Reasoning Under Uncertainty

Uncertainty material is covered in chapters 13 and 14 of R&N and Chapter 8 of https://artint.info.

Chapter 13 gives some basic background on probability from the point of view of A.I.

Chapter 14 talks about Bayesian Networks, exact reasoning in Bayes Nets as well as approximate reasoning, which will be main topics for us.

Note: Slides in this section draw from Faheim Bacchus, Craig Boutillier, Andrew Moore, Sheila McIlraith ...

Review: Probability Distributions over Finite Sets

A probability is a function defined over a set of atomic events U.

U represents the universe of all possible events.

Review: Probability over Finite Sets

Given **U** (a universe of events), a probability function is a function defined over subsets of **U** that maps each subset onto the real numbers and that satisfies the Axioms of Probability. These are:

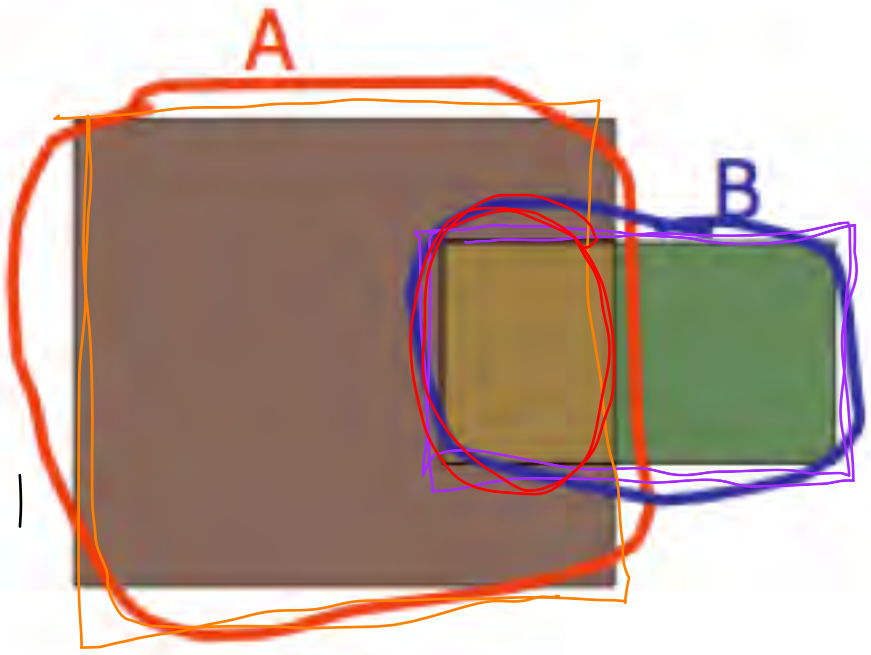
1. $\mathbf{\underline{P(U) =}\ 1}$
2. $\mathbf{\underline{P(A) \in [0,1]}}$
3. $\mathbf{P(\{\}) = 0}$
4. $\mathbf{\underline{P(A \cup B) = P(A) + P(B) - P(A \cap B)}}$

NB: if $\boldsymbol{A \cap B = \{\}}$ ***then*** $\boldsymbol{P(A \cup B) = P(A) + P(B)}$

Review: Probability over Finite Sets

$$P(\underline{A} \cup \underline{B}) = P\underline{(A)} + P\underline{(B)} - \underline{\underline{P(A \cap B)}}$$

$\mathcal{U}$



In a presidential election, there are four candidates. Call them A, B, C, and D. Based on polls, we estimate that A has a 20 percent chance of winning the election, while B has a 40 percent chance of winning. What is the probability that A or B win the election?

Vote!
http://etc.ch/ekXi

VIEW POLL RESULTS HERE https://directpoll.com/r?XDbzPBd3ixYqg8Ave5sE9MkTtkd5vAYDdCsIT7a9h

$$P(A \text{ or } B) = \underset{20\%}{\underline{P(A)}} + \underset{40\%}{\underline{P(B)}} - \underline{P(A \cap B)}$$

Notation: Properties and Sets

We often write

$A \lor B$: to represent the set of events with either property A or B, i.e. the set $A \cup B$

$A \land B$: to represent the set of events both property A and B, i.e. the set $A \cap B$

$\neg A$: to represent the set of events that do not have property A: the set U-A (i.e., the complement of A w.r.t. the universe of events U)

Review: Probability over Feature Vectors

CS student = {Gender, eyesight, height, eye color}

As we move forward, ee will model sets of events in our universe as vectors of feature values.

Like CSPs, we have

1. a set of variables $V_1, V_2, \ldots, V_n$
2. a finite domain of values for each variable, $Dom[V_1]$, $Dom[V_2]$, …, $Dom[V_n]$.

The universe of events U is the set of all vectors of values for the variables

$$\langle d_1, d_2, \ldots, d_n \rangle : d_i \in Dom[V_i]$$

When we write P(A=a, B=b), we will mean the probability that variable A has been assigned value ‘a’ **and** variable B has been assigned value ‘b’. Note that here, sets of events are induced by a given value assignment. So, P(A=a) represents a set of events in which A holds the value ‘a’.

Review: Probability over Feature Vectors

Example:

$$P(\{V_1 = 1\}) = \sum_{x_2 \in \mathbf{Dom}[V_2]}, \sum_{x_3 \in \mathbf{Dom}[V_3]} \underbrace{P(\{V_1 = 1, V_2 = x_2, V_3 = x_3\})}_{0.1}.$$

U →

(V1 = 1, V2 = 1, V3 = 1)
(V1 = 1, V2 = 1, V3 = 2)
(V1 = 1, V2 = 1, V3 = 3)
(V1 = 1, V2 = 2, V3 = 1)
(V1 = 1, V2 = 2, V3 = 2)
(V1 = 1, V2 = 2, V3 = 3)
(V1 = 1, V2 = 3, V3 = 1)
(V1 = 1, V2 = 3, V3 = 2)
(V1 = 1, V2 = 3, V3 = 3)

(V1 = 2, V2 = 1, V3 = 1)
(V1 = 2, V2 = 1, V3 = 2)
(V1 = 2, V2 = 1, V3 = 3)
(V1 = 2, V2 = 2, V3 = 1)
(V1 = 2, V2 = 2, V3 = 2)
(V1 = 2, V2 = 2, V3 = 3)
(V1 = 2, V2 = 3, V3 = 1)
(V1 = 2, V2 = 3, V3 = 2)
(V1 = 2, V2 = 3, V3 = 3)

(V1 = 3, V2 = 1, V3 = 1)
(V1 = 3, V2 = 1, V3 = 2)
(V1 = 3, V2 = 1, V3 = 3)
(V1 = 3, V2 = 2, V3 = 1)
(V1 = 3, V2 = 2, V3 = 2)
(V1 = 3, V2 = 2, V3 = 3)
(V1 = 3, V2 = 3, V3 = 1)
(V1 = 3, V2 = 3, V3 = 2)
(V1 = 3, V2 = 3, V3 = 3)

0.02

P(U) = 1

Review: Probability over Feature Vectors

Example:

$$P(\{\underline{\underline{V_1 = 1, V_3 = 2}}\}) = \sum_{x_2 \in \mathrm{Dom}[V_2]} P(\{V_1 = 1, \underline{\underline{V_2 = x_2}}, V_3 = 2\}).$$

$\cup$

(V1 = 1, V2 = 1, V3 = 1)
(V1 = 1, V2 = 1, V3 = 2)
(V1 = 1, V2 = 1, V3 = 3)
(V1 = 1, V2 = 2, V3 = 1)
(V1 = 1, V2 = 2, V3 = 2)
(V1 = 1, V2 = 2, V3 = 3)
(V1 = 1, V2 = 3, V3 = 1)
(V1 = 1, V2 = 3, V3 = 2)
(V1 = 1, V2 = 3, V3 = 3)

(V1 = 2, V2 = 1, V3 = 1)
(V1 = 2, V2 = 1, V3 = 2)
(V1 = 2, V2 = 1, V3 = 3)
(V1 = 2, V2 = 2, V3 = 1)
(V1 = 2, V2 = 2, V3 = 2)
(V1 = 2, V2 = 2, V3 = 3)
(V1 = 2, V2 = 3, V3 = 1)
(V1 = 2, V2 = 3, V3 = 2)
(V1 = 2, V2 = 3, V3 = 3)

(V1 = 3, V2 = 1, V3 = 1)
(V1 = 3, V2 = 1, V3 = 2)
(V1 = 3, V2 = 1, V3 = 3)
(V1 = 3, V2 = 2, V3 = 1)
(V1 = 3, V2 = 2, V3 = 2)
(V1 = 3, V2 = 2, V3 = 3)
(V1 = 3, V2 = 3, V3 = 1)
(V1 = 3, V2 = 3, V3 = 2)
(V1 = 3, V2 = 3, V3 = 3)

In these examples we are "summing out" some variables, which is also known as "marginalizing" our distribution

In a presidential election, there are four candidates. Call them A, B, C, and D. Polling data and some features of the candidates are in the table below. What's P(Winner Age > 65)? What's P(Winner Gender = Male)?

Candidate	Gender	Region	Age	P(win)
A	Female	Midwestern	Over 65	0.2
B	Male	Midwestern	Over 65	0.4
C	Female	Eastern	Over 65	0.3
D	Male	Western	Under 65	0.1

Vote!
http://etc.ch/ekXi

VIEW POLL RESULTS HERE https://directpoll.com/r?XDbzPBd3ixYqg8Ave5sE9MkTtkd5vAYDdCsIT7a9h

$$\sum_{id} \sum_{gender} \sum_{region} P(id, gender, region, age > 65)$$

Review: Probability over Feature Vectors

Problem:

There is an exponential number of atomic probabilities to specify.

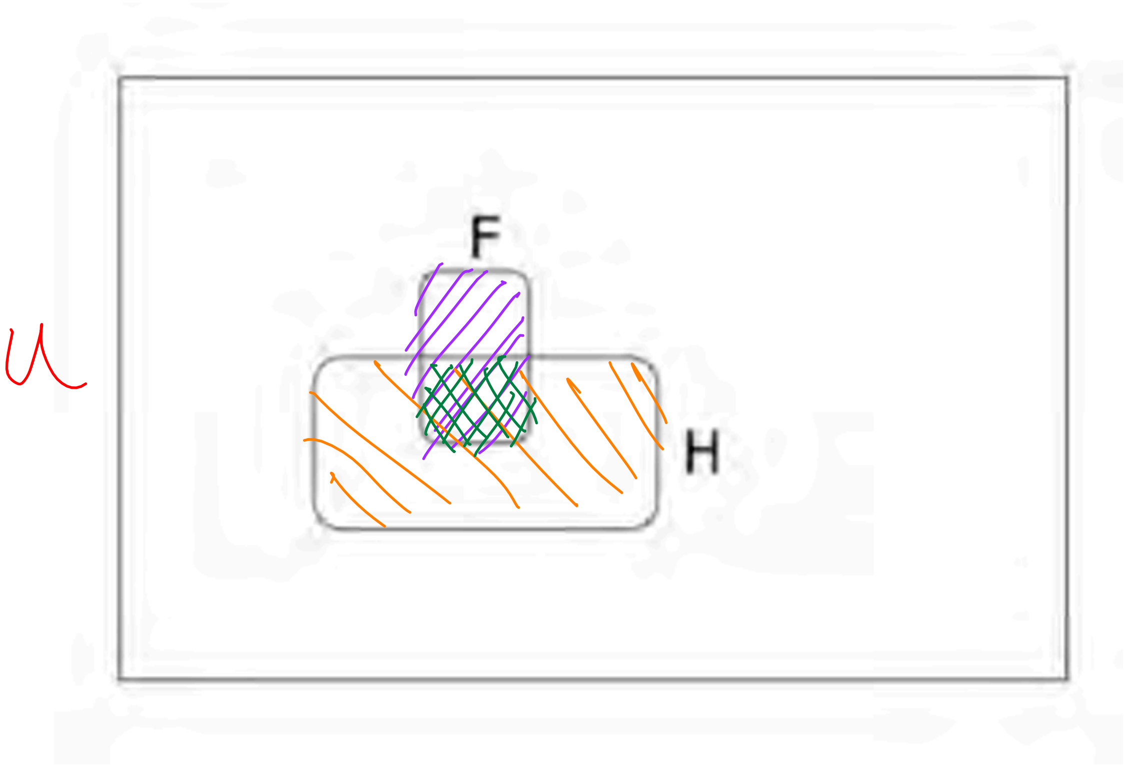
Requires summing up an exponential number of items.

To evaluate the probability of sets containing a particular subset of variable assignments we can do much better. Improvements come from the use of:

1. **probabilistic independence, especially conditional independence.**
2. **approximation techniques, many of which depend on distributions structured by independence.**

What is a Conditional Probability?

P(Headache=true|Flu=true) represents the fraction of flu-infected worlds in which you have a headache.



$$\frac{\text{Gren area}}{\text{purple area}}$$

= # worlds with flu and headache/#worlds with flu

= area of flu and headache/area of flu

= P(Headache=true,Flu=true)/P(Flu=true)

$$\frac{P(H=t \wedge F=t)}{P(F=t)}$$

Axioms of Conditional Probability

A conditional probability is a also probability function, but now over a *subset* of events in the universe instead of over the entire universe. Similar axioms hold:

$$\left\{\begin{array}{l} \mathbf{P(A|A) = 1} \\[1em] \underline{\mathbf{P(B|A)}} \in \mathbf{[0,1]} \\[1em] \underline{\mathbf{P(C} \cup \mathbf{B|A)}} = \underline{\mathbf{P(C|A)}} + \underline{\mathbf{P(B|A)}} - \underline{\mathbf{P(C} \cap \mathbf{B|A)}} \end{array}\right.$$

Review: Independence

Probability density is a measure of likelihood. Assume you pick an element at random from U. Density (i.e. the value of P(B) is a measure as to how likely is it to also be in set B.

It could be that the density (i.e. likelihood) of B given (or conditioned on) A is **identical** to its density (or likelihood) in U.

Alternately, the density of B given A could be very **different** that its density (or likelihood) in U.

In the first case we say that B is **independent** of A. While in the second case B is **dependent** on A.

Review: Independence

A and B are **independent** properties:

$$\underline{P(B|A)} = \underline{P(B)}$$

A and B are **dependent**:

$$P\underline{(B|A) \neq P}(B)$$

Computational Impact

We will soon see in more detail how independence allows us to speed up computations related to inference. But the fundamental insight is that

If A and B are independent properties then

$$\mathbf{P(\underline{A \wedge B}) = P(\underline{B}) * \underline{P(A}) }$$

Proof:

$$P(A \mid B) = P(A)$$

$$\frac{P(A \,\&\, B)}{P(B)} = P(A) \Rightarrow P(A \,\&\, B) = P(A) * P(B)$$

Computational Impact

We will soon see in more detail how independence allows us to speed up computations related to inference. But the fundamental insight is that

If A and B are independent properties then

$$\mathbf{P(A \wedge B) = P(B) * P(A)}$$

Proof:

$$P(B|A) = P(B) \qquad \text{(def'n of independence)}$$
$$P(A \wedge B)/P(A) = P(B)$$
$$P(A \wedge B) = P(B) * P(A)$$

What is Conditional Independence?

If $P(V_1|V_2=b,V_3=c) = P(V_1|V_2=b)$, we have not gained any additional information about V_1 from knowing $V_3=c$.

In this case we say that V_1 is conditionally independent of V_3 *given* V_2.

That is, once we know V_2, additionally knowing V_3 is irrelevant (it will give us no more information as to the value of V_1).

Note we could have $P(V_1|V_3=c) \neq P(V_1)$. But once we learn $V_2=b$, the value of V_3 becomes irrelevant.

Computational Impact

Similar results hold for conditional independence. If B and C are conditionally independent given A, then

$$\underline{\underline{P(B\wedge C|A)}} = \underline{\underline{P(B|A)}} * \underline{\underline{P(C|A)}}$$

$$P(B|CA) = P(B|A)$$

$$\frac{P(BCA)}{P(CA)} = \frac{P(BA)}{P(A)}$$

Proof:

$$\frac{P(\cancel{A})}{P(CA)} \cdot \frac{P(B,C,A)}{P(\cancel{A})} =$$

$$P(B|A)$$

Computational Impact

Similar results hold for conditional independence. If B and C are conditionally independent given A, then

$$P(B \wedge C|A) = P(B|A) * P(C|A)$$

Proof:

$P(B|C \wedge A) = P(B|A)$ (def'n of conditional independence)

$P(B \wedge C \wedge A)/P(C \wedge A) = P(B \wedge A)/P(A)$

$P(B \wedge C \wedge A)/P(A) = P(C \wedge A)/P(A) * P(B \wedge A)/P(A)$

$P(B \wedge C|A) = P(B|A) * P(C|A)$

Computational Impact

As with independence, conditional independence allows us to break up our computation onto distinct parts

	A	
	1	1
B	1	2

A	B	C
1	1	1
1	2	1
1	1	2
1	2	2

$$\circledast \quad \boxed{P(B \wedge C|A)} = \boxed{P(B|A)} * \boxed{P(C|A)}$$

It also allows us to ignore certain pieces of information during computations

$$P(B|A \wedge C) = P(B|A)$$

Review: Chain Rule

$$\boxed{P(A_1 \wedge A_2 \wedge \ldots \wedge A_n) =}$$

$$\rightarrow \underline{P(A_1 | A_2 \wedge \ldots \wedge A_n)} * \underline{P(A_2 | A \;\, _3 \wedge \ldots \wedge A_n)}$$

$$* \ldots * \underline{P(A_{n-1} | A_n)} * \underline{P(A_n)}$$

Proof:

$$\frac{P(A_1 \wedge A_2 \wedge \ldots \wedge A_n)}{P(A_2 \cancel{\ldots} \ldots A_n)} \cdot \frac{P(A_2 \wedge \cancel{\wedge} \ldots \wedge A_n)}{P(A_2 \wedge \cancel{\ldots} \ldots A_n)} \cdots \frac{P(A_{\cancel{n}-1} \wedge A_n)}{P(A_{\cancel{n}})} \cdot P(\cancel{A}_n)$$

$$= P(A_n | A_{n-1} \ldots A_1) \cdot P(A_{n-1} | A_{n-2} \wedge \ldots A_1) \ldots * P(A_1)$$

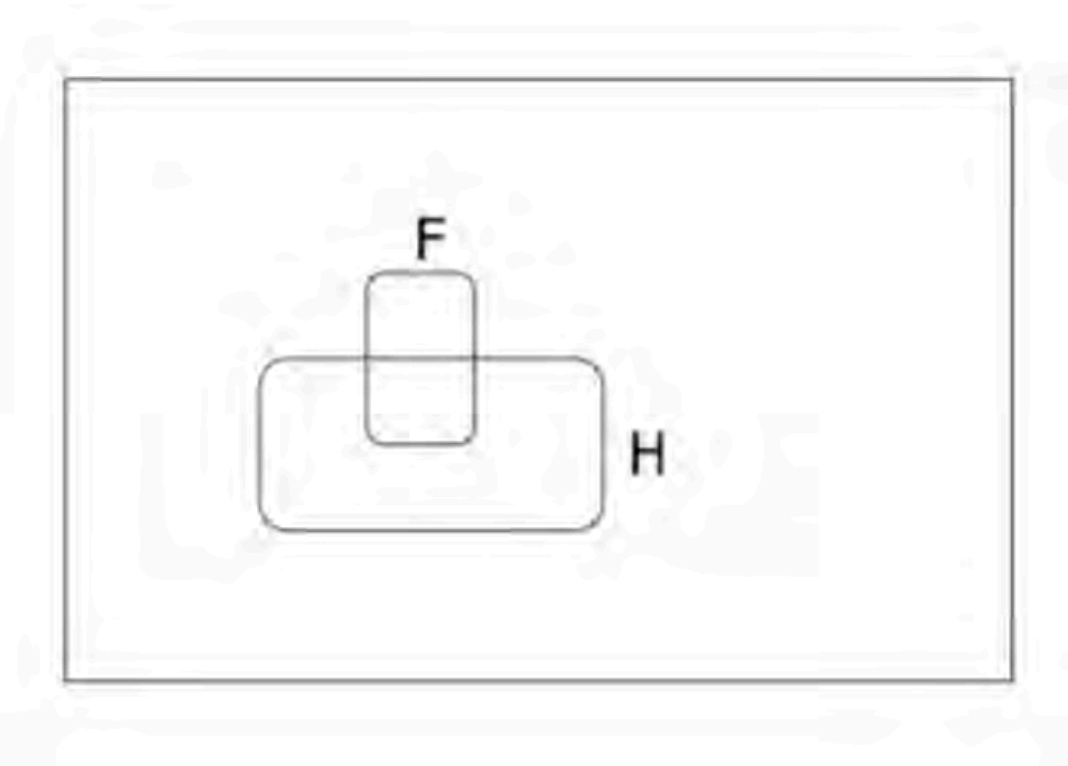
Review: Chain Rule

$$\underline{P(A_1 \wedge A_2 \wedge \ldots \wedge A_n)} =$$
$$P(A_1 | A_2 \wedge \ldots \wedge A_n) * P(A_2 | A_3 \wedge \ldots \wedge A_n)$$
$$* \ldots * P(A_{n-1} | A_n) * P(A_n)$$

Proof:

$$P(A_1 | A_2 \wedge \ldots \wedge A_n) * P(A_2 | A_3 \wedge \ldots \wedge A_n)$$
$$* \ldots * P(A_{n-1} | A_n)$$
$$= P(A_1 \wedge A_2 \wedge \ldots \wedge A_n) / P(A_2 \wedge \ldots \wedge A_n) *$$
$$P(A_2 \wedge \ldots \wedge A_n) / P(A_3 \wedge \ldots \wedge A_n) * \ldots *$$
$$P(A_{n-1} \wedge A_n) / P(A_n) * P(A_n)$$

Back to Flu World



P(Headache=true) = 1/10

P(Flu=true) = 1/40

P(Headache=true|Flu=true) = 1/2

Headaches are rare and having flu is rarer. But, given flu, there is a 50/50 chance you have a headache.

$$\frac{P(F,H)}{P(H)} = \frac{P(H|F)P(F)}{P(H)} \rightarrow$$

What is P(Flu=true|Headache=true)?

$$\frac{\frac{1}{2} \cdot \frac{1}{40}}{\frac{1}{10}}$$

Vote!
http://etc.ch/ekXi

VIEW POLL RESULTS HERE https://directpoll.com/r?XDbzPBd3ixYqg8Ave5sE9MkTtkd5vAYDdCsIT7a9h

Vote!
http://etc.ch/ekXi

[VIEW POLL RESULTS HERE https://directpoll.com/r?XDbzPBd3ixYqg8Ave5sE9MkTtkd5vAYDdCsIT7a9h](https://directpoll.com/r?XDbzPBd3ixYqg8Ave5sE9MkTtkd5vAYDdCsIT7a9h)

posterior

prior

$$\underline{P(F|H)} = \frac{P(H|F)\,P(F)}{P(H)}$$

Bayes rule

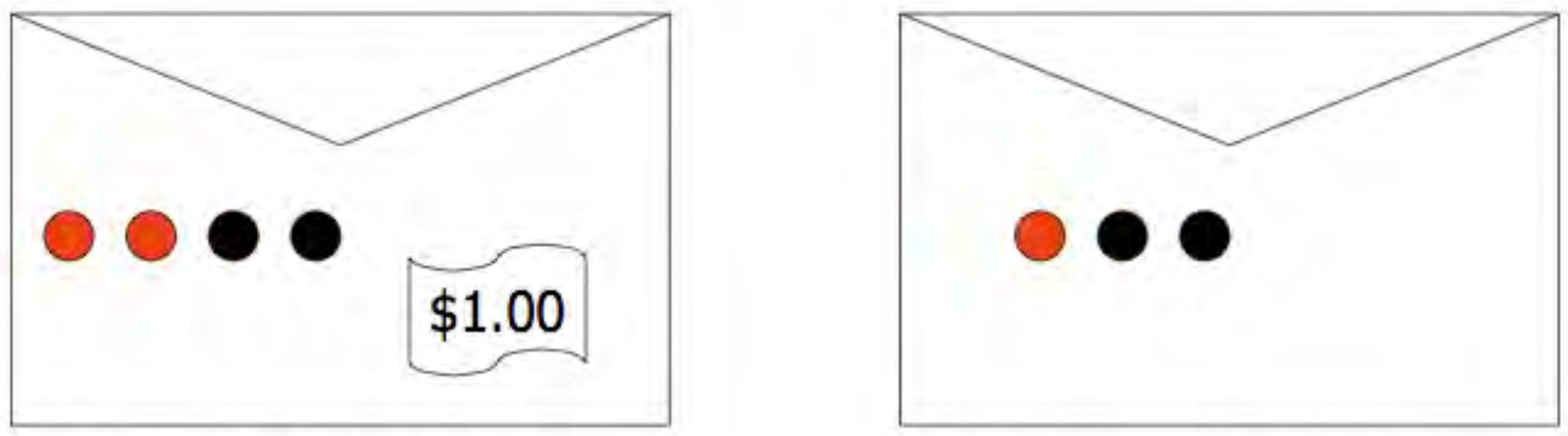
What we just did

We Derived Bayes' Rule.

$$P(Y|X) = P(X|Y)P(Y)/P(X)$$

$$\begin{aligned} P(Y|X) &= P(Y \wedge X)/P(X) \\ &= P(Y \wedge X)/P(X) * P(Y)/P(Y) \\ &= P(Y \wedge X)/P(Y) * P(Y)/P(X) \\ &= P(X|Y)P(Y)/P(X) \end{aligned}$$

Using Bayes Rule to gamble



The "Win" envelope has a dollar and four beads in it

The "Lose" envelope has three beads and no money

Trivial question: Someone picks an envelope and random and asks you to bet as to whether or not it holds a dollar. What are your odds?

VIEW POLL RESULTS HERE https://directpoll.com/r?XDbzPBd3ixYqg8Ave5sE9MkTtkd5vAYDdCsIT7a9h

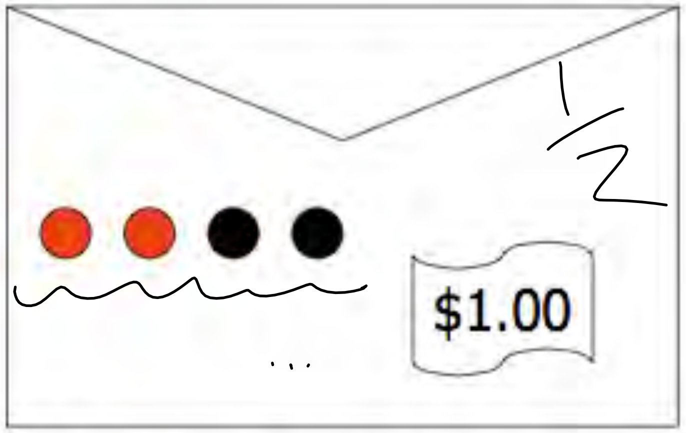
Vote!
http://etc.ch/ekXi

Using Bayes Rule to gamble

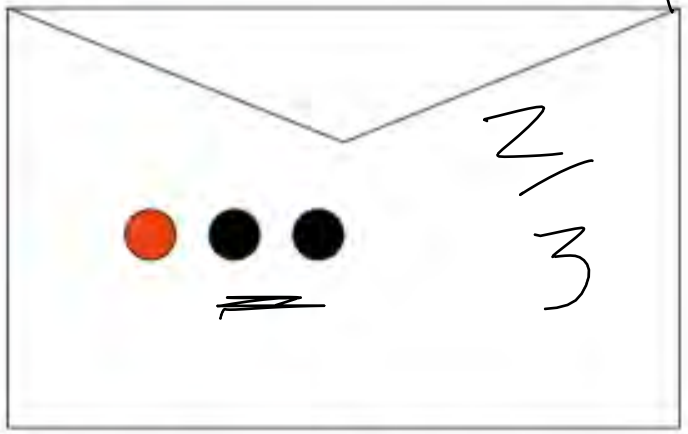
$P(E=\$ \mid B=black)$?

$$* \frac{P(B=black \mid E=\$)P(\$)}{P(B=black)}$$

$$\frac{\frac{1}{2}\cdot\frac{1}{2}}{\frac{1}{2}\cdot\frac{1}{2}+\frac{2}{3}\cdot\frac{1}{2}}$$



The "Win" envelope has a dollar and four beads in it



The "Lose" envelope has three beads and no money

$$P(B=bl \mid E=\$) = \frac{\frac{1}{2}\cdot\frac{1}{2}}{P(B=bl \mid E=\$)P(E=\$) + P(B=Bl \mid E=\bar{\$}) * P(E=\bar{\$})}$$

Not trivial question: Someone lets you take a bead out of the envelope before you bet. If it is black, what are your odds? If it is red, what are your odds?

$$\textcircled{*} \; \sum_E P(B=bl, E) = P(B=bl)$$

Vote!
http://etc.ch/ekXi

VIEW POLL RESULTS HERE https://directpoll.com/r?XDbzPBd3ixYqg8Ave5sE9MkTtkd5vAYDdCsIT7a9h

Review: Normalizing

To **normalize** a vector of k numbers or a column in our table, e.g., <3, 4, 2.5, 1, 10, 21.5> we must sum them and divide each number by the sum:

3 + 4 + 2.5 +1 +10 + 21.5 = 42

Normalized vector:
= <3/42, 4/42, 2.5/42, 1/42, 10/42, 21.5/42>
= <0.071, 0.095, 0.060, 0.024, 0.238, 0.512>

After normalizing the vector of numbers sums to 1

It therefore can be used to specify a probability distribution.

In a presidential election, there are four candidates. Call them A, B, C, and D. Polling data and some features of the candidates are in the table below. Given the winning candidate is female, what's the probability that they are Midwestern?

Candidate	Gender	Region	Age	P(win)
A	Female	Midwestern	Over 65	0.2
B	Male	Midwestern	Over 65	0.4
C	Female	Eastern	Under 65	0.3
D	Male	Western	Under 65	0.1

P(Midwest | ♀)

0.5

Vote!
http://etc.ch/ekXi

VIEW POLL RESULTS HERE https://directpoll.com/r?XDbzPBd3ixYqg8Ave5sE9MkTtkd5vAYDdCsIT7a9h

P(Midwest + ♀) | P(♀)

0.2/0.5

DO SOME MORE PROBLEMS!!! https://www.teach.cs.toronto.edu//~csc384h/summer/Lectures/probability-problems.html