Artificial Intelligence

IV. Uncertain Knowledge and Reasoning

1. Uncertainty

F. C. Langbein

School of Computer Science
and Informatics
Cardiff University

Overview

- Making decisions with uncertain knowledge
  - Objective probability
  - Subjective/Bayesian probability
- Probability theory
  - Probability spaces
  - Random variables
  - Propositions
  - Conditional probability
  - Bayes' rule, product rule, chain rule
  - Probability distributions

Uncertainty

- Let action A_t be "leave for airport t minutes before flight"
  - Question: will A_t get me there on time?
- Problems
  - Partial observability (road state, other drivers' plans, ...)
  - Noisy sensors (radio traffic reports)
  - Uncertainty in action outcomes (flat tire, ...)
  - Immense complexity of modelling and predicting traffic
- Hence, a purely logical approach either
  - (1) risks *falsehood*: "A_{25} will get me there on time" *or*
  - (2) leads to *weak* conclusions (unsuitable for decisions):
    "A_{25} will get me there on time if there's no accident, it does not rain, my tires remain intact, ..."
- A_{1440} might reasonably get me there on time, but I'd have to stay overnight in the airport...

Handling Uncertainty

- *Non-monotonic logic*: conclusion may be retracted if new information becomes available
  - How to handle contradictions? What assumptions are reasonable?
- Probability
  - Deals with *laziness* (omitted data) and *ignorance* (lack of data)
  - Given *available evidence*, A_{60} will be on time with $P(A_{60}) = 0.6$
  - Probabilities change with new evidence

$$P(A_{60}|3\text{am}) = 0.9 \qquad P(A_{60}|3\text{am}, \text{accident\_report}) = 0.7$$
$$P(A_{60}|9\text{am}) = 0.3 \qquad P(A_{60}|\text{accident\_report}) = 0.1$$

  ➜ Update beliefs according to observations
- Choose action according to preferences
  - Missing flight vs. airport cuisine, . . .
  - Utility theory: represent and infer preferences
- Decision theory = utility theory + probability theory

Probability

- Objective probability
  - Averages over repeated experiments of random events
    - E.g. estimate $\mathrm{P}(\text{Rain})$ from historical observation
  - Makes assertions about future experiments
  - New evidence changes the reference class
- Subjective / Bayesian probability
  - *Degrees of belief* about unobserved event: *state of knowledge*
    - E.g. agent's belief that it will rain, given the season
  - Estimate probabilities from past experience
  - New evidence updates beliefs
- Fuzzy logic handles *degrees of truth*, not uncertainty
    - $\text{Wet}(\text{Grass})$ is true to degree 0.2
  - Fuzzy sets: degree of membership—rough vs. crisp (usual) sets

Probability Basics

- Probability space / probability model
  - Sample space: a set Ω (all models)
    - $\omega \in \Omega$: sample point / possible world / atomic event
  - With function $P : \Omega \to [0, 1]$, such that $\sum_{\omega \in \Omega} P(\omega) = 1$
- An event A is any subset of Ω
  - $P(A) = \sum_{\omega \in A} P(\omega)$
- E.g. die roll probability space
  - $\Omega = \{1, 2, 3, 4, 5, 6\}$
  - $P(1) = P(2) = P(3) = P(4) = P(5) = P(6) = 1/6$
  - $P(\{1, 2, 3\}) = 1/6 + 1/6 + 1/6 = 1/2$

Random Variables

- A random variable is function $\mathbf{X} : \Omega \to \mathbb{X} = \{x_1, \ldots, x_n\}$
  - from sample space to some values $\mathbb{X}$ (reals, integers, Booleans)
- P induces a probability distribution $\mathbf{P}(\mathbf{X})$ for $\mathbf{X}$:

$$\mathbf{P}(\mathbf{X}) = \langle P(\mathbf{X} = x_1), \ldots, P(\mathbf{X} = x_n) \rangle$$

with $P(\mathbf{X} = x_l) = \sum_{\omega \in \Omega \wedge \mathbf{X}(\omega) = x_l} P(\omega)$

- E.g. $\mathbf{P}(\mathbf{DieRoll}) = \langle \underbrace{1/6}_{1}, \underbrace{1/6}_{2}, \underbrace{1/6}_{3}, \underbrace{1/6}_{4}, \underbrace{1/6}_{5}, \underbrace{1/6}_{6} \rangle$
  - $P(\mathbf{DieRoll} < 4) = 1/2$
- Or $\mathbf{P}(\mathbf{Odd}) = \langle \underbrace{1/2}_{\text{true}}, \underbrace{1/2}_{\text{false}} \rangle$
  - $P(\mathbf{Odd} = \mathrm{true}) = 1/6 + 1/6 + 1/6 = 1/2$

Propositions: Semantics

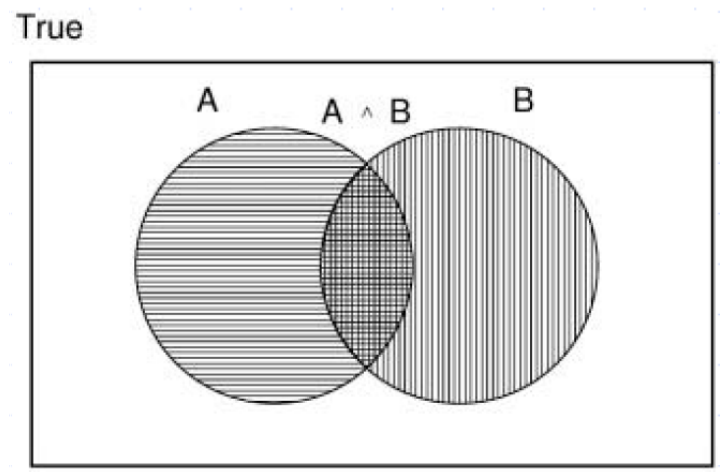
- A proposition is an *event* where proposition is true
  - E.g. given Boolean random variables $\mathbf{A}$ and $\mathbf{B}$
    - $a = \{\omega \in \Omega : \mathbf{A}(\omega) = \text{true}\}$
    - $\neg a = \{\omega \in \Omega : \mathbf{A}(\omega) = \text{false}\}$
    - $a \wedge b = \{\omega \in \Omega : \mathbf{A}(\omega) = \text{true} \wedge B(\omega) = \text{true}\}$
- Given only Boolean (random) variables:
  Proposition = disjunction of atomic events in which it is true
  - $(a \vee b) \equiv (\neg a \wedge b) \vee (a \wedge \neg b) \vee (a \wedge b)$
    $\Rightarrow P(a \vee b) = P(\neg a \wedge b) + P(a \wedge \neg b) + P(a \wedge b)$
- Often in AI: sample points are defined by the values of a set of random variables
  - ⇒ Sample space is Cartesian product of variable domains

Propositions: Syntax

- **Boolean** random variables
  - E.g. $Cavity$ = "Do I have a cavity?"
  - $Cavity$ = false or ¬cavity is a proposition
- **Discrete** random variables (*finite* or *infinite*)
  - E.g. $Weather$ is one of ⟨sun,rain,cloud,snow⟩
  - $Weather$ = rain is a proposition
  - Values must be exhaustive and mutually exclusive
- **Continuous** random variables (*bounded* or *unbounded*)
  - E.g. $Temperature = 21.6$
  - Or inequalities, e.g., $Temperature < 22.0$
- Arbitrary Boolean combinations of atomic propositions give additional propositions

Probability and Logic

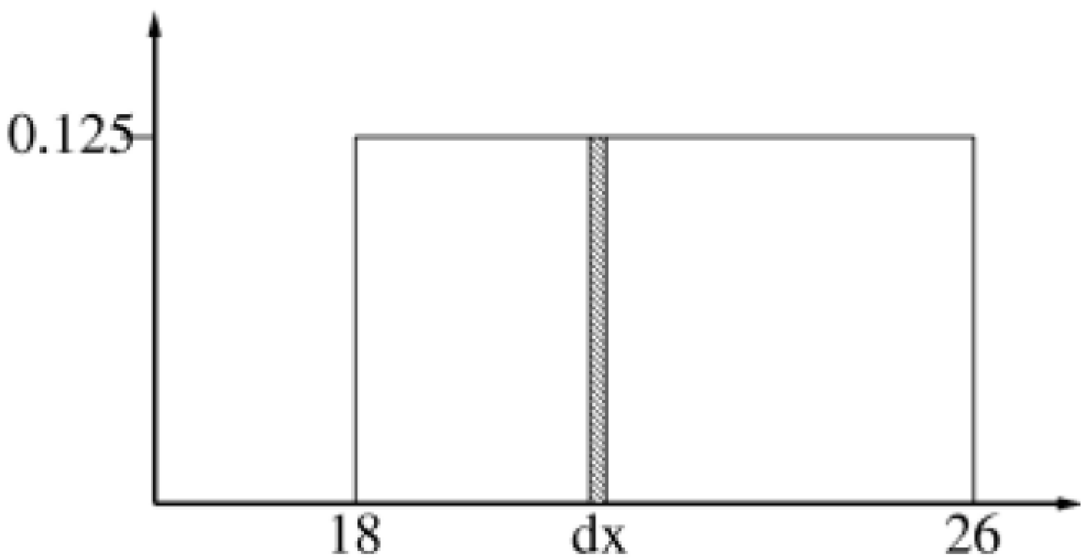
- The definitions imply that certain logically related events must have related probabilities
  - $0 \leq P(a) \leq 1$
  - $P(\text{true}) = 1$ and $P(\text{false}) = 0$
  - $P(a \vee b) = P(a) + P(b) - P(a \wedge b)$

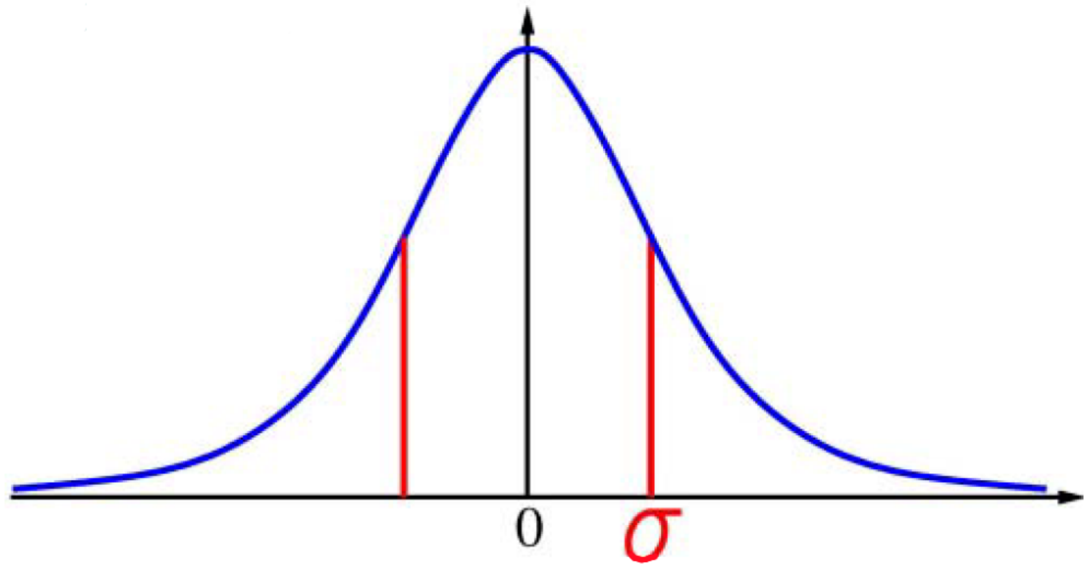


- de Finetti (1931):
  An agent who bets according to probabilities that violate these axioms can be forced to bet so as to lose money regardless of outcome.

Probability for Continuous Variables

- Express distribution as a *parameterised function of value*:
  - $P(\mathbf{X}_U = x) = U[18, 26](x)$: uniform density between 18 and 26
  - $P_G(\mathbf{X}_G = x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(\frac{-(x-\mu)^2}{2\sigma^2}\right)$: Gaussian density with mean μ and variance σ^2





- $\mathbf{P}(\mathbf{X})$ is a probability density such that $\int_{\mathbb{X}} P(X = x)\ dx = 1$
- $P(\mathbf{X} = 20.5) = 0.125$ really means

$$\lim_{\epsilon \to 0} \frac{P(20.5 \leq \mathbf{X} \leq 20.5 + \epsilon)}{\epsilon} = 0.125$$

Unconditional Probability

- Unconditional or prior probabilities of propositions correspond to arrival of any (new) evidence
  - E.g. $P(\text{cavity}) = 0.1$, $P(\mathbf{Weather} = \text{sun}) = 0.72$
- *Probability distribution*: values for all possible assignments
  - $\mathbf{P}(\mathbf{Weather}) = \langle 0.72, 0.1, 0.08, 0.1 \rangle$ (normalised, i.e. sums to 1)
- Joint *probability distribution* for a set of random variables: probability of every combination of values of those variables
  - $\mathbf{P}(\mathbf{Weather}, \mathbf{Cavity}) =$ 2×4 table

$\mathbf{Weather} =$	sun	rain	cloud	snow
$\mathbf{Cavity} =$ true	0.144	0.02	0.016	0.02
$\mathbf{Cavity} =$ false	0.576	0.08	0.064	0.08

- *Joint distribution answers every question about a domain because every event is a sum of sample points*

Conditional Probability

- Conditional or posterior probabilities
  - E.g. $P(\text{cavity}|\text{toothache}) = 0.8$
  - I.e. "given that toothache is all I know",
    not "if toothache then 80% chance of cavity"
- Probabilities change with new evidence / observations
  - E.g. $P(\text{cavity}|\text{tootchache} \wedge \text{cavity}) = 1$
  - Note, the less specific belief *remains valid* after more evidence arrives, but it is not always *useful*
- New evidence may be irrelevant, allowing simplification
  - E.g. $P(\text{cavity}|\text{toothache} \wedge \text{rain}) = P(\text{cavity}|\text{toothache})$
  - Simplification sanctioned by domain knowledge, is crucial

Conditional Probability Formulae

- Definition of conditional probability

$$P(a|b) = \frac{P(a \wedge b)}{P(b)} \text{ if } P(b) \neq 0$$

  – Probability of observing event a given evidence (knowledge / observation) of event b

$\neg a \wedge \neg b$	b
a	$b \wedge a$

$P(b) = 1/2$, $P(a) = 1/2$
$P(a|b) = 1/2$

- Product rule

$$P(a \wedge b) = P(a|b)P(b) = P(b|a)P(a) = P(b \wedge a)$$

- Bayes' rule

$$P(a|b) = P(b|a)\frac{P(a)}{P(b)}$$

Conditional Distributions

- Conditional distributions of random variables
  - $\mathbf{P}(Cavity|Toothache, Rain) =$

redundant

Toothache	Rain	P(cavity)	P(¬cavity)
true	true	.8	.2
true	false	.8	.2
false	true	.1	.9
false	false	.1	.9

- Formulae for distributions (per variable value combination)

$$\mathbf{P}(X|Y) = \frac{\mathbf{P}(X,Y)}{\mathbf{P}(Y)} \text{ (definition)}$$

$$\mathbf{P}(X,Y) = \mathbf{P}(X|Y)\mathbf{P}(Y) = \mathbf{P}(Y|X)\mathbf{P}(X) \text{ (product rule)}$$

$$\mathbf{P}(X|Y) = \mathbf{P}(Y|X)\frac{\mathbf{P}(X)}{\mathbf{P}(Y)} \text{ (Bayes' rule)}$$

Bayes' Rule

- Bayes' rule is useful for assessing *diagnostic* probability from *causal* probability

$$P(Cause|Effect) = P(Effect|Cause)\frac{P(Cause)}{P(Effect)}$$

  - E.g. let M be "menengitis?", and S be "stiff neck?"

$$P(m|s) = P(s|m)\frac{P(m)}{P(s)} = .8 \times .0001/.1 = .0008$$

    – Note, posterior probability of meningitis still very small

- Often causal probability $P(Effect|Cause)$ is easier to determine

Renormalisation

- Normalisation constant α
  - Multiply non-negative function $f : \mathbb{X} \to \mathbb{R}_0^+$ with α s.t.

$$\alpha \sum_{x \in \mathbb{X}} f(x) = 1 \quad \text{or} \quad \alpha \int_{\mathbb{X}} f(x)\, dx = 1$$

- Bayes' rule $\quad \mathbf{P}(\mathbf{X}|\mathbf{Y}) = \mathbf{P}(\mathbf{Y}|\mathbf{X}) \dfrac{\mathbf{P}(\mathbf{X})}{\mathbf{P}(\mathbf{Y})} = \alpha \underbrace{\mathbf{P}(\mathbf{Y}|\mathbf{X})\mathbf{P}(\mathbf{X})}_{\mathbf{Q}(\mathbf{Y},\mathbf{X})}$
- Here $\alpha = 1/\mathbf{P}(\mathbf{Y})$ hard to compute on its own
  - But we can sum over all cases $\{x_l\}$ where $\mathbf{Y} = y_k$

$$1/\alpha_k = P(y_k) = \sum_l P(y_k|x_l)P(x_l) = \sum_l Q(y_k|x_l)$$

➡ $$\alpha = 1/\mathbf{P}(\mathbf{Y}) = 1 \Big/ \sum_l \mathbf{P}(\mathbf{Y}|x_l)\mathbf{P}(x_l) = 1 \Big/ \sum_l \mathbf{Q}(\mathbf{Y}, x_l)$$

Chain Rule

- Chain rule derived by successive application of product rule

$$\mathbf{P}(X_1,\ldots,X_n) = \mathbf{P}(X_1,\ldots,X_{n-1})\mathbf{P}(X_n|X_1,\ldots,X_{n-1})$$
$$= \mathbf{P}(X_1,\ldots,X_{n-2})\mathbf{P}(X_{n-1}|X_1,\ldots,X_{n-2})\mathbf{P}(X_n|X_1,\ldots,X_{n-1})$$
$$= \cdots = \prod_{l=1}^{n} \mathbf{P}(X_l|X_1,\ldots,X_{l-1})$$

  - Takes joint probability distribution apart into conditional probability distributions
    - May be useful when computing probabilities