COMP307/AIML420 INTRODUCTION TO ARTIFICIAL INTELLIGENCE

Reasoning under uncertainty: Probability Basics

Outline

1. Introduction
2. Probability basics
3. Product Rule
4. Sum Rule
5. Normalisation Rule
6. Independence

Reasoning under uncertainty

- **Medical diagnosis**: Medical Doctors often need to make diagnoses based on **incomplete information** and **uncertain test** results.
- **Weather forecasting:** Meteorologists use **complex** models to provide probabilities of different weather outcomes
- **Financial decision making:** Investors often need to make decisions based on **uncertain market** conditions and future projections
  - The quants use lots of probability theory
- **Machine learning:** Most new learning algorithms are based on probability theory. Many directly output probabilities.

Uncertainty

- In most practical problems, there are **unknown** or **not precisely known** things
- Classes of uncertainty (in practice not so important)
  - **Aleatoric (AKA statistical) uncertainty:** inherent randomness or variability:
    - Quantum mechanics
    - Natural disasters
    - Geopolitical uncertainty
  - **Epistemic uncertainty:** results from a lack of knowledge:
    - Model parameters that are not precisely known
    - Finite-element computation of a tsunami

- **Uncertainty** plays a fundamental role in AI
- **Probability theory**: a mathematical framework for quantifying uncertainty

Belief about Propositions* / Events

- Instead of evaluating the truth or falsehood of a proposition, reason about the **degree of belief** that a proposition or event is true or false
- For each primitive proposition or event, attach a **degree of belief** to the sentence

- **Probability theory** provides a formal framework for manipulating these **degrees of belief**
- Example propositions / events
  - It will rain tomorrow
  - One possible prediction from a classification model ("it is a crow")

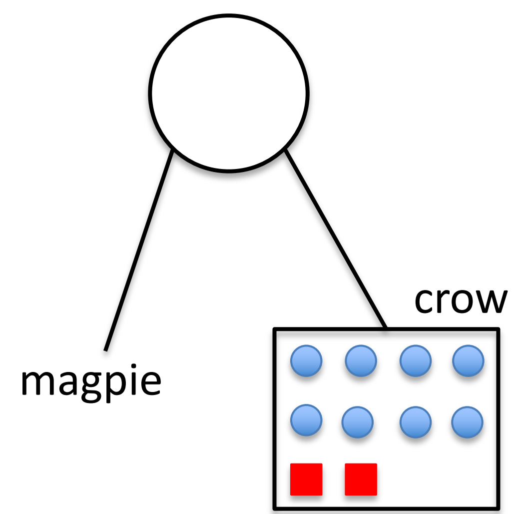
* **Proposition:** a statement that expresses a fact or a judgment that is either true or false.

Belief about Propositions* / Events

- Instead of evaluating the truth or falsehood of a proposition, reason about the **degree of belief** that a <u>proposition</u> or <u>event</u> is true or false
- To each primitive proposition or event, attach a **degree of belief**

- **Probability theory** provides a formal framework for manipulating these **degrees of belief**
- Example propositions / events
  - It will rain tomorrow
  - A prediction choice from a classification model

“it is a crow”



* **Proposition:** a statement that expresses a fact or a judgment that is either true or false.

Probability

- Given a **proposition** A, where A is either true or false (binary)
  - The probability that A is true is written as $P(A)$
  - $0 \leq P(A) \leq 1$

- **Sample space**: set of experiment outcomes
- **Event**: set of one or more experiment outcomes
- **Random variable (RV)**: formally a function mapping outcomes to numerical values
- Forms of **propositions**:
  - **Event** (happens) **-** one of the corresponding set of experiment outcomes happens
  - **random variable** = **specific numerical value**
    - Probability that proposition $X = x$ is true is written as $P(X = x)$, and sometimes as $p_X(x)$

- Example for weather:
  - **Sample space**: $\{rainy, sunny, cloudy, \ other\}$
  - **Events/Propositions**: {*rainy*}, {*rainy or cloudy*}, {*rainy or cloudy or other*}
  - **Random variable** $\boldsymbol{X}$: rainy → 1, sunny → 2, cloudy → 3, other → 4
  - **Random variable** $\boldsymbol{Y}$: rainy → 1, sunny → 2, cloudy → 1, other → 3
  - $\boldsymbol{P(sunny\ or\ cloudy) = 0.5}$: probability weather will be sunny or cloudy is 50%.

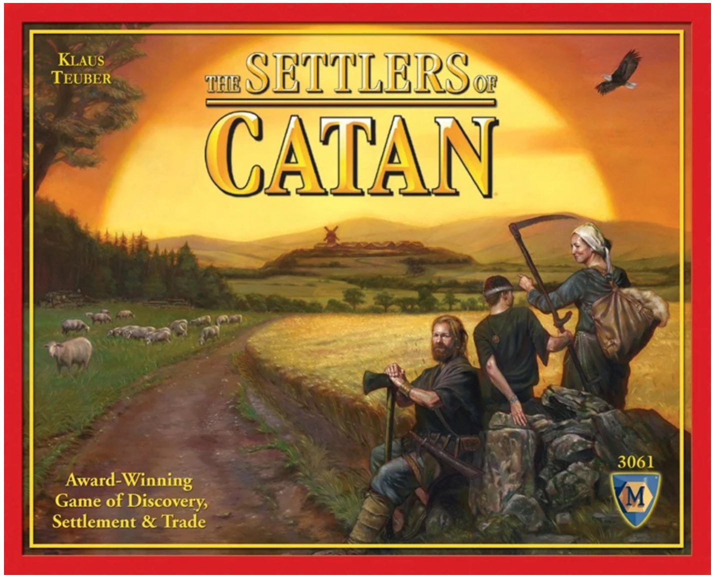
- What is the sample space of the outcome of a die?
  - Here we can use random variable = experiment outcome (RV is identity map)
  - Example proposition: random variable = 3 (the RV takes the value 3, $X = 3$)

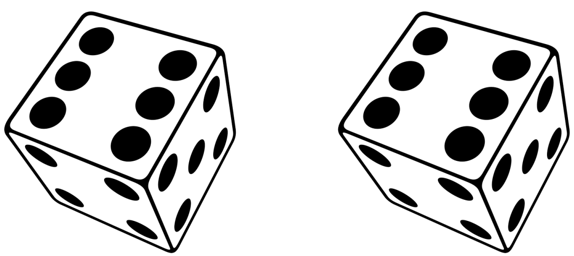
Probability

- **Notation**
  - **AND:** $\boldsymbol{A} \cap \boldsymbol{B}$. The probability that **both $\boldsymbol{A}$ and $\boldsymbol{B}$ are true**: $P(A \cap B)$
    - Sometimes written as $\boldsymbol{A} \wedge \boldsymbol{B}$
  - **OR:** $\boldsymbol{A} \cup \boldsymbol{B}$. The probability that **either $\boldsymbol{A}$ or $\boldsymbol{B}$ is true**: $P(A \cup B)$
    - Sometimes written as $\boldsymbol{A} \vee \boldsymbol{B}$
  - **NOT:** $\neg \boldsymbol{A}$. The probability that $\boldsymbol{A}$ **is false ($\neg \boldsymbol{A}$ is true)**: $P(\neg A)$

- **Axioms** of probability theory (Kolmogorov)
  - $\boldsymbol{P}(\boldsymbol{A}) \geq \mathbf{0}$
  - $\sum_{A \in \boldsymbol{\Omega}} \boldsymbol{P}(\boldsymbol{A}) = \mathbf{1}$, where $\boldsymbol{\Omega}$ is the sample space
  - $\boldsymbol{P}(\boldsymbol{A} \cup \boldsymbol{B} \cup \boldsymbol{C} \cdots) = \boldsymbol{P}(\boldsymbol{A}) + \boldsymbol{P}(\boldsymbol{B}) + \boldsymbol{P}(\boldsymbol{C}) + \cdots$, for mutually exclusive events $A, B, C, \cdots$

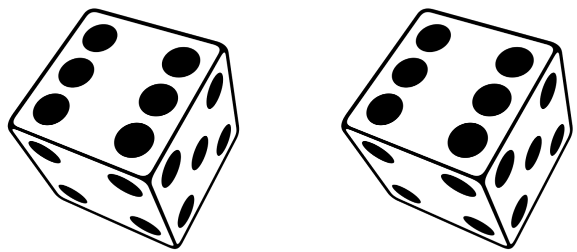
Example: Catan





Example 1

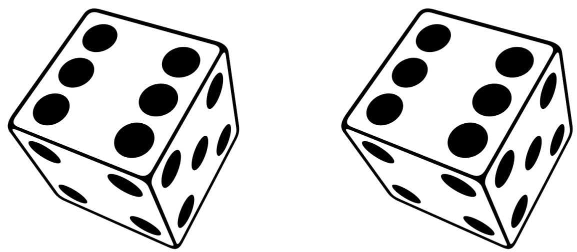
- If we roll two fair dice, what is the probability that the sum of their outcomes is 11?



Dice1	Dice2	Sum
1	1	2
1	2	3
1	3	4
1	4	5
1	5	6
1	6	7
…	…	…
6	5	11
6	6	12

Example 1

- If we roll two fair dice, what is the probability that the sum of their outcomes is 11?

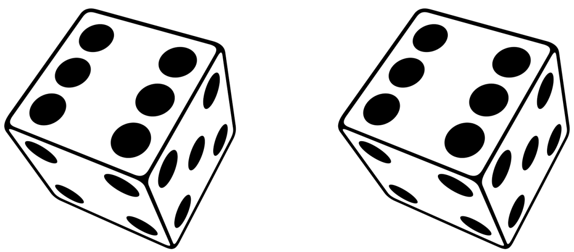


- 36 possible outcomes in total

Dice1	Dice2	Sum
1	1	2
1	2	3
1	3	4
1	4	5
1	5	6
1	6	7
…	…	…
6	5	11
6	6	12

Example 1

- If we roll two fair dice, what is the probability that the sum of their outcomes is 11?

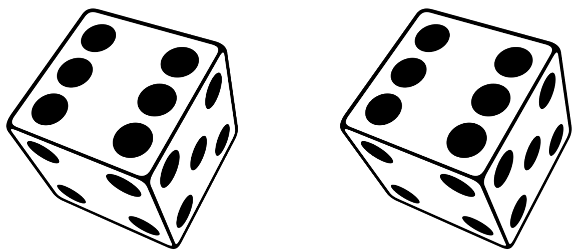


- Sample space contains 36 pairs
- 36 possible outcomes
- 2 outcomes **(5,6)** and **(6,5)** give the total number of 11

Dice1	Dice2	Sum
1	1	2
1	2	3
1	3	4
1	4	5
1	5	6
1	6	7
…	…	…
6	**5**	**11**
6	6	12

Example 1

- If we roll two fair dice, what is the probability that the sum of their outcomes is 11?



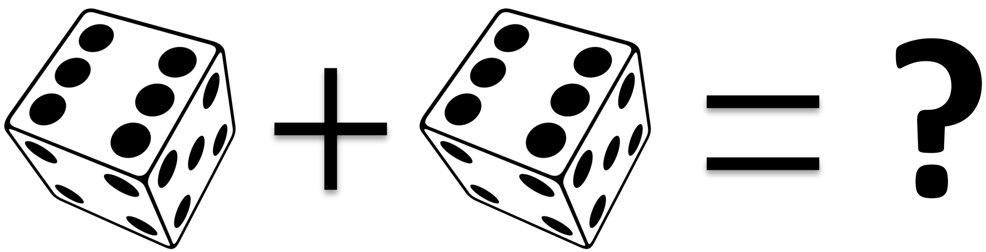
Dice1	Dice2	Sum
1	1	2
1	2	3
1	3	4
1	4	5
1	5	6
1	6	7
…	…	…
6	5	11
6	6	12

- Sample space contains 36 pairs
- 36 possible outcomes
- 2 outcomes **(5,6)** and **(6,5)** give the total number of 11
- $P(Sum = 11) =$

  $P(die\_1 = 5, die\_2 = 6 \cup die\_1 = 6, die\_2 = 5) =$
  $P(die\_1 = 5, die\_2 = 6) + P(die\_1 = 6, die\_2 = 5) =$
  $2/36 = 1/18$

Example 2

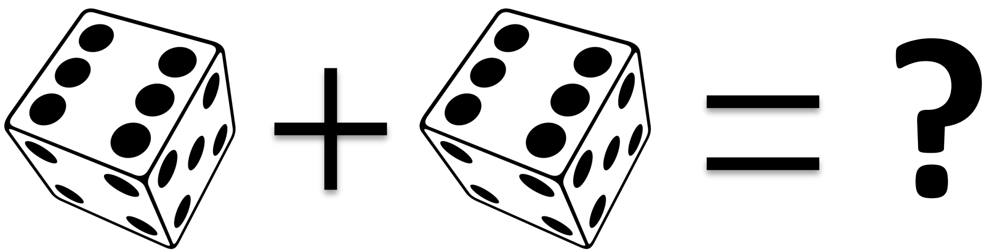
- If we roll two fair dice, what is the most probable value obtained when we sum their outcomes?



Dice1	Dice2	Sum
1	1	2
1	2	3
1	3	4
1	4	5
1	5	6
1	6	7
2	1	3
…	…	…
6	6	12

Example 2

- If we roll two fair dice, what is the most probable value obtained when we sum their outcomes?

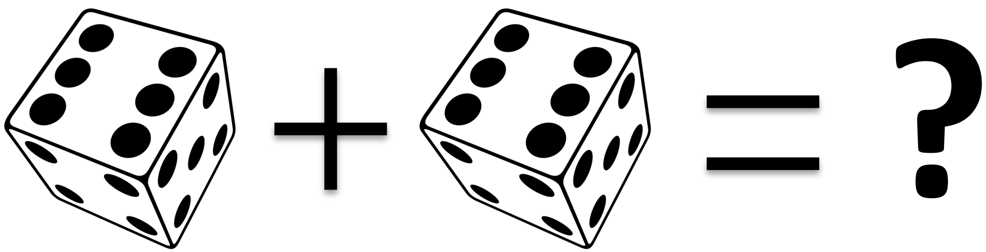


Dice1	Dice2	Sum
1	1	2
1	2	3
1	3	4
1	4	5
1	5	6
1	6	7
2	1	3
…	…	…
6	6	12

		Dice1					
		1	**2**	**3**	**4**	**5**	**6**
	1	2	3	4	5	6	7
	2	3	4	5	6	7	8
Dice2	**3**	4	5	6	7	8	9
	4	5	6	7	8	9	10
	5	6	7	8	9	10	11
	6	7	8	9	10	11	12

Example 2

- If we roll two fair dice, what is the most probable value obtained when we sum their outcomes?



Dice1	Dice2	Sum
1	1	2
1	2	3
1	3	4
1	4	5
1	5	6
1	6	7
2	1	3
…	…	…
6	6	12

		Dice1					
		1	2	3	4	5	6
Dice2	1	2	3	4	5	6	7
	2	3	4	5	6	7	8
	3	4	5	6	7	8	9
	4	5	6	7	8	9	10
	5	6	7	8	9	10	11
	6	7	8	9	10	11	12

$P(Sum = 7)$ = 6/36 = 1/6

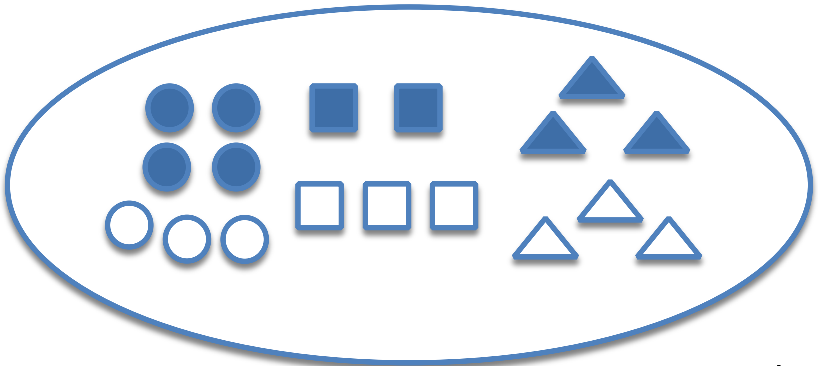
Unconditional/Conditional/Joint Probability

- Unconditional/Prior probability: degrees of belief in propositions in the absence of any other information
  - Example $P(Sum = 11)$

- Conditional/Posterior probability: degrees of belief in propositions given other information (evidence)
  - $P(A \mid B)$: the conditional probability that A is true given that B is true
  - Example $P(Sum = 11 \mid Die_1 = 6)$, the conditional probability that the total number is 11 given that the first dice outcome is 6

- Joint probability $P(A, B) ::= P(A \cap B)$: the probability that A is true and B is true

Example



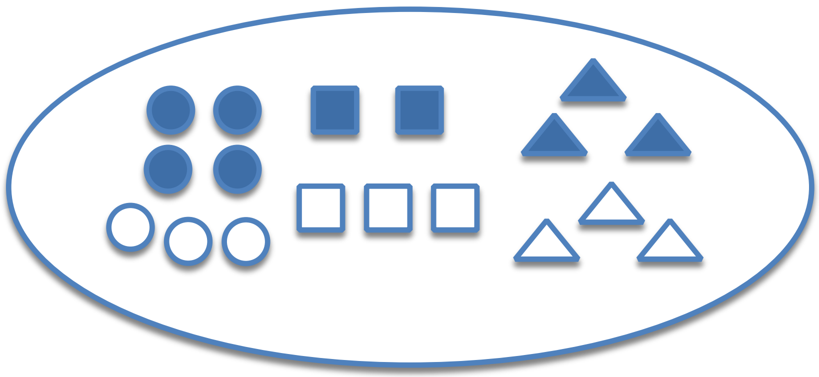
A sample space; in this example all outcomes have equal probability

Shape

Colour

	Circle	Square	Triangle
Blue			
White			

Example

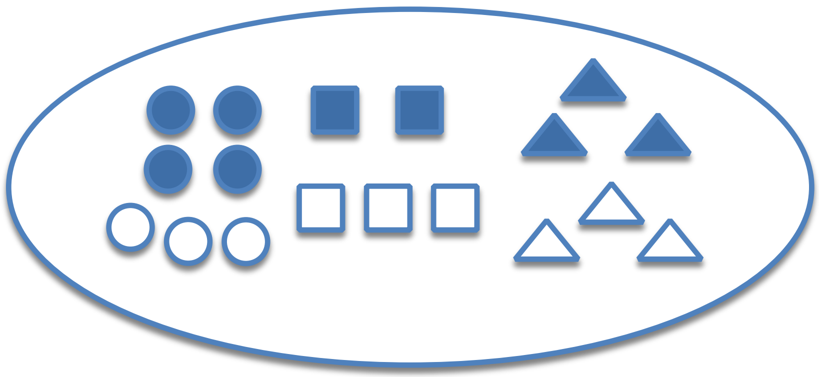


- P(Blue, Circle)

Shape

		Circle	Square	Triangle	
Colour	Blue	**4**	**2**	**3**	**9**
	White	**3**	**3**	**3**	**9**
		7	**5**	**6**	**18**

Example

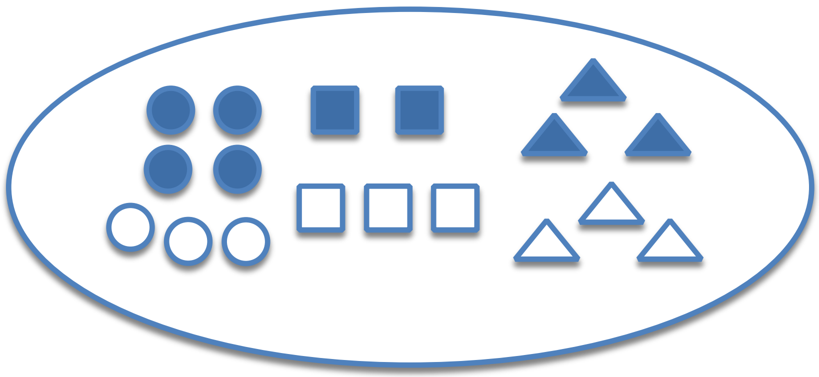


- P(Blue, Circle) = 4 / 18

Shape

Colour		Circle	Square	Triangle	
	Blue	**4**	**2**	**3**	**9**
	White	**3**	**3**	**3**	**9**
		7	**5**	**6**	**18**

Example



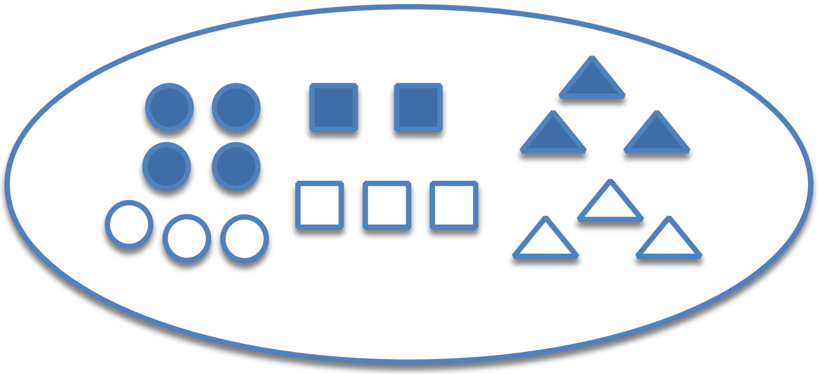
- P(Blue, Circle) = 4 / 18
- P(White, Square)

Shape

Colour

	Circle	Square	Triangle	
Blue	**4**	**2**	**3**	**9**
White	**3**	**3**	**3**	**9**
	7	**5**	**6**	**18**

Example



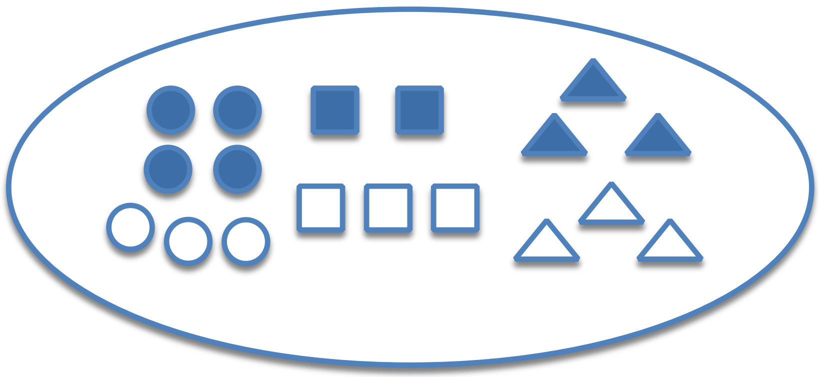
- P(Blue, Circle) = 4 / 18
- P(White, Square) = 3 / 18

Shape

Colour

	Circle	Square	Triangle	
Blue	**4**	**2**	**3**	**9**
White	**3**	**3**	**3**	**9**
	7	**5**	**6**	**18**

Example

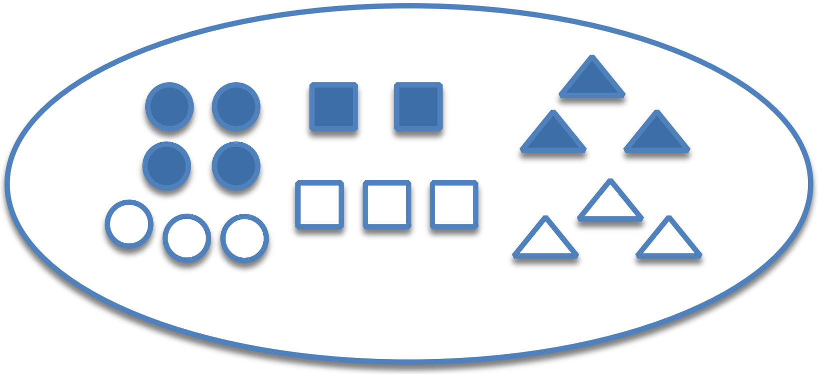


- P(Blue, Circle) = 4 / 18
- P(White, Square) = 3 / 18
- P(Circle)

Shape

		Circle	Square	Triangle	
Colour	Blue	**4**	**2**	**3**	**9**
	White	**3**	**3**	**3**	**9**
		7	**5**	**6**	**18**

Example

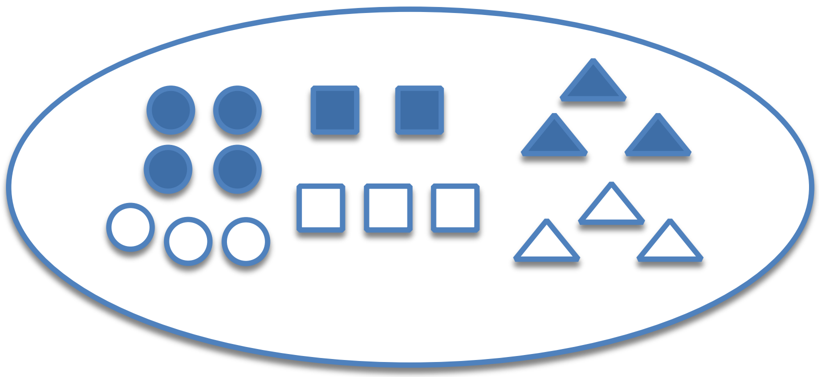


- P(Blue, Circle) = 4 / 18
- P(White, Square) = 3 / 18
- P(Circle) = 7 / 18

Shape

Colour		Circle	Square	Triangle	
	Blue	**4**	**2**	**3**	**9**
	White	**3**	**3**	**3**	**9**
		7	**5**	**6**	**18**

Example

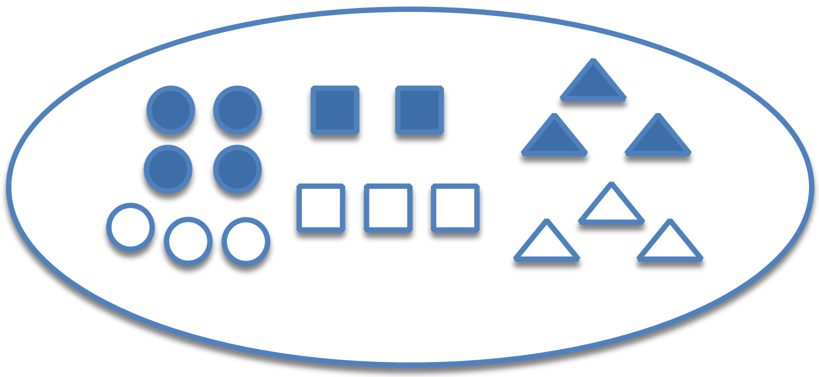


- P(Blue, Circle) = 4 / 18
- P(White, Square) = 3 / 18
- P(Circle) = 7 / 18
- P(Circle | Blue)

Shape

		Circle	Square	Triangle	
Colour	Blue	**4**	**2**	**3**	**9**
	White	**3**	**3**	**3**	**9**
		7	**5**	**6**	**18**

Example



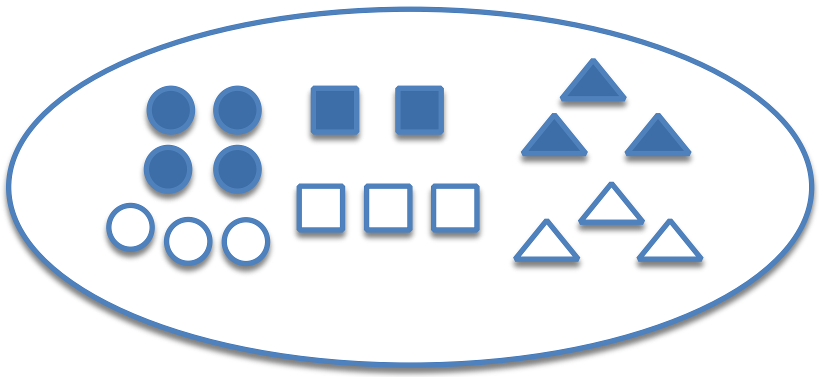
- P(Blue, Circle) = 4 / 18
- P(White, Square) = 3 / 18
- P(Circle) = 7 / 18
- P(Circle | Blue) = 4 / 9

Shape

Colour

	Circle	Square	Triangle	
Blue	**4**	**2**	**3**	**9**
White	**3**	**3**	**3**	**9**
	7	**5**	**6**	**18**

Example

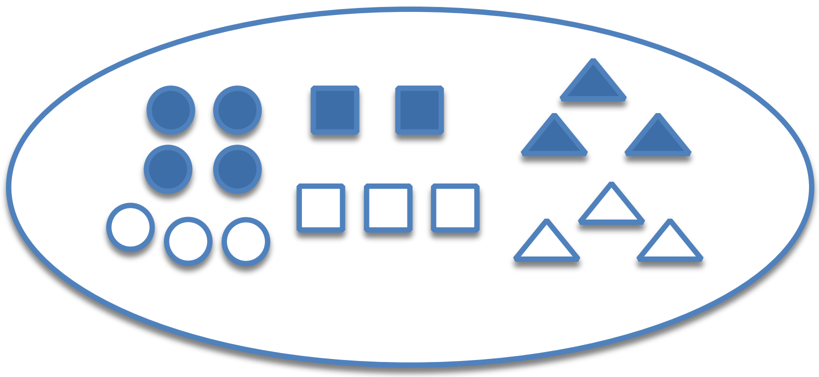


- P(Blue, Circle) = 4 / 18
- P(White, Square) = 3 / 18
- P(Circle) = 7 / 18
- P(Circle | Blue) = 4 / 9
- P(Blue | Triangle)

Shape

		Circle	Square	Triangle	
Colour	Blue	**4**	**2**	**3**	**9**
	White	**3**	**3**	**3**	**9**
		7	**5**	**6**	**18**

Example



- P(Blue, Circle) = 4 / 18
- P(White, Square) = 3 / 18
- P(Circle) = 7 / 18
- P(Circle | Blue) = 4 / 9
- P(Blue | Triangle) = 3 / 6

Shape

Colour		Circle	Square	Triangle	
	Blue	**4**	**2**	**3**	**9**
	White	**3**	**3**	**3**	**9**
		7	**5**	**6**	**18**

Product Rule

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

- Check the propositions
  - Simultaneously: P(A, B)
  - One by one: P(B) * P(A | B) or P(A) * P(B | A)

- Note that the product rule also means that

$$P(B|A) = P(A|B) * P(B)/P(A)$$

Example

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

Shape

Colour		Circle	Square	Triangle	
	Blue	**4**	**2**	**3**	**9**
	White	**3**	**3**	**3**	**9**
		7	**5**	**6**	**18**

- P(Blue, Circle) = 4 / 18

- P(Blue) = 9 / 18
- P(Circle | Blue) = 4 / 9

- P(Circle) = 7 / 18
- P(Blue | Circle) = 4 / 7

Law of Total Probability and Normalisation Rule

- **Law of total probability:** the probability that a random variable takes a certain value the sum of the joint probability of other variables over their values:

$$P(X = x) = \sum_{y \in \Omega} P(X = x, Y = y)$$

- **The normalisation rule:** the probabilities of all values a random variable can take sums to one:

$$\sum_x P(X = x) = 1$$

$$\sum_x P(X = x \mid Y = y) = 1$$

Independence

- The product rule: P(A, B) = P(B) * P(A | B) = P(A) * P(B | A)
- If A and B are independent ($A \perp B$) to each other, then
  - P(A | B) = P(A)
  - P(B | A) = P(B)
  - P(A, B) = P(A) * P(B)

- Flip coins twice, flip1 and flip2 are independent
- Weather and crop yield are dependent

Independent or Dependent?

- Rolling a die and flipping a coin?
- Flipping a coin twice?
- Picking colored balls from a bag without replacement?
- Medical diagnoses for brothers?
- A customer and the purchase of a product?

Summary

- **Uncertainty** is present in almost every worthwhile problem/decision
- **Probability theory** can be used to quantify and find relations for uncertainty
- There are many online resources, examples:
  - Khan academy
  - A lecture at the University of Chicago (first few slides discuss frequentist vs Bayesian; we use Bayesian view)