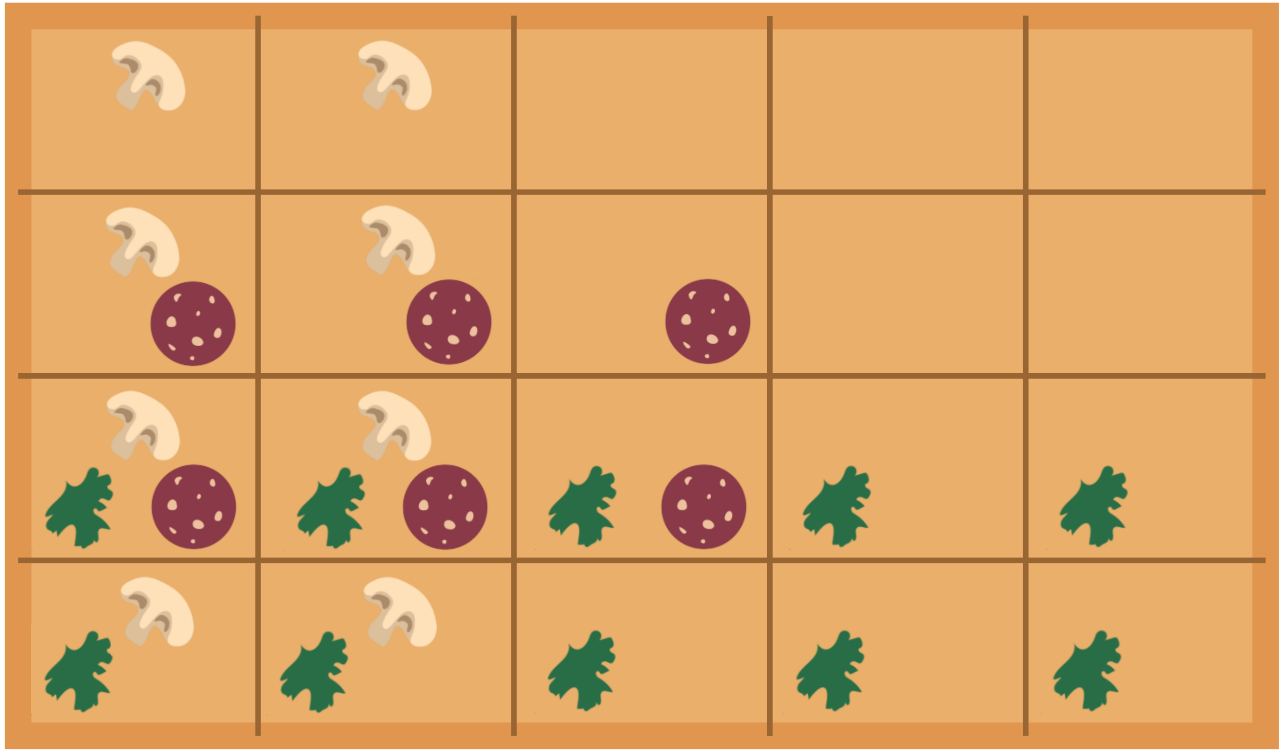
Walk-in warm-up: Omega Pizzeria

Three questions: What is the probability of getting a slice with:

1) No mushrooms
2) Spinach and no mushrooms
3) Spinach, when asking for slice with no mushrooms



Write down your answers.

We'll get to them in a few minutes.

Icons: CC, https://openclipart.org/detail/296791/pizza-slice

Announcements

Assignments:

- P4
  - Due Thu 3/28, 10 pm
- HW9 (written)
  - Out tonight, due Tue 4/2

Optional Probability (online)

Course Feedback:

- See Piazza post for mid-semester survey
- Note the separate TA posts

AI: Representation and Problem Solving

Probability

Instructors: Pat Virtue & Stephanie Rosenthal

Slide credits: CMU AI and http://ai.berkeley.edu

Previous Exposure to Probability

Previous topics

- Some version of notation
- Sample spaces, outcomes, sets, events
- Three axioms
- Theorem of total probability
- Partitions

$$\sum_{i=1}^{n} P(\mathrm{a}_i) = 1$$

Outline

1. Conditional probability
   - Pizza
   - Notation
   - Definition
2. Product rule & Bayes' theorem
   - Algebra
   - Why Bayes' theorem

Our probability toolbox

- Algebra
- Three axioms of probability
- Theorem of total probability
- Definition of conditional probability
- Product rule
- Bayes' theorem
- Chain rule

Piazza Poll 1

Three questions: What is the probability of getting a slice with:

1) No mushrooms

2) Spinach and no mushrooms

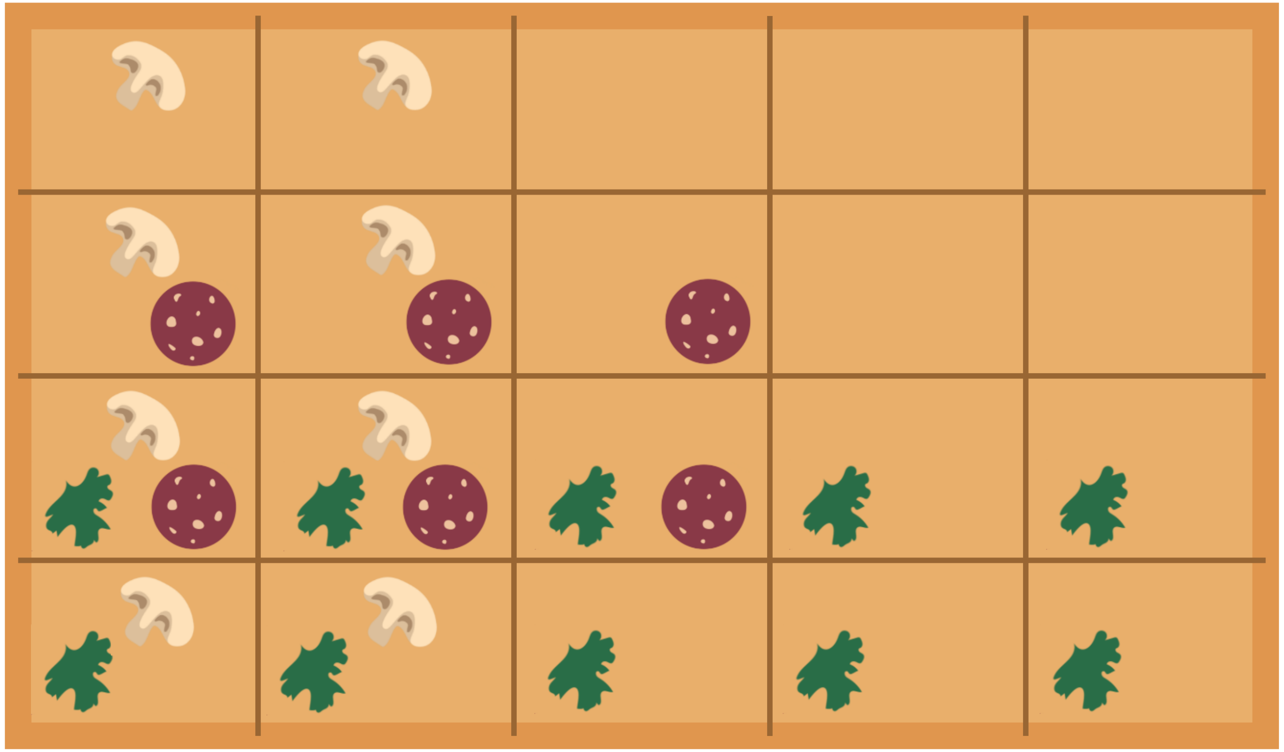
3) Spinach, when asking for slice with no mushrooms

How comfortable are you with this?

A: I’m good

B: So-so

C: A little freaked out



Icons: CC, https://openclipart.org/detail/296791/pizza-slice

Omega Pizzeria

Three questions: What is the probability of getting a slice with:

1) No mushrooms

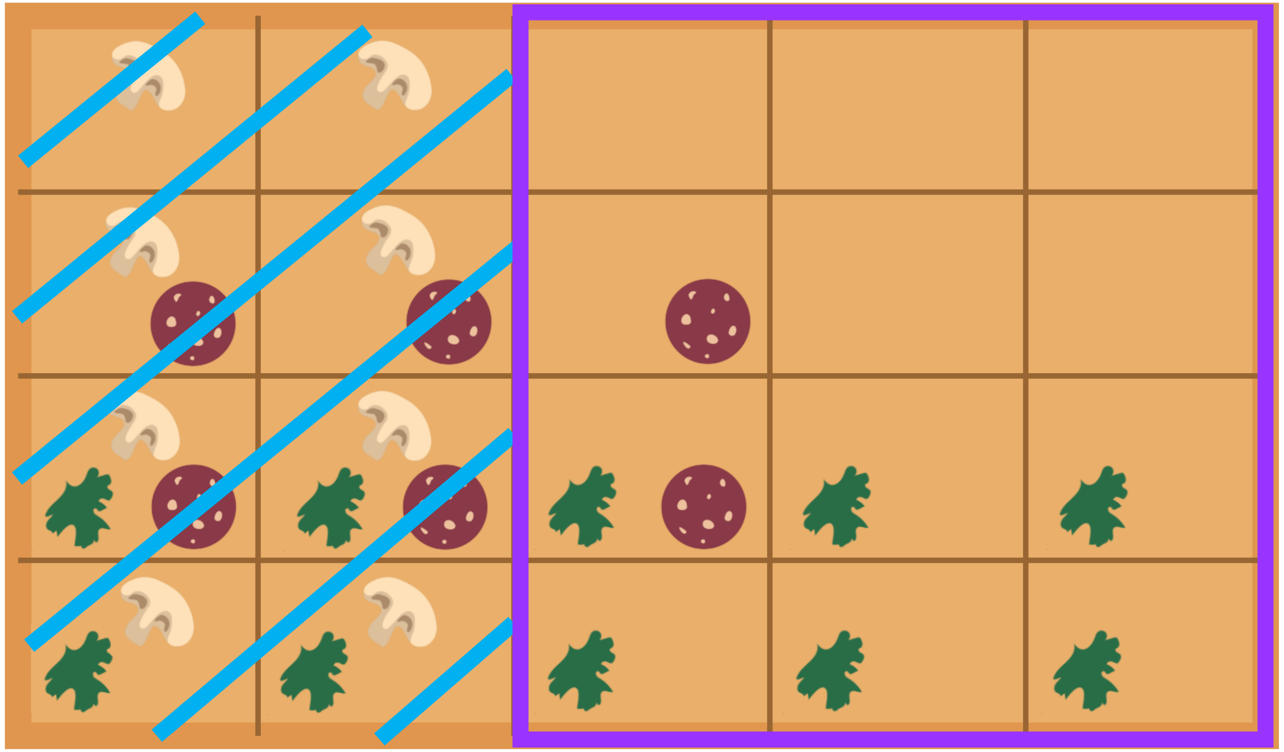
   3/5

2) Spinach and no mushrooms

   3/10

3) Spinach, when asking for slice with no mushrooms

   6/12



New information (condition)

Adjust sample space

Icons: CC, https://openclipart.org/detail/296791/pizza-slice

Omega Pizzeria

Formalize this a bit

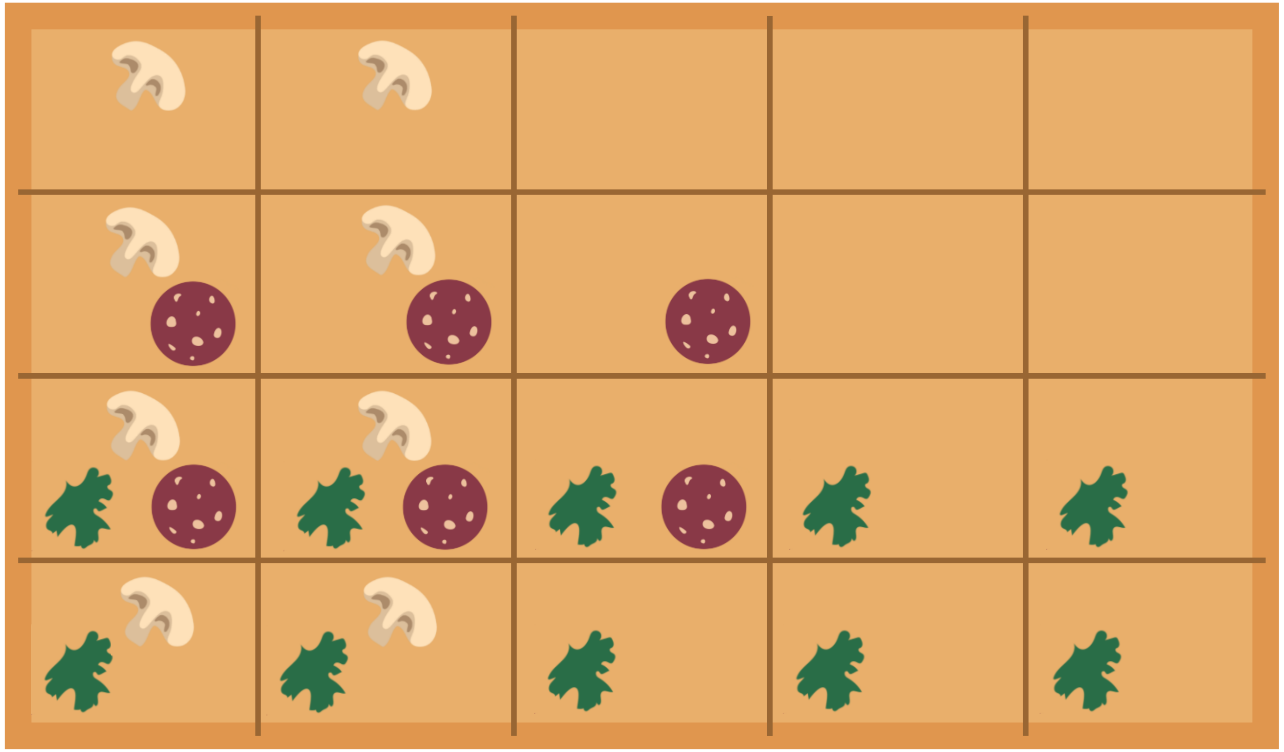
- $\underline{\Omega}$: whole pizza
- $\underline{S}$: Spinach random variable

  s_1: no spinach $-s$

  s_2: spinach $+s$
- M: Mushroom random variable

  m_1: no mushrooms

  m_2: mushrooms



Icons: CC, https://openclipart.org/detail/296791/pizza-slice

Omega Pizzeria

Formalize this a bit

- Ω: whole pizza
- Spinach partition

  s_1: no spinach

  s_2: spinach
- Mushroom partition

  m_1: no mushrooms

  m_2: mushrooms

Questions?

1) No mushrooms

   marginal

   $P(\underline{M} = \underline{m_1})$ $P(-m)$

2) Spinach and no mushrooms

   Joint

   $P(S = s_2, M = m_1)$ $P(+s) \cap P(-m)$

3) Spinach, when asking for slice with no mushrooms

   Conditional

   $P(S = s_2 \mid M = m_1)$

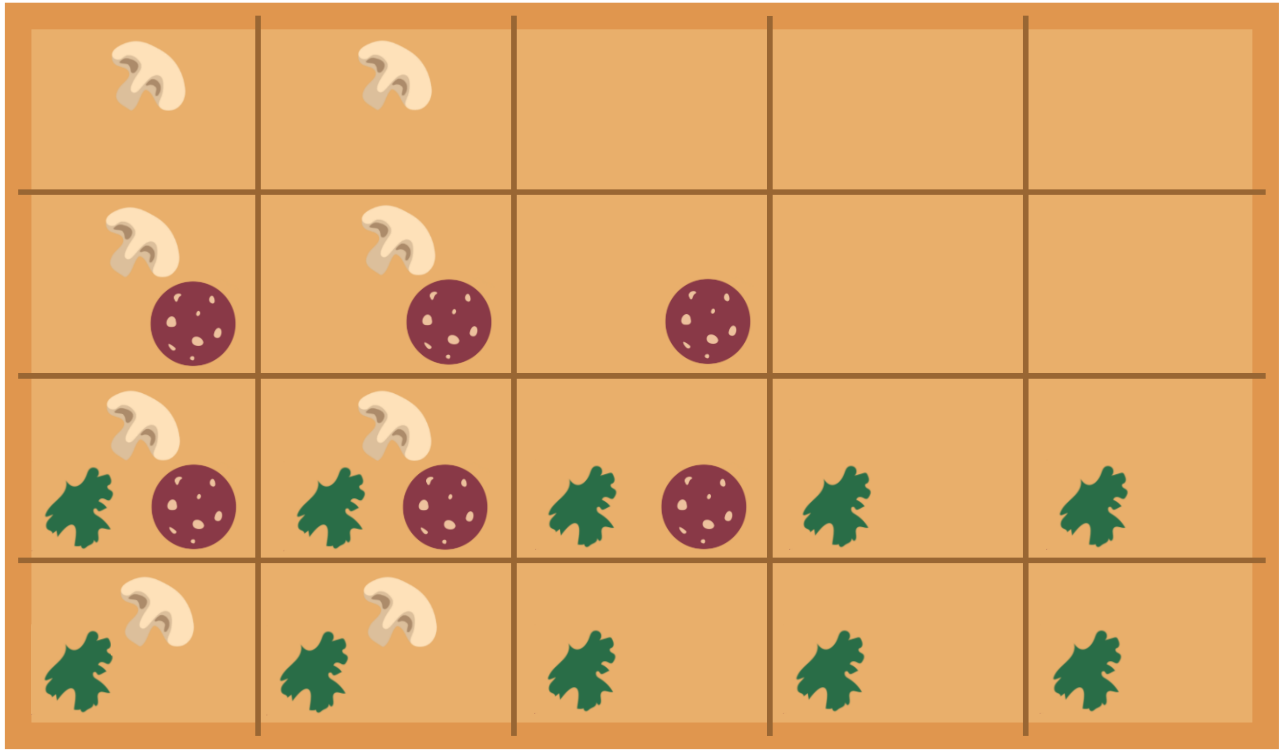
New notation alert!

Omega Pizzeria

More questions: What is the probability of getting a slice with:

1) No mushrooms
2) Spinach and no mushrooms
3) Spinach, when asking for slice with no mushrooms

- Mushrooms
- Spinach
- No spinach
- No spinach and mushrooms
- No spinach when asking for no mushrooms
- No spinach when asking for mushrooms
- Spinach when asking for mushrooms
- No mushrooms and no spinach



Icons: CC, https://openclipart.org/detail/296791/pizza-slice

Omega Pizzeria

You can answer all of these questions:

$P(M)$

m_1	12/20
m_2	

$P(S)$

s_1	
s_2	

$P(M,S)$

m_1	s_1	
m_1	s_2	6/20
m_2	s_1	
m_2	s_2	

$P(M|s_1)$

m_1	
m_2	

$P(M|s_2)$

m_1	
m_2	

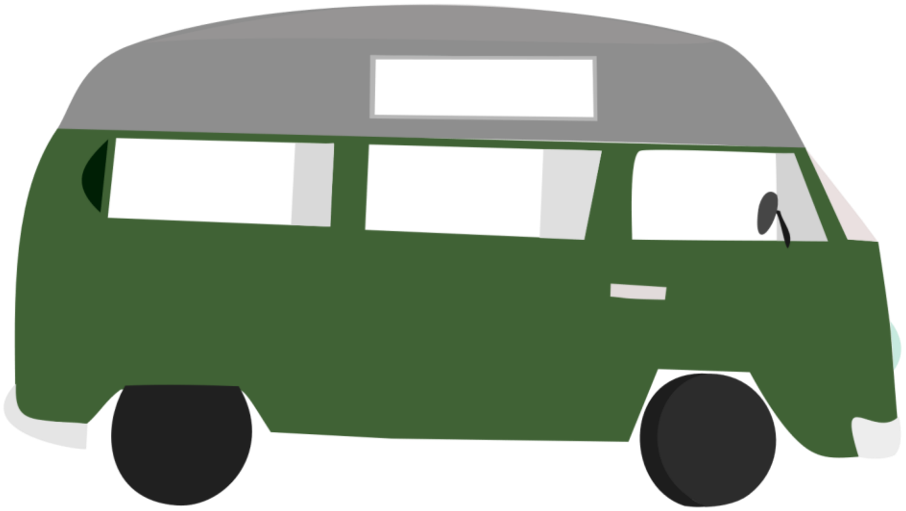
$P(S|m_1)$

s_1	
s_2	6/12

$P(S|m_2)$

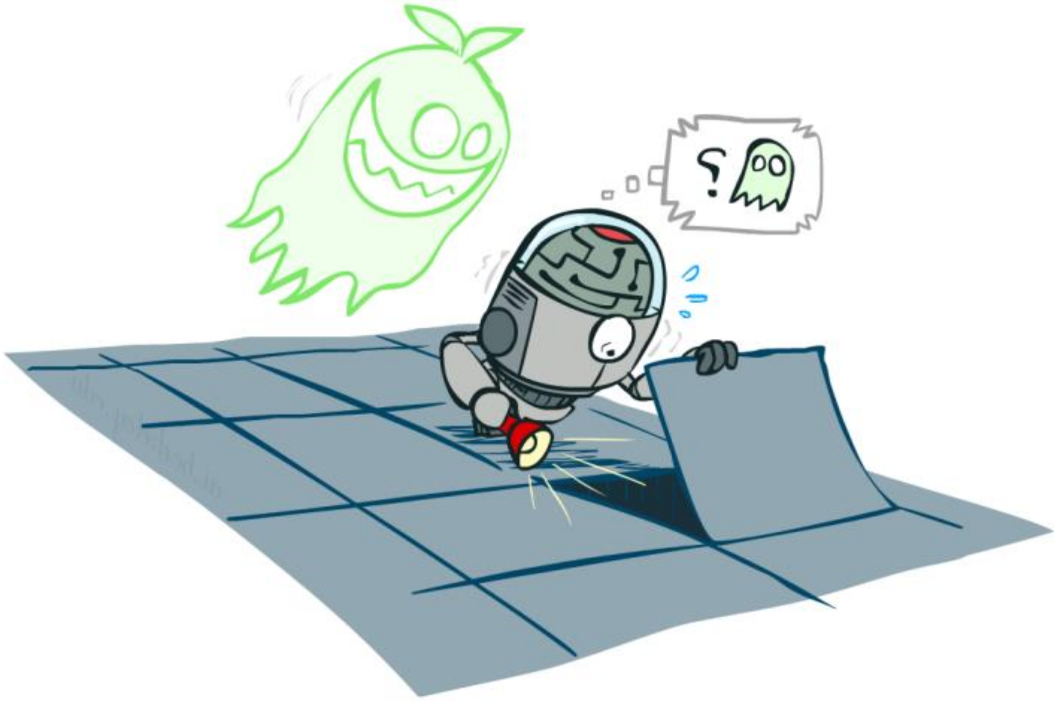
s_1	
s_2	

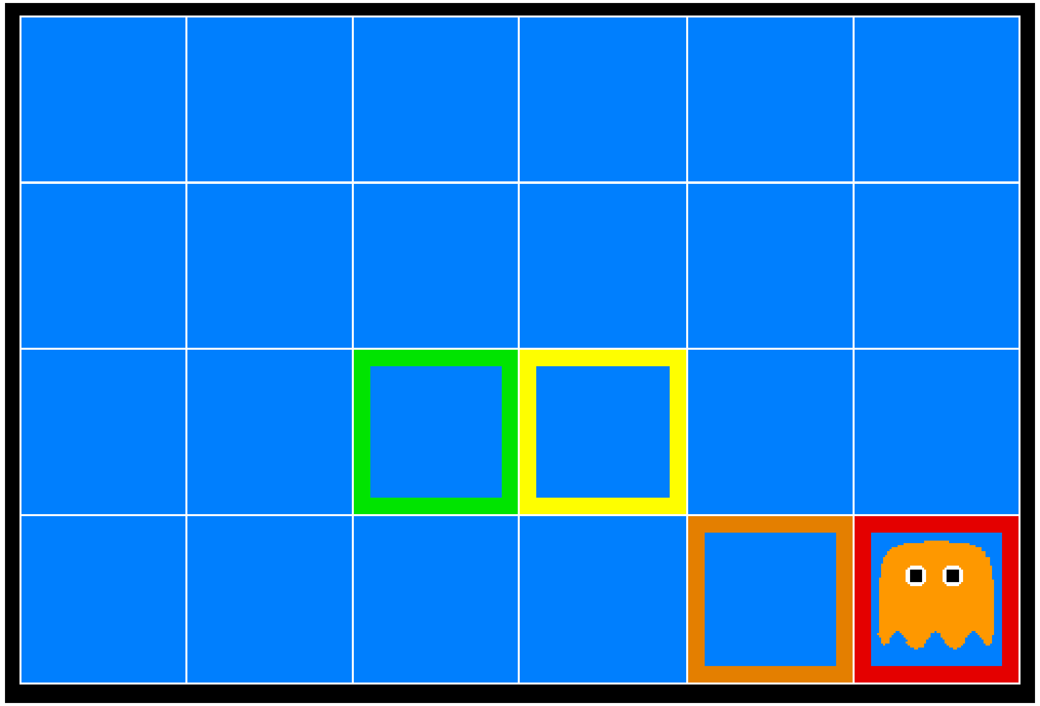
Sensor Probability



Sensor Probability

Ghostbusters project



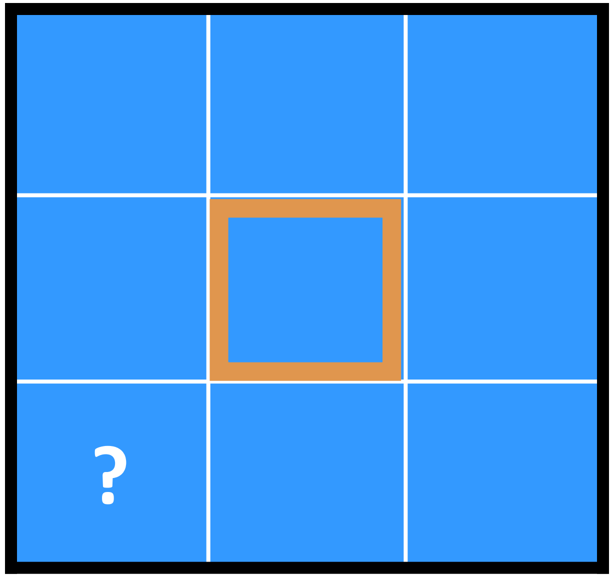


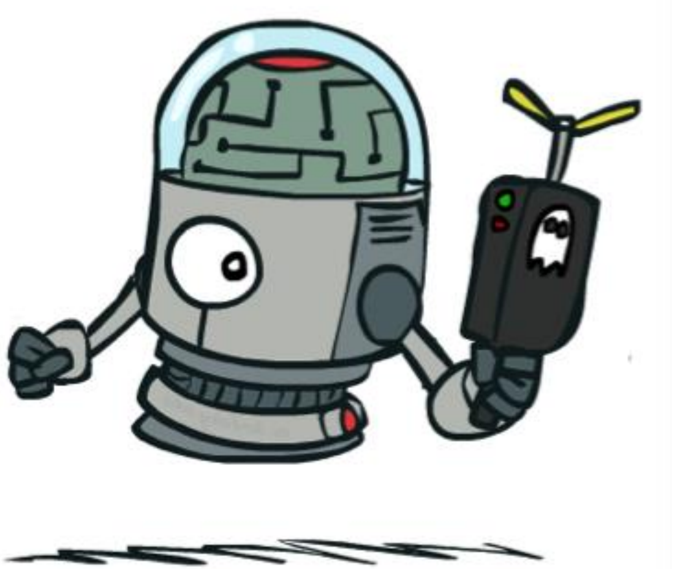
[ai.berkeley.edu demo: Ghostbuster – no probability (L12D1)]

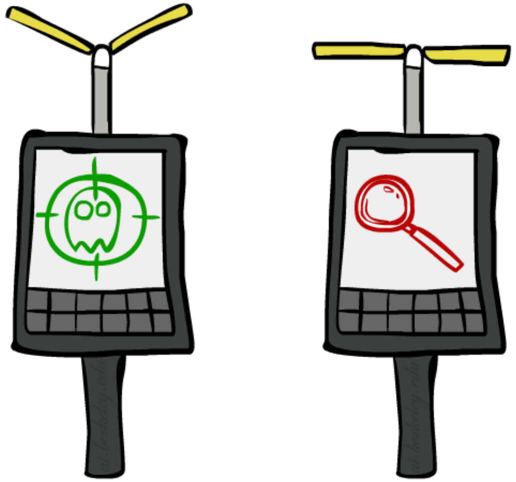
Sensor Probability

Ghostbusters project

- Want $P\left(\underline{g_{3,1}} \mid s_{2,2,Orange}\right)$
- Have $P\left(s_{2,2,Orange} \mid g_{3,1}\right)$
- We need more tools!





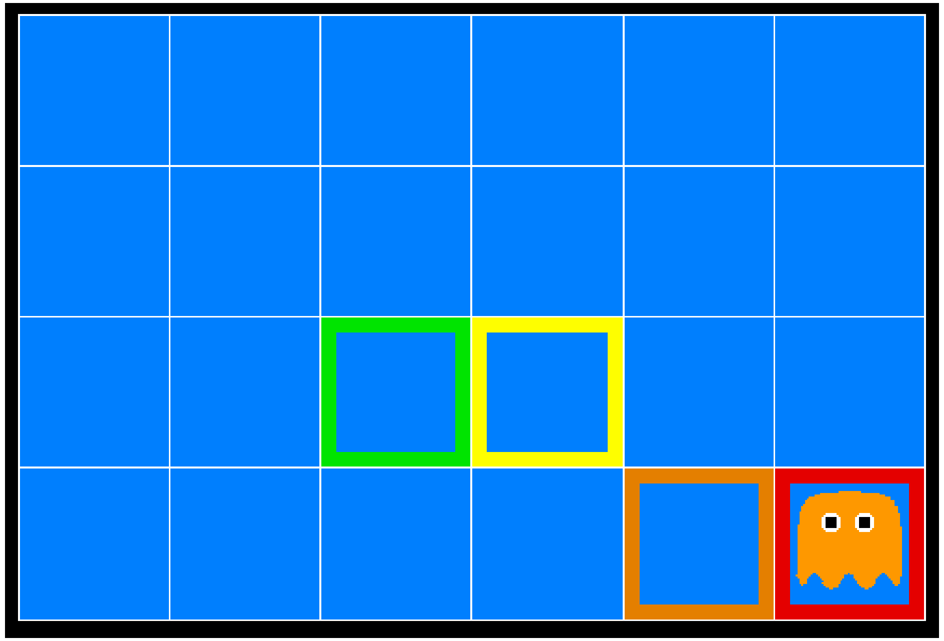


Inference in Ghostbusters

A ghost is in the grid somewhere

Sensor readings tell how close a square is to the ghost

- On the ghost: red
- 1 or 2 away: orange
- 3 or 4 away: yellow
- 5+ away: green



- Sensors are noisy, but we know P(Color(x,y) | DistanceFromGhost(x,y))

P(red | 3)	P(orange | 3)	P(yellow | 3)	P(green | 3)
0.05	0.15	0.5	0.3

Ghostbusters Demo

[Demo: Ghostbuster – no probability (L12D1)]

Definition of Conditional Probability

Definition:

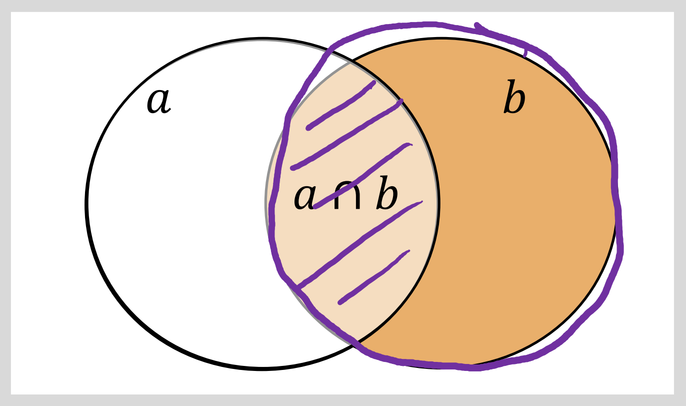
If $P(b) > 0$*, then the* **conditional probability** *of* a *given* b *is:*

$$P(a|b) = \frac{P(a,b)}{P(b)}$$

Counting: proportions

$$P(a) = \frac{Count(a)}{Count(\Omega)}$$

$$P(a|b) = \frac{Count(a \cap b)}{Count(b)}$$



Omega Pizzeria

Apply definition of conditional probability

- No mushrooms

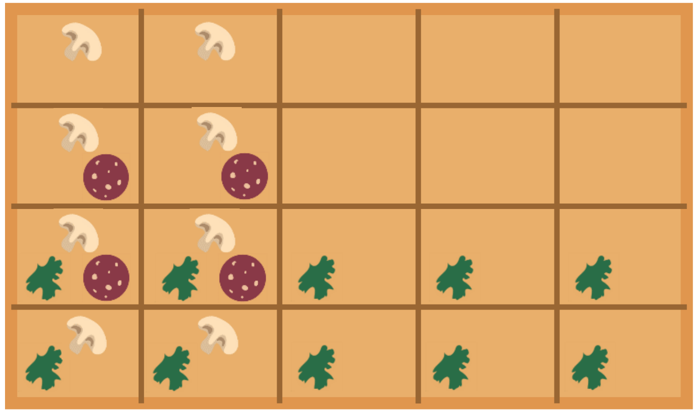
$$P(m_1) = \frac{12}{20}$$

- Spinach and no mushrooms

$$P(s_2, m_1) = \frac{6}{20}$$

- Spinach, when asking for slice with no mushrooms

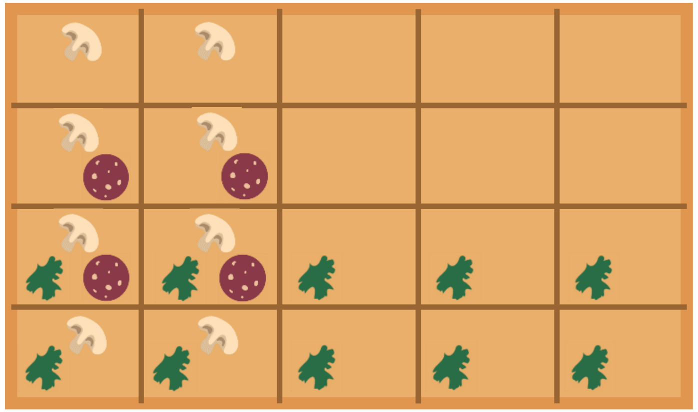
$$P(s_2 | m_1) = \frac{6}{12}$$



Conditional Probability: $P(a|b) = \frac{P(a,b)}{P(b)}$

Omega Pizzeria

Apply definition of conditional probability



- No mushrooms

$$\mathrm{P}(\mathrm{m}_1) = \frac{12}{20}$$

- Spinach and no mushrooms

$$\mathrm{P}(\mathrm{s}_2, \mathrm{m}_1) = \frac{6}{20}$$

Conditional Probability: $P(a|b) = \dfrac{P(a,b)}{P(b)}$

- Spinach, when asking for slice with no mushrooms

$$\mathrm{P}(\mathrm{s}_2|\mathrm{m}_1) = \frac{6}{12}$$

$$P(s_2|m_1) = \frac{P(s_2, s_1)}{P(s_1)} = \frac{\frac{6}{20}}{\frac{12}{20}} = \frac{6}{12}$$

Definition of Conditional Probability

Definition:

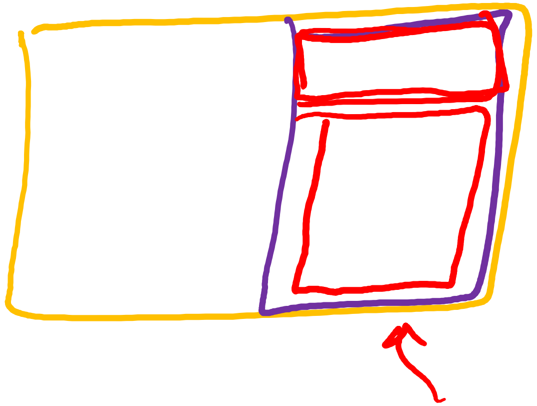
If $P(B) > 0$, then the **conditional probability** *of A given B is:*

$$P(A|B) = \frac{P(A,B)}{P(B)}$$

http://achievementgen.com/360/

Normalization Trick

P(X | Y=-y) ?



$P(X,Y)$

X	Y	P
~~+x~~	~~+y~~	~~0.2~~
+x	-y	0.3
~~-x~~	~~+y~~	~~0.4~~
-x	-y	0.1

SELECT the joint probabilities matching the evidence

+x -y 0.3
-x -y 0.1

0.4

NORMALIZE the selection (make it sum to one)

+x -y 3/4
-x -y 1/4

To Normalize

(Dictionary) To bring or restore to a normal condition

All entries sum to ONE

Procedure:

- Step 1: Compute Z = sum over all entries
- Step 2: Divide every entry by Z

Example 1

W	P
sun	0.2
rain	0.3

Normalize $\longrightarrow$ Z = 0.5

W	P
sun	0.4
rain	0.6

- ## Example 2

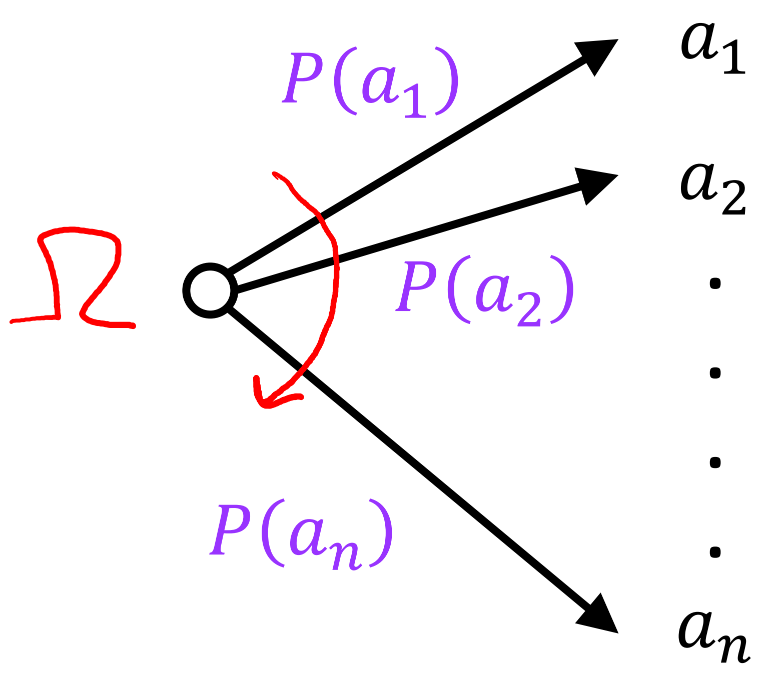
T	W	P
hot	sun	20
hot	rain	5
cold	sun	10
cold	rain	15

Normalize $\longrightarrow$ Z = 50

T	W	P
hot	sun	0.4
hot	rain	0.1
cold	sun	0.2
cold	rain	0.3

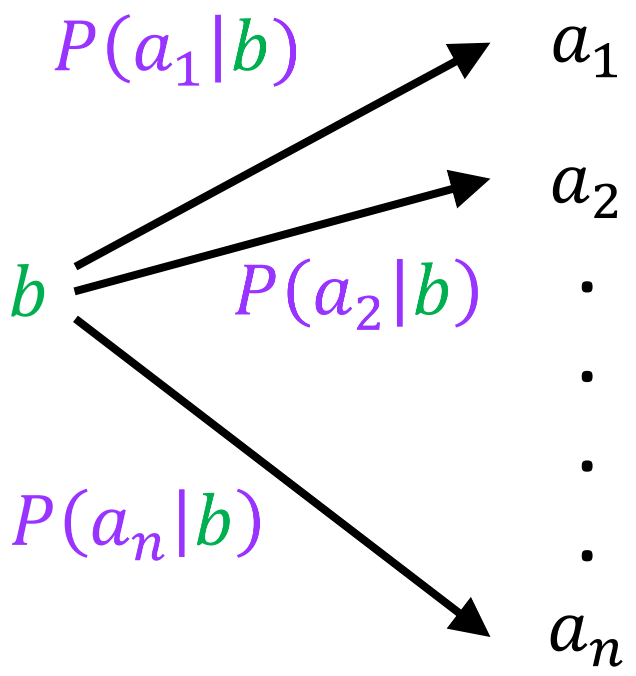
Sum over all values of a random variable

For all possible disjoint values of a random variable, A: $a_1, a_2, \ldots a_n$ (a partition of the sample space Ω), $\sum_{i=1}^{n} P(a_i) = 1$.



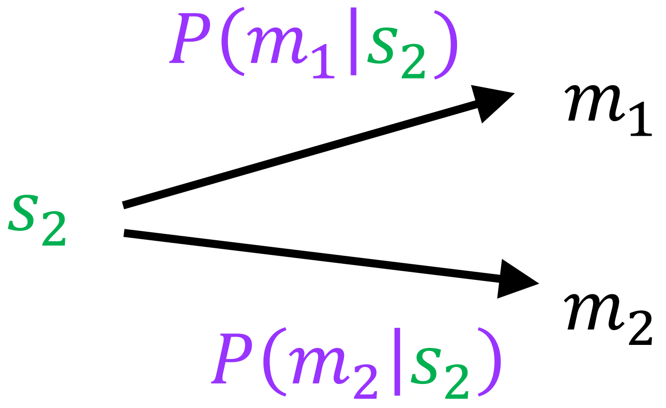
Partition given Event, Still Sums to One

For any event b and partition $a_1, a_2, \dots a_n$ of the sample space Ω, $\sum_{i=1}^{n} P(a_i \mid b) = 1$:

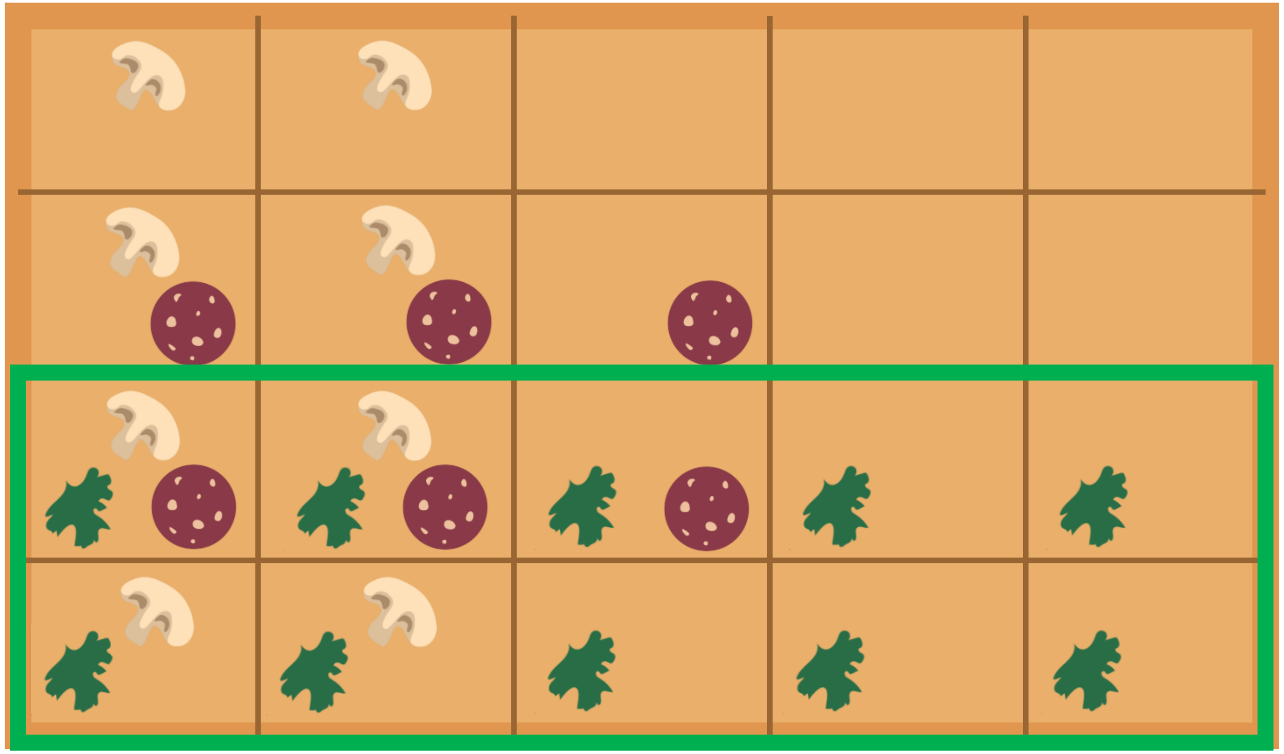


Partition given Event, Still Sums to One

For any event b and partition $a_1, a_2, \ldots a_n$ of the sample space Ω,
$\sum_{i=1}^{n} P(a_i \mid b) = 1$:



$$P(m_1|s_2) + P(m_2|s_2)$$



Icons: CC, https://openclipart.org/detail/296791/pizza-slice

Piazza Poll 2

How many valid equations can we compose using:

$$P(x),\ P(y),\ P(x,y),\ P(x|y),\ P(y|x) \text{ and } =,\ \times,\ \div$$

First one: $P(x|y) = P(x,y)/P(y)$

$P(y|x) = P(x,y)/P(x)$

A $\underline{P(x|y)P(y)} = \underline{P(x,y)}$

B $\underline{P(y|x)P(x)} = \underline{P(x,y)}$

$P(x|y) = P(y|x)P(x)/P(y)$

A $P(x|y)P(y) = P(y|x)P(x)$

$P(y|x) = P(x|y)P(y)/P(x)$

A) 2

B) 4

C) 7

D) Other

At most one use per probability term

e.g. Not $P(x) = P(x)$

Must be different

e.g. Cannot also use

$P(x,y)/P(y) = P(x|y)$

Product Rule and Bayes’ Theorem

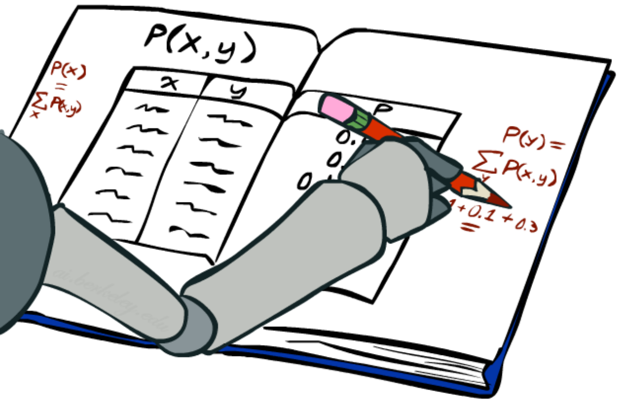
Reformulations of definition of conditional probability

Product rule:

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

Bayes’ theorem:

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

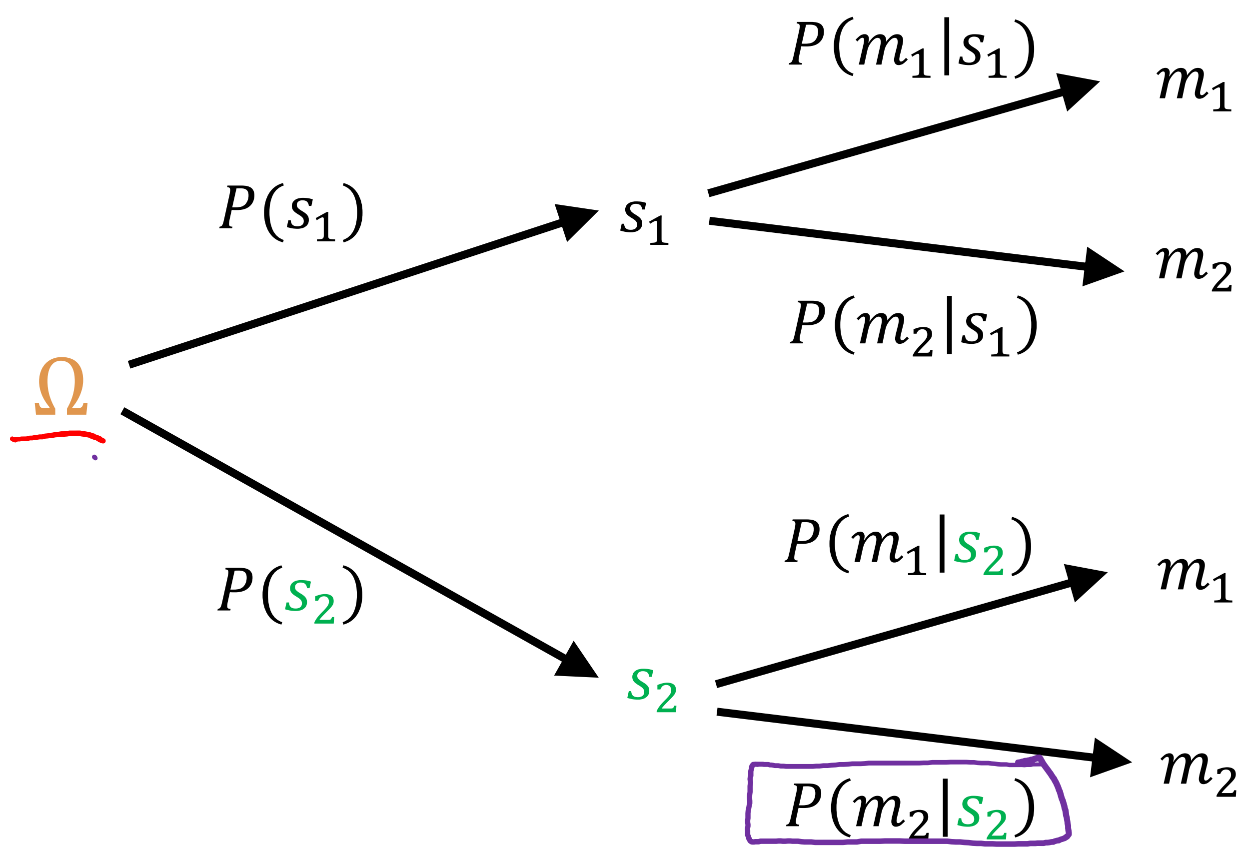


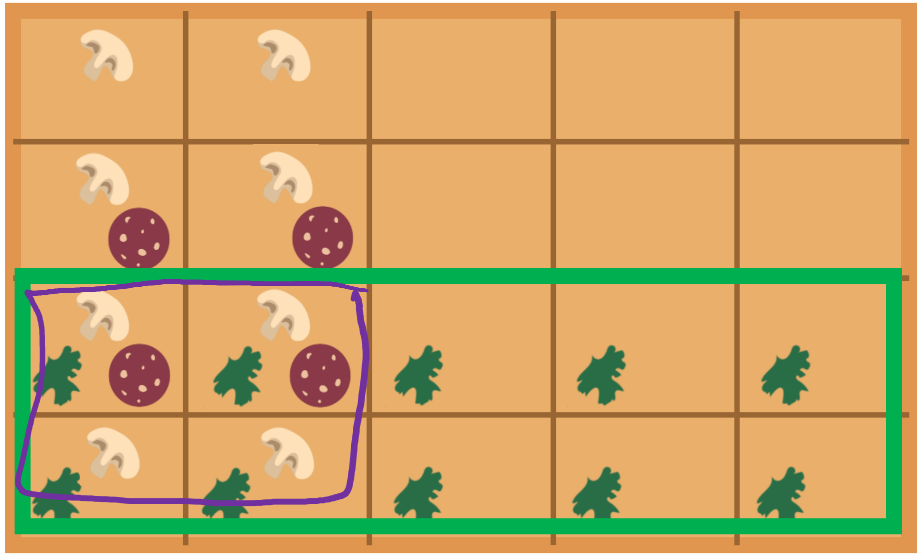
http://achievementgen.com/360/

Product Rule: Tree

Product rule:

$$\underline{P(a,b)} = \underline{P(a|b)}\,\underline{P(b)}$$





$$P(s_2)\,P(m_1|s_2) = \frac{1}{2}\cdot\frac{6}{10}$$

$$P(m_1, s_2) = \frac{6}{20}$$

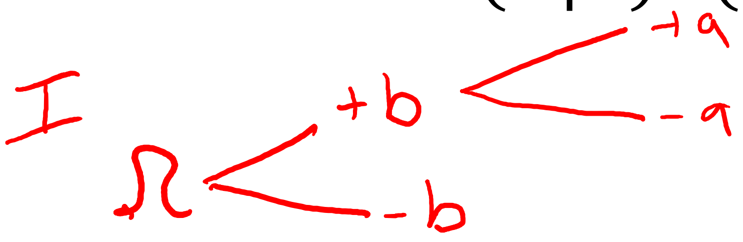
Icons: CC, https://openclipart.org/detail/296791/pizza-slice

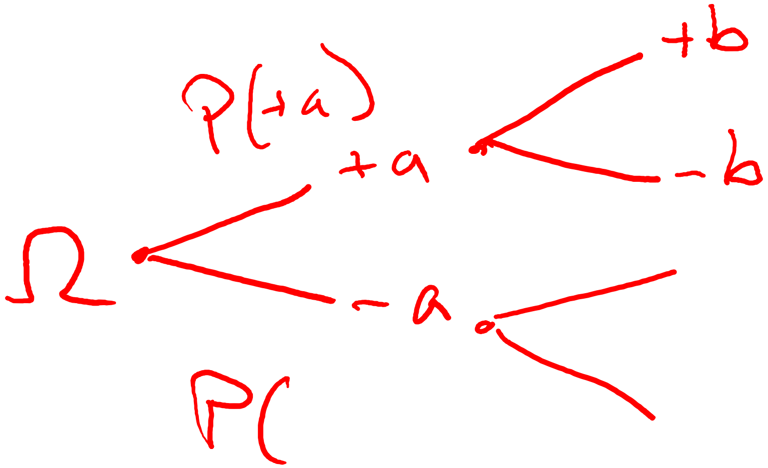
Exercise: Product Rule: Tree

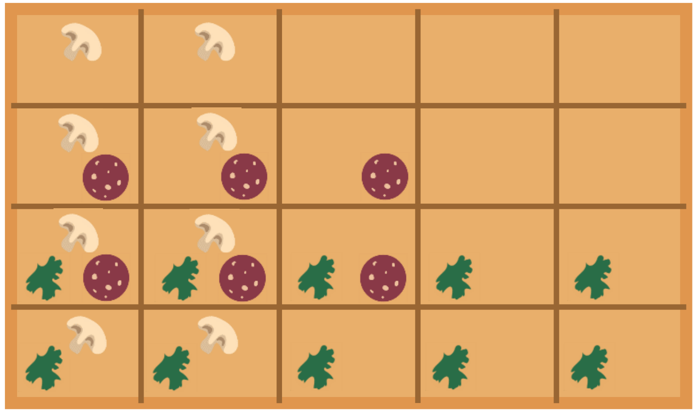
Demonstrate, using trees, that product rule works both ways:

$$\underline{P(A,B)} = P(A|B)P(B) \quad \mathrm{I}$$

$$= P(B|A)P(A) \quad \mathrm{II}$$







Icons: CC, https://openclipart.org/detail/296791/pizza-slice

$$P(b) = \sum_i \underbrace{P(b|a_i)\,P(a_i)}$$
$$= \sum_i P(a_i, b) = P(b)$$

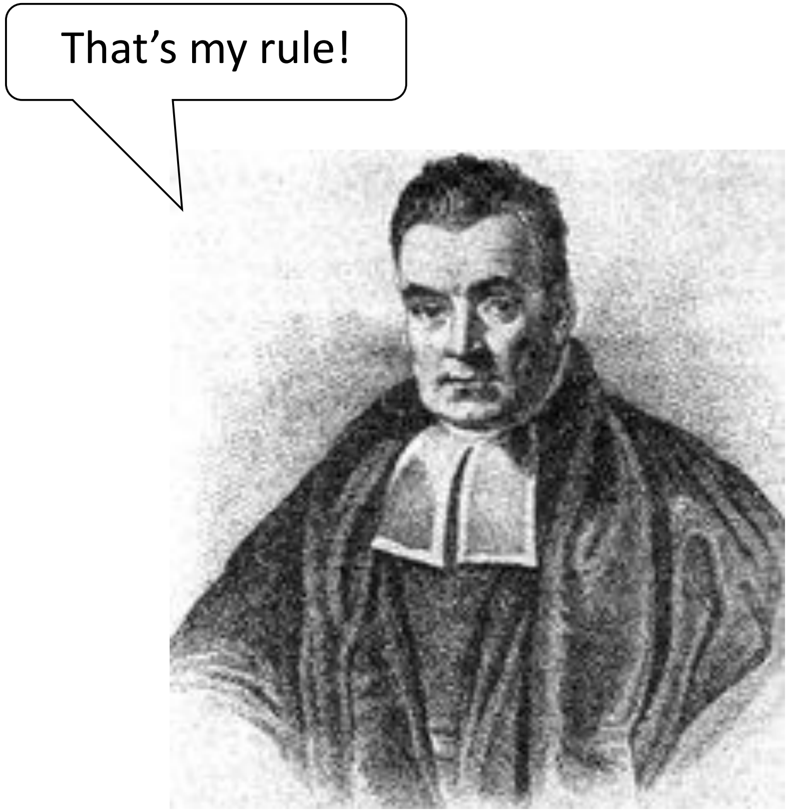
Bayes' Theorem

Bayes' theorem:

$$P(a_1|b) = \frac{P(b|a_1)P(a_1)}{P(b)}$$

Also:

$$P(a_1|b) = \frac{P(b|a_1)P(a_1)}{\sum_{i=1}^{n} P(b|a_i)P(a_i)}$$



Why is this at all helpful?

- Lets us build one conditional from its reverse
- Often one conditional is tricky but the other one is simple
- Describes an "update" step from prior $P(a)$ to posterior $P(a \mid b)$ ←
- Foundation of many probabilistic systems

Inference with Bayes' Theorem

Example: Diagnostic probability from *causal probability:*

$$P(cause \mid effect) = \frac{P(effect \mid cause)\, P(cause)}{P(effect)}$$

Example:

- Your friend has a stiff neck ($+s$)
- Knowledge:

$$P(+s) = 0.01$$
$$P(+m) = 0.0001$$
$$P(+s \mid +m) = 0.8$$

$$P(+m \mid +s) = \frac{P(+s|+m)\, P(+m)}{P(+s)}$$

$$= \frac{0.8 \times 0.0001}{0.01} = 0.008$$

- What are the chances your friend has meningitis ($+m$)?

Probabilistic Inference

Probabilistic inference: compute a desired probability from other known probabilities (e.g. conditional from joint)

We generally compute conditional probabilities

- P(on time | no reported accidents) = 0.90
- These represent the agent's *beliefs* given the evidence

Probabilities change with new evidence:

- P(on time | no accidents, 5 a.m.) = 0.95
- P(on time | no accidents, 5 a.m., raining) = 0.80
- Observing new evidence causes *beliefs to be updated*

Inference by Enumeration

P(W)?

P(W)
sun
rain

P(W | winter)?

P(W | winter, hot)?

P(S, T, W)

S	T	W	P
summer	hot	sun	0.30
summer	hot	rain	0.05
summer	cold	sun	0.10
summer	cold	rain	0.05
winter	hot	sun	0.10
winter	hot	rain	0.05
winter	cold	sun	0.15
winter	cold	rain	0.20

Inference by Enumeration

General case:

- Evidence variables: $E_1 \dots E_k = e_1 \dots e_k$
- Query* variable: Q
- Hidden variables: $H_1 \dots H_r$

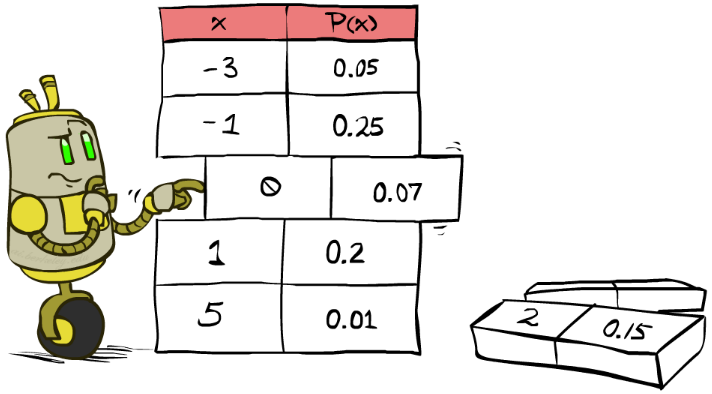
$$\left.\begin{array}{l} E_1 \dots E_k = e_1 \dots e_k \\ Q \\ H_1 \dots H_r \end{array}\right\} \begin{array}{c} X_1, X_2, \dots X_n \\ \textit{All variables} \end{array}$$

- We want:

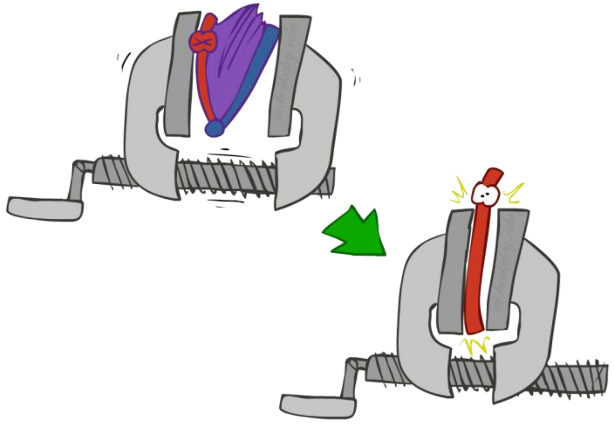
** Works fine with multiple query variables, too*

$$P(Q|e_1 \dots e_k)$$

- Step 1: Select the entries consistent with the evidence



- Step 2: Sum out H to get joint of Query and evidence



$$P(Q, e_1 \dots e_k) = \sum_{h_1 \dots h_r} P(\underbrace{Q, h_1 \dots h_r, e_1 \dots e_k}_{X_1, X_2, \dots X_n})$$

- Step 3: Normalize

$$\times \frac{1}{Z}$$

$$Z = \sum_q P(Q, e_1 \cdots e_k)$$

$$P(Q|e_1 \cdots e_k) = \frac{1}{Z} P(Q, e_1 \cdots e_k)$$

Inference by Enumeration

- Obvious problems:
  - Worst-case time complexity $O(d^n)$
  - Space complexity $O(d^n)$ to store the joint distribution

Tools Summary

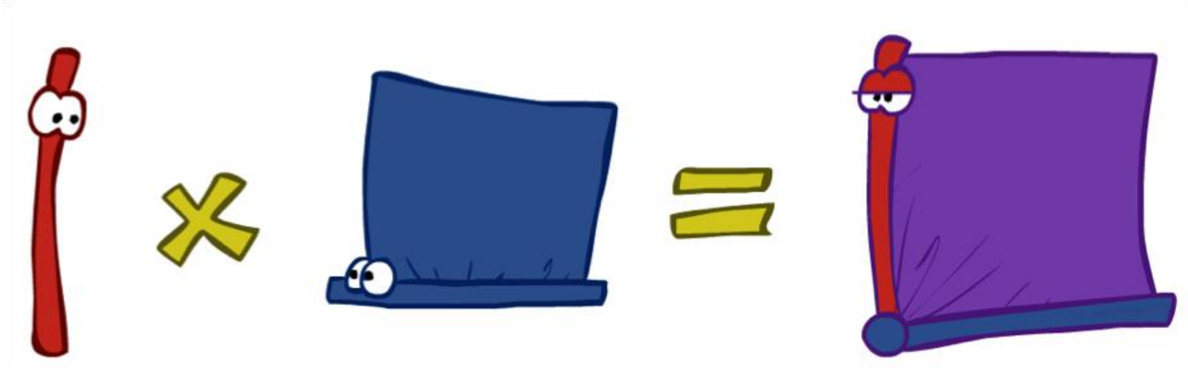
Adding to our toolbox

1. Definition of conditional probability $$P(A|B) = \frac{P(A,B)}{P(B)}$$
2. Product Rule $$P(A,B) = P(A|B)P(B)$$
3. Bayes' theorem $$P(B|A) = \frac{P(A|B)P(B)}{P(A)}$$
4. Chain Rule…

The Product Rule

Sometimes have conditional distributions but want the joint

$$P(y)P(x|y) = P(x,y) \qquad \Longleftrightarrow \qquad P(x|y) = \frac{P(x,y)}{P(y)}$$



The Product Rule

$$P(y)P(x|y) = P(x, y)$$

Example:

$P(W)$

R	P
sun	0.8
rain	0.2

$P(D|W)$

D	W	P
wet	sun	0.1
dry	sun	0.9
wet	rain	0.7
dry	rain	0.3

$P(D, W)$

D	W	P
wet	sun	
dry	sun	
wet	rain	
dry	rain	

The Chain Rule

More generally, can always write any joint distribution as an incremental product of conditional distributions

$$P(x_1, x_2, x_3) = P(x_1)P(x_2|x_1)P(x_3|x_1, x_2)$$

$$P(x_1, x_2, \ldots x_n) = \prod_i P(x_i|x_1 \ldots x_{i-1})$$

Ghostbusters, Revisited

Let’s say we have two distributions:

- Prior distribution over ghost location: P(G)
  - Let’s say this is uniform
- Sensor reading model: P(R | G)
  - Given: we know what our sensors do
  - R = reading color measured at (1,1)
  - E.g. P(R = yellow | G=(1,1)) = 0.1

0.11	0.11	0.11
0.11	0.11	0.11
0.11	0.11	0.11

We can calculate the posterior distribution P(G|r) over ghost locations given a reading using Bayes’ rule:

$$P(g|r) \propto P(r|g)P(g)$$

0.17	0.10	0.10
0.09	0.17	0.10
<0.01	0.09	0.17

[Demo: Ghostbuster – with probability (L12D2)]

Demo Ghostbusters with Probability