"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
- identify conditional independence
- 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.

Using Uncertain Knowledge

- Agents don't have complete knowledge about the world.
- Agents need to make (informed) decisions given their uncertainty.

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.

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.

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.
- When an agent makes an action under uncertainty, it is gambling $\Longrightarrow$ probability.

Probability

- Probability is an agent's measure of belief in some proposition — subjective probability.

Probability

- Probability is an agent's measure of belief in some proposition — subjective probability.
- An agent's belief depends on its prior belief and what it observes.
- Example: An agent's probability of a particular bird flying
  - Other agents may have different probabilities
  - An agent's belief in a bird's flying ability is affected by what the agent knows about that bird.

Random Variables

- A random variable starts with upper case.
- The domain of a variable X, written $domain(X)$, is the set of values X can take. (Sometimes use "range", "frame", "possible values").

Random Variables

- A random variable starts with upper case.
- The domain of a variable X, written $domain(X)$, is the set of values X can take. (Sometimes use "range", "frame", "possible values").
- A tuple of random variables $\langle X_1, \ldots, X_n \rangle$ is a complex random variable with domain
  $domain(X_1) \times \cdots \times domain(X_n)$.
  Often the tuple is written as $X_1, \ldots, X_n$.

Random Variables

- A random variable starts with upper case.
- The domain of a variable X, written $domain(X)$, is the set of values X can take. (Sometimes use "range", "frame", "possible values").
- A tuple of random variables $\langle X_1, \dots, X_n \rangle$ is a complex random variable with domain
  $domain(X_1) \times \cdots \times domain(X_n)$.
  Often the tuple is written as $X_1, \dots, X_n$.
- Assignment $X = x$ means variable X has value x.

Random Variables

- A random variable starts with upper case.
- The domain of a variable X, written $domain(X)$, is the set of values X can take. (Sometimes use "range", "frame", "possible values").
- A tuple of random variables $\langle X_1, \ldots, X_n \rangle$ is a complex random variable with domain $domain(X_1) \times \cdots \times domain(X_n)$.
  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 or inequality (e.g., $<$, $\leq$,...) between variables and values.

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 domain of the random variable.

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 domain of the random variable.
- $\omega \models X = x$
  means variable X is assigned value x in world ω.

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 domain 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

Probability defines a measure on sets of possible worlds.
A probability measure is a function μ from sets of worlds into the non-negative real numbers such that:

- $\mu(\Omega) = 1$ Ω all worlds
- $\mu(S_1 \cup S_2) = \mu(S_1) + \mu(S_2)$
  if $S_1 \cap S_2 = \{\}$.

Semantics of Probability

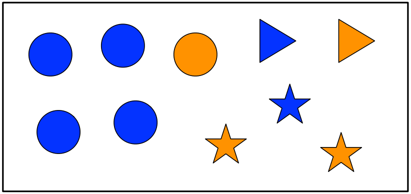
Probability defines a measure on sets of possible worlds.
A probability measure is a function μ from sets of worlds into the non-negative real numbers such that:

- $\mu(\Omega) = 1$
- $\mu(S_1 \cup S_2) = \mu(S_1) + \mu(S_2)$
  if $S_1 \cap S_2 = \{\}$.

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

Semantics

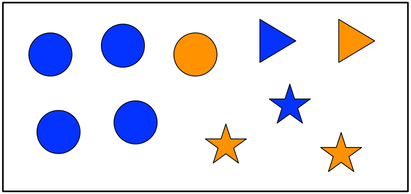
Possible Worlds:



Suppose the measure of each singleton world is 0.1.

Semantics

Possible Worlds:

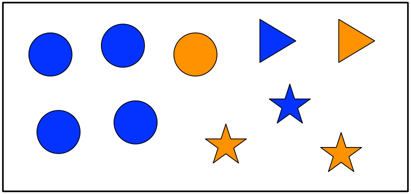


Suppose the measure of each singleton world is 0.1.

- What is the probability of circle? $5/_{10} = 0.5$

Semantics

Possible Worlds:

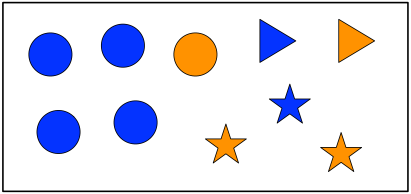


Suppose the measure of each singleton world is 0.1.

- What is the probability of circle?
- What us the probability of star? 3/10

Semantics

Possible Worlds:

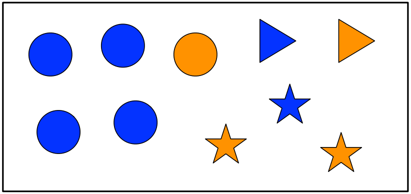


Suppose the measure of each singleton world is 0.1.

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

Semantics

Possible Worlds:

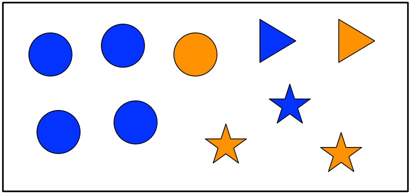


Suppose the measure of each singleton world is 0.1.

- 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? 2/10

Semantics

Possible Worlds:

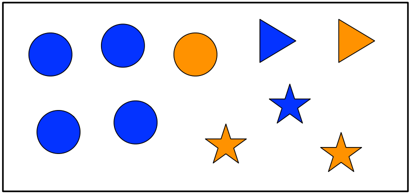


Suppose the measure of each singleton world is 0.1.

- 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? 1/10

Semantics

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 $P(\beta)$.

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.

Conditioning

- Probabilistic conditioning specifies how to revise beliefs based on new information.

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 $P(h \mid e)$ of h given e is the posterior probability of h.

Semantics of Conditional Probability

- Evidence e rules out possible worlds incompatible with e.

Semantics of Conditional Probability

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

Semantics of Conditional Probability

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

$$\mu_e(S) = \left\{ \right.$$

Semantics of Conditional Probability

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

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

e is false in all worlds in S.

Semantics of Conditional Probability

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

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

Semantics of Conditional Probability

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

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

Semantics of Conditional Probability

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

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

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

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

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

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

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

Semantics of Conditional Probability

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

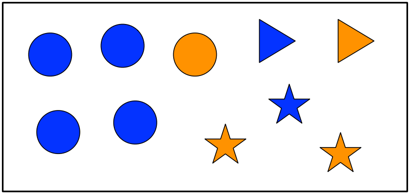
$$\mu_e(S) = \begin{cases} c \times \mu(S) & \text{if } \omega \models e \text{ for all } \omega \in S \\ 0 & \text{if } \omega \not\models e \text{ for all } \omega \in S \end{cases}$$

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

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

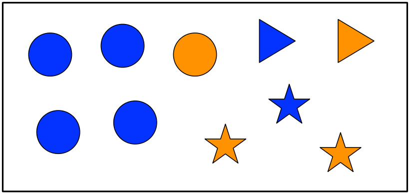
Conditioning

Possible Worlds:



Conditioning

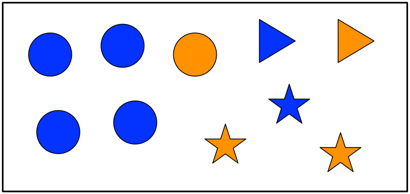
Possible Worlds:



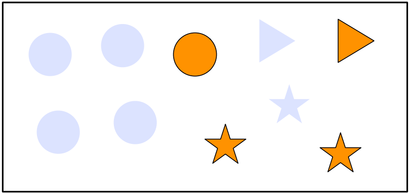
Observe *Color*=*orange*:

Conditioning

Possible Worlds:

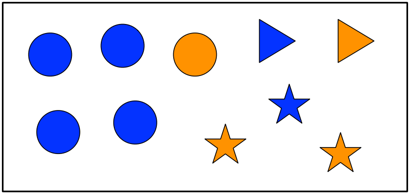


Observe *Color*=*orange*:

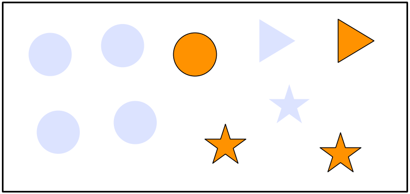


Conditioning

Possible Worlds:



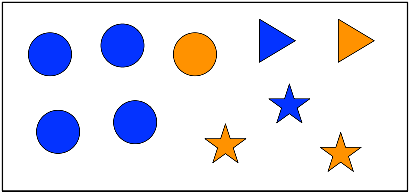
Observe *Color*=*orange*:



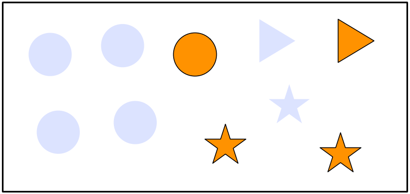
$P(Shape{=}star \mid Color{=}orange) = 0.5$

Conditioning

Possible Worlds:



Observe $Color{=}orange$:



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

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

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) $P(flu)$ top c

(d) $P(sneeze \mid flu)$ a/c

(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

Semantics of conditioning gives: $P(h \wedge e) = P(h \mid e) \times P(e)$

Chain Rule

Semantics of conditioning gives: $P(h \land e) = P(h \mid e) \times P(e)$

$$P(\overset{h}{f_n} \land \overbrace{f_{n-1} \land \ldots \land f_1}^{e})$$

$$=$$

Chain Rule

Semantics of conditioning gives: $P(h \wedge e) = P(h \mid e) \times P(e)$

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

Chain Rule

Semantics of conditioning gives: $P(h \wedge e) = P(h \mid e) \times P(e)$

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

Bayes' theorem

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

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

Bayes' theorem

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

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

Bayes' theorem

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

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

Bayes' theorem

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

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

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

$$P(h \mid e) =$$

Bayes' theorem

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

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

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

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

This is Bayes' theorem.

Why is Bayes' theorem interesting?

- Often you have causal knowledge:
  $P(symptom \mid disease)$

Why is Bayes' theorem interesting?

- Often you have causal knowledge:
  $P(symptom \mid disease)$
- and want to do evidential reasoning:

Why is Bayes' theorem interesting?

- Often you have causal knowledge:
  $P(symptom \mid disease)$
- and want to do evidential reasoning:
  $P(disease \mid symptom)$

Why is Bayes' theorem interesting?

- Often you have causal knowledge:
  $P(symptom \mid disease)$
  $P(light\ is\ off \mid status\ of\ switches\ and\ switch\ positions)$

- and want to do evidential reasoning:
  $P(disease \mid symptom)$

Why is Bayes' theorem interesting?

- Often you have causal knowledge:
  $P(symptom \mid disease)$
  $P(light\ is\ off \mid status\ of\ switches\ and\ switch\ positions)$

- and want to do evidential reasoning:
  $P(disease \mid symptom)$
  $P(status\ of\ switches \mid light\ is\ off\ and\ switch\ positions)$

Why is Bayes' theorem interesting?

- Often you have causal knowledge:
  $P(symptom \mid disease)$
  $P(light\ is\ off \mid status\ of\ switches\ and\ switch\ positions)$
  $P(alarm \mid fire)$

- and want to do evidential reasoning:
  $P(disease \mid symptom)$
  $P(status\ of\ switches \mid light\ is\ off\ and\ switch\ positions)$

Why is Bayes' theorem interesting?

- Often you have causal knowledge:
  $P(symptom \mid disease)$
  $P(light\ is\ off \mid status\ of\ switches\ and\ switch\ positions)$
  $P(alarm \mid fire)$

- and want to do evidential reasoning:
  $P(disease \mid symptom)$
  $P(status\ of\ switches \mid light\ is\ off\ and\ switch\ positions)$
  $P(fire \mid alarm)$

Why is Bayes' theorem interesting?

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

Why is Bayes' theorem interesting?

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

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