Why not use Predicate Calculus?

Eg: Consider diagnosing toothache:

1. $\forall p$ `Symptom(`p`, Toothache)` $\Rightarrow$ `Disease(`p`, Cavity)`

   Wrong – other factors cause toothaches:

2. $\forall p$ `Symptom(`p`, Toothache)` $\Rightarrow$ `Dis(`p`, Cavity)` $\vee$ `Dis(`p`, GumDisease)` $\vee$ `Dis(`p`, ImpactedWisdom)` $\vee \ldots$

   Too many! Maybe diagnostic:

3. $\forall p$ `Disease(`p`, Cavity)` $\Rightarrow$ `Symptom(`p`, Toothache)`

   Wrong – many other factors (on lhs)!

- Difficulties of Building Exhaustive KB

  Laziness: Just too many rules and contingencies.

  Theoretical Ignorance: No complete theory for the domain.

  Practical Ignorance: Don't have all the (patient) information available.

- **Probabilities** provide way of summarizing uncertainty from $\left\{\begin{array}{l}\text{laziness}\\ \text{ignorance}\end{array}\right\}$

Using Probability

- Not everyone with cavity has toothache

  $\neg$ [$\forall p$ `Disease(`p`, Cavity)` $\Rightarrow$ `Symptom(`p`, Toothache)`]

  but. . .
  perhaps 80% do.

- "80%" summarizes factors required for cavity to cause toothache

  + patient has cavity & toothache (but unrelated)

  Remaining 20% summarizes
  all other possible causes of toothache

Meaning: An individual with cavity either has toothache, or not.

In 80% of situations where x has `Cavity`
(ie, indistinguishable from this situation based on current knowledge)
x has `toothache`

Terms from Probability Theory

Random Variable:

`Weather` $\in$ {`Sunny, Rain, Cloudy, Snow` }

Domain: Possible values a random variable can take.

(. . . finite set, $\Re$, . . .)

Probability distribution: mapping from domain to values in $[0, 1]$

$$P(\texttt{Weather}) = \langle 0.7, 0.2, 0.08, 0.02\rangle$$

$$\text{means}\ \left\{\begin{array}{lcl} P(\texttt{Weather = Sunny}) & = & 0.7\\ P(\texttt{Weather = Rain}) & = & 0.2\\ P(\texttt{Weather = Cloudy}) & = & 0.08\\ P(\texttt{Weather = Snow}) & = & 0.02 \end{array}\right\}$$

Event: Each assignment
(eg, `Weather = Rain`)
is "event"

General Events

Boolean Combinations: Can have
Conjunction, Disjunction, Negation
of events:

$P($ `Weather = Rain` $\wedge$ `Card = 2S` $)$

$P($ `Weather = Rain` $\vee$ `Card = 2S` $)$

$P(\neg$`(Weather = Rain)` $)$

Atomic Event: "Complete specification"
Conjunction of assignments to EVERY variable

Joint Probability Distribution:
Probability of every possible atomic event

Toothache	Cavity	$P(\cdots)$
T	T	0.04
T	F	0.01
F	T	0.06
F	F	0.89

n binary variables: 2^n entries

($2^n - 1$ independent values, as sum = 1)

A huge table!

Joint Probability Distribution is Sufficient

t	c	$P(\texttt{Toothache} = t,\ \texttt{Cavity} = c)$
$+$	$+$	0.04
$+$	$-$	0.01
$-$	$+$	0.06
$-$	$-$	0.89

- $P(\texttt{Cavity} \vee \texttt{Toothache}) =$
  $P(\ \texttt{Cavity} \wedge\ \texttt{Toothache}) +$
  $P(\ \texttt{Cavity} \wedge \neg\texttt{Toothache}) +$
  $P(\neg\texttt{Cavity} \wedge\ \texttt{Toothache})$
  $= 0.04 + 0.01 + 0.06 = 0.11$

- $P(\texttt{Toothache}) =$
  $P(\texttt{Toothache}\wedge\texttt{Cavity}) + P(\texttt{Toothache}\wedge\neg\texttt{Cavity})$
  $= 0.04 + 0.01 = 0.05$

- Atomic Events are sufficient
  . . . but very unnatural

- Why not "connections"?

Conditional Probability

Conditional Probability: $P(A \mid B) =$
Probability of event A,
given that event B has happened.

$P(\texttt{Cavity} \mid \texttt{Toothache}) = 0.8$

In gen'l:
$$P(A \mid B) = \frac{P(A \wedge B)}{P(B)}$$
$$P(A \wedge B) = P(A \mid B)\, P(B)$$

Unconditional (prior) Probability: Probability of event *before evidence is presented*

$P(\texttt{Cavity}) = 0.01$
$\equiv$ probability that someone (from this population) has a cavity is 1 in 100

Evidence: Percepts that affects degree of belief in event

Conditional (posterior) Probability: Probability of event *after evidence is presented*

N.b., posterior probability can be COMPLETELY different than prior probability!

Bayes' Rule and Its Use

Diagnosis typically involves computing
$P(\texttt{Hypothesis} \mid \texttt{Symptoms})$

What is $P(\texttt{Meningitis} \mid \texttt{StiffNeck})$?
$\equiv$ prob that patient A has meningitis,
given that A has stiff neck?

Typically have . . .

- Prior prob of meningitis $P(\texttt{M}) = \frac{1}{50{,}000}$
- Prior prob of having a stiff neck $P(\texttt{SN}) = \frac{1}{20}$
- Prob that meningitis causes a stiff neck
  $P(\texttt{SN} \mid \texttt{M}) = \frac{1}{2}$

Bayes' Rule: $P(B \mid A) = \dfrac{P(A \mid B)\, P(B)}{P(A)}$

$$P(Y \mid X, E) = \frac{P(X \mid Y, E)\, P(Y \mid E)}{P(X \mid E)}$$

Eg: $P(\texttt{M} \mid \texttt{SN}) = \frac{P(\texttt{SN} \mid \texttt{M}) P(\texttt{M})}{P(\texttt{SN})} = \frac{0.5 \times 0.00002}{0.05} = 0.0002$

Note: Only 1 in 5000 stiff necks have meningitis. . .
even though SN is major symptom of M. . .

Comments on Bayes Rule

- Don't need $P(\texttt{SN})$ if have $P(\texttt{SN} \mid \neg\texttt{M})$:

$$\begin{aligned} P(\texttt{SN}) &= P(\texttt{SN},\ \texttt{M}) + P(\texttt{SN},\ \neg\texttt{M}) \\ &= P(\texttt{SN} \mid \texttt{M})\, P(\texttt{M}) + P(\texttt{SN} \mid \neg\texttt{M})\, P(\neg\texttt{M}) \\ P(\neg\texttt{M}) &= 1 - P(\texttt{M}) \end{aligned}$$

- Given "sore neck", want to compare

  prob of meningitis $P(\texttt{M} \mid \texttt{SN}) = \frac{P(\texttt{SN} \mid \texttt{M})\, P(\texttt{M})}{P(\texttt{SN})}$

  prob of whiplash $P(\texttt{W} \mid \texttt{SN}) = \frac{P(\texttt{SN} \mid \texttt{W})\, P(\texttt{W})}{P(\texttt{SN})}$

  To compute "relative likelihood", don't need $P(\texttt{SN})$:

  $\frac{P(\texttt{M} \mid \texttt{SN})}{P(\texttt{W} \mid \texttt{SN})} = \frac{P(\texttt{SN} \mid \texttt{M})\, P(\texttt{M})}{P(\texttt{SN} \mid \texttt{W})\, P(\texttt{W})} = \ldots$

- $P(Y \mid X) = \alpha\, P(X \mid Y)\, P(Y)$
  where $\alpha = \frac{1}{P(X)}$ is independent of Y.

Combining Evidence

- What is prob of Cavity, given {Toothache, Catch}?

  $P(\texttt{Cav} \mid \texttt{Ta}, \texttt{Ct})$

- Bayesian Update:

$$\begin{aligned} P(\texttt{Cav} \mid \{\}) &= P(\texttt{Cav}) \\ P(\texttt{Cav} \mid \texttt{Ta}) &= P(\texttt{Cav} \mid \{\}) \frac{P(\texttt{Ta} \mid \texttt{Cav})}{P(\texttt{Ta})} \\ P(\texttt{Cav} \mid \texttt{Ta}, \texttt{Ct}) &= P(\texttt{Cav} \mid \texttt{Ta}) \frac{P(\texttt{Ct} \mid \texttt{Ta}, \texttt{Cavity})}{P(\texttt{Ct} \mid \texttt{Ta})} \end{aligned}$$

  Each time new evidence is observed (Toothache; Catch; ...), belief in unknown (Cavity) is multiplied by factor that depends on new evidence.

  (Note: independent of order of observations)

Using Independence

Note: needs 3rd order information:
$P(\texttt{Ct} \mid \texttt{Ta}, \texttt{Cav})$

Not always available...

- But sometimes, INDEPENDENCE!

  $P(\texttt{Ct} \mid \texttt{Ta}, \texttt{Cav}) = P(\texttt{Ct} \mid \texttt{Cav})$

  (Prob of symptom2, given disease and symptom1 $\equiv$ Prob of symptom2, given disease)

  If so...

  $P(\texttt{Cav} \mid \texttt{Ta}, \texttt{Ct}) = P(\texttt{Cav}) \frac{P(\texttt{Ta} \mid \texttt{Cav})}{P(\texttt{Ta})} \frac{P(\texttt{Ct} \mid \texttt{Cav})}{P(\texttt{Ct} \mid \texttt{Ta})}$

- **ASSUMPTION is NOT ALWAYS TRUE!**

  But when it is, just need 2nd order statistics!

- Even better:
  Denominator is $P(\texttt{Ta}) P(\texttt{Ct} \mid \texttt{Ta}) = P(\texttt{Ta}, \texttt{Ct})$

  Independent of Cavity;
  just normalizing term!

Probability Theory

Axioms:

$$\boxed{\begin{array}{l} 0 \le P(A) \le 1 \\ P(\texttt{True}) = 1, \quad P(\texttt{False}) = 0 \\ P(A \vee B) = P(A) + P(B) - P(A \wedge B) \\ P(A) + P(\neg A) = 1 \end{array}}$$

Not arbitrary: If Agent1 assign prob that violate axioms, then
$\exists$ betting strategy s.t. Agent1 guaranteed to lose \$

Def'n: Independence of Variables:

Events A and B are independent $\Leftrightarrow$

$$\begin{aligned} P(A \wedge B) &= P(A) P(B) \\ P(A \mid B) &= P(A) \\ P(A \vee B) &= 1 - (1 - P(A))(1 - P(B)) \end{aligned}$$

Variables independent
$\Leftrightarrow$ independent for all values

$$\forall a, b \; P(A = a, \; B = b) = P(A = a) \times P(B = b)$$

Source of Numbers

Requires numbers: $P(X)$, $P(X \mid Y)$

Where do they come from?

Experiments: Empirical, frequentist approach

Prior prob's from actuary tables,
Conditional probability (symptom, given disease)
sensitivity/specificity (lab results)

Objectivist: Probabilities express some real aspects of the universe.

Subjectivist: Characterizes an agent's beliefs.

- Reference class problem...

  What are "equivalent cases"?

Motivation for Belief Nets

Challenge: To decide on proper action

- Which treatment, given symptoms?
- Where to move?
- Where to search for info?
- ...

Need to know dependencies in world

- between symptom and disease
- between symptom_1 and symptom_2
- between disease_1 and disease_2
- ...

Q: Full joint?

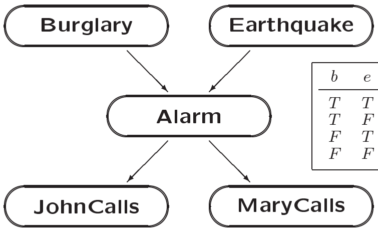
A: Too big ($\geq 2^n$)
Too slow (inference requires adding 2^k ...)

Better:
+ Encode dependencies
+ Encode *relevant* dependencies

Components of a Bayesian Net

$P(B)$
0.001

$P(E)$
0.002



b	e	$P(A \mid B=b, E=e)$
T	T	0.95
T	F	0.94
F	T	0.29
F	F	0.001

a	$P(J \mid A=a)$
T	0.90
F	0.05

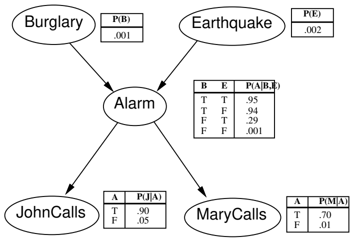
a	$P(M \mid A=a)$
T	0.70
F	0.01

Directed Acyclic Graph:

$$\mathcal{BN} = \left\{ \begin{array}{lll} \mathcal{N} & \text{Nodes} & \equiv \text{Variables} \\ \mathcal{A} & \text{Arcs} & \equiv \text{Dependencies} \\ \mathcal{C} & \text{CPTables} & \equiv \text{“weights”} \end{array} \right\}$$

- **Nodes:** one for each *random variable*
- **Arcs:** one for each *direct influence* between two random variables
- **CPT:** each node stores a conditional probability table
  $P(\texttt{Node} \mid \texttt{Parents(Node)})$
  to quantify effects of "parents" on child

Causes, and Bayesian Net



- What "causes" `Alarm`?
  A: `Burglary`, `Earthquake`
- What "causes" `JohnCall`?
  A: `Alarm` $\boxed{\text{\textit{N.b., NOT Burglary, ...}}}$
- Why not `Alarm` $\Rightarrow$ `MaryCalls`?

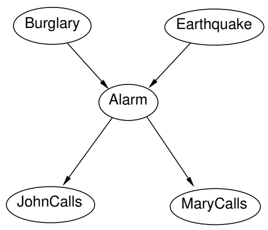
  (CPTable =

Alarm	$P(\texttt{MC} \mid \texttt{A})$
T	1.0
F	0.0

)

  A: Mary not always home
  ... may be playing loud music
  ... phone may be broken
  ...

Independence in a Bayesian Net



- `Burglary`, `Earthquake` independent
  (have no parents...)
- Given `Alarm`,
  `JohnCalls` and `MaryCalls` independent

`JohnCalls` is correlated with `MaryCalls` in general
as suggest `Alarm`
But given `Alarm`,
`JohnCalls` gives no NEW evidence wrt `MaryCalls`

Recovering Joint

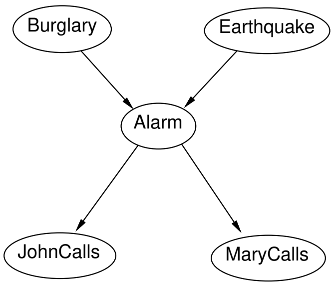
$P(\neg b,\, e,\, a,\, \neg j,\, m\,) \quad =$

$P(\neg b\,)\; P(\,e\,|\,\neg b\,)\; P(\,a\,|\,e, \neg b\,)\; P(\,\neg j\,|\,a, e, \neg b\,)\; P(\,m\,|\,\neg j, a, e, \neg b\,)$

$P(\neg b\,)\; P(\,e\,)\qquad P(\,a\,|\,e, \neg b\,)\; P(\,\neg j\,|\,a\,)\qquad P(\,m\,|\,a\,)$
0.99 × 0.02 × 0.29 × 0.1 × 0.70

Node independent of predecessors, given parents

$P(\,\neg b,\, e,\, a,\, \neg j,\, m\,) \;=$



$P(\,\neg b\,)\;\cdot\; P(\,e\,)\;\cdot$

$P(\,a\,|\,\neg b, e\,)\;\cdot$

$P(\,\neg j\,|\,a\,)\;\cdot\; P(\,m\,|\,a\,)$

Meaning of Bayesian Net

- A BN represents
  + joint distribution
  + condition independence statements

Eg: $P(\,J, M, A, \neg B, \neg E\,)$
$= P(\,J\,|\,A\,)\,P(\,M\,|\,A\,)\,P(\,A\,|\,\neg B, \neg E\,)\,P(\,\neg B\,)\,P(\,\neg E\,)$
$= \;0.90 \;\times\; 0.70 \;\times\; 0.001 \;\times 0.999 \times 0.998$
$= \;0.00062$

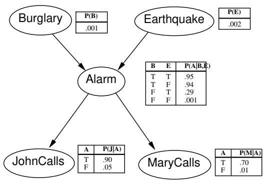
In gen'l, $P(\,X_1, X_2, \ldots, X_m\,) =$
$P(\,X_1\,|\,X_2, \ldots, X_m\,)\,P(\,X_2, \ldots, X_m\,) \;=$
$P(\,X_1\,|\,X_2, \ldots, X_m\,)\,P(\,X_2\,|\,X_3, \ldots, X_m\,)\,P(\,X_3, \ldots, X_m\,) =$
$\prod_i P(\,X_i\,|\,X_{i+1}, \ldots, X_m\,)$

- Independence means:
  $P(\,X_i\,|\,X_{i+1}, \ldots, X_m\,) = P(\,X_i\,|\,Parents(X_i)\,)$

Node independent of predecessors, given parents

So. . . $\boxed{P(\,X_1, X_2, \ldots, X_m\,) = \prod_i P(\,X_i\,|\,Parents(X_i)\,)}$

Comments



- BN used 10 entries
  . . . can recover full joint (2^5 entries)

  (Given structure, other $2^5 - 10$ entries
  are REDUNDANT)

⇒ Can compute
$P(\,$`Burglary` $|$ `JohnCalls`, $\neg$`MaryCalls`$\,)$:

Get joint, then marginalize, conditionalize, . . .

∃ better ways. . .

Note: Given structure, ANY CPT is consistent.

∄ redundancies in BN. . .

Conditional Independence

Node X is independent of its non-descendants
given assignment to immediate parents $parents(X)$

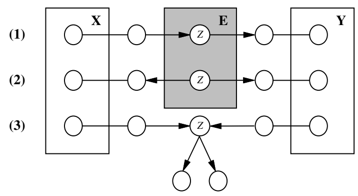
General question: "$X \perp Y \,|\, E$"
Are nodes X independent of nodes Y,
given assignments to (evidence) nodes E?

Answer: If every undirected path from X to Y
is d-separated by E, then $X \perp Y \,|\, E$

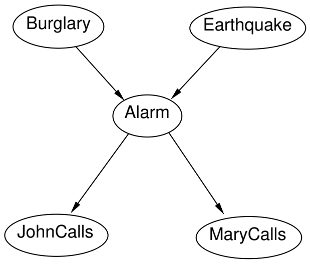
d-separated if every path from X to Y is **blocked** by E

. . . if ∃ node Z on path s.t.

1. $Z \in E$, and Z has 1 out-link (on path)
2. $Z \in E$, and Z has 2 out-link
3. Z has 2 in-links, $Z \notin E$, no child of Z in E



Explaining d-Separation



Case 1: `Burglary` and `JohnCalls` are
conditionally independent given `Alarm`

Case 2: `JohnCalls` and `MaryCalls` are
conditionally independent given `Alarm`

Case 3: `Burglary` and `Earthquake` are
independent given no other information

But... `Burglary` and `Earthquake` are
dependent given `Alarm`

Ie, `Earthquake` may "explain away" `Alarm`
decreasing prob of `Burglary`