Learning Analytics

Dr. Bowen Hui
Computer Science
University of British Columbia Okanagan

No Theoretical Probability Overview

Image taken from http://www.junkjungle.com/

No Theoretical Probability Overview

- Assume you took STAT 230
(and maybe STAT 303, even better: STAT 309)
- Our focus:
  - How to model the problem
  - How to compute probability estimations
  - Apply probabilistic inference algorithms for complex models
- Bring in Stats as needed

Cheatsheet of Statistics

- Probability distribution
  - All values must sum up to 1.0
- Conditional probability: $Pr(b|a) = \frac{Pr(b,a)}{Pr(a)}$
- Product rule: $Pr(a,b) = Pr(a|b)Pr(b)$
- Sum-out rule (marginalization): $Pr(a) = \sum_b \Pr(a, b)$
  $= \sum_b \Pr(a|b) \Pr(b)$
- Chain rule: $Pr(abcd) = Pr(a|bcd)Pr(b|cd)Pr(c|d)Pr(d)$
  - Applies to any number of variables
- Bayes rule: $Pr(b|a) = \frac{Pr(a|b)Pr(b)}{Pr(a)}$

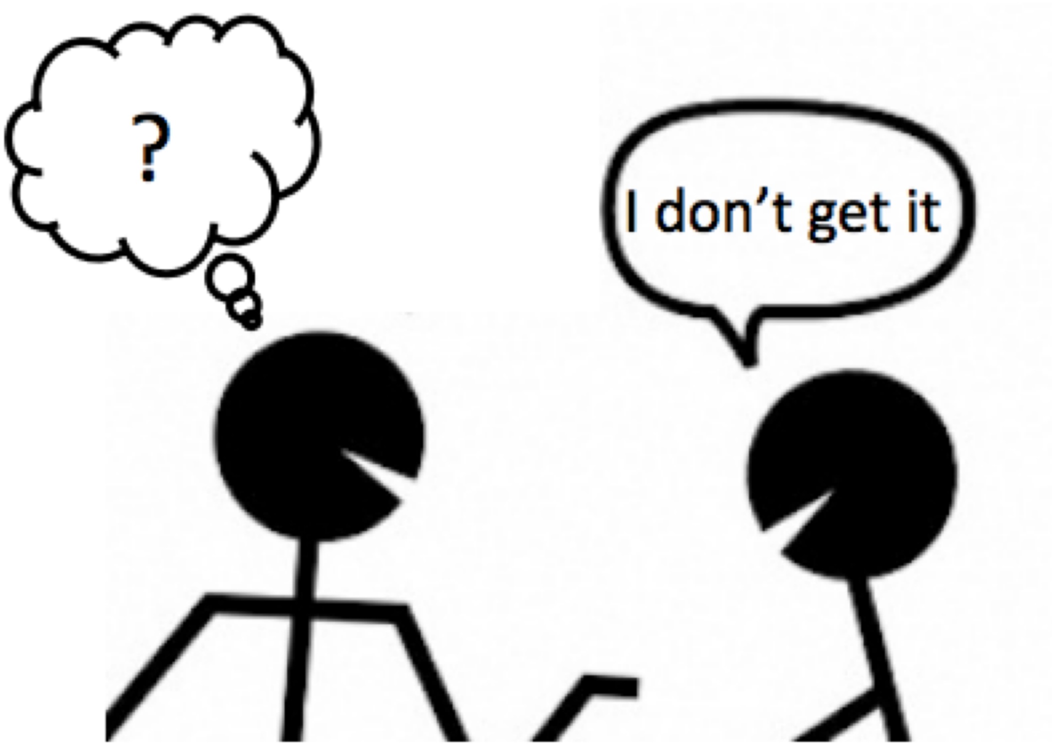
Short Exercise

- Given joint distribution of student's level of understanding and whether TA knows the answer:

	Understanding = High	Low
Know = Yes	0.3	0.4
Know = No	0.2	0.1

- Example: Pr(K=No, U=Low) = 0.1
- Compute:
  - Pr(Know = Yes, Understanding = High)
  - Pr(Know = Yes)
  - Pr(Know = No)

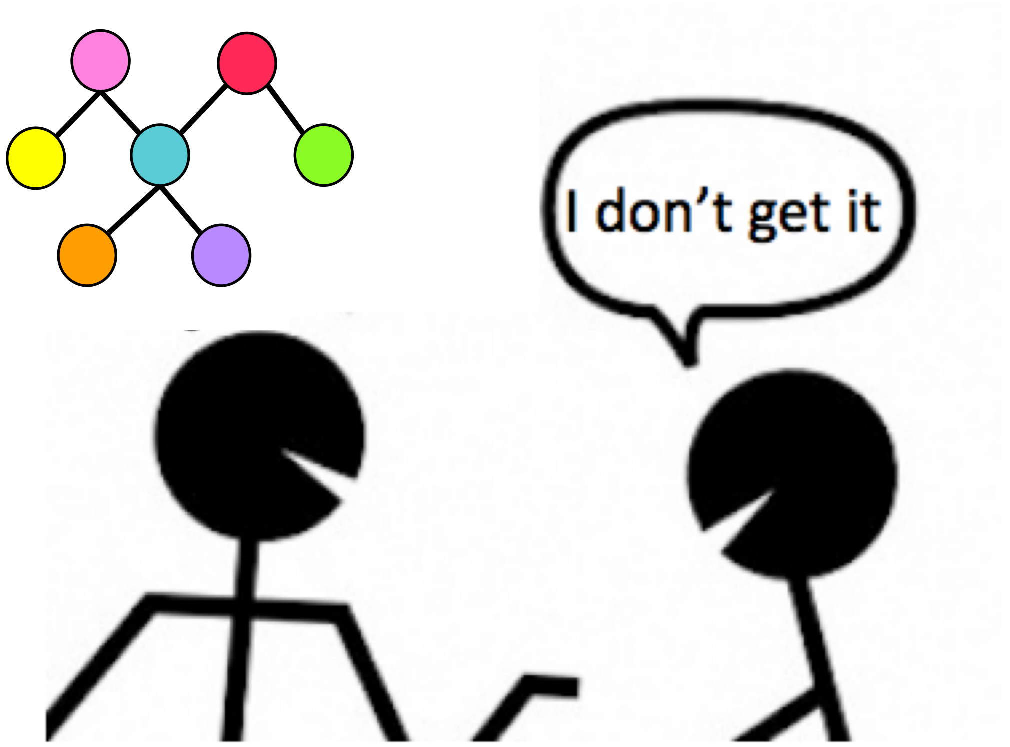
Multiagent Interaction



Person A Person B

What should A do?

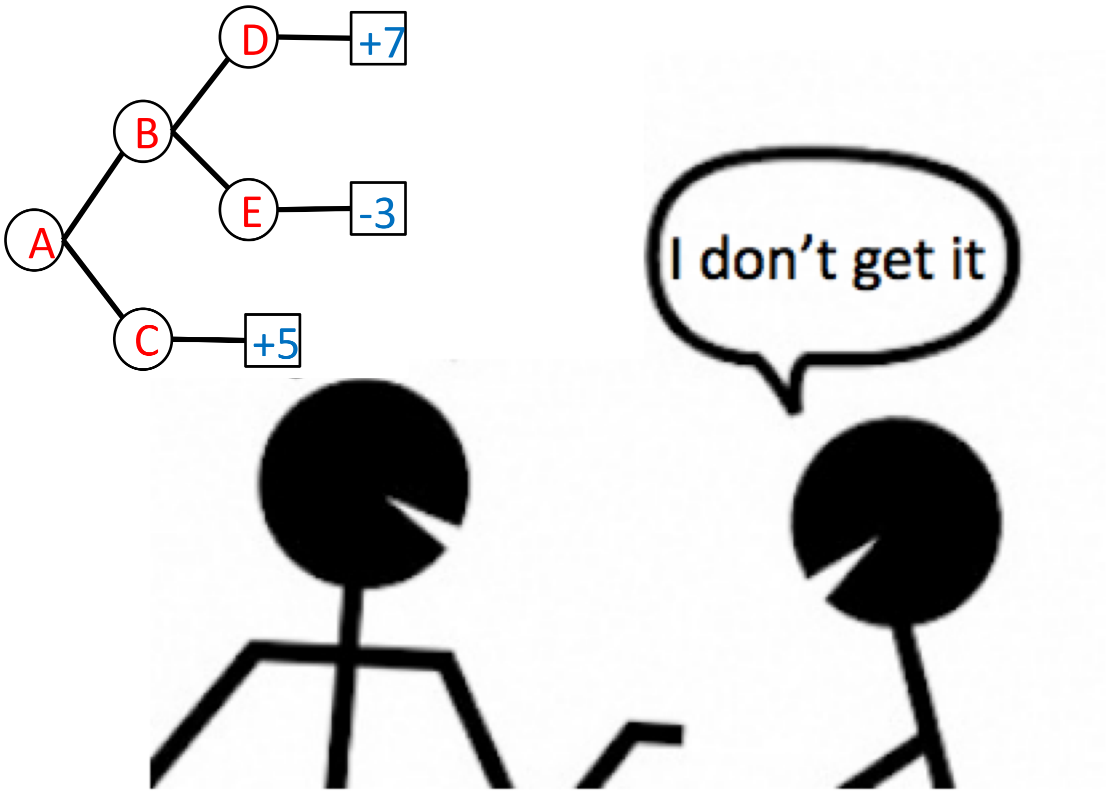
Multiagent Interaction



Person A Person B

Key note #1: Appropriate course of action is contingent upon the current state of affairs

Multiagent Interaction



Person A Person B

Key note #2: Decisions to act depend on the relative importance of competing objectives ("preferences")

Modeling the State in a Chess Game

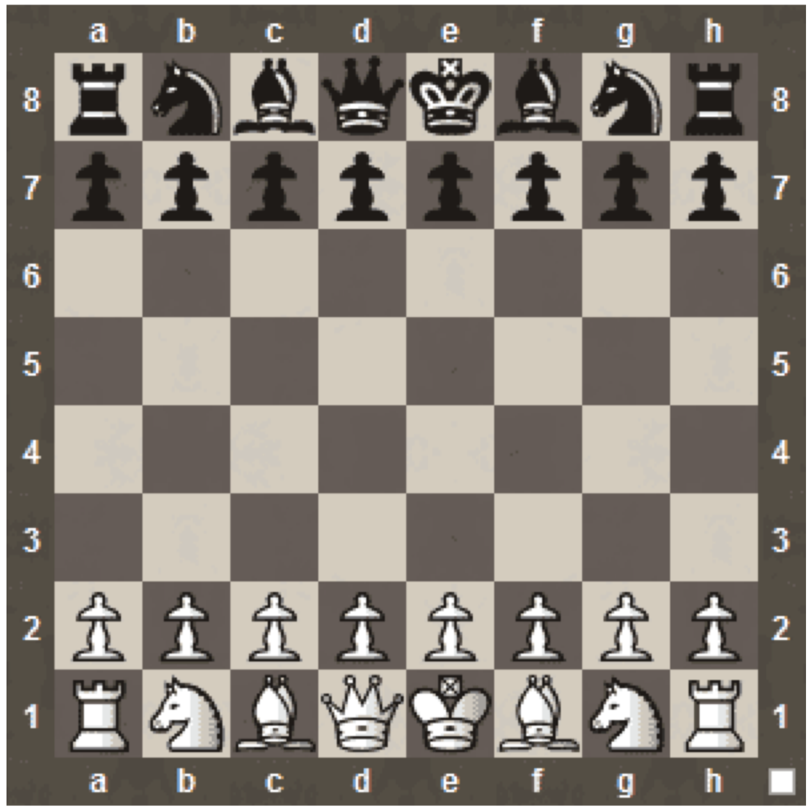


Image taken from computerchessonline.net

State of game: black horse at b7, etc.

Modeling Tic Tac Toe State

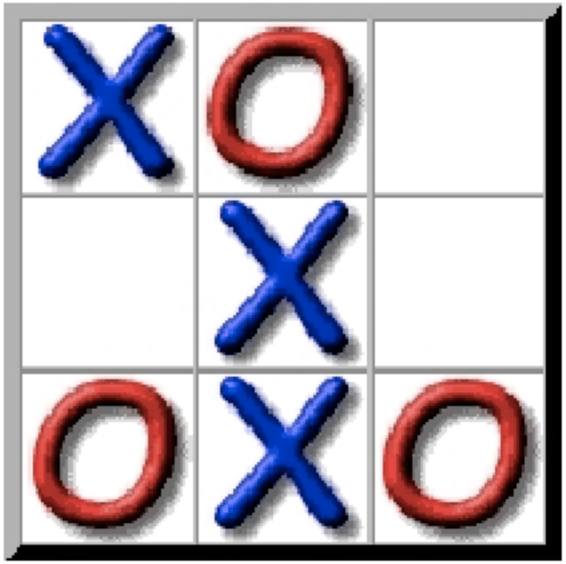
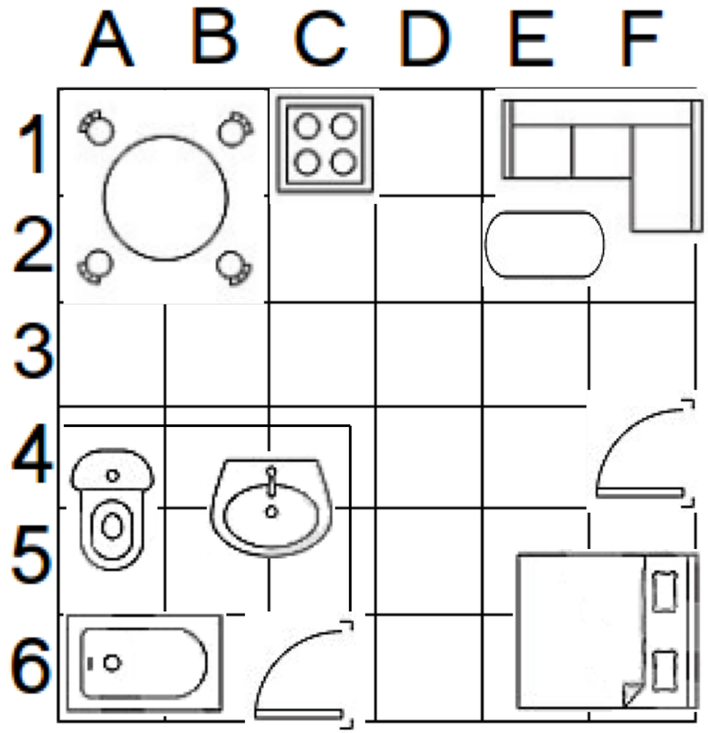


Image taken from www.mathematik.uni-ulm.de

State of game: x at {a1, b2, b3}, o at {a3, b1, c3}

Modeling an Apartment

Will furniture move?



State of apartment: Stove at C1, coffee table at E2, etc.

State of the World

- In games and simple worlds, we can assume or expect:
  - Everything is fully observable (that means, there are no hidden variables)
  - The world has no uncertainty

- Is this always going to be feasible?

Modeling Person's Emotional State

Image taken from www.pinterest.com

State of user: currently 'happy' ... Are you sure?

Beliefs over State

- Uncertainty arises when:
  - You can’t observe the value of a state
  - When the state changes
- When there’s uncertainty:
  - We have beliefs of the world
  - Need to quantify level of uncertainty
- Then we can make decisions properly
- Use probability theory to model our beliefs

Want Something Like This

- Simple world:
  stove is on or off,
  time is 9:00am, 9:01am, …

- Variables:
  Stove, Time

- Sample state:
  Stove=on and Time=9:00am

- Sample estimations:
  Pr(Stove=on)
  Pr(Stove=on|Time=noon)

Random Variables

- Assume set **V** of random variables: X, Y, etc.
  - Each RV X has a domain of values Dom(X)
  - X can take on any value from Dom(X)
  - Assume **V** and Dom(X) are finite

- Examples:
  - Dom(X) = $\{x_1, x_2, x_3\}$
  - Dom(Weather) = {sunny, cloudy, rainy}
  - Dom(IamHappy) = {true, false}

Modeling Example

- Student asks you (TA) a programming question
- You consider how to answer the question
  - What are RVs of the student?
  - What are RVs about you?

Which variables are observable,
which are hidden?

State

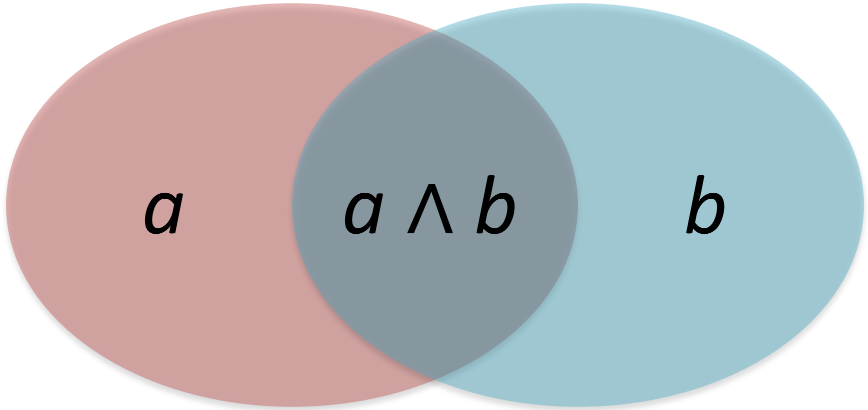
- A formula is a logical combination of variable assignments
  - E.g. $(X = x_2 \lor X = x_3) \land Y = y_2$
- A state is an assignment of values to each variable
  - One state represents one possible world
  - The set of states denote the set of possible worlds
  - Note: Think truth tables for discrete RVs

Modeling Example cont.

- Student asks you (TA) a programming question
- You consider how to answer the question
  - What are RVs of the student?
  - What are RVs about you?
- Draw out the set of states in truth table format

Probability Distributions

- A probability distribution Pr: $\mathcal{L} \rightarrow [0,1]$ s.t.
  - $0 \leq \Pr(a) \leq 1$
  - *Pr(a) = Pr(b)* if *a* is logically equivalent to *b*
  - *Pr(a) = 1* if *a* is a tautology
  - $\Pr(a \vee b) = \Pr(a) + \Pr(b) - \Pr(a \wedge b)$



Probability Distributions

- A probability distribution $\Pr: \mathcal{L} \rightarrow [0,1]$ s.t.
  - $0 \leq \Pr(a) \leq 1$
  - *Pr(a) = Pr(b)* if *a* is logically equivalent to *b*
  - *Pr(a) = 1* if *a* is a tautology
  - $\Pr(a \vee b) = \Pr(a) + \Pr(b) - \Pr(a \wedge b)$

Caution: Probability axioms not always followed in theories that use probability!

Probability Distributions

- A probability distribution Pr: $\mathcal{L} \rightarrow [0,1]$ s.t.
  - $0 \leq \Pr(a) \leq 1$
  - *Pr(a) = Pr(b)* if *a* is logically equivalent to *b*
  - *Pr(a) = 1* if *a* is a tautology
  - $\Pr(a \vee b) = \Pr(a) + \Pr(b) - \Pr(a \wedge b)$
- *Pr(a)* denotes our degree of belief in *a*
  - *Pr(a) = 0* if you consider it to be impossible
- The sum of the distribution must be 1.0

Examples

- $\Pr(X = x_1) = 0.9$
  - *Pr(Stove = on) = 0.9*
  - *Pr(Stove = off) = 1 – Pr(Stove = on) = 0.1*
  - *Pr(Time = noon) = 0.001*

Visualizing Probability Distribution

- X-axis: set of possible outcomes
- Y-axis: probability

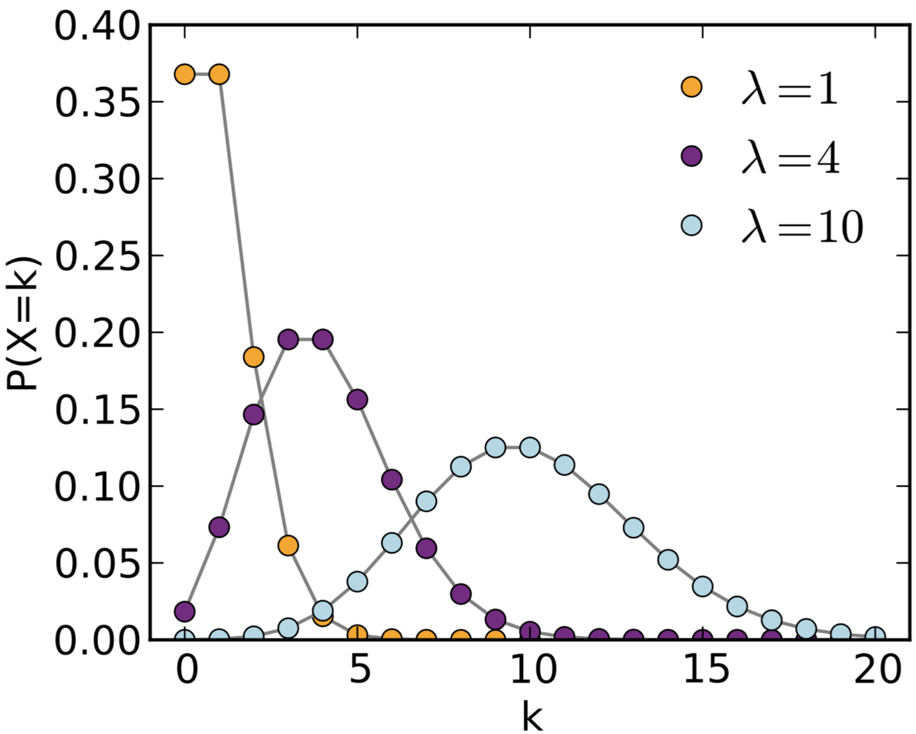


Image taken from wikipedia.org

Commonly used distributions:

- Normal (Gaussian)
- Uniform

Joint Probability Distribution

- Probability distribution
  - Involves one RV to describe state space
  - Probabilities must sum up to 1.0

- Joint probability distribution
  - When the state is described by two or more RVs
  - Specify probabilities for all combinations of events
  - Probabilities must sum up to 1.0

Examples

- *Pr*($X = x_1 \wedge Y = y_2$) *= 0.6*
  - $\Pr(Stove = on \wedge Time = noon) = 0.6$
  - $\Pr(Stove = on \wedge Time = 12{:}01pm) = 0.2$
  - $\Pr(Stove = on \wedge Time = 12{:}02pm) = 0.1$
  - …

Modeling Example cont.

- Student asks you (TA) a programming question
- You consider how to answer the question
  – What are RVs of the student?
  – What are RVs about you?
- Draw out the set of states in truth table format
- Assign probabilities to each state in the table

The Summing Out Property

- $$\Pr(x) = \sum_{y \in Dom(Y)} \Pr(x \wedge y)$$
- Also called marginalization
- Example:
  Pr(Stove=on) = Pr(Stove = on ∧Time = 9:00am)
  + Pr(Stove = on ∧ Time = 9:01am)
  + Pr(Stove = on ∧ Time = 9:02am)
  + Pr(Stove = on ∧ Time = 9:03am)
  …
  for all values in Dom(Y)

Modeling Example cont.

- Student asks you (TA) a programming question
- You consider how to answer the question
  - What are RVs of the student?
  - What are RVs about you?
- Draw out the set of states in truth table format
- Assign probabilities to each state in the table
- Pick an RV, sum it out

The Inference Task

- General structure:
  - You have general knowledge about the world
  - You observe an event (or series of events)
  - You want to estimate the probability of an event (or several events) that you cannot observe

Example

- You're programming and suddenly you get a headache. You think: Argh! 50% of my headaches are caused by annoying bugs, so there's a 50% chance there's a bug in the code ☹

Is this reasoning correct?

Example

- You’re programming and suddenly you get a headache. You think: Argh! 50% of my headaches are caused by annoying bugs, so there’s a 50% chance there’s a bug in the code ☹

- H = have headache
- B = have bugs in code

Example

- You're programming and suddenly you get a headache. You think: Argh! 50% of my headaches are caused by annoying bugs, so there's a 50% chance there's a bug in the code ☹
- Also given:
  - $\Pr(H) = 1/10$
  - $\Pr(B) = 1/40$
  - $\Pr(H|B) = 1/2$

We observe H.

Task is to compute $\Pr(B|H)$.

Conditional Probability, *Pr(b|a)*

- Recall definition: What is the probability of *b* when *a* has already occurred?

- Conditional probability is critical in inference
  - E.g. Pr(Stove = on | Time = noon)?
  - E.g. Pr(PhysicsSkills = poor | MathSkills = good)?
  - E.g. Pr(MathSkills = good | NumAlgebraErrors = low ∧ NumCalculusErrors = high)?

- Effectively, compute *Pr(b|a)* when we observe *a*

Calculating Conditional Probability

- $$\Pr(b \mid a) = \frac{\Pr(b \wedge a)}{\Pr(a)}$$
- If $Pr(a) = 0$, we set $Pr(b|a) = 1$ by convention
- Intuition:
  - Numerator: What is the probability both events occur together?
  - Denominator: What is the probability a occurs at all (regardless of what other events that are happening)?
  - $Pr(b|a)$ gives relative weight of *b*-worlds among *a*-worlds

Related Properties

- Given conditional probability: $\Pr(b \mid a) = \dfrac{\Pr(b \wedge a)}{\Pr(a)}$
- Product rule: $Pr(b \wedge a) = Pr(b|a)Pr(a)$

  or

  $Pr(a \wedge b) = Pr(a|b)Pr(b)$
- Sum out rule: $Pr(a) = \sum_b Pr(a \wedge b)$

  or

  $Pr(a) = \sum_b Pr(a|b)Pr(b)$
- Chain rule:

  $Pr(abcd) = Pr(a|bcd)Pr(b|cd)Pr(c|d)Pr(d)$

  – Holds for any number of variables

Note: Pr(ab) is shorthand for $Pr(a \wedge b)$

Example

- You're programming and suddenly you get a headache. You think: Argh! 50% of my headaches are caused by annoying bugs, so there's a 50% chance there's a bug in the code ☹

$\Pr(H) = 1/10$
$\Pr(B) = 1/40$
$\Pr(H|B) = 1/2$

- **Want:** $\Pr(B|H) = ?$

Joint Distribution Example

	sunny	
	cold	~cold
headache	0.108	0.012
~headache	0.016	0.064

	~sunny	
	cold	~cold
headache	0.072	0.008
~headache	0.144	0.576

All values sum up to 1.0

Joint Distribution Example

sunny

	cold	~cold
headache	0.108	0.012
~headache	0.016	0.064

~sunny

	cold	~cold
headache	0.072	0.008
~headache	0.144	0.576

- Pr(headache ∧ cold | sunny) = ?

Joint Distribution Example

sunny		
	cold	~cold
headache	0.108	0.012
~headache	0.016	0.064

~sunny		
	cold	~cold
headache	0.072	0.008
~headache	0.144	0.576

- $\Pr(\text{headache} \wedge \text{cold} \mid \sim\text{sunny}) = ?$

Asian Flu Example

- Doctor X knows that Asian flu causes fever 95% of the time.
- X knows that a random person has a 10^{-7} chance of having Asian flu.
- X knows that 1 in 100 people suffer from a fever.
- Joe has a fever: what are the chances that Asian flu is the cause of the fever?

Evidence is symptom (F)

Hypothesis is illness causing symptom (A)

- A = Asian flu
- F = fever
- Pr(A|F) = ?

What information is given?

Asian Flu Example

Pr(F|A)

- Doctor X knows that Asian flu causes fever 95% of the time.
- X knows that a random person has a 10^{-7} chance of having Asian flu. Pr(A)
- X knows that 1 in 100 people suffer from a fever. Pr(F)
- Joe has a fever: what are the chances that Asian flu is the cause of the fever?

Evidence is symptom (F)

Hypothesis is illness causing symptom (A)

- A = Asian flu
- F = fever
- Pr(A|F) = ?

What information is given?

Recall conditional probability is Pr(A|F) = Pr(A,F) / Pr(F)

Can we use this to solve for Pr(A|F)?

Gets its own slide!!!

Bayes Rule

- Note: $Pr(ab) = Pr(ba)$
- We have: $Pr(ab) = Pr(a|b)Pr(b)$
- So: $Pr(a|b)Pr(b) = Pr(ab) = Pr(ba) = Pr(b|a)Pr(a)$

- Bayes rule states:

$$Pr(b|a) = \frac{Pr(a|b)Pr(b)}{Pr(a)}$$

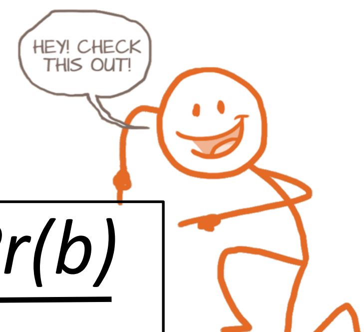


Image taken from giphy.com

- Why is this so important?

Using Bayes Rule for Inference

- We may want to form a hypothesis (H) about the world based on the evidence (e) we observe
- Bayes rule expresses this notion as the belief of H given e

$$\Pr(H|e) = \frac{\Pr(e|H)\Pr(H)}{\Pr(e)}$$

Using Bayes Rule for Inference

- We may want to form a hypothesis (H) about the world based on the evidence (e) we observe
- Bayes rule expresses this notion as the belief of H given e

Likelihood

Prior probability

Posterior probability

$$\Pr(H|e) = \frac{\Pr(e|H)\Pr(H)}{\Pr(e)}$$

Normalizing constant

Need for Simplifying Assumptions

- Previously: compute posterior distribution
- More often: compute posterior joint distribution

- Problem: joint distribution is usually too big
  - Exponential in # variables
- Solution: use independence
  - To simplify computational needs
  - To simplify model

Independence

- Two variables *A* and *B* are independent if knowledge of *A* does not the change uncertainty of *B* (and vice versa)
  - Pr(A|B) = Pr(A)
  - Pr(B|A) = Pr(B)
  - Pr(AB) = Pr(A)Pr(B)
  - In general:

$$Pr(X_1,\ldots,X_n) = \prod_{i=1}^{n} \Pr(Xi)$$

Only need n numbers to specify the joint!

Independence Example

- Consider: Bennett smiles and squint eyes

Image taken from iemoji.com

- If Pr(Smile|Squint) = Pr(Smile)
  – Chance of him smiling when he squints
  – Chance of him smiling in anyway

- And Pr(Squint|Smile) = Pr(Squint)
  – Chance of him squinting when he smiles
  – Chance of him squinting no matter what else he's doing

- Then Smile and Squint are independent

What does Independence Buy Us?

- Product rule changes:

$$\Pr(ab) = \Pr(a|b)\Pr(b)$$

$$\Pr(ab) = \Pr(a)\Pr(b)$$

- Chain rule changes:

$$\Pr(abcd) = \Pr(a|bcd)\Pr(b|cd)\Pr(c|d)\Pr(d)$$

$$\Pr(abcd) = \Pr(a)\Pr(b)\Pr(c)\Pr(d)$$

Conditional Independence

- To loosen the independence assumption, we can use conditional independence

- Two variables *A* and *B* are conditionally independent given *C* if:
  - $\Pr(a|b,c) = \Pr(a|c) \qquad \forall\ a, b, c$

- Knowing the value of B does not change the prediction of A given the presence of C

Conditional Independence Example

Image taken from pinterest.com

- Consider: Want tea, pink cup, and rainy

- If Pr(Tea|Pink,Rainy) = Pr(Tea|Rainy)
  - Chance of wanting tea on rainy days in pink cup is the same as chance of wanting tea on rainy days in any cup

- And Pr(Tea|Pink,~Rainy) = Pr(Tea|~Rainy)
  And Pr(Tea|~Pink,Rainy) = Pr(Tea|Rainy)
  And Pr(Tea|~Pink,~Rainy) = Pr(Tea|~Rainy)
  And Pr(~Tea|Pink,Rainy) = Pr(~Tea|Rainy)
  And …
  - Check equivalence for all other combinations

- Then Tea is independent of Pink given Rainy

Cheatsheet of Statistics

- Probability distribution
  - All values must sum up to 1.0

- Conditional probability: $\Pr(b|a) = \dfrac{\Pr(b,a)}{\Pr(a)}$
- Product rule: $\Pr(a,b) = \Pr(a|b)\Pr(b)$
- Sum-out rule (marginalization): $\Pr(a) = \sum_b \Pr(a,b)$
  $= \sum_b \Pr(a|b) \Pr(b)$
- Chain rule: $\Pr(abcd) = \Pr(a|bcd)\Pr(b|cd)\Pr(c|d)\Pr(d)$
  - Applies to any number of variables

- Bayes rule: $\Pr(b|a) = \dfrac{\Pr(a|b)\Pr(b)}{\Pr(a)}$

Key Ideas

- Main concept
  - Using probability to model uncertainty
- Representation:
  - States as an assignment of values to each RV
  - Beliefs over states as probability distributions
- Computational issues:
  - Joint distributions are often too large to compute
  - Assume independence and conditional independence