Fundamentals of Artificial Intelligence

Reasoning under Uncertainty

Shyamanta M Hazarika
Mechanical Engineering
Indian Institute of Technology Guwahati
s.m.hazarika@iitg.ac.in

http://www.iitg.ac.in/s.m.hazarika/

Reasoning under Uncertainty

- **Probability theory** provides the **basis for our treatment of systems that reason under uncertainty**.
- **Utility Theory** provides **ways and means of weighing up the desirability of goals** and the likelihood of achieving them.
  - Actions are no longer certain to achieve goals.
- **Probability theory and utility theory** put together constitute decision theory; **take decisions within uncertain domain**.
  - Build rational agents for uncertain worlds.

Reasoning under Uncertainty

- Basics of probability theory, including the representation language for uncertain beliefs.
  - Acting Under Uncertainty
  - Rational Decisions
  - Basic Probability Notation
  - Bayes' Rule
- Belief networks, a powerful tool for representing and reasoning with uncertain knowledge.

A Logical Agent believes a sentence is either True or False.
A Probabilistic Agent has a degree of belief between 0 and 1

Acting under Uncertainty

- Within a logical-agent approach, **agents almost never have access to the whole truth** about their environment.
  - Some **sentences can be ascertained directly from the agent's percepts**, and others can be **inferred from current and previous percepts** together with knowledge about the environment.
  - However, **for almost every case, there will be important questions to which the agent cannot find a categorical answer**.
- The **agent must therefore act under uncertainty**.
- Uncertainty can also **arise because of incompleteness and incorrectness in the agent's understanding** of the properties of the environment.

© Shyamanta M Hazarika, ME, IIT Guwahati

Handling uncertain knowledge

Trying to use first-order logic to cope with complex domain like medical diagnosis fails for three main reasons:

1. **Laziness**: It is too much work to list the complete set of antecedents or consequents needed to ensure an exception less rule, and too hard to use the enormous rules that result.
2. **Theoretical ignorance**: Expertise of the area may not be sufficient to have complete theory for the domain.
3. **Practical ignorance**: Even if we know all the rules, we may be uncertain about particular cases because all the necessary tests have not or cannot be run.

Handling uncertain knowledge

- **Agent's knowledge can at best provide only a degree of belief** in the relevant sentences.
  - True for medical domain, as well as most other judgmental domains: law, business, design, automobile repair, gardening, dating, and so on
- Dealing with **degrees of belief is through probability theory**, which assigns a numerical degree of belief between 0 and 1 to sentences.
- **Probability provides a way of summarizing the uncertainty** that comes from our laziness and ignorance.

Uncertainty and Rational Decisions

- To make such choices, an agent must first have preferences, between possible outcomes of the plans.
  - Use the utility theory to represent and reason with "preference"

1. **Preference**: options, choices, what is more preferred.
2. **Outcome**: Completely specified state.
3. **Utility Theory**: "The quality of being useful" - theory says that every state has a degree of usefulness, or utility, to an agent and that the agent will prefer states with higher utility.

Basic Probability Notation

- Notation for **describing degrees of belief.**
  - Formal language for representing and reasoning with uncertain knowledge.
- The version of probability theory we present uses an extension of **propositional logic for its sentences**.
- The **dependence on experience** is reflected in the syntactic distinction between
  - **prior probability statements**, which apply before any evidence is obtained, and
  - **conditional probability statements**, which include the evidence explicitly.

Basic Probability Notation

- Sample Space: The **set of all possible worlds i.e., all possible outcomes** is referred to as sample space.
- Notation
  - Ω - Sample Space
  - ω - An element in the sample space
  - φ - An event or a proposition
    - An event φ is a subset of sample space Ω: $\varphi \subseteq \Omega$
      Example: Two Dice adding up to 11 is an event
      φ = { (5,6) , (6,5) }

Probability Model

- A probability model associates a numerical probability P(ω) with each possible world:
  - Every possible world must have a probability between 0 and 1
  - Total probability of the set of all possible worlds is 1
- Unconditional Probability
  - Unconditional Probability is when you **don't consider any other information** except for the object in question.
    - Example – Two dices – red and blue; consider only one – red.
- Conditional Probability
  - In Conditional Probability **we have evidence i.e., extra information** already revealed.
    - Example – Rolling two dices – one is a 6; Sum cannot be 5!

Prior Probability

Use notation **P(A) for the unconditional or prior probability** that proposition A is true.

For example, if *Fever* denotes the proposition that a particular patient has a fever,

P(Fever) = 0.1

P(A) can only be used when there is no other information. As soon as some new information B is known, we have to reason with the conditional probability of A given B instead of P(A).

means that **in the absence of any other information**, the agent will assign a probability of 0.1 (a 10% chance) to the event of the patient having a fever.

Random Variables

- The proposition that is the subject of a probability statement can be represented by a proposition symbol, as in the P(A) example.
- **Propositions can also include equalities involving random variables**.
- Every Random Variable has a domain - a set of possible values that it can take.
  - For example, lets say we have the random variable Total that calculates the sum of two dice:
    - Then the domain is the set {2, …, 12}
  - A Boolean random variable has the domain {True, False}

Random Variables

- For propositions involving random variables; For example, if we are **concerned about the random variable Weather**, we might have
  - P(Weather=Sunny) = 0.7
  - P(Weather=Cloudy) = 0.08
- Can view **proposition symbols as random variables** as well, if we assume that they have a domain [true,false).
  - For example: Expression **P(Fever) can be viewed as shorthand for P(Fever = true)**.
  - Similarly, **P(¬ Fever) is shorthand for P(Fever =false)**.

Probability Distribution

- A probability distribution is when we want to talk about **all the possible values of a random variable**. Usually **indicated by a bold P**.
- A **Discrete Random Variable** is a random variable that takes a **finite number of distinct values**.

  For example,

  An expression such as P(Weather), denotes a vector of values for the probabilities of each individual state of the weather.

  For example, we would write

  P(Weather) = (0.7,0.2,0.08,0.02)
- This statement defines a **probability distribution.**

Probability Density Function

- A **Continuous Random Variable** is a random variable that **takes an infinite number of distinct values**.

  For Example:

  $P(Temp = x) = Uniform_{[18C,\ 26C]}(x)$

  Expresses that the temperature is distributed uniformly between 18 and 26 degrees.

- This is called a **probability density function**.

Conditional Probability

- Conditional or posterior probabilities is expressed with the notation P(A|B).
  - This is read as "the probability of A given that all we know is B."
- For example

  P(Cavity|Toothache) = 0.8

  - Indicates that if a patient is observed to have a toothache, and no other information is yet available, then the probability of the patient having a cavity will be 0.8.

Once the agent has obtained some evidence concerning the previously unknown propositions making up the domain, prior probabilities are no longer applicable.

Conditional Probability

- P(X | Y) is a two-dimensional table giving the values of $P(X=x_i|Y = y_j)$ for each possible i, j.
- Conditional probabilities can be **defined in terms of unconditional probabilities.**

$$P(A|B) = \frac{P(A \cap B)}{P(B)}$$

- This **equation can also be written as follows, which is called the product rule**.

The product rule is perhaps easier to remember: it comes from the fact that for A and B to be true, we need B to be true, hen A to be true given B.

$$P(A \cap B) = P(A|B)\ P(B)$$

Axioms of Probability

To define properly the semantics of statements in probability theory, we will need to describe how probabilities and logical connectives interact.

1. All probabilities are between 0 and 1.

   $0 < P(A) < 1$

2. Necessarily true (i.e., valid) propositions have probability 1, and necessarily false (i.e., unsatisfiable) propositions have probability 0.

   P(True) = 1; P(False) = 0

3. The probability of a disjunction is given by

   $P(A \vee B) = P(A) + P(B) - P(A \wedge B)$

Axioms of Probability

To define properly the semantics of statements in probability theory, we will need to describe how probabilities and logical connectives interact.

1. All probabilities are between 0 and 1.

   $0 < P(A) < 1$

2. Necessarily true (i.e., valid) propositions have probability 1, and necessarily false (i.e., unsatisfiable) propositions have probability 0.

   $P(True) = 1; \; P(False) = 0$

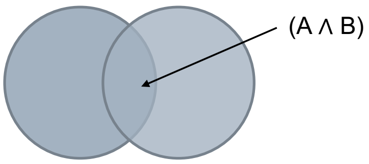
The FIRST TWO axioms **serve to define the probability scale.**

Axioms of Probability

To define properly the semantics of statements in probability theory, we will need to describe how probabilities and logical connectives interact.

3. The probability of a disjunction is given by

   $P(A \vee B) = P(A) + P(B) - P(A \wedge B)$



The figure depicts each proposition as a set, which can be thought of as the set of all possible worlds in which the proposition is true.

The **total probability of (A ∨ B)** is seen to be the **sum of the probabilities assigned to A and B,** but with **P(A ∧ B) subtracted out so that those cases are not counted twice.**

Axioms of Probability

- From these three axioms, we can **derive all other properties of probabilities**.

  For example,

  If we let B be ¬A in the last axiom, we obtain an expression for the **probability of the negation of a proposition** in terms of the probability of the proposition itself:

  $P(A \lor \neg A) = P(A) + P(\neg A) - P(A \land \neg A)$

  $P(True) = P(A) + P(\neg A) - P(False)$

  $1 = P(A) + P(\neg A)$

  $P(\neg A) = 1 - P(A)$

Joint Probability Distribution

- Joint probability distribution **completely specifies an agent's probability assignments to all propositions** in the domain (both simple and complex).
- The joint probability distribution assigns probabilities to all possible atomic events.
  - An n-dimensional table with a value in every cell giving the probability of that specific state occurring.

	Toothache	¬Toothache
Cavity	0.04	0.06
¬Cavity	0.01	0.89

Joint Probability Distribution

	Toothache	¬Toothache
Cavity	0.04	0.06
¬Cavity	0.01	0.89

- Atomic events are mutually exclusive, any conjunction of atomic events is necessarily false. Because they are collectively exhaustive, their disjunction is necessarily true.
  - Hence, from the second and third axioms of probability, the entries in the table sum to 1.
  - The joint probability distribution can be used to compute any probabilistic statement.

Joint Probability Distribution

	Toothache	¬Toothache
Cavity	0.04	0.06
¬Cavity	0.01	0.89

- Adding across a row or column gives the unconditional probability of a variable,
  - P(Cavity) = 0.06 + 0.04 = 0.10.
  - P(Cavity ∨ Toothache) = 0.04 + 0.01 + 0.06 = 0.11
- Conditional probabilities can be found from the joint,
  - $P(\text{Cavity}|\text{Toothache}) = \frac{P(\text{Cavity} \wedge \text{Toothache})}{P(\text{Toothache})} = \frac{0.04}{0.04+0.01} = 0.8$

Bayes' Rule

- Recall the **two forms of the product rules**

  $P(A \wedge B) = P(A|B)\ P(B)$

  $P(A \wedge B) = P(B|A)\ P(A)$
- Equating the two right-hand sides and dividing by P(A), we get

$$P(B|A) = \frac{P(A|B)\ P(B)}{P(A)}$$

- This **equation is known as Bayes' rule** (also Bayes' law or Bayes' theorem). This simple equation underlies all modern AI systems for probabilistic inference.

Bayes' Rule

$$P(B|A) = \frac{P(A|B)\ P(B)}{P(A)}$$

- Bayes' rule requires three terms
  - two prior probabilities and
  - a conditional probability

  to compute the fourth an conditional probability.
- In practice, Bayes' rule is useful, **we have good probability estimates for these three quantities** and need to compute the fourth.

Applying Bayes' Rule: Simple Case

A doctor knows that the disease **meningitis causes the patient to have a stiff neck, say, 50% of the time**. The doctor also knows some unconditional facts: the **prior probability of a patient having meningitis is 1/50,000**, and the **prior probability of any patient having a stiff neck is 1/20**.

S be the proposition that the patient has a stiff neck
M be the proposition that the patient has meningitis.

$P(S|M) = 0.5$

$P(M) = \frac{1}{50000} = 0.00002$

$P(S) = \frac{1}{20} = 0.05$

$P(M|S) = \frac{P(S|M)P(M)}{P(S)} = 0.0002$

Notice that even though a stiff neck is strongly indicated by meningitis (probability 0.5), the probability of meningitis in the patient remains small.

This is because the prior on stiff necks is much higher than that for meningitis.

Using Bayes' Rule: Combining Evidence

Suppose we have two conditional probabilities relating to cavities:

$P(Cavity|Toothache) = 0.8$
$P(Cavity|Catch) = 0.95$

What can a dentist conclude if her nasty steel probe catches in the aching tooth of a patient?

$$P(Cavity|Toothache \wedge Catch) = \frac{P(Toothache \wedge Catch|Cavity)\ P(Cavity)}{P(Toothache \wedge Catch)}$$

Although it seems feasible to estimate conditional probabilities for n different individual variables, it is a daunting task to come up with numbers for n^2 pairs of variables.

- Application of Bayes' rule - simplified to a form that requires fewer probabilities in order to produce a result.

Using Bayes' Rule: Combining Evidence

- The process of Bayesian updating incorporates evidence one piece at a time, modifying the previously held belief in the unknown variable.

$$P(Cavity|Toothache) = \frac{P(Toothache\ |Cavity)\ P(Cavity)}{P(Toothache)}$$

- When Catch is observed, we can apply Bayes' rule with **Toothache as the constant conditioning context**.

$$P(Cavity|Toothache \wedge Catch) = \frac{P(Cavity|Tootache)\ P(Catch|Toothache \wedge Cavity)}{P(\ Catch|Toothache)}$$

- In Bayesian updating, as new piece of evidence is observed, the belief in the unknown variable is multiplied by a factor that depends on the new evidence.

Using Bayes' Rule: Combining Evidence

The cavity is the direct cause of both the toothache and the probe catching in the tooth. Given a cavity, the probability of the probe catching does not depend on the presence of a toothache; similarly, the probe catching is not going to change the probability that the cavity is causing a toothache.

- Exploit **conditional independence** of Toothache and Catch given Cavity.
  - Given conditional independence, we can simplify the equation for updating.
- Combining many pieces of evidence may require assessing a large number of conditional probabilities.
- **Conditional independence** brought about by direct causal relationships in the domain **allows Bayesian updating to work effectively** even with multiple pieces of evidence.

Reasoning under Uncertainty

- Uncertainty arises because of both laziness and ignorance. It is inescapable in complex, dynamic, or inaccessible worlds.
  - Many of the simplifications that are possible with deductive inference are no longer valid.
- Probabilities express the agent's inability to reach a definite decision regarding the truth of a sentence, and summarize the agent's beliefs.
  - Basic probability statements include prior probabilities and conditional probabilities over simple and complex propositions.

Reasoning under Uncertainty

- Axioms of probability specify constraints on reasonable assignments of probabilities to propositions.
  - An agent that violates the axioms will behave irrationally in some circumstances.
  - The joint probability distribution specifies the probability of each complete assignment of values to random variables. It is usually far too large to create or use.
- Bayes' rule allows unknown probabilities to be computed from known, stable ones.
- Conditional independence allows Bayesian updating to work effectively even with multiple pieces of evidence.