Bayesian Networks 2
Reasoning Patterns, Independencies

Reasoning Patterns

Reasoning Patterns

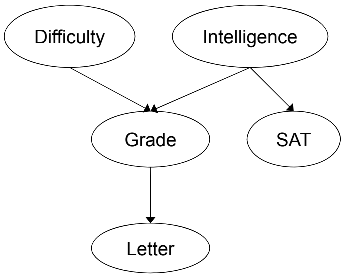
- A joint probability distribution allows us to calculate probabilities like:

  $P(\mathbf{Y} = \mathbf{y} \mid \mathbf{E} = \mathbf{e})$

  Evidence

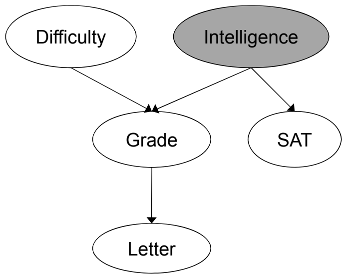
- Bayes nets allow us to see how this probability changes as we observe different evidence

Reasoning Patterns



P(Letter = Strong) = 0.502

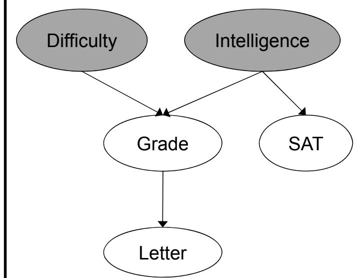
Reasoning Patterns



P(Letter = Strong) = 0.502

P(Letter = Strong | Intelligence = low) = 0.389

Reasoning Patterns

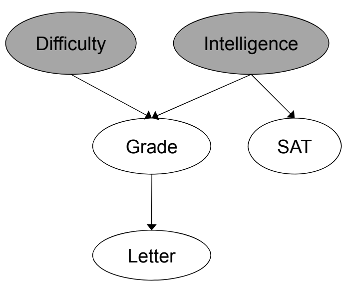


P(Letter = Strong) = 0.502

P(Letter = Strong | Intelligence = low) = 0.389

P(Letter = Strong | Intelligence = low, Difficulty = low) = 0.513

Reasoning Patterns



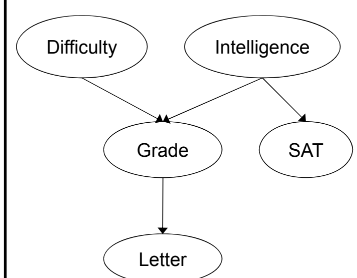
Predicting the "downstream" effects of evidence – instances of causal reasoning or prediction

P(Letter = Strong) = 0.502

P(Letter = Strong | Intelligence = low) = 0.389

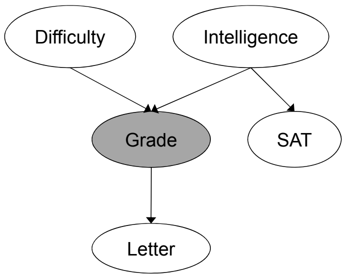
P(Letter = Strong | Intelligence = low, Difficulty = low) = 0.513

Reasoning Patterns



P(Intelligence = high) = 0.30

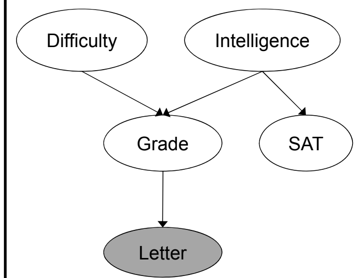
Reasoning Patterns



P(Intelligence = high) = 0.30

P(Intelligence = high | Grade = C) = 0.079

Reasoning Patterns

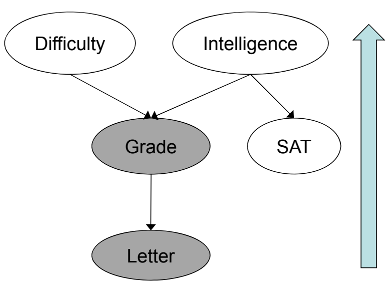


P(Intelligence = high) = 0.30

P(Intelligence = high | Grade = C) = 0.079

P(Intelligence = high | Letter = Weak) = 0.14

Reasoning Patterns



Reasoning from effects to causes are instances of evidential reasoning or explanation

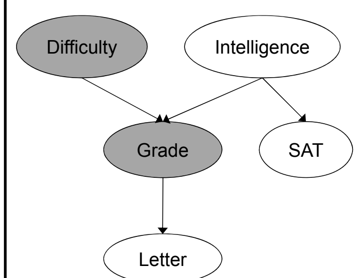
P(Intelligence = high) = 0.30

P(Intelligence = high | Grade = C) = 0.079

P(Intelligence = high | Letter = Weak) = 0.14

P(Intelligence = high | Grade = C, Letter = Weak) = 0.079

Reasoning Patterns



Why does observing Difficulty=high make the probability 0.34?

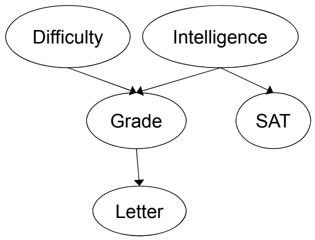
Notice how Difficulty (causal factor for Grade) gave us information about Intelligence (another causal factor for Grade).

This is called "explaining away" (more about this in the next few lectures)

P(Intelligence = high | Grade = B) = 0.079

P(Intelligence = high | Grade = B, Difficulty = high) = 0.34

Independencies



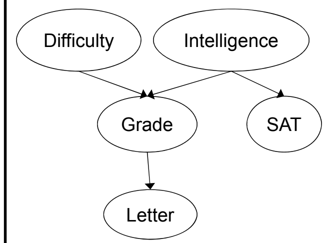
What are some conditional independence statements in this network?

$(L \perp \{I, D, S\} \mid G)$
Once we know the student's grade, our beliefs about the quality of his recommendation letter are not influenced by any other variable

$(S \perp \{D, G, L\} \mid I)$
SAT score is conditionally independent of all other nodes given I

13

Independencies



What about : $(G \perp L \mid D, I)$? (conditioning on parents of G only)

Intuitively (and using our model), this is false. Suppose we have a smart student in a difficult class. If the student gets a strong letter, then we expect

P(Grade = A | Intelligence = high, Difficulty = high, Letter = strong) >
P(Grade = A | Intelligence = high, Difficulty = high)

14

Independencies

- Knowing the value of a variable's parents "shield" it from information relating directly or indirectly to its other ancestors
- Information about the variable's descendants can change its probability
- What's the general pattern?

15

Independencies

Definitions

- A Bayesian network structure $\mathcal{G}$ is a directed acyclic graph whose nodes represent random variables $X_1, \ldots, X_n$.
- Let $Parents(X_i, \mathcal{G})$ denote the parents of X_i in $\mathcal{G}$,
- *Let* $NonDescendants(X_i)$, denote the variables in the graph that are not descendants of X_i.

16

Independencies

Then $\mathcal{G}$ encodes the following set of conditional independence assumptions, called the local independencies, and denoted by $\mathcal{I}_\ell(\mathcal{G})$:

The ℓ stands for "local"

For each variable X_i:
$(X_i \perp NonDescendants(X_i) \mid Parents(X_i, \mathcal{G}))$

Informally: X_i is conditionally independent of its nondescendants given its parents

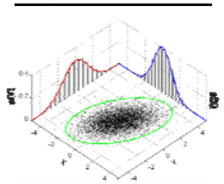
17

Graphs and Distributions

18

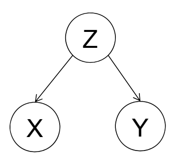
Graphs and Distributions

Distribution P



Has some set of independence relationships $\mathcal{I}$(P) eg. $(\boldsymbol{X} \perp \boldsymbol{Y} \mid \boldsymbol{Z})$

Graph $\mathcal{G}$



Has some set of local independence relationships $\mathcal{I}_\ell(\mathcal{G})$

How do we represent P using $\mathcal{G}$?

19

Graphs and Distributions

Distribution P

- Let P be a distribution over $\mathcal{X}$.
- Let $\mathcal{I}(P)$ be the set of independence assertions of the form $(\boldsymbol{X} \perp \boldsymbol{Y} \mid \boldsymbol{Z})$ that hold in P.

Graph $\mathcal{G}$

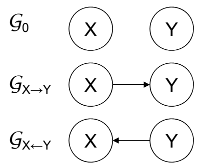
- Let $\mathcal{G}$ be any graph associated with a set of independencies $\mathcal{I}(\mathcal{G})$.
- $\mathcal{G}$ is an I-map for a set of independencies $\mathcal{I}(P)$ if $\mathcal{I}(\mathcal{G}) \subseteq \mathcal{I}(P)$.

Note: any independencies that $\mathcal{G}$ asserts must hold in P but P may have additional independencies that are not reflected in $\mathcal{G}$.

20

Graphs and Distributions

Three graphs with 2 variables X, Y:



$\mathcal{G}_0$: Independence assumption: $X \perp Y$

$\mathcal{G}_{X\to Y}$: No independence assumptions encoded

$\mathcal{G}_{X\leftarrow Y}$: No independence assumptions encoded

Suppose we have the following 2 distributions:

P_{left}

X	Y	P(X,Y)
0	0	0.08
0	1	0.32
1	0	0.12
1	1	0.48

P_{right}

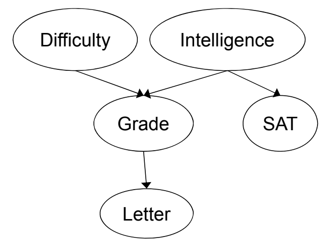
X	Y	P(X,Y)
0	0	0.4
0	1	0.3
1	0	0.2
1	1	0.1

All 3 graphs are I-maps of P_{left}.

$\mathcal{G}_0$ is not an I-map of P_{right} since $(X \perp Y) \notin I(P)$

Graphs and Distributions

- Suppose we have a distribution *P* for which the student Bayes net is an I-map
- From the student Bayes net, we can see examples of the conditional independencies in $\mathcal{I}(\mathcal{G})$ (and hence in $\mathcal{I}(P)$):



<u>$\mathcal{I}(P)$</u>
$(L \perp \{I, D, S\} \mid G)$
$(S \perp \{D, G, L\} \mid I)$
$(G \perp L \mid D, I)$
$(I \perp D)$
…

Graphs and Distributions

- We can decompose the joint distribution for the student Bayes net :

$$\boxed{\begin{aligned}&P(I,D,G,L,S)\\&= P(I)P(D\,|\,I)P(G\,|\,I,D)P(L\,|\,I,D,G)P(S\,|\,I,D,G,L)\end{aligned}}$$

[Chain Rule]

- But some of these conditional probability distributions are quite big eg. *P(S|I,D,G,L)*

Graphs and Distributions

Using the conditional independence assumptions:

- $(I \perp D) \in \mathcal{I}(P)$ implies $P(D \mid I) = P(D)$
- $(L \perp \{I, D\} \mid G) \in \mathcal{I}(P)$ implies $P(L \mid I, D, G) = P(L \mid G)$
- $(S \perp \{D, G, L\} \mid I) \in \mathcal{I}(P)$ implies $P(S \mid I, D, G, L) = P(S \mid I)$

$$\boxed{\begin{aligned}&P(I,D,G,L,S)\\&= P(I)P(D\,|\,I)P(G\,|\,I,D)P(L\,|\,I,D,G)P(S\,|\,I,D,G,L)\\&= P(I)P(D)P(G\,|\,I,D)P(L\,|\,G)P(S\,|\,I)\end{aligned}}$$

Graphs and Distributions

$$P(I,D,G,L,S) = P(I)P(D)P(G \mid I,D)P(L \mid G)P(S \mid I)$$

- The joint distribution can be computed as a product of factors, one for each variable.
- Each factor represents a conditional probability of the variable given its parents in the network.
- This factorization applies to any distribution P for which $\mathcal{G}_{student}$ is an I-Map.

25

Graphs and Distributions

The chain rule for Bayesian networks

- Let $\mathcal{G}$ be a Bayes net graph over the variables $X_1, \ldots, X_n$. We say that a distribution P over the same space factorizes according to $\mathcal{G}$ if P can be expressed as a product

$$P(X_1,\ldots,X_n) = \prod_{i=1}^{n} P(X_i \mid Parents(X_i,\mathcal{G}))$$

- A Bayesian network is a pair $\mathcal{B} = (\mathcal{G}, P)$ where P factorizes over $\mathcal{G}$, and where P is specified as a set of CPDs associated with $\mathcal{G}$'s nodes. The distribution is often annotated $P_{\mathcal{B}}$.

26

Graphs and Distributions

From Theorems 3.1 and 3.2:

$\mathcal{G}$ is an I-map for P ie. $(X_i \perp NonDescendants(X_i) \mid Parents(X_i, \mathcal{G})$

$$\Updownarrow$$

P factorizes as:

$$\left(P(X_1,\ldots,X_n) = \prod_{i=1}^{n} P(X_i \mid Parents(X_i,\mathcal{G})) \right)$$

27