Review 6/30

* Probabilities

$P(X)$ unconditional
$P(Y|X)$ conditional
$P(X,Y)$ joint

* Rules

$P(A,B,C,\ldots) = P(A)\,P(B|A)\,P(C|A,B)\ldots$ product rule

$P(X|Y) = \dfrac{P(Y|X)\,P(X)}{P(Y)}$ Bayes rule

$P(X) = \sum_y P(X, Y=y)$ marginalization

* Conditionalized versions

$P(A,B,C,\ldots|E) = P(A|E)\,P(B|A,E)\,P(C|A,B,E)\ldots$

$P(X|Y,E) = P(Y|X,E)\,P(X|E)\,/\,P(Y|E)$

$P(X|E) = \sum_y P(X, Y=y \mid E)$

* Independence

$$\left.\begin{aligned} P(X,Y) &= P(X)\,P(Y) \\ P(X|Y) &= P(X) \\ P(Y|X) &= P(Y) \end{aligned}\right\}$$ each of these implies the others

* Conditional independence

$$P(X, Y \mid E) = P(X \mid E)\, P(Y \mid E)$$

$$P(X \mid Y, E) = P(X \mid E)$$

$$P(Y \mid X, E) = P(Y \mid E)$$

each of these implies the others

Probabilistic inference

- Today: do probabilities capture patterns of common sence reasoning?

Examples – reasoning about:

1) multiple explanations of a single event
2) multiple events - common explanation
3) Chain of intervening events

* Binary random variables

B = burglary?
E = earthquake?
A = alarm?

* Joint distribution.

$$P(B, E, A) = P(B)\, P(E \mid B)\, P(A \mid B, E)$$

product rule

* Domain knowledge – what we're assuming to be true

- $P(B=1) = 0.001 \longrightarrow P(B=0) = 1 - P(B=1) = 0.999$

- $\left.\begin{array}{l} P(E=1 \mid B=0) = 0.002 \\ P(E=1 \mid B=1) = 0.002 \end{array}\right\} P(E=1 \mid B) = P(E=1) = 0.002$

Assumption: B & E are independent random variables

-

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

1) reasoning about multiple explanations.

Compare: $P(B=1) = 0.001$

$P(B=1 \mid A=1) = ?$

$P(B=1 \mid A=1, E=1) = ?$

Bayes rule

$$P(B=1 \mid A=1) = \frac{P(A=1 \mid B=1)\ \overbrace{P(B=1)}^{0.001}}{P(A=1)}$$

Term in denominator :

$$P(A=1) = \sum_{\substack{b \in \{0,1\} \\ e \in \{0,1\}}} P(B=b, E=e, A=1) \quad \text{marginalization}$$

productrule

$$= \sum_{b,e} P(B=b)\, P(E=e \mid B=b)\, P(A=1 \mid E=e, B=b)$$

$$= \sum_{b,e} P(B=b)\, P(E=e)\, P(A=1 \mid B=b, E=e)$$

independence

$$\begin{aligned} = \; & P(B=0)\, P(E=0)\, P(A=1 \mid B=0, E=0) \\ + \; & P(B=1)\, P(E=0)\, P(A=1 \mid B=1, E=0) \\ + \; & P(B=0)\, P(E=1)\, P(A=1 \mid B=0, E=1) \\ + \; & P(B=1)\, P(E=1)\, P(A=1 \mid B=1, E=1) \end{aligned}$$

$$= 0.00252$$

Term in numerator

$$P(A=1 \mid B=1) = \sum_{e \in \{0,1\}} P(A=1, E=e \mid B=1)$$

conditionalized marginali

$$= \sum_{e} P(A=1 \mid E=e, B=1)\, P(E=e \mid B=1)$$

conditionalized product rule

$$= \sum_{e} P(A=1 \mid E=e, B=1)\, P(E=e) \quad \text{independence}$$

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

$= 0.94002$

$$\therefore\ P(B=1 \mid A=1) = \frac{\overset{0.94002}{P(A=1 \mid B=1)}\ \overset{0.001}{P(B=1)}}{\underset{0.00252}{P(A=1)}}$$

$$= 0.37$$

Compare to $P(B=1) = 0.001$

What about $P(B=1 \mid A=1, E=1)$?

* Conditionalized Bayes rule

$$P(B=1 \mid A=1, E=1) = \frac{P(A=1 \mid B=1, E=1)\ P(B=1 \mid E=1)}{P(A=1 \mid E=1)}$$

$$= \frac{0.95 \times 0.001}{P(A=1 \mid E=1) \leftarrow 0.29} = 0.0033$$

Term in denominator

$$P(A=1 \mid E=1) = \sum_b P(A=1, B=b \mid E=1) \quad \text{conditionalized marginalization}$$

$= \ldots$ conditional product rule

$= \ldots$ independence

$$= \sum P(A=1 \mid B=b, E=1)\ P(B=b) = 0.29$$

Summary

$P(B=1) = 0.001$

$P(B=1 \mid A=1) = 0.37$ $\uparrow$

non-monotonic

$P(B=1 \mid A=1, E=1) = 0.0033$ $\downarrow$

$\Rightarrow$ earthquake "explains away" the alarm, decreasing our belief in burglary.

Arises from multiple (causal) explanations of observed event.

2) Multiple events with a common explanation.

Two more binary random variables.

J = John calls?
M = Mary calls?

Assumptions:

$$\left.\begin{aligned} P(J \mid A) &= P(J \mid A, B, E) \\ P(M \mid A) &= P(M \mid A, B, E, J) \end{aligned}\right\} \text{conditional independence}$$

* Joint distribution

Product ru

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

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

conditional indepe

* Conditional probabilities (domain knowledge)

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

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

Compare $P(A=1) = 0.00252$ (from previous example)

$P(A=1 \mid J=1) = ?$ $\quad 0.0435$ ↑

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

* Bayes rule

$$P(A=1 \mid J=1) = \cancel{P(J=1)\,P} \; \frac{\overset{0.9}{P(J=1 \mid A=1)}\;\overset{0.00252}{P(A=1)}}{\underset{0.0521}{P(J=1)}}$$

Term in denominator:

$$P(J=1) = \sum_a P(A=a, J=1) \quad \text{marginalization}$$

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

$$= P(A=1)\, P(J=1 \mid A=1) + P(A=0)\, P(J=1 \mid A=0)$$

$$= (0.00252)(0.9) + (1-0.00252)(0.05)$$

$$= 0.0521$$

$\therefore \quad P(A=1 \mid J=1) = 0.0435$

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

* Bayes rule with multiple pieces of evidence

$$P(A=1 \mid J=1, M=0) = \frac{P(J=1, M=0 \mid A=1)\, P(A=1)}{P(J=1, M=0)}$$

$$= \frac{\overset{0.9}{P(J=1 \mid A=1)}\ \overset{1-0.7}{P(M=0 \mid A=1)}\ \overset{0.00252}{P(A=1)}}{\underset{0.05}{P(J=1, M=0)}}$$

(b/c J & M are conditionally ind (C.I.) given A)

Term in denominator

$$P(J=1, M=0) = \sum_a P(A=a, J=1, M=0) \quad \text{marginalization}$$

$$= \sum_a P(A=a)\, P(J=1 \mid A=a)\, P(M=0 \mid A=a, J=1)$$

product rule

$$= \sum_a P(A=A)\, P(J=1 \mid A=a)\, P(M=0 \mid A=a)$$

conditional ind.

$$= 0.05$$

$$P(A=1 \mid J=1, M=0) = \frac{(0.9)(1-0.7)(0.00252)}{0.05}$$

$$= 0.0136$$

$P(A=1) = 0.00252$

$P(A=1 \mid J=1) = 0.0435 \uparrow$ reproduces common sense.

$P(A=1 \mid J=1, M=0) = 0.0136 \downarrow$ non-monotonic

3) Reasoning about intervening events.

Compare $P(A=1) = 0.00252$

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

$P(A=1 \mid J=1, B=1) = ?$

$$P(A=1 \mid J=1, B=1) = \frac{P(J=1 \mid A=1, B=1)\, P(A=1 \mid B=1)}{P(J=1 \mid B=1)}$$

conditionalized Bayes rule

$$= \frac{P(J=1 \mid A=1)\, P(A=1 \mid B=1)}{P(J=1 \mid B=1)}$$ conditional independence

$$= \frac{(0.9)\,(0.94002)}{(\quad\quad)}$$ from example #1

Denominator:

$$P(J=1 \mid B=1) = \sum_a P(A=a, J=1 \mid B=1)$$ conditionalized marginalization

$$= \sum_a P(A=a \mid B=1)\, P(J=1 \mid A=a, B=1)$$ conditionalized product rule

$= \sum_a P(A=a \mid B=1)\, P(J=1 \mid A=a)$ conditional independence

$=$ plug & chug ?

$= 0.849$

$\therefore P(A=1 \mid J=1, B=1) = \dfrac{(0.9)(0.94002)}{(0.849)}$

$= \underline{\underline{0.9965}}$ ↑↑

—

Motivation

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

- More compact representations

- More efficient algorithms for inference ?

Alarm example

Binary variables $\{0, 1\}$

B burglary

E earthquake

A alarm

J John calls

M Mary calls

* Joint distribution

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

product rule

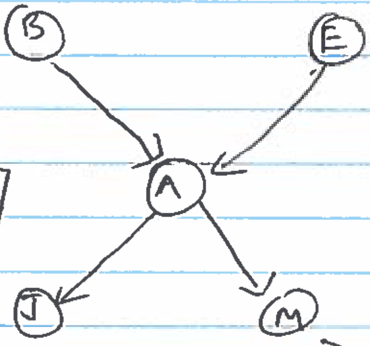
$= P(B)\,P(E)\,P(A|B,E)\,P(J|A)\,P(M|A)$

conditional independence

* Directed acyclic graph (DAG)

$P(B=1) = 0.001$

$P(E=1) = 0.002$



A	P(J=1|A)
0	0.05
1	0.9

B	E	P(A=1|B,E)
0	0	0.001
1	0	0.94
0	1	0.29
1	1	0.95

* Conditional probability tables (CPTs)

A	P(M=1|A)
0	0.01
1	0.7

* Joint probabilities

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

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

$= (0.001)(1-0.002)(0.94)(0.9)(0.7)$

* Any query can be answered from joint distribution

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

From product rule: $P(B=1, E=0 \mid M=1) = \dfrac{P(B=1, E=0, M=1)}{P(M=1)}$

From marginalization

numerator: $P(B=1, E=0, M=1) = \sum_{a,j} P(B=1, E=0, A=a, J=j, M=1)$

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

More efficient algorithms? Yes

Exploit structure of DAG.

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

* A BN is a DAG in which
  (i) nodes represent random variables
  (ii) edges represent conditional dependencies
  (iii) CPTs describe how each node depends on its parents.

* Conditional independence

Generally true in any domain that

product rule

$$P(X_1, X_2 \ldots X_n) = P(X_1)\, P(X_2 \mid X_1) \ldots P(X_n \mid X_1 \ldots X_{n-1})$$

$$= \prod_{i=1}^{n} P(X_i \mid X_1, X_2 \ldots X_{i-1})$$

In a given domain, suppose that

(*) $P(X_1 X_2 \ldots X_n) = \prod_{i=1}^{n} P(X_i \mid \text{parents}(X_i))$ where

parents (X_i) is some subset of $\{X_1, X_2, \ldots, X_{i-1}\}$.

BIG IDEA : represent dependencies by a graph!

* Constructing a BN :

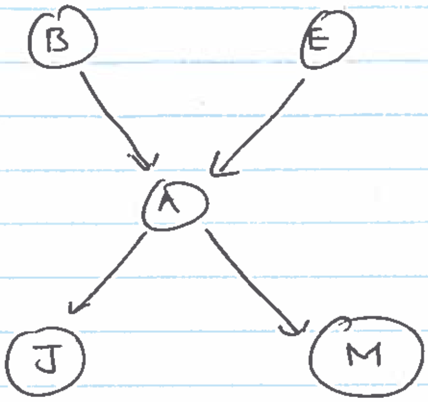
1) choose random variables

2) choose Ordering

3) while there are variables left :
   (a) add node X_i to the BN
   (b) set parents of X_i to the minimal subset satisfying (*).
   (c) define CPT $P(X_i \mid pa(X_i))$

parents of X_i

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



* advantages

- complete ~~model~~, compact, consistent representation of joint probs

Ex: for binary variables, if k = max # parent of any node (in-degree of graph),

then $O(n2^k)$ numbers will be needed to writ out CPTs,

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

if $k << n$, huge savings!

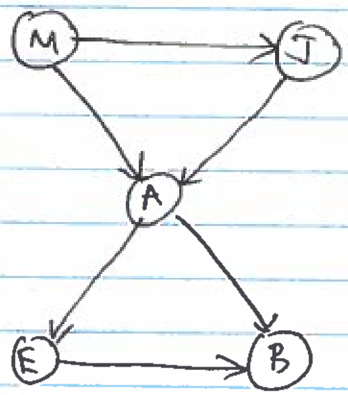
- clean separation of qualitative vs quantitativ knowledge.

DAG encodes conditional independencies
CPTs encode numerical influences

* Node ordering

- Best order is to add "root causes", then the variables they influence, and so on.

- Ex: "wrong order" {M, J, A, E, B}. What BN do we get?



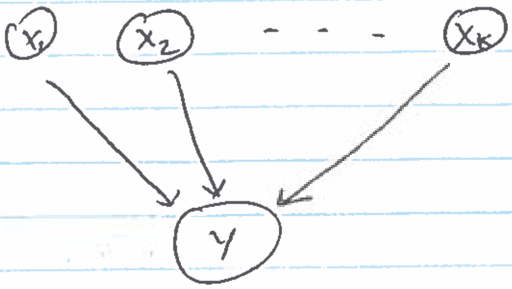
from misordered graph.

- (conditional) independences in world not obvious

- more numbers in CPT to specify same joint distribution.

- less natural, more difficult to assess or learn CPTs from data.

* Representing CPTs



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

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

<u>Possible answers</u> :

1) lookup table $O(2^k)$ can store arbitrary CPT

2^k rows

X_1	X_2	- -	X_k	$P(Y=1 \| X_1, X_2, \ldots X_k)$
0	0	- - -	0	0.6
1	0	- - -	0	0.2
⋮				⋮
1	1	1 1 1	1	0.5

What if k is too large