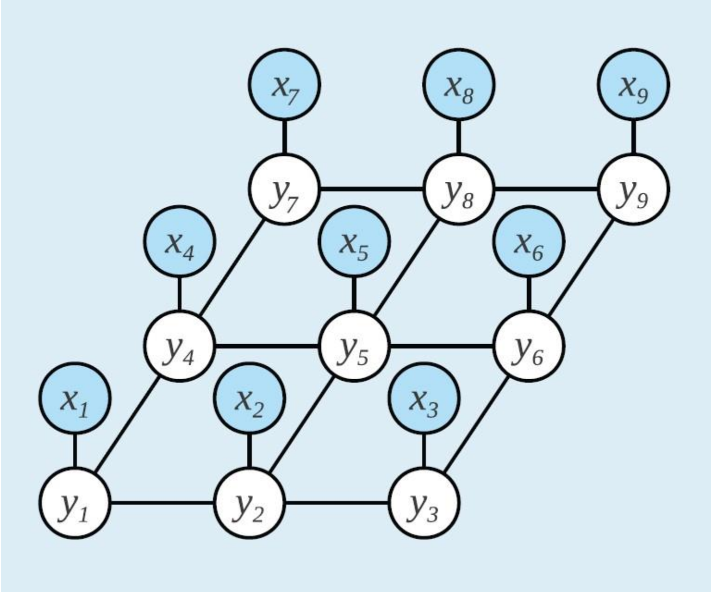
Probabilistic Graphical Models in Bioinformatics

Lecture 2: Bayesian network representation

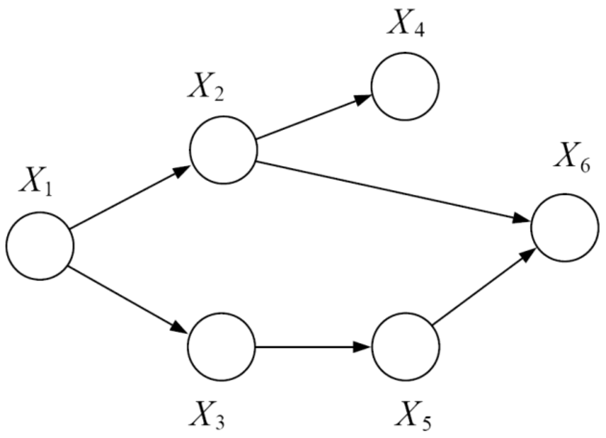


Hesam Montazeri
Department of Bioinformatics, IBB, University of Tehran

Bahman 29, 1397
(Feb 18, 2019)

Representation, inference, learning

- **Representation:** represent the problem as a collection of random variables $X_1, \ldots, X_n$ with joint distribution $P(X_1, \ldots, X_n)$ with some conditional independences.



- **Inference:** compute conditional probabilities given some evidences $P(X_i \mid E = e)$.

- **Learning:** estimate the parameters and structure of a Bayesian network from data

Outline

- Exploiting Independence properties
- Bayesian networks
- Independencies in Graphs

Exploiting Independence properties

Independent Random Variables

- Let X_i represent the outcome of a toss of coin *i*
  - Under the assumption coin tosses are independent

$$P(X_1, X_2, \ldots., X_n) = P(X_1)P(X_2) \ldots P(X_n)$$

- By exploiting independence information, we only require *n* parameters: $\theta_1, \ldots, \theta_n$ $\quad \theta_i = P(X_i = 1)$

$$P(x_1, \ldots, x_n) = \prod_i P(x_i) = \prod_i \theta_i^{x_i}(1-\theta_i)^{1-x_i}$$

This representation is limited for most real-world data

Conditional Parameterization

- Ex. A company trying to hire an **intelligent** college graduate
  - Have access to the student's SAT scores

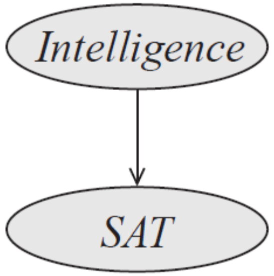
- Random variables:
  - Intelligence: $Val(I) = \{i^1, i^0\}$, represents high and low intelligence.
  - SAT score: $Val(S) = \{s^1, s^0\}$, represents high and low score.

- One possible way to represent the joint distr.
  - requires 3 parameters

I	S	$P(I, S)$
i^0	s^0	0.665
i^0	s^1	0.035
i^1	s^0	0.06
i^1	s^1	0.24.

Alternative representation using the chain rule

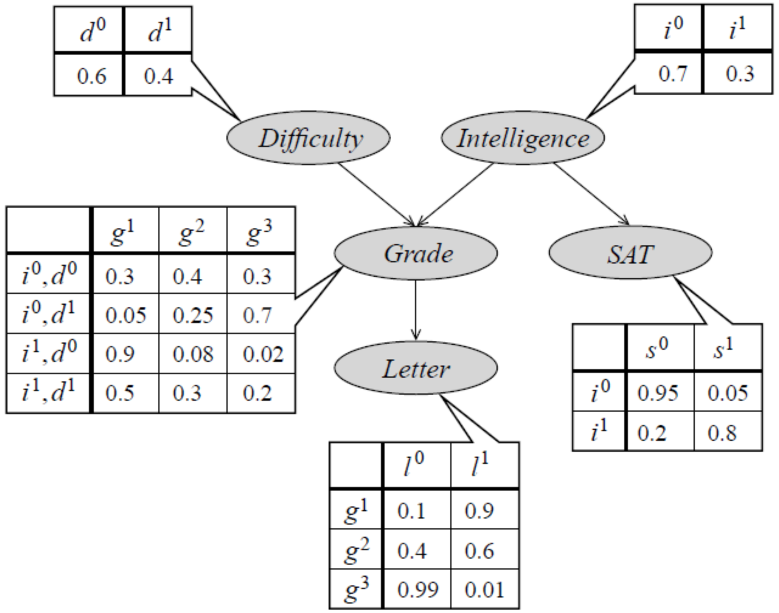
- More natural way of representing the same joint distribution:



$$P(I, S) = P(I)P(S \mid I)$$

i^0	i^1
0.7	0.3

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



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

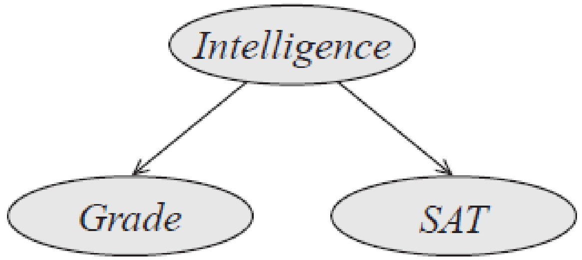
- How to parameterize it?
  - Using 3 distributions; one for $P(I)$, two for $P(S \mid i^0)$ and $P(S \mid i^1)$
  - Hence, we need 3 independent parameters: θ_{i^1}, $\theta_{s^1 \mid i^0}$ and $\theta_{s^1 \mid i^1}$

Naïve Bayes Model-the *Student* example

- Assumption: given intelligence, SAT and Grade do not give information about each other

$$S \perp G \mid I$$

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

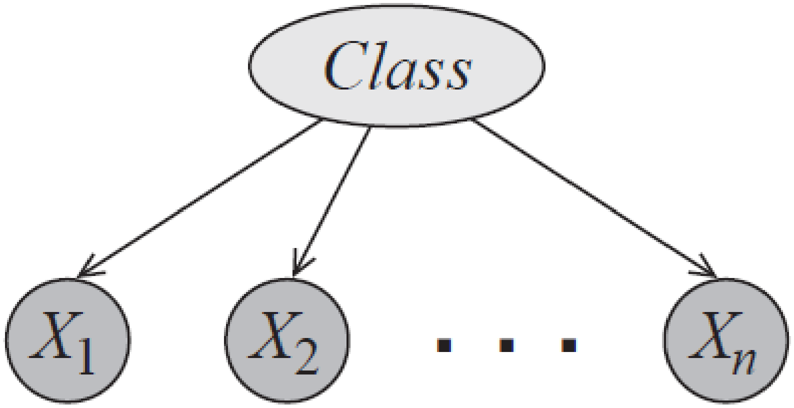


- Joint distr. can be written as?

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

Naïve Bayes Model-the general model

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



This model is generally used in the classification.

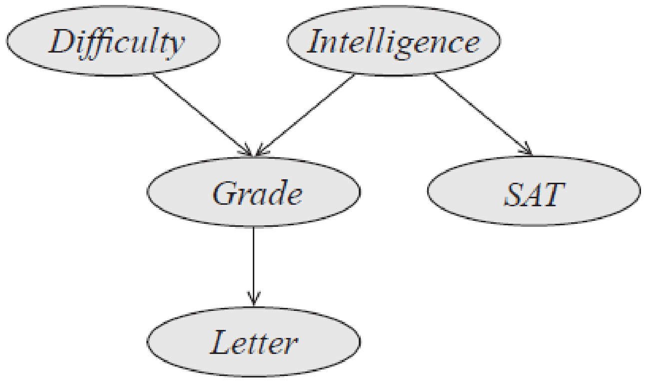
$$\frac{P(C = c^1 \mid x_1, \dots, x_n)}{P(C = c^2 \mid x_1, \dots, x_n)} = \frac{P(C = c^1)}{P(C = c^2)} \prod_{i=1}^{n} \frac{P(x_i \mid C = c^1)}{P(x_i \mid C = c^2)}$$

University of TEHRAN

Bayesian networks

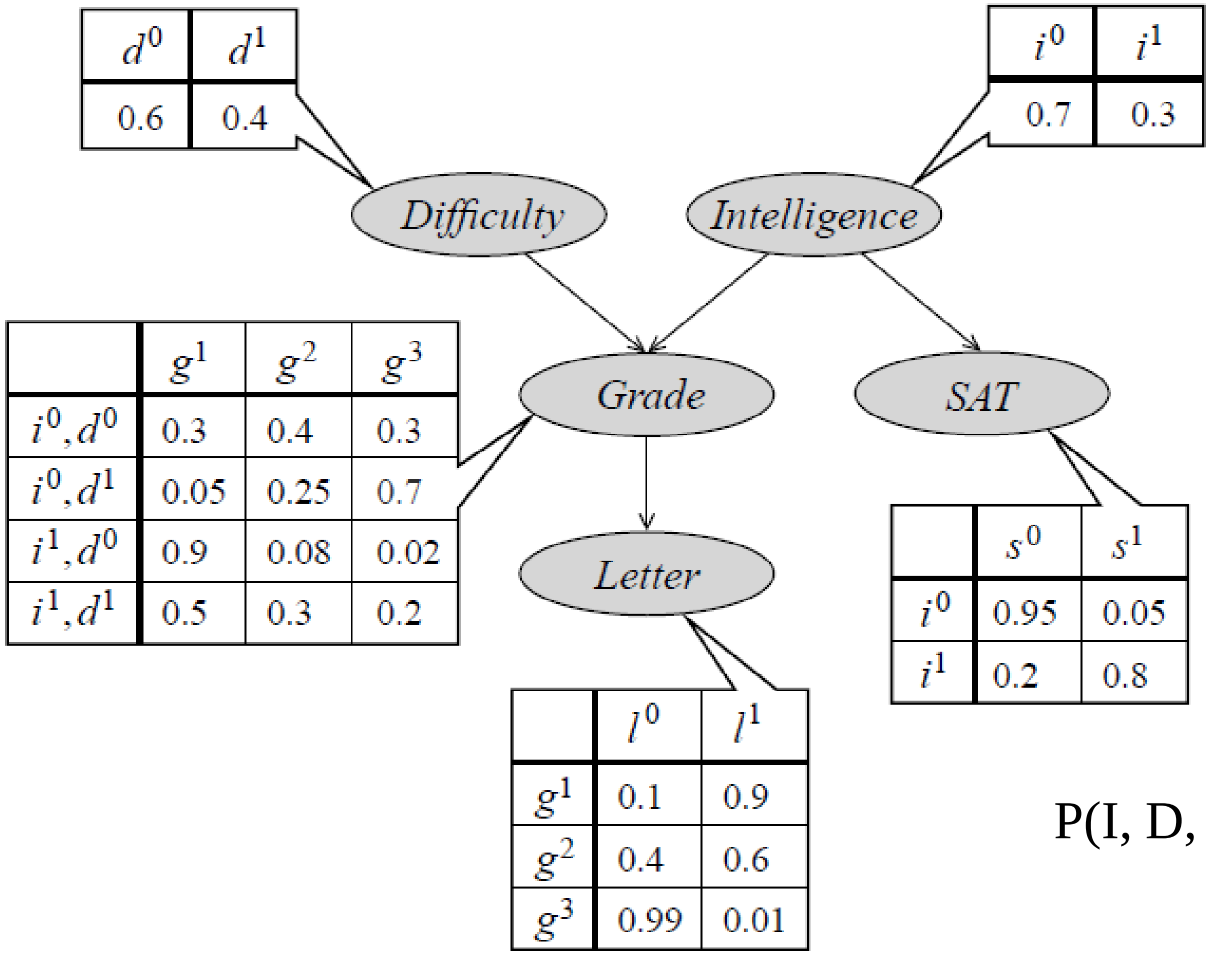
Bayesian networks

- The core of the Bayesian network representation is a directed acyclic graph (DAG) G
  - Nodes: random variables
  - Edges: direct influence of one node on another



- Two different views on graph G
  - A graphical representation of a joint distribution compactly in a factorized way
  - compact representation for a set of conditional independence assumptions about a distribution.

Two views are, in a strong sense, equivalent!



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(I, D, G, S, L) = P(I) P(D) P(G|I, D) P(S|I) P(L|G)

chain rule for Bayesian networks

$$P(i^0, d^0, g^2, s^1, l^0) = ?$$

Question: Prove P(I, D, G, S, L) is a legal distribution. You need to show

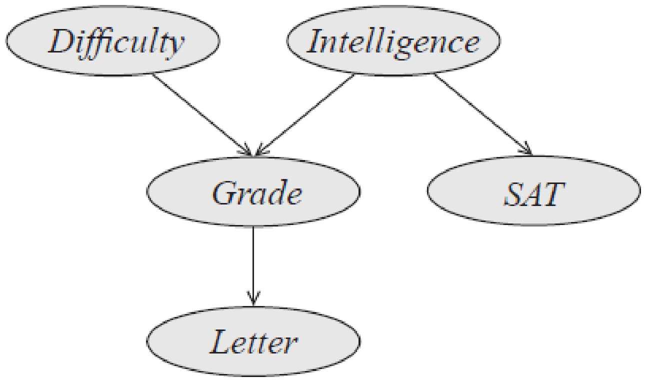
a) $P \geq 0$, b) $\sum_{I,D,G,S,L} P(I,D,G,S,L) = 1$

BN - Local independencies

- Bayesian network G encodes the following set of local independencies for each variable X_i

$$X_i \perp \mathrm{NonDescendants}_{\,X_i} \mid \mathrm{Pa_G(X_i)}$$

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



Question: write down local independencies for other nodes in the network.

BN - Factorization

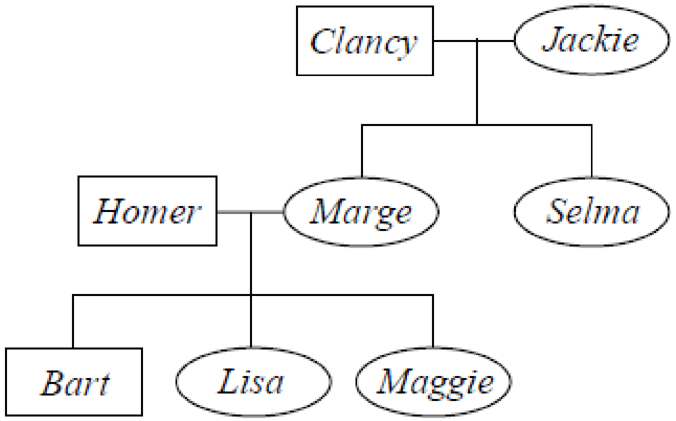
- Each node X_i is assigned to a conditional probability distribution (CPD) or a local probability distribution $P(X_i \mid \mathrm{Par}_G(X_i))$

- The BN represents a joint distribution via the chain rule

$$P(X_1, \ldots, X_n) = \prod_{i=1}^{n} P(X_i \mid \mathrm{Par}_G(X_i))$$

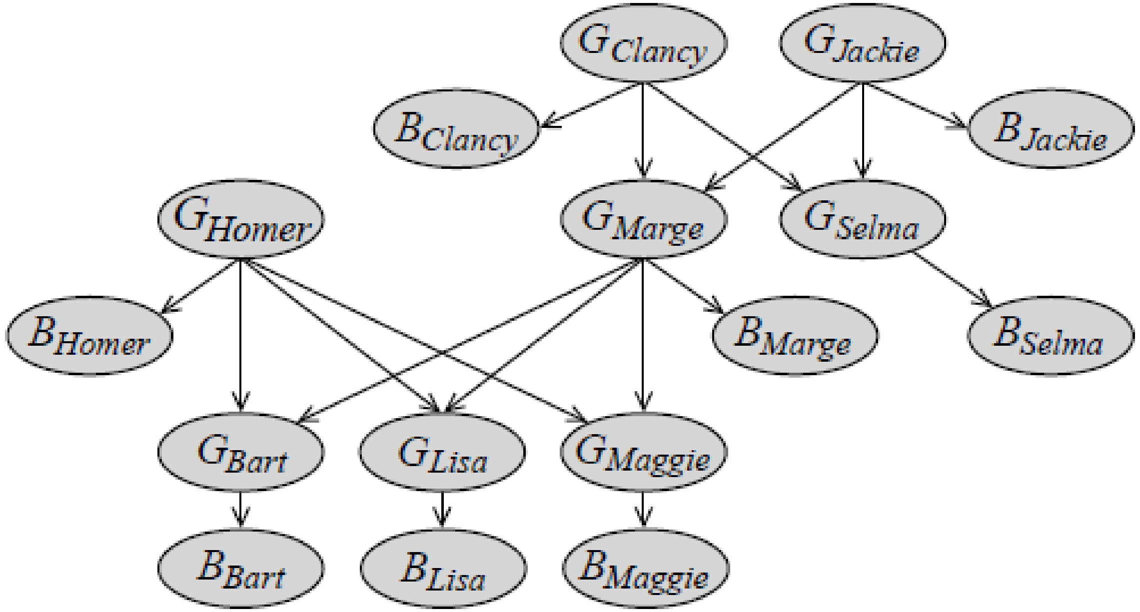
Case study- Modeling genetic inheritance

University of TEHRAN



Family tree

Parent Genes	AA	BB	AB	OO	AO	BO
AA	AA	AB	AA, AB	AO	AA, AO	AB, AO
BB	AB	BB	AB, BB	BO	AB, BO	BB, BO
AB	AA, AB	AB, BB	AA, BB, AB	AO, BO	AA, AO, BO, AB	BB, BO, AB, AO
OO	AO	BO	AO, BO	OO	AO, OO	BO, OO
AO	AA, AO	AB, BO	AA, AB, AO, BO	AO, OO	AA, AO, OO	AO, BO, AB, OO
BO	AO, AB	BB, BO	AB, AO, BB, BO	BO, OO	AB, AO, BO, OO	BB, BO, OO



Genetic inheritance

G: genotype
B: blood type

$$P(B_p \mid G_p), P(G_p \mid G_m, G_f)?$$

Graphs and distributions

- Let *P* be a distribution over some random variables.
- ***I(P)*** : the set of independence assertions of the form $X \perp Y \mid Z$ that hold in *P*.

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

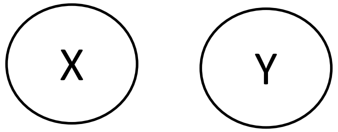
$I(P_1) = \{X \perp Y\}$

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

$I(P_2) = \{\}$

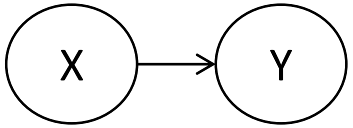
- ***I(G)***: the set of independence assertions in graph *G*

G0:



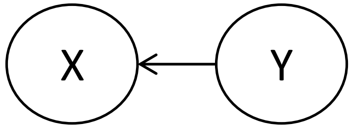
I(G0)={X⊥Y}

G1:



I(G1)={}

G2:



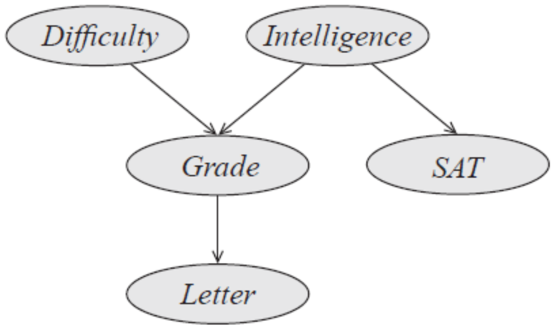
I(G2)={}

- *G* is an I-map of *P* if $I(G) \subseteq I(P)$
  - P may have additional independencies that are not reflected in *G*
  - *G0, G1, G2 are I-maps of P1. Only G1 and G2 are I-maps of P2.*

I-Map to Factorization

- **Theorem:** if G is an I-map for P, then P factorizes according to G.

- Example:
  - Without any assumption, one can write the joint distribution



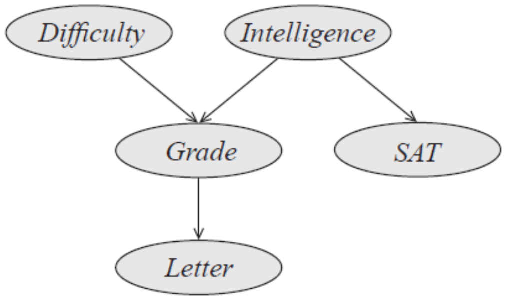
$$P(I,D,G,L,S) = P(I)P(D \mid I)P(G \mid I,D)P(L \mid I,D,G)P(S \mid I,D,G,L)$$

By the conditional independences in the graph we obtain

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

Factorization to I-Map

- **Theorem:** if *P* factorizes according to *G*, then *G* is an I-map for *P*.



- Given

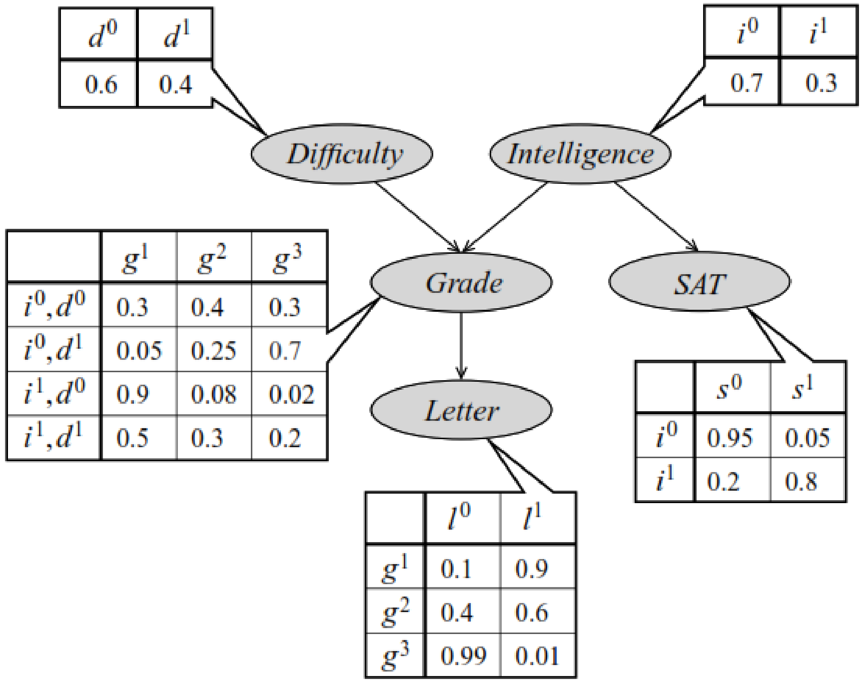
$$P(I, D, G, L, S) = P(I)\, P(D)\, P(G \mid I, D)\, P(L \mid G)\, P(S \mid I)$$

- We want to prove independencies in *G* exists in *P.* For example we want to prove $S \perp D, G, L \mid I$; we need to show

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

Question: prove the above statement.

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

causal reasoning or prediction: predicting the downstream effects of various factors or causes.

Causal reasoning: How likely George is to get a strong recommendation letter from his professor in Econ101

Case 1: knowing nothing about George and Econ101

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

Case 2: knowing George is not so intelligent

$$P(l^1 \mid i^0) = 0.389$$

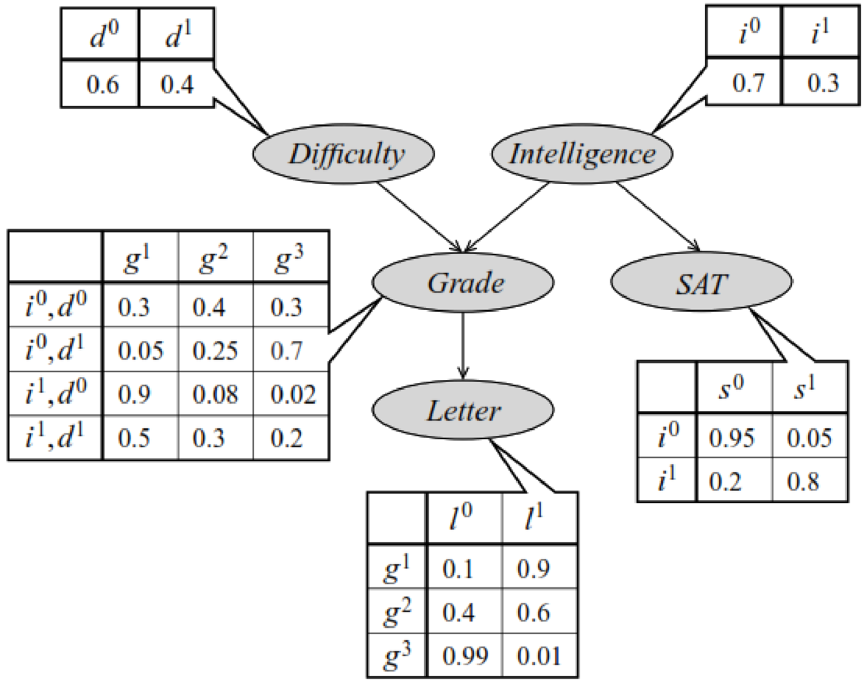
Causal reasoning

Case 3: additionally knowing Econ101 is an easy course

$$P(l^1 \mid i^0, d^0) = 0.513$$

Causal reasoning

Reasoning Patterns



evidential reasoning: reasoning from effects to causes.

Intercausal reasoning: interaction of different causes of the same effect. **Explaining away** is a special type of this reasoning.

Query: a recruiter is trying to decide whether to hire George

Case 1: knowing nothing about George and Econ101

$$P(i^1) = 0.30$$

Case 2: knowing George received grade *C*

$$P(i^1 \mid g^3) = 0.079$$

evidential reasoning

Case 3: knowing George received grade *C* & the letter is weak

$$P(i^1 \mid g^3, l^0) = 0.079$$

evidential reasoning

Case 4: knowing George received grade *C* & the course is difficult

$$P(i^1 \mid g^3, d^1) = 0.11$$

Intercausal reasoning

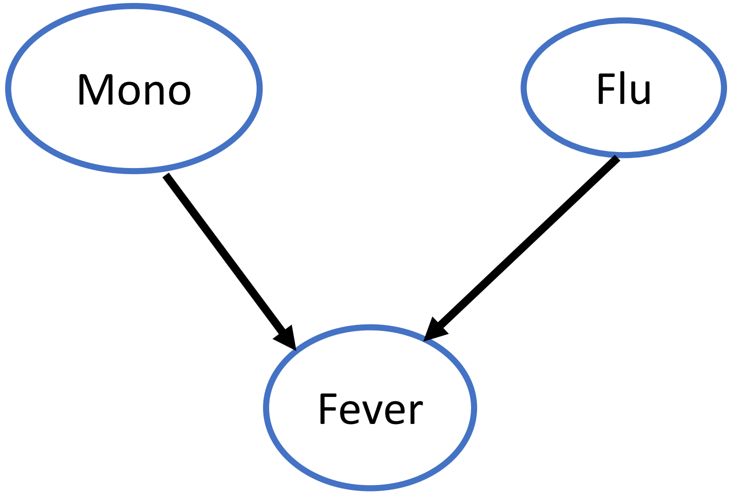
Case 5: knowing George received grade *C* & his SAT score is high

$$P(i^1 \mid g^3, s^1) = 0.578$$

evidential reasoning?

Explaining away-another example

- Common in human reasoning

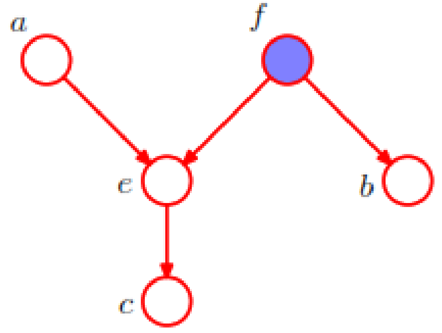


- Having flu reduces the probability of having mononucleosis

Independencies in Graphs-introduction

Independencies in Graphs

- Question: In a general DAG, is *X* dependence of Y given Z?
  - In the following DAG, are *a* and *b* independent given *f*?

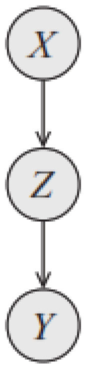


- D-separation
  - to check whether an independence assertion $X \perp Y \mid Z$ is implied by a given DAG.

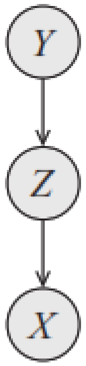
Independencies in Graphs- simple examples

- Direct connection
  - $X \rightarrow Y$ or $Y \rightarrow X$

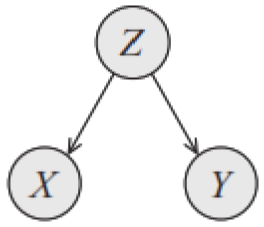
- Three-node networks



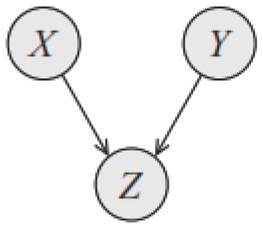
Indirect causal effect



Indirect evidential effect



common cause



common effect
or v-structure

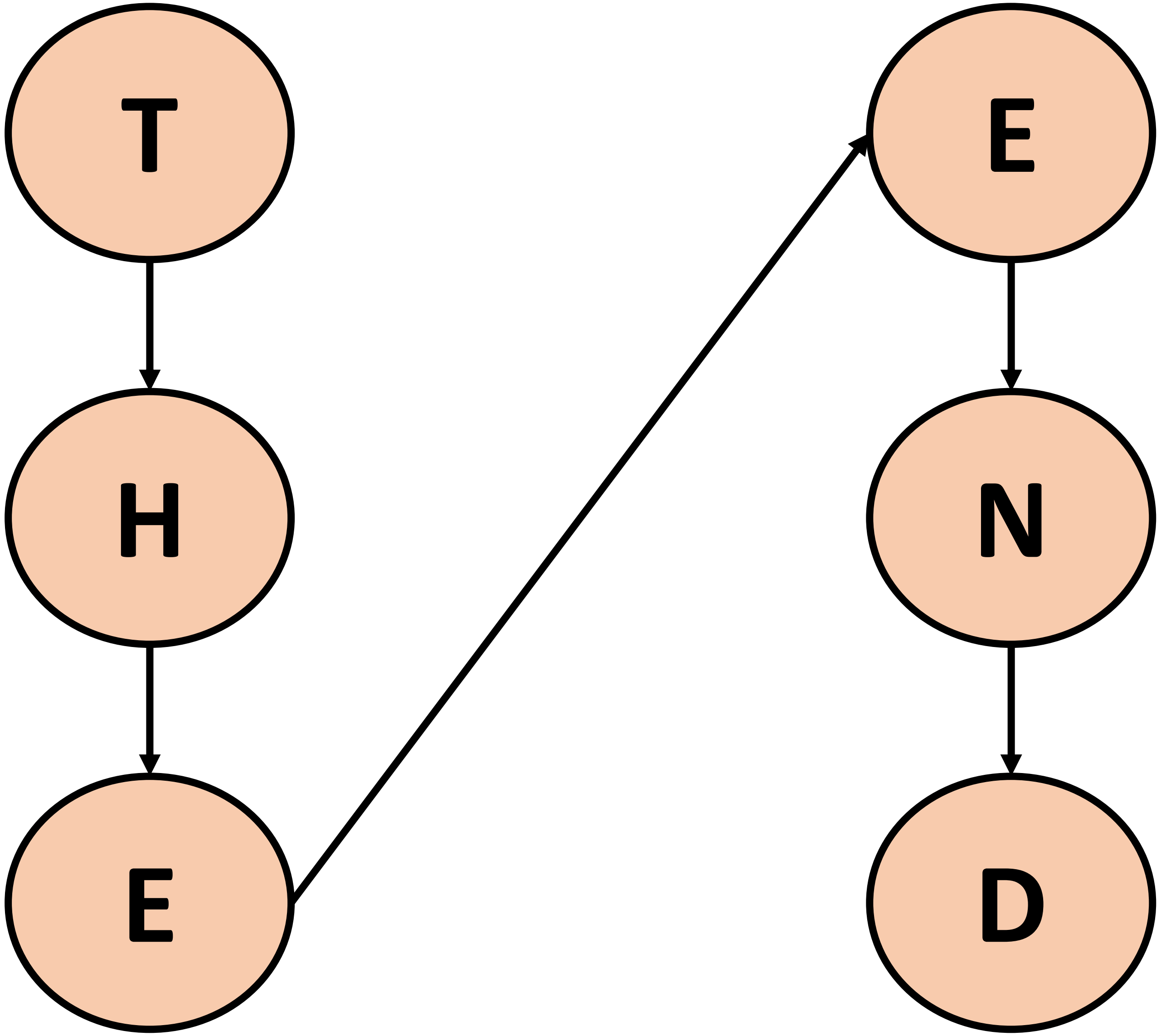
Not independence symbol
in PowerPoint

$X \perp Y \mid Z$

$!\,(X \perp Y)$

$!\,(X \perp Y \mid Z)$

$X \perp Y$


T
H
E
E
N
D