Lecture 1

- A probabilistic model is a quantitative description of an uncertain situation
- To describe the possible outcomes of an experiment we need to define a Sample space.
- Then we need to define a probability law, which tells us if one outcome is more likely than another.
- probabilities must satisfy axioms (basic rules) in order to be meaningful. E.g. probabilities cannot be negative.
- Probability axioms are few, but powerful, as they infer a lot of additional information.

Sample Space

• Two steps
1. Describe the possible outcomes
2. Describe our beliefs about the likelihood of an outcome.

Step 1

• List (set) of possible outcomes = Ω = Sample space
- Mutually exclusive (outcomes don't overlap).
- Collectively exhaustive. e.g. add up to 1
- At the right granularity. e.g. don't add irrelevant variables that barely effect the experiment.

Ω

H. .T

Ω

- H and its raining
- H and its sunny
- T and its raining
- T and its sunny

If weather is unlikely to effect coin flips, is it a relevant variable?

Sample space: discrete / finite example.

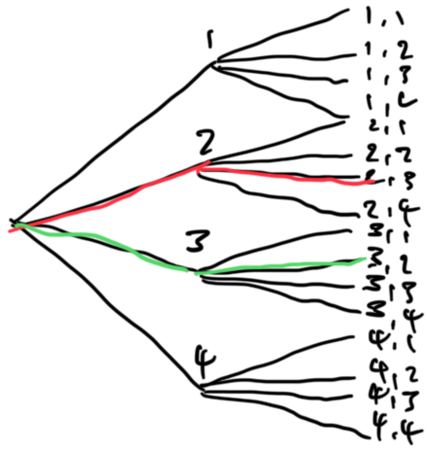
- Two rolls of a tetrahedral die

Sequential description

Y = Second roll				
4	1,4	2,4	3,4	4,4
3	1,3	2,3	3,3	4,3
2	1,2	2,2	3,2	4,2
1	1,1	2,1	3,1	4,1
	1	2	3	4

X = first roll

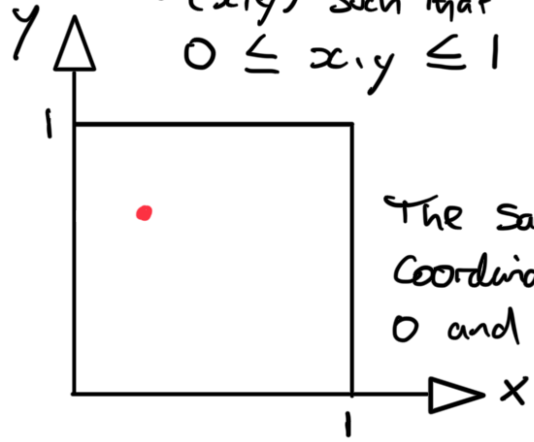
$4 \times 4 = 16$ outcome



Roll 2,3
Roll 3,2

Sample space: Continous example

- (x,y) Such that
$0 \leq x,y \leq 1$



The sample space is the set of coordinates (x,y pairs) that lie between 0 and 1.

Probability axioms

- A a single coordinate in a continuous sample space has a probability of 0.

- Therefore probabilities can only be calculated for a set of coordinates, within the sample space.

- These subsets are called an Event.

$P(A)$

A did occur.
A did not occur.

Axioms

- Nonnegativity: $P(A) \geq 0$
- Normalisation: $P(\Omega) = 1$
- Finite additivity

$\cap$ = intersection
$\cup$ = union

$A \cap B$
$A \cup B$

If $A \cap B = \emptyset$, then
$P(A \cup B) = P(A) + P(B)$

$\emptyset$ = Empty set.

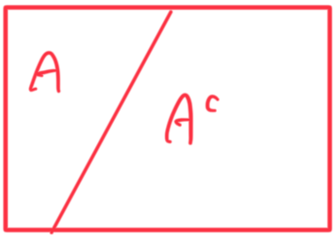
- If the intersection of A and B is empty. Then you can add their probabilities together. (disjoint)

Consequences of the axioms.

Axiom	Consequence
a) $P(A) \geq 0$	$P(A) \leq 1$
b) $P(\Omega) = 1$	$P(\emptyset) = 0$

c) $P(A) + P(A^c) = 1$

- The probability that an event happened + the probability that event did not happen = 1.



$A \cup A^c = \Omega$
$A \cap A^c = \emptyset$

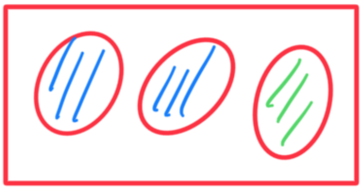
$1 \overset{(b)}{=} P(\Omega) = P(A \cup A^c)$

$$1 = P(\Omega) = P(A \cap A^c)$$
$$\overset{(c)}{=} P(A) + P(A^c)$$

$$P(A) = 1 - P(A^c) \overset{(a)}{\leq} 1$$

$$1 = P(\Omega) + P(\Omega^c)$$
$$1 = 1 + P(\emptyset) \Rightarrow P(\emptyset) = 0$$

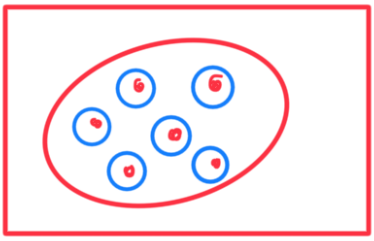
- A, B, C disjoint. $P(A \cup B \cup C) = P(A) + P(B) + P(C)$



$$P(A \cup B \cup C) = P((A \cup B) \cup C)$$
$$= P(A \cup B) + P(C)$$
$$= P(A) + P(B) + P(C)$$

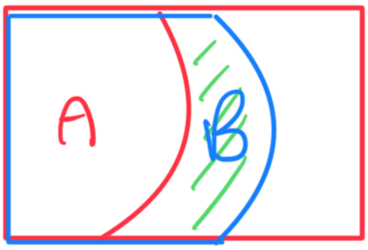
If $A_1, \dots, A_k$ disjoint $\Rightarrow P(A_1 \cup \dots \cup A_k) = \sum_{i=1}^{k} P(A_i)$

$$P(\{s_1, s_2, \dots, s_k\}) = (\{s_1\} \cup \{s_2\} \cup \dots \cup \{s_k\})$$



$$= P(\{s_1\}) + \dots + P(\{s_k\})$$
$$= P(s_1) + \dots + P(s_k)$$

- If $A \subset B$, then $P(A) \leq P(B)$



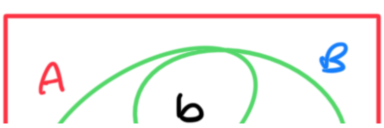
$\subset$ = smaller

$A \subset B$ = A smaller than B

$$B = A \cup (B \cap A^c)$$

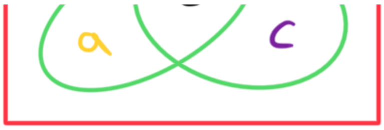
$$P(B) = P(A) + P(B \cap A^c)$$
$$\geq P(A)$$

- $P(A \cup B) = P(A) + P(B) - P(A \cap B)$ ≥ 0



$a = P(A \cap B^c)$

$b = P(A \cap B)$



$c = P(B \cap A^c)$

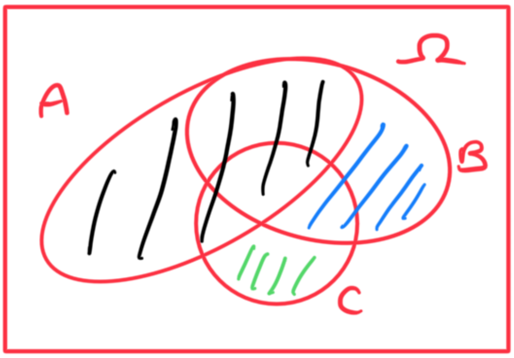
$P(A \cup B) = a + b + c$

Union Bound

$P(A \cup B) \leq P(A) + P(B)$

$$P(A+B) - P(A \cap B)$$
$$= (a+b) + (\cancel{b}+c) - \cancel{b}$$
$$= (a+b) + c$$

- $P(A \cup B \cup C) = \boxed{P(A) + P(A^c \cap B) + P(A^c \cap B^c \cap C)}$



$A \cup B \cup C$
$= A \cup (B \cap A^c)$
$\cup (C \cap A^c \cap B^c)$

- Collecting terms in this way makes each set disjoint meaning we can sum them.

Y = second roll

Y \ X	1	2	3	4
4	1,4	2,4	3,4	4,4
3	1,3	2,3	3,3	4,3
2	1,2	2,2	3,2	4,2
1	1,1	2,1	3,1	4,1

X = first roll

$$P(X=1) = 4 \cdot \frac{1}{16} = \frac{1}{4}$$

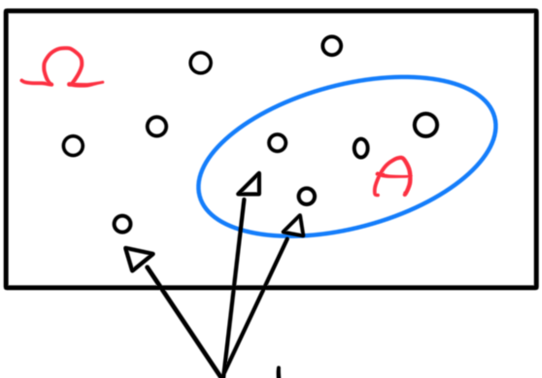
Let $z = \min(x, y)$

$$P(Z=4) = \frac{1}{16}$$

$$P(Z=2) = 5 \cdot \frac{1}{16} = \frac{5}{16}$$

Discrete Uniform Law

- Assume Ω consists of n equally likely elements
- Assume A consists of k elements



$\text{Prob} = \frac{1}{n}$ because of

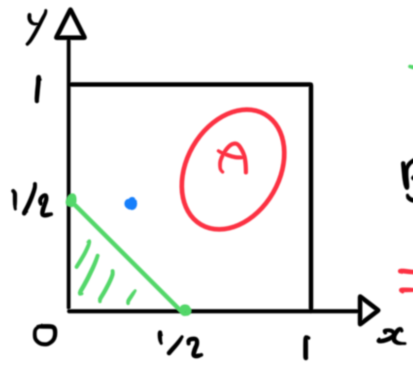
$P(A) = k \cdot \frac{1}{n}$

the axiom of normalisation.

E.g. $\underbrace{\frac{1}{2} + \frac{1}{4} + \frac{1}{8} + \frac{1}{8}}_{k} = 1.$

Continuous example

- (x, y) such that $0 \leq x, y \leq 1$
- Uniform probability law: Probability = Area



$P(\{(x,y) \mid x+y \leq 1/2\}) =$

Base × Height / 2

$= (\frac{1}{2} \cdot \frac{1}{2}) / 2 = \frac{1}{2} \cdot \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{8}$

$P(\{(0.5, 0.3)\}) = 0$ (single point)

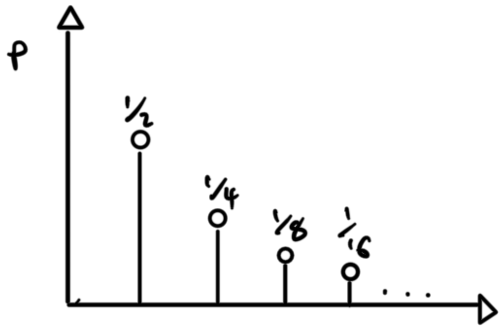
Probability Calculation Steps

- Specify the sample space
- Specify the probability law
- Identify an event of interest
- Calculate ...

— Describe the steps in words
— Then mathematically
— Then if possible, with a picture

Often probability laws are implicit / indirect. In this case we have to do extra work to find the probability of the event we care about.

Discrete but infinite sample space



- Sample space $\{1, 2, \ldots\}$

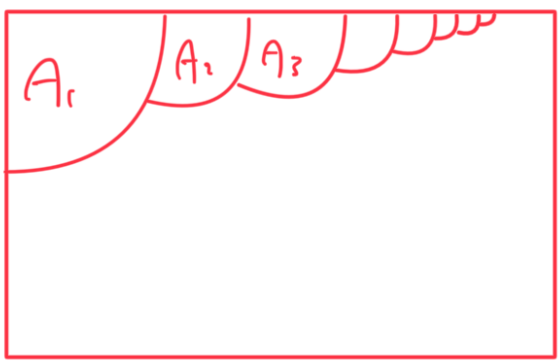
— We are given $P(n) = \frac{1}{2^n}$
$n = 1, 2, \ldots$

$$\sum_{n=1}^{\infty} \frac{1}{2^n} = \frac{1}{2} \sum_{n=0}^{\infty} \frac{1}{2^n} = \boxed{\frac{1}{2} \cdot \frac{1}{1 - (1/2)} = 1}$$

- We factor out $\frac{1}{2}$ from every term which reduces the exponent from n to -1. Which is the same as starting the sum from $n = 0$.

— This turns the sum into a usual geometric series, which there is a standard formula for.

Countable Additivity Axiom



— The additivity axiom only applies to a sequence of disjoint events. As a single continuous point is 0.

- It only works for countable sequences.
* The unit square / real line is not countable.
- Their elements cannot be arranged in a countable sequence.

○ Proofs can be found in "Measure Theory".

Interpretation of probability theory

- Probability can be viewed as frequencies. OR a description of beliefs.
- It gives us a framework for analysing situations with uncertain outcomes, and inturn making predictions.

- Predictions are only good if the model is good. So we use data to try and create good models.