Today's lecture:

- Refresher on probability basics
- Final project for the class

Probability in AI:

- The agent has a knowledge base where it encodes the state of the world. Here we have a measure of belief in some proposition: <u>subjective probability</u>

Numerical Measures of Belief:

- A probability value of f in $[0,1]$
- The probability is a measure of the agent's ignorance, NOT the DEGREE of truth. Only how much we believe the proposition is true within some state of the world.

Random Variable:

- Term in a language that can take one of a number of different values
- The $range$ of X, denoted $range(X)$ is the set of values X can take
- <u>Complex random variable</u>: a tuple of random variables (complex rand var), where $< X_1, \dots, X_n > \in range(X_1) \times \dots \times range(X_n)$
- <u>Assignment</u>: $X = x$

Possible World Semantics

- A specific world ω entails $X = x$ means that in that world, X is assigned x
- Axioms: in the slide

Semantics of Probability: Finite Case

For a finite number of possible worlds:

- $\mu(\omega) > 0$ is a measure such that this measure for every possible world sums up to 1
- it specifies how much an agent thinks the world ω is like the real world
- the probability of proposition f is: $\Pr[f] = \sum_{\omega \text{ } entails \text{ } f} \mu(\omega)$

For instance, the 1 ball 3 cups game, say:

- $\Pr[ball \text{ } under \text{ } C_1] = 0.2$
- …

Where $ball \text{ } under \text{ } C_1$ is one of 3 worlds in this example.

Axioms of Probability

- $\forall f: \Pr[f] \geq 0$
- $\Pr[\tau] = 1$ iff τ is a tautology
- $\Pr[f \vee g] = \Pr[f] + \Pr[g]$ if $\sim(f \vee g)$ is a tautology

Semantics of Probability: General Case:

In the general case, we define $\mu(\omega)$ as a measurement of sets of worlds $S \subseteq \Omega$:

- $\mu(S) \geq 0$
- $\mu(\Omega) = 1$
- The rest is in the slide

Probability Distributions

- A distribution of a random var X is a function $range(X) \to [0,1]$ that maps a value to a corresponding probability.
- If the range is infinite, this is called a density function

Conditioning:

- Probabilistic conditioning specifies how to revise beliefs based on new information
- Prior probability: the probabilistic model built based on all background information
- Posterior probability of h: the conditional probability $P(h|e)$ where e is evidence obtained subsequently

Semantics of Conditional Probability:

- Evidence e rules out worlds incompatible with it
- So we have a new measure μ_e:

$$\mu_e = \begin{cases} c \times \mu(S), if\ \omega\ entails\ e\ \forall \omega \in S \\ 0, \quad otherwise \end{cases}$$

Where c is scaling w.r.t. the evidence. What is c?

$$c = \frac{1}{\Pr[e]}$$

This derives from Bayes rule: $P(h|e) = \mu_e(\{\omega | \omega\ entails\ h\}) = P(h \wedge e)/P(e)$

Chain Rule

$$P(f_1 \wedge \dots \wedge f_n) = P(f_n | f_1 \wedge \dots \wedge f_{n-1}) \times P(f_1 \wedge \dots \wedge f_{n-1}) = \cdots = \prod_{i=1}^{n} P(f_i | f_1 \wedge \dots \wedge f_{i-1})$$

Bayes Theorem

$P(h \wedge e) = P(h|e) \times P(e) = P(e|h) \times P(h)$

And when $P(e) \neq 0$ we divide by $P(e)$ and get:

$$\boxed{P(h|e) = P(e|h) \times P(h) / P(e)}$$