

*It is remarkable that a science which began with the consideration of games of chance should become the most important object of human knowledge ... The most important questions of life are, for the most part, really only problems of probability ...*

*The theory of probabilities is at bottom nothing but common sense reduced to calculus.*

*– Pierre Simon de Laplace [1812]*

At the end of the class you should be able to:

- justify the use and semantics of probability
- know how to compute marginals and apply Bayes' theorem
- identify conditional independence
- build a belief network for a domain

# Using Uncertain Knowledge

- Agents don't have complete knowledge about the world.
- Agents need to make (informed) decisions given their uncertainty.
- It isn't enough to assume what the world is like.  
**Example:** wearing a seat belt.
- An agent needs to reason about its uncertainty.

# Why Probability?

- There is lots of uncertainty about the world, but agents still need to act.
- Predictions are needed to decide what to do:
  - ▶ definitive predictions: you will (not) be run over tomorrow
  - ▶ point probabilities: probability you will be run over tomorrow is 0.002
  - ▶ probability ranges: you will be run over with probability in range  $[0.001, 0.34]$
- Acting is gambling: agents who don't use probabilities will lose to those who do — Dutch books.
- Probabilities can be learned from data.  
Bayes' rule specifies how to combine data and prior knowledge.

# Numerical Measures of Belief

- Belief in proposition,  $f$ , can be measured in terms of a number between 0 and 1 — this is the **probability of  $f$** .
  - ▶ The probability  $f$  is 0 means that  $f$  is believed to be definitely false.
  - ▶ The probability  $f$  is 1 means that  $f$  is believed to be definitely true.
- Using 0 and 1 is purely a convention.
- $f$  has a probability between 0 and 1, means the agent is ignorant of its truth value.
- Probability is a measure of an agent's ignorance.
- Probability is *not* a measure of degree of truth.

# Possible World Semantics

- $\Omega$  is a (possibly infinite) set of possible worlds.
- A **random variable** is a function possible worlds
- A random variable with a countable (or finite) range is a **discrete** random variable.
- A variable with range  $\{true, false\}$  is a **Boolean** random variable.
- A variable with range the real numbers is a **continuous** random variable.
- $\omega \models X = v$  means variable  $X$  has value  $v$  in world  $\omega$ .
- $\omega \models X > v$  means variable  $X$  is greater than  $v$  in world  $\omega$ .
- Logical connectives have their standard meaning:
  - $\omega \models \alpha \wedge \beta$  if  $\omega \models \alpha$  and  $\omega \models \beta$
  - $\omega \models \alpha \vee \beta$  if  $\omega \models \alpha$  or  $\omega \models \beta$
  - $\omega \models \neg\alpha$  if  $\omega \not\models \alpha$

(Sometimes the range of a random variable is called its **domain** – it is the domain of a function on random variables.)

- Probability defines a measure on sets of possible worlds.  
(Not all sets have measures; just those that can be described.)
- A **probability measure** is a function  $\mu$  from sets of worlds into the non-negative real numbers such that:

- ▶  $\mu(\Omega) = 1$

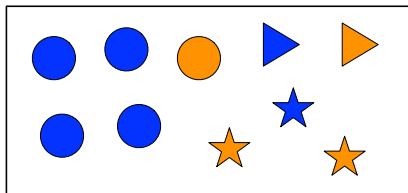
- ▶  $\mu(S_1 \cup S_2) = \mu(S_1) + \mu(S_2)$   
if  $S_1 \cap S_2 = \{\}$ .

Extended to countable unions ( $\sigma$ -additivity):

$$\mu\left(\bigcup_i S_i\right) = \sum_i \mu(S_i) \text{ if } S_i \cap S_j = \{\} \text{ for } i \neq j$$

- $P(\alpha) = \mu(\{\omega \mid \omega \models \alpha\})$ .

Possible Worlds:



Suppose the measure of each singleton world is 0.1.

- What is the probability of circle?
- What us the probability of star?
- What is the probability of orange?
- What is the probability of orange and star?
- What is the probability of orange and circle?

Note that  $P(\alpha \wedge \beta)$  is **not** a function of  $P(\alpha)$  and  $P(\beta)$ .



# Expected Value

- The **expected value** of numerical random variable  $X$  with respect to probability  $P$  is

$$\mathbb{E}_P(X) = \sum_{v \in \text{domain}(X)} v * P(X=v)$$

when the domain of  $X$  is finite or countable.

- When the domain is continuous, the sum becomes an integral.
- If  $\alpha$  is a proposition, treating *true* as 1 and *false* as 0:

$$\mathbb{E}_P(\alpha) = P(\alpha)$$

- Probabilistic conditioning specifies how to revise beliefs based on new information.
- An agent builds a probabilistic model taking all background information into account.  
This gives the **prior probability**.
- All other information must be conditioned on.
- If **evidence**  $e$  is the all of the information obtained subsequently, the **conditional probability**  $P(h | e)$  of  $h$  given  $e$  is the **posterior probability** of  $h$ .

# Semantics of Conditional Probability

- Evidence  $e$  rules out possible worlds incompatible with  $e$ .
- Evidence  $e$  induces a new measure,  $\mu_e$ , over possible worlds:

$$\mu_e(S) = \begin{cases} c * \mu(S) & \text{if } \omega \models e \text{ for all } \omega \in S \\ 0 & \text{if } \omega \not\models e \text{ for all } \omega \in S \end{cases}$$

To derive  $c$ :

$$\begin{aligned} \mu_e(\Omega) &= \mu_e(\{\omega \mid \omega \models e\}) + \mu_e(\{\omega \mid \omega \not\models e\}) \\ &= c * P(e) \end{aligned}$$

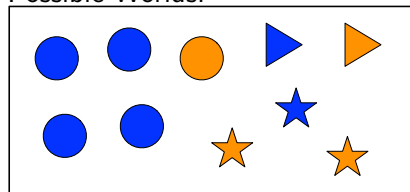
therefore  $c = \frac{1}{P(e)}$ .

- The conditional probability of formula  $h$  given evidence  $e$  is

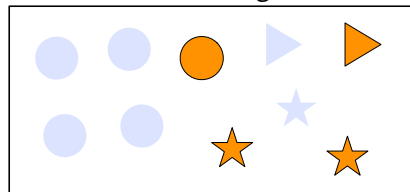
$$\begin{aligned} P(h \mid e) &= \mu_e(\{\omega : \omega \models h\}) \\ &= \frac{P(h \wedge e)}{P(e)} \end{aligned}$$

# Conditioning

Possible Worlds:



Observe  $Color=orange$ :



$$P(\text{Shape}=\textit{star}) = 0.3$$

$$P(\text{Shape}=\textit{circle}) = 0.5$$

$$P(\text{Shape}=\textit{star} \mid \\ \text{Color}=\textit{orange}) = 0.5$$

$$P(\text{Shape}=\textit{circle} \mid \\ \text{Color}=\textit{orange}) = 0.25$$

## Chain Rule: probability of conjunctions

$$P(h | e) = \frac{P(h \wedge e)}{P(e)}$$

Therefore

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

Semantics of conditioning gives:  $P(h \wedge e) = P(h | e) * P(e)$

$$\begin{aligned} & P(f_n \wedge f_{n-1} \wedge \dots \wedge f_1) \\ &= P(f_n | f_{n-1} \wedge \dots \wedge f_1) * \\ & \quad P(f_{n-1} \wedge \dots \wedge f_1) \\ &= P(f_n | f_{n-1} \wedge \dots \wedge f_1) * \\ & \quad P(f_{n-1} | f_{n-2} \wedge \dots \wedge f_1) * \\ & \quad P(f_{n-2} \wedge \dots \wedge f_1) \\ &= P(f_n | f_{n-1} \wedge \dots \wedge f_1) * \\ & \quad P(f_{n-1} | f_{n-2} \wedge \dots \wedge f_1) \\ & \quad * \dots * P(f_3 | f_2 \wedge f_1) * P(f_2 | f_1) * P(f_1) \\ &= \prod_{i=1}^n P(f_i | f_1 \wedge \dots \wedge f_{i-1}) \end{aligned}$$

# Bayes' theorem

The chain rule and commutativity of conjunction ( $h \wedge e$  is equivalent to  $e \wedge h$ ) gives us:

$$\begin{aligned}P(h \wedge e) &= P(h | e) * P(e) \\ &= P(e | h) * P(h).\end{aligned}$$

If  $P(e) \neq 0$ , divide the right hand sides by  $P(e)$ :

$$P(h | e) = \frac{P(e | h) * P(h)}{P(e)}.$$

This is **Bayes' theorem**.

# Why is Bayes' theorem interesting?

- Often you have causal knowledge:

$$P(\textit{symptom} \mid \textit{disease})$$

$$P(\textit{light is off} \mid \textit{status of switches and switch positions})$$

$$P(\textit{alarm} \mid \textit{fire})$$

$$P(\textit{image looks like } \img alt="tree icon" data-bbox="380 440 415 495" \mid \textit{a tree is in front of a car})$$

- and want to do evidential reasoning:

$$P(\textit{disease} \mid \textit{symptom})$$

$$P(\textit{status of switches} \mid \textit{light is off and switch positions})$$

$$P(\textit{fire} \mid \textit{alarm})$$

$$P(\textit{a tree is in front of a car} \mid \textit{image looks like } \img alt="tree icon" data-bbox="735 715 770 770" )$$