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]

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

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.

Why Probability?

- There is lots of uncertainty about the world, but agents still need to act.

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]

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.

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.

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.

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.

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

- Ω is a (possibly infinite) set of possible worlds.
- A random variable is

Possible World Semantics

- Ω is a (possibly infinite) set of possible worlds.
- A random variable is a function possible worlds

Possible World Semantics

- Ω 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

Possible World Semantics

- Ω 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.

Possible World Semantics

- Ω 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

Possible World Semantics

- Ω 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.

Possible World Semantics

- Ω 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

Possible World Semantics

- Ω 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.

Possible World Semantics

- Ω 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 ω.

Possible World Semantics

- Ω 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 ω.

Possible World Semantics

- Ω 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 ω.
- $\omega \models X > v$ means variable X is greater than v in world ω.

Possible World Semantics

- Ω 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 ω.
- $\omega \models X > v$ means variable X is greater than v in world ω.
- Logical connectives have their standard meaning:

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

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

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

Possible World Semantics

- Ω 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 ω.
- $\omega \models X > v$ means variable X is greater than v in world ω.
- Logical connectives have their standard meaning:

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

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

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

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

Semantics of Probability

- Probability defines a measure on sets of possible worlds.
  (Not all sets have measures; just those than can be described.)

Semantics of Probability

- Probability defines a measure on sets of possible worlds. (Not all sets have measures; just those than can be described.)
- 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 $S_1 \cap S_2 = \{\}$.

Semantics of Probability

- Probability defines a measure on sets of possible worlds. (Not all sets have measures; just those than can be described.)
- 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 $S_1 \cap S_2 = \{\}$.
    Extended to countable unions (σ-additivity):

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

Semantics of Probability

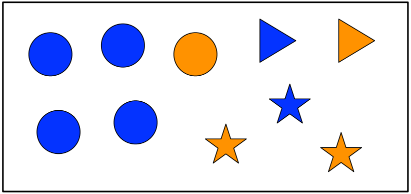
- Probability defines a measure on sets of possible worlds. (Not all sets have measures; just those than can be described.)
- A probability measure is a function μ from sets of worlds into the non-negative real numbers such that:
  - $\mu(\Omega) = 1$
  - $\mu(S_1 \cup S_2) = \mu(S_1) + \mu(S_2)$ if $S_1 \cap S_2 = \{\}$. Extended to countable unions (σ-additivity):

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

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

Semantics

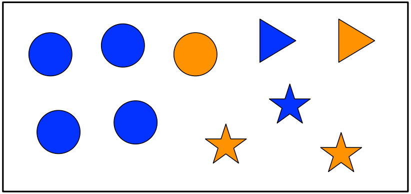
Possible Worlds:



Suppose the measure of each singleton world is 0.1.

Semantics

Possible Worlds:

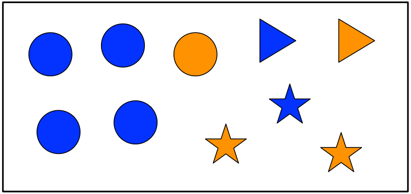


Suppose the measure of each singleton world is 0.1.

- What is the probability of circle?

Semantics

Possible Worlds:

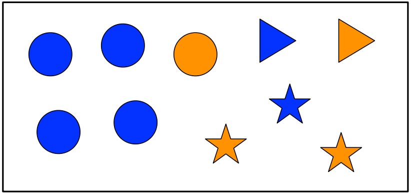


Suppose the measure of each singleton world is 0.1.

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

Semantics

Possible Worlds:

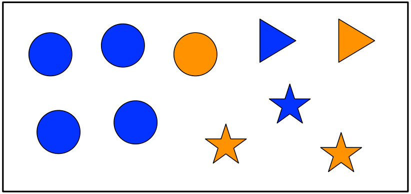


Suppose the measure of each singleton world is 0.1.

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

Semantics

Possible Worlds:

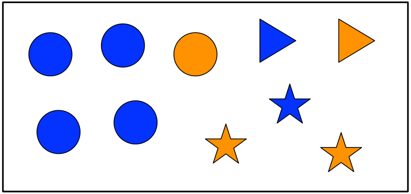


Suppose the measure of each singleton world is 0.1.

- 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?

Semantics

Possible Worlds:

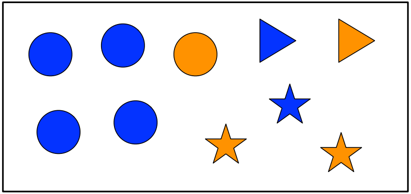


Suppose the measure of each singleton world is 0.1.

- 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?

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 $P(\beta)$.

Expected Value

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

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

when the domain is X is finite or countable.

Expected Value

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

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

  when the domain is X is finite or countable.
- When the domain is continuous, the sum becomes an integral.

Expected Value

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

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

when the domain is X is finite or countable.

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

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

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.

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:

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

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:

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

Semantics of Conditional Probability

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

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

To derive c:

Semantics of Conditional Probability

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

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

To derive c:

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

Semantics of Conditional Probability

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

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

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

Semantics of Conditional Probability

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

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

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 $c =$

Semantics of Conditional Probability

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

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

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

Semantics of Conditional Probability

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

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

  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

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

Semantics of Conditional Probability

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

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

  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

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

Semantics of Conditional Probability

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

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

  To derive c:

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

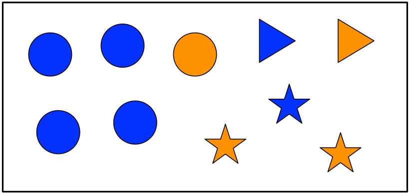
  therefore $c = \frac{1}{P(e)}$.

- 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:

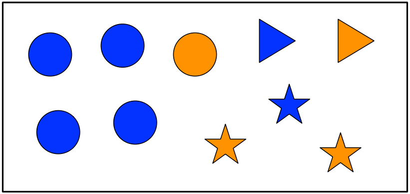


$P(Shape{=}star) = 0.3$

$P(Shape{=}circle) = 0.5$

Conditioning

Possible Worlds:



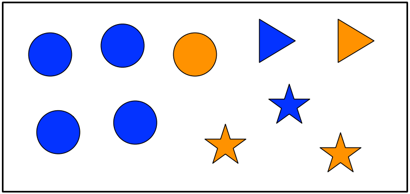
Observe $Color{=}orange$:

$P(Shape{=}star) = 0.3$

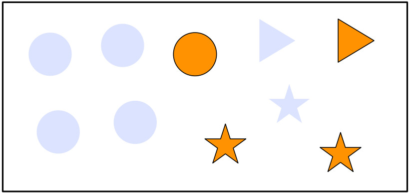
$P(Shape{=}circle) = 0.5$

Conditioning

Possible Worlds:



Observe $Color{=}orange$:

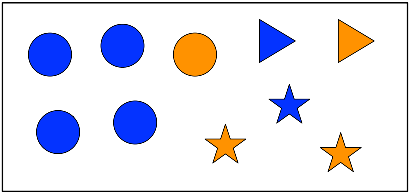


$P(Shape{=}star) = 0.3$

$P(Shape{=}circle) = 0.5$

Conditioning

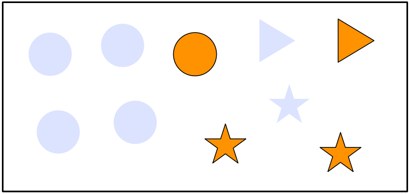
Possible Worlds:



$$P(Shape{=}star) = 0.3$$

$$P(Shape{=}circle) = 0.5$$

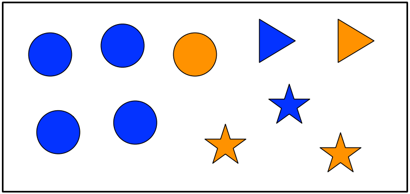
Observe *Color=orange*:



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

Conditioning

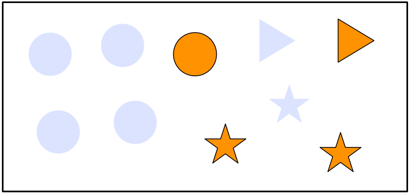
Possible Worlds:



$P(Shape{=}star) = 0.3$

$P(Shape{=}circle) = 0.5$

Observe *Color=orange*:



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

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

Clicker Question

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) $P(flu \wedge sneeze)$

A: 0.04
B: 0.16
C: 0.24
D: 0.4
E: 0.8

Clicker Question

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

A: 0.04
B: 0.16
C: 0.24
D: 0.4
E: 0.8

What is:

(a) $P(flu \wedge sneeze)$ 0.16

Clicker Question

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

A: 0.04
B: 0.16
C: 0.24
D: 0.4
E: 0.8

What is:

(a) $P(flu \wedge sneeze)$ 0.16

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

Clicker Question

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

A: 0.04
B: 0.16
C: 0.24
D: 0.4
E: 0.8

What is:

(a) $P(flu \wedge sneeze)$ 0.16

(b) $P(flu \wedge \neg sneeze)$ 0.04

Clicker Question

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) $P(flu \wedge sneeze)$ 0.16

(b) $P(flu \wedge \neg sneeze)$ 0.04

(c) $P(flu)$ (not clicker)

A: 0.04
B: 0.16
C: 0.24
D: 0.4
E: 0.8

Clicker Question

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) $P(flu \wedge sneeze)$ 0.16

(b) $P(flu \wedge \neg sneeze)$ 0.04

(c) $P(flu)$ (not clicker) 0.2

A: 0.04
B: 0.16
C: 0.24
D: 0.4
E: 0.8

Clicker Question

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) $P(flu \land sneeze)$ 0.16

(b) $P(flu \land \neg sneeze)$ 0.04

(c) $P(flu)$ (not clicker) 0.2

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

A: 0.04
B: 0.16
C: 0.24
D: 0.4
E: 0.8

Clicker Question

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

A: 0.04
B: 0.16
C: 0.24
D: 0.4
E: 0.8

What is:

(a) $P(flu \wedge sneeze)$ 0.16

(b) $P(flu \wedge \neg sneeze)$ 0.04

(c) $P(flu)$ (not clicker) 0.2

(d) $P(sneeze \mid flu)$ 0.8

Clicker Question

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

A: 0.04
B: 0.16
C: 0.24
D: 0.4
E: 0.8

What is:

(a) $P(flu \wedge sneeze)$ 0.16

(b) $P(flu \wedge \neg sneeze)$ 0.04

(c) $P(flu)$ (not clicker) 0.2

(d) $P(sneeze \mid flu)$ 0.8

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

Clicker Question

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

A: 0.04
B: 0.16
C: 0.24
D: 0.4
E: 0.8

What is:

(a) $P(flu \wedge sneeze)$ 0.16

(b) $P(flu \wedge \neg sneeze)$ 0.04

(c) $P(flu)$ (not clicker) 0.2

(d) $P(sneeze \mid flu)$ 0.8

(e) $P(\neg flu \wedge sneeze)$ 0.24

Clicker Question

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

A: 0.04
B: 0.16
C: 0.24
D: 0.4
E: 0.8

What is:

(a) $P(flu \wedge sneeze)$ 0.16

(b) $P(flu \wedge \neg sneeze)$ 0.04

(c) $P(flu)$ (not clicker) 0.2

(d) $P(sneeze \mid flu)$ 0.8

(e) $P(\neg flu \wedge sneeze)$ 0.24

(f) $P(sneeze)$

Clicker Question

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) $P(flu \land sneeze)$ 0.16

(b) $P(flu \land \neg sneeze)$ 0.04

(c) $P(flu)$ (not clicker) 0.2

(d) $P(sneeze \mid flu)$ 0.8

(e) $P(\neg flu \land sneeze)$ 0.24

(f) $P(sneeze)$ 0.4

A: 0.04
B: 0.16
C: 0.24
D: 0.4
E: 0.8

Clicker Question

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

A: 0.04
B: 0.16
C: 0.24
D: 0.4
E: 0.8

What is:

(a) $P(flu \wedge sneeze)$ 0.16

(b) $P(flu \wedge \neg sneeze)$ 0.04

(c) $P(flu)$ (not clicker) 0.2

(d) $P(sneeze \mid flu)$ 0.8

(e) $P(\neg flu \wedge sneeze)$ 0.24

(f) $P(sneeze)$ 0.4

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

Clicker Question

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

A: 0.04
B: 0.16
C: 0.24
D: 0.4
E: 0.8

What is:

(a) $P(flu \wedge sneeze)$ 0.16

(b) $P(flu \wedge \neg sneeze)$ 0.04

(c) $P(flu)$ (not clicker) 0.2

(d) $P(sneeze \mid flu)$ 0.8

(e) $P(\neg flu \wedge sneeze)$ 0.24

(f) $P(sneeze)$ 0.4

(g) $P(flu \mid sneeze)$ 0.4

Clicker Question

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

A: 0.04
B: 0.16
C: 0.24
D: 0.4
E: 0.8

What is:

(a) $P(flu \wedge sneeze)$ 0.16

(b) $P(flu \wedge \neg sneeze)$ 0.04

(c) $P(flu)$ (not clicker) 0.2

(d) $P(sneeze \mid flu)$ 0.8

(e) $P(\neg flu \wedge sneeze)$ 0.24

(f) $P(sneeze)$ 0.4

(g) $P(flu \mid sneeze)$ 0.4

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

Clicker Question

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

A: 0.04
B: 0.16
C: 0.24
D: 0.4
E: 0.8

What is:

(a) $P(flu \wedge sneeze)$ 0.16

(b) $P(flu \wedge \neg sneeze)$ 0.04

(c) $P(flu)$ (not clicker) 0.2

(d) $P(sneeze \mid flu)$ 0.8

(e) $P(\neg flu \wedge sneeze)$ 0.24

(f) $P(sneeze)$ 0.4

(g) $P(flu \mid sneeze)$ 0.4

(h) $P(sneeze \mid flu \wedge snore)$ 0.8

Clicker Question

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

A: 0.04
B: 0.16
C: 0.24
D: 0.4
E: 0.8

What is:

(a) $P(flu \wedge sneeze)$ 0.16

(b) $P(flu \wedge \neg sneeze)$ 0.04

(c) $P(flu)$ (not clicker) 0.2

(d) $P(sneeze \mid flu)$ 0.8

(e) $P(\neg flu \wedge sneeze)$ 0.24

(f) $P(sneeze)$ 0.4

(g) $P(flu \mid sneeze)$ 0.4

(h) $P(sneeze \mid flu \wedge snore)$ 0.8

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

Clicker Question

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

A: 0.04
B: 0.16
C: 0.24
D: 0.4
E: 0.8

What is:

(a) $P(flu \wedge sneeze)$ 0.16

(b) $P(flu \wedge \neg sneeze)$ 0.04

(c) $P(flu)$ (not clicker) 0.2

(d) $P(sneeze \mid flu)$ 0.8

(e) $P(\neg flu \wedge sneeze)$ 0.24

(f) $P(sneeze)$ 0.4

(g) $P(flu \mid sneeze)$ 0.4

(h) $P(sneeze \mid flu \wedge snore)$ 0.8

(i) $P(flu \mid sneeze \wedge snore)$ 0.4

Chain Rule: probability of conjunctions

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

Therefore

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

Chain Rule: probability of conjunctions

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

Therefore

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

Chain Rule

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

Chain Rule

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

$$P(f_n \wedge f_{n-1} \wedge \ldots \wedge f_1)$$

Chain Rule

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

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

Chain Rule

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

$$\begin{aligned}
& P(f_n \wedge f_{n-1} \wedge \ldots \wedge f_1) \\
&\quad = \quad P(f_n \mid f_{n-1} \wedge \cdots \wedge f_1) * \\
&\qquad\qquad P(f_{n-1} \wedge \cdots \wedge f_1) \\
&\quad = \quad P(f_n \mid f_{n-1} \wedge \cdots \wedge f_1) * \\
&\qquad\qquad P(f_{n-1} \mid f_{n-2} \wedge \cdots \wedge f_1) * \\
&\qquad\qquad P(f_{n-2} \wedge \cdots \wedge f_1)
\end{aligned}$$

Chain Rule

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

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

Chain Rule

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

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

Bayes' theorem

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

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

Bayes' theorem

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

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

Bayes' theorem

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

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

Bayes' theorem

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

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

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:

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

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

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

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

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 colors 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.]