Uncertainty reasoning

Dec 26th 2020

Reasoning under Uncertainty

- Basics of probability theory, including the representation language for uncertain beliefs.
  - Acting Under Uncertainty
  - Rational Decisions
  - Basic Probability Notation
  - Bayes' Rule
- Belief networks, a powerful tool for representing and reasoning with uncertain knowledge.

A Logical Agent believes a sentence is either True or False.
A Probabilistic Agent has a degree of belief between 0 and 1

Acting under Uncertainty

- Within a logical-agent approach, **agents almost never have access to the whole truth** about their environment.
  - Some **sentences can be ascertained directly from the agent's percepts**, and others can be **inferred from current and previous percepts** together with knowledge about the environment.
  - However, **for almost every case, there will be important questions to which the agent cannot find a categorical answer**.
- The **agent must therefore act under uncertainty**.
- Uncertainty can also **arise because of incompleteness and incorrectness in the agent's understanding** of the properties of the environment.

Need for Reasoning w/ Uncertainty

- The world is full of uncertainty
  - chance nodes/sensor noise/actuator error/partial info..
  - Logic is brittle
    - can’t encode exceptions to rules
    - can‘t encode statistical properties in a domain
  - Computers need to be able to handle uncertainty
- Probability: new foundation for AI (& CS!)

- Massive amounts of data around today
  - Statistics and CS are both about data
  - Statistics lets us summarize and understand it
  - Statistics is the basis for most learning
- Statistics lets data do our work for us

Logic *vs.* Probability

Symbol: Q, R …	**Random variable**: Q …
Boolean values: T, F	**Domain**: you specify e.g. {heads, tails} [1, 6]
State of the world: Assignment to Q, R … Z	**Atomic event**: complete specification of world: Q… Z • Mutually exclusive • Exhaustive
	Prior probability (aka Unconditional prob: P(Q)
	Joint distribution: Prob. of every atomic event

Probability Basics

- Begin with a set S: the sample space
  - e.g., 6 possible rolls of a die.
- x ϵ S is a sample point/possible world/atomic event
- A probability space or probability model is a sample space with an assignment P(x) for every x s.t. $0 \leq P(x) \leq 1$ and $\sum P(x) = 1$
- An event A is any subset of S
  - e.g. A= ‘die roll < 4’
- A random variable is a function from sample points to some range, e.g., the reals or Booleans

Types of Probability Spaces

Propositional or Boolean random variables

e.g., $Cavity$ (do I have a cavity?)

Discrete random variables (*finite* or *infinite*)

e.g., $Weather$ is one of $\langle sunny, rain, cloudy, snow \rangle$

$Weather = rain$ is a proposition

Values must be exhaustive and mutually exclusive

Continuous random variables (*bounded* or *unbounded*)

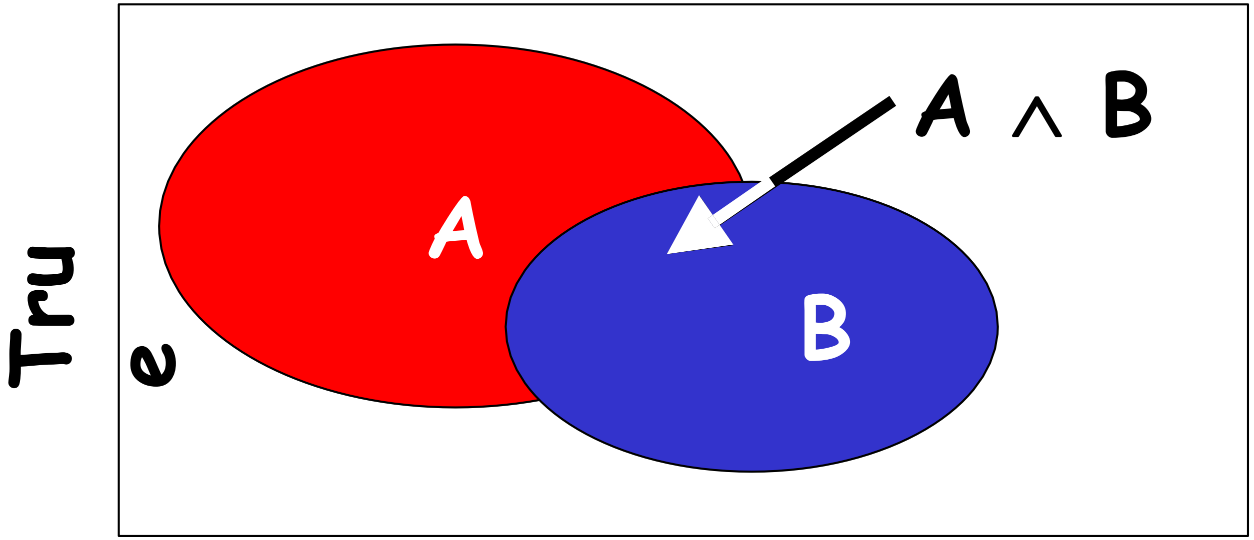
e.g., $Temp = 21.6$; also allow, e.g., $Temp < 22.0$.

Arbitrary Boolean combinations of basic propositions

Axioms of Probability Theory

- All probabilities between 0 and 1
  - $0 \leq P(A) \leq 1$
  - P(true) = 1
  - P(false) = 0.
- The probability of disjunction is:

$$P(A \vee B) = P(A) + P(B) - P(A \wedge B)$$



Prior Probability

Prior or unconditional probabilities of propositions

e.g., $P(Cavity = true) = 0.1$ and $P(Weather = sunny) = 0.72$

correspond to belief prior to arrival of any (new) evidence

Probability distribution gives values for all possible assignments:

$\mathbf{P}(Weather) = \langle 0.72, 0.1, 0.08, 0.1 \rangle$ (*normalized*, i.e., sums to 1)

Joint probability distribution for a set of r.v.s gives the probability of every atomic event on those r.v.s

$\mathbf{P}(Weather, Cavity) =$ a 4×2 matrix of values:

Joint distribution can answer any question

Conditional probability

- Conditional or posterior probabilities
  e.g., P(*cavity* | *toothache*) = 0.8
  i.e., given that *toothache* is all I know there is 80% chance of cavity

- Notation for conditional distributions:
  P(*Cavity* | *Toothache*) = 2-element vector of 2-element vectors)

- If we know more, e.g., *cavity* is also given, then we have
  P(*cavity* | *toothache, cavity*) = 1

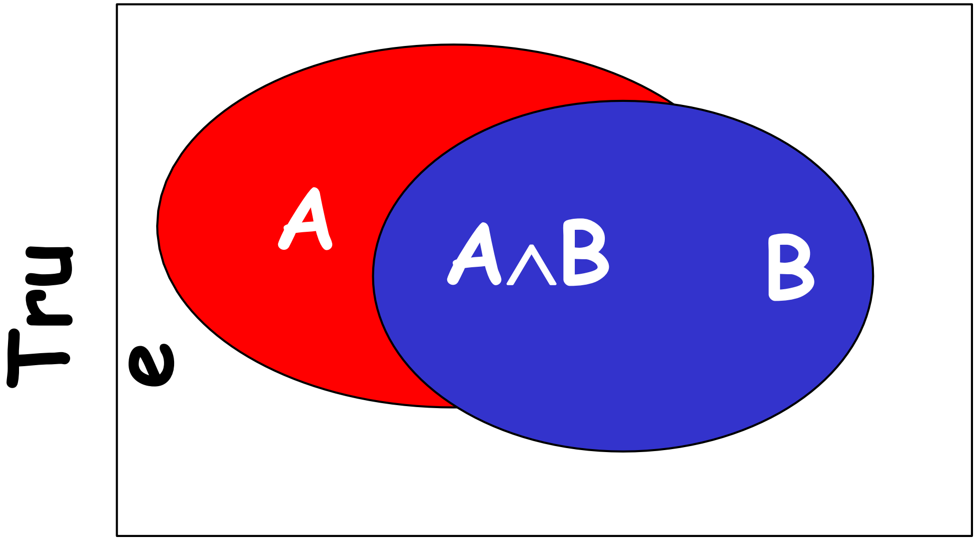
- New evidence may be irrelevant, allowing simplification:
  P(*cavity* | *toothache, sunny*) = P(*cavity* | *toothache*) = 0.8

- This kind of inference, sanctioned by domain knowledge, is crucial

Conditional Probability

- P(*A* | *B*) is the probability of *A* given *B*
- Assumes that *B* is the only info known.
- Defined by:

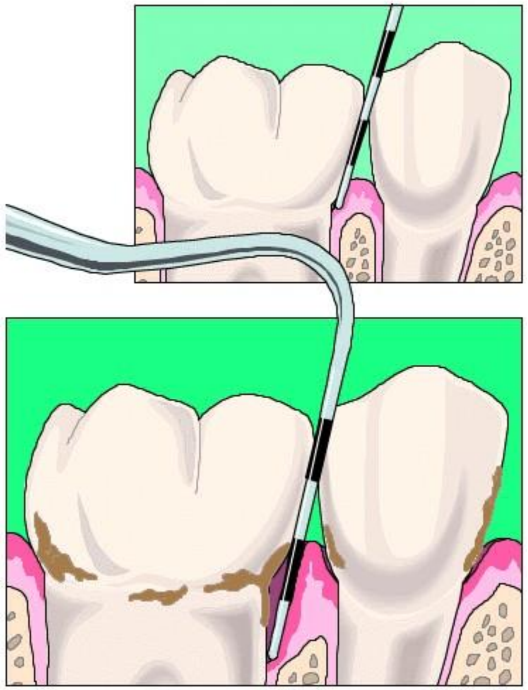
$$P(A|B) = \frac{P(A \wedge B)}{P(B)}$$



Chain Rule/Product Rule

- $P(X_1, \ldots, X_n) = P(X_n | X_1..X_{n-1})P(X_{n-1} | X_1..X_{n-2})\ldots\ P(X_1)$

Dilemma at the Dentist's



What is the probability of a cavity given a toothache?
What is the probability of a cavity given the probe catches?

Inference by Enumeration

Start with the joint distribution:

	toothache		*¬ toothache*	
	catch	*¬ catch*	*catch*	*¬ catch*
cavity	.108	.012	.072	.008
¬ cavity	.016	.064	.144	.576

For any proposition ϕ, sum the atomic events where it is true:

$$P(\phi) = \Sigma_{\omega:\omega \models \phi} P(\omega)$$

P(toothache)=.108+.012+.016+.064

= .20 or 20%

Inference by Enumeration

Start with the joint distribution:

	toothache		$\neg$ *toothache*	
	catch	$\neg$ *catch*	*catch*	$\neg$ *catch*
cavity	.108	.012	.072	.008
$\neg$ *cavity*	.016	.064	.144	.576

For any proposition ϕ, sum the atomic events where it is true:

$$P(\phi) = \Sigma_{\omega:\omega\models\phi} P(\omega)$$

P(toothache$\vee$cavity) = .20 + .072 + .008

.28

Inference by Enumeration

Start with the joint distribution:

	toothache		¬ *toothache*	
	catch	¬ *catch*	*catch*	¬ *catch*
cavity	.108	.012	.072	.008
¬ *cavity*	.016	.064	.144	.576

Can also compute conditional probabilities:

$$P(\neg cavity | toothache) = \frac{P(\neg cavity \wedge toothache)}{P(toothache)}$$

$$= \frac{0.016 + 0.064}{0.108 + 0.012 + 0.016 + 0.064} = 0.4$$

Complexity of Enumeration

- Worst case time: $O(d^n)$
  - Where d = max arity
  - And n = number of random variables
- Space complexity also $O(d^n)$
  - Size of joint distribution

Independence

- *A* and *B* are *independent* iff:

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

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

These two constraints are logically equivalent

- Therefore, if *A* and *B* are independent:

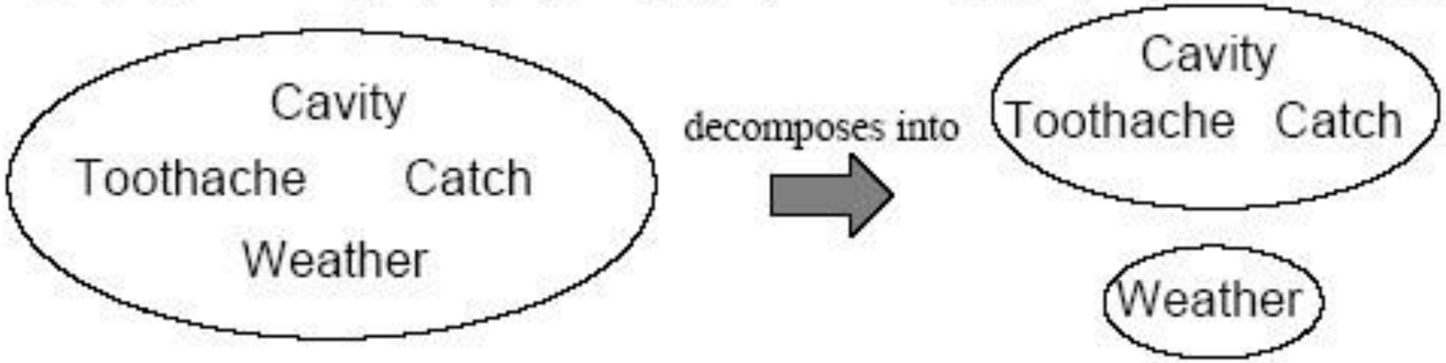
$$P(A \mid B) = \frac{P(A \wedge B)}{P(B)} = P(A)$$

$$P(A \wedge B) = P(A)P(B)$$

Independence

A and B are independent iff

$\mathbf{P}(A|B) = \mathbf{P}(A)$ or $\mathbf{P}(B|A) = \mathbf{P}(B)$ or $\mathbf{P}(A, B) = \mathbf{P}(A)\mathbf{P}(B)$



$$\mathbf{P}(Toothache, Catch, Cavity, Weather)$$
$$= \mathbf{P}(Toothache, Catch, Cavity)\mathbf{P}(Weather)$$

32 entries reduced to 12; for n independent biased coins, $2^n \to n$

Complete independence is powerful but rare

What to do if it doesn't hold?

Conditional Independence

$\mathbf{P}(Toothache, Cavity, Catch)$ has $2^3 - 1 = 7$ independent entries

If I have a cavity, the probability that the probe catches in it doesn't depend on whether I have a toothache:

(1) $P(catch|toothache, cavity) =$

The same independence holds if I haven't got a cavity:

(2) $P(catch|toothache, \neg cavity) = P(catch|\neg cavity)$

Conditional Independence II

P(catch | toothache, cavity) = P(catch | cavity)
P(catch | toothache,¬cavity) = P(catch |¬cavity)

Equivalent statements:

$$\mathbf{P}(Toothache|Catch, Cavity) = \mathbf{P}(Toothache|Cavity)$$

$$\mathbf{P}(Toothache, Catch|Cavity) = \mathbf{P}(Toothache|Cavity)\mathbf{P}(Catch|Cavity)$$

Why only 5 entries in table?

Write out full joint distribution using chain rule:

$$\mathbf{P}(Toothache, Catch, Cavity)$$

$$=$$

$$=$$

$$=$$

Power of Cond. Independence

- Often, using conditional independence reduces the storage complexity of the joint distribution from exponential to linear!!

- Conditional independence is the most basic & robust form of knowledge about uncertain environments.

Bayes Rule

Bayes rules!

posterior

$$P(x, y) = P(x \mid y)P(y) = P(y \mid x)P(x)$$

$$\boxed{P(x \mid y) = \frac{P(y \mid x)\, P(x)}{P(y)} = \frac{\text{likelihood} \cdot \text{prior}}{\text{evidence}}}$$

Useful for assessing diagnostic probability from causal probability:

$$P(Cause|Effect) = \frac{P(Effect|Cause)P(Cause)}{P(Effect)}$$

Computing Diagnostic Prob. from Causal Prob.

$$P(Cause|Effect) = \frac{P(Effect|Cause)P(Cause)}{P(Effect)}$$

E.g. let M be meningitis, S be stiff neck

P(M) = 0.0001,
P(S) = 0.1,
P(S|M)= 0.8

$$P(m|s) = \frac{P(s|m)P(m)}{P(s)} = \frac{0.8 \times 0.0001}{0.1} = 0.0008$$

Note: posterior probability of meningitis still very small!

Other forms of Bayes Rule

$$P(x|y) = \frac{P(y|x)\,P(x)}{P(y)} = \frac{\text{likelihood} \cdot \text{prior}}{\text{evidence}}$$

$$P(x|y) = \frac{P(y|x)P(x)}{\sum_x P(y|x)\,P(x)}$$

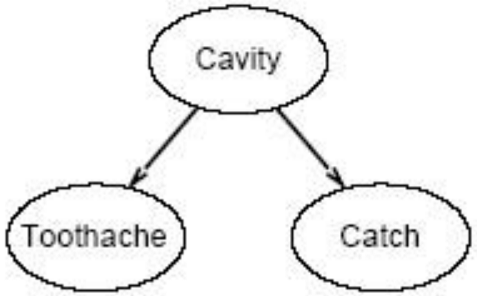
$$\boxed{\begin{array}{c} P(x|y) = \alpha P(y|x)P(x) \\ \text{posterior} \propto \text{likelihood} \cdot \text{prior} \end{array}}$$

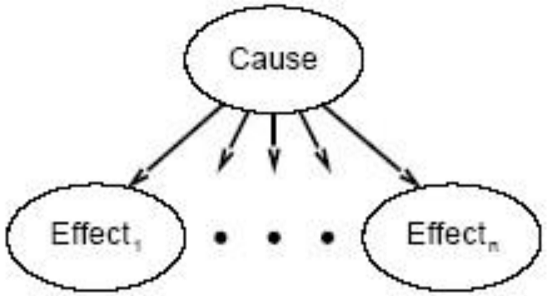
Bayes' Rule & Cond. Independence

$$\begin{aligned}
&\mathbf{P}(Cavity|toothache \wedge catch) \\
&= \alpha\,\mathbf{P}(toothache \wedge catch|Cavity)\mathbf{P}(Cavity) \\
&= \alpha\,\mathbf{P}(toothache|Cavity)\mathbf{P}(catch|Cavity)\mathbf{P}(Cavity)
\end{aligned}$$

This is an example of a *naive Bayes* model:

$$\mathbf{P}(Cause, Effect_1, \ldots, Effect_n) = \mathbf{P}(Cause)\Pi_i\mathbf{P}(Effect_i|Cause)$$





Total number of parameters is *linear* in n

Burglars and Earthquakes

- You are at a “Done with the AI class” party.
- Neighbor John calls to say your home alarm has gone off (but neighbor Mary doesn't).
- Sometimes your alarm is set off by minor earthquakes.
- Question: Is your home being burglarized?
- Variables: Burglary, Earthquake, Alarm, JohnCalls, MaryCalls
- Network topology reflects "causal" knowledge:
  - A burglar can set the alarm off
  - An earthquake can set the alarm off
  - The alarm can cause Mary to call
  - The alarm can cause John to call

Example

- Pearl lives in Los Angeles. It is a high-crime area. Pearl installed a burglar alarm. He asked his neighbors John & Mary to call him if they hear the alarm. This way he can come home if there is a burglary. Los Angeles is also earth-quake prone. Alarm goes off when there is an earth-quake.

Burglary => Alarm

Earth-Quake => Alarm

Alarm => John-calls

Alarm => Mary-calls

If there is a burglary, will Mary call?

If Mary didn’t call, is it possible that Burglary occurred?

Example (Real)

- Pearl lives in Los Angeles. It is a high-crime area. Pearl installed a burglar alarm. He asked his neighbors John & Mary to call him if they hear the alarm. This way he can come home if there is a burglary. Los Angeles is also earth-quake prone. Alarm goes off when there is an earth-quake.
- Pearl lives in real world where (1) burglars can sometimes disable alarms (2) some earthquakes may be too slight to cause alarm (3) Even in Los Angeles, Burglaries are more likely than Earth Quakes (4) John and Mary both have their own lives and may not always call when the alarm goes off (5) Between John and Mary, John is more of a slacker than Mary.(6) John and Mary may call even without alarm going off

- Burglary => Alarm Earth-Quake => Alarm Alarm => John-calls Alarm => Mary-calls

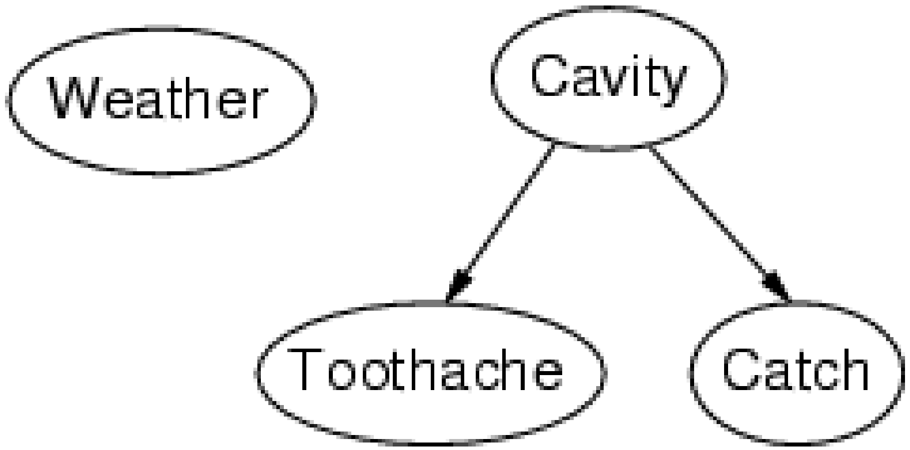
- If there is a burglary, will Mary call?
- Check KB & E |= M
- If Mary didn't call, is it possible that Burglary
  - occurred?
- Check KB & ~M *doesn't entail* ~B
- John already called. If Mary also calls, is it more l that Burglary occurred?
- You now also hear on the TV that there was an earthquake. Is Burglary more or less likely now?

Bayes Nets

- In general, joint distribution *P* over set of variables ($X_1 \times \ldots \times X_n$) requires exponential space for representation & inference
- BNs provide a graphical representation of *conditional independence* relations in *P*
  - usually quite compact
  - requires assessment of fewer parameters, those being quite natural (e.g., causal)
  - efficient (usually) inference: query answering and belief update

Back at the dentist's

Topology of network encodes conditional independence assertions:



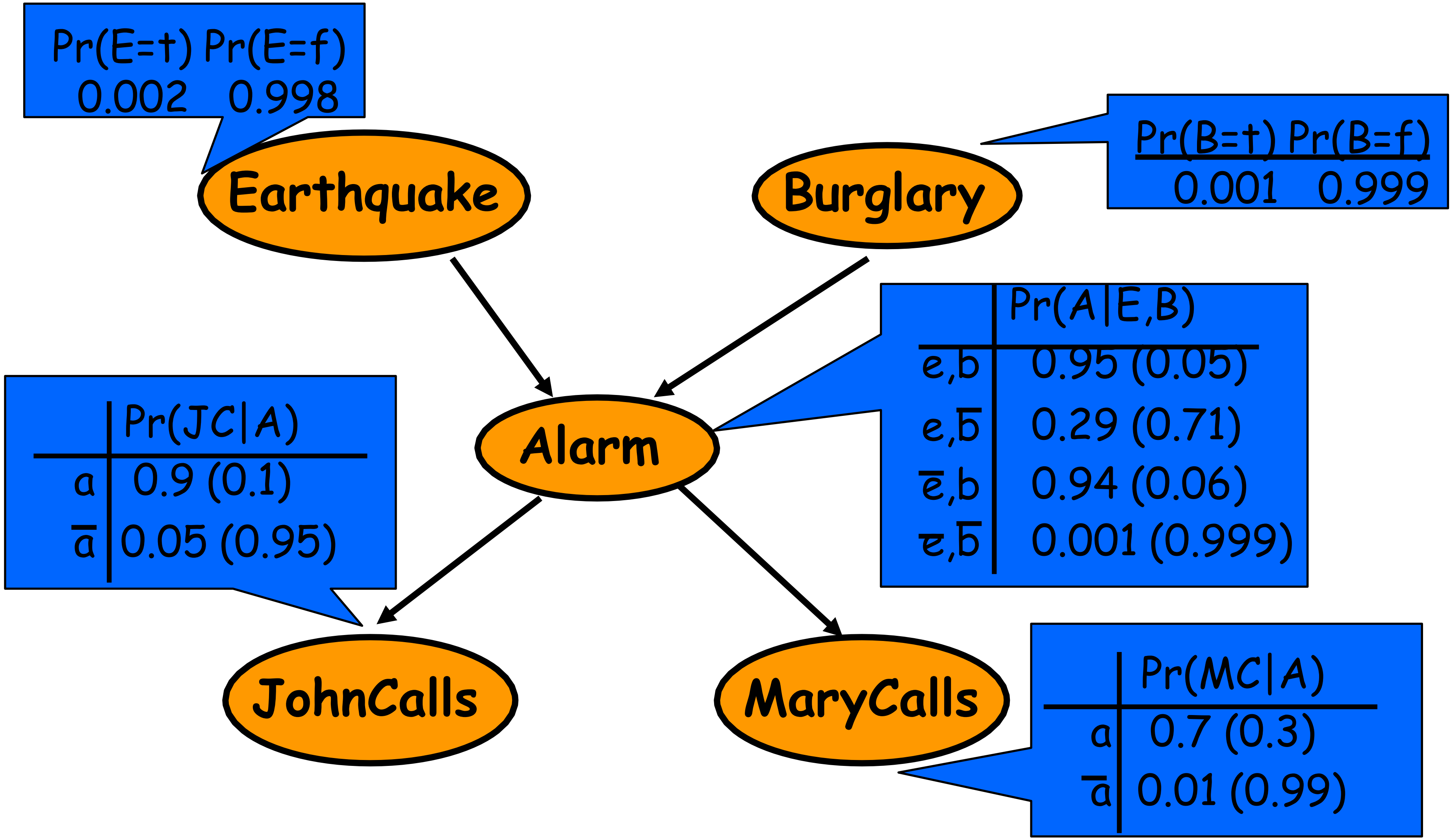
Weather is independent of the other variables

Toothache and Catch are <u>conditionally independent</u> of each other given Cavity

Syntax

- a set of nodes, one per random variable
- a directed, acyclic graph (link ≈"directly influences")
- a conditional distribution for each node given its parents: $P(X_i \mid \text{Parents}(X_i))$
  - For discrete variables, conditional probability table (CPT)= distribution over X_i for each combination of parent values

Burglars and Earthquakes



Pr(E=t)	Pr(E=f)
0.002	0.998

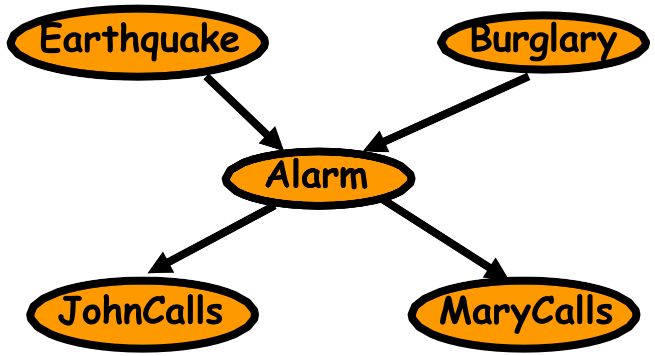
Pr(B=t)	Pr(B=f)
0.001	0.999

	Pr(A|E,B)
e,b	0.95 (0.05)
e,$\bar{b}$	0.29 (0.71)
$\bar{e}$,b	0.94 (0.06)
$\bar{e}$,$\bar{b}$	0.001 (0.999)

	Pr(JC|A)
a	0.9 (0.1)
$\bar{a}$	0.05 (0.95)

	Pr(MC|A)
a	0.7 (0.3)
$\bar{a}$	0.01 (0.99)

Earthquake Example (cont'd)



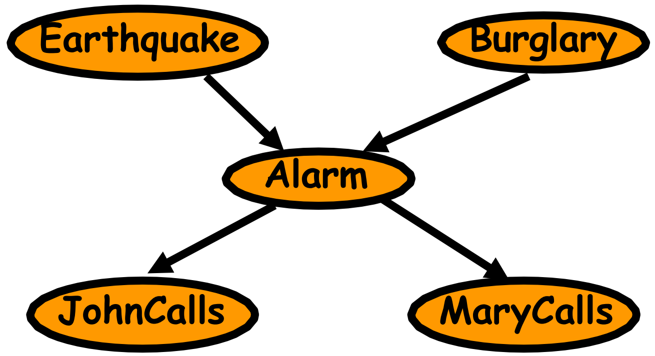
- **If we know *Alarm*, no other evidence influences our degree of belief in *JohnCalls***
  - $P(JC|MC,A,E,B) = P(JC|A)$
  - also: $P(MC|JC,A,E,B) = P(MC|A)$ and $P(E|B) = P(E)$
- **By the chain rule we have**

$$P(JC,MC,A,E,B) = P(JC|MC,A,E,B) \cdot P(MC|A,E,B) \cdot P(A|E,B) \cdot P(E|B) \cdot P(B)$$

$$= \mathbf{P(JC|A) \cdot P(MC|A) \cdot P(A|B,E) \cdot P(E) \cdot P(B)}$$

- **Full joint requires only 10 parameters (cf. 32)**

Earthquake Example (Global Semantics)



- **We just proved**

  $P(JC,MC,A,E,B) = \mathbf{P(JC|A)} \cdot \mathbf{P(MC|A)} \cdot \mathbf{P(A|B,E)} \cdot \mathbf{P(E)} \cdot \mathbf{P(B)}$

- **In general full joint distribution of a Bayes net is defined as**

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

BNs: Qualitative Structure

- Graphical structure of BN reflects conditional independence among variables
- Each variable *X* is a node in the DAG
- Edges denote *direct probabilistic influence*
  - usually interpreted *causally*
  - parents of *X* are denoted *Par(X)*
- ***Local semantics: X* is conditionally independent of all nondescendents given its parents**
  - Graphical test exists for more general independence
  - "Markov Blanket"

THANK YOU