

Uncertainty

AIMA2e Chapter 13

1

Outline

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

2

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 (KCBS traffic reports)
- 3) uncertainty in action outcomes (flat tire, etc.)
- 4) immense complexity of modelling and predicting traffic

Hence 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 . . .)

Methods for handling uncertainty

Default or nonmonotonic 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:

$A_{25} \mapsto_{0.3}$ get there on time

$Sprinkler \mapsto_{0.99} WetGrass$

$WetGrass \mapsto_{0.7} Rain$

Issues: Problems with combination, e.g., *Sprinkler* causes *Rain*??

Probability

Given the available evidence, A_{25} will get me there on time with probability 0.04

Mahaviracarya (9th C.), Cardano (1565) theory of gambling

(Fuzzy logic handles *degree of truth* NOT uncertainty e.g.,

WetGrass is true to degree 0.2)

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

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

These are *not* claims of some **probabilistic tendency** in the current situation

(but might be learned from past experience of similar situations)

Probabilities of propositions change with new evidence:

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

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

Making decisions under uncertainty

Suppose I believe the following:

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

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

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

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

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., $P(1) = P(2) = P(3) = P(4) = P(5) = P(6) = 1/6$.

An *event* A is any subset of Ω

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

E.g., $P(\text{die roll} < 4) = 1/6 + 1/6 + 1/6 = 1/2$

Random variables

A *random variable* is a function from sample points to some range, e.g., the reals or Booleans

e.g., $\text{Odd}(1) = \text{true}$.

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

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

e.g., $P(\text{Odd} = \text{true}) = 1/6 + 1/6 + 1/6 = 1/2$

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 $A(\omega) = true$

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

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

Often in AI applications, the sample points are *defined* by the values of a set of random variables, i.e., the sample space is the Cartesian product of the ranges of the variables

With Boolean variables, sample point = propositional logic model

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

Proposition = disjunction of atomic events in which it is true

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

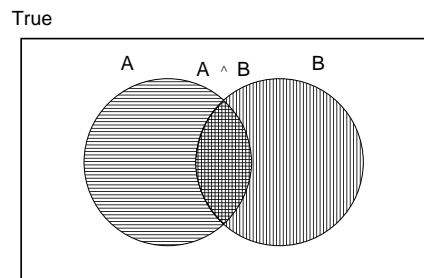
$\Rightarrow P(a \vee b) = P(\neg a \wedge b) + P(a \wedge \neg b) + P(a \wedge b)$

9

Why use probability?

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

E.g., $P(a \vee b) = P(a) + P(b) - P(a \wedge b)$



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.

10

Syntax for propositions

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 $\langle \textit{sunny}, \textit{rain}, \textit{cloudy}, \textit{snow} \rangle$

$\textit{Weather} = \textit{rain}$ is a proposition

Values must be exhaustive and mutually exclusive

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

e.g., $\textit{Temp} = 21.6$; also allow, e.g., $\textit{Temp} < 22.0$.

Arbitrary Boolean combinations of basic propositions

11

Prior probability

Prior or **unconditional probabilities** of propositions

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

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

Probability distribution gives values for all possible assignments:

$\mathbf{P}(\textit{Weather}) = \langle 0.72, 0.1, 0.08, 0.1 \rangle$ (*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 (i.e., every sample point)

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

<i>Weather</i> =	<i>sunny</i>	<i>rain</i>	<i>cloudy</i>	<i>snow</i>
<i>Cavity</i> = <i>true</i>	0.144	0.02	0.016	0.02
<i>Cavity</i> = <i>false</i>	0.576	0.08	0.064	0.08

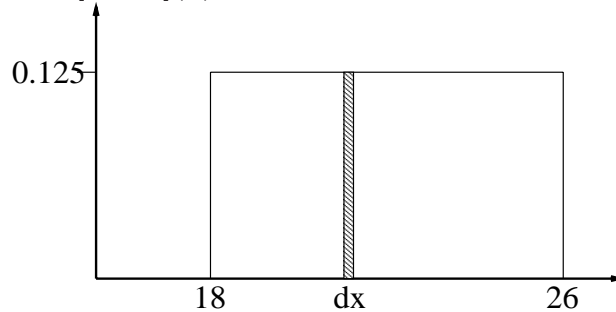
Every question about a domain can be answered by the joint distribution because every event is a sum of sample points

12

Probability for continuous variables

Express distribution as a parameterized function of value:

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



Here P is a *density*; integrates to 1.

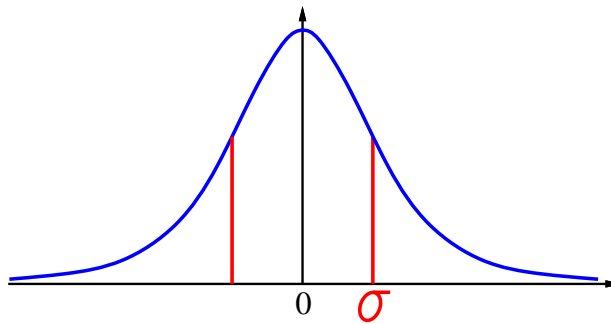
$P(X = 20.5) = 0.125$ really means

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

13

Gaussian density

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



14

Conditional probability

Conditional or posterior probabilities

e.g., $P(\text{cavity}|\text{toothache}) = 0.8$

i.e., given that *toothache* is all I know

NOT “if *toothache* then 80% chance of *cavity*”

(Notation for conditional distributions:

$\mathbf{P}(\text{Cavity}|\text{Toothache}) = 2\text{-element vector of } 2\text{-element vectors}$)

If we know more, e.g., *cavity* is also given, then we have

$P(\text{cavity}|\text{toothache}, \text{cavity}) = 1$

Note: the less specific belief *remains valid* after more evidence arrives, but is not always *useful*

New evidence may be irrelevant, allowing simplification, e.g.,

$P(\text{cavity}|\text{toothache}, 49ersWin) = P(\text{cavity}|\text{toothache}) = 0.8$

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

15

Conditional probability

Definition of conditional probability:

$$P(a|b) = \frac{P(a \wedge b)}{P(b)} \text{ if } P(b) \neq 0$$

Product rule gives an alternative formulation:

$$P(a \wedge b) = P(a|b)P(b) = p(b|a)P(a)$$

A general version holds for whole distributions, e.g.,

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

(View as a 4×2 set of equations, *not* matrix mult.)

Chain rule is derived by successive application of product rule:

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

16

Inference by enumeration

Start with the joint distribution:

	<i>toothache</i>		\neg <i>toothache</i>	
	<i>catch</i>	\neg <i>catch</i>	<i>catch</i>	\neg <i>catch</i>
<i>cavity</i>	.108	.012	.072	.008
\neg <i>cavity</i>	.016	.064	.144	.576

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

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

17

Inference by enumeration

Start with the joint distribution:

	<i>toothache</i>		\neg <i>toothache</i>	
	<i>catch</i>	\neg <i>catch</i>	<i>catch</i>	\neg <i>catch</i>
<i>cavity</i>	.108	.012	.072	.008
\neg <i>cavity</i>	.016	.064	.144	.576

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

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

$$P(\textit{toothache}) = 0.108 + 0.012 + 0.016 + 0.064 = 0.2$$

18

Inference by enumeration

Start with the joint distribution:

	<i>toothache</i>		\neg <i>toothache</i>	
	<i>catch</i>	\neg <i>catch</i>	<i>catch</i>	\neg <i>catch</i>
<i>cavity</i>	.108	.012	.072	.008
\neg <i>cavity</i>	.016	.064	.144	.576

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

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

$$P(\text{cavity} \vee \text{toothache}) =$$

$$0.108 + 0.012 + 0.072 + 0.008 + 0.016 + 0.064 = 0.28$$

19

Inference by enumeration

Start with the joint distribution:

	<i>toothache</i>		\neg <i>toothache</i>	
	<i>catch</i>	\neg <i>catch</i>	<i>catch</i>	\neg <i>catch</i>
<i>cavity</i>	.108	.012	.072	.008
\neg <i>cavity</i>	.016	.064	.144	.576

Can also compute conditional probabilities:

$$\begin{aligned}
 P(\neg \text{cavity} | \text{toothache}) &= \frac{P(\neg \text{cavity} \wedge \text{toothache})}{P(\text{toothache})} \\
 &= \frac{0.016 + 0.064}{0.108 + 0.012 + 0.016 + 0.064} = 0.4
 \end{aligned}$$

20

Normalization

	<i>toothache</i>		\neg <i>toothache</i>	
	<i>catch</i>	\neg <i>catch</i>	<i>catch</i>	\neg <i>catch</i>
<i>cavity</i>	.108	.012	.072	.008
\neg <i>cavity</i>	.016	.064	.144	.576

Denominator can be viewed as a *normalization constant* α

$$\begin{aligned}
 \mathbf{P}(\text{Cavity}|\text{toothache}) &= \alpha \mathbf{P}(\text{Cavity}, \text{toothache}) \\
 &= \alpha [\mathbf{P}(\text{Cavity}, \text{toothache}, \text{catch}) + \mathbf{P}(\text{Cavity}, \text{toothache}, \neg \text{catch})] \\
 &= \alpha [\langle 0.108, 0.016 \rangle + \langle 0.012, 0.064 \rangle] \\
 &= \alpha \langle 0.12, 0.08 \rangle = \langle 0.6, 0.4 \rangle
 \end{aligned}$$

General idea: compute distribution on query variable
by fixing **evidence variables** and summing over **hidden variables**

21

Inference by enumeration, contd.

Typically, we are interested in
the posterior joint distribution of the **query variables** \mathbf{Y}
given specific values \mathbf{e} for the **evidence variables** \mathbf{E}

Let the **hidden variables** be $\mathbf{H} = \mathbf{X} - \mathbf{Y} - \mathbf{E}$

Then the required summation of joint entries is done by summing out the
hidden variables:

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

The terms in the summation are joint entries because \mathbf{Y} , \mathbf{E} , and \mathbf{H} together
exhaust the set of random variables

Obvious problems:

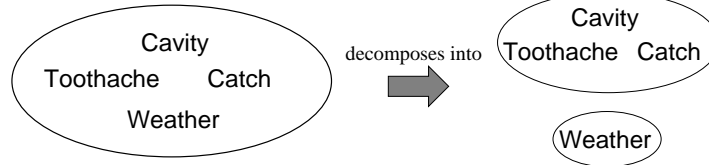
- 1) Worst-case time complexity $O(d^n)$ where d is the largest arity
- 2) Space complexity $O(d^n)$ to store the joint distribution
- 3) How to find the numbers for $O(d^n)$ entries???

22

Independence

A and B are independent iff

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



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

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

Absolute independence powerful but rare

Dentistry is a large field with hundreds of variables,
none of which are independent. What to do?

23

Conditional independence

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

$$(1) \mathbf{P}(\textit{catch}|\textit{toothache}, \textit{cavity}) = \mathbf{P}(\textit{catch}|\textit{cavity})$$

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

$$(2) \mathbf{P}(\textit{catch}|\textit{toothache}, \neg\textit{cavity}) = \mathbf{P}(\textit{catch}|\neg\textit{cavity})$$

\textit{Catch} is **conditionally independent** of $\textit{Toothache}$ given \textit{Cavity} :

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

Equivalent statements:

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

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

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

24

Conditional independence contd.

Write out full joint distribution using chain rule:

$$\begin{aligned} & \mathbf{P}(\textit{Toothache}, \textit{Catch}, \textit{Cavity}) \\ &= \mathbf{P}(\textit{Toothache}|\textit{Catch}, \textit{Cavity})\mathbf{P}(\textit{Catch}, \textit{Cavity}) \\ &= \mathbf{P}(\textit{Toothache}|\textit{Catch}, \textit{Cavity})\mathbf{P}(\textit{Catch}|\textit{Cavity})\mathbf{P}(\textit{Cavity}) \\ &= \mathbf{P}(\textit{Toothache}|\textit{Cavity})\mathbf{P}(\textit{Catch}|\textit{Cavity})\mathbf{P}(\textit{Cavity}) \end{aligned}$$

I.e., $2 + 2 + 1 = 5$ independent numbers (equations 1 and 2 remove 2)

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.

25

Bayes' Rule

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

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

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:

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

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!

26

Bayes' Rule and conditional 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:

$$\mathbf{P}(Cause, Effect_1, \dots, Effect_n) = \mathbf{P}(Cause)\prod_i \mathbf{P}(Effect_i|Cause)$$



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

$P_{ij} = true$ 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 $\mathbf{P}(P_{1,1}, \dots, P_{4,4}, B_{1,1}, B_{1,2}, B_{2,1})$

Apply product rule:

$$\mathbf{P}(B_{1,1}, B_{1,2}, B_{2,1} | P_{1,1}, \dots, P_{4,4}) \mathbf{P}(P_{1,1}, \dots, P_{4,4})$$

(Do it this way to get $P(\text{Effect}|\text{Cause})$.)

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

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

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

for n pits.

29

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

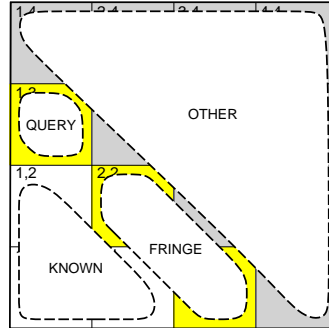
$$\mathbf{P}(P_{1,3} | known, b) = \alpha \sum_{unknown} \mathbf{P}(P_{1,3}, unknown, known, b)$$

Grows exponentially with number of squares!

30

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!

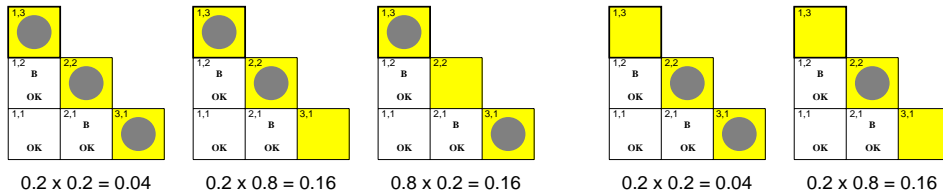
31

Using conditional independence contd.

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

32

Using conditional independence contd.



$$\mathbf{P}(P_{1,3}|known, b) = \alpha' < 0.2(0.04 + 0.16 + 0.16), 0.8(0.04 + 0.16) >$$

$$\approx < 0.31, 0.69 >$$

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

33

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

34