CSE422: Artificial Intelligence

Probability

You would probably need it

BRAC UNIVERSITY
Inspiring Excellence

Motivation

Why do we need probability in AI?

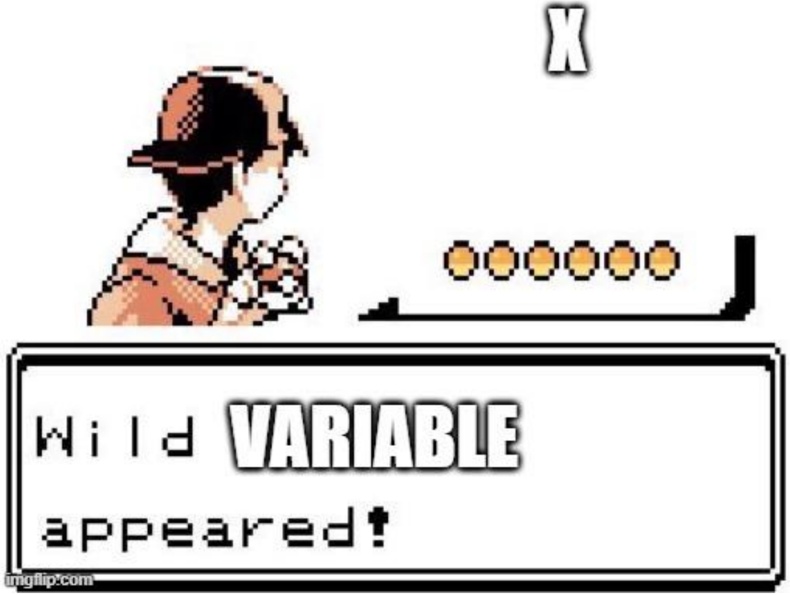
To be certain of how uncertain we are.

Applications:

- Speech Recognition
- Object Tracking
- Diagnostics
- Genetics
- ... and countless others!
- To ensure that STA201 does not go to waste

BRAC UNIVERSITY
Inspiring Excellence

Random Variables



Random Variables

- Random variables are a way to represent an event
- The values of the variable are the possible scenarios that can take place in the event
- A probability of occurrence is associated with each scenario
- The sum of probability of all the scenarios in the event is always 1

Random Variables

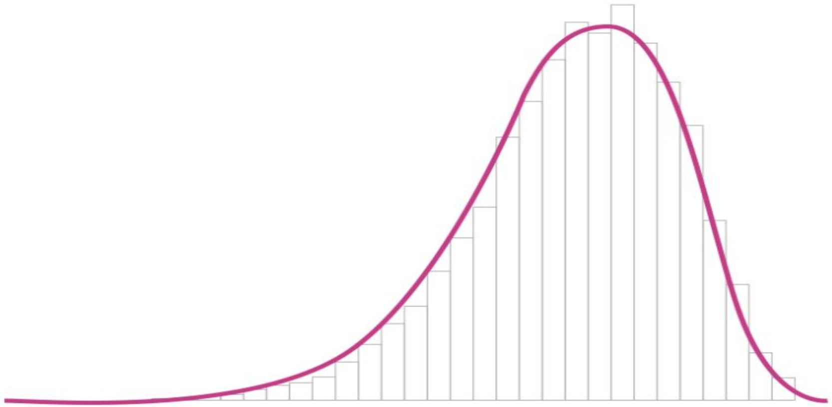
Discrete Random Variable

Let ***X*** be a random variable which represents the event how you are feeling right now

- P(***X*** = Bored) = 0.15
- P(***X*** = Excited) = 0.05
- P(***X*** = Sleepy) = 0.1
- P(***X*** = Depressed because you still have 30+ more lectures to watch) = 0.70

Continuous Random Variable

Let ***Y*** be a random variable which represents the score that you will receive in the next quiz



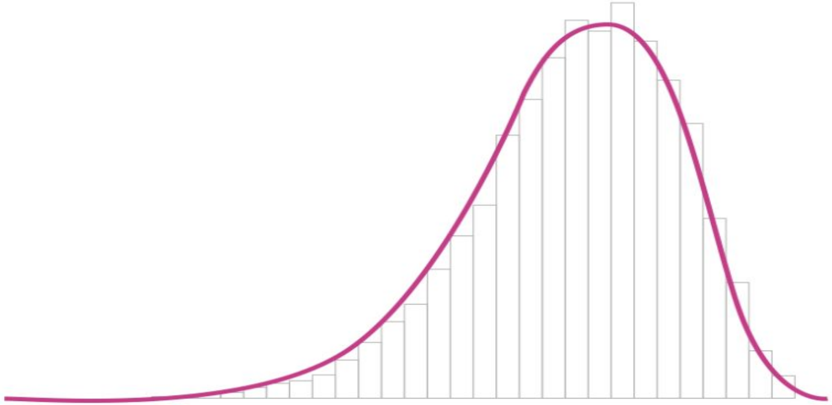
Probability Distributions

It is the distribution of the probabilities of each of the scenarios of the event.

Let ***X*** be a random variable which represents the event how you are feeling right now

P(X)	
Bored	0.15
Excited	0.05
Sleepy	0.1
Depressed because you still have 30 more lectures to watch	0.70

Let ***Y*** be a random variable which represents the score that you will receive in the next quiz



Joint Distributions

These are probability distributions where two or more random variables are being considered.

Will I eat right now?

Considerations:

- Am I hungry right now?
- Am I bored right now?
- Am I going to stream a movie/series right now?

BRAC UNIVERSITY

Inspiring Excellence

Joint Distributions

RV ***H*** represents if I am hungry and it assumes the following values:

- Yes represented by **+h**
- No represented by **-h**

RV ***B*** represents if I am bored and assumes the following values:

- Yes represented by **+b**
- No represented by **-b**

RV ***S*** represents if I am going to stream and assumes the following values:

- Yes represented by **+s**
- No represented by **-s**

Joint Distributions

H	*B*	*S*	**P**
+h	**+b**	**+s**	0.44
+h	**+b**	**-s**	0.15
+h	**-b**	**+s**	0.14
+h	**-b**	**-s**	0.07
-h	**+b**	**+s**	0.10
-h	**+b**	**-s**	0.04
-h	**-b**	**+s**	0.05
-h	**-b**	**-s**	0.01

P(+h, -b, +s) = ?
P(+h, -b, +s) = 0.14

P(+b) = ?
P(+b) = 0.44 + 0.15 + 0.10 + 0.04
= 0.73

P((+h, -b) OR +s) = ?
P((+h, -b) OR +s) = P(+h, -b) + P(+s) - P(+h, -b, +s)
= (0.14 + 0.07) + (0.44 + 0.14 + 0.10 + 0.05) - 0.14
= 0.80

BRAC UNIVERSITY
Inspiring Excellence

Marginal Distributions

H	*B*	*S*	P
+h	+b	+s	0.44
+h	+b	-s	0.15
+h	-b	+s	0.14
+h	-b	-s	0.07
-h	+b	+s	0.10
-h	+b	-s	0.04
-h	-b	+s	0.05
-h	-b	-s	0.01

H	P(H)
+h	
-h	

B	*S*	P(H)
+b	+s	
+b	-s	
-b	+s	
-b	-s	

Marginal Distributions

H	*B*	*S*	P
+h	+b	+s	0.44
+h	+b	-s	0.15
+h	-b	+s	0.14
+h	-b	-s	0.07
-h	+b	+s	0.10
-h	+b	-s	0.04
-h	-b	+s	0.05
-h	-b	-s	0.01

H	P(*H*)
+h	0.80
-h	0.20

B	*S*	P
+b	+s	0.54
+b	-s	0.19
-b	+s	0.19
-b	-s	0.08

Marginal Distributions

H	*B*	*S*	P
+h	+b	+s	0.44
+h	+b	-s	0.15
+h	-b	+s	0.14
+h	-b	-s	0.07
-h	+b	+s	0.10
-h	+b	-s	0.04
-h	-b	+s	0.05
-h	-b	-s	0.01

H	P(*H*)
+h	0.80
-h	0.20

B	P(*B*)
+b	
-b	

S	P(*S*)
+s	
-s	

Marginal Distributions

H	*B*	*S*	P
+h	+b	+s	0.44
+h	+b	-s	0.15
+h	-b	+s	0.14
+h	-b	-s	0.07
-h	+b	+s	0.10
-h	+b	-s	0.04
-h	-b	+s	0.05
-h	-b	-s	0.01

H	P(*H*)
+h	0.80
-h	0.20

B	P(*B*)
+b	0.73
-b	

S	P(*S*)
+s	
-s	

Marginal Distributions

H	*B*	*S*	P
+h	+b	+s	0.44
+h	+b	-s	0.15
+h	-b	+s	0.14
+h	-b	-s	0.07
-h	+b	+s	0.10
-h	+b	-s	0.04
-h	-b	+s	0.05
-h	-b	-s	0.01

H	P(*H*)
+h	0.80
-h	0.20

B	P(*B*)
+b	0.73
-b	0.27

S	P(*S*)
+s	
-s	

Marginal Distributions

H	*B*	*S*	P
+h	+b	+s	0.44
+h	+b	-s	0.15
+h	-b	+s	0.14
+h	-b	-s	0.07
-h	+b	+s	0.10
-h	+b	-s	0.04
-h	-b	+s	0.05
-h	-b	-s	0.01

H	P(*H*)
+h	0.80
-h	0.20

B	P(*B*)
+b	0.73
-b	0.27

S	P(*S*)
+s	0.73
-s	

Marginal Distributions

H	*B*	*S*	P
+h	+b	+s	0.44
+h	+b	-s	0.15
+h	-b	+s	0.14
+h	-b	-s	0.07
-h	+b	+s	0.10
-h	+b	-s	0.04
-h	-b	+s	0.05
-h	-b	-s	0.01

H	P(*H*)
+h	0.80
-h	0.20

B	P(*B*)
+b	0.73
-b	0.27

S	P(*S*)
+s	0.73
-s	0.27

Conditional Probability

These are ways to calculate the probability of an event ***X*** assuming the value **x** given that another event ***Y*** already has the value **y**.

A quick example:
If ***X*** represents that you will be late to class and ***Y*** represents if there is heavy traffic, what is the probability that you will be late to class (meaning ***X*** = **+x**) if there is already heavy traffic (meaning ***Y*** = **+y**).

Symbolic representation:

$$P(\boldsymbol{X} = \mathbf{+x} \mid \boldsymbol{Y} = \mathbf{+y})$$

$$P(\mathbf{+x} \mid \mathbf{+y})$$

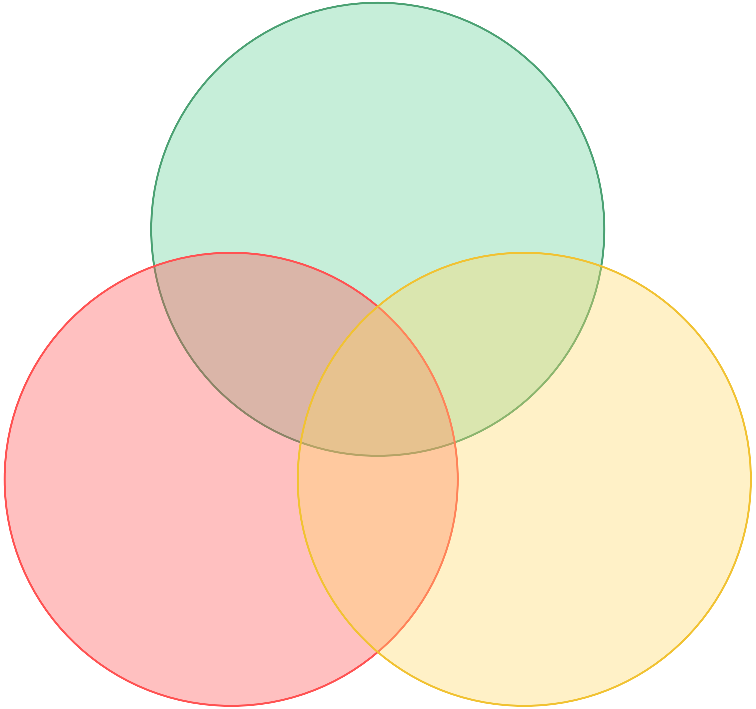
BRAC UNIVERSITY
Inspiring Excellence

Conditional Probability

H

B

S



Conditional Probability

	+b		-b	
	+s	-s	+s	-s
+h	0.44	0.15	0.14	0.07
-h	0.10	0.04	0.05	0.01

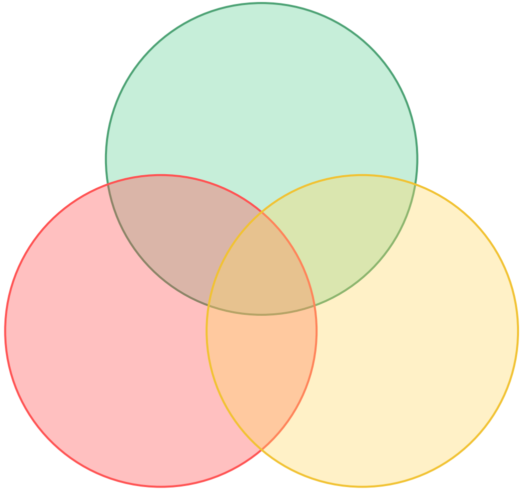
$P(+h \mid +b) = ?$

$P(+h \mid +b) = P(+h, +b) / P(+b)$

$= (0.44 + 0.15) / (0.44 + 0.15 + 0.10 + 0.04)$

$= 0.59 / 0.73$

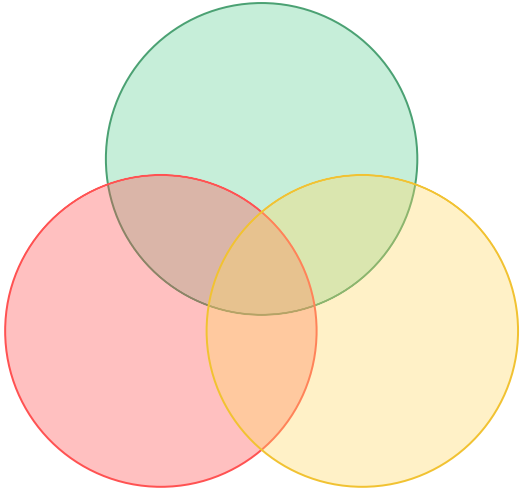
$= 0.81$



Conditional Probability

	+b		-b	
	+s	-s	+s	-s
+h	0.44	0.15	0.14	0.07
-h	0.10	0.04	0.05	0.01

$$\begin{aligned}
P(-h, -s \mid +b) &= ? \\
P(-h, -s \mid +b) &= P(-h, -s, +b) / P(+b) \\
&= 0.04 / (0.44 + 0.15 + 0.10 + 0.04) \\
&= 0.04 / 0.73 \\
&= 0.05
\end{aligned}$$

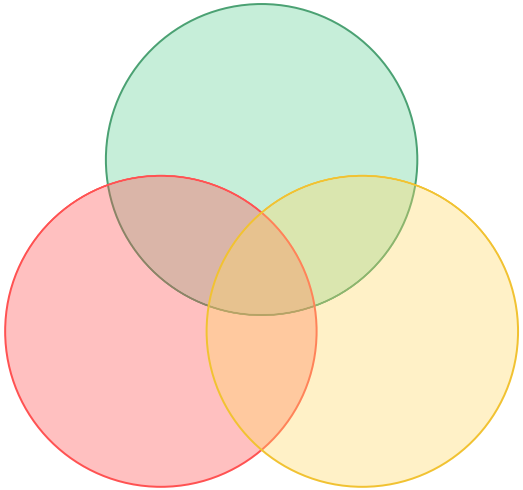


Conditional Probability

	+b		-b	
	+s	-s	+s	-s
+h	0.44	0.15	0.14	0.07
-h	0.10	0.04	0.05	0.01

P(-h | +s, -b) = ?

P(-h | +s, -b) = P(-h, +s, -b) / P(+s, -b)
= 0.05 / (0.05 + 0.14)
= 0.05 / 0.19
= 0.26



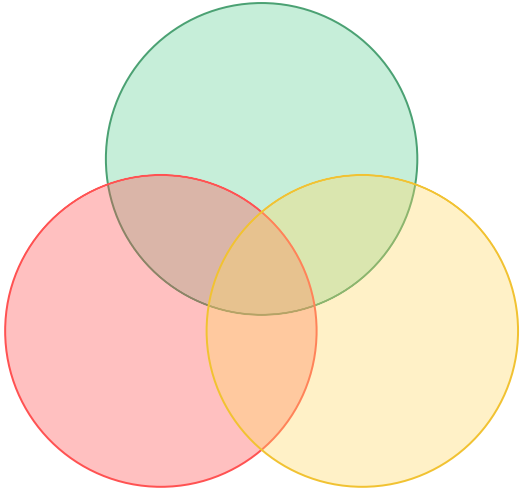
BRAC UNIVERSITY
Inspiring Excellence

Conditional Probability

	+b		-b	
	+s	-s	+s	-s
+h	0.44	0.15	0.14	0.07
-h	0.10	0.04	0.05	0.01

$$P(+h \vee -s \mid -b) = ?$$

$$\begin{aligned}
P(+h \vee -s \mid -b) &= (P(+h, -b) + P(-s, -b) - P(+h, -s, -b)) \,/\, P(-b) \\
&= ((0.14 + 0.07) + (0.07 + 0.01) - 0.07) \,/\, (0.14 + 0.07 + 0.05 + 0.01) \\
&= (0.21 + 0.08 - 0.07) \,/\, 0.27 \\
&= 0.81
\end{aligned}$$



Formulae, Rules and Principles

$0 \leq P(\boldsymbol{X} = \mathbf{x}) \leq 1$

$\sum P(\boldsymbol{X} = \mathbf{x}_i) = 1$

$P(x \mid y) = P(x, y) / P(y)$

Conditional Probability

$P(x \vee y) = P(x) + P(y) - P(x \wedge y)$

Inclusion-Exclusion Principle

$P(x, y) = P(x \mid y) * P(y)$

Product Rule

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

Absolute Independence

$P(x_1, x_2, \ldots x_n) = P(x_1) * P(x_2 \mid x_1) * P(x_3 \mid x_1, x_2) \ldots$
$= \prod P(x_i \mid x_1, \ldots x_{i-1})$

Chain Rule

$P(x, y \mid z) = P(x \mid z) * P(y \mid z)$

Conditional Independence

Absolute and Conditional Independence

Absolute Independence

When a RV ***X*** does not affect another RV ***Y*** at any time.

Example:
X => the probability of me eating right now
Y => the probability of a mosquito biting you right now

Since they are independent
P(**x**, **y**) = P(**x**) * P(**y**)

Conditional Independence

When a RV ***X*** does not affect another RV ***Y*** ONLY when ***Z*** is observed.

Example:
X => the probability that you will be late to class
Y => the probability of heavy traffic
Z => the probability of the class being cancelled

Given that the class is cancelled, ***X*** and ***Y*** becomes conditionally independent, so
P(**x**, **y** | **z**) = P(**x** | **z**) * P(**y** | **z**)