Chapter 13

Quantifying Uncertainty

1

Outline

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

2

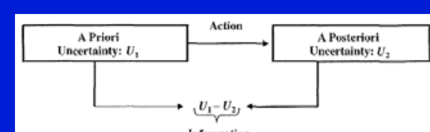
Theory of Uncertainty

- To develop a fully operational theory for dealing with uncertainty of some conceived type requires the issues be addressed at the following levels:
  - Find an appropriate *mathematical formalization* of the conceived type of uncertainty
  - Develop a *calculus* by which this type of uncertainty is manipulated.
  - Find a meaningful way of *measuring* the amount of relevant uncertainty.
  - Develop *methodological aspects* of the theory, including procedures of making the various uncertainty principles.

3

Information and Uncertainty

- ***Uncertainty_Based Information (UBI)***
  - Uncertainty in any problem-solving situation caused by information deficiency
  - Suppose that the amount of uncertainty conceptualized in a mathematical theory can be measured.
  - *Information Gain:* suppose that the amount of uncertainty can be reduced by obtaining relevant information as a result of action.
  - *Information as Uncertainty Reduction:* the amount of information obtained by the action may be measured by the reduction of uncertainty.



4

Continued..
Information and Uncertainty

- ***Uncertainty_Based Information (UBI)***
  - Problem: the restricted notion of information.
    The rich notion of information in human communication and cognition, or the algorithmic conception of information, etc. is not captured in the information measured by the reduction of relevant uncertainty.
  - *Cf) Descriptive information,* or *Algorithmic information:*
    The concept of information in terms of the *theory of computability:*
    The amount of information represented by an computable object
    is measured by the length of the shortest program written in language.
  - UBI : information conceived in terms of uncertainty reduction.
  - important in nondeterministic system.

Continued..
Information and Uncertainty

- The nature of UBI depends on the *mathematical theory* within which uncertainty pertaining to various problem-solving situation is formalized. – each formalization of uncertainty is a mathematical model of the problem-solving situation.
- The modeling becomes limited by the constraints of the theory when we commit ourselves to a particular mathematical theory.

Generalized Information Theory (GIT)

- Two classical roots of a formal treatment of UBI:
  - *Possibility*
  - *Probability*
- The (traditional) *Information Theory*: 1948 -
  - UBI was conceived in terms of *classical set theory* and *probability theory*.
  - a theory based upon the measure of probabilistic uncertainty
    – Shannon's measure.
- *The Generalized Information Theory:* 1980 -
  - based on the *broader conception* of UBI.
  - Goal: To capture properties of UBI formalized within any feasible mathematical framework.
  - Fuzzy set theory, Possibility theory, Evidence theory.
- The framework in GIT is based on two generalizations in mathematics:
  - classical measure theory → the theory of *monotone measures*
  - classical set theory → the theory of *fuzzy sets*.

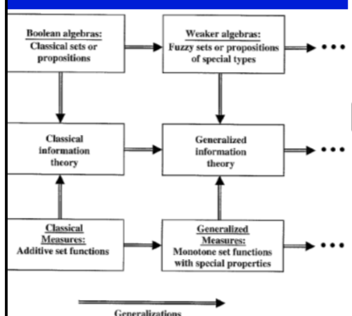


Figure 4.1. Classical information theory and its generalizations.

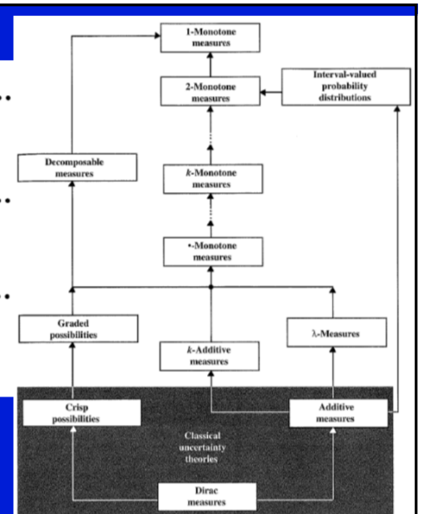


Figure 5.1. Ordering of monotone measures used for representing imprecise probabilities by their levels of generality. (Dirac measures are defined in Note 5.12.)

Uncertainty theories			Formalized languages						
			Classical Sets	Nonclassical Sets					
				Standard Fuzzy Sets	Nonstandard fuzzy sets				
					Interval Valued	Type 2	Level 2	Lattice Based	•••
Monotone Measures	Additive	Classical numerical probability							
	Nonadditive	Possibility/ necessity							
		Sugeno λ-measures							
		Belief/ plausibility (capacities of order ∞)							
		Capacities of various finite orders							
		Interval-valued probability distributions							
		⋮							
		General lower and upper probabilities							

continued.

- 3 Types of Uncertainty:
  - *Nonspecificity* (imprecision): connected with sizes of relevant sets of alternatives.
  - *Fuzziness* (vagueness): results from imprecise boundaries of sets.
  - *Strife* (discord): conflicts among the various sets of alternatives.
- 5 Theories of Uncertainty:
  - *Classical set Theory*: nonspecificity
  - *Fuzzy set Theory*: nonspecificity, fuzziness
  - *Possibility Theory*: nonspecificity, strife
  - *Evidence Theory*: nonspecificity, strife fuzziness (fuzzified ET)
  - *Probability Theory*: strife

Three basic types of uncertainty and Basic uncertainty measures

Figure 9.7 Three basic types of uncertainty.

TABLE 9.1 SUMMARY OF BASIC UNCERTAINTY MEASURES FOR FINITE SETS

Uncertainty Theory	Uncertainty Measure	Equation	Uncertainty Type	Year
Classical set theory	$U(A) = \log_2 \lvert A\rvert$	(9.1)	Nonspecificity	1928
Fuzzy set theory	$U(A) = \frac{1}{h(A)}\int_0^{h(A)} \log_2 \lvert {}^\alpha A\rvert d\alpha$	(9.22)	Nonspecificity	1983
Possibility theory	$U(r) = \sum_{i=2}^{n} r_i \log_2 \frac{i}{i-1}$	(9.26)	Nonspecificity	1983
Evidence theory	$N(m) = \sum_{A\in\mathcal{F}} m(A)\log_2 \lvert A\rvert$	(9.36)	Nonspecificity	1985
Probability theory	$H(m) = -\sum_{x\in X} m(\{x\})\log_2 m(\{x\})$	(9.37)	Strife	1948
Evidence theory	$S(m) = -\sum_{A\in\mathcal{F}} m(A)\log_2 \sum_{B\in\mathcal{F}} m(B)\frac{\lvert A\cap B\rvert}{\lvert A\rvert}$	(9.54)	Strife	1992
Possibility theory	$S(r) = \sum_{i=2}^{n}(r_i - r_{i+1})\log_2 \frac{i}{\sum_{j=1}^{i} r_j}$	(9.58)	Strife	1992
Evidence theory	$NS(m) = \sum_{A\in\mathcal{F}} m(A)\log_2 \frac{\lvert A\rvert^2}{\sum_{B\in\mathcal{F}} m(B)\lvert A\cap B\rvert}$	(9.59)	Total: $N(m) + S(m)$	1992
Possibility theory	$NS(r) = \sum_{i=2}^{n}(r_i - r_{i+1})\log_2 \frac{i^2}{\sum_{j=1}^{i} r_j}$	(9.60)	Total: $N(r) + S(r)$	1992
Fuzzy set theory	$f(A) = \sum_{x\in X}[1 - \lvert 2A(x) - 1\rvert]$	(9.34)	Fuzziness	1979
Fuzzified evidence theory	$F(m) = \sum_{A\in\mathcal{F}} m(A) f(A)$	(9.61)	Fuzziness	1988

Table 5.1. Probability Theory Versus Possibility Theory: Comparison of Mathematical Properties for Finite Sets

Probability Theory	Possibility Theory
Based on measures of *one type*: probability measures, *Pro*	Based on measures of *two types*: possibility measures, *Pos*, and necessity measures, *Nec*
Body of evidence consists of *singletons*	Body of evidence consists of a *family of nested subsets*
Unique representation of *Pro* by a *probability* distribution function $p: X \to [0, 1]$ via the formula $Pro(A) = \sum_{x\in A} p(x)$	Unique representation of *Pro* by a *basic possibility function* $r: X \to [0, 1]$ via the formula $Pos(A) = \max_{x\in A} r(x)$
Normalization: $\sum_{x\in X} p(x) = 1$	Normalization: $\max_{x\in X} r(x) = 1$
Additivity: $Pro(A \cup B) = Pro(A) + Pro(B) - Pro(A \cap B)$	Max/Min rules: $Pos(A \cup B) = \max\{Pos(A), Pos(B)\}$ $Pos(A \cap B) \le \min\{Pos(A), Pos(B)\}$ $Nec(A \cap B) = \min\{Nec(A), Nec(B)\}$ $Nec(A \cup B) \ge \max\{Nec(A), Nec(B)\}$
Not applicable	Duality: $Nec(A) = 1 - Pos(\bar{A})$ $Pos(A) < 1 \Rightarrow Nec(A) = 0$ $Nec(A) > 0 \Rightarrow Pos(A) = 1$
$Pro(A) + Pro(\bar{A}) = 1$	$Pos(A) + Pos(\bar{A}) \ge 1$ $Nec(A) + Nec(\bar{A}) \le 1$ $\max\{Pos(A), Pos(\bar{A})\} = 1$ $\min\{Nec(A), Nec(\bar{A})\} = 0$
Total ignorance: $p(x) = 1/\lvert X\rvert$ for all $x \in X$	Total ignorance: $r(x) = 1$ for all $x \in X$
Conditional probabilities: $p_{X\mid Y}(x\mid y) = \frac{p(x,y)}{p_Y(y)}$ $p_{X\mid Y}(y\mid x) = \frac{p(x,y)}{p_X(x)}$	Conditional possibilities: $r_{X\mid Y}(X\mid Y) = \begin{cases} r(x,y) & \text{when } r_Y(y) > r(x,y) \\ [r(x,y), 1] & \text{when } r_Y(y) = r(x,y) \end{cases}$ $r_{Y\mid X}(Y\mid X) = \begin{cases} r(x,y) & \text{when } r_X(x) > r(x,y) \\ [r(x,y), 1] & \text{when } r_X(y) = r(x,y) \end{cases}$
Probabilistic noninteraction: $p(x, y) = p_X(x) \cdot p_Y(y)$ (a) Probabilistic independence: $p_{X\mid Y}(x\mid y) = p_X(x)$, $p_{Y\mid X}(y\mid x) = p_Y(y)$ (b) (a) ⇒ (b) and (b) ⇒ (a)	Possibilistic noninteraction: $r(x, y) = \min\{r_X(x), r_Y(y)\}$ (α) Possibilistic independence: $r_{X\mid Y}(x\mid y) = r_X(x)$, $r_{Y\mid X}(y\mid x) = r_Y(y)$ (β) (β) ⇒ (α) but (α) ⇏ (β)

Uncertainty

- Agent's act under uncertainty:
  - agents almost never have access to the whole truth about their environment.
  - A complex real world than a toy world.
    -- it might be impossible to construct a complete and correct description of how its actions will work.
- What an agent should do when not all is crisply clear?
- The rational decision depends on both the relative importance of various goals and the likelihood that, and degree to which, they will be achieved.

13

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 (traffic reports)
3. uncertainty in action outcomes (flat tire, etc.)
4. immense complexity of modeling 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 …)

14

Methods for Handling Uncertainty

- ***Rules with (Un)certainty factors:***
  - A_{25} |-- $_{0.3}$ get there on time
  - $Sprinkler$ |-- $_{0.99}$ $WetGrass$
  - $WetGrass$ |— $_{0.7}$ $Rain$
  - Issues: Problems with combination, e.g., *Sprinkler* causes *Rain*??
- ***Probability***
  - Model agent's *degree of belief* :
    The agent's knowledge can at best provide only a degree of belief in the relevant sentences.
  - Given the available evidence,
    A_{25} will get me there on time with probability 0.04
- ***Evidence Theory: Dempster-Shafer Theory***
  - Uncertainty of ignorance
  - Interval-valued degrees of belief ([Bel, Pl]) to represent an agent's knowledge of the *probability of a proposition*.
- ***Fuzzy Set and Logic***
  - Uncertainty of vagueness :

15

Probability

- Probabilistic assertions *summarize* effects of
  - *laziness:* failure to enumerate exceptions, qualifications, etc.
    -- Too much work to list the complete set of antecedents/consequents.
  - *ignorance:* lack of relevant facts, initial conditions, etc.
    - Theoretical ignorance, practical ignorance
- *Subjective* or *Bayesian* probability:
  Probabilities relate propositions to *agent's own state of knowledge*
  e.g., $P(A_{25}$ | *no reported accidents*$) = 0.06$
- These are *not* assertions about the world
  -- not claims of some probabilistic tendency in the current situation but might be learned from past experience of similar situation
- Probabilities of propositions change with new evidence:
  e.g., $P(A_{25}$ | no reported accidents, 5 a.m.$) = 0.15$

  (analogous to logical entailment status KB |= α, not truth)

16

Making Decisions under Uncertainty

- Suppose I believe the following:
  $P(A_{25}$ gets me there on time | …) = *0.04*
  $P(A_{90}$ gets me there on time | …) = *0.70*
  $P(A_{120}$ gets me there on time | …) = *0.95*
  $P(A_{1440}$ gets me there on time | …) = *0.9999*
- Which action to choose?
  depends on my *preferences* for missing flight vs. waiting time, etc.
- *Utility theory* is used to represent and infer *preferences*
  -- the degree of usefulness
- *Decision theory* = Probability theory + Utility theory

17

Probability Basics

- A degree of belief
- Begin with a set Ω - the *sample space*
  - e.g. 6 possible rolls of a die
  - $\omega \in \Omega$ is a *sample point*, a possible world or an *atomic event*
- A *probability space* or *probability model* is a sample space with an assignment $P(\omega)$ for every $\omega \in \Omega$ s.t.
  $0 \le P(\omega) \le 1$
  $\Sigma_{\omega} P(\omega) = 1$
  e.g. $P(1)=P(2)=P(3)=P(4)=P(5)=P(6)=1/6$
- An *event* A is any subset of Ω
  $P(A)=\Sigma_{\{\omega \in A\}} P(\omega)$
  e.g.) P(die roll $< 4) = P(w=1 \cup w=2 \cup w=3) = 1/6+1/6+1/6=1/2$

18

Random Variable

- A *random variable* is a function from sample points to some range, e.g. the reals or booleans
  - e.g. $Odd(1)=true$.
  - *Boolean random variable:*
    the value of r.v. is from the domain <true, false>
    e.g.) $Cavity = true.$
  - *Discrete random variable*
    the value of r.v. is from the countable domain.
    e.g.) $Weather = v$, where $v \in$ <sunny, rainy, cloudy, snow>
  - *Continuous random variable*
    the value of r.v. is from the real numbers
    e.g.) $x \le 4.02$
- P induces a *probability distribution for any random variable X:*
  $P(X=x_i) = \Sigma_{\{\omega: X(\omega)=xi\}} P(\omega)$

e.g.) $P(\text{Odd=true}) = P(1)+P(3)+P(5) = 1/6+1/6+1/6=1/2$

19

Propositions

- Think of a *proposition* as the event (set of sample points) where the proposition is true *-- an assertion that such-and-such is the case*.
- Given *Boolean* random variables A and B:
  - Event a = set of sample points (ω's) 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.) $(a \vee b) \equiv (\neg a \wedge b) \vee (a \wedge \neg b) \vee (a \wedge b)$
    $\Rightarrow P(a \vee b) = P(\neg a \wedge b) + P(a \wedge \neg b) + P(a \wedge b)$

20

Syntax for Propositions

- Basic element: *random variable*
- Similar to propositional logic: possible worlds defined by assignment of values to random variables.
- *Propositional or Boolean* random variables
  e.g., *Cavity* ∈ <true, false> (do I have a cavity?)
- *Discrete* random variables (*finite* or *infinite*)
  e.g., *Weather* ∈ *<sunny, rainy, cloudy, snow>*
  – *Weather = rain* is a *proposition*
  – Domain values must be exhaustive and mutually exclusive
  – Elementary proposition constructed by assignment of a value to a random variable:
    e.g., *Weather = sunny*, *Cavity = false* (abbreviated as ¬*Cavity*)
  – Complex propositions formed from elementary propositions and standard logical connectives: e.g., (*Weather = sunny*) ∨ (*Cavity = false*)
- *Continuous* random variables (*bounded* or *unbounded*)
  – Takes on values from the real number ∈ R, or ∈ the interval (e.g. [0, 1]).
  – e.g. X=4.02 asserts that the random variable X has the exact value 4.02
  – *E..g Temp=21.6;* also allows, *e.g. Temp < 22.0*

21

Continued.. Syntax

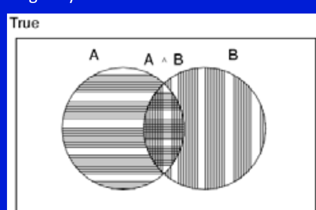
- *Atomic event*:
  A *complete* specification of the state of the world about which the agent is uncertain
  E.g.) if the world consists of only two Boolean variables *Cavity* and *Toothache*, then there are 4 distinct atomic events:
  Cavity = false ∧*Toothache = false*
  Cavity = false ∧ *Toothache = true*
  Cavity = true ∧ *Toothache = false*
  Cavity = true ∧ *Toothache = true*
- Atomic events are *mutually exclusive* and *exhaustive*

22

Axioms of Probability: Kolmogorov's Axiom

- For any propositions *A, B*
  – $0 \le P(A) \le 1$
  – $P(true) = 1$ and $P(false) = 0$
  – $P(A \vee B) = P(A) + P(B) - P(A \wedge B)$: the probability of a disjunction -- Additivity.

  -- certain logically related events must have related probabilities



23

Prior Probability

- *Prior* or *Unconditional Probabilities* of propositions
  e.g., $P(Cavity = \text{true}) = 0.1$ and $P(Weather = \text{sunny}) = 0.72$
  correspond to *belief prior to arrival of any (new) evidence*
- *Probability Distribution* gives values for all possible assignments:
  $\mathbf{P}(Weather) = <0.72, 0.1, 0.08, 0.1>$ (*normalized*, i.e., sums to 1)

  Cf) **Probability Density Function (pdf)**: a PD for a *continuous* variable.
- *Joint Probability Distribution* for a *set of random variables* gives the probability of every atomic event on those random variables
  $\mathbf{P}(Weather, Cavity)$ = 4 × 2 matrix of values:

Weather =	sunny	rainy	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*:
  A JPD that covers the complete set of random variables.
- *Every question about a domain can be answered by the JPD because every event is a sum of sample points!*

24

Posterior Probability

- ***Conditional*** or ***Posterior Probabilities***
  e.g., $P(cavity \mid toothache) = 0.8$
  i.e., *given that toothache is all I know*
  - If a patient is observed to have a toothache and no other information is yet available, then the probability of the patient's having a cavity will be 0.8.
- If we know more, e.g., *cavity* is also given, then we have
  $P(cavity \mid toothache, cavity) = 1$
- 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

25

Continued.. Posterior Probability

- Definition of conditional probability:
  $P(a \mid b) = P(a \wedge b) / P(b) \quad \text{if } P(b) \neq 0$
  - the probability of *a*, given that all we know is *b*.
- ***Product Rule*** gives an alternative formulation:
  $P(a \wedge b) = P(a|b)\, P(b) = P(b|a) \cdot P(a)$
- A general version holds for whole distributions, e.g.,
  $\mathbf{P}(Weather, Cavity) = \mathbf{P}(Weather \mid Cavity)\, \mathbf{P}(Cavity)$
- ***Chain Rule*** is derived by successive application of product rule:

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

26

Inference by Enumeration using Full Joint Distribution

- Start with the Joint Probability 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 ω where it is true:
  $P(\phi) = \Sigma_{\omega:\omega \models \phi} P(\omega)$

27

Continued.. Inference by Enumeration using FJD

- Start with the Joint Probability 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:
  $P(\phi) = \Sigma_{\omega:\omega \models \phi} P(\omega)$
- *P(toothache)* = P(catch ∧ cavity ∧ toothache)
  + P(catch ∧ ¬cavity ∧ toothache)
  + P(¬ catch ∧ cavity ∧ toothache)
  + P(¬ catch ∧ ¬ cavity ∧ toothache)
  $= 0.108 + 0.012 + 0.016 + 0.064 = 0.2$

28

Continued.. Inference by Enumeration using FJD

- Start with the Joint Probability 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:
  $P(\varphi) = \Sigma_{\omega:\omega \models \varphi} P(\omega)$
- $P(cavity \vee toothache)$
  $= 0.108 + 0.012 + 0.072 + 0.008 + 0.016 + 0.064 = 0.28$

29

Continued.. Inference by Enumeration using FJD

- Start with the Joint Probability Distribution:

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

- Can also compute conditional probabilities:

$$P(\neg cavity \mid toothache) = \frac{P(\neg cavity \wedge toothache)}{P(toothache)} = \frac{0.016+0.064}{0.108 + 0.012 + 0.016 + 0.064} = 0.4$$

30

Normalization

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

- Denominator can be viewed as a ***normalization constant*** α

  $P(Cavity \mid toothache) = \alpha\, P(Cavity, toothache)$
  $= \alpha\, [P(Cavity, toothache, catch) + P(Cavity, toothache, \neg catch)]$
  $= \alpha\, [<0.108, 0.016> + <0.012, 0.064>]$ where $Cavity{=}true$, i.e. $cavity$
  $= \alpha <0.12, 0.08> = <0.6, 0.4>$ and $Cavity{=}false$, i.e. $\neg cavity$
  where $\alpha = P^{-1}(toothache) = .108 + .012 + .016 + .064 = 1/.2{=}5$
  -- no need to compute α !!

  $P(cavity \mid toothache) = \alpha\, P(cavity, toothache)$?
  $P(\neg cavity \mid toothache) = \alpha\, P(\neg cavity, toothache)$?
- General idea: Compute distribution on *Query variable* by fixing ***Evidence variables*** and summing over ***Hidden variables***

31

Inference by Enumeration, contd.

- Typically, we are interested in
  the *Posterior Joint Distribution* of the *Query variables* **Y**
  given specific values **e** for the *Evidence variables* ***E*** : $\mathbf{P}(Y \mid E{=}e)$
- Let the *Hidden variables* be ***H*** = ***X*** - ***Y*** - ***E***
- Then, the required summation of joint entries is done by summing out the hidden variables:
  $P(Y \mid E = e) = \alpha\, P(Y, E = e) = \alpha\, \Sigma_h P(Y, E = e, H = h)$
- The terms in the summation are joint entries because **Y**, **E** and **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?

32

Independence

- *A* and *B* are **independent** iff
  $P(A|B) = P(A)$ or $P(B|A) = P(B)$ or $P(A, B) = P(A)\,P(B)$

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

- $32 (=2^3 * 4)$ entries reduced to 12; for n independent biased coins, $O(2^n) \rightarrow O(n)$
- Absolute independence is powerful but rare
  -- Independence assertions can help in reducing the size of the domain representation and the complexity of the inference problem,
  but *no clean cut* b/t entire sets of variables.
- Dentistry is a large field with hundreds of variables, none of which are independent. What to do?

33

Conditional 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) $\mathbf{P}(catch \mid toothache, cavity) = \mathbf{P}(catch \mid cavity)$
- The same independence holds if I haven't got a cavity:
  (2) $\mathbf{P}(catch \mid toothache, \neg cavity) = \mathbf{P}(catch \mid \neg cavity)$
- *Catch* is *conditionally independent* of *Toothache* given *Cavity*:
  $\mathbf{P}(Catch \mid Toothache, Cavity) = \mathbf{P}(Catch \mid Cavity)$
- Equivalent statements:
  $\mathbf{P}(Toothache \mid Catch, Cavity) = \mathbf{P}(Toothache \mid Cavity)$
  $\mathbf{P}(Toothache, Catch \mid Cavity) = \mathbf{P}(Toothache \mid Cavity)\,\mathbf{P}(Catch \mid Cavity)$

34

Continued.. Conditional Independence

- Write out Full Joint Distribution using chain rule:
  $\mathbf{P}(Toothache, Catch, Cavity)$
  $= \mathbf{P}(Toothache \mid Catch, Cavity)\,\mathbf{P}(Catch, Cavity)$
  $= \mathbf{P}(Toothache \mid Catch, Cavity)\,\mathbf{P}(Catch \mid Cavity)\,\mathbf{P}(Cavity)$
  $= \mathbf{P}(Toothache \mid Cavity)\,\mathbf{P}(Catch \mid Cavity)\,\mathbf{P}(\text{Cavity})$

  i.e., 2 + 2 + 1 = 5 independent numbers

- *In most cases, the use of conditional independence reduces the size of the representation of the joint distribution from $O(2^n)$ to $O(n)$.*
- *Conditional independence is our most basic and robust form of knowledge about uncertain environments.*

35

Bayes' Rule

- Product rule $P(a \wedge b) = P(a \mid b)\,P(b) = P(b \mid a)\,P(a)$
  $\Rightarrow$ Bayes' rule: $P(a \mid b) = P(b \mid a)\,P(a) / P(b)$

  or in distribution form
  $\mathbf{P}(Y|X) = \mathbf{P}(X|Y)\,\mathbf{P}(Y) / \mathbf{P}(X) = \alpha\,\mathbf{P}(X|Y)\,\mathbf{P}(Y)$
- Useful for assessing diagnostic probability from causal probability:
  $P(Cause|Effect) = P(Effect|Cause)\,P(Cause) / P(Effect)$
- E.g., let *M* be meningitis, *S* be stiff neck:
  $P(m|s) = P(s|m)\,P(m) / P(s) = 0.8 \times 0.0001 / 0.1 = 0.0008$
- Note: posterior probability of meningitis still very small!

36

Bayes' Rule and conditional independence: Combining Evidence

$\mathbf{P}(Cavity \mid toothache \wedge catch)$
$= \alpha \cdot \mathbf{P}(toothache \wedge catch \mid Cavity)\ \mathbf{P}(Cavity)$
$= \alpha \cdot \mathbf{P}(toothache \mid Cavity)\ \mathbf{P}(catch \mid Cavity)\ \mathbf{P}(Cavity)$

- This is an example of a *Naïve Bayes* model (Bayesian classifier):
  $\mathbf{P}(Cause, Effect_1, \ldots, Effect_n) = \mathbf{P}(Cause)\ \Pi_i\ \mathbf{P}(Effect_i|Cause)$

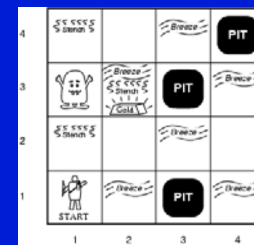
-- the effect variables are conditionally independent given the cause variable.

- Total number of parameters is *linear* in n : $O(n)$

37

Wumpus World: PEAS description

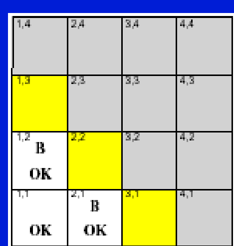
- **Performance measure**
  - gold +1000, death -1000
  - -1 per step, -10 for using the arrow
- **Environment (forms the rules/facts of KB)**
  - Squares adjacent to wumpus are smelly
  - Squares adjacent to pit are breezy
  - Glitter iff gold is in the same square
  - Shooting kills wumpus if you are facing it
  - Shooting uses up the only arrow
  - Grabbing picks up gold if in same square
  - Releasing drops the gold in same square



- **Sensors**: Stench, Breeze, Glitter, Bump, Scream
- **Actuators**: Left turn, Right turn, Forward, Grab, Release, Shoot

38

Wumpus World



- Aim: Probability each of [1,3], [2,2] and [3,1] contains a pit?
- 1. Identify the set of *random variables* needed:
  - P_{ij}*=true* iff $[i,j]$ contains a pit
  - B_{ij}*=true* iff $[i,j]$ is breezy
- Include only B_{11}, B_{12}, B_{21} in the probability model

39

Specifying the probability model

- 2. The *full joint distribution* is $\mathbf{P}(P_{11}, \ldots, P_{44}, B_{11}, B_{12}, B_{21})$
- Apply product rule: $\mathbf{P}(B_{11}, B_{12}, B_{21} | P_{11}, \ldots, P_{44})\ \mathbf{P}(P_{11}, \ldots, P_{44})$

  (Do it this way to get $\mathbf{P}(Effect|Cause)$.)
- First term: $\mathbf{P}(B_{11}, B_{12}, B_{21} | P_{11}, \ldots, P_{44})$
  - conditional probability of a breeze configuration, given a pit configuration.
  - 1 if pits are adjacent to breezes, 0 otherwise
- Second term: $\mathbf{P}(P_{11}, \ldots, P_{44})$
  - prior probability of a pit configuration
  - pits are placed randomly, probability 0.2 per square: $P(P_{ij}) = .2\ \forall i,j$

    $\mathbf{P}(P_{11}, \ldots, P_{44}) = \Pi_{(i,j)=(1,1)\ \ldots(4,4)}\ \mathbf{P}(P_{ij}) = 0.2^n * 0.8^{16-n}$ for n pits.

40

Observations and query

- We know the following facts:

  $b = \neg b_{11} \wedge b_{12} \wedge b_{21}$

  $known = \neg p_{11} \wedge \neg p_{12} \wedge \neg p_{21}$

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

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

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

- Grows exponentially with number of squares.
  - There are 12 unknown squares; 2^{12}=4096 terms.

41

Using conditional independence

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

- Define $Unknown = Frontier \cup Other$
- $\mathbf{P}(b \mid P_{1,3}, known, Unknown) = \mathbf{P}(b \mid P_{1,3}, known, Frontier, \cancel{Other})$
- Manipulate query into a form where we can use this.

42

Continued.. Using conditional independence

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

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

$= \alpha\, \Sigma_{frontier}\, \Sigma_{other}\, P(b \mid known, P_{1,3}, Frontier, Other)\, P(P_{1,3}, known, Frontier, Other)$

$= \alpha\, \Sigma_{frontier}\, \Sigma_{other}\, P(b \mid known, P_{1,3}, Frontier)\, P(P_{1,3}, known, Frontier, Other)$

$= \alpha\, \Sigma_{frontier}\, P(b \mid known, P_{1,3}, Frontier)\, \Sigma_{other}\, P(P_{1,3}, known, Frontier, Other)$

$= \alpha\, \Sigma_{frontier}\, P(b \mid known, P_{1,3}, Frontier)\, \Sigma_{other}\, P(P_{1,3}) P(known)\, P(Frontier)\, P(Other)$

$= \underline{\alpha\, P(known)}\, P(P_{1,3})\, \Sigma_{frontier}\, P(b \mid known, P_{1,3}, Frontier)\, P(Frontier)\, \underline{\Sigma_{other}\, P(Other)}$

$= \underline{\alpha'}\, P(P_{1,3})\, \Sigma_{frontier}\, P(b \mid known, P_{1,3}, Frontier)\, P(Frontier)$

$= \underline{\alpha'}\, P(P_{1,3})\, \Sigma_{P2,2,P3,1}\, P(b \mid known, P_{1,3}, P_{2,2}, P_{3,1})\, P(P_{2,2}, P_{3,1})$

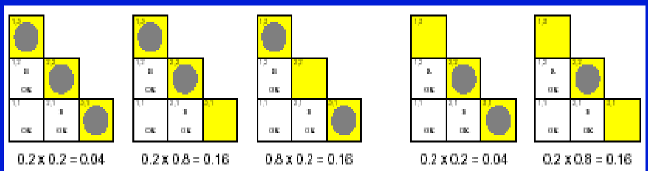
$= \underline{\alpha'}\, P(P_{1,3})\, \Sigma_{P2,2,P3,1}\, P(b \mid known, P_{1,3}, P_{2,2}, P_{3,1})\, P(P_{2,2})\, P(P_{3,1})$

$P(b \mid known, P_{1,3}, P_{2,2}, P_{3,1}) = 1$ when the $P_{2,2}, P_{3,1}$ is consistent with the breeze observations,
$= 0$ otherwise.

43

Continued…

$P(P_{1,3} \mid known, b) = \underline{\alpha'}\, P(P_{1,3})\, \Sigma_{frontier}\, P(b \mid known, P_{1,3}, P_{2,2}, P_{3,1})\, P(P_{2,2})\, P(P_{3,1})$



$P(b \mid known, P_{1,3}, P_{2,2}, P_{3,1}) = 1 \text{ or } 0$

Consistent models for the frontier variable $P_{2,2}$ and $P_{3,1}$, showing $P(frontier)$ for each model:

(a) 3 models with $P_{1,3}=true$ (i.e. $p_{1,3}$) showing 2 or 3 pits

(b) 2 models with $P_{1,3}=false$ (i.e. $\neg p_{1,3}$) showing 1 or 2 pits.

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

$= \alpha' < 0.2 \cdot 1 \cdot (0.2 \cdot 0.2 + 0.2 \cdot 0.8 + 0.8 \cdot 0.2), 0.8 \cdot (0.2 \cdot 0.2 + 0.2 \cdot 0.8) >$

$= \alpha' < .072, .160 > \approx < 0.31, 0.69 >$

Similarly,

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

$P(P_{3,1} \mid known, b)$? $\approx < 0.31, 0.69 >$

44

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

45