Relationship: DAGs and probability distributions

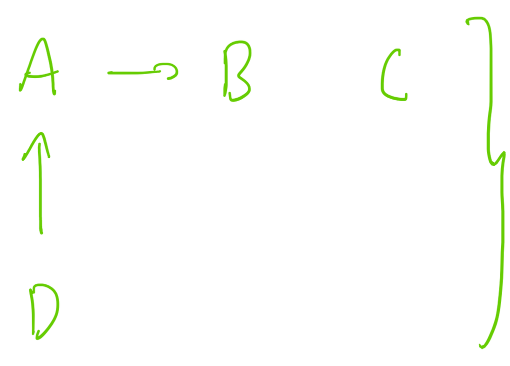
21.04.21 05:46

- DAGs encode assumptions (dependencies) between variables / nodes / vertices

DAG tells us 3 things:

1) btw. which variables we have independence
2) which variables are conditionally independent from each other
3) ways that we can factorize and simplify a joint dist.

example:



encodes information about the joint dist. $p(A,B,C,D)$

⇒ this DAG implies:

(1) $p(C \mid A,B,D) = p(C)$ (C is independent of all variables)

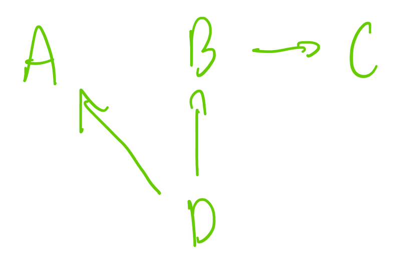
(2) $p(B \mid A,C,D) = p(B \mid A)$ (i.e. $B \perp\!\!\!\perp D, C \mid A$)

(3) $p(B \mid D) \neq p(B)$ (i.e. B and D are (marginally) dependent b/c

different $\begin{cases} p(B) = \int p(B \mid A,D) \cdot p(A \mid D) \cdot p(D) \; dD \, dA \\ p(B \mid D) = \int p(B \mid A,D) \cdot p(A \mid D) \; dA \end{cases}$

(4) $p(D \mid A,B,C) = p(D \mid A)$

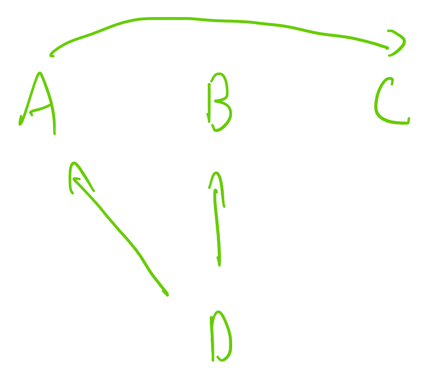
Example 2:



(1) $p(A \mid B,C,D) = p(A \mid D)$
↳ b/c $A \perp\!\!\!\perp B, C \mid D$

(2) $p(D \mid A,B,C) = p(D \mid A,B)$
↳ b/c $D \perp\!\!\!\perp C \mid B$
↳ we need A, B b/c both are affected by D, thus A, B should tell us sth. independently about D

Example 3:



(1) $p(A \mid B,C,D) = p(A \mid C,D)$
↳ $A \perp\!\!\!\perp B \mid C, D$

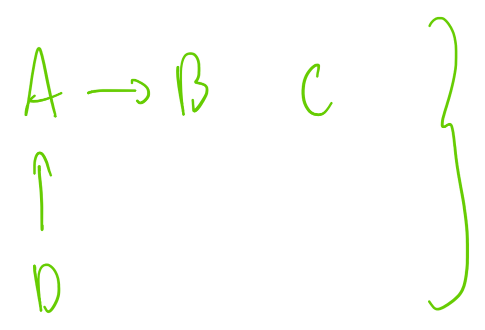
(2) $p(D \mid A,B,C) = p(D \mid A,B)$
↳ $D \perp\!\!\!\perp C \mid A,B$

Decomposition of a joint distribution:

- decompose joint into product of conditional dist. by sequentially only conditioning on parents only.
  step 1: start with root nodes (i.e. nodes w/o parents)
  step 2: proceed "down" the descendent line, always conditioning on parents

⇓

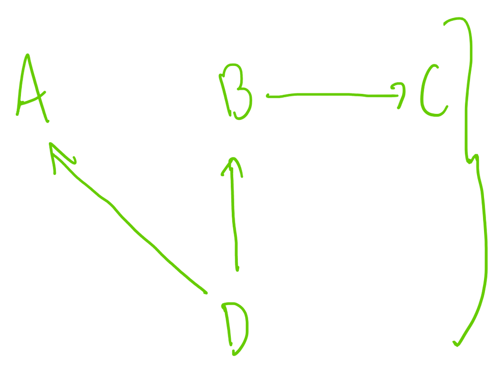
example joint dist. Decomposition using DAG:



$p(A,B,C,D) = \underbrace{p(C) \cdot p(D)}_{\text{step 1: root nodes}} \cdot \overbrace{p(A \mid D)}^{\text{look at step 2: children of roots}} \cdot \underbrace{p(B \mid A)}_{\text{step 3: look at children of nodes from step 2}}$

iff this DAG is correct, then $p(A,B,C,D) = p(C)\,p(D)\,p(A \mid D)\,p(B \mid A)$ is a proper / correct / truthful decomposition of the joint distribution.

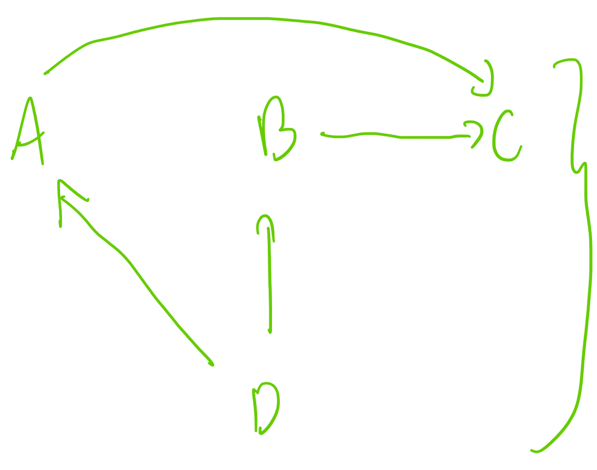
example #2



DAG

$p(A,B,C,D) = \underbrace{p(D) \cdot p(A \mid D) \cdot p(B \mid D) \cdot p(C \mid B)}_{\text{based on DAG to the left}}$

example #3 decomposition of a joint dist.:



$p(A,B,C,D) = \underbrace{p(D) \cdot p(A \mid D) \cdot p(B \mid D) \cdot p(C \mid A,B)}_{\text{the DAG admits this factorization}}$

the DAG and the probability fct. are compatible.

Compatibility btw. DAGs and prob. dists.:

DAGs that are compatible w/ a particular prob. fct. are not necessarily unique!

⇒ a decomposed joint prob. dist. doesn't necessarily imply a unique DAG !!!

↳ simple example

DAG 1: $A \to B$ — $p(A) \cdot p(B \mid A) = p(A,B)$

both DAGs convey dependence btw. A and B

DAG 2: $B \to A$ — $p(B) \cdot p(A \mid B) = p(A,B)$

which we know implies: $p(A,B) \neq p(A) \cdot p(B)$ (!)

this statement is compatible with both DAGs, DAG 1 & DAG 2.

⇓

lesson: probability statements do not imply a unique DAG!!!

(✓) DAG ⇒ unique prob. dist. statement

(✗) prob. dist. statement $\not\Rightarrow$ unique DAG

(BUT not other direction)