Review

- Alarm example

  Binary random variables

  B = burglary

  E = earthquake

  A = alarm

  J = John calls

  M = Mary calls

- Probability captures common sense patterns of reasoning

  (1) Explaining away: $P(B=1|A=1) > P(B=1)$

  but $P(B=1|A=1, E=1) < P(B=1|A=1)$

  (2) Conflicting "symptoms": $P(A=1|J=1) > P(A=1)$

  but $P(A=1|J=1, M=0) < P(A=1|J=1)$

  (3) Intervening events (later today)

Today - from probabilities to graphs

Motivation

- Joint distribution $P(X_1 = x_1, X_2 = x_2, \ldots, X_n = x_n)$

  $O(2^n)$ numbers for n binary random variables

- more compact representations
- more efficient algorithms for inference

Alarm Example

- Joint distribution

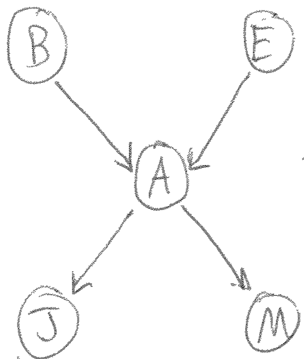
  $P(B,E,A,J,M) = P(B)P(E|B)P(A|B,E)P(J|B,E,A)P(M|B,E,A,J)$

- Conditional independence

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

Directed acyclic graph (DAG)

$P(B=1)=0.001$



$P(E=1)=0.002$

- Conditional probability tables (CPTs)

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

- Joint distribution

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

$= 0.001 \times (1-0.002) \times 0.94 \times 0.9 \times (1-0.7) \approx 0.00025$

- Any query can be answered from joint distribution

3) Reasoning about intervening events

Want to compare:
1. $P(A=1) = 0.00252$
2. $P(A=1 \mid J=1) = 0.0435$
3. $P(A=1 \mid J=1, B=1) = ?$

Brute force calculation

- From product rule: $P(A=1|B=1, J=1) = \dfrac{P(A=1, B=1, J=1)}{P(B=1, J=1)}$

- From marginalization:

numerator: $P(A=1, B=1, J=1) = \sum_{e,m} P(A=1, B=1, J=1, E=e, M=m)$

denominator: $P(B=1, J=1) = \sum_{a,e,m} P(B=1, J=1, A=a, E=e, M=m)$

More efficient algorithm – Exploit structure of DAG

$$P(A=1|B=1, J=1) = \frac{P(J=1|A=1, B=1)\, P(A=1|B=1)}{P(J=1|B=1)} \quad \text{(conditionalized Bayes rule)}$$

$$= \frac{\overset{0.9}{P(J=1|A=1)}\, \overset{\text{from last lecture } (0.94002)}{P(A=1|B=1)}}{P(J=1|B=1)} \quad \text{(cond. ind.)}$$

Denominator:
$$P(J=1|B=1) = \sum_a P(J=1, A=a|B=1) \quad \text{(cond. marginalization)}$$

$$= \sum_a P(A=a|B=1)P(J=1|A=a, B=1) \quad \text{(product rule)}$$

$$= \sum_a \underbrace{P(A=a|B=1)}_{\text{computed last time}} \underbrace{P(J=1|A=a)}_{\text{CPT}}$$

$$= 0.849$$

$\Rightarrow P(A=1|J=1, B=1) = \underline{0.9965} \rightarrow 3$

Now compare:

(1) $P(A=1) \qquad = 0.00252$

(2) $P(A=1 \mid J=1) \qquad = 0.0435 \uparrow$

(3) $P(A=1 \mid J=1, B=1) \approx 0.9965 \uparrow\uparrow$

So $\quad P(A=1) < P(A=1 \mid J=1) \ll P(A=1 \mid J=1, B=1)$

Also note $\quad P(A=1 \mid J=1, B=1) > P(A=1 \mid B=1)$ (0.94002)

Belief Network (BN)

A BN is a DAG

(1) nodes represent random variables

(2) edges represent conditional dependencies

(3) CPTs describe how each node depends on parents.

- Conditional independence

Generally true that

$$P(X_1, X_2, \dots, X_n) = P(X_1)P(X_2 \mid X_1)P(X_3 \mid X_1, X_2) \cdots P(X_n \mid X_1, X_2, \dots, X_{n-1})$$
$$= \prod_{i=1}^{n} P(X_i \mid X_1, X_2, \dots, X_{i-1}) \qquad (\star)$$

In a given domain, suppose that

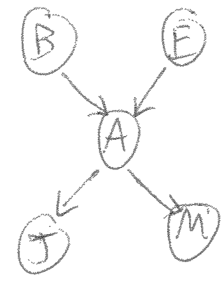
$$P(X_1, X_2, \dots, X_n) = \prod_i P(X_i \mid \text{parents}(X_i)) \qquad (\star\star)$$

where $\text{parents}(X_i)$ is some subset of $\{X_1, X_2, \dots, X_{i-1}\}$

Big idea: represent dependence relations by a DAG.

- How to construct a BN?

Ex: Alarm example
$\{B, E, A, J, M\}$



① choose random variables

② choose ordering

③ while there are variables left:

(a) add node X_i

(b) set the parents of X_i to the minimal subset satisfying (**)

(c) define CPT $P(X_i \mid pa(X_i))$

* advantages: (1) complete, compact, consistent representation of joint distribution

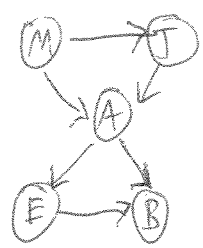
Ex: for binary variables, if k = max # of parents of graph, then:
$O(n \cdot 2^k)$ numbers will appear in CPTs vs.
$O(2^n)$ for joint distribution

(2) Clean separation of qualitative and quantitative knowledge

DAG encodes conditional independence
CPTs encode numerical influences

- Node ordering
  - Best order is to add "root" causes, then the variables they influence, and so on....
  - From misordered graph,
    *conditional independences in world not obvious

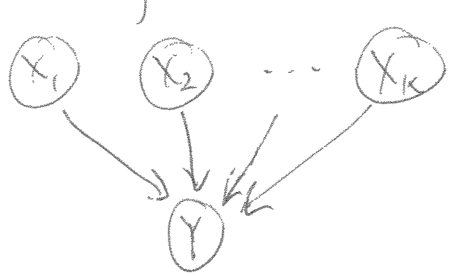
Ex: wrong ordering $\{M, J, A, E, B\}$



two additional edges!

- more numbers (larger CPTs) to specify the same joint distribution
- less natural (more difficult) to assess the CPTs or learn CPT from data

• Representing CPTs



for simplicity, assume $X_i \in \{0,1\}$, $Y \in \{0,1\}$

How to represent $P(Y=1 \mid X_1, X_2, \ldots, X_K)$?