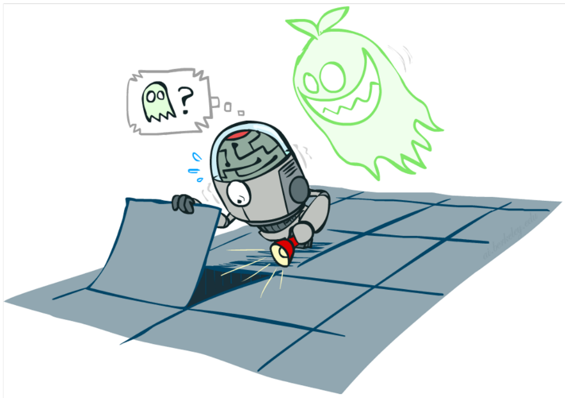
CS440 – Probabilistic Reasoning



Purpose: Consider if pacman can only detect the location of the ghost with a noisy sensor. The sensor does not always give exact readings. Pacman would need to use probability and reasoning techniques to determine the most likely position of the ghost.

The purpose of this module is to introduce the basics of probability and how we can reason with probabilistic results.

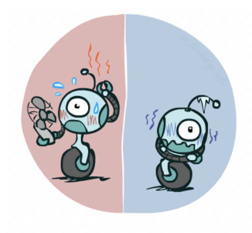
Knowledge: This module will help you become familiar with the following content knowledge:

- Probability basics
- Inference by enumeration
- Bayes' rule

Activity1 - Probability Basics:
With your group perform the following tasks and answer the questions. You will be reporting your answers back to the class in 30 minutes.

1. A probability distribution associates a probability with each value. For example:

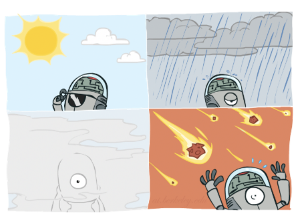
Temperature:



$P(T)$

T	P
hot	0.5
cold	0.5

Weather:



$P(W)$

W	P
sun	0.6
rain	0.1
fog	0.3
meteor	0.0

What do we know about the sum of all the probabilities in a distribution?

2. A joint distribution over a set of variable $X_1, X_2, ...X_n$ specifies a real number for each assignment (or outcome) such that:

$$P(x_1, x_2, ...x_n) \geq 0$$
$$\sum_{(x_1, x_2, ..., x_n)} P(x_1, x_2, ..., x_n) = 1$$

For example:

$P(T, W)$

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

This says that the probability that it is hot and sunny is .4

What would be the size of the distribution if we had n variables each with a domain size of d?

An event is a set of outcomes and the probability of an event is the sum of the probabilities of all the outcomes in the event.

From the joint distribution table above give the following:

The probability that it is hot and rainy.
The probability that it is hot.
The probability that it is hot or sunny.

3. Marginal distributions are sub-tables which eliminate variables. Marginalization (summing out) combines collapsed rows by summing:

$P(T, W)$

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

$$P(t) = \sum_s P(t, s)$$

$$P(s) = \sum_t P(t, s)$$

$P(T)$

T	P
hot	0.5
cold	0.5

$P(W)$

W	P
sun	?
rain	?

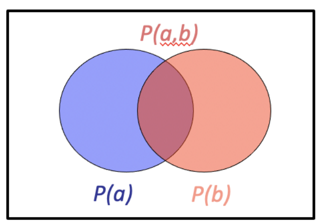
$$P(X_1 = x_1) = \sum_{x_2} P(X_1 = x_1, X_2 = x_2)$$

What is the probability of sun?
What is the probability of rain?

4. A conditional probability is the probability of an event a given the fact that another event b has occurred. The notation is $P(a|b)$.

There is a simple relationship between joint and conditional probabilities:

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



$P(T, W)$

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

$$P(W = s | T = c) = \frac{P(W = s, T = c)}{P(T = c)} \quad = ?$$

What is the probability of sunny weather given that it is cold?

5. A conditional distribution is a probability distribution over some variables given fixed values of other variables:

Conditional Distributions

$P(W|T)$

$P(W|T = hot)$

W	P
sun	0.8
rain	0.2

$P(W|T = cold)$

W	P
sun	?
rain	?

Joint Distribution

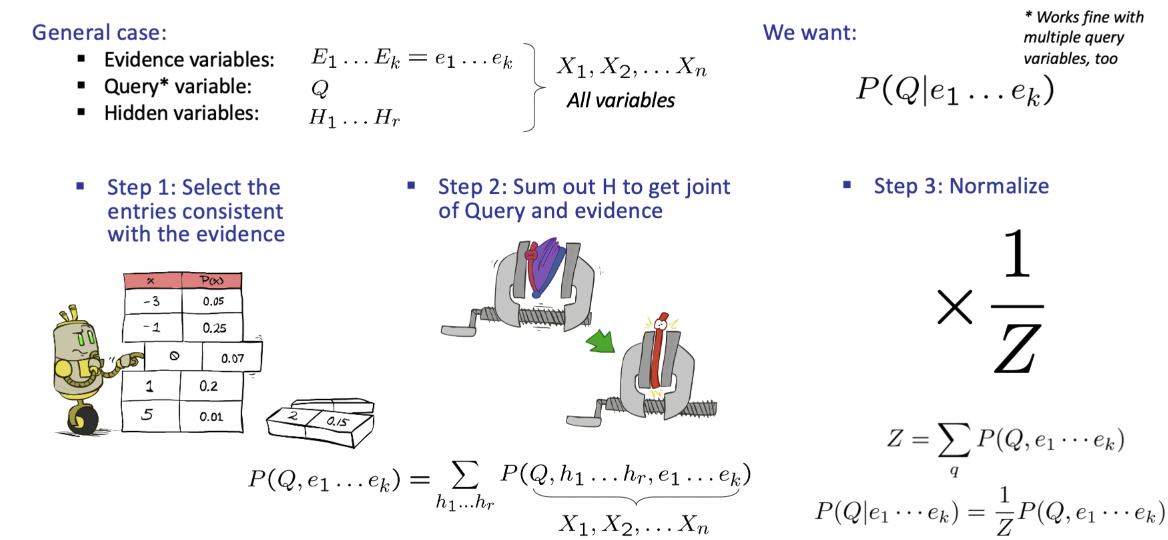
$P(T, W)$

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

What is the probability of sun given it is cold?
What is the probability of rain given it is cold?

Activity2 - Inference by Enumeration:
With your group perform the following tasks and answer the questions. You will be reporting your answers back to the class in 15 minutes.

1. Given evidence $e_1, e_2, ...e_k$ that has been observed, we want to infer the probability of some other variable(s) Q. The enumeration method is diagramed below:



So given the joint distribution:

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

Give the joint distribution table for $P(W)$
Give the table for $P(W|winter)$
Give the table for $P(W|winter, hot)$
What is the time and space complexity of this algorithm if n variables with d possible values each?

Activity3 - The Product, Chain, and Bayes' Rules:
With your group perform the following tasks and answer the questions. You will be reporting your answers back to the class in 30 minutes.

1. Sometimes we have conditional distributions but we want the joint distribution. The product rule allows us to construct the joint:

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

   or equivalently $P(x|y) = \frac{P(x,y)}{P(y)}$

   Complete the following joint distribution:

   $P(W)$

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	

2. A more general version is the chain rule which says:

   $$P(x_1, x_2, x_3) = P(x_1)P(x_2|x_1)P(x_3|x_1, x_2)$$
   $$P(x_1, x_2, ..., x_n) = \prod_i P(x_i|x_1...x_{i-1})$$

   Use the product rule multiple times to show the chain rule.

3. There are two ways to factor a joint distribution over two variables:

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

   Dividing we get Bayes' rule: $P(x|y) = \frac{P(y|x)}{P(y)}P(x)$

   Given:

   $P(W)$

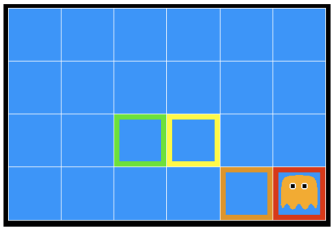
R	P
sun	0.8
rain	0.2

   $P(D|W)$

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

   What is $P(W|dry)$?

4. We can now apply our knowledge to the following scenario: Consider a ghost is in the grid somewhere and sensor readings tell how close a square is to the ghost. The sensor reads "red" if the ghost is on the square. The sensor reads "orange" if the ghost is 1 or 2 squares away. The sensor reads "yellow" if the ghost is 3 or 4 squares away. The sensor reads "green" if the ghost is 5 or more squares away.



The sensors, however, are noisy, which means that they are not always exactly accurate. We do know P(Color | Distance) which is the probability that the sensor reads a color given the distance of the ghost. For example:

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

So suppose that we have two distributions: a prior distribution over ghost location, $P(G)$ and a sensor reading model, $P(R|G)$.

The sensor model uses what we know our sensors do. So if R is the reading color, like yellow measured at (1,1) then we can lookup $P(R = yellow|G = (1, 1))$

We can use the tables to calculate the *posterior distribution* $P(G|r)$ over the ghost locations to give us the probabilities of there being a ghost given the sensor reading that we received.

By Bayes' rule: $P(g|r) \propto P(r|g)P(g)$.

Note that we are only using proportionality rather than the exact probability value since we just care about how the probabilities in the various locations compare with each other to give the most likely ghost location.