Part 7: Representing Uncertainty – Some Basic Concepts of Probability Theory

Univ.-Prof. Dr. Gerhard Widmer
Department of Computational Perception
Johannes Kepler University Linz

gerhard.widmer@jku.at
http://www.cp.jku.at/people/widmer

Overview

Motivation: Uncertainty

Basics of probability theory and probability distributions

Probabilistic inference using full joint probability distributions: marginalisation, normalisation, and inference by enumeration

Independence and Conditional Independence

An example of probabilistic reasoning: The Wumpus World Revisited

Limitations of Reasoning in Pure Logic

Reasoning in Logic:

- Propositions are either true, false, or unknown
- Given all the relevant facts about its environment and a world that is *deterministic,* a logical agent can derive plans that are guaranteed to work

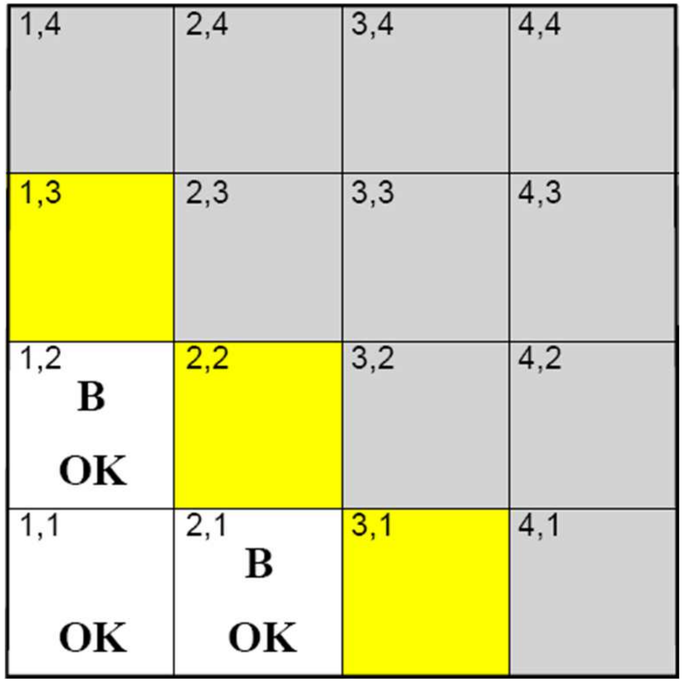
Problem:

- Agents almost never have access to the whole truth about their environment
- Example: Wumpus agent
  - can get only local information
  - most of the world is not immediately observable
  - may not be able to decide which of two adjacent squares contains a pit
- The agent's world may be *non-deterministic* (things don't always work ..)

➔ **A strict logic fails in such a case**

➔ **Need to deal with uncertainty**

Uncertainty in the Wumpus World



Two *breezes* have been observed (in [1,2] and [2,1])
– where are the *pits* ? – in [1,3], [2,2], and/or [3,1] ?

- Pure logical inference cannot determine which square is safe
  ➔ **logical agent** would have to choose randomly
- **Probabilistic agent** can do better: compute probabilities and choose room that is least likely to contain a pit (➔ risk minimisation)

Representing Uncertain Knowledge

Need to be able to express three things:

- Statements about the world
- Degree of certainty (belief) of each statement
- How degrees of certainty of some facts depend on certainty of other facts (e.g., observations)

In this class:

- Statements about the world: sentences in a kind of extended propositional logic (based on **random variables**)
- Degree of certainty/belief: **probabilities**
- Dependencies between facts and other facts (e.g., observations): **conditional probability statements**

Modelling Uncertainty via Probability Theory

Probability theory assigns to each sentence in the knowledge base a numerical degree of belief between 0 and 1

Example:

$$P(\,cavity) = 0.2$$

"For a random person walking into a dentist's office, there is a 20% chance that the person has a cavity (if nothing else is known about the person)"

Degree of belief may change as new **evidence (observations)** is collected:

$$P(\,cavity \mid toothache) = 0.7$$

*"There is a 70% chance that a patient has a cavity **if** she has a toothache"*

$$P(\,cavity \mid toothache, catch) = 0.95$$

*"If the patient has a toothache **and** the dentist's steel probe catches in the tooth, then the probability of a cavity is 95%"*

Note:

The fact itself is either true or false (the patient has a cavity or not)

➔ "degree of belief" is different from "degree of truth" (=> fuzzy logic)

A Short Digression: Interpretations of Probability

Frequentist position (“classical” probability theory):

- Probability numbers come only from experiments
- $P(heads)=0.5$ means that if I tossed a coin an infinite number of times, then it would come up heads in 50% of the cases

Objectivist view:

- Probabilities are real aspects of the universe (tendencies of objects to behave in certain ways)
- The fact that a fair coin comes up heads with a probability of 0.5 is a property of the coin itself

Subjectivist (Bayesian) view:

in this class

- Probabilities are simply a way of characterising an agent’s beliefs
- Have no external physical significance
- “I believe the probability that our team will win tonight is 80%”
  - ➔ either it will, or will not
  - ➔ this is a statement about the agent’s current belief about a singular event, not about something that can be repeated an infinite number of times …

Probabilistic Knowledge Representation: Syntax (1)

Basic element: Random variable:

- Refers to some specific aspect of the world
- Takes certain values with certain probabilities
- Example: $Cavity$ refers to whether my lower left wisdom tooth has a cavity
- Notation: Random variables start with upper-case letter ($Cavity$)

Each random variable has a domain (set of possible values):

- Example: the domain of $Cavity$ is $\{true, false\}$

Types of random variables:

- **Boolean random variables:** domain $\{true, false\}$
  Notation: instead of $Cavity = true$ / $false$, we will often write $cavity$ / $\neg cavity$

- **Discrete random variables:** countable domain of discrete values
  Example: $Weather \in \{sunny, rainy, cloudy, snow\}$
  Notation: instead of $Weather = snow$, we will often write $snow$

- **Continuous random variables:** domain is (subinterval of) real numbers.
  Propositions can involve equality ($T = 0.42$) or inequalities ($T \leq 0.42$)

In this class: will not deal with continuous variables (spare you a lot of integrals ☺)

Probabilistic Knowledge Representation: Syntax (2)

Atomic sentence:

- Statement about the specific value of a random variable
- Examples: $Cavity = true$ (or simply $cavity$); $Weather = sunny$ (or simply $sunny$)

➔ essentially corresponds to a proposition symbol (an elementary statement) in propositional logic

Complex sentence:

- Logical combination of propositions
- Example: $Cavity = true \wedge Toothache = false$ or simply $cavity \wedge \neg toothache$

States of the World: Events

Atomic event ("atomares Ereignis"):

- is a ***complete specification*** of the state of the world
- is an assignment of specific values to ***all*** the variables of which the agent's world consists
- Example: if my world only consists of the Boolean variables $Cavity$ and $Toothache$, there are four distinct atomic events:

$$cavity \wedge toothache$$
$$\neg cavity \wedge toothache$$
$$cavity \wedge \neg toothache$$
$$\neg cavity \wedge \neg toothache$$

≈ models in propositional logic

Event ("Ereignis"):

- is an assignment of specific values to ***some of*** the variables
- equivalent to the ***set of*** atomic events consistent with this assignment
- Example: $Cavity = true$ is equivalent to the two atomic events

$$cavity \wedge toothache$$
$$cavity \wedge \neg toothache$$

Prior Probability

The unconditional (prior) probability associated with a sentence a is the degree of belief attributed to it if no other information is available

Notation: $P(Cavity = true) = 0.2$ or $P(cavity) = 0.2$

Notation for probabilities of all possible values of a random variable:

$\mathbf{P}(Weather) = [0.7, 0.2, 0.08, 0.02]$ (note: sums to 1)

is shorthand for

$$P(Weather = sunny) = 0.7$$
$$P(Weather = rain) = 0.2$$
$$P(Weather = cloudy) = 0.08$$
$$P(Weather = snow) = 0.02$$

➔ This is called the **probability distribution** of random variable *Weather*

Prior Probability Distributions

Joint Probability Distribution:

- $\mathbf{P}(X_1, X_2, \ldots)$ denotes probabilities over **all combinations of values** of a ***set of* random variables**
- Example: $\mathbf{P}(\, \mathit{Toothache}, \mathit{Cavity})$ can be represented as a 2 x 2 table:

	toothache	¬ *toothache*
cavity	0.12	0.08
¬ *cavity*	0.08	0.72

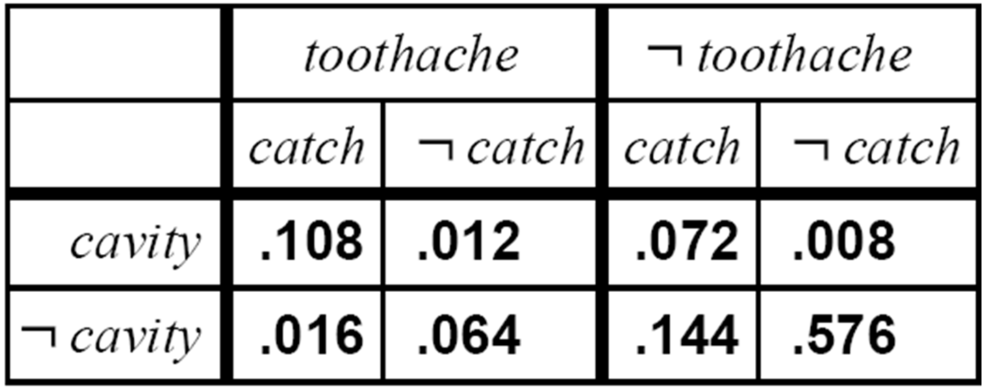
Prior Probability Distributions

Full Joint Probability Distribution:

- joint probability distribution over ***all*** random variables in the agent’s world
- Example: if the world consists of the three variables *Cavity*, *Catch*, and *Toothache*, the full joint distribution is

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

  and can be represented as a 2 x 2 x 2 table (i.e., 8 entries)



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

Full joint PD specifies the probability of every atomic event.
Will see:

- is a complete specification of the agent’s knowledge about the world
- any probabilistic query can be answered from the full joint distribution

Conditional Probability

Conditional Probability ("bedingte Wahrscheinlichkeit") specifies the degree of belief in some proposition a, given some additional evidence (known facts):

Notation: $P(a \mid b)$, where a and b are any propositions
Read: "the probability of a, given that we have observed b
(i.e., given that we know that b is true)"

Example: $P(\mathit{cavity} \mid \mathit{toothache}) = 0.6$

Notation for conditional distributions: $\mathbf{P}(X \mid Y)$ stands for the set of probability statements $P(X{=}x_i \mid Y{=}y_j)$ for all possible i, j. For example:

$$\mathbf{P}(\mathit{Cavity} \mid \mathit{Toothache})$$

… 2x2 table, equivalent to

$$\begin{aligned}
&P(\mathit{Cavity} = \mathit{true} \mid \mathit{Toothache} = \mathit{true}) = \ldots \\
&P(\mathit{Cavity} = \mathit{false} \mid \mathit{Toothache} = \mathit{true}) = \ldots \\
&P(\mathit{Cavity} = \mathit{true} \mid \mathit{Toothache} = \mathit{false}) = \ldots \\
&P(\mathit{Cavity} = \mathit{false} \mid \mathit{Toothache} = \mathit{false}) = \ldots
\end{aligned}$$

Conditional Probability

Definition of conditional probability:

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

Can be rewritten as: $P(a \wedge b) = P(a \mid b)P(b)$

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

("Product rule")

For distributions: $\mathbf{P}(X, Y) = \mathbf{P}(X \mid Y)\mathbf{P}(Y)$

(meaning: holds for all possible value combinations of X and Y)

Generalisation for conjunction of n variables: "Chain rule":

$$\mathbf{P}(X_1, X_2, \ldots, X_n) = \mathbf{P}(X_1 \mid X_2, \ldots, X_n)\mathbf{P}(X_2 \mid X_3, \ldots, X_n) \cdots \mathbf{P}(X_n)$$

Note: Due to commutativity of conjunction, this can be decomposed in any order

Background: The Axioms of Probability

Kolmogorov’s axioms (one version):

1. All probabilities are between 0 and 1:

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

2. Propositions that are necessarily true (i.e., valid) have probability 1, those that are necessarily false (i.e., unsatisfiable) have prob. 0:

$$P(true) = 1; \quad P(false) = 0$$

3. Probabilities of logical connections:

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

This is the complete set of axioms from which probability theory can be built …

Andrey Nikolaevich Kolmogorov (1903 – 1987)

Kolmogorov was one of the broadest of this century's mathematicians. He laid the mathematical foundations of probability theory and the algorithmic theory of randomness and made crucial contributions to the foundations of statistical mechanics, stochastic processes, information theory, fluid mechanics, and nonlinear dynamics. Kolmogorov graduated from Moscow State University in 1925 and then became a professor there in 1931. In 1939 he was elected to the Soviet Academy of Sciences, receiving the Lenin Prize in 1965 and the Order of Lenin on seven separate occasions.

His work on reformulating probability started with a 1933 paper in which he built up probability theory in a rigorous way from fundamental axioms, similar to Euclid's treatment of geometry. Kolmogorov went on to study the motion of the planets and turbulent fluid flows, later publishing two papers in 1941 on turbulence that even today are of fundamental importance.

In 1954 he developed his work on dynamical systems in relation to planetary motion, thus demonstrating the vital role of probability theory in physics and re-opening the study of apparent randomness in deterministic systems, much along the lines originally conceived by Henri Poincare.

In 1965 he introduced the algorithmic theory of randomness via a measure of complexity, now referred to Kolmogorov Complexity. According to Kolmogorov, the complexity of an object is the length of the shortest computer program that can reproduce the object. Random objects, in his view, were their own shortest description. Whereas, periodic sequences have low Kolmogorov complexity, given by the length of the smallest repeating "template" sequence they contain. Kolmogorov's notion of complexity is a measure of randomness, one that is closely related to Claude Shannon's entropy rate of an information source.

Kolmogorov had many interests outside mathematics research, notable examples being the quantitative analysis of structure in the poetry of the Russian author Pushkin, studies of agrarian development in 16th and 17th century Novgorod, and mathematics education.

Probabilistic Inference Using Full Joint Distributions

Assumption:

- The agent knows (or estimates) the full joint probability distribution over the variables in its world
- In other words: it has a probability estimate **for each possible state of the world** (atomic event)
  ➔ the full joint PD is the agent's "knowledge base"

Task: "Probabilistic Inference"

- Compute the probability of a proposition, *given some observations* (generalisation of *logical* inference, where we wish to determine the *truth* of a proposition, given some observations)

In the following:

- We will show that ***any probabilistic query can be answered from the full joint distribution*** (through "Inference by Enumeration")

Probabilistic Inference Using Full Joint Distributions: Inference by Enumeration (“Aufzählen”)

Example 1: a simple two-variable world:

- World defined by two random variables: *Toothache* and *Cavity*
- Full joint probability distribution (2 x 2 table):

	toothache	¬ *toothache*
cavity	0.12	0.08
¬ *cavity*	0.08	0.72

Note:

1. The probabilities in the full joint PD add up to 1.0
2. *The probability of any proposition is equal to the sum of the probabilities of all the atomic events in which the proposition is true*
   (see examples on following pages for intuitive explanation)

Probabilistic Inference Using Full Joint Distributions: Inference by Enumeration

	toothache	¬ *toothache*	
cavity	0.12	0.08	*0.2*
¬ *cavity*	0.08	0.72	
	0.2		

"marginal probabilities"

$$P(cavity) = \quad P(cavity \wedge (toothache \vee \neg toothache)) = \boxed{0.12 + 0.08} = 0.2$$

$$P(toothache) = \boxed{0.12 + 0.08} = 0.2$$

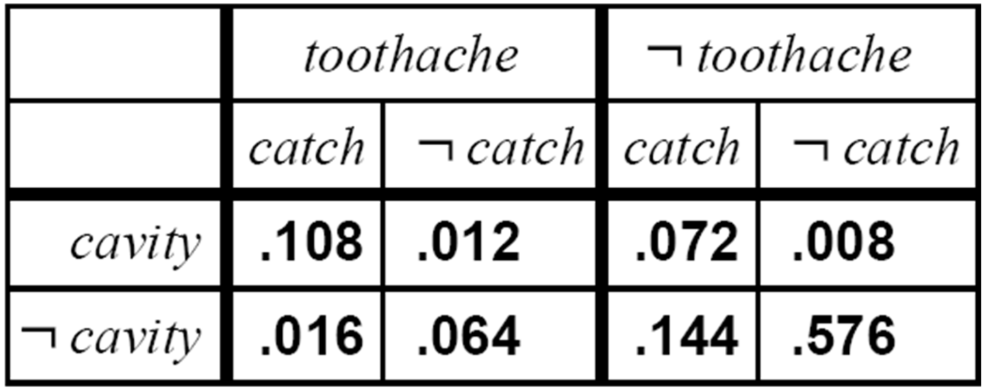
➔ **"Marginalisation"**

= computing the unconditional probability over a single variable ("marginal probability", "Randwahrscheinlichkeit")

Probabilistic Inference Using Full Joint Distributions: Inference by Enumeration

Example 2: a three-variable world:

- World defined by three random variables: *Toothache, Cavity, Catch*
- Full joint probability distribution (2 x 2 x 2 table):



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

Probabilistic Inference Using Full Joint Distributions: Inference by Enumeration

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

0.2

Summing over all value combinations of variables that are not fixed in the query (➔ "enumeration")

$$P(cavity) =$$

$$P(cavity \wedge toothache \wedge catch) + P(cavity \wedge toothache \wedge \neg catch) +$$

$$+ P(cavity \wedge \neg toothache \wedge catch) + P(cavity \wedge \neg toothache \wedge \neg catch) =$$

$$= 0.108 + 0.012 + 0.072 + 0.008 = 0.2$$

Probabilistic Inference Using Full Joint Distributions: Inference by Enumeration

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

0.2

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

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

➔ We can compute the probability of *any* (simple or complex) proposition by summing over all Full Joint PD entries (= atomic event probabilities) in which the proposition is true.

Probabilistic Inference Using Full Joint Distributions: Inference by Enumeration

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

Computation of *conditional* probabilities and distributions:

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

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

Summing over all variables not fixed in the query ("enumeration")

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

$$= \boxed{(.108 + .012)} \,/\, \boxed{(.108 + .012 + .016 + .064)} = 0.6$$

$$P(\neg cavity \mid toothache) = \boxed{(.016 + .064)} \,/\, \boxed{(.108 + .012 + .016 + .064)} = 0.4$$

$$\overline{\underline{\underline{1.0}}}$$

Normalisation

$$P(cavity \mid toothache) = \frac{P(cavity \wedge toothache)}{P(toothache)} = (.108 + .012) / \boxed{(.108 + .012 + .016 + .064)} = 0.6$$

$$P(\neg cavity \mid toothache) = \frac{P(\neg cavity \wedge toothache)}{P(toothache)} = (.016 + .064)/\boxed{(.108 + .012 + .016 + .064)} = 0.4$$

$$\underline{\underline{\quad 1.0}}$$

Note:

- Term $1/P(toothache) = 1 / (.108 + .012 + .016 + .064)$ remains constant for all queries about the different values of $Cavity$
- Can be regarded as a **normalisation constant α** for the **conditional distribution $\mathbf{P}(Cavity \mid toothache)$** (to ensure that it adds up to 1)

➔ Need not be computed for each query; can be computed at the end (such that the distribution adds up to one)

➔ Don't bother computing this any more;
simpler notation (see next page) from now on …

Normalisation

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

Note:
"," means "and"
(joint probability)

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

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

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

$$= \alpha\ \ [\mathbf{P}(Cavity,\ toothache,\ catch) + \mathbf{P}(Cavity,\ toothache,\ \neg catch)] =$$

$$= \alpha\left(\begin{bmatrix}.108\\ .016\end{bmatrix}+\begin{bmatrix}.012\\ .064\end{bmatrix}\right) = \alpha\begin{bmatrix}.12\\ .08\end{bmatrix} = \begin{bmatrix}\mathbf{.6}\\ \mathbf{.4}\end{bmatrix} \quad \begin{matrix}\leftarrow Cavity = true\\ \leftarrow Cavity = false\end{matrix}$$

where the normalisation constant α is computed afterwards such that the individual probabilities sum up to *1.0:*

$$\alpha = \frac{1}{.12 + .08} = 5.0$$

Important slide:
Make sure you understand this!

General Procedure for Inference by Enumeration from the Full Joint PD

The General Probabilistic Inference Problem:

We want the **conditional joint probability distribution** of some set of query variables $\mathbf{X}$, given **specific values** $\mathbf{e}$ for some evidence variables $\mathbf{E}$: $\mathbf{P}(\mathbf{X} \mid \mathbf{E}=\mathbf{e})$

$\mathbf{X}$ … the set of query variables ($\{Cavity\}$, in our previous example)

$\mathbf{E}$ … the set of evidence variables ($\{Toothache\}$, in our example)

$\mathbf{e}$ … the observed values for them ($\{true\}$)

$\mathbf{Y}$ … the remaining unobserved variables in the world ($\{Catch\}$)

Algorithm for computing $\mathbf{P}(\mathbf{X} \mid \mathbf{E}=\mathbf{e})$:

$$\mathbf{P}(\mathbf{X} \mid \mathbf{E} = \mathbf{e}) = \alpha\, \mathbf{P}(\mathbf{X}, \mathbf{E} = \mathbf{e}) = \alpha \sum_{\mathbf{y}} \overbrace{\mathbf{P}(\mathbf{X}, \mathbf{E} = \mathbf{e}, \mathbf{Y} = \mathbf{y})}^{\textit{entries from the Full Joint PD}}$$

where the sum is over all possible $\mathbf{y}$s
(i.e., all possible combinations of values of the hidden variables $\mathbf{Y}$)

General Procedure for Inference by Enumeration from the Full Joint PD

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

Properties of this algorithm:

- Complete algorithm for answering probabilistic queries for discrete variables
- **But:** for a world described by n Boolean variables, the Full Joint PD table has size $O(2^n)$!

➔ Completely impractical for realistic problems (table too large for memory; and how could the agent ever learn (estimate) 2^n probabilities? or how could a human domain expert ever specify 2^n probabilities?)

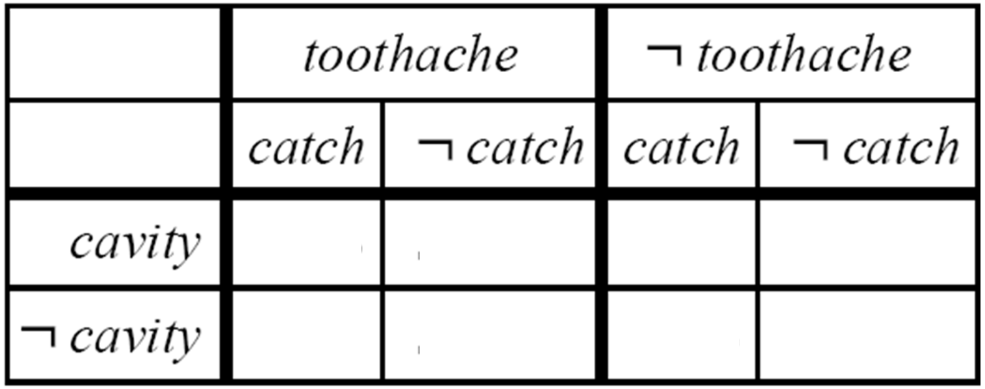
➔ Full joint distribution in tabular form is not a practical tool for building reasoning systems

➔ Need to somehow reduce the complexity ➔ see next slides …

Independence

Example:

- Add fourth random variable $Weather \in \{sunny, rain, cloudy, snow\}$ to our world
- Full joint distribution $\mathbf{P}(Toothache, Catch, Cavity, Weather)$ has 2 x 2 x 2 x 4 = 32 entries
- Table contains four “editions” of 3-variable tables like this one, one for each value of $Weather$:



	toothache		*¬ toothache*	
	catch	*¬ catch*	*catch*	*¬ catch*
cavity				
¬ cavity				

Question:

- What is the relationship between these four “sub-tables”, and their relationship to the original three-variable table?
- For example: how are $P(toothache, catch, cavity, Weather{=}cloudy)$ and $P(toothache, catch, cavity)$ related?

Independence

Question:

How are $P(\mathit{toothache, catch, cavity, cloudy})$ and $P(\mathit{toothache, catch, cavity})$ related?

Answer (via product rule):

$$P(\mathit{toothache, catch, cavity, cloudy}) = P(\mathit{cloudy} \mid \mathit{toothache, catch, cavity}) \, P(\mathit{toothache, catch, cavity})$$

But:

Unrealistic to believe that tooth problems influence the weather

➔ Can assume that

$$P(\mathit{cloudy} \mid \mathit{toothache, catch, cavity}) = P(\mathit{cloudy})$$

From this follows:

$$P(\mathit{toothache, catch, cavity, cloudy}) = P(\mathit{cloudy}) \, P(\mathit{toothache, catch, cavity})$$

➔ Can decompose joint probability over 4 variables into product of simpler probabilities (over 1 and 3 variables, respectively)

Independence

More generally:

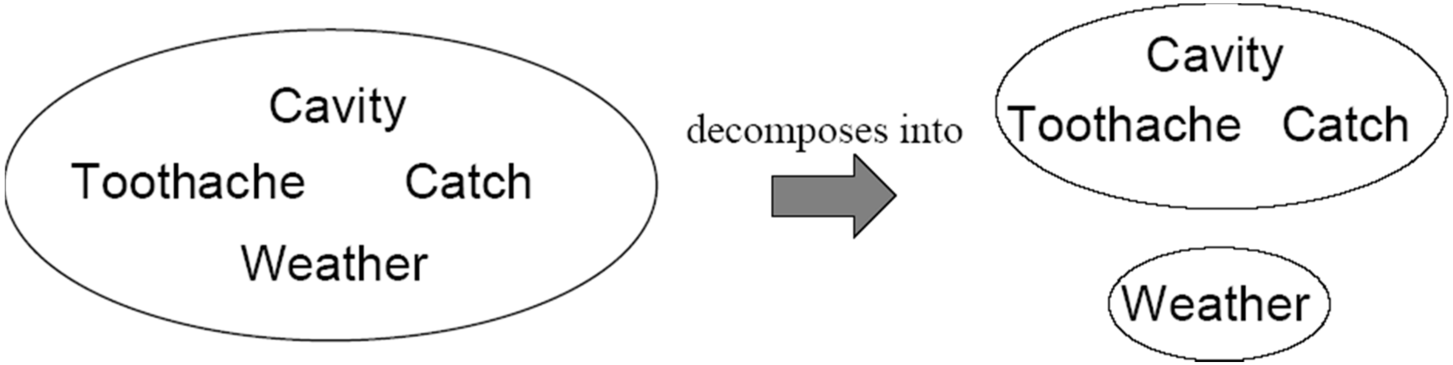
All values of variable *Weather* are independent of tooth problems

➔ Can generally write:

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

➔ 32-element table for four variables can be reconstructed from one 8-element table and one four-element table

➔ Need only 12 probabilities instead of 32 to represent the world:



Independence

Definition:

Two propositions a and b are **independent** if $P(a \mid b) = P(a)$ or (equivalently) $P(b \mid a) = P(b)$ or (equivalently) $P(a \wedge b) = P(a)P(b)$.

(Exercise: show that these 3 statements are all equivalent)

Generalisation to random variables (i.e., to probability distributions):

Two variables X and Y are **independent** if $\mathbf{P}(X \mid Y) = \mathbf{P}(X)$ or (equivalently) $\mathbf{P}(Y \mid X) = \mathbf{P}(Y)$ or (equivalently) $\mathbf{P}(X, Y) = \mathbf{P}(X)\mathbf{P}(Y)$.

➔ Independence between variables permits us to reduce complex joint PD tables to combinations of simpler ones ➔ ***reduction of complexity***
(e.g., from 32 to 12 table entries in previous example,
or from $\mathrm{O}(2^n)$ to $O(n)$ in the case of tossing n independent coins !)

Unfortunately: absolute independence is rare in a given world (e.g., dentistry) …

Important slide:
Make sure you understand this!

Conditional Independence

Consider the world of dentistry $\mathbf{P}(\, Toothache, Catch, Cavity)$:

Are *Toothache* and *Catch* independent?

+: none is a direct cause of the other

But: If the patient has a toothache, the tooth probably has a cavity, and that will probably cause the steel probe to catch …

➔ Knowing that the patient has a *toothache* increases the probability of *catch* (or my "degree of belief" in *catch*) – and vice versa

➔ $P(\, catch \mid toothache) \neq P(\, catch)$

➔ *Toothache* and *Catch* are **not independent !**

But:

Both Toothache and Catch are caused by a Cavity, but neither has a direct effect on the other

➔ If I *know* the patient has a cavity, the *additional* information that she has a toothache does not change my degree of belief that the probe will catch:

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

(and also: $P(\, catch \mid toothache, \neg cavity) = P(\, catch \mid \neg cavity)$)

Conditional Independence

Definition:

Two variables X and Y are **conditionally independent, given a third variable** Z, if $\mathbf{P}(X \mid Y, Z) = \mathbf{P}(X \mid Z)$ or (equivalently) $\mathbf{P}(Y \mid X, Z) = \mathbf{P}(Y \mid Z)$ or (equivalently)

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

In intuitive terms:

X and Y are conditionally independent, given Z, if Y gives me no additional Information about X, if I already know the value of Z.
Or: If I know Z, then learning about the value of Y will not make me change my mind („degree of belief") about X .

Typical case of conditional independence:

Z is a direct cause of both X and Y
(e.g., in our dentistry world, *cavity* is the direct cause of a *toothache*, and also of the fact that the dentist's steel probe *catch*es in the tooth)

Conditional Independence

Example:

Catch is conditionally independent of *Toothache* given *Cavity*:

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

Equivalent statements:

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

Again: Can use conditional independence to reduce a full joint distribution to a combination of simpler joint (conditional) distributions:

$$\begin{aligned}
&\mathbf{P}(\,Toothache, Catch, Cavity) = && \textit{[chain rule]}\\
&\quad \mathbf{P}(\,Toothache \mid Catch, Cavity)\, \mathbf{P}(\,Catch, Cavity) = && \textit{[chain rule]}\\
&\quad \mathbf{P}(\,Toothache \mid Catch, Cavity)\, \mathbf{P}(\,Catch \mid Cavity)\, \mathbf{P}(\,Cavity) = && \textit{[def. of cond. indep.]}\\
&\quad \mathbf{P}(\,Toothache \mid Cavity)\, \mathbf{P}(\,Catch \mid Cavity)\, \mathbf{P}(\,Cavity)
\end{aligned}$$

➔ Can replace 2 x 2 x 2 probability table with 2 x 2 plus 2 x 2 plus 2 x 1

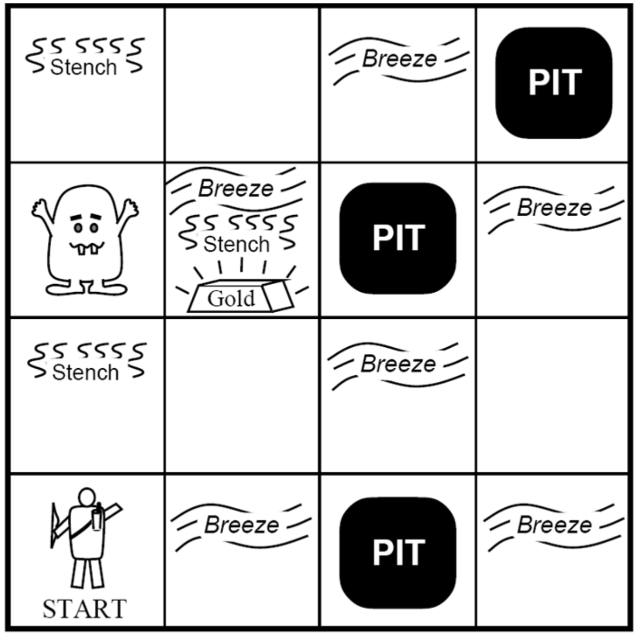
The Importance of Conditional Independence

In many practical cases, the use of conditional independence reduces the **size of the representation** of the full joint distribution from **exponential** in n (e.g., 2^n) to **linear** in n !

Conditional independence is the most fundamental and useful knowledge about uncertain environments

➔ Basis for Bayesian Belief Networks

Uncertainty in the Wumpus World

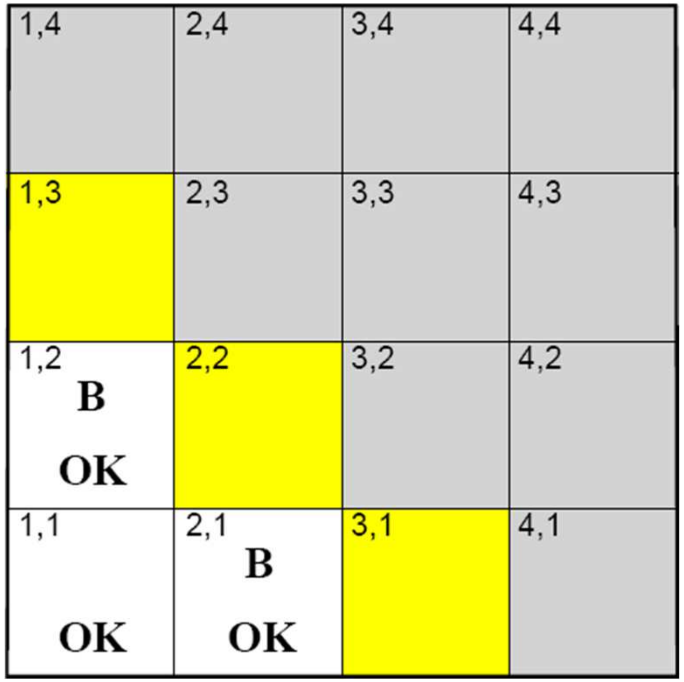


Sources of uncertainty:

The agent’s sensors give only

- **local information** (e.g., can only determine *Breezes* and *Stench* in already visited squares)
- **partial information** (e.g., a breeze leaves several options as to where the corresponding pit may be)

Uncertainty in the 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

Example:

Two *breezes* have been observed (in [1,2] and [2,1])
– where are the *pits* ? – in [1,3], [2,2], and/or [3,1] ?

- Pure logical inference cannot conclude that any of these three rooms is safe
➔ **logical agent** would have to choose randomly
- **Probabilistic agent** can do better …

This part not treated in class –
ignore for exam!

A Probabilistic Model of the Wumpus World

Variables in the knowledge base:

- model the world with Boolean ***random*** variables P_{ij} and B_{ij}
- $P_{ij} = true$ 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 (for simplicity)

The full probability model of this world = full joint distribution:

$$\mathbf{P}(P_{11}, P_{12}, \ldots, P_{44}, B_{11}, B_{12}, B_{21}) =$$

[use product rule to get $P(Effect|Cause)$]

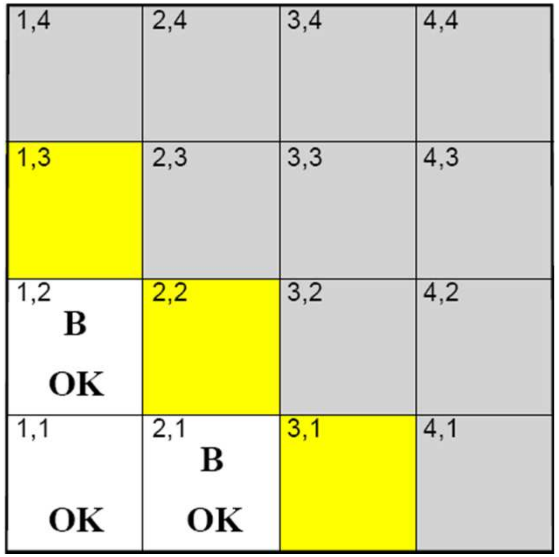
$$\underbrace{\mathbf{P}(B_{11}, B_{12}, B_{21} \mid P_{11}, P_{12}, \ldots, P_{44})}\,\underbrace{\mathbf{P}(P_{11}, P_{12}, \ldots, P_{44})}$$

- Conditional prob. of a breeze config., given a pit configuration;
- Conceptually: table with 2^3x2^{16} entries;
- Entry is 1 if breezes are next to pits, 0 otherwise

- Prior probability of any pit configuration;
- Conceptually: table with 2^{16} entries
- Assuming each square contains a pit with probability 0.2 (given information), entry is 0.2^n x 0.8^{16-n} if there are n pits (i.e., if n of the the P_{ij} are true)

This part not treated in class – ignore for exam!

Uncertainty in the Wumpus World



Evidence:

- observed breezes and non-breezes: $b \;:=\; \neg b_{11} \wedge b_{12} \wedge b_{21}$
- known absence of pits in visited squares: $known \;:=\; \neg p_{11} \wedge \neg p_{12} \wedge \neg p_{21}$

Queries of interest relative to given evidence:

- $\mathbf{P}(P_{13} \mid known, b) = ?$
- $\mathbf{P}(P_{22} \mid known, b) = ?$
- $\mathbf{P}(P_{31} \mid known, b) = ?$

This part not treated in class – ignore for exam!

Uncertainty in the Wumpus World

$$\begin{aligned} b &:= \neg b_{11} \wedge b_{12} \wedge b_{21} \\ known &:= \neg p_{11} \wedge \neg p_{12} \wedge \neg p_{21} \end{aligned}$$

$$\mathbf{P}(P_{13} \mid known, b) = ?$$

Remember:

this kind of problem can be solved by enumeration over full joint PD

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

Let as define *Unknown* as the set of P_{ij} other than P_{13} and *Known*

$$\rightarrow \quad \mathbf{P}(P_{13} \mid known, b) = \alpha \sum_{unknown} \mathbf{P}(P_{13}, known, b, unknown)$$

But:

P(*Unknown*) contains 2^{12} entries → 4096 summation terms → can't we do better?

This part not treated in class – ignore for exam!

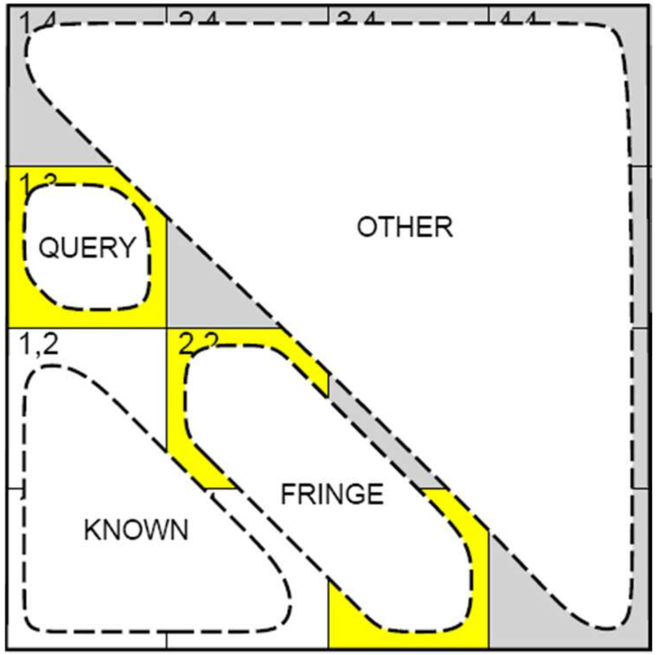
Conditional Independence in the Wumpus World

Basic insight:

Breezes in 3 known squares (B) are **conditionally independent** of pits in other hidden squares ($Other$), given their direct neighbour squares ($Query$ (i.e., P_{13}) + $Fringe$)

Define/decompose: $Unknown = Fringe \cup Other$

➔ $\boxed{\mathbf{P}(B \mid P_{13}, Known, Unknown) = \mathbf{P}(B \mid P_{13}, Known, Fringe)}$



Using Conditional Independence in the Wumpus World

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

→ *Unknown* completely eliminated!

prior prob. of p_{13} / $\neg p_{13}$ $= \{0.2, 0.8\}$

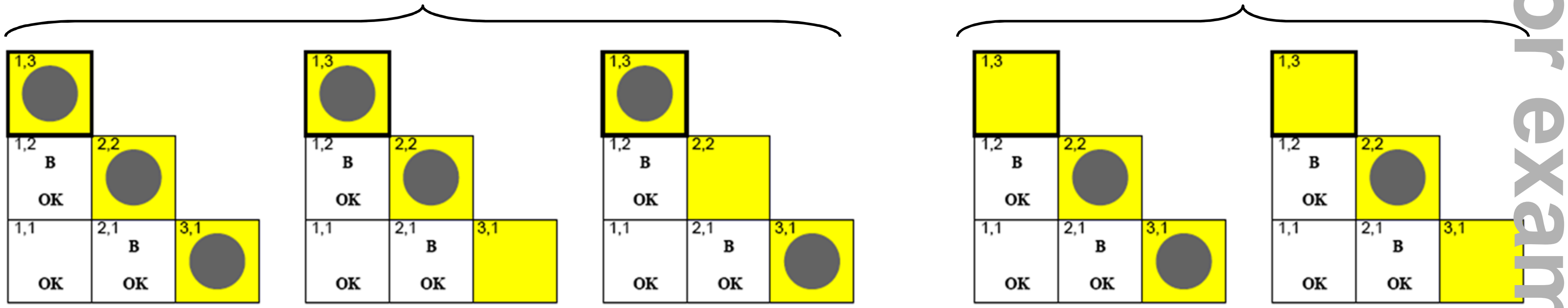
0 or 1 (depending on whether *fringe* is consistent with breezes b and p_{13} / $\neg p_{13}$)

This part not treated in class – ignore for exam!

Solving the Query …

3 (out of four possible) *fringe* configurations consistent with b and p_{13}

2 *fringe* configurations consistent with b and $\neg p_{13}$



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

$$\uparrow$$

$P(fringe)$

$$\mathbf{P}(P_{13} \mid known, b) = \alpha'[0.2(0.04+0.16+0.16), 0.8(0.04+0.16)] \approx [0.31, 0.69]$$

$P(true)$

$P(false)$

$$\mathbf{P}(P_{22} \mid known, b) \approx [0.86, 0.14]$$

$$\mathbf{P}(P_{31} \mid known, b) \approx [0.31, 0.69]$$

➔ **Agent: Avoid square [2,2] !**

This part not treated in class – ignore for exam!

Summary

- **Probability theory** is a rigorous formalism for **uncertain knowledge**
- The **joint probability distribution** specifies the probability of every atomic event
- Queries can be answered by summing over atomic events **(inference by enumeration)**
- For non-trivial domains, we must find a way to **reduce the size of the joint distribution** (which is exponential)
- **Independence** and **conditional independence** make that possible
- Next: **Bayesian networks:** systematic way of modeling conditional independence relations, and efficient algorithms for probabilistic reasoning