# Uncertainty

Handling Uncertainty with Degrees of Belief

Uncertainty due to

- partial observability - agent may not know for certain what state it is in
- non-determinism - agent may not know for certain where it will end up after a sequence of actions
- or both

Problem Solving Agents/logical agents handle uncertainty by

- keeping track of a belief state –a representation of the set of all possible world states that it might be in
- and generating a contingency plan that
- handles every possible eventuality that its sensors may report during execution

# Uncertainty

Drawbacks

When interpreting partial sensor information, a logical agent must consider every logically
- possible explanation for the observations, no matter how unlikely.
- This leads to impossible large and complex belief-state representations.

A correct contingent plan that handles every eventuality can grow arbitrarily large
- And  must consider arbitrarily unlikely contingencies.

Sometimes there is no plan that is guaranteed to achieve the goal
- yet the agent must act.
- It must have some way to compare the merits of plans that are not guaranteed.

# Uncertainty

Example

an automated taxi has the goal of delivering a passenger to the airport on time. Airport is 7 km away

The agent forms a plan,

A90, that involves leaving home 90 minutes before the flight departs and driving at a reasonable speed

a logical taxi agent will not be able to conclude with certainty that

“Plan A90 will get us to the airport in time.”

it reaches the weaker conclusion

“Plan A90 will get us to the airport in time, as long as the car doesn’t break down or run out of gas, and I don’t get into an accident, and there are no accidents on the bridge, and the plane doesn’t leave early, and no meteorite hits the car, and . . . .”

None of these conditions can be deduced for sure, so the plan’s success cannot be inferred - Qualification Problem

# Uncertainty

In some sense A90 is in fact the right thing to do out of all the plans that could be executed -

A90 is expected to maximize the agent's performance measure

(the expectation is relative to the agent's knowledge about the environment)

The performance measure includes

getting to the airport in time for the flight,

avoiding a long, unproductive wait at the airport,

avoiding speeding tickets along the way.

**The agent's knowledge cannot guarantee any of these outcomes for A90, but it can provide some degree of belief that they will be achieved.**

Other plans, such as A180, might increase the agent's belief that it will get to the airport on

time, but also increase the likelihood of a long wait.

The right thing to do—the rational decision—

depends on both the relative importance of various goals and

the likelihood that, and degree to which, they will be achieved

# Uncertainty

**Toothache ⇒ Cavity**

This rule is wrong as not all patients with toothaches have cavities; some of them have gum disease, an abscess, or one of several other problems:

**Toothache ⇒ Cavity ∨ GumProblem ∨ Abscess . . .**

in order to make the rule true, we have to add an almost unlimited list of possible problems.

try turning the rule into a causal rule:

**Cavity ⇒ Toothache**

this rule is not right either; not all cavities cause pain!

way to fix the rule is to make it logically exhaustive:

# Uncertainty

Trying to use logic to cope with a domain like medical diagnosis thus fails for three main reasons:

**Laziness:** It is too much work to list the complete set of antecedents or consequents needed to ensure an exception less rule and too hard to use such rules.

**Theoretical ignorance:** Medical science has no complete theory for the domain.

**Practical ignorance:** Even if we know all the rules, we might be uncertain about a particular patient because not all the necessary tests have been or can be run.

The connection between toothaches and cavities is just not a logical consequence in either direction.

typical of the medical domain, as well as most other judgmental domains

The Agents knowledge can at best provide only a **degree of belief** in the relevant sentences.

main tool for dealing with degrees of belief is probability theory

# Uncertainty

a logical agent believes each sentence to be true or false or has no opinion

a probabilistic agent may have a numerical degree of belief

between 0 (for sentences that are certainly false)

and 1 (certainly true)

Probability provides a way of summarizing the uncertainty that comes from our laziness and ignorance,

thereby solving the qualification problem.

We might not know for sure what afflicts a particular patient,

but we believe that there is, say, an 80% chance—that is, a probability of 0.8—that the patient who has a toothache has a cavity

That is, we expect that out of all the situations that are indistinguishable from the current situation as far as our knowledge goes, the patient will have a cavity in 80% of them.

This belief could be derived from statistical data—80% of the toothache patients seen so far have had cavities—or from some general dental knowledge, or from a combination of evidence sources.

# Probability

Probability statements are made with respect to a knowledge state, not with respect to the real world
at the time of our diagnosis, there is no uncertainty in the actual world: the patient either has a cavity or doesn't.

"The probability that the patient has a cavity, given that she has a toothache, is 0.8."

If we later learn that the patient has a history of gum disease, a different statement -
"The probability that the patient has a cavity, given that she has a toothache and a history of gum disease, is 0.4."

If we gather further conclusive evidence against a cavity, we can say
"The probability that the patient has a cavity, given all we now know, is almost 0."

these statements do not contradict each other;
each is a separate assertion about a different knowledge state.

## Uncertainty and rational decisions

A90 plan for getting to the airport. - Suppose it gives us a 97% chance of catching our flight - a rational choice?
there might be other plans, such as A180, with higher probabilities

If it is vital not to miss the flight, then it is worth risking the longer wait at the airport.

A1440, a plan that involves leaving home 24 hours in advance? - not a good choice,
it almost guarantees getting there on time, but it involves an intolerable wait

an agent must first have preferences between the different possible outcomes of the various plans.
An outcome is a completely specified state, including such factors
as whether the agent arrives on time and the length of the wait at the airport.

# Uncertainty and rational decisions

Utility theory to represent and reason with preferences:

Utility theory says that every state has a degree of usefulness, or utility, to an agent and that the agent will prefer states with higher utility.

Decision Theory: Probability Theory + Utility Theory

MEU: Principle of Maximum Expected Utility :

An agent is rational if and only if it chooses the action that yields the highest **expected** utility, averaged over all the possible outcomes of the action

expected = statistical mean of the outcomes, weighted by the probability of the outcome

# Assertions with Probability

Like logical assertions, probabilistic assertions are about possible worlds.

Logical assertions say which possible worlds are strictly ruled out (all those in which the assertion is false),

**Probabilistic assertions talk about how probable the various worlds are.**

**Sample space:** set of all possible worlds

possible worlds are mutually exclusive and exhaustive

example

if we are about to roll two (distinguishable) dice,

 there are 36 possible worlds to consider: (1,1), (1,2), . . ., (6,6).

 **Ω** (uppercase omega) refers to the sample space

 **ω** (lowercase omega) refers to elements of the space - particular possible worlds.

# Probability

probability model associates a numerical probability P(ω) with each possible world.

**Basic axioms of probability theory:**

every possible world has a probability between 0 and 1 and the total probability of the set of possible worlds is 1:

$$0 \le P(\omega) \le 1 \text{ for every } \omega \qquad \text{and} \qquad \sum_{\omega \in \Omega} P(\omega) = 1$$

example

assume that each die is fair and the rolls don't interfere with each other,

then each of the possible worlds (1,1), (1,2), . . ., (6,6)

has probability 1/36.

Probabilistic assertions and queries are not usually about particular possible worlds,

but about sets of them - **Events**

example: sets in the cases where the two dice add up to 11, the cases where doubles are rolled.

(1,1), (2,2), (3, 3), (4, 4), (5, 5), (6, 6)

Sets are described by propositions

# Probability

For each proposition,

the corresponding set contains just those possible worlds in which the proposition holds.

The probability associated with a proposition is sum of the probabilities of the worlds in which it holds:

**For any proposition φ,** $$P(\varphi) = \sum_{\omega \in \Omega} P(\omega)$$

example, when rolling fair dice, we have

P(Total = 11) = P((5, 6)) + P((6, 5)) = 1/36 + 1/36 = 1/18

**Unconditional Probabilities** :

P(Total =11) and P(doubles) are called unconditional or prior probabilities

they refer to **degrees of belief** in propositions **in the absence of any other information.**

**Evidence**: Most of the time, we have some information, that has already been revealed.

# Probability

Case: the first die may already be showing a 5 and the other one is yet to show

we are interested not in the unconditional probability of rolling doubles,

but the **conditional or posterior probability** of rolling doubles given that the first die is a 5.

**Conditional Probability: P(doubles | Die1 =5)**

example

going to the dentist for a regular checkup, the probability

P(cavity) = 0.2

but going to the dentist because of a toothache,

it's P(cavity | toothache) = 0.6

P(cavity)=0.2 is still valid after toothache is observed

it just isn't especially useful.

**When making decisions, an agent needs to condition on all the evidence it has observed**

# Probability

The assertion that

$P(\text{cavity} \mid \text{toothache}) = 0.6$

does not mean

"Whenever toothache is true, conclude that cavity is true with probability 0.6"

rather it means

**"Whenever toothache is true and we have no further information, conclude that cavity is true with probability 0.6."**

Example

if we had the further information that the dentist found no cavities,

we do not conclude that cavity is true with probability 0.6

instead we use

$P(\text{cavity} \mid \text{toothache} \wedge \neg\text{cavity}) = 0.$

# Conditional Probability

**Conditional Probability:**

for any propositions a and b,

$$P(a \mid b) = P(a \wedge b) \;/\; P(b) \qquad \text{whenever } P(b) > 0$$

**b rules out all those possible worlds where b is false**.

The set now is a set whose **Total P is just P(b)**

**Within this set, 'a' worlds satisfy a ∧ b** and

constitute a **fraction P(a ∧ b) / P(b)**

example

$$P(\text{doubles} \mid \text{Die1} = 5) = P(\text{doubles} \wedge \text{Die1} = 5) \;/\; P(\text{Die1} = 5)$$

**PRODUCT RULE : $P(a \wedge b) = P(a \mid b)\, P(b)$**

**For a ∧ b to be true, we need b to be true and we also need a to be true given b**

# Propositions in Probability assertions

factored representation: a possible world is represented by a set of variable/value pairs.

**random variables**: Variables in probability theory

names begin with an uppercase letter

**DOMAIN** : Every random variable has a domain—the set of possible values it can take on.

example

Total and Die1 are random variables.

domain of Total for two dice is the set {2, . . . , 12}

domain of Die1 is {1, . . . , 6}.

A Boolean random variable has the domain {true, false}

**values** are in lowercase

# Propositions in Probability assertions

**Proposition**: Doubles are rolled:

Doubles = true or Doubles

Doubles = false or ¬ Doubles

Domains: sets of arbitrary tokens

Domain of Weather = {sunny, rain, cloudy, snow} finite domains

Weather =sunny value assignment when no ambiguity

Variables can have infinite domains, either discrete (like the integers) or continuous (like the reals).

For any variable with an ordered domain, inequalities are also allowed

$$NumberOfAtomsInUniverse \geq 10^{70}$$

**elementary propositions can be combined using the connectives of propositional logic**

"The probability that the patient has a cavity, given that she is a teenager with no toothache, is 0.1"

$$\mathbf{P(cavity \mid \neg toothache \land teen) = 0.1}$$

# Propositions in Probability assertions

probabilities of all the possible values of a random variable:

P(Weather =sunny) = 0.6

P(Weather =rain) = 0.1

P(Weather =cloudy) = 0.29

P(Weather =snow) = 0.01 ,

Or

**P** (Weather) = ‹ 0.6, 0.1, 0.29, 0.01 ›

bold **P** indicates that the result is a vector of numbers,

and predefined ordering ‹ sunny, rain, cloudy, snow › on the domain of Weather is assumed

**P** defines a probability distribution for the random variable Weather

**for conditional distributions:**

**P(**X | Y **) gives the values of P(**X = xi | Y = yj**) for each possible i, j pair.**

# Probability

Distributions on multiple variables

example

**joint probability distribution** of Weather and Cavity

$\mathbf{P}$(Weather , Cavity) = $\mathbf{P}$(Weather | Cavity) $\mathbf{P}$(Cavity) denotes

the probabilities of all combinations of the values of Weather and Cavity (4 x 2 table?)

$\mathbf{P}$(sunny, Cavity) mixing variables with and without values

4 x 2 = 8 equations

$$
\begin{aligned}
P(W=\text{sunny} \wedge C=\text{true}) &= P(W=\text{sunny} \mid C=\text{true})\, P(C=\text{true}) \\
P(W=\text{rain} \wedge C=\text{true}) &= P(W=\text{rain} \mid C=\text{true})\, P(C=\text{true}) \\
P(W=\text{cloudy} \wedge C=\text{true}) &= P(W=\text{cloudy} \mid C=\text{true})\, P(C=\text{true}) \\
P(W=\text{snow} \wedge C=\text{true}) &= P(W=\text{snow} \mid C=\text{true})\, P(C=\text{true}) \\
P(W=\text{sunny} \wedge C=\text{false}) &= P(W=\text{sunny} \mid C=\text{false})\, P(C=\text{false}) \\
P(W=\text{rain} \wedge C=\text{false}) &= P(W=\text{rain} \mid C=\text{false})\, P(C=\text{false}) \\
P(W=\text{cloudy} \wedge C=\text{false}) &= P(W=\text{cloudy} \mid C=\text{false})\, P(C=\text{false}) \\
P(W=\text{snow} \wedge C=\text{false}) &= P(W=\text{snow} \mid C=\text{false})\, P(C=\text{false})
\end{aligned}
$$

# Probability

**$0 \leq P(\omega) \leq 1$ for every $\omega$ and $\sum_{\omega \in \Omega} P(\omega) = 1$ (1)**

**For any proposition $\varphi$, $P(\varphi) = \sum_{\omega \in \Omega} P(\omega)$ (2)**

probability of a proposition is the sum of the probabilities of worlds in which it holds.
possible world: an assignment of values to all of the random variables under consideration.

Example: if the random variables are Cavity, Toothache, and Weather ,
then there are $2 \times 2 \times 4 = 16$ possible worlds

**full joint probability distribution:**
probability model is completely determined by the joint distribution for all of the random variables

**Every proposition's probability is a sum over possible worlds.**
**a full joint distribution suffices, for calculating the probability of any proposition.**

Probability

$$P(\neg a) = \sum_{\omega \in \neg a} P(\omega)$$

$$= \sum_{\omega \in \neg a} P(\omega) + \sum_{\omega \in a} P(\omega) - \sum_{\omega \in a} P(\omega)$$

$$= \sum_{\omega \in \Omega} P(\omega) - \sum_{\omega \in a} P(\omega)$$

$$= 1 - P(a)$$

Probability of a disjunction/ inclusion–exclusion principle:

$$P(a \vee b) = P(a) + P(b) - P(a \wedge b) \qquad (3)$$

Product Rule

$$P(a \wedge b) = P(a \mid b)\, P(b) \qquad (4)$$

(1),(2), (3), (4) - Kolmogorov's axioms

# Probability - Inference

Probabilistic inference

computation of posterior probabilities for query propositions given observed evidence.

Knowledge base : full joint distribution

A full joint distribution for the Toothache, Cavity, Catch world.

| | toothache | | ¬toothache | |
|---|---|---|---|---|
| | catch | ¬catch | catch | ¬catch |
| cavity | 0.108 | 0.012 | 0.072 | 0.008 |
| ¬cavity | 0.016 | 0.064 | 0.144 | 0.576 |

Domain: 3 Boolean variables: Toothache, Cavity, Catch

Full joint distribution is a $2 \times 2 \times 2$ table

Probabilities in the joint distribution sum to 1,

# Probability of a proposition

Calculating the probability of any proposition, simple or complex:

simply identify those possible worlds in which the proposition is true and add up their probabilities

| | toothache | | ¬toothache | |
|---|---|---|---|---|
| | catch | ¬catch | catch | ¬catch |
| cavity | 0.108 | 0.012 | 0.072 | 0.008 |
| ¬cavity | 0.016 | 0.064 | 0.144 | 0.576 |

**cavity ∨ toothache holds in six possible worlds**

$$P(\text{cavity} \lor \text{toothache}) = 0.108 + 0.012 + 0.072 + 0.008 + 0.016 + 0.064 = 0.28$$

**Unconditional/marginal probability of cavity** (extract the distribution over some subset of variables or a single variable.)

$$P(\text{cavity}) = 0.108 + 0.012 + 0.072 + 0.008 = 0.2$$

**Marginalization/ summing out:**

sum up the probabilities for each possible value of the other variables, thereby taking them out of the equation

## Probability of a proposition – inference using joint probabilities

general marginalization rule for any sets of variables **Y** and **Z**

$$\mathbf{P}(\mathrm{Y}) = \sum_{z \in \mathbf{Z}} \mathbf{P}(\mathrm{Y}, z) = \sum_{z} \mathbf{P}(\mathrm{Y}, z)$$

$\sum_{z \in \mathbf{Z}}$ means to sum over all the possible combinations of values of the set of variables Z.

$$\mathbf{P}(\mathrm{Cavity}) = \sum_{z \in \{\mathrm{Catch},\ \mathrm{Toothache}\}} \mathbf{P}(\mathrm{Cavity}, z) \qquad (\text{Joint probabilities})$$

**Using Conditional Probabilities , the same can be expressed as**

$$\mathbf{P}(\mathrm{Y}) = \sum_{z \in \mathbf{Z}} \mathbf{P}(\mathrm{Y} \mid z)\, \mathrm{P}(z)$$

In most cases, we are interested in computing conditional probabilities of some variables , given evidence about others.

# Probability of a proposition – inference using joint probabilities

Conditional probabilities can be found by using

Conditional Probability: for any propositions a and b, ( in terms of unconditional probabilities)

$$P(a \mid b) = P(a \land b) \ / \ P(b)$$

|  | toothache |  | ¬toothache |  |
|---|---|---|---|---|
|  | catch | ¬catch | catch | ¬catch |
| cavity | 0.108 | 0.012 | 0.072 | 0.008 |
| ¬cavity | 0.016 | 0.064 | 0.144 | 0.576 |

1. probability of a cavity, given evidence of a toothache

$$P(cavity \mid toothache) = P(cavity \land toothache) \ / \ P(toothache)$$

$$= (0.108 + 0.012) \ / \ (0.108 + 0.012 + 0.016 + 0.064)$$

$$= 0.6$$

2. probability that there is no cavity, given a toothache

$$P(\neg cavity \mid toothache) = P(\neg cavity \land toothache) / \ P(toothache)$$

$$= (0.016 + 0.064) \ / (0.108 + 0.012 + 0.016 + 0.064)$$

$$= 0.4$$

$$P(cavity \mid toothache) + P(\neg cavity \mid toothache) = 1$$

in these two calculations the term 1/P(toothache ) remains constant, no matter which value of Cavity we calculate.

can be viewed as a normalization constant for the distribution P(Cavity | toothache)

# Probability of a proposition – inference using joint probabilities

$\mathbf{P}(Cavity \mid toothache) = \alpha\, \mathbf{P}(Cavity, toothache)$

$= \alpha\, [\mathbf{P}(Cavity, toothache, catch) + \mathbf{P}(Cavity, toothache, \neg catch)]$

$= \alpha\, [\langle 0.108, 0.016 \rangle + \langle 0.012, 0.064 \rangle] = \alpha\, \langle 0.12, 0.08 \rangle = \langle 0.6, 0.4 \rangle$

| | toothache | | ¬toothache | |
|---|---|---|---|---|
| | catch | ¬catch | catch | ¬catch |
| cavity | 0.108 | 0.012 | 0.072 | 0.008 |
| ¬cavity | 0.016 | 0.064 | 0.144 | 0.576 |

$\mathbf{P}(Cavity \mid toothache)$ can be calculated even if the value of P(toothache)! Is not known!

Ignoring the factor 1/P(toothache ) and adding up the values for cavity and ¬cavity,
Gets us 0.12 and 0.08.

Those are the correct relative proportions, but they don't sum to 1.
Normalizing them by dividing each one by 0.12 + 0.08, gets the true probabilities of 0.6 and 0.4.

Normalization makes the computation easier and
allows us to proceed when some probability assessment (such as P(toothache)) is not available.

general inference procedure.

Assuming query involves a single variable, X (Cavity)

Let E be the list of evidence variables (Toothache),

let e be the list of observed values for them, and

let Y be the remaining unobserved variables (Catch ).

The query is P(X | e) and can be evaluated as

$$\mathbf{P}(X \mid e) = \alpha\, P(X, e) = \alpha \sum_{y \in \mathbf{Y}} \mathbf{P}(X, e, y)$$

together the variables X, E, and Y constitute the complete set of variables for the domain,

so P(X, e, y) is simply a subset of probabilities from the full joint distribution.

Given the full joint distribution to work with, can answer probabilistic queries for discrete variables.

> for a domain described by n Boolean variables, it requires an input table of size $O(2^n)$ and takes $O(2^n)$ time to process the table!

not a practical tool for building reasoning systems

# Independence

full joint distribution for 4 variables

$$P(\text{Toothache, Catch, Cavity, Weather}) \qquad \text{with entries: } 2 \times 2 \times 2 \times 4 = 32$$

how are P(toothache, catch, cavity, cloudy) and P(toothache, catch, cavity) related? Using product rule

$$P(\text{toothache, catch, cavity, cloudy}) = P(\text{cloudy} \mid \text{toothache, catch, cavity})\, P(\text{toothache, catch, cavity})$$

one should not imagine that one's dental problems influence the weather. And for indoor dentistry, at least, it seems safe to say that the weather does not influence the dental variables. Therefore, the following assertion seems reasonable:

$$\mathbf{P(\text{cloudy} \mid \text{toothache, catch, cavity}) = P(\text{cloudy})}$$

**property of independence/ marginal independence/ absolute independence**

From this, we can deduce

$$\mathbf{P(\text{toothache, catch, cavity, cloudy}) = P(\text{cloudy})\, P(\text{toothache, catch, cavity})}$$

# Independence

General equation

$$P(Toothache, Catch, Cavity, Weather) = P(Toothache, Catch, Cavity)\, P(Weather)$$

the 32-element table for four variables can be constructed from one 8-element table and one 4-element table.
weather is independent of one's dental problems.

**Independence** between propositions a and b

$P(a \mid b) = P(a)$ or

$P(b \mid a) = P(b)$ or

$P(a \wedge b) = P(a)\, P(b)$ ( all are equivalent)

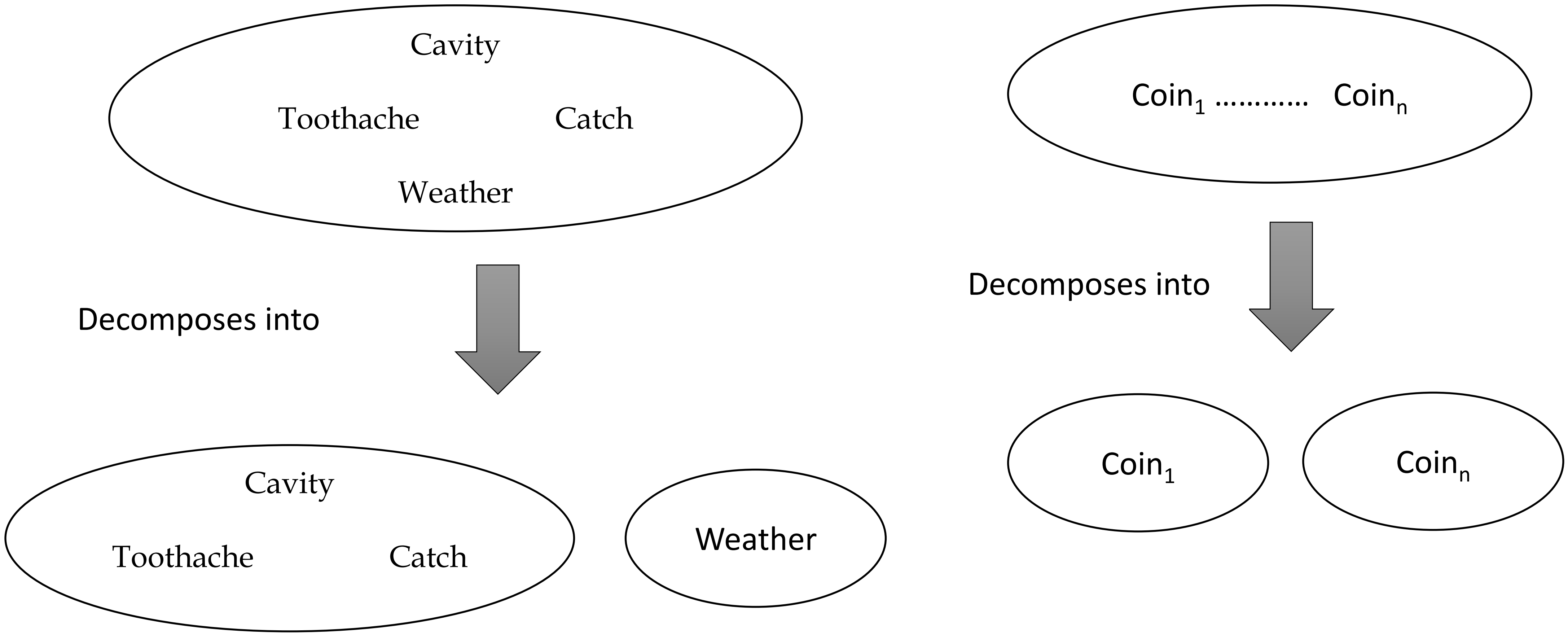
Independence between variables X and Y can be written as

$P(X \mid Y) = P(X)$ or

$P(Y \mid X) = P(Y)$ or

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

factoring a large joint distribution into smaller distributions, using absolute independence. (a) Weather and dental problems are independent. (b) Coin flips are independent.



# Independence

Independence of assertions reduces the amount of information necessary to specify the full joint distribution.

If the complete set of variables can be divided into independent subsets, then the full joint distribution can be factored into separate joint distributions on those subsets.

example

the full joint distribution on the outcome of n independent coin flips, P(C1, . . . ,Cn), has $2^n$ entries,

but it can be represented as the product of n single-variable distributions P(Ci)

even independent subsets can be quite large

example,

dentistry might involve dozens of diseases and hundreds of symptoms, all of which are interrelated.

# Bayes' Rule

Product Rule:

$P(a \wedge b) = P(a \mid b) P(b)$

$P(a \wedge b) = P(b \mid a) P(a)$ $\qquad$ $P(a \wedge b) = P(b \wedge a)$

Equating RHS and dividing by P(a)

$P(b \mid a) = P(a \mid b) P(b) / P(a)$ $\qquad$ Bayes' Rule

$\mathbf{P}(Y \mid X) = \mathbf{P}(X \mid Y) \mathbf{P}(Y) / \mathbf{P}(X)$ $\qquad$ general case for multivalued variables

$P(Y \mid X, e) = P(X \mid Y, e) P(Y \mid e) / P(X \mid e)$ $\qquad$ more general version conditionalized on some background evidence e

## Bayes’ rule:

Often, we perceive as evidence the effect of some unknown cause and we would like to determine that cause. In that case, Bayes’ rule becomes

P(cause | effect) = P(effect | cause) P(cause) / P(effect)

The conditional probability

P(effect | cause) quantifies the relationship in the causal direction,

whereas

P(cause | effect) describes the diagnostic direction.

In medical diagnosis, we often have conditional probabilities on causal relationships

the doctor knows P(symptoms | disease)) and

want to derive a diagnosis, P(disease | symptoms)

.

# Bayes' Rule

example,

a doctor knows that the disease meningitis causes the patient to have a stiff neck, say, 70% of the time.

The doctor also knows some unconditional facts:

the prior probability that a patient has meningitis is 1/50,000, and

the prior probability that any patient has a stiff neck is 1%.

proposition that the patient has a stiff neck: s

proposition that the patient has meningitis: m

$P(s/m) = 0.7$

$P(m) = 1/50000$

$P(s) = 0.01$

$P(m/s) = P(s \mid m)\, P(m) \,/\, P(s) = (0.7 \times 1/50000) \,/\, 0.01 = 0.0014$

# Bayes' Rule

we expect less than 1 in 700 patients with a stiff neck to have meningitis.
Though a stiff neck is quite strongly indicated by meningitis (with probability 0.7),
The probability of meningitis in the patient remains small,
Because the prior probability of stiff necks is much higher than that of meningitis.

We can avoid assessing the prior probability of the evidence P(s) by
computing a posterior probability for each value of the query variable (here, m and ¬m) and
then normalizing the results.

$$P(M \mid s) = \alpha \langle P(s \mid m)\, P(m),\ P(s \mid \neg m)\, P(\neg m) \rangle$$

$P(Y \mid X) = \alpha\, P(X \mid Y) P(Y)$ general form of Bayes' rule with normalization

$\alpha$ is the normalization constant needed to make the entries in $P(Y \mid X)$ sum to 1.

# Bayes' Rule

Combining evidence

what can a dentist conclude if the steel probe catches in the aching tooth of a patient?

(we know the full joint distribution)

$P(Cavity \mid toothache \wedge catch) = \alpha < 0.108, 0.016 > \approx <0.871, 0.129 >$

| | toothache | | ¬toothache | |
|---|---|---|---|---|
| | catch | ¬catch | catch | ¬catch |
| cavity | 0.108 | 0.012 | 0.072 | 0.008 |
| ¬cavity | 0.016 | 0.064 | 0.144 | 0.576 |

does not scale up to larger numbers of variables.

Using Bayes Rule -

$$P(Cavity \mid toothache \wedge catch) = \alpha \ P(toothache \wedge catch \mid Cavity) \ P(Cavity)$$

need to know the conditional probabilities of the conjunction toothache ∧ catch for each value of Cavity.

does not scale up.

If there are n possible evidence variables (X rays, diet, oral hygiene, etc.), then there are $2^n$ possible combinations of observed values for which we would need to know conditional probabilities.

# Simplifying expressions

Toothache and Catch - independent ?

if the probe catches in the tooth, then it is likely that the tooth has a cavity and that the cavity causes a toothache.

These variables are independent, however, given the presence or the absence of a cavity.

Each is directly caused by the cavity, but neither has a direct effect on the other:

toothache depends on the state of the nerves in the tooth,

the probe's accuracy depends on the dentist's skill,

to which the toothache is irrelevant

**Conditional independence** of toothache and catch given Cavity

$$\mathbf{P}(toothache \land catch \mid Cavity) = \mathbf{P}(toothache \mid Cavity)\, \mathbf{P}(catch \mid Cavity)$$

Substituting in Bayes' rule:

$$\mathbf{P}(Cavity \mid toothache \land catch) = \alpha\, \mathbf{P}(toothache \land catch \mid Cavity)\, \mathbf{P}(Cavity)$$

$$\mathbf{P}(Cavity \mid toothache \land catch) = \alpha\, \mathbf{P}(toothache \mid Cavity)\, \mathbf{P}(catch \mid Cavity)\, \mathbf{P}(Cavity)$$

## Simplifying expressions

information requirements are the same as for inference, using each piece of evidence separately

the prior probability P(Cavity) for the query variable and

the conditional probability of each effect, given its cause.

general definition of conditional independence of two variables X and Y , given a third variable Z :

$$\mathbf{P}(X, Y \mid Z) = \mathbf{P}(X \mid Z)\, \mathbf{P}(Y \mid Z)$$

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

absolute independence assertions allow a decomposition of the full joint distribution into much smaller pieces

same is true for conditional independence assertions

$$\begin{aligned}
&\mathbf{P}(\text{Toothache}, \text{Catch}, \text{Cavity}) \\
&= \mathbf{P}(\text{Toothache}, \text{Catch} \mid \text{Cavity})\, \mathbf{P}(\text{Cavity}) \qquad \text{(product rule)} \\
&= \mathbf{P}(\text{Toothache} \mid \text{Cavity})\, \mathbf{P}(\text{Catch} \mid \text{Cavity})\, \mathbf{P}(\text{Cavity})
\end{aligned}$$

# Simplifying expressions

the original large table is decomposed into three smaller tables.

the size of the representation grows as $O(n)$ instead of $O(2^n)$

conditional independence assertions can allow probabilistic systems to scale up

They are much more commonly available than absolute independence assertions

decomposition of large probabilistic domains into weakly connected subsets through conditional independence

Useful for AI.

commonly occurring pattern: a single cause directly influences a number of effects, all of which are conditionally independent, given the cause.

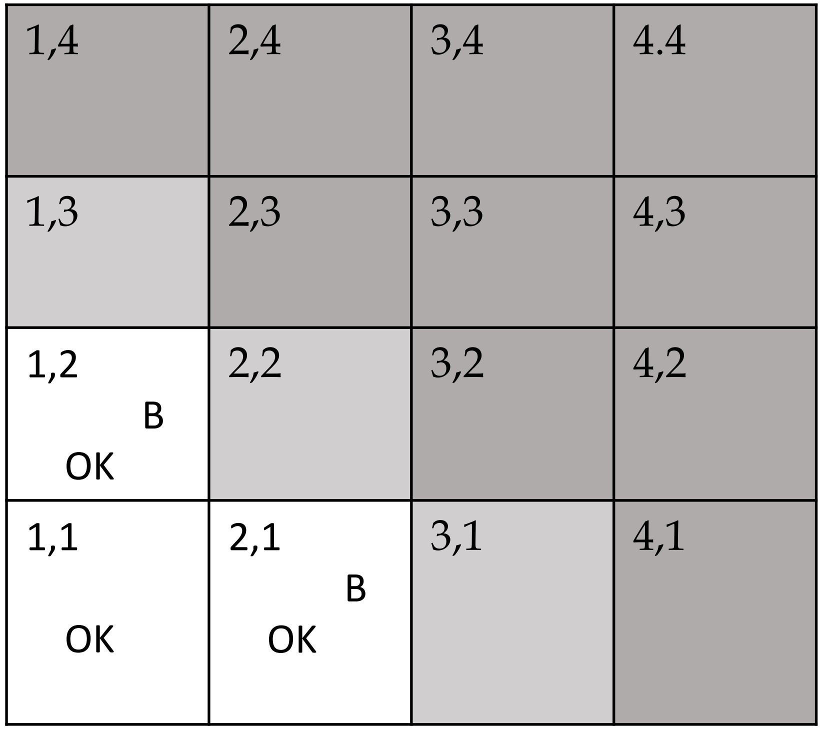
full joint distribution

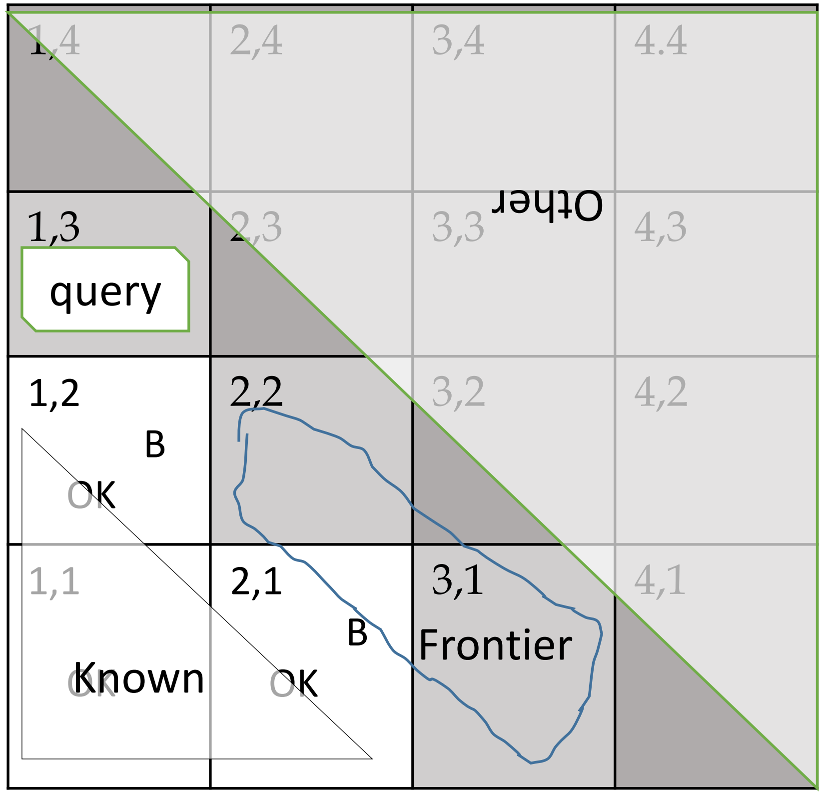
$$P(\text{Cause}, \text{Effect}_1, \ldots, \text{Effect}_n) = \mathbf{P}(\text{Cause}) \prod_i P(\text{Effect}_i \mid \text{Cause})$$

naive Bayes model/Bayesian Classifier

# Wumpus World Example – Probabilistic Agent

Uncertainty in Wumpus World:  agent's sensors give only partial information about the world.




Each of the three squares -  [1,3], [2,2], [3,1]  - might contain a pit.

Which square is most likely  to be safe? Calculate the probability that each of the three squares contains a pit.

(1)  a pit causes breezes in all neighboring squares, and

(2)  each square other than [1,1] contains a pit with probability 0.2.

# Wumpus World Example – Probabilistic Agent

one Boolean variable Pij for each square, which is true if square [i, j] actually contains a pit.

Boolean variables Bij that are true if square [i, j] is breezy

Consider B ij , only for the observed squares—in this case, [1,1], [1,2], and [2,1]

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

Applying the product rule -

$$\mathbf{P}(P_{1,1}, \ldots, P_{4,4}, B_{1,1}, B_{1,2}, B_{2,1}) = \mathbf{P}(B_{1,1}, B_{1,2}, B_{2,1} \mid P_{1,1}, \ldots, P_{4,4})\, \mathbf{P}(P_{1,1}, \ldots, P_{4,4})$$

$\mathbf{P}(B_{1,1}, B_{1,2}, B_{2,1} \mid P_{1,1}, \ldots, P_{4,4})$ - first term

conditional probability distribution of a breeze configuration, given a pit configuration;

its values are 1 if the breezes are adjacent to the pits and 0 otherwise.

# Wumpus World Example – Probabilistic Agent

$\mathbf{P}(P1,1, \ldots, P4,4)$ - $2^{nd}$ term

prior probability of a pit configuration.

Each square contains a pit with probability 0.2, independently of the other squares

$$\mathbf{P}(P1,1, \ldots, P4,4) = \prod_{i,j=1,1}^{4,4} \mathbf{P}(Pi,j)$$

For a particular configuration with exactly n pits,

$$P(P1,1, \ldots, P4,4) = 0.2^{n} \times 0.8^{16-n}$$

Evidence observed breeze ( absence of it) in each square that is visited, combined with the fact that each such square contains no pit

$$b = \neg b1,1 \wedge b1,2 \wedge b2,1 \qquad known = \neg p1,1 \wedge \neg p1,2 \wedge \neg p2,1$$

general inference procedure.

Assuming query involves a single variable, X (Cavity)

Let E be the list of evidence variables (Toothache),

let e be the list of observed values for them, and

let Y be the remaining unobserved variables (Catch ).

The query is P(X | e) and can be evaluated as

$$\mathbf{P}(X \mid e) = \alpha\, P(X, e) = \alpha \sum_{y \in \mathbf{Y}} \mathbf{P}(X, e, y)$$

together the variables X, E, and Y constitute the complete set of variables for the domain,

so P(X, e, y) is simply a subset of probabilities from the full joint distribution.

Given the full joint distribution to work with, can answer probabilistic queries for discrete variables.

> for a domain described by n Boolean variables, it requires an input table of size $O(2^n)$ and takes $O(2^n)$ time to process the table!

not a practical tool for building reasoning systems

# Wumpus World Example – Probabilistic Agent

Query

$P(P1,3 \mid \text{known}, b)$: how likely is it that [1,3] contains a pit, given the observations so far?

Standard approach:

summing over entries from the full joint distribution

Let Unknown be the set of $P_{i,j}$ variables for squares other than the Known squares and the query square [1,3].

$$\mathbf{P}(P_{1,3} \mid \text{known}, b) = \alpha \sum_{\text{unknown}} \mathbf{P}(P_{1,3}, \text{unknown}, \text{known}, b)$$

12 unknown squares ➔ summation contains $2^{12} = 4096$ terms.

Other squares are irrelevant? How could [4,4] affect whether [1,3] has a pit?

# Wumpus World Example – Probabilistic Agent

Let Frontier be the pit variables (other than the query variable) that are adjacent to visited squares,

in this case just [2,2] and [3,1]

Let Other be the pit variables for the other unknown squares

in this case, there are 10 other squares

key insight is that the observed breezes are

**conditionally independent of the other variables, given the known, frontier, and query variables**

$\mathbf{P}(P_{1,3} \mid known, b)$

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

$$= \alpha \sum_{unknown} \mathbf{P}(b \mid P_{1,3}, known, unknown)\, \mathbf{P}(P_{1,3},, known, unknown) \quad \text{: product rule}$$

$$= \alpha \sum_{frontier} \sum_{other} \mathbf{P}(b \mid known, P_{1,3}, frontier, other)\, \mathbf{P}(P_{1,3},, known, frontier, other)$$

$$= \alpha \sum_{frontier} \sum_{other} \mathbf{P}(b \mid known, P_{1,3}, frontier)\, \mathbf{P}(P_{1,3},, known, frontier, other) \quad \text{: conditional independence}$$

# Wumpus World Example – Probabilistic Agent

$\mathbf{P}(P_{1,3} \mid known, b)$

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

$$= \alpha \sum_{unknown} \mathbf{P}(b \mid P_{1,3}, known, unknown)\, \mathbf{P}(P_{1,3}, known, unknown) \quad \text{: product rule}$$

$$= \alpha \sum_{frontier} \sum_{other} \mathbf{P}(b \mid known, P_{1,3}, frontier, other)\, \mathbf{P}(P_{1,3}, known, frontier, other)$$

$$= \alpha \sum_{frontier} \sum_{other} \mathbf{P}(b \mid known, P_{1,3}, frontier)\, \mathbf{P}(P_{1,3}, known, frontier, other) \quad \text{: conditional independence}$$

$$= \alpha \sum_{frontier} \mathbf{P}(b \mid known, P_{1,3}, frontier) \sum_{other} \mathbf{P}(P_{1,3}, known, frontier, other) \quad \text{: first term does not depend on other}$$

$$= \alpha \sum_{frontier} \mathbf{P}(b \mid known, P_{1,3}, frontier) \sum_{other} \mathbf{P}(P_{1,3})\, P(known)\, P(frontier)\, P(other) \quad :$$

$$= \alpha\, P(known)\, \mathbf{P}(P_{1,3}) \sum_{frontier} \mathbf{P}(b \mid known, P_{1,3}, frontier)\, P(frontier) \sum_{other} P(other)$$

$$= \alpha'\, \mathbf{P}(P_{1,3}) \sum_{frontier} \mathbf{P}(b \mid known, P_{1,3}, frontier)\, P(frontier)$$

# Wumpus World Example – Probabilistic Agent

the last step folds P(known) into the normalizing constant $\alpha'$

and uses the fact $\sum_{other} P(other) = 1$

$$\mathbf{P}(P_{1,3} \mid known, b) = \alpha' \, \mathbf{P}(P_{1,3}) \sum_{frontier} \mathbf{P}(b \mid known, P_{1,3}, frontier) \, P(frontier)$$

Frontier variables : $P_{2,2}$, $P_{3,1}$ 4 terms

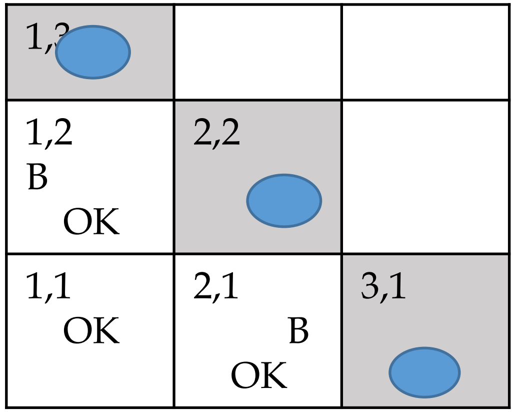
The use of independence and conditional independence has completely eliminated the other squares from consideration.

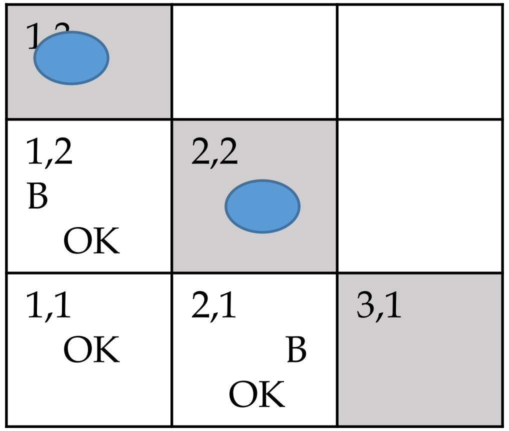
$P(b \mid known, P_{1,3}, frontier)$ is 1 when the frontier is consistent with the breeze observations, and 0 otherwise.

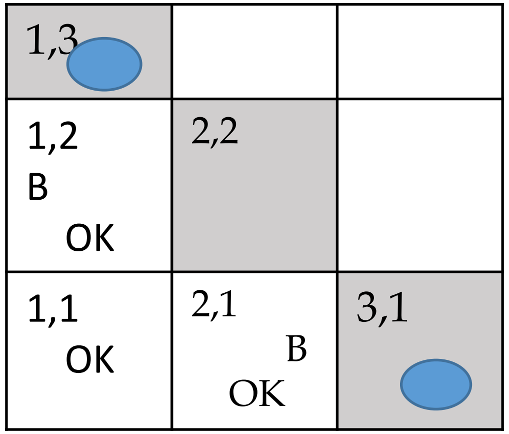
for each value of P1,3, sum over the logical models for the frontier variables that are consistent with the known facts.

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

# Wumpus World Example – Probabilistic Agent



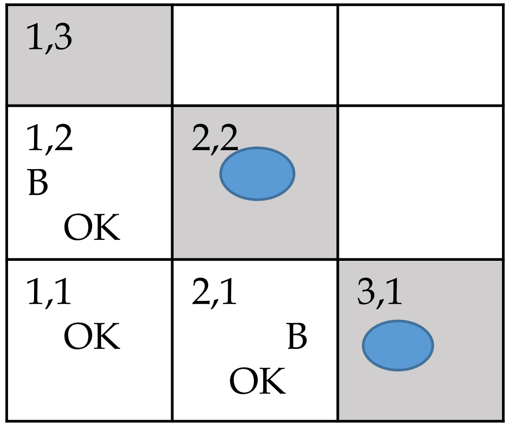



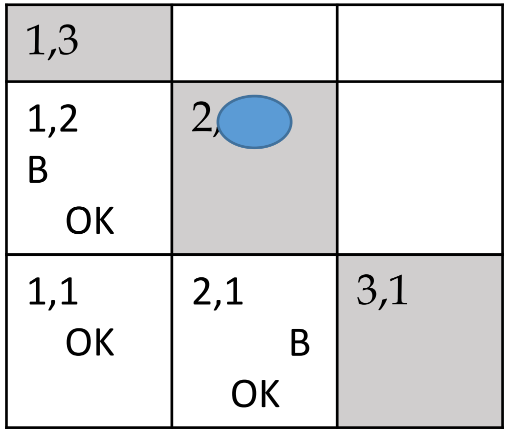




P(frontier) = 0.2 x 0.2 = 0.04 0.2 x 0.8 = 0.16 0.8 x 0.2 = 0.16

Consistent models for frontier variables $P_{2,2}$, $P_{3,1}$ with $P_{1,3}$ = true showing 2 or 3 pits

With with $P_{1,3}$ = false showing 1 or 2 pits




P(frontier) = 0.2 x 0.2 = 0.04 0.2 x 0.8 = 0.16

# Wumpus World Example – Probabilistic Agent

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

[1,3] (and [3,1] by symmetry) contains a pit with roughly 31% probability

[2,2] contains a pit with roughly 86% probability. Calculate.

logical agent did not know that [2,2] was worse than the other squares

Logic can tell us that it is unknown whether there is a pit in [2, 2], but we need probability to tell us how likely it is.

# Wumpus World Example – Probabilistic Agent

# Wumpus World Example – Probabilistic Agent

# Wumpus World Example – Probabilistic Agent