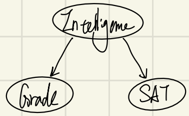
Intelligence $I = \begin{cases} i^0 & \text{low} \\ i^1 & \text{high} \end{cases}$

SAT $S = \begin{cases} s^0 & \text{low} \\ s^1 & \text{high} \end{cases}$

Grade $G = \begin{cases} g^1 & A \\ g^2 & B \\ g^3 & C \end{cases}$

$P(S, I) \longrightarrow P(I) \cdot P(S|I)$



$P(I,S,G) = P(S,G|I) \cdot P(I)$

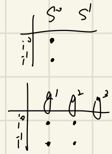
2 values, 2, 3

$2 \times 2 \times 3 - 1 = 11$ params

$= P(I) \cdot P(S|I) \cdot P(G|I)$

1, 2, 4

$1 + 2 + 4 = 7$ params



- **Bayesian Network**

The skeleton for representing a joint distribution compactly in a factorized way

The representation for a set of conditional independence assumption about a distribution

Intelligence $I = \begin{cases} i^0 & \text{low} \\ i^1 & \text{high} \end{cases}$

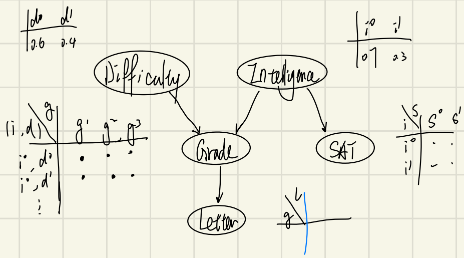
SAT $S = \begin{cases} s^0 & \text{low} \\ s^1 & \text{high} \end{cases}$

Grade $G = \begin{cases} g^1 & A \\ g^2 & B \\ g^3 & C \end{cases}$

Test Difficulty $D = \begin{cases} \text{easy} \\ \text{hard} \end{cases}$

Recommendation $L = \begin{cases} \text{strong} \\ \text{weak} \end{cases}$

The Bayesian Network represents a joint distribution via the chain rule for Bayesian Networks

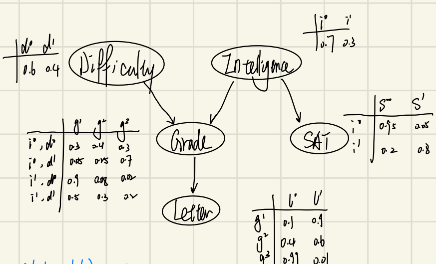


each variable is a stochastic function of its parents

$P(I,D,G,S,L) = P(I) \cdot P(D) \cdot P(S|I) \cdot P(G|I,D) \cdot P(L|G)$ see this as a factor product

P factorizes over G if $P(X_1 \cdots X_n) = \prod P(X_i | \text{parents}(X_i))$

- **Reasoning Patterns**



	d^0	d^1
	0.6	0.4

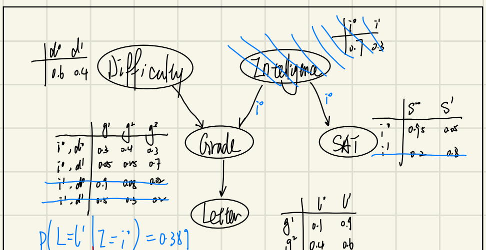
	i^0	i^1
	0.7	0.3

	g^1	g^2	g^3
i^0, d^0	0.3	0.4	0.3
i^0, d^1	0.05	0.25	0.7
i^1, d^0	0.9	0.08	0.02
i^1, d^1	0.5	0.3	0.2

	s^0	s^1
i^0	0.95	0.05
i^1	0.2	0.8

	l^0	l^1
g^1	0.1	0.9
g^2	0.4	0.6
g^3	0.99	0.01

$P(L = l^1) = 0.502$



$P(L = l^1 | I = i^0) = 0.389$

causal reasoning

evidential reasoning

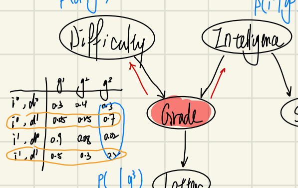
$P(d^1) = 0.4$

$P(d^1 | g^3) \approx 0.63$

$P(i^1) = 0.3$

$P(i^1 | g^3) \approx 0.08$

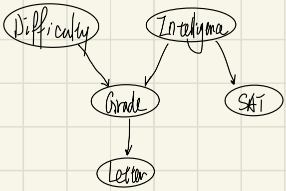
$P(i^1 | g^3, d^1) \approx 0.11$



$P(\ | g^3, d^1)$

$P(\ | g^3)$

- Dependencies



$(L \perp I, D, S \mid G)$

$(S \perp D, G, L \mid I)$

$(G \perp S \mid I, D)$

$(I \perp D)$

$(D \perp I, S)$

parents of variable "shield" it from probabilistic influence that is causal in nature

$(X_i \perp NonDescendants(X_i) \mid Parents(X_i))$

each node X_i is conditionally independent of its nondescendants given its parents

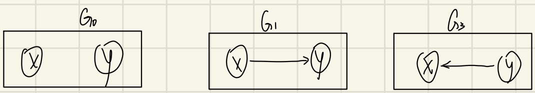
- I-maps

P is a distribution over X, $I(P)$ is the set of independence $\{(X \perp Y \mid Z)\}$

if P satisfies local independences associated with graph G, then $I_\ell(G) \subseteq I(P)$

G is a I-map for P

eg



X	Y	$P_1(X,Y)$
x^0	y^0	0.08
x^0	y^1	0.32
x^1	y^0	0.12
x^1	y^1	0.48

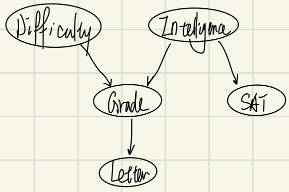
X	Y	$P_2(X,Y)$
x^0	y^0	0.4
x^0	y^1	0.3
x^1	y^0	0.2
x^1	y^1	0.1

$I(P_1) = \{X \perp Y\}$ $I(P_2) = \emptyset$

$I(G_0) = \{X \perp Y\}$ $I(G_1) = \emptyset$ $I(G_2) = \emptyset$

G_0, G_1 and G_2 are all I-maps for P_1

· I-maps to Factorization



for any distribution P of which G is a I-map, $I_\ell(G) \subseteq I(P)$
we can decompose P by local independence of G

$$P(I,D,G,L,S) = p(I)\cdot P(D|I)\cdot P(G|D,I)\cdot P(L|I,D,G)\cdot P(S|L,I,D,G)$$ total chain rule

$$= P(I)\cdot P(D)\cdot p(G|D,I)\cdot P(L|G)\cdot p(S|I)$$ independences

the factorization applies to any distribution P of which G is a I-map

P factorizes according to G if P can be expressed as

$$P(X_1 \cdots X_n) = \prod_i P(X_i \mid \mathrm{Parent}(X_i))$$

A Bayesian network $B = (G, P)$ where P factorizes over G,
and P is specified as CPDs associated with nodes of G

if G is a I-map for P, then P factorizes according to G

$$P(X_1, \dots, X_n) = \prod_{i=1}^{n} P(X_i \mid X_1, \dots X_{i-1})$$

$$= \prod_{i=1}^{n} P(X_i \mid \mathrm{parent}(X_i))$$

assume $X_1, \cdots X_n$ is a topological ordering.
$(X_i \perp \mathrm{Nondescendants}(X_i) \mid \mathrm{parents}(X_i))$

$\therefore$ P factorizes over G

G is an I-map for P
$\Updownarrow$
P factorizes over G

· Factorization to I-maps

if P factorizes according to G, then G is an I-map for P
$(\text{d-sep}(X,Y|Z) \Rightarrow (X \perp Y | Z))$

Suppose $X_1 \cdots X_n$ is a topological ordering

$$p(X_i \mid X_1 \cdots X_{i-1}) = \frac{P(X_1 \cdots X_{i-1}, X_i)}{P(X_1 \cdots X_{i-1})}$$

$$= \frac{\prod_{j=1}^{i} P(X_j \mid \mathrm{parents}(X_j))}{\prod_{j=1}^{i-1} P(X_j \mid \mathrm{parents}(X_j))}$$

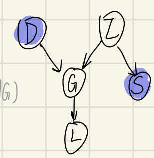
$$= p(X_i \mid \mathrm{parents}(X_i))$$

$\therefore p(X_i \mid X_1 \cdots X_{i-1}) = p(X_i \mid \mathrm{parents}(X_i))$

$\therefore$ P satisfies the local independencies of G $\quad I_\ell(G) \subseteq I(P)$

$\therefore$ G is a I-map for P

DLS



$$P(D,I,G,S,L) = P(D)\cdot P(I)\cdot p(G|D,I)\cdot P(S|I)\cdot p(L|G)$$

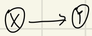
$$P(D,S) = \sum_{I,G,L} P(D)\cdot P(I)\cdot p(G|D,I)\cdot P(S|I)\cdot p(L|G)$$

$$= \sum_I P(D)\, P(I)\, P(S|I) \sum_G P(G|D,I) \sum_L p(L|G)$$

$$= P(D)\; P(S)$$

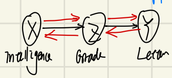
· D-separation

1. Direct connection



$(X \not\perp Y \mid Z)$

2. Indirect causal effect / Indirect evidential effect

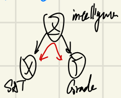


$(X \not\perp Y)$ $(X \perp Y \mid Z)$

$(Y \perp X \mid Z)$

causal trail / evidential trail is active iff Z is not observed

3. Common Cause



$(X \not\perp Y)$ $(X \perp Y \mid Z)$

active iff Z is not observed

4. Common effect

$(X \perp Y)$ $(X \not\perp Y \mid Z)$

active iff (Z or one of Z's descendants) is observed

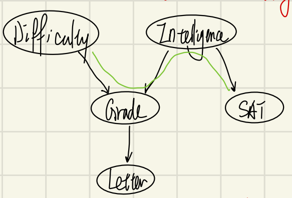
Let G be a Bayesian network structure. $X_1 \rightleftharpoons \cdots \rightleftharpoons X_n$ a trail in G

Let Z be a subset of observed variable

the trail $X_1 \rightleftharpoons \cdots \rightleftharpoons X_n$ is active given Z if

whenever we have a v-structure $X_{i-1} \rightarrow X_i \leftarrow X_{i+1}$, X_i or one of its descendants is observed

no other node along the trail is in Z



not active when neither G or L is observed

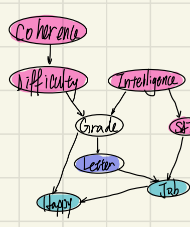
not active when $\{L, I\}$ are observed

one node can influence another if there is any trail along which influence can flow

Let $\mathbf{X}, \mathbf{Y}, \mathbf{Z}$ be 3 sets of nodes in G. We say that $\mathbf{X}$ and $\mathbf{Y}$ are d-separated given $\mathbf{Z}$ $\left(\text{d-sep}_G(\mathbf{X}; \mathbf{Y} \mid \mathbf{Z})\right)$

if there is no active trail between any node $X \in \mathbf{X}$ and $Y \in \mathbf{Y}$ given $\mathbf{Z}$

$I(G) = \{(\mathbf{X} \perp \mathbf{Y} \mid \mathbf{Z}) : \text{d-sep}_G(\mathbf{X}; \mathbf{Y} \mid \mathbf{Z})\}$ the set of global Markov independencies



not descendant not parent

descendant

any node is d-separated from its non-descendants given its parents and not given other nodes

trials from top are blocked by giving the parents

trials from bottom are blocked by v-structure

· Soundness and Completeness

① If a distribution P factorizes according to G, then $\mathcal{I}(G) \subseteq \mathcal{I}(P)$
(if X and Y are d-separated given Z, then $(X \perp Y \mid Z)$ in P)

A distribution P is faithful to G if $(X \perp Y \mid Z) \in \mathcal{I}(P) \Rightarrow \text{d-sep}_G(X; Y \mid Z)$
(if any dependence in P is reflected in d-separation of G)

Even if a distribution factorizes over G, it can still contain independencies that are not reflected in G.

eg.

$P(A, B) = P(A, B)$. $\qquad A \to B$ (G)

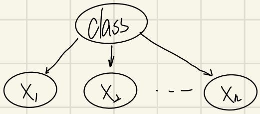
	b^0	b^1
a^0	0.4	0.6
a^1	0.4	0.6

there's no independencies in G.
$\therefore$ P factorizes over G

② if $(X \perp Y \mid Z)$ is in all distributions P that factorize over G, then $\text{d-sep}_G(X; Y \mid Z)$

③ if X and Y are not d-sep given Z in G, then X and Y are dependent in some P that factorizes over G

· Naive Bayes



$(X_i \perp X_j \mid C)$

$P(C, X_1, \ldots X_n) = P(C) \cdot \prod_{i=1}^{n} P(X_i \mid C)$

$$\frac{P(C=c^1 \mid X_1 \ldots X_n)}{P(C=c^2 \mid X_1 \ldots X_n)} = \frac{P(C=c^1, X_1 \ldots X_n) / P(X_1, \ldots X_n)}{P(C=c^2, X_1 \ldots X_n) / P(X_1 \ldots X_n)}$$

$$= \frac{P(C=c^1) \cdot P(X_1 \mid C=c^1) \; P(X_2 \mid X_1, C=c^1) \cdots}{P(C=c^2) \cdot P(X_1 \mid C=c^2) \cdot P(X_2 \mid X_1, C=c^2) \cdots}$$

$$= \frac{P(C=c^1)}{P(C=c^2)} \cdot \prod_{i=1}^{n} \frac{P(X_i \mid C=c^1)}{P(X_i \mid C=c^2)}$$

	X_i^0	X_i^1
$C=c^1$	$1-p_i$	p_i
$C=c^2$	$1-q_i$	q_i

$$\log \frac{P(C=c^1 \mid X_1 \ldots X_n)}{P(C=c^2 \mid X_1 \ldots X_n)} = \log \frac{P(C=c^1)}{P(C=c^2)} + \sum_{i=1}^{n} \log \frac{P(X_i \mid C=c^1)}{P(X_i \mid C=c^2)}$$

$$= \left[\log P(C=c^1) - \log P(C=c^2)\right] + \sum_{i=1}^{n} \left[\log P(X_i \mid C=c^1) - \log P(X_i \mid C=c^2)\right]$$

$$= \left[\log P(C=c^1) - \log P(C=c^2)\right] + \sum_{i=1}^{n} \left[\log p_i^{X_i} (1-p_i)^{(1-X_i)} - \log q_i^{X_i} (1-q_i)^{(1-X_i)}\right]$$

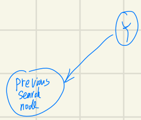
$$= \cdots + \sum_{i=1}^{n} \left[X_i \log p_i + (1-X_i) \log (1-p_i) - X_i \log q_i - (1-X_i) \log (1-q_i) \right]$$

$$= \alpha_0 + \sum_{i=1}^{n} \alpha_i X_i$$

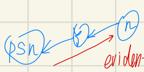
- Algorithm for finding reachable nodes via active trails

```
def d_sep_reachable(G: DirectedGraph, X: node, Z: Set<Node>):
    # phase I: mark all ancestor of Z
    Z_cp = Z.copy(); Z_ancestor = ∅
    while Z_cp ≠ ∅:
        Y = Z_cp[0]   Z_cp = Z_cp[1:]           like breadth-first search, but reversing the direction of edges
        if Y ∉ Z_ancestor:
            Z_cp = Z_cp ∪ Parents(Y)
            Z_ancestor = Z_ancestor ∪ (Y)
    # phase II: traverse active trails
    L = {(X, ↑)};          # (X, ↑) means searching X from X's children
    visited = ∅;   R = ∅   # R keeps track of all reachable nodes from X given Z
    while L ≠ ∅:
        Y, d = L[0];   L = L[1:]             # Y is a reachable node from X given Z
        if (Y, d) ∈ visited:
            continue
        visited = visited ∪ (Y, d)

        if Y ∉ Z:
            R = R ∪ (Y)
```



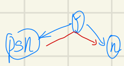

```
        if d = ↑ and Y ∉ Z:
            for n ∈ parents(Y):
                L = L ∪ (n, ↑)
```



is an active trail if Y is not given

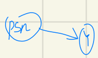
evidential effect

```
            for n ∈ children(Y)
                L = L ∪ (n, ↓)
```

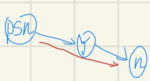


common cause

```
        else if d = ↓:
```

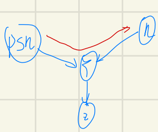



```
            if Y ∉ Z then:
                for n ∈ children(Y):
                    L = L ∪ (n, ↓)
```



causal effect

```
            if Y ∈ Z_ancestor:
                for n in parents(Y)
                    L = L ∪ (n, ↑)
    return R
```



common effect (V-structure)

active if Y is ancestor of observations

· I-equivalence

different BN structures can be equivalent in that they encode same conditional independence assertions

$X \to Z \to Y$ $\quad$ $X \leftarrow Z \leftarrow Y$ $\quad$ $X \leftarrow Z \to Y$

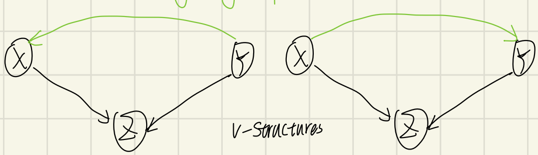
$(X \perp Y \mid Z)$

2 graph structures K_1 K_2 are I-equivalent if $I(K_1) = I(K_2)$
if P factorizes over K_1, it factorizeds over K_2

the skeleton of a graph is the node and edges without directions

if G_1 and G_2 have the same skeleton and same set of v-structures, they are I-equivalent

covering edges for v-structures



a v-structure $X \to Z \leftarrow Y$ is an immorality if there is no direct edge between X and Y
if there is such an edge, it's called an covering edge for the v-structure

G_1 and G_2 have the same skeleton and same set of immoralities $\Leftrightarrow$ G_1 and G_2 are I-equivalent

· Minimal I-Map

A graph K is a minimal I-map for a set of independencies I if

$\begin{cases} ① \text{ it is an I-map for I} \\ ② \text{ removing and edge renders it not an I-map} \end{cases}$

· Algorithm for finding Minimal I-map Bayesian Network of Distribution P

```
def build_minimal_I_map(𝒳: list<r.v>, I: list<Independences>)
    set G to an empty graph over 𝒳
    for i = 1, ... n
        U = {X1 ... Xi-1}    # candidates for parents of Xi
        for U' ⊆ {X1, ... , Xi-1}
            if U' ⊂ U and (Xi ⊥ {X1, ... , Xi-1} - U' | U') ∈ I
                U = U'
        // at this stage U is a minimal set satisfying {Xi ⊥ {X1 ... Xi-1} - U | U}
        for Xj ∈ U
            add edge Xj → Xi
    return G
```

· Conclusion

A Bayesian Network is a $\begin{cases} \text{factorization of } P \\ \text{I-map of } P \end{cases}$