Lecture 2: Basic Probability Theory

Yunwen Lei

School of Computer Science, University of Birmingham

1 Sample Space and Events

1.1 Random process and random experiments

There are two types of process: deterministic process and random process.

For the deterministic process, the outcome can be predicted exactly in advance. That is, when we know the initialization and the rule governing the process, we can know the state of the process at any time. An example is the Newton's *law of motion*

$$\text{force} = \text{mass} \times \text{acceleration}.$$

If we are given values for mass and acceleration, we exactly know the value of **force**. Roughly speaking, the deterministic process is under our control.

For the random processes, the outcome is not known exactly, but we can still describe the *probability distribution* of possible outcomes. Here is an example: suppose we toss 10 coin tosses. We do not know exactly how many heads we will get, but we can calculate the probability of getting a certain number of heads. We know it is possible that we get 5 heads and 5 tails. But it is very unlikely that we get 10 heads. In this example, there is randomness in the experiment which is not under our control. But we can still use the language of probability to do some inference.

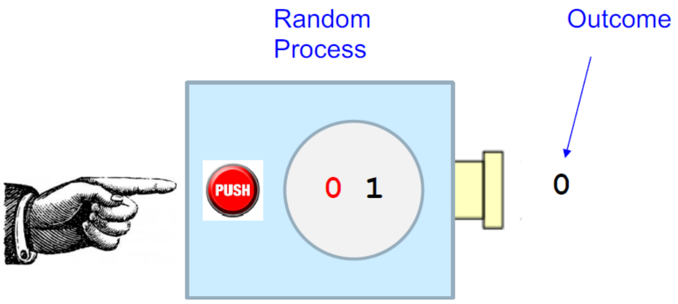
Now we give the definition of random experiments.

Definition 1 (Random Experiments)**.** A Random Experiment is a process that produces uncertain outcomes from a well-defined set of possible outcomes.

According to the definition, we know there are some requirements:

(1) well-defined set of outcomes: we know the *possible* outcome of a random experiment, e.g. the coin must be either a hear or a tail.

(2) uncertain outcomes: even we know the possible value of the outcome, we are not sure the outcome of a single experiment, e.g., it can be either a head or a tail. We know this until we have finished the experiment. The outcome can be different for repetitions of the same experiment. For example, it is possible that the first toss leads to a head while the second toss leads to a tail.



Here are examples of random experiments:

(1) we randomly toss a coin: the outcome must be either a head or a tail

(2) we randomly throw a die: the outcome must be a number in $\{1, 2, \ldots, 6\}$

(3) we randomly draw a card from deck: the outcome can take 52 values

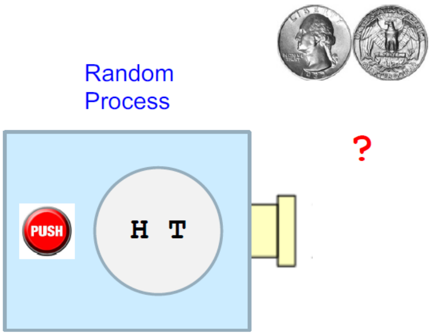
(4) we randomly draw a marble from a bag: the color can be either red, black or white.

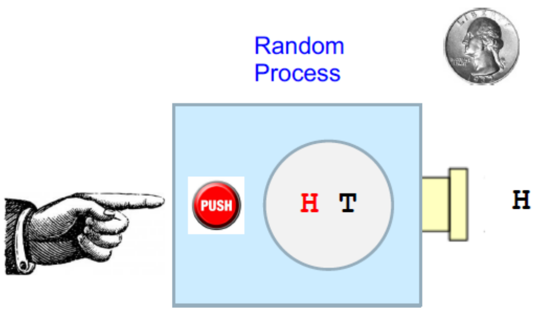
1.2 Sample Space

A very fundamental concept in probability is *sample space*, which indicates all the possible outcomes of an experiment.

Definition 2 (Sample Space)**.** The set of all possible outcomes of an experiment is called the Sample Space and is denoted by Ω. Any individual outcome is called a Sample Point.

Example 1. Toss a coin, is it heads or tails?





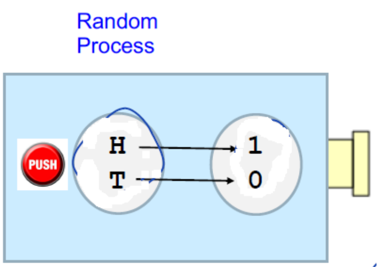
Before Experiment

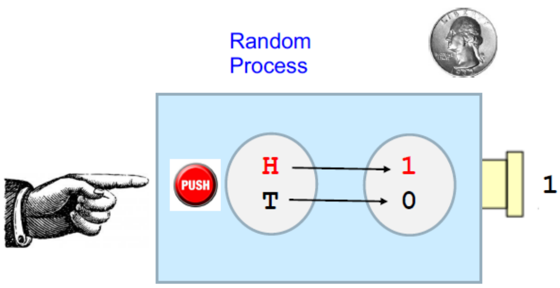
After Experiment

In this example, the sample space is $\Omega = \{H, T\}$.

The description of the random experiment may indicate that there is more than one stage to creating the outcome, for example, a physical analogy may produce results that have to be interpreted as "outcomes"

Example 2. Toss a coin, count how many heads show.





Before Experiment

After Experiment

In this example, the experiment outputs the number 1 if observing a head and the number 0 otherwise. Then the sample space becomes $\Omega = \{0, 1\}$.

1.3 Event

Another fundamental concept is the event. You can imagine an event as a set of outcomes. In probability, the outcome is an atom which can not be further divided. We can combine several outcomes to form an event.

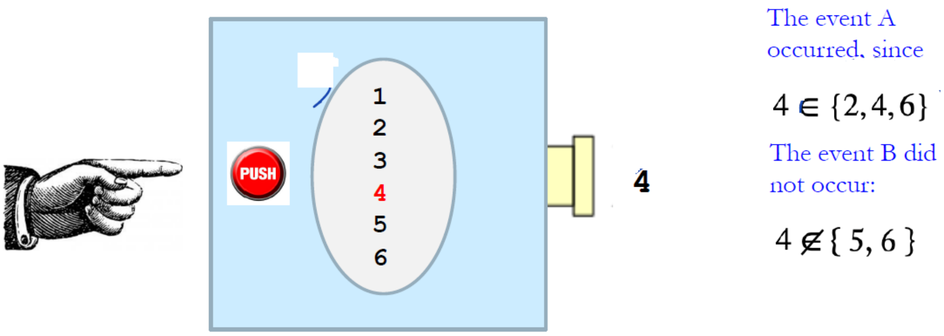
Definition 3 (Event)**.** An Event is any subset of the Sample Space. An event A is said to have occurred if the outcome of the random experiment is a member of A.

We say an event happens if the real outcome (also called realization) of the experiment belongs to this event. We can define many events for an experiment.

Here are some examples. We use $|A|$ to denote the cardinality of A.

Example 3. Toss a die and output the number of dots showing. Let A = "there are an even number of dots showing" and B = "there are at least 5 dots showing."

$$\Omega = \{1, 2, \ldots, 6\}, \quad A = \{2, 4, 6\}, \quad B = \{5, 6\}.$$



Example 4. Tossing two coins ($|S|$ denotes the cardinality of a set S)

- Sample space: $\Omega = \{(H, H), (H, T), (T, H), (T, T)\}, |\Omega| = 4$
- An event: $A = \big\{\text{the two coins come up different}\big\} = \{(H, T), (T, H)\}$

In the following example, horse A, horse B and horse C are taking a race and we are interested in the order. We use (A, B, C) to mean that horse A is ranked first, horse B is the second and horse C is the last.

Example 5. Finishing order of a race of 3 horses

- Sample space: $\Omega = \big\{(A, B, C), (A, C, B), (B, A, C), (B, C, A), (C, A, B), (C, B, A)\big\}$
- An event: $\{\text{horse } B \text{ wins}\} = \{(B, A, C), (B, C, A)\}$.

In the following example, we repeatedly roll a die until we first see 6. Then, the outcome of the experiment can be any sequence of numbers between 1 and 5, and the last number is 6. We can not see 6 before the last number since otherwise we would have already stopped the experiment. Note the sample space is infinite but still countable. You can refer to `https://en.wikipedia.org/wiki/Countable_set` for the definition of countable set.

Example 6. Repeatedly rolling a die until we first see 6

- Sample space: $\Omega = \{\text{sequences of numbers between 1 and 5, and then a 6}\}$
- An event: $(|A| = 5)$

$$A = \big\{\text{roll 4 first, get 6 on the third roll}\big\} = \big\{(4, 1, 6), (4, 2, 6), (4, 3, 6), (4, 4, 6), (4, 5, 6)\big\}.$$

Example 7. Rolling two dice

- Sample space: the set of all possible outcomes (36 sample points)
$$\Omega = \Big\{(1,1),(1,2),(1,3),(1,4),\dots,(6,3),(6,4),(6,5),(6,6)\Big\}$$
- An event of "double" (both numbers are the same)
$$A = \big\{(1,1),(2,2),(3,3),(4,4),(5,5),(6,6)\big\}, \quad |A| = 6$$
- Another event of sum being 4
$$A = \{(1,3),(2,2),(3,1)\}, \quad |A| = 3$$

In the following example, we are interested in the lifetime of a device, which can take any nonnegative number. This is a continuous experiment in the sense that the possible outcomes are uncountably infinite.

Example 8. Lifetime of a device (measured in years)

- Sample space: $\Omega = [0, \infty)$
- An event $A = \big\{\text{device lasts for at least 5 years}\big\} = [5, \infty)$
- Another event $A = \big\{\text{device is dead by its } 6^{\text{th}} \text{ birthday}\big\} = [0, 6)$

The Sample Points can be just about anything (numbers, letters, words, people, etc.) and the Sample Space (= set of all sample points) can be

- Finite

  Example: Toss a coin $\Omega = \{\text{head, tail}\}$
- Countably Infinite

  Example: Toss a coin until heads appears, and report the number of tosses $\Omega = \{1, 2, 3, 4, ...\}$
- Uncountably Infinite

  Example: Lifetime of a device $\Omega = [0, \infty)$

1.4 Event Space

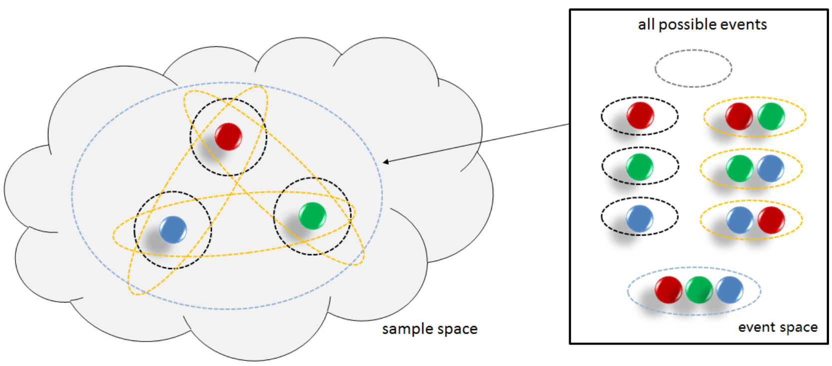
Each element in an event space is an event.

Definition 4 (Event Space)**.** The collection of all events is called the Event Space, denoted as $\mathcal{F}$.

Question: If you have n elements in the sample space, how many elements are there in the event space?

Solution: each element can either be in or not in the event, and therefore we have
$$2 \times 2 \times 2 \cdots \times 2 = 2^n$$

To help you understand we consider the following example. In this example, the sample space has 3 outcomes. Each one can be either in or not in an event. Therefore, the cardinality of the event space is 8. The right-hand side lists all the possible events in the event space.

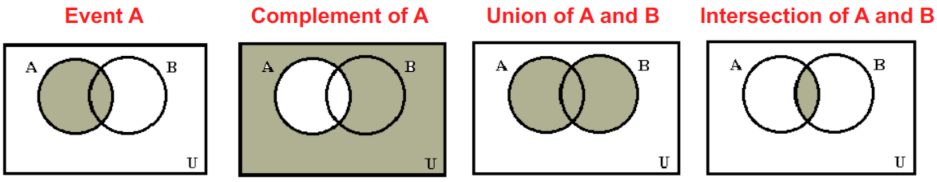


1.5 Operation on Events

Recall that events are essentially sets. Therefore the operations on sets also apply to events. Three common operations are: complement, union and intersection.

- The complement A^c of an event A is the event that A does not occur
- The union $A \cup B$ of two events A and B is the event that either A or B or both occurs
- The intersection $A \cap B$ of two events A and B is the event that both A and B occur

As you see, these definitions coincide with those for the sets.



The above figure gives an explanation. Below is an example to show how to implement the operations.

Example 9 (Toss a die and output the number of dots showing)**.** Let A = "there are an even number of dots showing" and B = "there are at least 5 dots showing."

$$\Omega = \{1, 2, \ldots, 6\}, \quad A = \{2, 4, 6\}, \quad B = \{5, 6\}.$$

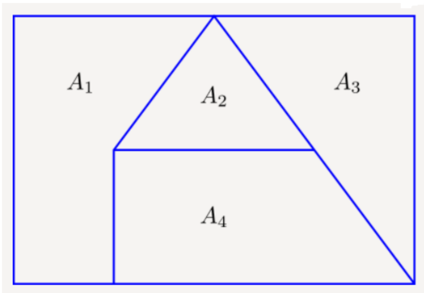
It is then clear

$$A \cup B = \{2, 4, 5, 6\} \qquad A \cap B = \{6\}$$
$$A^c = \{1, 3, 5\} \qquad A^c \cap B = \{5\}$$

In the forthcoming study, we will use the definition of a partition of a sample space. Note there are many ways to partition a sample space.

Definition 5 (Partition)**.** A collection of sets $\{A_1, \ldots, A_n\}$ is a partition to the universal set Ω if it satisfies the following conditions:

- (non-overlap) $\{A_1, \ldots, A_n\}$ is disjoint
- (decompose) $A_1 \cup A_2 \cup \ldots \cup A_n = \Omega$



Example 10. If $\Omega = \{1, 2, 3, 4, 5, 6\}$. Then

- $A_1 = \{1, 3, 4\}, A_2 = \{2, 5\}, A_3 = \{6, 7\}$ is not a partition of Ω
- $A_1 = \{1, 3, 4\}, A_2 = \{2, 5\}, A_3 = \{1, 6\}$ is not a partition of Ω
- $A_1 = \{1, 3, 4\}, A_2 = \{2, 5\}, A_3 = \{6\}$ is a partition of Ω

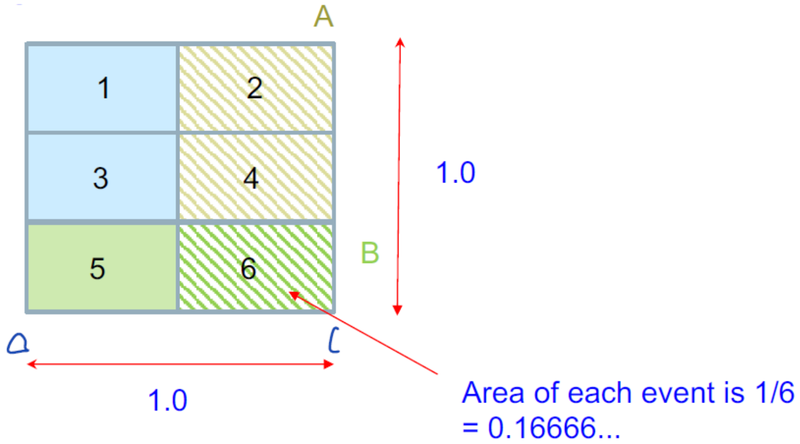
2 Probability

To give a formal definition of probability, it would be helpful to give an intuitive understanding of probability.

- It would be intuitive to use area as a proxy for probability.

- "The area of Ω" $= 1$ and "the area of an event" $=$ "probability of that event"

Example 11. Toss a die and output the number of dots showing. Let A = "there are an even number of dots showing" and B = "there are at least 5 dots showing."



In this example, the sample space is partitioned into 6 blocks with each having an area of $1/6$. We can assign the probability of $1/6$ to each block. The event A has three blocks and has an area of $3/6$, while the event B has two blocks and has an area of $2/6$. We can intuitively think the event A happens with probability $1/2$, while the event B happens with probability $1/3$.

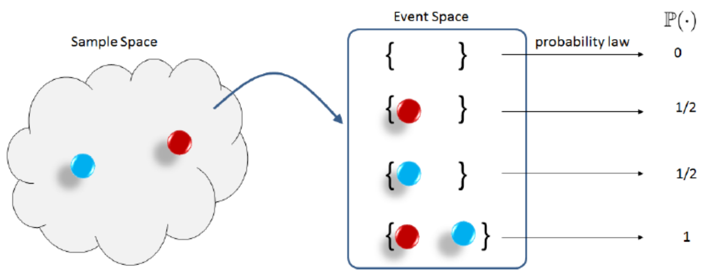
2.1 Definition of Probability

Now we provide the definition of probability. The rigorous theoretical foundation of probability was established by the great mathematician Andrey Kolmogorov at the age of 30. Kolmogorov is widely recognized as one of the greatest mathematician in the 20th century [1].

Definition 6 (Probability Law)**.** A Probability Law is a function $\mathbb{P} : \mathcal{F} \mapsto [0, 1]$ that maps an event A to a real number in $[0, 1]$. It satisfies the following Kolmogorov axioms:

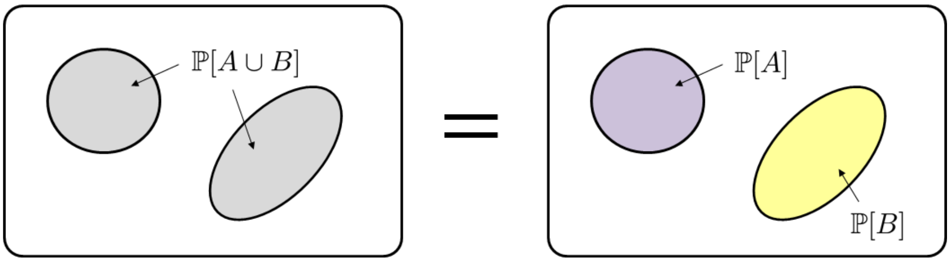
- Non-negativity: for any event $A \in \mathcal{F}$, $\mathbb{P}(A) \geq 0$ (nonnegative area of event)
- Unit measure: $\mathbb{P}(\Omega) = 1$ (the area of the whole sample space is 1)
- Additivity of disjoint events: if $A_1, A_2, \ldots$ is a collection of disjoint events then (if two regions do not overlap, then the area of the combined region is the sum of the area of each region)

$$\mathbb{P}\big(\cup_{i=1}^{\infty} A_i \big) = \sum_{i=1}^{\infty} \mathbb{P}(A_i)$$



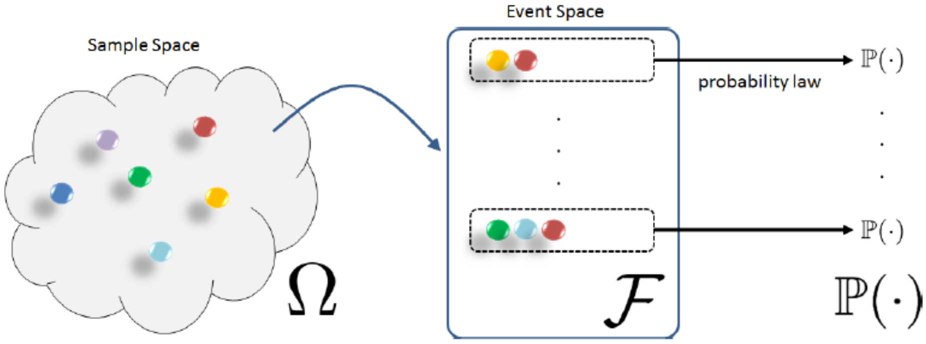
Explanations: Unit measure means the outcome of the experiment must belong to the sample space, which is consistent with the definition of sample space. Additivity of disjoint events is natural: if we are given several disjoint set, then the area of the union should be the summation of the area of each set. This property tells us how to compute the probability of an event once we know the probability of each outcome (note outcomes are atoms in probability).

[1] https://valeman.medium.com/andrey-kolmogorov-one-of-the-greatest-mathematicians-of-the-xxst-century-4167ad02d10

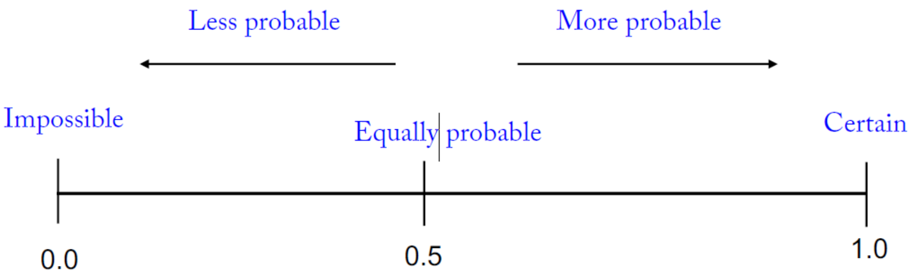


2.2 Probability Space

A probability space consists of a triplet $(\Omega, \mathcal{F}, \mathbb{P})$, i.e., sample space, event space and probability law. The sample space indicates all the possible outcomes, the sample space contains the events we are interested in, and the probability law associates a likelihood value to each event. We need to give these three components in defining a probability space.



So we measure the probability of events on a real-number scale from 0 to 1.

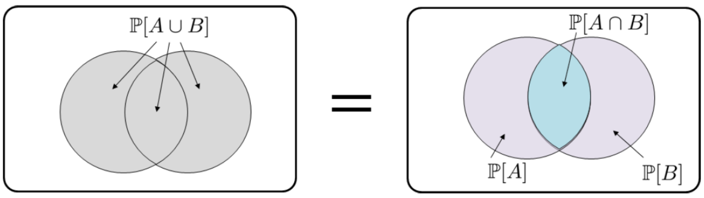


2.3 Properties of Probability Laws

Proposition 1. *Properties of Probability Laws Consider a probability law, and let A and B be events*

(a) If $A \subset B$ then $\mathbb{P}(A) \leq \mathbb{P}(B)$

(b) $\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B)$



(c) $\mathbb{P}(A^c) = 1 - \mathbb{P}(A)$.

Here are some explanations. Part (a) means that a larger event happens with bigger probability. Part (b) tells us how to compute probability for an union of two events. It recovers the additivity of probability law if these two events are disjoint. Part (c) shows the relationship between the probability of an event and that of its complement. We leave the proof as exercises.

2.4 Equally Likely Outcomes

Some sample spaces have equally likely outcomes, meaning that each outcome happens with the same probability. Here are some examples

- Coin toss: $\Omega = \{\text{Head,Tail}\}$
- Tossing two coins: $\Omega = \{(H,H),(H,T),(T,H),(T,T)\}$
- Roll of 6-sided die: $\Omega = \{1,2,3,4,5,6\}$

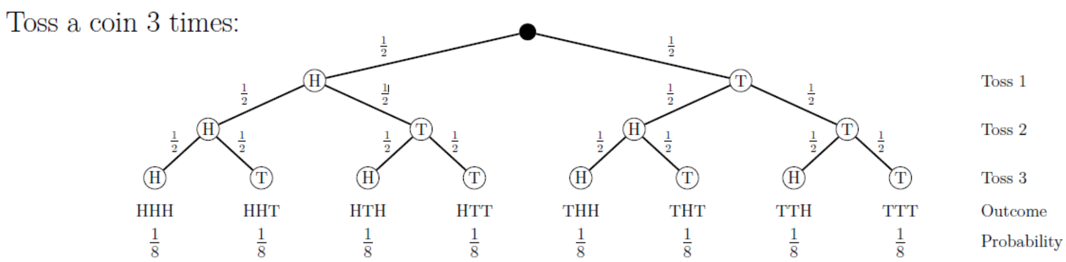
If we have equally likely outcomes, then $\mathbb{P}(\{\text{each outcome}\}) = 1/|\Omega|$, i.e., each outcome happens with the same probability.

By additivity of disjoint events, we know

$$\mathbb{P}(A) = \frac{\#\text{ outcomes in } A}{\#\text{ outcomes in } \Omega} = \frac{|A|}{|\Omega|}$$

That is, the probability of an event is proportional to its cardinality.

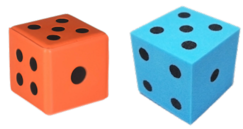
Example 12 (Toss Three Coins)**.** Suppose we toss a coin 3 times. We can show this process by a tree. In the beginning, we are at the root node. Then we can go to either the left child or the right child, depending on the realization of the first coin. Suppose the first coin is a "H", then we go to the left child. We further go down along the tree depending on the realization of the second coin. When finishing the experiment, we arrive at a left node, each of which is associated with a three character sequence. This is an experiment with equally likely outcomes.



We can compute the probability of "2 heads" as follows

$$\mathbb{P}(\text{"2 heads"}) = \frac{\text{number of sequences with 2 heads}}{|\Omega|} = \frac{3}{8}$$

Example 13 (Roll Two Dice)**.** Roll two 6-sided fair dice. What is $\mathbb{P}(\text{sum} = 7)$?



The sample space consists of all pairs of numbers in $\{1,2,3,4,5,6\}$.

$$
\Omega = \left\{
\begin{array}{cccccc}
(1,1) & (1,2) & (1,3) & (1,4) & (1,5) & \mathbf{(1,6)} \\
(2,1) & (2,2) & (2,3) & (2,4) & \mathbf{(2,5)} & (2,6) \\
(3,1) & (3,2) & (3,3) & \mathbf{(3,4)} & (3,5) & (3,6) \\
(4,1) & (4,2) & \mathbf{(4,3)} & (4,4) & (4,5) & (4,6) \\
(5,1) & \mathbf{(5,2)} & (5,3) & (5,4) & (5,5) & (5,6) \\
\mathbf{(6,1)} & (6,2) & (6,3) & (6,4) & (6,5) & (6,6)
\end{array}
\right\}
$$

This is an experiment with equally likely outcome. The probability of an event is equal to the cardinality of that event divided by the cardinality of the sample space.

$$\mathbb{P}(A) = \frac{|A|}{|\Omega|} = \frac{6}{36} = \frac{1}{6}$$