"The mind is a neural computer, fitted by natural selection with combinatorial algorithms for causal and probabilistic reasoning about plants, animals, objects, and people.

"In a universe with any regularities at all, decisions informed about the past are better than decisions made at random. That has always been true, and we would expect organisms, especially informavores such as humans, to have evolved acute intuitions about probability. The founders of probability, like the founders of logic, assumed they were just formalizing common sense."

Steven Pinker, How the Mind Works, 1997, pp. 524, 343.

Learning Objectives

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
- build a belief network for a domain
- predict the inferences for a belief network
- explain the predictions of a causal model

Using Uncertain Knowledge

- Agents don't have complete knowledge about the world.
- Agents need to make decisions based on 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 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.

Probability

- Probability is an agent's measure of belief in some proposition — subjective probability.
- An agent's belief depends on its prior assumptions and what the agent observes.

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.

Random Variables

- A random variable is a term in a language that can take one of a number of different values.
- The range of a variable X, written $range(X)$, is the set of values X can take.
- A tuple of random variables $\langle X_1, \ldots, X_n \rangle$ is a complex random variable with range $range(X_1) \times \cdots \times range(X_n)$. Often the tuple is written as $X_1, \ldots, X_n$.
- Assignment $X = x$ means variable X has value x.
- A proposition is a Boolean formula made from assignments of values to variables.

Possible World Semantics

- A possible world specifies an assignment of one value to each random variable.
- A random variable is a function from possible worlds into the range of the random variable.
- $\omega \models X = x$
  means variable X is assigned value x in world ω.
- Logical connectives have their standard meaning:

$$\omega \models \alpha \wedge \beta \text{ if } \omega \models \alpha \text{ and } \omega \models \beta$$

$$\omega \models \alpha \vee \beta \text{ if } \omega \models \alpha \text{ or } \omega \models \beta$$

$$\omega \models \neg\alpha \text{ if } \omega \not\models \alpha$$

- Let Ω be the set of all possible worlds.

Semantics of Probability

For a finite number of possible worlds:

- Define a nonnegative measure $\mu(\omega)$ to each world ω so that the measures of the possible worlds sum to 1.
- The probability of proposition f is defined by:

$$P(f) = \sum_{\omega \models f} \mu(\omega).$$

Axioms of Probability: finite case

Three axioms define what follows from a set of probabilities:

Axiom 1 $0 \leq P(a)$ for any proposition a.

Axiom 2 $P(true) = 1$

Axiom 3 $P(a \vee b) = P(a) + P(b)$ if a and b cannot both be true.

- These axioms are sound and complete with respect to the semantics.

Semantics of Probability: general case

In the general case, probability defines a measure on sets of possible worlds. We define $\mu(S)$ for some sets $S \subseteq \Omega$ satisfying:

- $\mu(S) \geq 0$
- $\mu(\Omega) = 1$
- $\mu(S_1 \cup S_2) = \mu(S_1) + \mu(S_2)$ if $S_1 \cap S_2 = \{\}$.
  Or sometimes σ-additivity:

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

Then $P(\alpha) = \mu(\{\omega | \omega \models \alpha\})$.

Probability Distributions

- A probability distribution on a random variable X is a function $range(X) \rightarrow [0, 1]$ such that

$$x \mapsto P(X = x).$$

  This is written as $P(X)$.
- This also includes the case where we have tuples of variables. E.g., $P(X, Y, Z)$ means $P(\langle X, Y, Z \rangle)$.
- When $range(X)$ is infinite sometimes we need a probability density function...

Conditioning

- 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, μ_e, over possible worlds

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

We can show $c =$

Semantics of Conditional Probability

- Evidence e rules out possible worlds incompatible with e.
- Evidence e induces a new measure, μ_e, over possible worlds

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

  We can show $c = \frac{1}{P(e)}$.
- The conditional probability of formula h given evidence e is

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

Semantics of Conditional Probability

- Evidence e rules out possible worlds incompatible with e.
- Evidence e induces a new measure, μ_e, over possible worlds

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

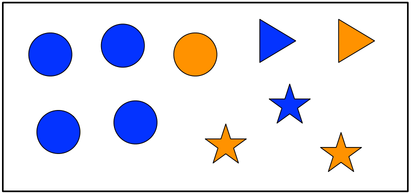
  We can show $c = \frac{1}{P(e)}$.

- The conditional probability of formula h given evidence e is

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

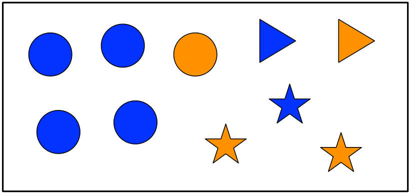
Conditioning

Possible Worlds:

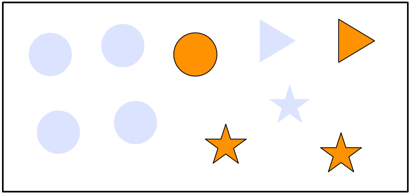


Conditioning

Possible Worlds:



Observe *Color* $=$ *orange*:



Exercise

Flu	$Sneeze$	$Snore$	μ
true	true	true	0.064
true	true	false	0.096
true	false	true	0.016
true	false	false	0.024
false	true	true	0.096
false	true	false	0.144
false	false	true	0.224
false	false	false	0.336

What is:

(a) $P(flu \wedge sneeze)$

(b) $P(flu \wedge \neg sneeze)$

(c) $P(flu)$

(d) $P(sneeze \mid flu)$

(e) $P(\neg flu \wedge sneeze)$

(f) $P(flu \mid sneeze)$

(g) $P(sneeze \mid flu \wedge snore)$

(h) $P(flu \mid sneeze \wedge snore)$

Chain Rule

$$P(f_1 \wedge f_2 \wedge \ldots \wedge f_n)$$

$$=$$

Chain Rule

$$
\begin{aligned}
& P(f_1 \wedge f_2 \wedge \ldots \wedge f_n) \\
& \quad = \quad P(f_n | f_1 \wedge \cdots \wedge f_{n-1}) \times \\
& \qquad\qquad P(f_1 \wedge \cdots \wedge f_{n-1}) \\
& \quad =
\end{aligned}
$$

Chain Rule

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

Bayes' theorem

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

$$P(h \wedge e) \quad =$$

Bayes' theorem

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

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

Bayes' theorem

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

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

Bayes' theorem

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

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

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

$$P(h|e) =$$

Bayes' theorem

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

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

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

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

This is Bayes' theorem.

Why is Bayes' theorem interesting?

- Often you have causal knowledge:
  $P(\mathit{symptom} \mid \mathit{disease})$
  $P(\textit{light is off} \mid \textit{status of switches and switch positions})$
  $P(\mathit{alarm} \mid \mathit{fire})$
  $P($*image looks like*  $\mid$ *a tree is in front of a car*$)$
- and want to do evidential reasoning:
  $P(\mathit{disease} \mid \mathit{symptom})$
  $P(\textit{status of switches} \mid \textit{light is off and switch positions})$
  $P(\mathit{fire} \mid \mathit{alarm})$.
  $P($*a tree is in front of a car* $\mid$ *image looks like* $)$

Exercise

A cab was involved in a hit-and-run accident at night. Two cab companies, the Green and the Blue, operate in the city. You are given the following data:

- 85% of the cabs in the city are Green and 15% are Blue.
- A witness identified the cab as Blue. The court tested the reliability of the witness in the circumstances that existed on the night of the accident and concluded that the witness correctly identifies each one of the two colours 80% of the time and failed 20% of the time.

What is the probability that the cab involved in the accident was Blue?

Exercise

A cab was involved in a hit-and-run accident at night. Two cab companies, the Green and the Blue, operate in the city. You are given the following data:

- 85% of the cabs in the city are Green and 15% are Blue.
- A witness identified the cab as Blue. The court tested the reliability of the witness in the circumstances that existed on the night of the accident and concluded that the witness correctly identifies each one of the two colours 80% of the time and failed 20% of the time.

What is the probability that the cab involved in the accident was Blue?

From D. Kahneman, Thinking Fast and Slow, 2011, p. 166.

Exercise

A cab was involved in a hit-and-run accident at night. Two cab companies, the Green and the Blue, operate in the city. You are given the following data:

- The two companies operate the same number of cabs, but Green cabs are involved in 85% of the accidents.
- A witness identified the cab as Blue. The court tested the reliability of the witness in the circumstances that existed on the night of the accident and concluded that the witness correctly identifies each one of the two colours 80% of the time and failed 20% of the time.

What is the probability that the cab involved in the accident was Blue?

Exercise

A cab was involved in a hit-and-run accident at night. Two cab companies, the Green and the Blue, operate in the city. You are given the following data:

- The two companies operate the same number of cabs, but Green cabs are involved in 85% of the accidents.
- A witness identified the cab as Blue. The court tested the reliability of the witness in the circumstances that existed on the night of the accident and concluded that the witness correctly identifies each one of the two colours 80% of the time and failed 20% of the time.

What is the probability that the cab involved in the accident was Blue?

From D. Kahneman, Thinking Fast and Slow, 2011, p. 167. Chapter 16 “Causes trump statistics”