Uncertainty

School of EECS
Washington State University
Sometimes the truth or falsity of facts in the world is unknown

Sources of uncertainty

- Incompleteness of rules used by agent
- Incorrectness of rules used by agent
- Limited and ambiguous sensors
- Imperfection and noise of agent’s actions
- Unpredictable and dynamic nature of agent’s environment
- Approximate nature of internal models and algorithms used by agent
Environment Properties

- Fully observable vs. partially observable
- Deterministic vs. stochastic
- Episodic vs. sequential
- Static vs. dynamic
- Discrete vs. continuous
- Single agent vs. multiagent
Wumpus World

- $P(Pit_{1,3}) = ?$
- $P(Pit_{2,2}) = ?$
- $P(Pit_{3,1}) = ?$
Logic-based Approach

- Maintain set of all possible world states

Partially observable vacuum world (a)

Nondeterministic vacuum world (b)
Pure logic fails for three main reasons

Laziness
- Too much work to list all conditions needed to ensure an exceptionless rule

Theoretical ignorance
- Science has no complete theory for the domain

Practical ignorance
- Even if we know all the rules, we may be uncertain about them
- I.e., we only have a degree of belief in them
Our main tool for dealing with degrees of belief is probability theory

Probability summarizes uncertainty due to our laziness and ignorance

<table>
<thead>
<tr>
<th>Logic</th>
<th>World composed of facts that do or do not hold.</th>
<th>Each sentence is true or false or unknown.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Probability Theory</td>
<td>World composed of facts that do or do not hold.</td>
<td>Numerical degree of belief between 0 (for sentences that are certainly false) and 1 (certainly true)</td>
</tr>
</tbody>
</table>
Choose action A that maximizes expected utility

I.e., maximizes Prob(A) * Utility(A)

Prob(A) = probability A will succeed
Utility(A) = value to agent of A’s outcomes
A probability model associates a numeric probability $P(w)$ with each possible world $w$, where

$$0 \leq P(w) \leq 1 \quad \text{and} \quad \sum_{w} P(w) = 1$$
The probability $P(a)$ of a proposition ‘a’ is the sum of the probabilities of the worlds in which ‘a’ is true

$$P(a) = \sum_{w \in a} P(w)$$
What is the probability of the proposition that a two-dice roll totals 7?

Consider the 36 possible outcomes of rolling two dice, each with probability 1/36

- $(1, 1)$ (1, 2) (1, 3) (1, 4) (1, 5) (1, 6)
- $(2, 1)$ (2, 2) (2, 3) (2, 4) (2, 5) (2, 6)
- $(3, 1)$ (3, 2) (3, 3) (3, 4) (3, 5) (3, 6)
- $(4, 1)$ (4, 2) (4, 3) (4, 4) (4, 5) (4, 6)
- $(5, 1)$ (5, 2) (5, 3) (5, 4) (5, 5) (5, 6)
- $(6, 1)$ (6, 2) (6, 3) (6, 4) (6, 5) (6, 6)

$P(\text{Total}=7) = P(\text{Die}_1=1 \land \text{Die}_2=6) + P(\text{Die}_1=2 \land \text{Die}_2=5) + \ldots + P(\text{Die}_1=6 \land \text{Die}_2=1) = ?$
Probability

- An **unconditional** or **prior probability** is the degree of belief without any other information
  - E.g., \( P(\text{Total} = 7) = 1/6 \)
- A **conditional** or **posterior probability** is the degree of belief given some evidence
  - E.g., given that the first die is a 5, what is the probability that the total will be 7?
  - \( P(\text{Total} = 7 \mid \text{Die}_1 = 5) = ? \)
Conditional Probability

\[ P(a | b) = \frac{P(a \land b)}{P(b)} \quad \text{assumes} \quad P(b) > 0 \]

- Fraction of worlds in which ‘a’ and ‘b’ are true out of the worlds in which ‘b’ is true
- \( P(\text{Total}=7 \mid \text{Die}_1=5) = \frac{P(\text{Total}=7 \land \text{Die}_1=5)}{P(\text{Die}_1=5)} = \frac{1/36}{6/36} = 1/6 \)
- Product rule: \( P(a \land b) = P(a \mid b) \cdot P(b) \)
Probability Theory

- A variable in probability theory is called a **random variable**
  - E.g., Total, Die₁
- The **domain** of a random variable is the set of possible values it can take on
  - E.g., domain(Die₁) = {1,2,3,4,5,6}
- A **probability distribution** over a random variable gives the probability of each value in the random variable’s domain
  - E.g., P(Die₁=1)=1/6, ..., P(Die₁=6)=1/6
  - Or, P(Die₁)=⟨1/6,1/6,1/6,1/6,1/6,1/6⟩

Note: Boldfaced
Conditional probability distribution
- P(X | Y) gives the values of P(X=x_i | Y=y_j) for all possible i,j pairs

Joint probability distribution
- P(X,Y) denotes the probabilities of all possible combinations of X and Y
- E.g., P(Die_1,Die_2) gives values for P(Die_1=1,Die_2=1), P(Die_1=1,Die_2=2), ..., P(Die_1=6,Die_2=6)
- E.g., P(Die_1=5,Die_2=1) abbreviated P(5,1)

Full joint probability distribution
- Joint probability distribution over all random variables in the world
More Probability Axioms

- \( P(\neg a) = 1 - P(a) \)
- \( P(a \lor b) = P(a) + P(b) - P(a \land b) \)
Probabilistic Inference

- Approach #1: Just ask Spock…

“Errand of Mercy” (1967)
Given full joint probability distribution, can answer any question about probabilities

Example: Tooth World
- Random variables: Toothache, Cavity, Catch

<table>
<thead>
<tr>
<th></th>
<th>toothache</th>
<th>¬toothache</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>catch</td>
<td>¬catch</td>
</tr>
<tr>
<td>cavity</td>
<td>.108</td>
<td>.012</td>
</tr>
<tr>
<td>¬cavity</td>
<td>.016</td>
<td>.064</td>
</tr>
</tbody>
</table>
Probabilistic Inference

- Answer questions (perform probabilistic inference) by summing probabilities

- Examples
  - \( P(\text{cavity}) = 0.108 + 0.012 + 0.072 + 0.008 = 0.2 \)
  - \( P(\text{cavity} \land \text{toothache}) = 0.108 + 0.012 = 0.12 \)

<table>
<thead>
<tr>
<th></th>
<th>toothache</th>
<th>¬toothache</th>
</tr>
</thead>
<tbody>
<tr>
<td>catch</td>
<td>0.108</td>
<td>0.012</td>
</tr>
<tr>
<td>¬catch</td>
<td>0.072</td>
<td>0.008</td>
</tr>
<tr>
<td>cavity</td>
<td>0.016</td>
<td>0.064</td>
</tr>
<tr>
<td>¬cavity</td>
<td>0.144</td>
<td>0.576</td>
</tr>
</tbody>
</table>
Marginalization is the process of finding the probability distribution over a subset of variables $Y$:

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

For example:

$$P(\text{Cavity}) = \sum_{z \in \{\text{Catch,Toothache}\}} P(\text{Cavity}, z)$$

Conditioning:

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

$Z$ is all possible combinations of values

$z$ in {$(\text{catch, tooth}), (\text{catch, } \neg\text{tooth}), (\neg\text{catch, tooth}), (\neg\text{catch, } \neg\text{tooth})$}
Conditional probabilities

\[ P(\text{cavity}|\text{toothache}) = \frac{P(\text{cavity} \land \text{toothache})}{P(\text{toothache})} = \frac{0.108 + 0.016}{0.108 + 0.012 + 0.016 + 0.064} = 0.6 \]

\[ P(\neg\text{cavity}|\text{toothache}) = \frac{P(\neg\text{cavity} \land \text{toothache})}{P(\text{toothache})} = \frac{0.016 + 0.064}{0.108 + 0.012 + 0.016 + 0.064} = 0.4 \]

Note that denominator same in both (normalization constant)
Don’t have to know normalization constant to compute probability distribution

- For example
  - \( P(\text{cavity} \mid \text{toothache}) = \alpha P(\text{cavity} \land \text{toothache}) = \alpha 0.12 \)
  - \( P(\neg \text{cavity} \mid \text{toothache}) = \alpha P(\neg \text{cavity} \land \text{toothache}) = \alpha 0.08 \)
  - Since these have to sum to 1, \( \alpha = 1 / 0.2 = 5 \)
  - Also means \( P(\text{toothache}) = 1 / 5 = 0.2 \), but we didn’t need to know this
  - Also means we can determine which is more likely (cavity or \( \neg \text{cavity} \)) without knowing \( P(\text{toothache}) \)
Probabilistic Inference

- General rule
- Want to know the probability distribution over $X$ given observed variables $e$ (evidence) and unobserved variables $y$

$$P(X \mid e) = \alpha P(X, e) = \alpha \sum_y P(X, e, y)$$
Independence

- Full joint probability distributions are typically infeasible to write down
  - E.g., $2^n$ table entries for $n$ Boolean variables
- If we know some variables are independent of others, then we can decompose the full table into smaller tables
- If two variables $X$ and $Y$ are independent, then
  - $P(X,Y) = P(X)P(Y)$
  - $P(X|Y) = P(X)$
  - $P(Y|X) = P(Y)$
Independence

- For example, add Weather variable to Tooth World
  - domain(Weather) = \{sunny, cloudy, rain, snow\}
  - 8 entry table now a 32 entry table
- Assuming your teeth don’t affect the weather
  - I.e., Weather is independent of Toothache, Catch and Cavity
  - So, \( P(\text{toothache}, \text{catch}, \text{cavity}, \text{cloudy}) = P(\text{toothache}, \text{catch}, \text{cavity}) \cdot P(\text{cloudy}) \)
  - Full joint distribution described by two tables of 8 and 4 entries
Bayes Rule

- Foundational rule of probabilistic reasoning
- Example: Diagnosing cancer
- Given
  - 1% of population has cancer
  - Test has 20% false positive rate
  - Test has 10% false negative rate
  - Patient tests positive for cancer
  - What is the probability that patient has cancer?

Thomas Bayes (1701–1761)

Artificial Intelligence 26
Bayes Rule

Example: Diagnosing cancer

Variables
- Diagnosis $\in \{\text{cancer, healthy}\}$
- Test $\in \{\text{pos, neg}\}$

Given
- $P(\text{cancer}) = 0.01$, $P(\text{healthy}) = 0.99$
- $P(\text{neg}|\text{cancer}) = 0.1$, $P(\text{pos}|\text{cancer}) = 0.9$
- $P(\text{pos}|\text{healthy}) = 0.2$, $P(\text{neg}|\text{healthy}) = 0.8$

$P(\text{cancer}|\text{pos}) = ?$

$P(\text{cancer}|\text{pos}) = \frac{P(\text{cancer} \land \text{pos})}{P(\text{pos})} = ?$
Bayes Rule

\[ P(b \mid a) = \frac{P(a \mid b)P(b)}{P(a)} \]

\[ P(Y \mid X) = \frac{P(X \mid Y)P(Y)}{P(X)} \]

- Turns a diagnostic question into a causal one

\[ P(cause \mid effect) = \frac{P(effect \mid cause)P(cause)}{P(effect)} \]

- In general, given some evidence \( e \):

\[ P(Y \mid X, e) = \frac{P(X \mid Y, e)P(Y \mid e)}{P(X \mid e)} \]
Back to our example…

\[
P(cancer \mid pos) = \frac{P(pos \mid cancer)P(cancer)}{P(pos)}
\]

- We know \(P(pos \mid cancer)=0.9\) and \(P(cancer)=0.01\)
- Can compute \(P(pos)\) by marginalization
  - \(P(pos) = P(pos \land cancer) + P(pos \land healthy)\)
  - \(P(pos) = P(pos \mid cancer)P(cancer) + P(pos \mid healthy)P(healthy)\)
  - \(P(pos) = (0.90)(0.01) + (0.20)(0.99) = 0.207\)

\[
P(cancer \mid pos) = \frac{(0.9)(0.01)}{(0.207)} = 0.043
\]

Bayes Rule
Could also compute $P(\text{pos})$ via normalization

$P(\text{cancer}|\text{pos}) = \alpha P(\text{pos}|\text{cancer})P(\text{cancer}) = \alpha(0.9)(0.01) = \alpha(0.009)$

$P(\text{healthy}|\text{pos}) = \alpha P(\text{pos}|\text{healthy})P(\text{healthy}) = \alpha(0.2)(0.99) = \alpha(0.198)$

$\alpha = 1 / (0.009 + 0.198) = 4.83$

$P(\text{pos}) = 1 / \alpha = 0.207$

Many times, $P(\text{effect}|\text{cause})$ is easier to determine than $P(\text{cause}|\text{effect})$
Combining Evidence in Bayes Rule

- How to compute $P(a \mid b \land c \land \ldots)$?
  - E.g., $P(\text{Cavity} \mid \text{toothache} \land \text{catch}) = ?$
- Easy if we have full joint probability distribution
  - E.g., $P(\text{Cavity} \mid \text{toothache} \land \text{catch}) = \alpha \langle P(\text{cavity} \land \text{toothache} \land \text{catch}), P(\neg \text{cavity} \land \text{toothache} \land \text{catch}) \rangle = \alpha \langle 0.108, 0.016 \rangle = \langle 0.871, 0.129 \rangle$
- Or, using Bayes rule: $\alpha [P(b \land c \land \ldots \mid a) \cdot P(a)]$
  - E.g., $\alpha P(\text{toothache} \land \text{catch} \mid \text{Cavity}) \cdot P(\text{Cavity})$
  - Still need to know many probabilities
If b, c, ... are “caused” by ‘a’, but not by each other, then
- \( P(b \land c \land ... \mid a) = P(b \mid a) P(c \mid a) \ldots \)
- \( P(a \mid b \land c \land ...) = \alpha P(b \land c \land ... \mid a) P(a) \)
- \( P(a \mid b \land c \land ...) = \alpha P(b \mid a) P(c \mid a) \ldots P(a) \)

I.e., b, c, ... are independent given ‘a’

Two propositions (effects) are conditionally independent if they are independent given a third proposition (cause)

Now we only need to know the probabilities of individual effects given the cause \( P(b \mid a) \) and the prior probability of the cause \( P(a) \)
Combining Evidence in Bayes Rule

Example (assuming toothache and catch are conditionally independent given Cavity)
- $P(\text{toothache} \land \text{catch} \mid \text{Cavity}) = P(\text{toothache} \mid \text{Cavity}) \cdot P(\text{catch} \mid \text{Cavity})$

Using this in Bayes rule
- $P(\text{Cavity} \mid \text{toothache} \land \text{catch}) = \alpha P(\text{toothache} \land \text{catch} \mid \text{Cavity}) \cdot P(\text{Cavity})$
- $P(\text{Cavity} \mid \text{toothache} \land \text{catch}) = \alpha P(\text{toothache} \mid \text{Cavity}) \cdot P(\text{catch} \mid \text{Cavity}) \cdot P(\text{Cavity})$
Combining Evidence in Bayes Rule

In general, if random variables $X$ and $Y$ are conditionally independent given $Z$

- $P(X,Y \mid Z) = P(X \mid Z) \cdot P(Y \mid Z)$
- $P(X \mid Y,Z) = P(X \mid Z)$
- $P(Y \mid X,Z) = P(Y \mid Z)$
Naïve Bayes

- Want \( P(\text{Cause} \mid \text{Effect}_1, \text{Effect}_2, \ldots, \text{Effect}_n) \)
- Assume \( \text{Effect}_1, \text{Effect}_2, \ldots, \text{Effect}_n \) are conditionally independent given \( \text{Cause} \)
- Applying Bayes Rule
  - \( P(\text{Cause} \mid \text{Effect}_1, \text{Effect}_2, \ldots, \text{Effect}_n) = \alpha P(\text{Effect}_1, \text{Effect}_2, \ldots, \text{Effect}_n \mid \text{Cause}) P(\text{Cause}) = \alpha P(\text{Cause}) \prod_i P(\text{Effect}_i \mid \text{Cause}) \)
Using Bayes Rule

- Be careful
  - Conditional independence rarely absolutely true
  - Probabilities estimated from data

- Bayes Rule thrown out of court
Wumpus World Revisited

- $P(\text{Pit}_{1,3}) = ?$
- $P(\text{Pit}_{2,2}) = ?$
- $P(\text{Pit}_{3,1}) = ?$

<table>
<thead>
<tr>
<th></th>
<th>1,1</th>
<th>1,2</th>
<th>1,3</th>
<th>1,4</th>
</tr>
</thead>
<tbody>
<tr>
<td>2,1</td>
<td>B</td>
<td>OK</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2,2</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3,1</td>
<td>B</td>
<td>OK</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3,2</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4,1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4,2</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4,3</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4,4</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
“query” = Pit$_{1,3}$
“frontier” = \{Pit$_{2,2}$, Pit$_{3,1}$\}
“other” = other 10 pit variables
“known” = ¬pit$_{1,1} \land ¬$pit$_{1,2} \land ¬$pit$_{2,1}$
“breeze” = ¬breeze$_{1,1} \land$ breeze$_{1,2} \land$ breeze$_{2,1}$
Note: “breeze” is conditionally independent of “other” given “known”, “frontier” and “query”
And given independence of pits $P_{i,j}$

- $P(P_{1,3} | \text{known, breeze}) = \alpha P(P_{1,3}) \sum_{\text{frontier}} P(\text{breeze} | \text{known, } P_{1,3}, \text{frontier}) P(\text{frontier}) = \alpha \langle 0.2(0.04+0.16+0.16), 0.8(0.04+0.16) \rangle = \langle 0.31, 0.69 \rangle$

**P_{1,3}=true**

**P_{1,3}=false**
Wumpus World Revisited

- So, $P(P_{1,3} = \text{true}) = P(P_{3,1} = \text{true}) = 0.31$
- $P(P_{2,2} = \text{true}) = 0.86$
- Probabilistic agent “knows more” than logical agent
Summary: Uncertainty

- Probability theory allows us to reason about uncertainty
- Bayes rule allows us to change diagnostic questions to causal questions
- Conditional independence allows us to simplify complexity
- Probabilistic agent can outperform logical agent in partially observable and stochastic environments