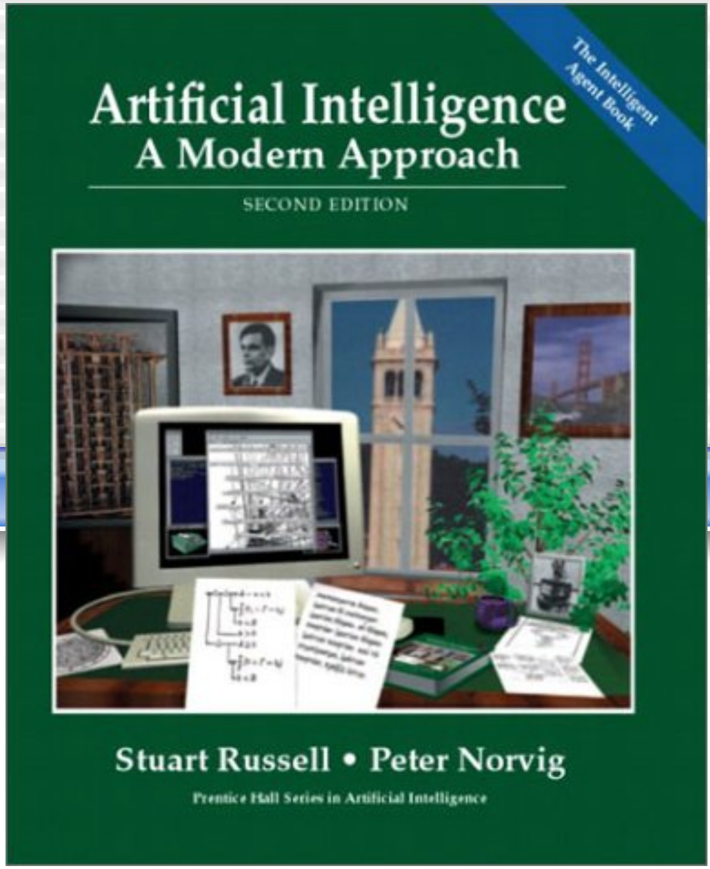


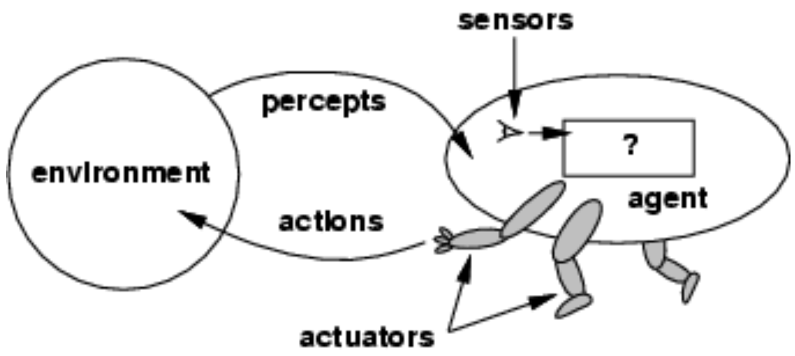
Uncertainty

Chapter 13

Outline

- Agents
- Uncertainty
- Probability
- Syntax and Semantics
- Inference
- Independence and Bayes' Rule
- Bayesian Networks

Agents and environments



- The agent function maps from percept histories to actions: $[f: \mathcal{P}^\star \rightarrow \mathcal{A}]$
- The agent program runs on the physical architecture to produce f
- agent = architecture + program
  architecture: PC, robotic car, …

Uncertainty

Let action A_t = leave for airport t minutes before flight
Will A_t get me there on time?

Problems:

1. partial observability (road state, other drivers' plans, etc.)
2. noisy sensors (traffic reports)
3. uncertainty in action outcomes (flat tire, etc.)
4. immense complexity of modeling and predicting traffic

Hence, it seems that a purely logical approach either

1. risks falsehood: "A_{25} will get me there on time", or
2. leads to conclusions that are too weak for decision making:

"A_{25} will get me there on time if there's no accident on the bridge and it doesn't rain and my tires remain intact etc etc."

(A_{1440} might reasonably be said to get me there on time but I'd have to stay overnight in the airport …)

Problems of Logic in certain Domains

- Diagnosis:
  - Toothache => Cavity *v GumProblem v Abscess v …*
  - Cavity => Toothache
- The connection between toothaches and cavities is just not a logical consequence. For medical diagnosis logic does not seem to be appropriate.

Methods for handling uncertainty

- Logic:
  - Assume my car does not have a flat tire
  - Assume A_{25} works unless contradicted by evidence
- Issues: What assumptions are reasonable? How to handle contradiction?
- Rules with fudge factors (belief in the rule):
  - A_{25} $|\!\rightarrow_{0.3}$ get there on time
  - *Sprinkler* $|\!\rightarrow_{0.99}$ *WetGrass*
  - *WetGrass* $|\!\rightarrow_{0.7}$ *Rain*
- Issues: Problems with combination, e.g., *Sprinkler* causes *Rain*??

- Probability
  - Model agent's degree of belief
  - Given the available evidence,
  - A_{25} will get me there on time with probability 0.04

Probability

Probabilistic assertions summarize effects of

- laziness: failure to enumerate exceptions, qualifications, etc.
- theoretical ignorance: no complete theory
- practical ignorance: lack of relevant facts, initial conditions, tests, etc.

Subjective probability:

- Probabilities relate propositions to agent's own state of knowledge

  e.g., $P(A_{25} \mid \text{no reported accidents}) = 0.06$

These are not assertions about the world

Probabilities of propositions change with new evidence:

e.g., $P(A_{25} \mid \text{no reported accidents, 5 a.m.}) = 0.15$

Making decisions under uncertainty

Suppose I believe the following:

$P(A_{25}$ gets me there on time | …) $= 0.04$

$P(A_{90}$ gets me there on time | …) $= 0.70$

$P(A_{120}$ gets me there on time | …) $= 0.95$

$P(A_{1440}$ gets me there on time | …) $= 0.9999$

- Which action to choose?

  Depends on my preferences for missing flight vs. time spent waiting, etc.

  - Utility theory is used to represent and infer preferences
  - Decision theory = probability theory + utility theory

Decision theoretic agent

function DT-AGENT(*percept*) **returns** an *action*
  static: *belief_state*, probabilistic beliefs about the current state of the world
    action, the agent's action

  update *belief_state* based on *action* and *percept*
  calculate outcome probabilities for actions,
    given action descriptions and current *belief_state*
  select *action* with highest expected utility
    given probabilities of outcomes and utility information
  return *action*

Figure 13.1 A decision-theoretic agent that selects rational actions. The steps will be fleshed out in the next five chapters.

Probability theory: syntax

- Basic element: random variable
- Similar to propositional logic: possible worlds defined by assignment of values to random variables.
- Boolean random variables
  e.g., *Cavity* (do I have a cavity?). Domain is <true, false>
- Discrete random variables
  e.g., *Weather* is one of <*sunny,rainy,cloudy,snow*>
- Domain values must be exhaustive and mutually exclusive
- Elementary proposition constructed by assignment of a value to a random variable: e.g.,
  - *Weather* = *sunny*,
  - *Cavity* = *false* (abbreviated as $\neg$*cavity*)
  - *Cavity* = *true* (abbreviated as *cavity*)
- Complex propositions formed from elementary propositions and standard logical connectives e.g., *Weather* = *sunny* $\vee$ *Cavity* = *false*

Syntax

- Atomic event: A complete specification of the state of the world about which the agent is uncertain

  E.g., if the world is described by only two Boolean variables *Cavity* and *Toothache*, then there are 4 distinct atomic events:

  $Cavity = false \wedge Toothache = false$

  $Cavity = false \wedge Toothache = true$

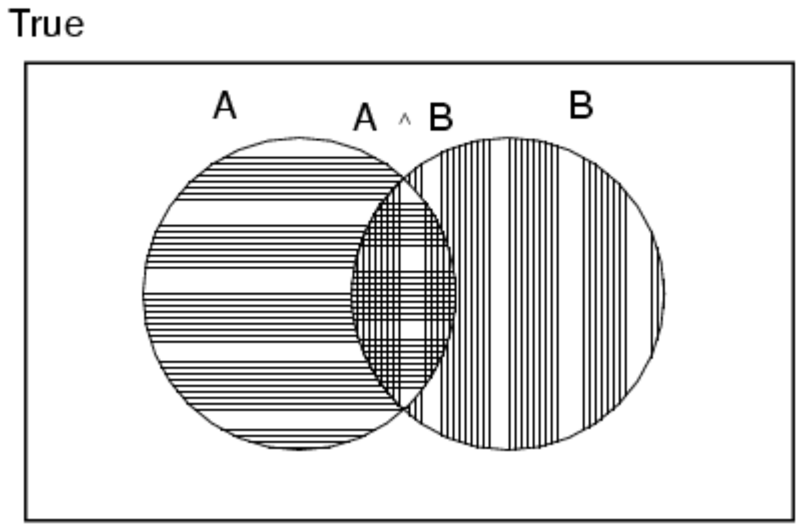
  $Cavity = true \wedge Toothache = false$

  $Cavity = true \wedge Toothache = true$

- Atomic events are mutually exclusive and exhaustive

Axioms of probability

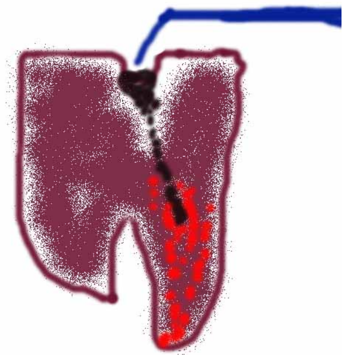
- For any propositions *A*, *B*
  - $0 \leq P(A) \leq 1$
  - $P(true) = 1$ and $P(false) = 0$
  - $P(A \lor B) = P(A) + P(B) - P(A \land B)$



Example world

Example: ***Dentist problem*** **with four variables:**

- ***Toothache*** **(I have a toothache)**
- ***Cavity*** **(I have a cavity)**
- ***Catch*** **(steel probe catches in my tooth)**
- ***Weather*** **(*sunny,rainy,cloudy,snow*)**



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:

  P(*Weather*) = <0.72,0.1,0.08,0.1>

  (normalized, i.e., sums to 1 because one must be the case)

Full joint probability distribution

- Joint probability distribution for a set of random variables gives the probability of every atomic event on those random variables

  P(*Weather,Cavity*) = a 4 × 2 matrix of values:

Weather =	sunny	rainy	cloudy	snow	
Cavity = true		0.144	0.02	0.016	0.02
Cavity = false		0.576	0.08	0.064	0.08

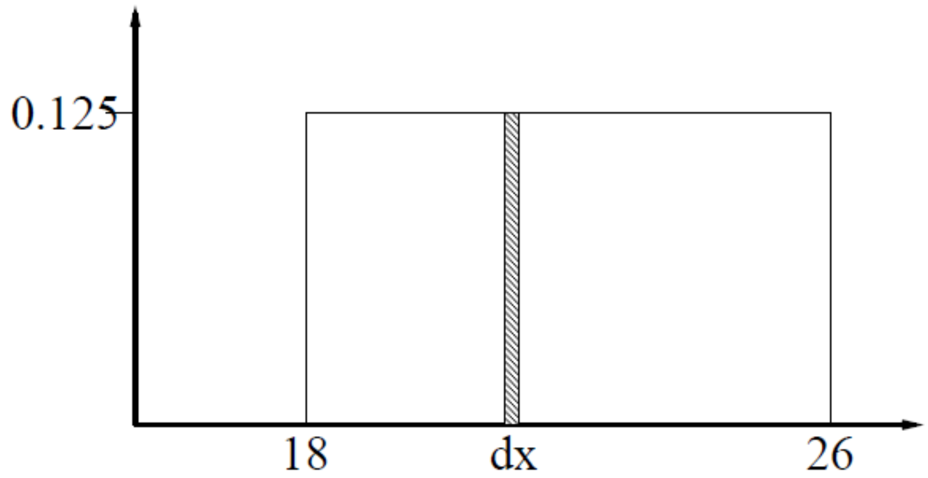
- Full joint probability distribution: all random variables involved
  - **P**(Toothache, Catch, Cavity, Weather)

- Every question about a domain can be answered by the full joint distribution

Probability for continuous variables

Express distribution as a parameterized function of value:

$P(X = x) = U[18, 26](x)$ = uniform density between 18 and 26

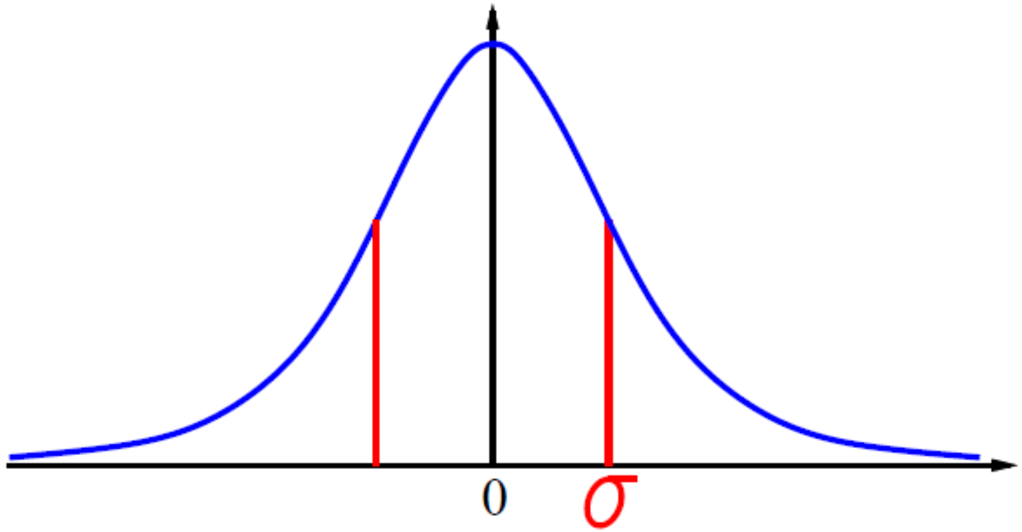


Here P is a density; integrates to 1.
$P(X = 20.5) = 0.125$ really means

$$\lim_{dx \to 0} P(20.5 \leq X \leq 20.5 + dx)/dx = 0.125$$

Gaussian density

$$P(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-(x-\mu)^2/2\sigma^2}$$



Conditional probability

- Conditional or posterior probabilities
  e.g., P(*cavity* | *toothache*) = 0.8
  or: <0.8>
  i.e., given that *toothache* is all I know

- (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, e.g.,
  P(*cavity* | *toothache, sunny*) = P(*cavity* | *toothache*) = 0.8

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

Conditional probability

- Definition of conditional probability (in terms of uncond. prob.):
  $P(a \mid b) = P(a \wedge b) / P(b)$ if $P(b) > 0$
- Product rule gives an alternative formulation ($\wedge$ is commutative):
  $P(a \wedge b) = P(a \mid b) P(b) = P(b \mid a) P(a)$
- A general version holds for whole distributions, e.g.,
  $\mathbf{P}(Weather, Cavity) = \mathbf{P}(Weather \mid Cavity)\, \mathbf{P}(Cavity)$
  View as a set of 4 × 2 equations, not matrix mult.
  (1,1) $P(Weather{=}sunny \mid Cavity{=}true)\ P(Cavity{=}true)$
  (1,2) $P(Weather{=}sunny \mid Cavity{=}false)\ P(Cavity{=}false)$, ….
- Chain rule is derived by successive application of product rule:

$$
\begin{aligned}
\mathbf{P}(X_1, \ldots, X_n) &= \mathbf{P}(X_1,\ldots,X_{n-1})\, \mathbf{P}(X_n \mid X_1,\ldots,X_{n-1}) \\
&= \mathbf{P}(X_1,\ldots,X_{n-2})\, \mathbf{P}(X_{n-1} \mid X_1,\ldots,X_{n-2})\, \mathbf{P}(X_n \mid X_1,\ldots,X_{n-1}) \\
&= \ldots \\
&= \prod_{i=1}^{n} \mathbf{P}(X_i \mid X_1, \ldots, X_{i-1})
\end{aligned}
$$

Inference by enumeration

- Start with the joint probability 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: $P(\varphi) = \Sigma_{\omega:\omega \models \varphi} P(\omega)$

Inference by enumeration

- Start with the joint probability 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: $P(\varphi) = \Sigma_{\omega:\omega \models \varphi} P(\omega)$
- P(*toothache*) = 0.108 + 0.012 + 0.016 + 0.064 = 0.2
- Unconditional or **marginal probability** of toothache
- Process is called marginalization or summing out

Marginalization and conditioning

- Let Y, Z be sequences of random variables s.th. Y ∪ Z denotes all random variables describing the world

- Marginalization
  - $\mathbf{P}(\mathbf{Y}) = \Sigma_{\mathbf{z} \text{ in } \mathbf{Z}} \mathbf{P}(\mathbf{Y},\mathbf{z})$

- Conditioning
  - $\mathbf{P}(\mathbf{Y}) = \Sigma_{\mathbf{z} \text{ in } \mathbf{Z}} \mathbf{P}(\mathbf{Y}|\mathbf{z}) P(\mathbf{z})$

Inference by enumeration

- Start with the joint probability 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:
$P(\varphi) = \Sigma_{\omega:\omega \models \varphi} P(\omega)$

- P(cavity ∨ *toothache*) = 0.108 + 0.012 + 0.072 + 0.008+ 0.016 + 0.064 = 0.28

(P(*cavity* ∨ *toothache*) = P(*cavity*) + P(*toothache*) – P(*cavity* ∧ *toothache*))

Inference by enumeration

- Start with the joint probability distribution:

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

- Can also compute conditional probabilities:

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

Product rule

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

$$= 0.4$$

$P(cavity \mid toothache) = 0.108+0.012/0.2 = 0.6$

Normalization

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

- Denominator **P**(**z**) (or P(toothache) in the example before) can be viewed as a normalization constant α

P(*Cavity* | *toothache*) = α **P**(*Cavity*,*toothache*)
= α [**P**(*Cavity*,*toothache*,*catch*) + **P**(*Cavity*,*toothache*,¬ *catch*)]
= α [<0.108,0.016> + <0.012,0.064>]
= α <0.12,0.08> = <0.6,0.4>

General idea: compute distribution on query variable by fixing evidence variables (Toothache) and summing over hidden variables (Catch)

Inference by enumeration, contd.

Typically, we are interested in
the posterior joint distribution of the query variables **Y**
given specific values **e** for the evidence variables **E**
(**X** are all variables of the modeled world)

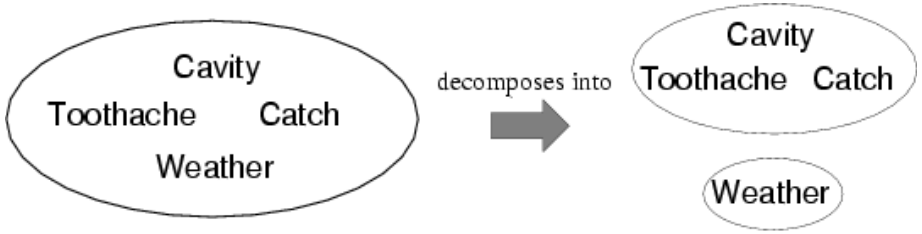
Let the hidden variables be **H** = **X** – **Y** – **E** then the required summation of joint entries is done by summing out the hidden variables:

$$\mathbf{P}(\mathbf{Y} \mid \mathbf{E} = \mathbf{e}) = \alpha \mathbf{P}(\mathbf{Y}, \mathbf{E} = \mathbf{e}) = \alpha \Sigma_h \mathbf{P}(\mathbf{Y}, \mathbf{E} = \mathbf{e}, \mathbf{H} = \mathbf{h})$$

- The terms in the summation are joint entries because **Y**, **E** and **H** together exhaust the set of random variables (**X**)
- Obvious problems:
  1. Worst-case time complexity $O(d^n)$ where d is the largest arity and n denotes the number of random variables
  2. Space complexity $O(d^n)$ to store the joint distribution
  3. How to find the numbers for $O(d^n)$ entries?

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, $O(2^n) \to O(n)$
- Absolute independence powerful but rare
- Dentistry is a large field with hundreds of variables, none of which are independent. What to do?

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) $\mathbf{P}(catch \mid toothache, cavity) = \mathbf{P}(catch \mid cavity)$
- The same independence holds if I haven't got a cavity:
  (2) $\mathbf{P}(catch \mid toothache, \neg cavity) = \mathbf{P}(catch \mid \neg cavity)$
- *Catch* is conditionally independent of *Toothache* given *Cavity*:
  $\mathbf{P}(Catch \mid Toothache, Cavity) = \mathbf{P}(Catch \mid Cavity)$
- Equivalent statements:
  $\mathbf{P}(Toothache \mid Catch, Cavity) = \mathbf{P}(Toothache \mid Cavity)$

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

Conditional independence contd.

- Write out full joint distribution using chain rule:
  P(*Toothache, Catch, Cavity*)
  = **P**(*Toothache* | *Catch, Cavity*) **P**(*Catch, Cavity*)

  = **P**(*Toothache* | *Catch, Cavity*) **P**(*Catch* | *Cavity*) **P**(*Cavity*)
  conditional independence
  = **P**(*Toothache* | *Cavity*) **P**(*Catch* | *Cavity*) **P**(Cavity)

  i.e., 2 + 2 + 1 = 5 independent numbers
- In most cases, the use of conditional independence reduces the size of the representation of the joint distribution from exponential in *n* to linear in *n*.
- Conditional independence is our most basic and robust form of knowledge about uncertain environments.

Bayes' Rule

Product rule $P(a \wedge b) = P(a|b)P(b) = P(b|a)P(a)$

$\Rightarrow$ Bayes' rule $P(a|b) = \dfrac{P(b|a)P(a)}{P(b)}$

or in distribution form

$$\mathbf{P}(Y|X) = \frac{\mathbf{P}(X|Y)\mathbf{P}(Y)}{\mathbf{P}(X)} = \alpha \mathbf{P}(X|Y)\mathbf{P}(Y)$$

Useful for assessing diagnostic probability from causal probability:

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

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

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

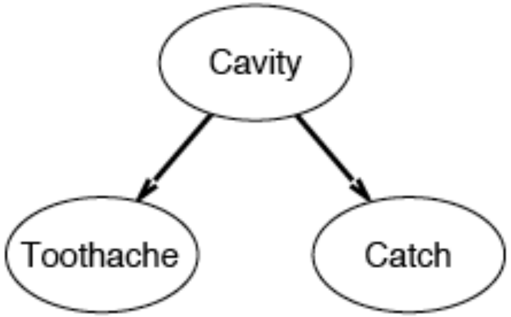
Note: posterior probability of meningitis still very small!

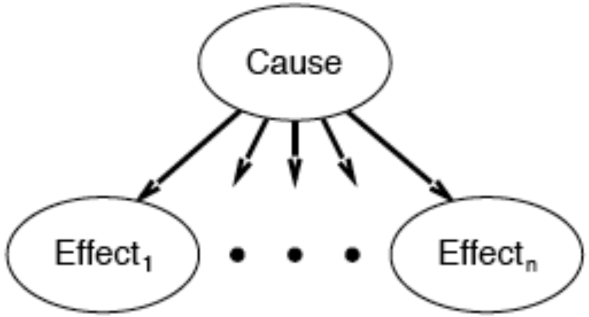
Bayes' Rule (2)

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

This is an example of a naive Bayes model:

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



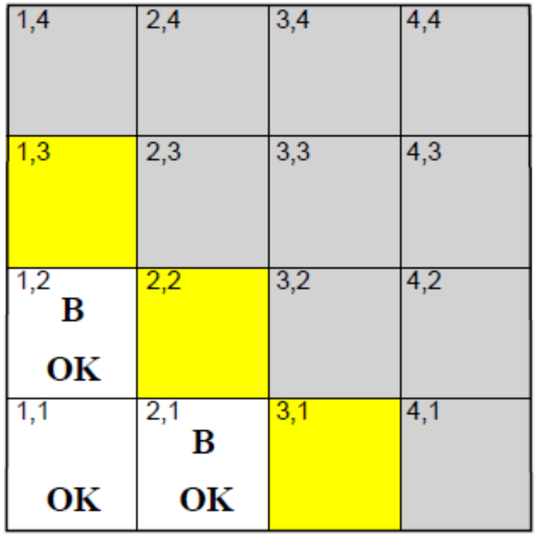


Total number of parameters is linear in n

Why is Bayes' theorem interesting?

- Often you have causal knowledge:
  $P(\mathit{symptom} \mid \mathit{disease})$
  $P(\mathit{light\ is\ off} \mid \mathit{status\ of\ switches\ and\ switch\ positions})$
  $P(\mathit{alarm} \mid \mathit{fire})$
  $P(\mathit{image\ looks\ like}$  $\mid \mathit{a\ tree\ is\ in\ front\ of\ a\ car})$
- and want to do evidential reasoning:
  $P(\mathit{disease} \mid \mathit{symptom})$
  $P(\mathit{status\ of\ switches} \mid \mathit{light\ is\ off\ and\ switch\ positions})$
  $P(\mathit{fire} \mid \mathit{alarm})$.
  $P(\mathit{a\ tree\ is\ in\ front\ of\ a\ car} \mid \mathit{image\ looks\ like}$  $)$

Wumpus World



$P_{ij} = true$ iff $[i, j]$ contains a pit

$B_{ij} = true$ iff $[i, j]$ is breezy

Include only $B_{1,1}, B_{1,2}, B_{2,1}$ in the probability model

Specifying the probability model

The full joint distribution is $\mathbf{P}(P_{1,1}, \ldots, P_{4,4}, B_{1,1}, B_{1,2}, B_{2,1})$

Apply product rule: $\mathbf{P}(B_{1,1}, B_{1,2}, B_{2,1} \,|\, P_{1,1}, \ldots, P_{4,4})\mathbf{P}(P_{1,1}, \ldots, P_{4,4})$

(Do it this way to get $P(Effect|Cause)$.)

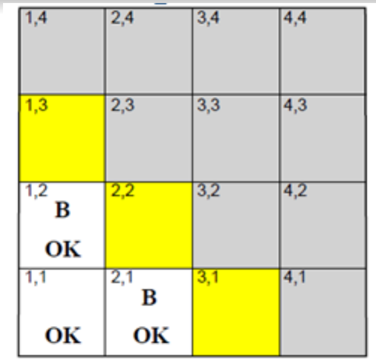
First term: 1 if pits are adjacent to breezes, 0 otherwise

Second term: pits are placed randomly, probability 0.2 per square:

$$\mathbf{P}(P_{1,1}, \ldots, P_{4,4}) = \Pi_{i,j=1,1}^{4,4}\mathbf{P}(P_{i,j}) = 0.2^n \times 0.8^{16-n}$$

for n pits.

Observations and query



We know the following facts:

$b = \neg b_{1,1} \wedge b_{1,2} \wedge b_{2,1}$

$known = \neg p_{1,1} \wedge \neg p_{1,2} \wedge \neg p_{2,1}$

Query is $\mathbf{P}(P_{1,3}|known, b)$

Define $Unknown = P_{ij}$s other than $P_{1,3}$ and $Known$

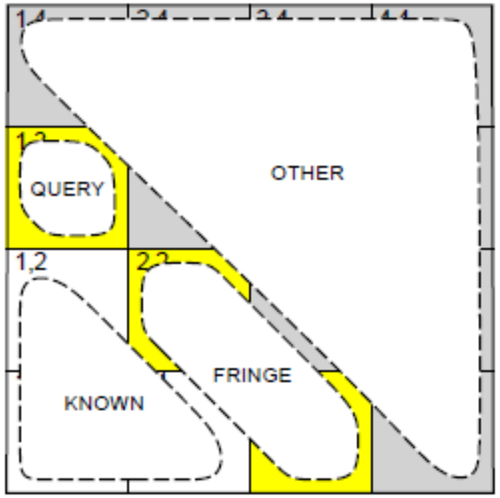
For inference by enumeration, we have

$$\mathbf{P}(P_{1,3}|known, b) = \alpha \Sigma_{unknown} \mathbf{P}(P_{1,3}, unknown, known, b)$$

Grows exponentially with number of squares!

Using conditional independence

Basic insight: observations are conditionally independent of other hidden squares given neighbouring hidden squares



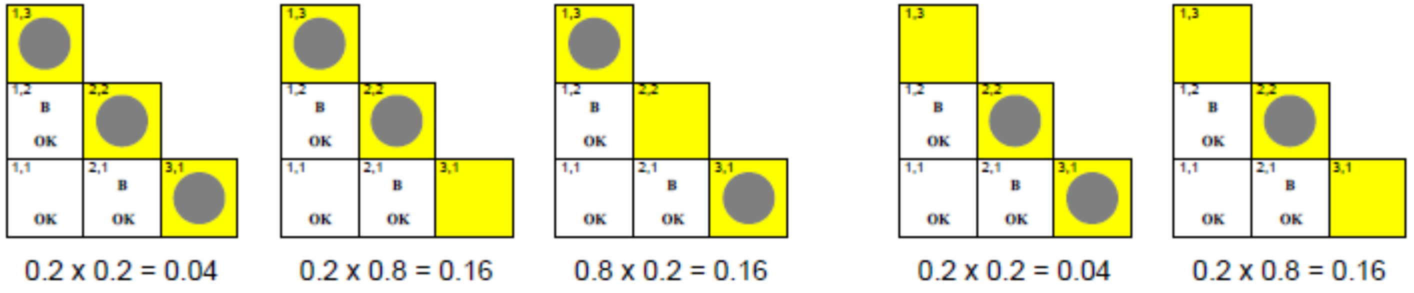
Define $Unknown = Fringe \cup Other$
$\mathbf{P}(b|P_{1,3}, Known, Unknown) = \mathbf{P}(b|P_{1,3}, Known, Fringe)$

Manipulate query into a form where we can use this!

Using conditional independence contd.

$$
\begin{aligned}
\mathbf{P}(P_{1,3}|known, b) &= \alpha \sum_{unknown} \mathbf{P}(P_{1,3}, unknown, known, b) \\
&= \alpha \sum_{unknown} \mathbf{P}(b|P_{1,3}, known, unknown)\mathbf{P}(P_{1,3}, known, unknown) \\
&= \alpha \sum_{fringe} \sum_{other} \mathbf{P}(b|known, P_{1,3}, fringe, other)\mathbf{P}(P_{1,3}, known, fringe, other) \\
&= \alpha \sum_{fringe} \sum_{other} \mathbf{P}(b|known, P_{1,3}, fringe)\mathbf{P}(P_{1,3}, known, fringe, other) \\
&= \alpha \sum_{fringe} \mathbf{P}(b|known, P_{1,3}, fringe) \sum_{other} \mathbf{P}(P_{1,3}, known, fringe, other) \\
&= \alpha \sum_{fringe} \mathbf{P}(b|known, P_{1,3}, fringe) \sum_{other} \mathbf{P}(P_{1,3})P(known)P(fringe)P(other) \\
&= \alpha \, P(known)\mathbf{P}(P_{1,3}) \sum_{fringe} \mathbf{P}(b|known, P_{1,3}, fringe)P(fringe) \sum_{other} P(other) \\
&= \alpha' \, \mathbf{P}(P_{1,3}) \sum_{fringe} \mathbf{P}(b|known, P_{1,3}, fringe)P(fringe)
\end{aligned}
$$

Using conditional independence contd.



$$\mathbf{P}(P_{1,3}|known, b) = \alpha' \langle 0.2(0.04 + 0.16 + 0.16),\ 0.8(0.04 + 0.16) \rangle$$
$$\approx \langle 0.31, 0.69 \rangle$$

$$\mathbf{P}(P_{2,2}|known, b) \approx \langle 0.86, 0.14 \rangle$$

Summary

Probability is a rigorous formalism for uncertain knowledge

Joint probability distribution specifies probability of every atomic event

Queries can be answered by summing over atomic events

For nontrivial domains, we must find a way to reduce the joint size

Independence and conditional independence provide the tools