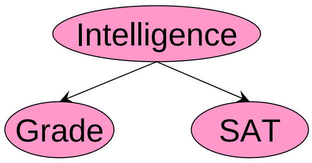
Mathematical Tools: Probabilistic Models

Lecture 3

Instructor: Su-In Lee

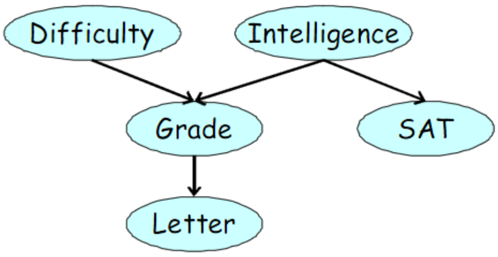
Bayesian network

- Directed acyclic graph (DAG)
  - Node: a random variable
  - Edge: direct influence of one node on another
- The *Student* example revisited
  - Intelligence (I), SAT (S), Grade (G)
  - Val (I) = $\{i^1, i^0\}$, Val (S) = $\{s^1, s^0\}$, Val (G) = $\{g^1, g^2, g^3\}$



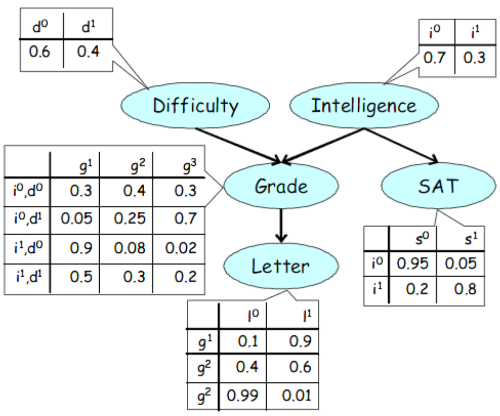
The *Student* example revisited

- More complex scenario
  - Course difficulty (D), quality of the recommendation letter (L), Intelligence (I), SAT (S), Grade (G)
  - Val(D) = {easy, hard}, Val(L) = {strong, weak}, Val (I) = $\{i^1, i^0\}$, Val (S) = $\{s^1, s^0\}$, Val (G) = $\{g^1, g^2, g^3\}$
  - Joint distribution requires 48 entries



The *Student* Bayesian network

- Joint distribution
  - P(I,D,G,S,L) = P(I)P(D)P(G|I,D)P(S|I)P(L|G)



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^2	0.99	0.01

[Koller and Friedman]

Basic independencies

- Given G, L is independent of I, D, S.
- Given I, S is independent of D, G, L.
- I and D are independent
- D is independent of I and S

  :

- The intuition
  - The parents of a variable “shield” it from probabilistic influence that is causal in nature.

Bayesian network semantics

- **Definition:** A Bayesian network structure G is a directed acyclic graph whose nodes represent random variables $X_1,\ldots,X_n$.

  PaX_i: parents of X_i in G

  NonDescendantsX_i: the variables in G that are not descendants of X_i.

  Then, G encodes the following set of conditional independence assumptions, called the local Markov assumptions, and denoted by $I_L(G)$:

  For each variable X_i:

[Koller and Friedman]

Bayesian network joint distribution

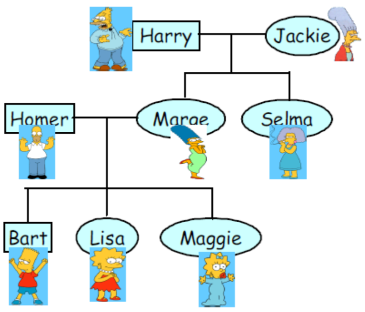
- **Definition:** Let G be a Bayesian network graph over the variables $X_1,\ldots,X_n$. We say that a distribution P over the sample space factorizes according to G if P can be expressed as a product:

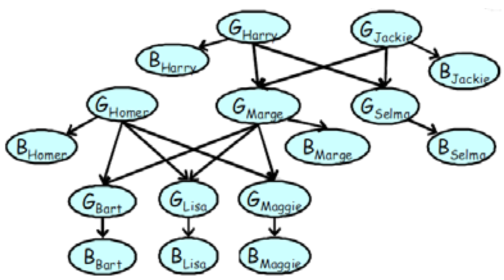
- **Definition:** A Bayesian network is a pair (G,P) where P *factorizes* over G, and where P is specified as a set of CPDs associated with G's nodes.

[Koller and Friedman]

The *Genetics* example

- Variables
  - B: blood type (a phenotype)
  - G: genotype of the gene that encodes a person's blood type; <A,A>, <A,B>, <A,O>, <B,B>, <B,O>,…





[Koller and Friedman]

The *Genetics* example

- CPDs
  - The penetrance model P (B(c) | C(c))
  - The transmission model P (G(c) | G(p),G(m))
  - Genotype priors P (G(c))

Acknowledgement

- This set of slides is based on the following materials:
  - "Probabilistic Graphical Models: Principles and Techniques" by Daphne Koller and Nir Friedman