Uncertainty

Contents

- Basic Probability Notation
  - Language of Propositions in Probability Assertions
- Inference Using Full Joint Distributions
- Independence
- Bayes' Rule and Its Use
- Wumpus World

Basic Probability Notation

- 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.

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

  $$\sum_{\omega} P(\omega) = 1$$

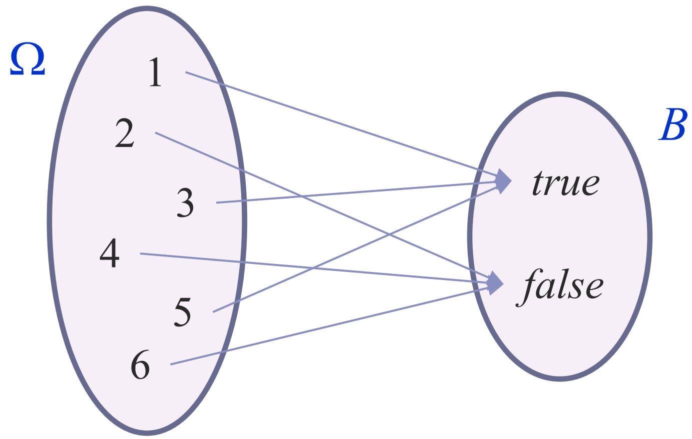
  e.g., $P(1) = P(2) = P(3) = P(4) = P(5) = P(6) = 1/6$
- An event A is any subset of Ω (i.e., a set of sample points or atomic events)

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

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

Language of Propositions in Probability Assertions

- A random variable is a function from sample points to some range e.g., the real numbers or Booleans
  - e.g., $Odd(1) = true$



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

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

e.g., $P(Odd = true) = P(1) + P(3) + P(5) = 1/6 + 1/6 + 1/6 = 1/2$

Language of Propositions in Probability Assertions

- Boolean random variables
  - e.g., $Cavity$ (do I have a cavity?)
  - $Cavity = true$ ($false$) is a proposition, also written $cavity$ ($\neg cavity$)
- Discrete random variables (finite or infinite)
  - e.g., $Weather$ is one of $\langle sunny, rain, cloudy, snow \rangle$
  - $Weather = rain$ is a proposition
  - Values must be exhaustive and mutually exclusive
- Continuous random variables (bounded or unbounded)
  - e.g., $Temp = 21.6$, also allow, e.g., $Temp < 22.0$
- A proposition is an arbitrary Boolean combination of basic propositions
  - e.g., $Weather = rain \wedge Temp < 22.0$

Language of Propositions in Probability Assertions

- Prior or unconditional probabilities of propositions

  e.g., $P(Cavity = true) = 0.2$ and $P(Weather = sunny) = 0.72$

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

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

- Joint probability distribution for a set of r.v.s gives the probability of every atomic event on those r.v.s (i.e., every sample point)
  - $\mathbf{P}(Weather, Cavity)$ is a 4×2 matrix of values:

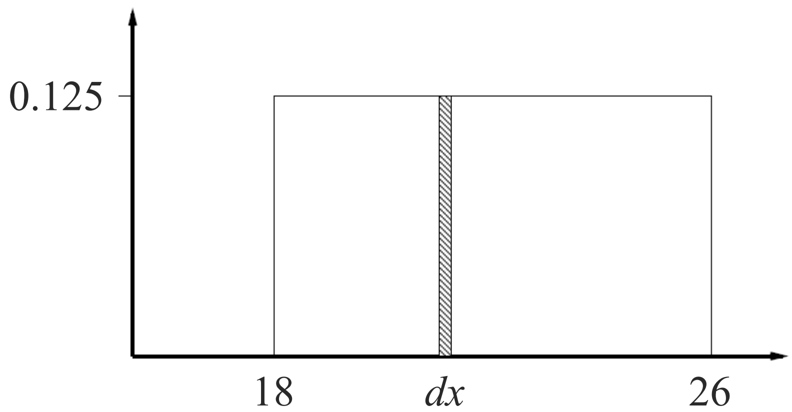
$Weather =$	$sunny$	$rain$	$cloudy$	$snow$
$Cavity = true$	0.144	0.02	0.016	0.02
$Cavity = false$	0.576	0.08	0.064	0.08

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

Language of Propositions in Probability Assertions

- 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

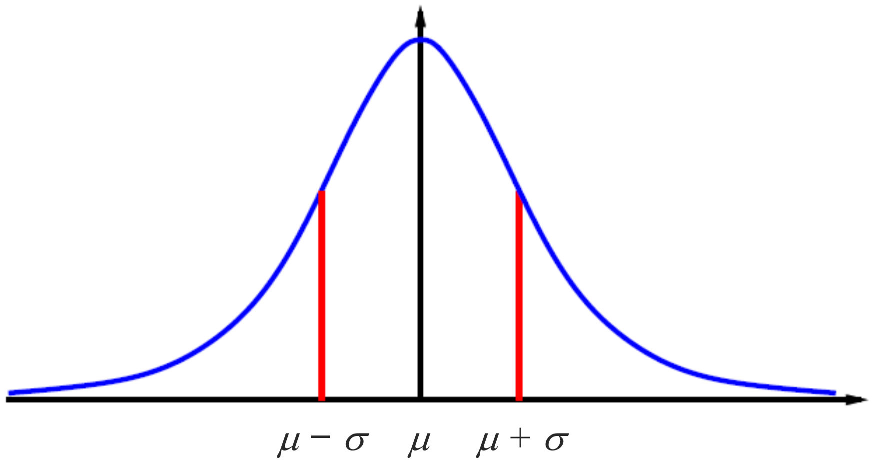


- Here P is a density; integrates to 1
  - $P(X = 20.5) = 0.125$ really means

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

Language of Propositions in Probability Assertions

- Gaussian density



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

Language of Propositions in Probability Assertions

- Conditional or posterior probabilities, e.g., $P(cavity \mid toothache) = 0.8$, i.e., given that toothache is all I know
  - NOT "**if** toothache **then** 80% chance of cavity"
- Notation for conditional distributions:
  - $\mathbf{P}(Cavity \mid Toothache)$ is a 2-element vector of 2-element vectors
- If we know more, e.g., $cavity$ is also given, then we have

  $$P(cavity \mid toothache, cavity) = 1$$

  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.,

  $$P(cavity \mid toothache, 49ersWin) = P(cavity \mid toothache) = 0.8$$

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

Language of Propositions in Probability Assertions

- Definition of conditional probability:

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

- Product rule gives an alternative formulation:

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

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

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

  - View as a 4×2 set of equations, not matrix multiplication

- Chain rule is derived by successive application of product rule:

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

$$\mathbf{P}(X_1, X_2, X_3, X_4) = \mathbf{P}(X_1)\, \mathbf{P}(X_2 \mid X_1)\, \mathbf{P}(X_3 \mid X_1, X_2)\, \mathbf{P}(X_4 \mid X_1, X_2, X_3)$$

Inference Using Full Joint Distributions

- 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**

Cavity	*Toothache*	*Catch*	
T	T	T	.108
T	T	F	.102
T	F	T	.072
T	F	F	.008
F	T	T	.016
F	T	F	.064
F	F	T	.144
F	F	F	.576

  - Joint probability distribution specifies probability of every atomic event

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

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

Inference Using Full Joint Distributions

- Start with the joint distribution

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

  - Joint probability distribution species probability of every atomic event

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

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

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

$\Updownarrow$

$Toothache = true$

Inference Using Full Joint Distributions

- Start with the joint distribution

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

  - Joint probability distribution species probability of every atomic event

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

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

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

$$= 0.28$$

Inference Using Full Joint Distributions

- Start with the joint distribution

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

- Can also compute conditional probabilities:

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

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

- $P(toothache)$ is not necessary if we consider the ratio

$P(cavity, toothache) : P(\neg cavity, toothache)$

Inference Using Full Joint Distributions

- Normalization

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

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

- Denominator can be viewed as a normalization constant α

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

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

Inference Using Full Joint Distributions

- General marginalization (summing out) rule for any sets of variables $\mathbf{Y}$ and $\mathbf{Z}$:

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

  - A distribution over $\mathbf{Y}$ can be obtained by summing out the probabilities for each possible value of the other variables

$$\begin{aligned} P(toothache) &= P(toothache, cavity) + P(toothache, \neg cavity) \\ &= P(toothache, cavity, catch) + P(toothache, cavity, \neg catch) \\ &\quad + P(toothache, \neg cavity, catch) + P(toothache, \neg cavity, \neg catch) \end{aligned}$$

- Conditioning rule:

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

Inference Using Full Joint Distributions

- Let $\mathbf{X}$ be all the variables. Typically, we want

  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:

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

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

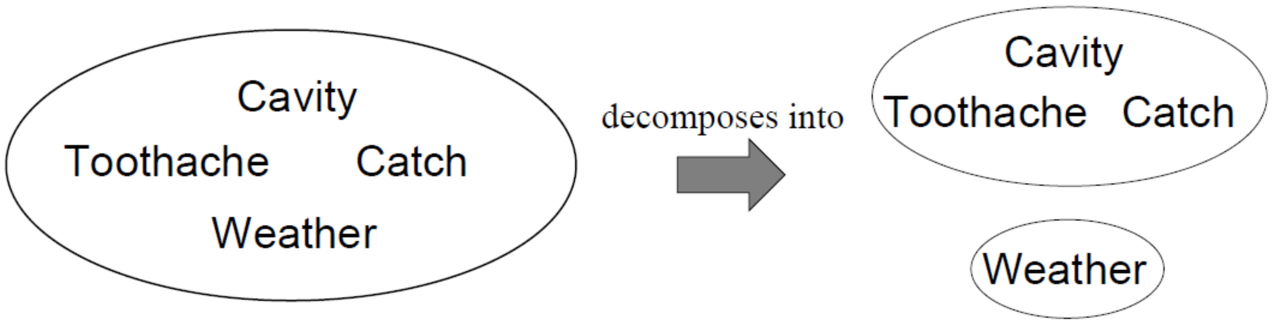
Inference Using Full Joint Distributions

- 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???

Independence

- A and B are independent if and only if

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



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

- 32 entries reduced to 12; for n independent biased coins, $2^n \to n$
- Absolute independence powerful but rare
- Dentistry is a large field with hundreds of variables, none of which are independent. What to do?

Independence

- $\mathbf{P}(Toothache, Cavity, Catch)$ has $2^3 - 1 = 7$ independent entries
- If I have a cavity, the probability that the probe catches in it doesn't depend on whether I have a toothache:

(1) $P(catch \mid toothache, cavity) = P(catch \mid cavity)$

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

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

- $Catch$ is conditionally independent of $Toothache$ given $Cavity$:

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

- Equivalent statements:

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

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

Independence

- Write out full joint distribution using chain rule:

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

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

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

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

  - 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

Bayes' Rule and Its Use

- Product rule $P(a \wedge b) = P(a \mid b)\, P(b) = P(b \mid a)\, P(a)$
  - Bayes' rule: $P(a \mid b) = \dfrac{P(b \mid a) P(a)}{P(b)}$

  or in distribution form

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

- Useful for assessing diagnostic probability from causal probability:

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

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

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

  - Note: posterior probability of meningitis still very small!

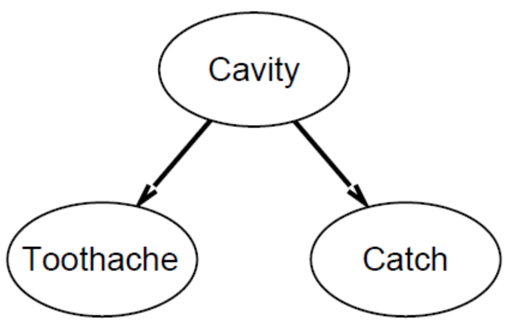
Bayes' Rule and Its Use

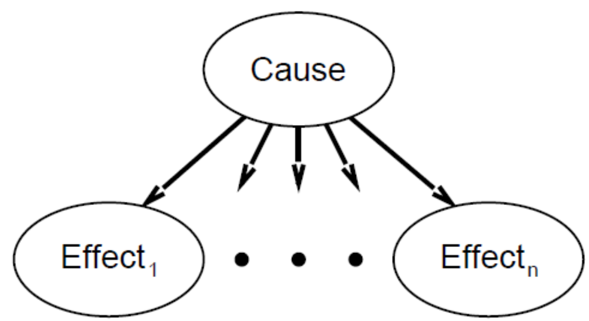
- Bayes' rule and conditional independence

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

This is an example of a Naïve Bayes model:

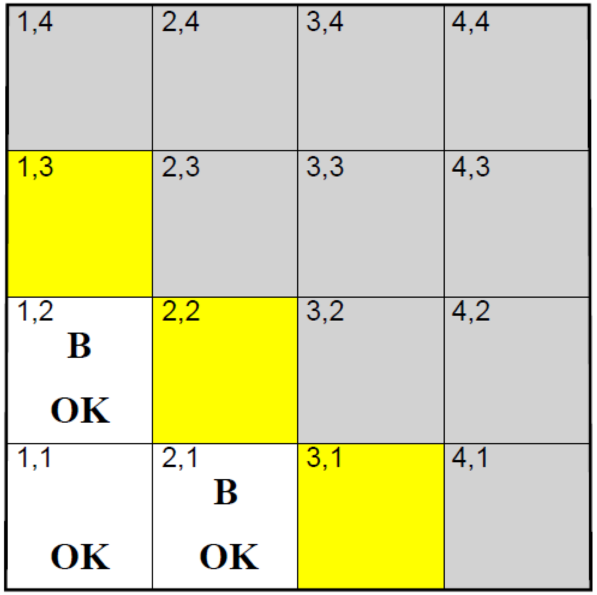
$$\mathbf{P}(Cause, Effect_1, \ldots, Effect_n) = \mathbf{P}(Cause) \prod_i \mathbf{P}(Effect_i \mid Cause)$$





- Total number of parameters is linear in n

Wumpus World

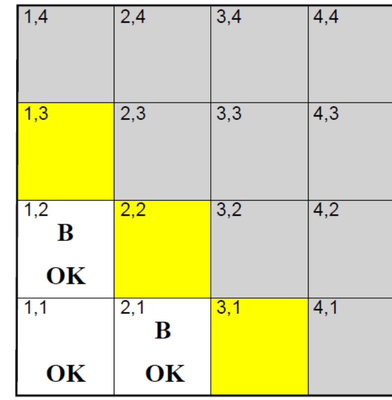


- $P_{i,j} = true$ iff $[i, j]$ contains a pit (probability of a pit is 0.2)
- $B_{i,j} = true$ iff $[i, j]$ is breezy (breezy if an adjacent square has a pit)

Will the yellow squares be O.K.?

Wumpus World

- 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: $\mathbf{P}(P_{1,3} \mid b, known)$
- Define $Unknown = P_{i,j}$s other than $P_{1,3}$ and $known$
- For inference by enumeration, we have

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

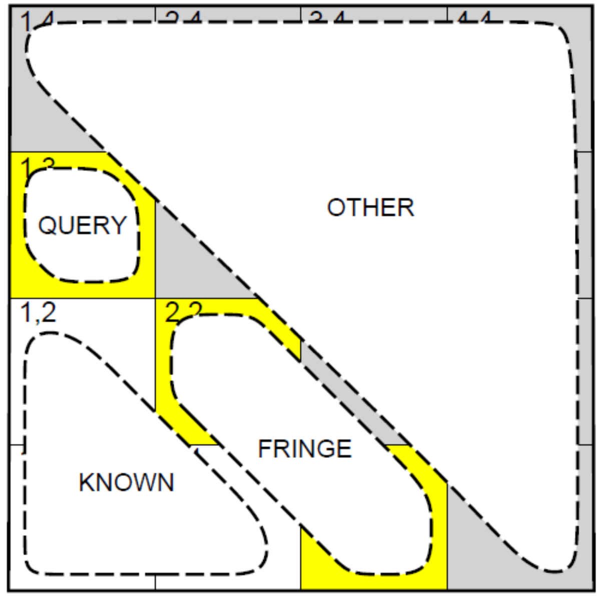
$$= \alpha \sum_{unknown} \underbrace{\mathbf{P}(b \mid P_{1,3}, known, unknown)}_{\text{0 or 1}} \mathbf{P}(P_{1,3}, known, unknown)$$

(Do it this way to get $\mathbf{P}(Effect \mid Cause)$)

But, grows exponentially with number of squares!

Wumpus World

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



- Define $Unknown = Fringe \cup Other$

$$\mathbf{P}(b \mid P_{1,3}, Known, Unknown) = \mathbf{P}(b \mid P_{1,3}, Known, Fringe)$$

  - Manipulate query into a form where we can use this!

Wumpus World

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

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

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

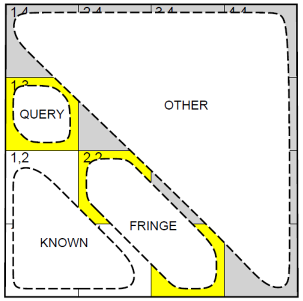
$$= \alpha \sum_{fringe} \mathbf{P}(b \mid P_{1,3}, known, fringe) \sum_{other} \mathbf{P}(P_{1,3}, known, fringe, other)$$

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

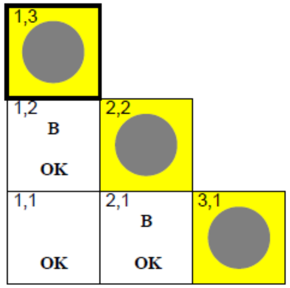
$$= \alpha \sum_{fringe} \mathbf{P}(b \mid P_{1,3}, known, fringe)\mathbf{P}(P_{1,3})P(known)P(fringe)$$

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

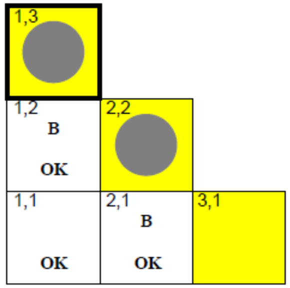
$$= \alpha' \mathbf{P}(P_{1,3}) \sum_{fringe} \mathbf{P}(b \mid P_{1,3}, known, fringe)P(fringe)$$



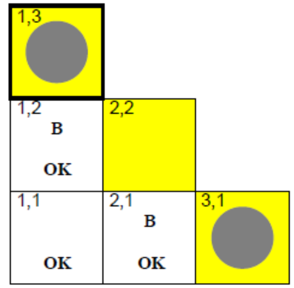
Wumpus World



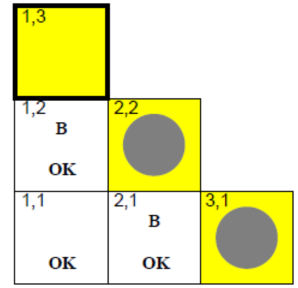
$0.2 \times 0.2 = 0.04$



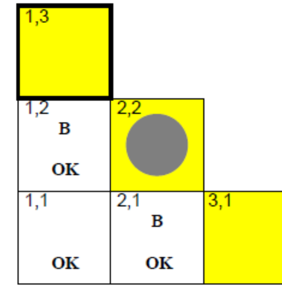
$0.2 \times 0.8 = 0.16$



$0.8 \times 0.2 = 0.16$



$0.2 \times 0.2 = 0.04$



$0.2 \times 0.8 = 0.16$

$$\begin{aligned}
\mathbf{P}(P_{1,3} \mid b, known) &= \alpha' \, \mathbf{P}(P_{1,3}) \sum_{fringe} \mathbf{P}(b \mid P_{1,3}, known, fringe) P(fringe) \\
&= \alpha' \langle 0.2, 0.8 \rangle \langle 0.04 + 0.16 + 0.16 + 0, 0.04 + 0.16 + 0 + 0 \rangle \\
&= \alpha' \langle 0.2(0.04 + 0.16 + 0.16), 0.8(0.04 + 0.16) \rangle \\
&\approx \langle 0.31, 0.69 \rangle
\end{aligned}$$

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