Reasoning with Uncertainty

Chapter 13

Outline

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

The real world is an uncertain place...

Example: I need a plan that will get me to airport on time

- 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 (ADOT/Google traffic reports and estimates)
  3. uncertainty in action outcomes (flat tire, detours, etc.)
  4. immense complexity of modeling and predicting traffic

- Hence a purely logical approach either:
  - Risks falsehood:
    - "Plan A_{90} leaves home 90 minutes early and airport is only 5 minutes away; A_{90} will get me there on time"
    - Does not take into account ***any*** uncertainties → is not realistic
  - or 2) leads to conclusions that are too weak for decision making:
    - "Plan A_{90} 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. etc."
    - Takes into account ***many*** (infinite?) uncertainties...none of which can be proven → no actionable plan.
  - Is irrationally cautious:
    - Plan A_{1440} leaves 24 hours early; might reasonably be said to get me there on time but I'd have to stay overnight in the airport . . .)

Dealing with Uncertainty

So what can we do? Need tools do we have to deal with this?

Belief States?

- Idea: generate and track all possible states of the world given uncertainty
  - Used for Problem-solving Agents (ch4) and Logical Agents (ch7)
  - Make a contingency plan that is guaranteed successful for all eventualities

- Nice idea, but not very realistic for complex, variable worlds:
  - For partially observable world, must consider every possible explanation for incoming sensor percepts...*no matter how unlikely*. → Huge belief states
  - A plan to handle *every* contingency gets arbitrarily large in a real world with essentially *infinite* contingencies.
  - Sometimes there is no plan that is guaranteed to achieve the goal...and yet we must act...rationally.

- Conclusion: We need some new tools!
  - Reasoning *rationally* under uncertainty. Takes into account:
    - *Relative importance* of various goals (performance measures of agent)
    - The *likelihood* of: contingencies, action success/failure, etc.

Dealing with Uncertainty

So how about building uncertainty into logical reasoning?

- Example: diagnosing a toothache
  - Diagnosis: classic example of a problem with inherent uncertainty
  - Attempt 1: Toothache ⇒ HasCavity
    - But: not all toothaches are caused by cavities. Not true!
  - Attempt 2: Toothache ⇒ Cavity ∨ GumDisease ∨ Abscess ∨ etc ∨ etc
    - To be true: would need nearly unlimited list of options...some unknown.
  - Attempt 3: Try make causal: Cavity ⇒ Toothache
    - Nope: not all cavities cause toothaches!

- Fundamental problems with using logic in uncertain domains:
    - Laziness: It's too much work to generate complete list of antecedents/consequents to cover all possibilities
    - Ignorance: You may not even *know* all of the possibilities.
      - Incomplete domain model. Common in real world...
    - Practical Ignorance: Even if domain model complete, I may not have all necessary percepts on hand
  - The connection between toothaches-cavities is just not a logical consequence!

- Need a new solution: Probability theory
  - Allow stating a *degree of belief* in various statements in the KB

Probability

- Probabilistic assertions (sentences in KB) essentially summarize effects of
  - laziness: failure to enumerate exceptions, qualifications, etc.
  - ignorance: lack of relevant facts, initial conditions, etc.

- Clearly a *subjective* technique!
  - Extensive familiarity with domain required to accurately state probabilities
  - Need for extensive fine-tuning. Probabilities are *conditional* on evolving facts

- Subjective or Bayesian probability:
  - Probabilities relate propositions to one's own current state of knowledge
    - e.g., $P(A_{25}|\text{no reported accidents}) = 0.06$
    - 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., $P(A_{25}|\text{no reported accidents, time=5 a.m.}) = 0.15$
  - Interesting: Analogous to logical entailment status
    - KB |= α → means α entailed by KB...which represents what you currently know.
    - Analogously: KB = "no reported accidents, time=5 a.m." → $KB \models_{(0.15)} \alpha$

Making Decisions under Uncertainty

- Probability theory seems effective at expressing uncertainty.
  - But how do I actually reason (make decisions) in an uncertain world?

- Suppose I believe the following:
  - P(A_{25} gets me there on time | etc etc etc) = 0.04
  - P(A_{90} gets me there on time | etc etc etc) = 0.70
  - P(A_{120} gets me there on time | etc etc etc) = 0.95
  - P(A_{1440} gets me there on time | etc etc etc) = 0.9999

- Accurately expresses uncertainty with probabilities. But which plan should I *choose*?
  - Depends on my preferences for:
    - missing flight risk vs. wait time in airport vs. (pro/con) vs. (pro/con) vs. etc.
  - Utility theory is used to represent and infer preferences
    - Reasons about how useful/valued various outcomes are to an agent

- Decision Theory = Utility Theory + Probability Theory
  - Complete basis for reasoning in an uncertain world!

Probability Theory Basics

- Like logic assertions, probabilistic assertions are about possible worlds
  - Logical assertion α: all possible worlds in which α is false ruled out.
  - Probabilistic assertion α: states how *probable* various worlds are given α.

- Defn: Sample space: a set Ω = all possible worlds that might exist
  - e.g., after two dice roll: 36 possible worlds (assuming distinguishable dice)
  - Possible worlds are *exclusive* and *mutually exhaustive*
    - Only one *can* be true (the actual world); at least one *must* be true
  - ω ∈ Ω is a sample point (possible world)

- Defn: probability space or probability model is a sample space with an assignment P(ω) for every ω∈ Ω such that:
  - $0 \le P(\omega) \le 1$
  - $\Sigma_\omega P(\omega) = 1$
  - e.g. for die roll: P(1,1) = P(1,2) = P(1,3) =... = P(6,6) = 1/36.

- An event A is any subset of Ω
  - Allows us to group possible worlds, e.g., "doubles rolled with dice"
  - $P(A) = \Sigma_{\{\omega \in A\}} P(\omega)$
  - e.g., P(doubles rolled) = P (1,1) + P (2,2) + ... + P (6,6)

Probability Theory Basics

- A proposition in the probabilistic world is then simply an assertion that some event (describing a set of possible worlds) is true.
  - θ="doubles rolled" → asserts event "doubles" is true → asserts {[1,1] $\vee$ [2,2] $\vee$ ... $\vee$ [6,6]} is true.
  - Propositions can be compound: $\theta = (\text{doubles} \wedge (\text{total} > 4))$
  - $P(\theta) = \Sigma_{\omega \in \theta} P(\omega)$ → probability of proposition is sum of its parts

- Nature of probability of some proposition θ being true can vary, depending:
  - Unconditional or prior probability = *a priori* belief in truth of some proposition *in the absence of other info.*
    - e.g. P(doubles) = 6 * (1/36) = 1/6 → odds given no other info.
    - But what if one die has already rolled a 5? Or I now know dice are loaded?
  - Conditional or posterior probability = probability *given* certain information
    - Maybe P(cavity) = 0.2 (the prior)... but P(cavity | toothache) = 0.6
    - Or could be: P(cavity | toothache $\wedge$ (dentist found no cavity)) = 0

Probability Theory Basics

- Syntax: how to actually write out a proposition
  - A factored representation: states all of the “things” that are asserted true.

  - “Things” = random variables (begin with upper case)
    - The features that together define a possible world by taking on values
    - E.g. “Cavity”, “Total-die-value”, “Die_1”

  - Every variable has domain = set of possible values
    - domain(Die_1) = {1,2,3,4,5,6} ; domain(Total-die-value)={1,2,...,12}

  - Variables with a boolean domain can (syntactic sugar) be compacted:
    - domain(Cavity) = {true, false} → instead of “Cavity=true”, just write “cavity”
    - conversely for Cavity=false → ¬cavity

  - Probability of proposition = summed probability of atomic events
    - P(DieSum=7) = P(6,1) + P(2,5) + P(5,2) + P(3,4) + etc etc

Probability Distributions

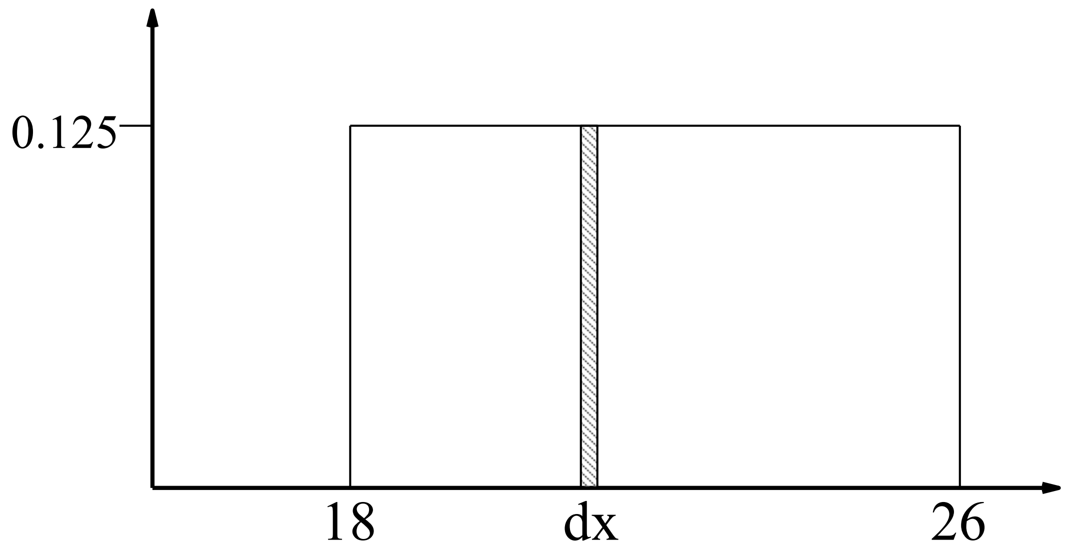
- So we can now express the probability of a proposition:
  - P(Weather=sunny) = 0.6 ; P(Cavity=false) = P(¬cavity)=0.1
- Probability Distribution expresses *all possible probabilities* for some event
  - So for: P(Weather=sunny) = 0.6; P(Weather=rain) = 0.1; etc etc →
  - **P**(Weather) = {0.72, 0.1, 0.29, 0.01} for Weather={sun, rain, clouds, snow}
    - Can be seen as total function that returns probabilities for all values of Weather
    - Is normalized, i.e., sum of all probabilities adds up to 1.
    - Note that bold **P** means prob. distr.; plain P means plain probability
- Joint Probability Distribution: for a set of random variables, gives probability for every combo of values of every variable.
  - Gives probability for every event within the sample space
  - **P**(Weather, Cavity) = a 4x2 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

- **Full** Joint Probability Distribution = joint distribution for all random variables in domain
  - Every probability question about a domain can be answered by full joint distribution
    - because every event is a sum of sample points (variable/value pairs)

Probability Distributions: for continuous variables

- What about continuous random variables?
  - Some variables are continuous, e.g. P(Temp=82.3) = 0.23; P(Temp=82.5)= 0.24; etc.
  - Also could assert ranges: P(Temp<85) ; P(40<Temp<67)
- We can express distributions as a parameterized function of value:
  - P (X = x) = U [18, 26](x) = uniform density between 18 and 26
  - Known as a probability density function (pdf)

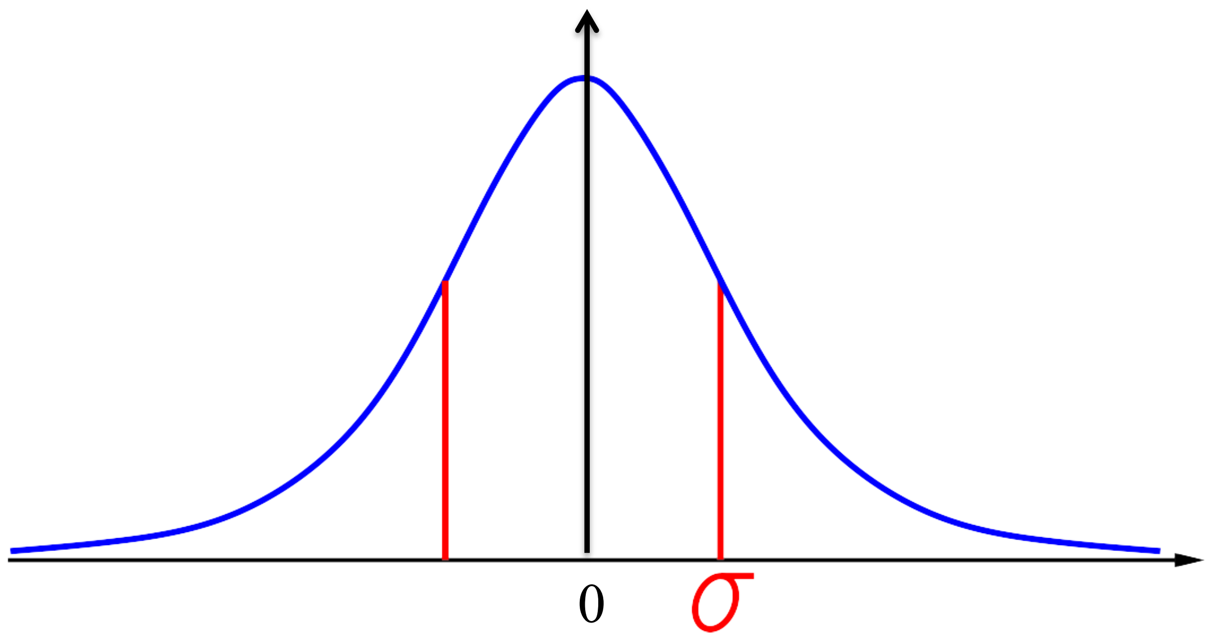


– Here P is a really a density distribution; the whole range integrates to 1.

- Probability of falling in 67-75 range is 100%
- Probability of NoonTemp at any single value is actually zero!
- P (X = 20.5) = 0.125 really means $\text{Lim}_{dx \to 0}\ P(20.5 \leq X \leq 20.5 + dx)/dx = 0.125$

Probability Distributions

Another example: simple Gaussian distribution: $\frac{1}{\sqrt{2\pi\sigma}} e^{-(x-\mu)^2/2\sigma^2}$



Conditional Probability

- Let's take a closer look now...
  - Precise meaning of: P(cavity | toothache) = 0.8 ?
    - **Not:** "if toothache, then 80% chance of cavity" !
    - That would be a hard fact: "whenever toothache, P(cavity) is 80%"
    - **Yes:** "P(cavity)=80% given that all I know is toothache"
    - Leaves room for P(cavity | (toothache ∧ fist-fight)) = 0.01
    - Less specific belief P(cavity | toothache)=0.8 remains true after more evidence arrives....but is less useful.
  - Some evidence may be "irrelevant", allowing simplification:
    - P(cavity | toothache, NAUjacksWin) = P(cavity | toothache) = 0.8
    - "Irrelevance" determined by detailed domain knowledge. We'll come back to this...
- Conditional Distributions
  - Concept of distributions can also by used for conditional probability
  - **P**(Cavity | Toothache) = probabilities for all values in range of Cavity, Toothache
    - = { P(cavity | toothache), P(¬cavity | toothache), P(cavity | ¬toothache), P(¬cavity | ¬toothache) }
  - So: **P**(X | Y) = gives values of $P(X=x_i \mid Y=y_j)$ for all possible i,j in ranges of X,Y

Computing with Conditional Probability

- Conditional probability can be defined in terms of unconditional probability:

$$P(a|b) = \frac{P(a \wedge b)}{P(b)} \qquad \text{E.g.:} \qquad P(doubles)|Die_1=5) = \frac{P(doubles \wedge Die_1=5)}{P(Die_1=5)}$$

  – can be rewritten, giving the product rule:
    - $P(a \wedge b) = P(a|b)\, P(b)$
    - Makes sense:
      – For $(a \wedge b)$ to be true, we need b to be true...and need a to be true given b

- Also works for distributions:
  – **P**(Weather, Cavity) = **P**(Weather|Cavity) **P**(Cavity)
    - Stands for a (4 values for Weather) x (2 values for Cavity) = 8 product equations

- The chain rule is derived by successive application of product rule:

$P(X_1, \ldots, X_n) = P(X_1, \ldots, X_{n-1})\, P(X_n|X_1, \ldots, X_{n-1})$

$= P(X_1, \ldots, X_{n-2})\, P(X_{n1}|X_1, \ldots, X_{n-2})\, P(X_n|X_1, \ldots, X_{n-1})$

= ...

$$\Pi_{i=1}^{n} \boldsymbol{P}(X_i|X_1 \ldots X_{i-1})$$

  – Note the recursive reduction joint P into a chained product of conditional P's

Inference in a probabilistic world

- Just need a couple more probabilistic rules:
  - Obvious: $P(\neg a) = 1 - P(a)$
  - Inclusion-Exclusion Principle: $P(a \vee b) = P(a) + P(b) - P(a \wedge b)$

- So how to do Inference?
  - Logical Inference = asking whether something is true (entailed), given the KB
  - Probabilistic Inference = asking *how likely something is*, given the KB
    - Just compute the posterior probability for query proposition, given KB!
  - We use the full joint probability distribution as the KB!
    - Contains the probability of all possible worlds!
  - Inference = look up the probability of a query proposition
    - Extract and sum up the appropriate “slice” of the joint distribution

- Example: Consider a world with just three boolean variables
  - Toothache (has one or not)
  - Cavity (has or not)
  - Catch (dentists tool catches or not)

Inference using full joint distribution

Start with the full joint distribution for this world:

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

For any proposition φ, the P(φ) = sum the atomic events where it is true:

$$P(\varphi) = \Sigma_{\omega:\omega|=\varphi} P(\omega)$$

Inference using full joint distribution

Start with the full joint distribution for this world:

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

For any proposition φ, the P(φ) = sum the atomic events where it is true:

$$P(\varphi) = \Sigma_{\omega:\omega|=\varphi} P(\omega)$$

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

This process is called summing out or marginalization

- Sum up probabilities across values of other (non-specified) variables
- In this case: Cavity and Catch
- Generally: $P(Y) = \Sigma_{z\in Z} P(Y,z)$,or also, by product rule: $P(Y)=\Sigma_{z\in Z} P(Y|z)\ P(z)$

Inference using full joint distribution

Start with the full joint distribution for this world:

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

For any proposition φ, the P(φ) = sum the atomic events where it is true:

$$P(\varphi) = \Sigma_{\omega:\omega\models\varphi} P(\omega)$$

Can also easily do compound propositional queries:

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

Inference using full joint distribution

Start with the full joint distribution for this world:

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

Can also compute conditional probabilities:

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

(Product rule)

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

$$= 0.4$$

Normalization

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

$$\text{P(cavity|toothache)} = \frac{P(cavity \wedge toothache)}{P(toothache)}$$

- Denominator can be viewed as a normalization constant α for the distribution **P**(Cavity|toothache)
  - Ensures that the probability of the distribution adds up to 1.

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

- Note that proportions between (0.12, 0.08) and (0.6, 0.4) are same
  - Latter are just normalized by application of α to add up to 1.
  - So if α just normalizes, I could also normalize "manually" → divide by sum of two.
  - Wow: I don't need to actually know P(toothache) → can just normalize manually!

Inference using full joint distribution

- **In Summary:** Compute distribution of query variable by fixing evidence variables (those in the "given" part) and summing over hidden (all other) variables
  - Let's analyze the implications more closely...

- Let X be all the variables.
  - Typically, we want the conditional joint distribution of the query variables Y given specific values e for the evidence variables E
  - Then the hidden variables are H = X − Y − E
- Then the required summation of joint entries is done by summing out the hidden variables:
  - $\mathbf{P}(Y|E = e) = \alpha\mathbf{P}(Y, E = e) = \alpha\Sigma_{h\in H}\mathbf{P}(Y, E = e, H = h)$

- Problem: works great, can answer all queries...but exponential complexity:
  - For world with n boolean variables:
    - Requires $O(2^n)$ to create store joint distribution table; $O(2^n)$ to process table lookup
  - Jumps to $O(d^n)$ for random variables with a range of d values!
  - Fine for toy worlds with three variables. Real worlds → >100 variables!

- Inefficiency! How to even find/define the probabilities for $O(d^n)$ table entries!
  - Especially given that you may never consult most of them!
  - We need some more tools!

Independence of variables

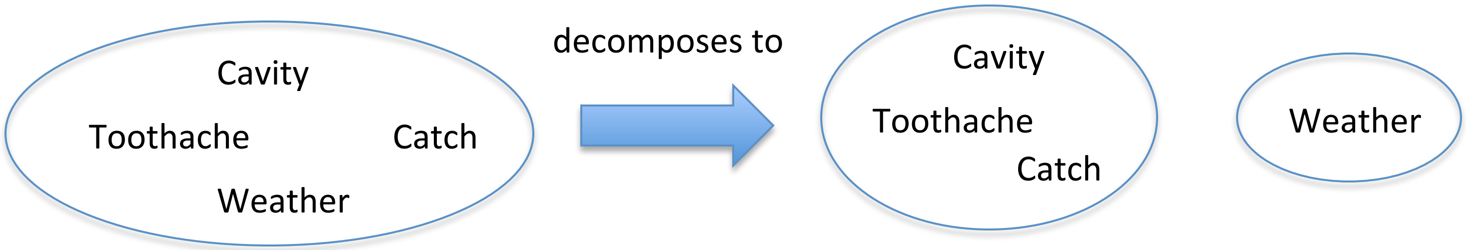
- **The problem:** full joint distribution get huge fast
  - the cross-product of all variables, all values in their range.
  - Different probability for every variables...conditional on all values of all other variables.

- But are all of these variables *really* related? Is every variable really related to all others?
  - Consider **P**(toothache, catch, cavity, cloudy) → 2 x 2 x 2 x 4 joint distr. = 32 entries
  - By product rule: P(toothache, catch, cavity, cloudy)
    = P(cloudy|toothache,catch,cavity) P(touchache,catch,cavity)
  - But it the weather really conditional on toothaches, cavities and dentist's tools? No!
  - So realistically: P(cloudy|toothache,catch,cavity) = P(cloudy)
  - So then actually: P(toothache, catch, cavity, cloudy) =
    P(cloudy) P(touchache,catch,cavity)
  - We say that cloudy and dental variables are independent (also absolute independence)
    - →probabilities separate → just multiplied simply.

- Effectively: the 32-element joint distribution table becomes one 8-element table + 4-element table

Independence of variables

- **Graphically:**



- Much easier to build/access 8-table + 4-table than 32-table!
  - 32 entries reduced to 12!
  - Generally: N dependent variables = 2^n vs. N independent variables = n **Wow!**

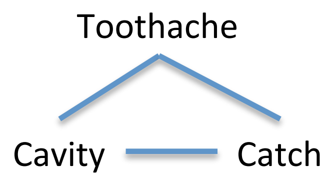
- Math: for independent variables X and Y:
- $\mathbf{P}(A|B) = \mathbf{P}(A)$ or $\mathbf{P}(B|A) = \mathbf{P}(B)$ or $\mathbf{P}(X,Y) = \mathbf{P}(X)\mathbf{P}(Y)$

- Independence assertions based on judgment, specific knowledge of domain
  - Can dramatically reduce information needed for full joint distribution ($2^n \rightarrow n$)
  - Sadly: absolute independence is quite rare in real world
    - Even an indirect connection must be accounted for as a conditional
  - Plus: even independent subset can still be large, e.g., real dentistry = 100's of variables

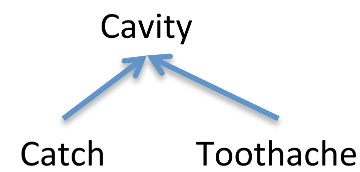
- Need more power!

Conditional Independence

- Consider again: Toothache, Catch, Cavity
  - Clearly **not** *independent*: toothache and tool and cavity obviously related
  - But what is the relationship?
    - Truly interconnected? No!



Or



  - Catch and Toothache are actually *halfway* independent of each other
    - They are related only *via* cavity. → they are both *caused by* the cavity
    - Formally: they are conditionally independent *given cavity*
    - Math notation: P(toothache ∧ catch | cavity) = P(toothache|cavity) P(catch|cavity)
  - Generally: given conditionally independent X, Y given some Z
    - **P**(X,Y|Z) = **P**(X|Z) **P**(Y|Z) and also **P**(X|Y,Z) = **P**(X|Z) and **P**(Y|X,Z)= **P**(Y|Z)
    - Allows same decomposition of large joint table to smaller ones as before:

      P(Toothache, Catch,Cavity)
      = P(Toothache,Catch|Cavity) P(Cavity) (prod. rule)
      = P(Toothache|Cavity) P(Catch|Cavity) P(Cavity) (using above)
  - One large table decomposed to three smaller ones. #entries: $O(2^n)$ → O(n) !

Conditional Independence

- Conditional independence is very common in real world!
  - Our basic and most robust form of knowledge about uncertain environments!

- A single cause often influences many conditionally independent effects
  - $\mathbf{P}(\text{Cause}, \text{Effect}_1, \text{Effect}_2, \ldots, \text{Effect}_n) = \mathbf{P}(\text{Cause}) \prod_i \mathbf{P}(\text{Effect}_i \mid \text{Cause})$
  - This probability distribution is a naive Bayes model
  - Naive: because it's often applied for simplicity...
    - Even when the effects are not strictly conditionally independent give the cause
    - Often works surprisingly well (i.e. "close enough" for good reasoning)

- Let's look at how we leverage conditional independence to reason...

Bayes Rule

- Recall the product rule:

  $P(a \wedge b) = P(a|b)\,P(b)$ or, conversely: $P(a \wedge b) = P(b|a)\,P(a)$
- equate and divide by P(a):

  Bayes rule: $$P(b|a) = \frac{P(a|b)\,P(b)}{P(a)}$$
- The basis for probabilistic inference in all modern AI systems!

- More generally, applied to probability distributions, we have:
-

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

  - As always, this represents a whole set of equations: every combo of var values

- And even more generally, conditioned on additional background info e :

$$P(Y|X,e) = \frac{P(X|Y,e)\,P(Y|e)}{P(X|e)}$$

Using Bayes Rule

- So: $P(Y|X) = \frac{P(X|Y)\,P(Y)}{P(X)}$
  - Doesn't seem super useful at first?
    - To calculate P(Y|X), I need P(X|Y) --- is that likely? Yes!
  - Very useful for cause-effect reasoning, e.g., diagnosis problems

$$P(\text{cause}|\text{effect}) = \frac{P(\text{effect}|\text{cause})\,P(\text{cause})}{P(\text{effect})}$$

- Example:

  A patient comes in with a stiff neck; one possible and very serious cause is meningitis. Epidemiological studies have shown that meningitis causes a stiff neck 70% of the time. It's also known that meningitis strikes about 1/50,000 people in general, and that about 1% of people have a stiff neck on any given day.

- So:
  - P(stiff|men) = 0.7
  - P(m) = 1/50,000 and P(stiff) = 1/100
  - P(men|stiff) = P(stiff|men) P(men) / P(stiff) = (0.7 * 1/50k)/0.01 = 0.0014

  - We often have probabilities in the causal direction…can compute probability in the diagnostic direction

Using Bayes Rule: a typical example

- Let's try this out:
  - Your doctor says you tested positive for a serious disease; test is 99% accurate. It's a rare disease though: only 1 in 10,000 people have it. Why should you be happy?

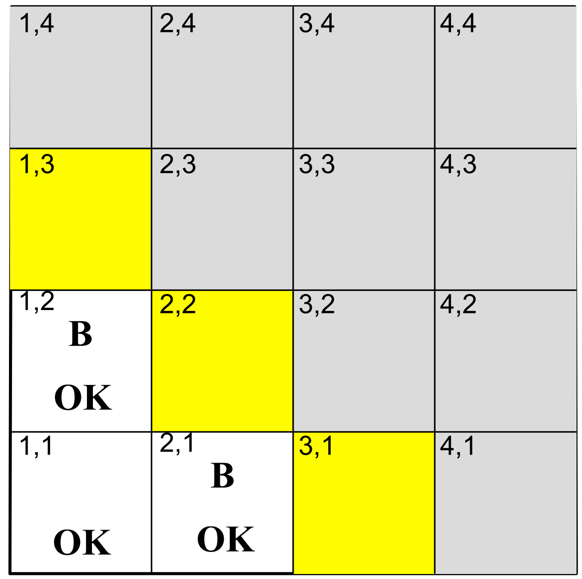
Summary

- Probability is a rigorous formalism for uncertain knowledge
  - Provide an entire mathematics for quantifying and calculating uncertainty
- Joint probability distribution specifies probability of every atomic event
  - Every combination of every variables across its whole range
  - Queries can be answered by summing over atomic events
- For nontrivial domains, we must find a way to reduce the joint table size
  - Size of joint distribution is $O(n^2)$ for n variables. Intractable.
  - Independence and conditional independence provide the tools
- Bayes Rule focuses probability calculus on forward diagnostic problems
  - Probability of a cause, given a set of conditionally independent effects
  - Useful for many “diagnosis” tasks
    - How likely is it that some event has occurred, given a set of observed evidence.
- Bayes rule provides the basis of probabilistic reasoning in AI
  - Basis for Bayesian networks (next chapter)

α β ⊆ ¬ ⇒ |= ∧ ∨
⇔

Extra slides…maybe next time…

Wumpus World



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

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

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

Specifying the probability model

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

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

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

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

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

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

for n pits.

Observations and query

We know the following facts:

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

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

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

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

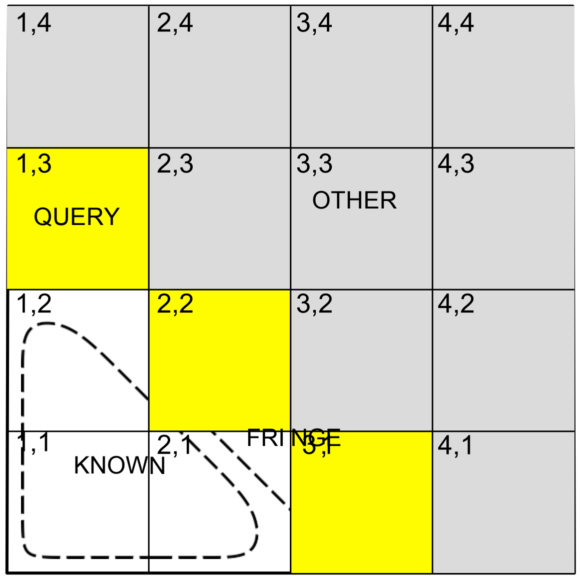
For inference by enumeration, we have

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

Grows exponentially with number of squares!

Using conditional independence

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



Define $Unknown = Fringe \cup Other$

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

Manipulate query into a form where we can use this!

Using conditional independence contd.

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

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

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

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

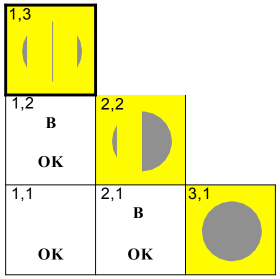
$$= \alpha \sum_{fringe} \mathbf{P}(b|known, P_{1,3}, fringe) \sum_{other} \mathbf{P}(P_{1,3}, known, fringe, other)$$

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

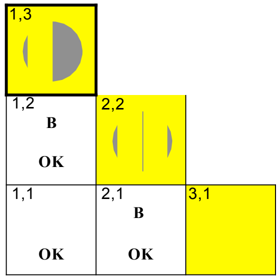
$$= \alpha\, P(known)\mathbf{P}(P_{1,3}) \sum_{fringe} \mathbf{P}(b|known, P_{1,3}, fringe)P(fringe) \sum_{other} P(other)$$

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

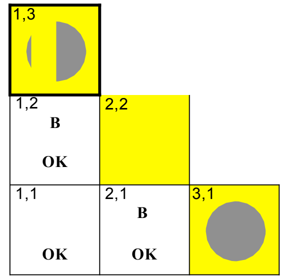
Using conditional independence contd.



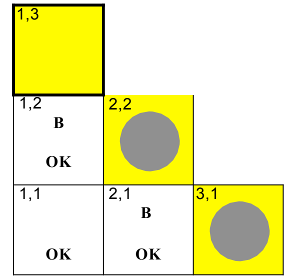
0.2 x 0.2 = 0.04



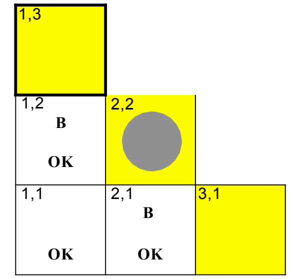
0.2 x 0.8 = 0.16



0.8 x 0.2 = 0.16



0.2 x 0.2 = 0.04



0.2 x 0.8 = 0.16

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

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