

Quantifying Uncertainty

Michael Rose

In which we see how an agent can tame uncertainty with degrees of belief

13.1 | Acting Under Uncertainty

To make choices with endless degrees of certainty or uncertainty, an agent must have preferences between the different possible outcomes of the various plans. We use **utility theory** to represent and reason with preferences. Utility theory states that every state has a degree of usefulness, or utility, to an agent and that the agent will prefer states with higher utility. Preferences, as expressed by utilities, are combined with probabilities in the general theory of rational decisions called **decision theory**. Decision theory = probability theory + utility theory. The fundamental idea of decision theory is that an agent is rational if and only if it chooses the action that yields the highest expected utility, averaged over all the possible outcomes of action.

13.2 | Basic Probability Notation

Basic axioms of probability:

$0 \leq P(\omega) \leq 1$ for every ω and $\sum_{\omega \in \Omega} P(\omega) = 1$.

In AI, the sets of events are described by propositions in a formal language. For each proposition, the corresponding set contains just those possible worlds in which the proposition holds. The probability associated with a proposition is defined to be the sum of the probabilities of the worlds in which it holds. For any proposition ϕ , $P(\phi) = \sum_{\omega \in \phi} P(\omega)$.

A possible world is defined to be an assignment of values to all of the random variables under consideration.

de Finetti proved the following:

informally: if an agent has some degree of belief in a proposition a , then the agent should be able to state odds at which it is indifferent to a bet for or against a .

more formally: If agent 1 expresses a set of degrees of belief that violate the axioms of probability theory then there is a combination of bets by agent 2 that guarantees that agent 1 will lose money every time.

De Finetti's theorem implies that no rational agent can have beliefs that violate the axioms of probability.

13.3 | Inference Using Full Joint Distributions

Frequentist - numbers come only from experiments. From any finite sample, we can estimate the true fraction and calculate how accurate our estimate is likely to be

Objectivist - Probabilities are real aspects of the universe - propensities for objects to behave in certain ways - rather than just descriptions of an observer's degree of belief.

Subjectivist - Probabilities are a way of characterizing an agent's beliefs, rather than having an external degree of significance.

Subjectivist Bayesian - allows any self-consistent ascription of prior probabilities to propositions, but insists on proper Bayesian updating as evidence arrives.

Even the strict frequentist position involves subjective analysis because of the **reference class problem**: in trying to determine the outcome probabilities of a particular experiment, the frequentist has to place it in a class of similar experiments with known frequencies.

The **principle of indifference** states that propositions that are syntactically symmetric with respect to the evidence should be accorded equal probability.

The Marginalization Rule: $P(Y) = \sum_{z \in Z} P(Y, z)$

The Conditioning Rule: $P(Y) = \sum_z P(Y|z)P(z)$

The full joint distribution in tabular form is just not a practical tool for building reasoning systems due to inefficient computational scaling. As a result, it should be viewed as a theoretical foundation for more effective approaches to be built.

13.4 | Independence

Independence assertions can help in reducing the size of the domain representation and the complexity of the inference problem. These allow us to decompose joint distributions into factored separate joint distributions on each subset. Unfortunately, clean independence is quite rare.

13.5 | Bayes' Rule and its Use

Bayes' Rule: $P(b|a) = \frac{P(a|b)P(b)}{P(a)}$

Conditional Independence: $P(X, Y|Z) = P(X|Z)P(Y|Z)$

Conditional independence assertions allow probabilistic systems to scale up; moreover, they are much more commonly available than absolute independence assertions. The decomposition of large probabilistic domains into weakly connected subsets through conditional independence is one of the most important developments in the recent history of AI.

Naive Bayes Model: $P(Cause, Effect_1, \dots, Effect_n) = P(Cause) \prod_i P(Effect_i|Cause)$

This is called naive because it is often used as a simplifying assumption in cases where the effect variables are not actually conditionally independent given the cause variable.

13.7 | Summary

- Uncertainty arises both due to laziness and ignorance. It is inescapable in complex, nondeterministic, or partially observable environments.
- Probabilities express an agent's inability to reach a definite decision regarding the truth of a sentence. Probabilities summarize the agent's beliefs relative to the evidence.
- Decision theory combines the agent's beliefs and desires, defining the best action as the one that maximizes expected utility
- Basic probability statements include prior probabilities and conditional probabilities over simple and complex propositions

- The axioms of probability constrain the possible assignments of probabilities to propositions. An agent that violates the axioms must behave irrationally in some cases.
- The full joint probability distribution specifies the probability of each complete assignment of values to random variables. It is usually too large to create or use in its explicit form, but when it is available it can be used to answer queries simply by adding up entries for the possible worlds corresponding to query propositions.
- absolute independence between subsets of random variables allows the full joint distribution to be factored into smaller joint distributions, greatly reducing its complexity. Absolute independence seldom occurs in practice.
- Baye's Rule allows unknown probabilities to be computed from known conditional probabilities, usually in the causal direction. Applying Bayes' rule with many pieces of evidence runs into the same scaling problems as the joint probability distribution.
- Conditional independence brought about by direct causal relationships in the domain might allow the full joint distribution to be factored into smaller, conditional distributions. The naive Bayes model assumes the conditional independence of all effect variables, given a single cause variable, and grows linearly with the number of effects.