COMP3411/9414: Artificial Intelligence

Module 4

Uncertainty

Russell & Norvig, Chapter 13

Outline

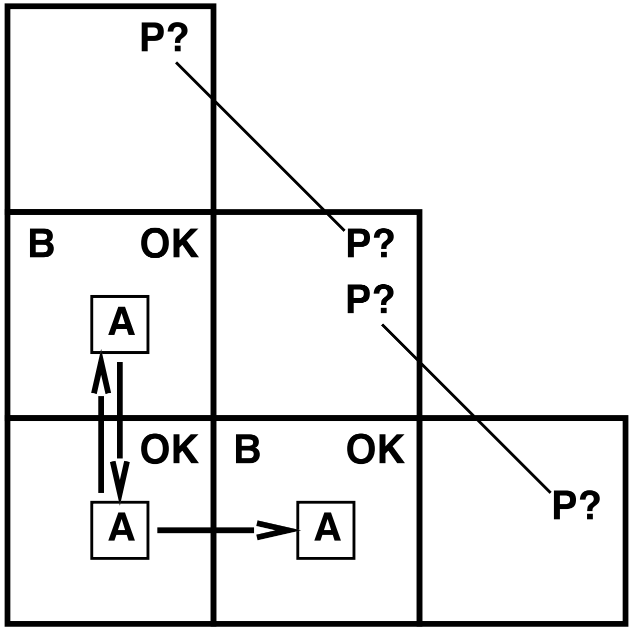
- Uncertainty
- Probability
- Syntax and Semantics
- Inference
- Conditional Independence
- Bayes' Rule

Uncertainty

In many situations, an AI agent has to choose an action based on incomplete information.

- stochastic environments (e.g. dice rolls in Backgammon)
- partial observability
  - some aspects of environment hidden from agent
  - robots can have noisy sensors, reporting quantities which differ from the “true” values

Uncertainty in the Wumpus World



In this situation no action is completely safe, because the agent does not know the location of the Pit(s).

Plannnig under Uncertainty

Let action At = leave for airport t minutes before flight

Will At get me there on time? Problems:

- partial observability, noisy sensors
- uncertainty in action outcomes (flat tyre, etc.)
- immense complexity of modelling and predicting traffic

Hence a purely logical approach either
1) risks falsehood: "*A*30 will get me there on time", or
2) leads to conclusions that are too weak for decision making:
"*A*30 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 be safe but I'd have to stay overnight in the airport ...)

Methods for handling Uncertainty

Default or nonmonotonic logic:

Assume my car does not have a flat tire, etc.
Assume *A*30 works unless contradicted by evidence
Issues: What assumptions are reasonable? How to handle contradiction?

Probability

Given the available evidence,
*A*30 will get me there on time with probability 0.04
Mahaviracarya (9th C.), Cardamo (1565) theory of gambling

Probability

Probabilistic assertions summarize effects of
Laziness: failure to enumerate exceptions, qualifications, etc.
Ignorance: lack of relevant facts, initial conditions, etc.

Subjective or Bayesian probability:
Probabilities relate propositions to one's own state of knowledge e.g.
$P(A30|\text{no reported accidents}) = 0.06$

These are not claims of a "probabilistic tendency" in the current situation (but might be learned from past experience of similar situations)

Probabilities of propositions change with new evidence: e.g. $P(A30|\text{no reported accidents, 5 a.m.}) = 0.15$

(Analogous to logical entailment status $KB \models \alpha$, not absolute truth)

Making decisions under uncertainty

Suppose I believe the following:

$$P(A30 \text{ gets me there on time}|\ldots) = 0.04$$
$$P(A90 \text{ gets me there on time}|\ldots) = 0.70$$
$$P(A120 \text{ gets me there on time}|\ldots) = 0.95$$
$$P(A1440 \text{ gets me there on time}|\ldots) = 0.9999$$

Which action to choose?

Depends on my preferences for missing flight vs. airport cuisine, etc.

Utility theory is used to represent and infer preferences

Decision theory = utility theory + probability theory

Probability basics

Begin with a set Ω – the sample space (e.g. 6 possible rolls of a die)

$\omega \in \Omega$ is a sample point/possible world/atomic event

A probability space or probability model is a sample space
with an assignment $P(\omega)$ for every $\omega \in \Omega$ s.t.

$0 \leq P(\omega) \leq 1$

$\sum_\omega P(\omega) = 1$
e.g. $P(1) = P(2) = P(3) = P(4) = P(5) = P(6) = \frac{1}{6}$.

An event A is any subset of Ω

$$P(A) = \sum_{\{\omega \in A\}} P(\omega)$$

e.g. $P(\text{die roll} < 4) = P(1) + P(2) + P(3) = \frac{1}{6} + \frac{1}{6} + \frac{1}{6} = \frac{1}{2}$

Random variables

A random variable (r.v.) is a function from sample points to some range (e.g. the Reals or Booleans)

For example, Odd(3) = true.

P induces a probability distribution for any r.v. X:

$$P(X = x_i) = \sum_{\{\omega : X(\omega) = x_i\}} P(\omega)$$

e.g., $P(\texttt{Odd} = \texttt{true}) = P(1) + P(3) + P(5) = \frac{1}{6} + \frac{1}{6} + \frac{1}{6} = \frac{1}{2}$

Propositions

Think of a proposition as the event (set of sample points) where the proposition is true

Given Boolean random variables A and B:

event a = set of sample points where $A(\omega) = \texttt{true}$

event $\neg a$ = set of sample points where $A(\omega) = \texttt{false}$

event $a \wedge b$ = points where $A(\omega) = \texttt{true}$ and $B(\omega) = \texttt{true}$

With Boolean variables, sample point = propositional logic model

e.g., $A = \texttt{true}$, $B = \texttt{false}$, or $a \wedge \neg b$.

Proposition = disjunction of atomic events in which it is true

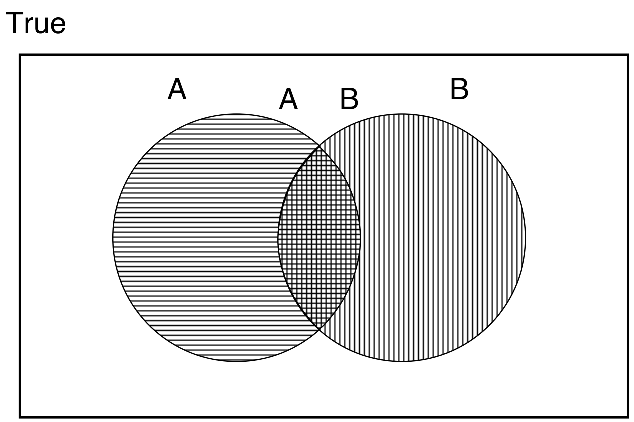
e.g., $(a \vee b) \equiv (\neg a \wedge b) \vee (a \wedge \neg b) \vee (a \wedge b)$

$\rightarrow P(a \vee b) = P(\neg a \wedge b) + P(a \wedge \neg b) + P(a \wedge b)$

Why use probability?

The definitions imply that certain logically related events must have related probabilities

For example, $P(a \vee b) = P(a) + P(b) - P(a \wedge b)$



de Finetti (1931): an agent who bets according to probabilities that violate these axioms can be forced to bet so as to lose money regardless of outcome.

Syntax for propositions

Propositional or Boolean random variables
e.g., `Cavity` (do I have a cavity?)
`Cavity = true` is a proposition, also written `Cavity`

Discrete random variables (finite or infinite)
e.g., `Weather` is one of ⟨`sunny,rain,cloudy,snow`⟩
`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.

Prior probability

Prior or unconditional probabilities of propositions
e.g. $P(\texttt{Cavity = true}) = 0.1$ and $P(\texttt{Weather = sunny}) = 0.72$
correspond to belief prior to arrival of any (new) evidence.

Probability distribution gives values for all possible assignments:
$P(\texttt{Weather}) = \langle 0.72, 0.1, 0.08, 0.1 \rangle$ (normalized, i.e., sums to 1)

Joint probability

Joint probability distribution for a set of r.v.'s gives the probability of every atomic event on those r.v's (i.e., every sample point)

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

$Weather =$	$sunny$	$rain$	$cloudy$	$snow$
$Cavity = \texttt{true}$	0.144	0.02	0.016	0.02
$Cavity = \texttt{false}$	0.576	0.08	0.064	0.08

What is the probability of being sunny?
What is the probability of having a Cavity?

Joint probability

Joint probability distribution for a set of r.v.'s gives the probability of every atomic event on those r.v's (i.e., every sample point)

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

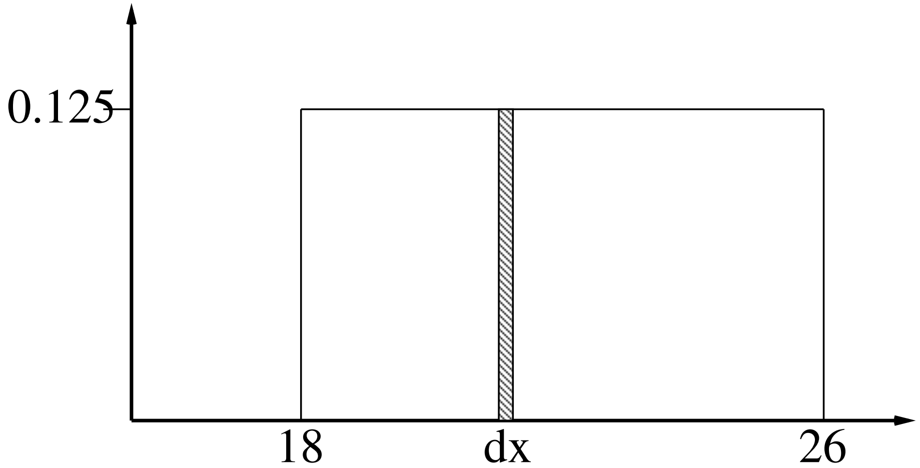
$Weather =$	$sunny$	$rain$	$cloudy$	$snow$
$Cavity = \texttt{true}$	0.144	0.02	0.016	0.02
$Cavity = \texttt{false}$	0.576	0.08	0.064	0.08

Every question about a domain can be answered by the joint distribution because every event is a sum of sample points.

Probability for continuous variables

- Express distribution as a parameterized function.

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



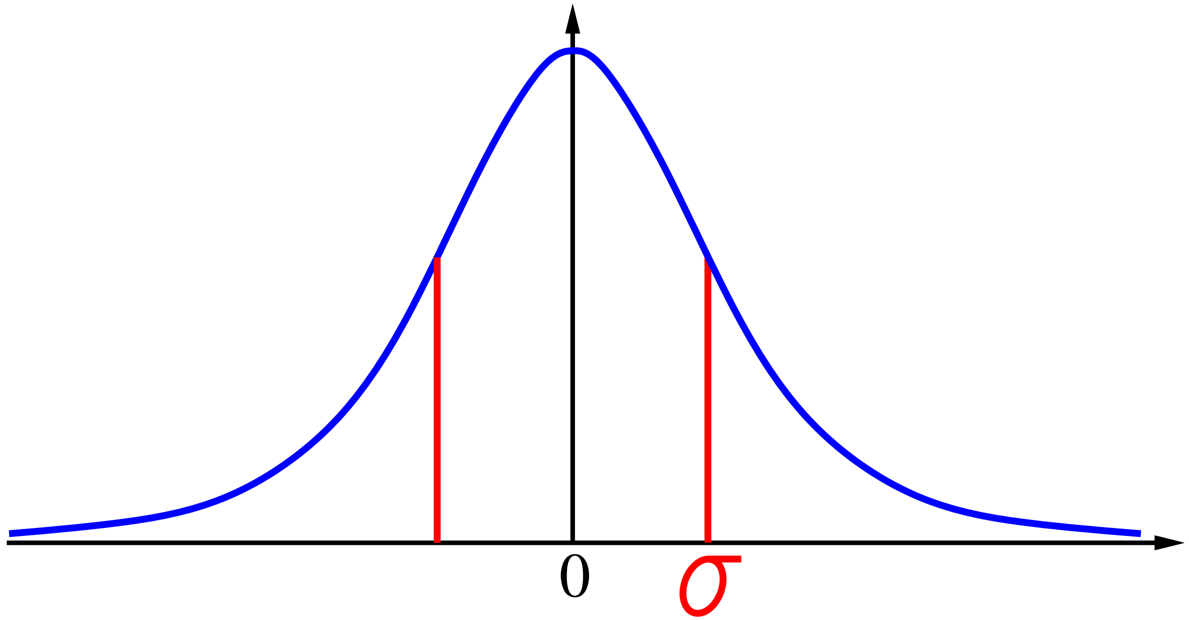
- Here P is a density; integrates to 1.

  $P(X=20.5) = 0.125$ really means

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

Gaussian density

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



Probabilistic Agents

We consider an Agent whose World Model consists not of a set of facts, but rather a set of probabilities of certain facts being true, or certain random variables taking particular values.

When the Agent makes an observation, it may update its World Model by adjusting these probabilities, based on what it has observed.

Example: Tooth Decay

Assume you live in a community where, at any given time, 20% of people have a cavity in one of their teeth which needs a filling from the dentist.

$$P(\texttt{cavity}) = 0.2$$

If you have a toothache, suddenly you will think it is much more likely that you have a cavity, perhaps as high as 60%. We say that the conditional probability of cavity, given toothache, is 0.6, written as follows:

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

If you go to the dentist, they will use a small hook-shaped instrument called a probe, and check whether this probe can catch on the back of your tooth. If it does catch, this information will increase the probability that you have a cavity.

Joint Probability Distribution

We assume there is some underlying joint probability distribution over the three random variables Toothache, Cavity and Catch, which we can write in the form of a table:

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

Note that the sum of the entries in the table is 1.0 .

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

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

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:

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

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:

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

$$P(toothache) = 0.108 + 0.012 + 0.016 + 0.064 = 0.2$$

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:

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

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

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**

Can also compute conditional probabilities:

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

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

Conditional Probability

If we consider two random variables a and b, with $P(b) \neq 0$, then the conditional probability of a given b is

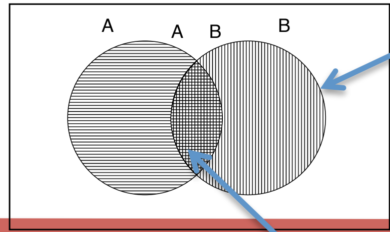
$$P(a|b) = \frac{P(a \wedge b)}{P(b)}$$

Alternative formulation: $P(a \wedge b) = P(a|b)P(b) = P(b|a)P(a)$

When an agent considers a sequence of random variables at successive time steps, they can be chained together using this formula repeatedly:

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

Conditional Probability



P(b)=true

$P(a \wedge b)$ =true

If we consider two random variables a and b, with $P(b) \neq 0$, then the conditional probability of a given b is

$$P(a|b) = \frac{P(a \wedge b)}{P(b)}$$

Alternative formulation: $P(a \wedge b) = P(a|b)P(b) = P(b|a)P(a)$

When an agent considers a sequence of random variables at successive time steps, they can be chained together using this formula repeatedly:

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

Independent Variables

Let's consider the joint probability distribution for Cavity and Weather.

$Weather =$	$sunny$	$rain$	$cloudy$	$snow$
$Cavity = \texttt{true}$	0.144	0.02	0.016	0.02
$Cavity = \texttt{false}$	0.576	0.08	0.064	0.08

Note that:

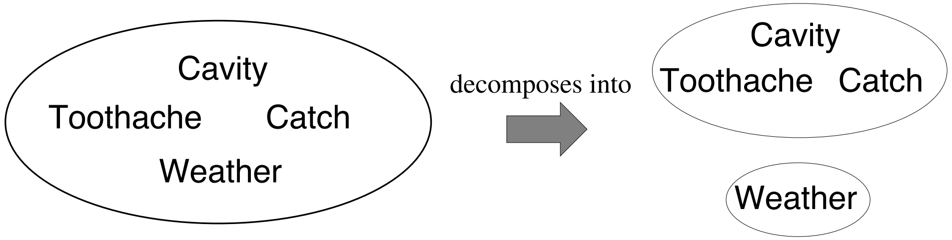
$$P(cavity\,|\,Weather = sunny) = \frac{0.144}{0.144 + 0.576} = 0.2 = P(cavity)$$

In other words, learning that the Weather is sunny has no effect on the probability of having a cavity (and the same for rain, cloudy and snow). We say that Cavity and Weather are independent variables.

Independence

A and B are independent iff

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



If variables not independent, would need 32 items in probability table.
Because Weather is independent of the other variables, only need two smaller tables, with a total of 8+4=12 items.

P(Toothache,Catch,Cavity,Weather)= *P*(Toothache,Catch,Cavity)*P*(Weather)

(Note: the number of free parameters is slightly less, because the values in each table must sum to 1).

Conditional independence

The variables Toothache, Cavity and Catch are not independent. But, they do exhibit conditional independence.

If you have a cavity, the probability that the probe will catch is 0.9, no matter whether you have a toothache or not.

If you don't have a cavity, the probability that the probe will catch is 0.2, regardless of whether you have a toothache. In other words,

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

We say that Catch is conditionally independent of Toothache given Cavity.

Conditional independence

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

The variables Toothache, Cavity and Catch are not independent. But, they do exhibit conditional independence.

If you have a cavity, the probability that the probe will catch is 0.9, no matter whether you have a toothache or not.

If you don't have a cavity, the probability that the probe will catch is 0.2, regardless of whether you have a toothache. In other words,

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

We say that Catch is conditionally independent of Toothache given Cavity.

Conditional independence

This conditional independence reduces the number of free parameters from 7 down to 5.

For larger problems with many variables, deducing this kind of conditional independence among the variables can reduce the number of free parameters substantially, and allow the Agent to maintain a simpler World Model.

Equivalent statements:

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

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

Bayes' Rule

The formula for conditional probability can be manipulated to find a relationship when the two variables are swapped:

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

$$\rightarrow \text{Bayes' rule } P(a \mid b) = \frac{P(b \mid a)P(a)}{P(b)}$$

This is often useful for assessing the probability of an underlying cause after an effect has been observed:

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

Bayes' Rule and Conditional Independence

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

This is an example of a naive Bayes model:

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

- Total number of parameters is linear in n

Wumpus World

1,4	2,4	3,4	4,4
1,3	2,3	3,3	4,3
1,2 **B** **OK**	2,2	3,2	4,2
1,1 **OK**	2,1 **B** **OK**	3,1	4,1

What is the probability of a Pit in (1,3) ? What about (2,2) ?

To answer this, we need a "prior" assumption about the placement of Pits. We will assume a 20% chance of a Pit in each square at the beginning of the game (independent of what Pits are in the other squares).

Specifying the Probability Model

We will use Bi, j to indicate a Breeze in square (i, j), and Piti, j to indicate a Pit in square (i, j).

We use `known` to represent what we know, i.e.

$$B_{1,2} \wedge B_{2,1} \wedge \neg B_{1,1} \wedge \neg \text{Pit}_{1,2} \wedge \neg \text{Pit}_{2,1} \wedge \neg \text{Pit}_{1,1}$$

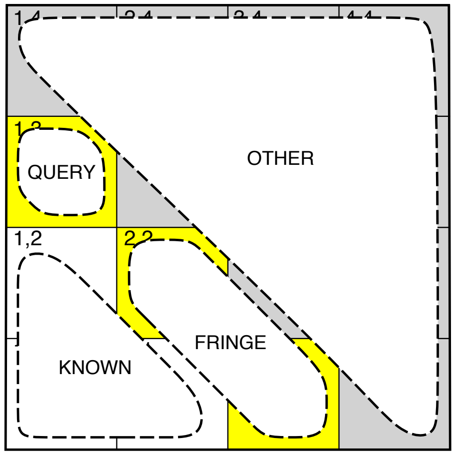
We use `Unknown` to represent the joint probability of Pits in all the other squares, i.e.

$$P(\texttt{Unknown}) = P(\text{Pit}_{1,4}, \ldots, \text{Pit}_{4,1})$$

We divide `Unknown into Fringe and Other`, where

$$P(\texttt{Fringe}) = P(\text{Pit}_{1,3}, \text{Pit}_{2,2}, \text{Pit}_{3,1})$$

and Other is all the other variables.



$$P(\text{Pit}_{1,3}|\,\texttt{known}) = \sum_{\texttt{unknown}} P(\text{Pit}_{1,3}, \texttt{unknown}\,|\,\texttt{known})$$

$$= \sum_{\texttt{fringe}} \sum_{\texttt{other}} P(\text{Pit}_{1,3}, \texttt{fringe}, \texttt{other}\,|\,\texttt{known})$$

$$= \sum_{\texttt{fringe}} \sum_{\texttt{other}} P(\text{Pit}_{1,3}|\,\texttt{fringe}, \texttt{other}, \texttt{known})\, P(\texttt{fringe}, \texttt{other}\,|\,\texttt{known})$$

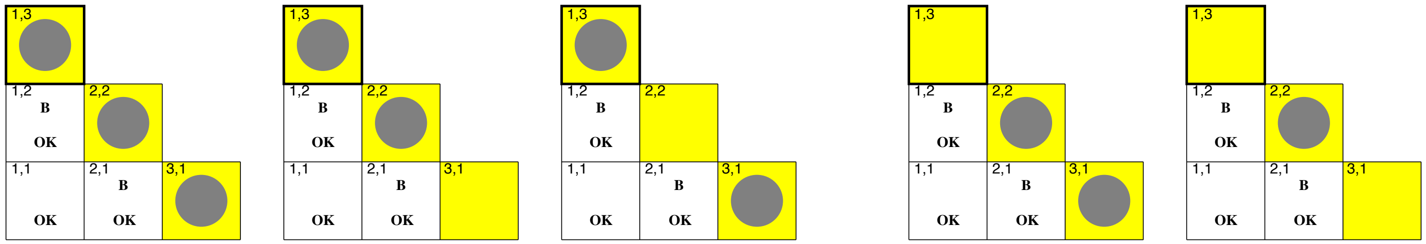
$$= \sum_{\texttt{fringe}} P(\text{Pit}_{1,3}|\,\texttt{fringe}) \sum_{\texttt{other}} P(\texttt{fringe}, \texttt{other}\,|\,\texttt{known})$$

$$= \sum_{\texttt{fringe}} P(\text{Pit}_{1,3}|\,\texttt{fringe}) \sum_{\texttt{other}} \frac{P(\texttt{known}\,|\,\texttt{fringe}, \texttt{other}) P(\texttt{fringe}, \texttt{other})}{P(\texttt{known})}$$

Note: have used the fact that $P1,3$ is independent of other, given fringe.

Fringe Models

Let's denote by F the set of fringe models compatible with the known facts:



0.2 x 0.2 = 0.04 0.2 x 0.8 = 0.16 0.8 x 0.2 = 0.16 0.2 x 0.2 = 0.04 0.2 x 0.8 = 0.16

$P(\texttt{known}\,|\,\texttt{fringe}, \texttt{other}) = 0$ outside F, so $P(\text{Pit}_{1,3}|\,\texttt{known})$ reduces to:

$$\frac{\sum_{\texttt{fringe} \in F} P(\text{Pit}_{1,3}|\,\texttt{fringe}) \sum_{\texttt{other}} P(\texttt{known}\,|\,\texttt{fringe}, \texttt{other}) P(\texttt{fringe}, \texttt{other})}{P(\texttt{known})}$$

Note also that

$$P(\texttt{known}) = \sum_{\texttt{fringe} \in F} \sum_{\texttt{other}} P(\texttt{known}\,|\,\texttt{fringe}, \texttt{other}) P(\texttt{fringe}, \texttt{other})$$

Using the Prior

- Because of the prior, `other` and `fringe` become independent, and known becomes independent of `other`, given fringe.

$$P(\texttt{known} \mid \texttt{fringe}, \texttt{other}) = P(\texttt{known} \mid \texttt{fringe}) = 1, \text{ for } \texttt{fringe} \in F, \text{so}$$

$$P(\texttt{known}) = \sum_{\texttt{fringe} \in F} P(\texttt{fringe}) = (0.2)^3 + 3 \times (0.2)^2(0.8) + (0.2)(0.8)^2$$

$$= 0.008 + 0.032 + 0.032 + 0.032 + 0.128 = 0.232$$

The numerator includes only those models for which $\text{Pit}_{1,3}$ is true, i.e.

$$P(\text{Pit}_{1,3} \mid \texttt{known}) = \frac{0.008 + 0.032 + 0.032}{0.232} = \frac{9}{29} \simeq 0.310$$

In a similar way,

$$P(\text{Pit}_{2,2} \mid \texttt{known}) = \frac{0.008 + 0.032 + 0.032 + 0.128}{0.232} = \frac{25}{29} \simeq 0.862$$

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