{By: Gabriel Khoury}

{Causal Basal Nets}

Definition 1.3.1 (Causal Bayesian Network)

let:
- $P(v)$ be a probability distribution on a set V of variables
- $P_x(v)$ be the distribution resulting from the intervention $do(X=x)$ that sets a subset of X variables to constants x.
- P_* be the set of all interventional distributions of $P_x(v)$ such that $X \subseteq V$ and $P(v)$ represents no intervention

$\Rightarrow$ a DAG G is a causal Bayesian network compatible with P_* if and only if:

(i) $P_x(v)$ is Markov relative to G

(ii) $P_x(v_i) = 1$ for all $V_i \in X$ such that V_i consistent with $X = x$

(iii) $P_x(v_i | pa_i) = P(v_i | pa_i)$ for all $V_i \notin X$ whenever pa_i is consistent with $X = x$

Some reminders & summary of previous lectures:

- previously we discussed if we have a distribution P of n discrete variables $X_1, X_2, \ldots, X_n$, the chain rule of probability gives us: $P(X_1, \ldots, X_n) = \prod_j P(X_j | X_1, \ldots, X_{j-1})$

such that x_j is sensitive only to a small subset of predecessors

$\Rightarrow P(x_j \mid x_1, \dots, x_{j-1}) = P(x_j \mid \underbrace{pa_j})$

⚠ Markov compatibility — also known as Markovian parent "x_j independent of all other predecessors besides set pa_j"

$\Downarrow$

if the value of each variable x_i is chosen at random with probability $P_i(x_i \mid pa_i)$ based solely on PA_i $\Rightarrow$ distribution P is Markov relative to G

$\Rightarrow$ this is why $P_x(v)$ Markov relative to G given $P_x(v) = \prod_{i=1}^{n} P_x(v_i \mid pa_i)$

$\longrightarrow$ how can we prove equation (1.37):

⚠⚠ $P_x(v) = \prod_{\{i \mid V \notin X\}} P(v_i \mid pa_i)$ for all v consistent with x

proof:

- first, assume $v_i = x_i$ for all i
- second, using property (i) $P_x(v) = \prod_{i=1}^{n} P_x(v_i \mid pa_i)$ since $P_x(v)$ Markov relative to G
- third, using property (ii), if $V_i \in X$, $P_x(v_i) = 1$ $\Rightarrow P_x(v_i \mid pa_i) = 1$
- once we eliminate the terms with $V_i \in X$ we are left with $V_i \notin X$ $\Rightarrow$ using property (iii) $P_x(v_i \mid pa_i) = P(v_i \mid pa_i)$

→ how can we obtain properties (1) and (2) using conditions (ii) and (iii)

$P(v_i | pa_i) = P_{pa_i}(v_i)$ $\qquad$ $P_{pa_i, s}(v_i) = P_{pa_i}(v_i)$ such that S is disjoint of $\{V_i, PA_i\}$

① let us start with property (1):

- by Markov's property, $P_{pa_i}(v_i) = P_{pa_i}(pa_i) \cdot P_{pa_i}(v_i | pa_i)$
- $P_{pa_i}(pa_i) = 1$ since using property (ii) we know $pa_i \in X$
- since we are left with the term such that $V_i \notin X$:

property (iii) $\Rightarrow \boxed{P_{pa_i}(v_i) = P(v_i | pa_i)}$

② moving on to property (2):

- we have a subset S of variables disjoint of $\{V_i, PA_i\}$
- using property (1), $P_{pa_i, s} = P_s(v_i | pa_i)$
- since S is consistent with pa_i because it is disjoint from $\{V_i, PA_i\}$, apply property (iii) to get

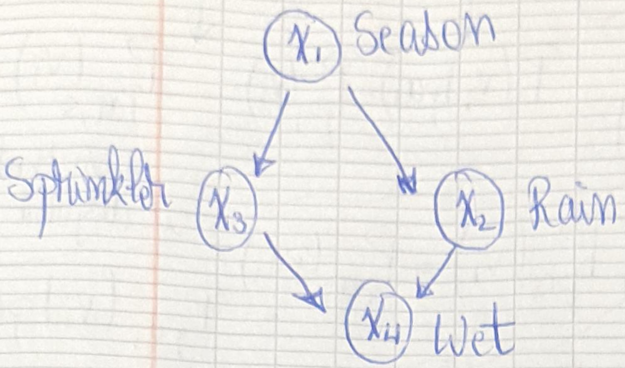
$$P_s(v_i | pa_i) = P(v_i | pa_i)$$

by property (1) once again

$$P_{pa_i}(v_i)$$

$\Rightarrow P_{pa_i, s}(v_i) = P_{pa_i}(v_i)$ when S is disjoint of $\{V_i, PA_i\}$

→ why do the properties make sense intuitively:



- learning the states of all the direct causes of the variable has the same effect as intervening to set all of those states.

- the direct causes of a wet floor are rain and sprinkler on
  - ↳ those two (X_2 and X_3) are all the direct causes

⟹ if you wake one day and you learn the state of what the rain is and the sprinkler is then your estimate of the probability that the road is wet should be the same as if you woke up and went and physically set the sprinkler to be in a particular state and the rain to be in a certain state

⟹ if you learn that it is raining today, your estimated probability of a wet floor will be the same as if you turned on the rain and turned off the sprinkler

⚠ this can only be valid if we are intervening on all of the parents of the variable.