Session 8

I. *Announcements [5 minutes]*

- **Feedback**: *I'd like to get feedback from everybody in section on how I can improve sections: 3 things you like, 3 things you dislike.*
    - I've sent out a link to a *Remailer* you can send email to me from anonymously. Also the link is posted on my webpage.
    - I know you're busy with Sudoku's, train tracks, and so forth, but take 10 minutes and send me feedback – it'll make the course better for you and me.

II. *Test Recap… briefly [10-20 minutes]*

- How did the test go?
- What questions did you have the most problems with?

III. *Introduction to Probability*

- Agents almost never have access to all of their environment's information.
    - This qualification problem is due to 3 things
        1. Laziness – too much work to explore all possibilities
        2. Incomplete Theory
        3. Practical Limitations – not all tests can be performed
    - Must be able to derive a plan that will work most of the time.
    - In uncertain environments, the rational decision depends on
        1. The relative importance of various goals
        2. The likelihood and degree of achievement of each goal
- Agent's knowledge provides a degree of belief ➔ Probability Theory
    - Probability Theory provides the degree to which an agent believes a statement given all plausible alternative situations that are indistinguishable from the current situation.
    - Probability Theory has the same ontological commitment as logic
        - *Facts either hold or do not hold in the world*
    - Probabilities depend on the evidence (precepts) the agent has acquired.
- Agents must have preferences towards the outcomes of different plans.
    - **Utility Theory** – Theory of preference based on an object's utility.
        - Preference indicates that different agents might have different goals so while goals might seem misguided, they are not necessarily irrational.
    - **Decision Theory** – Combination of Probability and Utility Theory.
        - **Principle of Maximum Expected Utility** - An agent is rational iff it chooses the action that yields the highest expected utility (averaged over all outcomes).
- **Belief State** – a representation of probabilities of all possible actual states of the world.

Probability Theory

- Language of Probability – ascribes degrees of belief to propositions.
    - **Random Variable** – refers to part of the world with initial unknown status
    - Domain – the values a random variable can take on.
        - Boolean – true/false
        - Discrete – countable, mutually exclusive, and exhaustive.
        - Continuous – uncountable
    - **Atomic Event –** a specific complete specification of the world about which the agent is uncertain.
        - The set of atomic events is mutually exclusive and exhaustive → forms a partition
        - Any atomic event entails true/false for every proposition.
        - Any proposition is equivalent to the disjunction of all atomic events that entail the proposition as true.
- **Prior Probability P(x)** – the degree of belief of proposition x in the absence of any evidence about other propositions.
    - **Probability Distribution**, P(X) – the vector of probabilities ascribed to each of the possible states of proposition X.
    - **Joint Probability Distribution**, $P(X_1,X_2,\ldots,X_n)$ – the probabilities of all combinations of values of variables X_1, X_2, …, X_n.
    - **Full Joint Probability Distribution** – includes the complete set of variables for the environment.
    - **Probability Density Function**, p(x)dx – The probability of a continuous variable on the interval [x,x+dx] in the limit as dx → 0.
- **Conditional Probability P(x|y)** – the degree of belief of proposition x given the evidence y.
- **Product Rule**: $P(X,Y) = P(X\mid Y)P(Y)$
- **Kolmogrov's Axioms**
    1. All probabilities are between 0 and 1:
       $$0 \le P(a) \le 1$$
    2. True propositions have probability 1; false propositions have 0:
       $$P(true) = 1 \qquad P(false) = 0$$
    3. Probability of a disjunction is given by
       $$P(a \vee b) = P(a) + P(b) - P(a \wedge b)$$
- Using these Axioms we can derive other important rules:
    - Let a discrete variable X have a domain $\{x_1, x_2, \ldots, x_n\}$ or χ:
      $$\sum_{i=1}^{n} P(X = x_i) = 1 \qquad \text{or} \qquad \int_{\chi} P(x)\,dx = 1$$
    - The probability of a proposition is equal to the sum of the probabilities in which it holds; the set **e**(a) for proposition a.
      $$P(a) = \sum_{e_i \in \mathbf{e}(a)} P(e_i)$$

- In probability, statements do not refer directly to the world, but rather to an agent's belief about the world, so why can't agent's beliefs violate the probability axioms?
    - **de Finetti Theorem** – If Agent 1 expresses a set of degrees of belief that violate Kolmogorov's Axioms, then there is a combination of bets Agent 2 can place that guarantees that Agent 1 will lose money every time.
    - Decision making is an unavoidable process (not making a decision is a decision).
- Probability Philosophy (Briefly)
    - **Frequentist –** probabilities are results of repeated experiments.
    - **Objectivist –** probabilities come from a propensity of objects to act in a certain way.
    - **Subjectivist –** probabilities are a way of characterizing beliefs.
    - **Principle of Indifference –** Propositions that are syntactically 'symmetric' w.r.t. evidence should be given equal probability.
    - **Inductive Logic** – a logic capable of computing the correct probability for any proposition from any collection of observations. Unclear if it exists.

Probabilistic Inference – computation of posterior probability of query proposition from observed evidence.

- **Marginalization (Conditioning)** – variables other than the query variable are summed out in order to obtain the probability of the query variable:

$$\text{joint}: \quad P(Y) = \sum_z P(Y, z)$$

$$\text{conditional}: \quad P(Y) = \sum_z P(Y \mid z) P(z)$$

- **Normalization –** introduction of constant α that normalizes the distribution to 1 instead of explicitly computing the constant probabilities that that do not depend on the variable being calculated
- Combinatorial Explosion – Full Joint Table is size $O(2^n)$ and requires $O(2^n)$ time to be processed.

(Marginal) Independence – variables independent of each other can be factored:

$$P(X,Y) = P(X)P(Y) \qquad P(X \mid Y) = P(X)$$

- If a complete set of variables can be divided into independent subsets, then the full joint distribution can be factored into separated joint distributions on those subsets.

Bayes' Rule – application of product rule that allows diagnostic beliefs to be derived from casual beliefs:

$$P(Y \mid X) = \frac{P(X \mid Y)P(Y)}{P(X)} \qquad P(Y \mid X, e) = \frac{P(X \mid Y, e)P(Y \mid e)}{P(X \mid e)}$$

- Diagnostic knowledge is often more fragile than casual knowledge ➔ direct casual (model-based) knowledge provides robustness needed for probabilistic systems to function in the real world.
- Conditional Independence – implies that two variables X,Y are independent given variable Z:

$$P(X, Y \mid Z) = P(X \mid Z)P(Y \mid Z) \qquad P(X \mid Y, Z) = P(X \mid Z)$$

- Given n symptoms that are conditionally independent given Z, the size of the representation grows as O(n) instead of O(2^n).
    - Allows probabilistic systems to scale up.
    - Conditional independence assertions are more common than marginal independence assertions.
- **Naïve Bayes Model –** a single cause Y directly influences a number of events X_i that are all conditionally independent given the cause:

$$P(Y, X_1, X_2, \ldots, X_n) = P(Y)\prod_i P(X_i \mid Y)$$

    - Often works in situations where conditional independence does not hold.