j | x_{\pi_j}) \stackrel{\triangle}{=} g(x_{1:i-1})$$

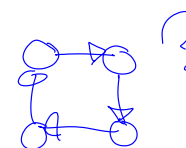
partition $1:i$ as $\{i\} \cup \pi_i \cup A$ [i.e. $A = 1:i-1 \setminus \pi_i$]

$$p(x_{1:i}) = p(x_i, x_{\pi_i}, x_A)$$

$$p(x_i | x_{\pi_i}) = \frac{\sum_{x_A} p(x_i, x_{\pi_i}, x_A)}{\sum_{x_A} \sum_{x_i'} p(x_i', x_{\pi_i}, x_A)} = \frac{f_i(x_i | x_{\pi_i}) \sum_{x_A} g(x_{1:i-1})}{\sum_{x_i'} \underbrace{f_i(x_i' | x_{\pi_i})}_{=1} \sum_{x_A} g(x_{1:i-1})} = f_i(x_i | x_{\pi_i}) //$$

* equivalently, now we can write $\mathcal{I}(G) = \{ p : p(x_v) = \prod_{i=1}^n p(x_i | x_{\pi_i}) \}$

note: adding edges \Rightarrow more distributions



i.e. $G=(V,E)$ & $G'=(V,E')$ with $E' \supseteq E$
 then $\mathcal{I}(G) \subseteq \mathcal{I}(G')$

examples:

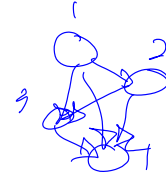
• $E=\emptyset$
 (trivial graph)

$\Rightarrow \mathcal{I}(G)$ contains only fully independent distributions
 i.e. $p(x_V) = \prod_{i=1}^n p(x_i)$

• complete digraph

for all $i \neq j$
 either $(i,j) \in E$
 or $(j,i) \in E$

+DAG
 : get $\text{wlog } p(x_V) = \prod_{i=1}^n p(x_i | x_{i \text{ parents}})$
 (like chain rule)



\Rightarrow all distributions are in $\mathcal{I}(G_{\text{complete}})$

⊛ removing edges impose more factorization constraints (and thus conditional indep. assumptions)

basic 3 nodes graphs

factorization w/ cond. indep. assumption

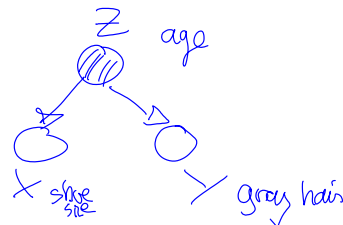
1) Markov chain:



here: "future" "past" "present"
 $X \perp\!\!\!\perp Y \mid Z$

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

2) latent cause (hidden variable)



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

but $X \not\perp\!\!\!\perp Y$

3) explaining away



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

(V-structure) ^{common cause}



but ~~\perp~~ ~~$\mid Z$~~

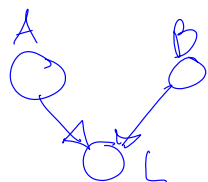
→ non-monotonic property of conditioning:

$p(\text{alien})$ is tiny

$p(\text{alien} \mid \text{late}) > p(\text{alien})$

$p(\text{alien} \mid \text{late, broken watch}) < p(\text{alien} \mid \text{late})$

concrete CPTs to see this



$A \rightarrow \text{alien?}$
 $B \rightarrow \text{broken watch?}$
 $L \rightarrow \text{late?}$

$$p(a, b, l) = p(l \mid a, b) p(a) p(b)$$

CPT's:

	$p(a)$
$A=1$	ϵ
$A=0$	$1-\epsilon$

	$p(b)$
$B=1$	0.1
$B=0$	0.9

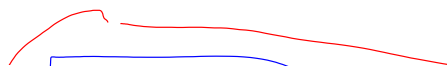
A	B	$p(L=1 \mid a, b)$
1	1	1
1	0	1
0	1	0.8
0	0	0.1

exercise. $P(A=1) = \epsilon$

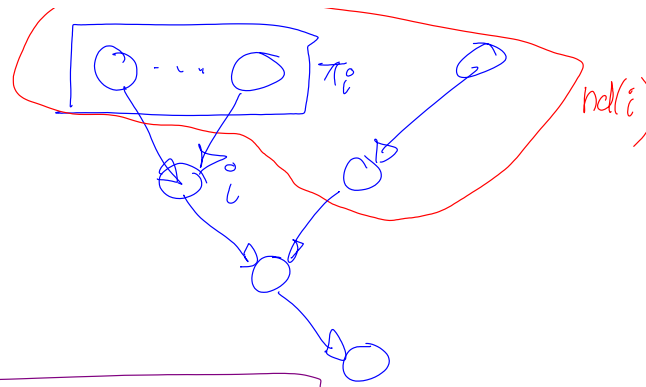
$$P(A=1 \mid L=1) = \frac{\epsilon}{0.17 + 0.83\epsilon} > \epsilon \quad (\text{if } \epsilon < 1)$$

$$P(A=1 \mid L=1, B=1) = \frac{\epsilon}{0.90 + 0.10\epsilon} < \frac{\epsilon}{0.17 + 0.83\epsilon}$$

cond. indep. statements in DGM:



let $nd(i) \triangleq \{j : \text{no path } i \rightarrow j\}$
 "non-descendants" of i



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

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

proof:
 \Rightarrow

key point: let i be fixed

then \exists a tp ordering s.t. $nd(i)$ are exactly before?

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

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

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

1 no x_i here

\Leftarrow (suppose that p satisfies cond. indep. statements)
 let $1:n$ be a top sort

then $1:i-1 \subseteq \text{nd}(i) \forall i$

$\Rightarrow X_i \perp\!\!\!\perp X_{1:i-1} \mid X_{\pi_i}$ (by decomposition)

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

$$= \prod_{i=1}^n p(x_i \mid x_{\pi_i}) \quad (\text{by C.I.})$$

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

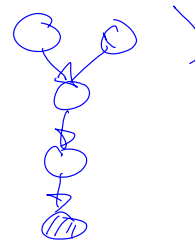
other C.I. statements?

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

d-separation : (intuition, is do not want this situation

sets A & B are said to be d-separated by C

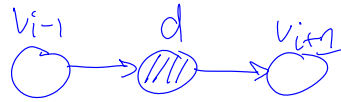
set of observed /
 conditioned R.V.s



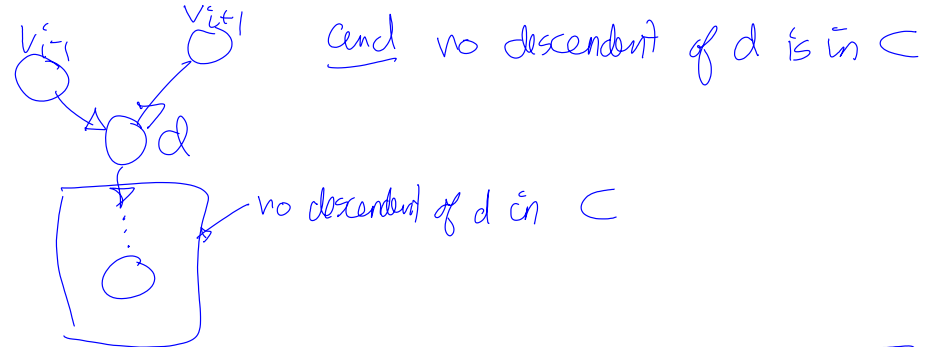
iff all chains from $a \in A$ to $b \in B$ are "blocked" given C

where a chain from a to b is blocked at node d \nwarrow subpath of chain

if a) either $d \in C$ and (v_{i-1}, d, v_{i+1}) is not a V-structure



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



prop:

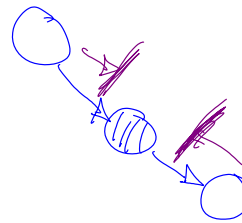
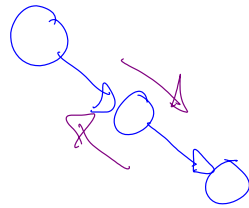
$$\text{if } p \in \mathcal{P}(G) \Leftrightarrow X_A \perp\!\!\!\perp X_B \mid X_C$$

$\forall A, B, C \ni A \not\sim B$ are d-separated by C

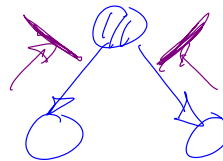
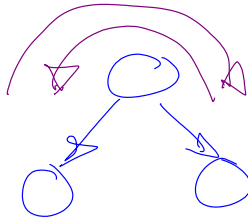
"Bayes - Ball" alg. : 'intuitive' alg. to check d-separation

rules for balls/chains being blocked

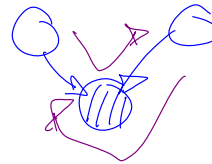
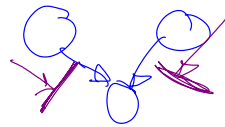
1)



2)

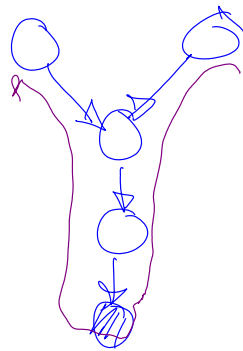
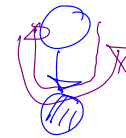


3) V-structure:



note that you can have
a same as b

ie.



see Alg. 3.1
in K & F book