CSE 250A Lecture 3.

Notation :

* Joint distribution $P(X_1 = x_1, X_2 = x_2 \ldots X_n = x_n)$ involves $O(2^n)$ numbers for binary random variables

* More Compact representations
  More efficient Algorithms

Example :

* Binary Variable : Burglary ? → B
  Earthquake ? → E
  Alarm ? → A
  John Calls ? → J
  Mary Calls ? → M

* Joint Distribution :

$$P(E, B, A, J, M) = P(B)\,P(E|B)\,P(A|B,E)\,P(J|B,E,A)\,P(M|J,B,E,A)$$

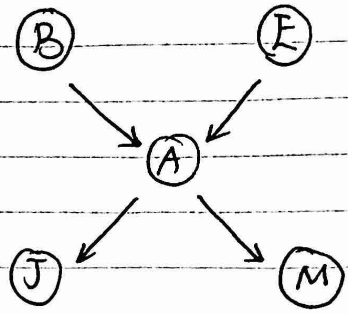
* Domain-Specific Assumption of (Marginal, Conditional) Independence.

$$P(B,E,A,J,M) = P(B)P(E)\,P(A|B,E)\,P(J|A)\,P(M|A)$$

J independence from B, E.

M independent from J, B. E.

* Direct Acyclic Graph (DAG)



$P(B=1)=0.001$ $P(E=1)=0.002$

B	E	$P(A=1 \mid B,E)$
0	0	0.001
1	0	0.94
0	1	0.29
1	1	0.95

A	$P(J=1 \mid A)$
0	0.05
1	0.9

A	$P(M=1 \mid A)$
0	0.01
1	0.7

These are called anditional probability table (CPT)

Joint probability are easy to compute :

$$P(B=1, E=0, A=1, J=1, M=0)$$

$$= P(B=1)P(E=0)P(A=1 \mid B=1, E=0)P(J=1 \mid A=1)P(M=0 \mid A=1)$$

$$= 0.001 \times 0.998 \times 0.94 \times 0.9 \times 0.3.$$

* Any "query" can be answered from joint distribution !

e.g. $$P(B=1, E=0 \mid M=1) = \frac{P(B=1, E=0, M=1)}{P(M=1)}$$

$$= \frac{\sum_{a,j} P(B=1, E=0, M=1, A=a, J=j)}{\sum_{b,j,e,a} P(M=1, B=b, E=e, A=a, J=j)}$$

* More efficient Algorithm?

Exploit structure of DAG (ordering of nodes, missing edges $\longleftrightarrow$ conditional independence)

$\boxed{\text{Belief Network (BNs)}}$

A Belief Network is a DAG which

1. Nodes $\longrightarrow$ random variables
2. Edges $\longrightarrow$ conditional dependence.
3. Tables (CPTs) $\longrightarrow$ how each node depends on parents.

* Conditional Independence

Generally True that $P(X_1, \dots X_n) = P(X_1)P(X_2|X_1) \cdots P(X_n|X_1, X_2 \dots)$

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

In a particular domain, suppose that:

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

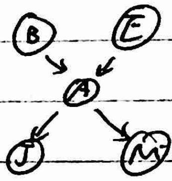
where $\text{parent}(X_i) \in \{X_1, X_2, \dots X_{i-1}\}$. $(*)$

* Big Idea: represent conditional independence by DAG

Constructing BN

(1) choose random variables
(2) choose ordering
(3) while there're variables left :
  (a) Add X_i to BN
  (b) Set parents of X_i to be minimal subset satisfying
  (c) Define CPT $P(X_i \mid parent(X_i))$

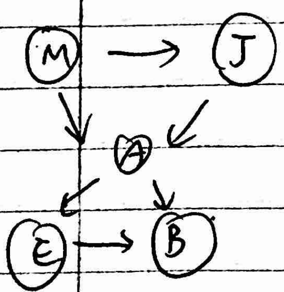
* ~~Note~~ Node ordering :
- Best ordering is to add "root causes", then variable they influence, so on.
- Ex : $\{B, E, J, A, M\}$.



- Ex : wrong ordering $\{M, J, A, E, B\}$.

$$P(M,J,A,E,B) = P(M)P(J|M)P(A|M,J)P(E|M,A,J)P(B|M,J,A,E)$$

$$= P(M)P(J|M)P(A|J,M)\cancel{P(B|J,A,M,E)}P(E|A)\cdot P(B|A,E)$$



Note : - 2 extra edges in BN
- Some independence not obvious in DAG
- more numbers to estimate for CPTs
- more difficult CPTs to access.

* Advantage of BNs :
  - complete , self - consistent , compact , non-redundant , representation of joint distribution .

    Ex : for n binery variables, if k is max # of parents of nodes in DAG (also called in-degree)

    then $O(n2^k)$ to represent joint distribution.
    versus $O(2^n)$

  - clean separation of qualitetive & quantetive knowledge.

  - DAG encodes conditional independence
  - CPTs encodes numerical influences.