Probability
a gentle introduction

A simple bet

versus

Yannis	Rajesh
Confident	Not as confident
\$2.00	\$1.00

Alternative representation: The predictive probabilities are:

Yankees win: 2/3 = 0.666….

Red sox win: 1/3 = 0.333….

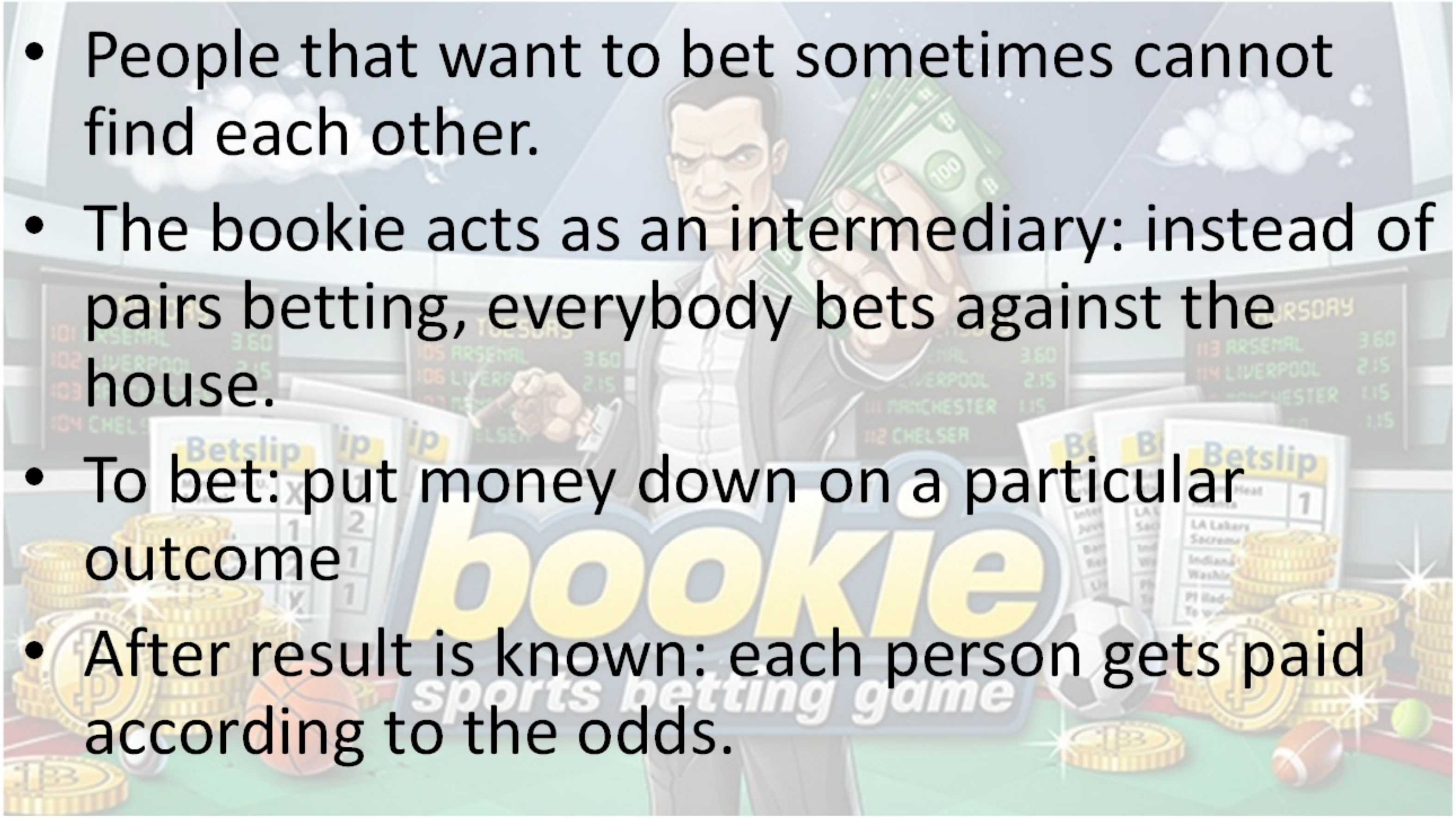
Probabilities and Odds

- Why do bets of $1 vs $2 corresponds to probabilities of 1/3 and 2/3?
- Because the expected amount of money Rajesh will pay Yannis is

$$\$1*(2/3) - \$2*(1/3) = 0$$

- Terms:
  - Probabilities
  - Fair betting odds
  - Actual amount
  - Expected amount

Bets against the house.



- People that want to bet sometimes cannot find each other.
- The bookie acts as an intermediary: instead of pairs betting, everybody bets against the house.
- To bet: put money down on a particular outcome
- After result is known: each person gets paid according to the odds.

Fair odds: in words

- In the betting games we will talk about, the probability of each <u>outcome</u> is known.
- The bet is <u>**fair if:**</u>
  - The long term average of gains/losses is zero.
  - In other words: the **expected value** is zero.

Fair odds: in symbols

n: number of outcomes

Probabilities of outcomes $p_1, p_2, \ldots, p_n$

money gained for each outcome: $g_1, g_2, \ldots, g_n$

price of ticket: T

At each iteration, player pays T and gains one of $g_1, g_2, \ldots, g_n$

The expected gain of the player is $\sum_{i=1}^{n} p_i g_i - T$

The game is fair if $\sum_{i=1}^{n} p_i g_i - T = 0$

Equivalently: the price is fair if $T = \sum p_i g_i$

First Example

- House flips an unbiased coin.
  - “heads” : house pays player $1
  - “tails”: house pays player $2
- Outcomes: “heads”,”tails”
- What is the fair ticket price?
  - $1.5
  - Why? Because 0.5*1 + 0.5*2 = 1.5

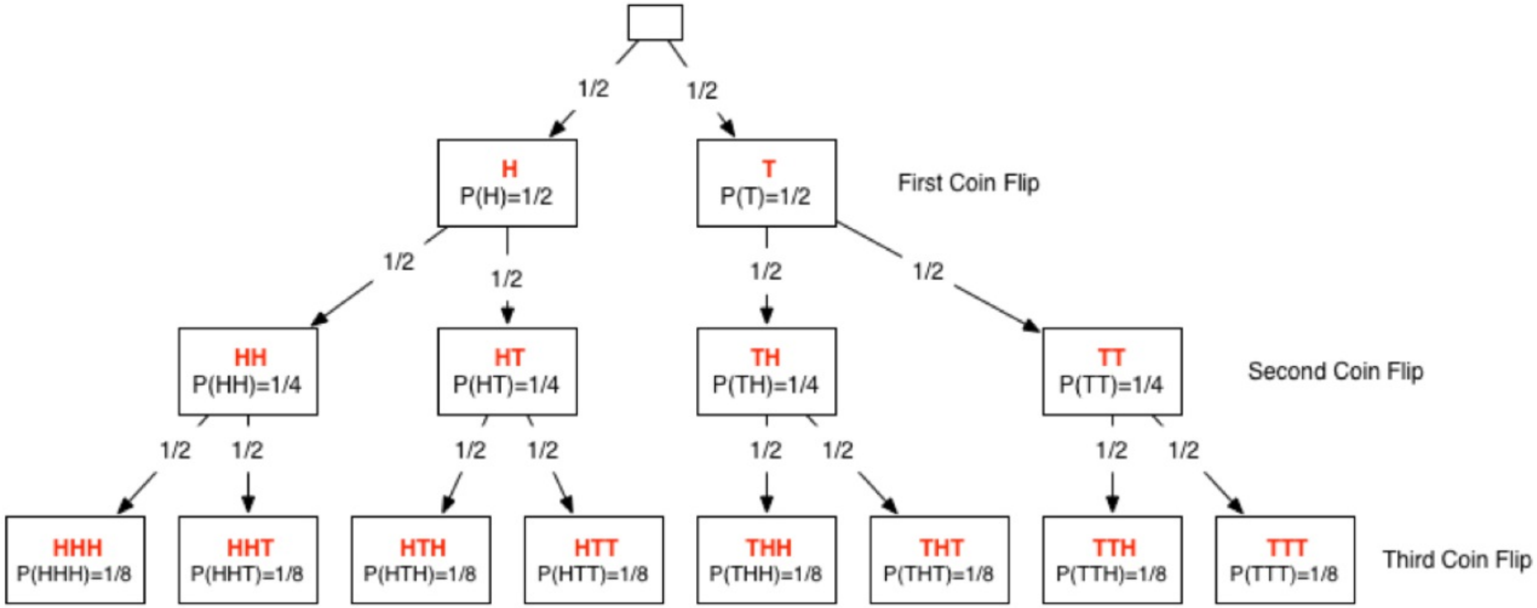
Second example

- The house flips the coin three times in a row.
- Eight outcomes:
  HHH,HHT,HTH,HTT, THH,THT,TTH,TTT
- Each outcome has probability 1/8
- Each outcome consists of three coin flips.
- You win $1 if there is exactly one T, 0$ otherwise, what is the fair price of the ticket?

Event Trees

- It sometimes helps our understanding to consider each coin flip separately, one by one.
- The result of considering all the possibilities is called the event tree.

The 3 coin flips event tree



What is an “Event” ?

- An event is a **set** of outcomes.
- Each node of an event tree defines an event
- The event of each node is a subset of the event of the parent.
  - The event “the first coin flip is H”. Corresponds to the set: {HHH,HHT,HTH,HTT}
  - The event “the first coin flip is T”. Corresponds to the set: {THH,THT,TTH,TTT}
  - The event “the first 2 coin flips are HH”. Corresponds to the set: {HHH,HHT}
  - The event “the first 2 coin flips are HT”. Corresponds to the set: {HTH,HTT}
  - …
  - The event “the three coin flips are HHH” corresponds to the set: {HHH}
- The probability of an event is the number of outcomes in the set, divided by 8.
- The event that contains all possible outcomes is called the “outcome space” and is denoted by Ω
- The probability of the whole outcome space is always 1:

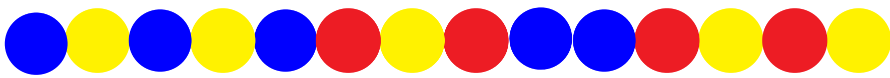
  Prob(Ω)=8/8=1

outcomes vs events

Gumball (Outcome) machine

Event = set of outcomes
or property of outcome

* Gumball is yellow
* Gumball is red
* Gumball is blue



Each gumball is an outcome

Also:
* Gumball is cracked
* Gumball is shiny.

Calculating probabilities of events

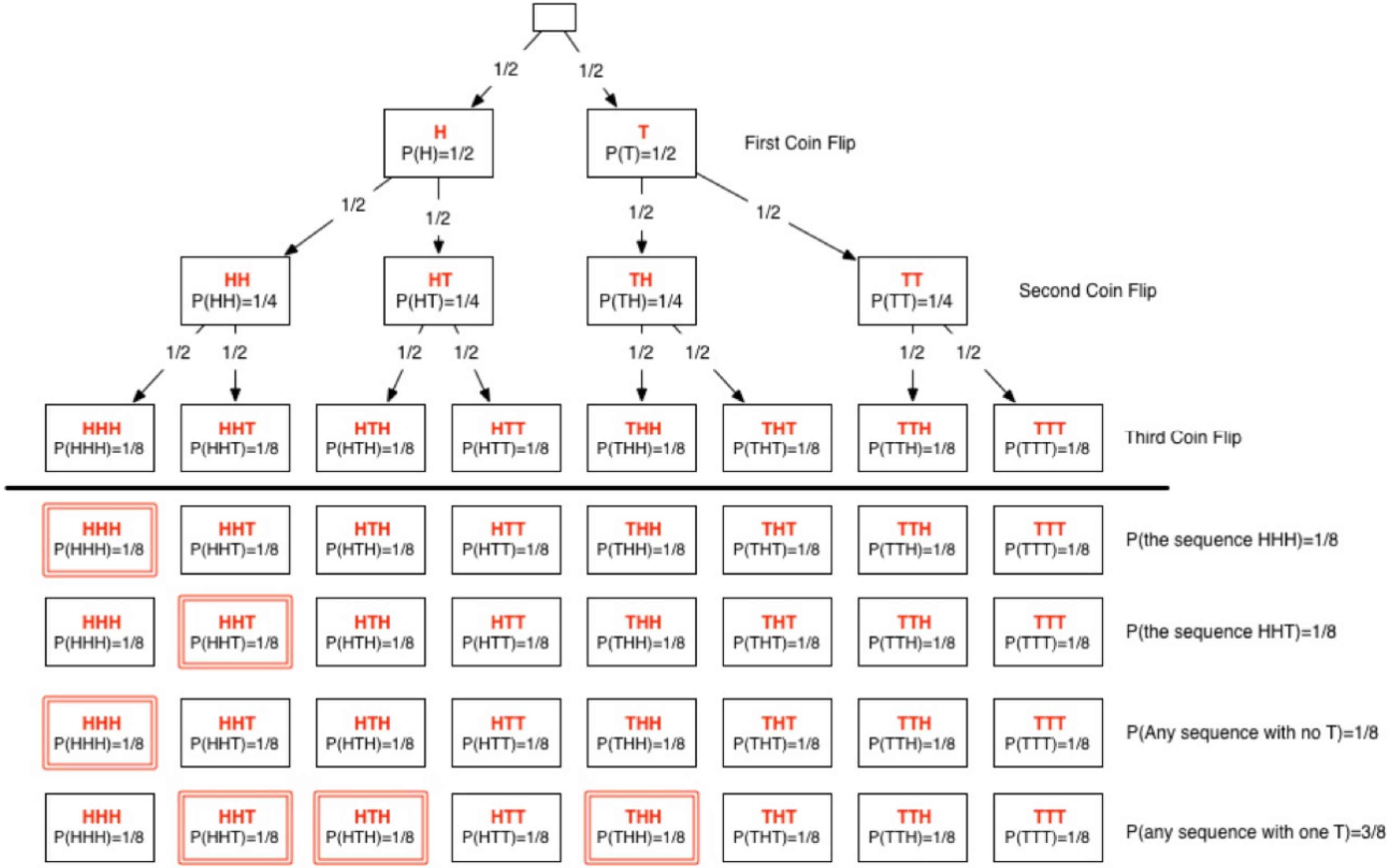
- The probability of an event (in the second example) is the number of outcomes in the set, divided by 8.
- The probability of an event is the number of outcomes in the event divided by the total number of outcomes in Ω (which is 8 in our case).
- Prob({HHH})=Prob({HHT})=1/8

The size of sets

- The number of elements in a set is also called the **size**, or **cardinality**, of the set.
- The size of the event A is denoted $|A|$
- The probability of the event A is $\mathrm{P}(A) = \frac{|A|}{|\Omega|}$
- *The fine print:* This is true for the special case of "uniform distribution over a finite sample space". Which will occupy us for the first 2 weeks or so.

Slightly more complex events

- P({The sequence contains no T})=P({HHH})=
- P({The sequence contains one T})=
  P({HHT,HTH,THH}) =
- While HHH,HHT,HTH,…. All have the same probability, the event defined by "one T" has three times the probability of "no T".
- The main task is to count the number of outcomes in the event. This is done using "combinatorics"



1/2
H P(H)=1/2
T P(T)=1/2
First Coin Flip
HH P(HH)=1/4
HT P(HT)=1/4
TH P(TH)=1/4
TT P(TT)=1/4
Second Coin Flip
HHH P(HHH)=1/8
HHT P(HHT)=1/8
HTH P(HTH)=1/8
HTT P(HTT)=1/8
THH P(THH)=1/8
THT P(THT)=1/8
TTH P(TTH)=1/8
TTT P(TTT)=1/8
Third Coin Flip
P(the sequence HHH)=1/8
P(the sequence HHT)=1/8
P(Any sequence with no T)=1/8
P(any sequence with one T)=3/8

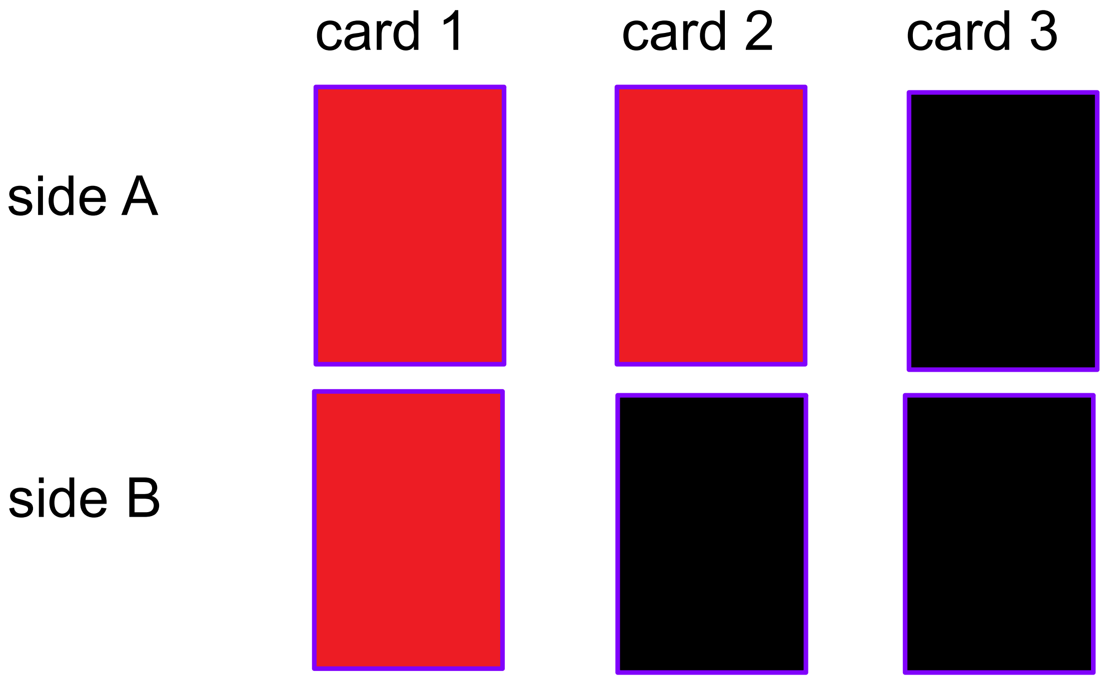
Ticket prices

- Suppose that the house pays you \$1 if a specified event happens, zero otherwise. What is the fair price?
- $T(E) = 1 * P(E) = P(E)$
  - In this special case, probability and expectation are the same.
- T({HHH})=T({HHT})=T({no T})=1/8=¢12.5
- T({one T})=\$3/8= ¢37.5

Do we really need all this?

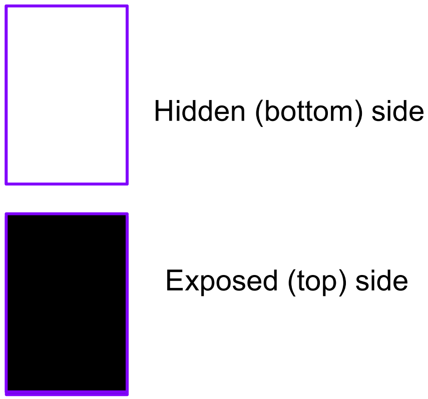
- Maybe you think:
  “Calculating the probabilities and expected values for what we have seen so far can be done intuitively, do we really need ‘events’ and ‘event trees’?”
- Here is an example which challenges the intuition.

The three card game



The cards are in a hat, pick one at random and
place it on one of the two sides at random

Choose one card at random, and put on random side.



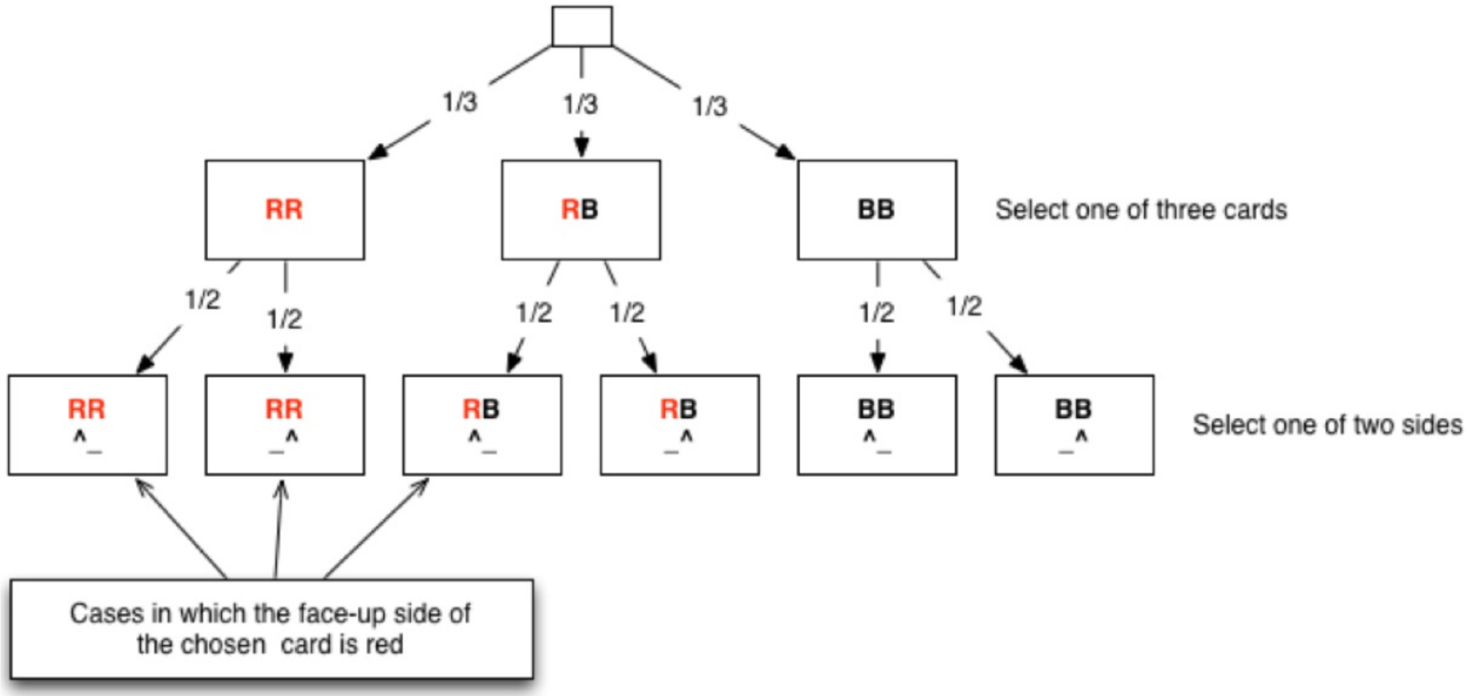
what is the color of the other side of the card?

What is the probability that the other side is red?

are the following odds fair?

Red - I give you 1$
Black - you give me 1$

Event tree for three cards



Conditional probability

- The probability that the seen color is R (B) is ½.
- The probability that the other side is R (B) given that the seen color is R(B) is 2/3.

For monday

* Read chapter 2 (this class)
* Read chapter 3 (next class)

* Register on edX
* Select consent
* Try solving problems, it might be hard,
  you'll do better after mon,wed classes.

* Come to discussions and ask questions!

We are often also interested in sets of tuples that cannot be expressed as products of individual sets. For instance, the points in the unit circle in $\mathbb{R}^2$ are

$$\{(x, y) : x^2 + y^2 \leq 1\},$$

a subset of $\mathbb{R}^2$ that cannot be written as a product $S_1 \times S_2$.

When S is product set of the form $S_1 \times \cdots \times S_k$, then the size of S is the product of the sizes of $S_1, S_2, \ldots, S_k$, i.e. $|S| = |S_1| \cdot |S_2| \cdots |S_k|$.

Chapter 3

Combinatorics

Probability theory is about sets. In the first lesson you encountered the set of possible outcomes of flipping two coins which contains four elements. If we use H to denote "Heads" and T to denote "Tails" then the set of possible outcomes, also known as the *outcome space* contains four elements: $\{(H,H),(H,T),(T,H),(T,T)\}$. We say that the outcome space is the set of four *tuples*. As the concepts of "set" and "tuple" will be used many times throughout the course, we take some pains to define these concepts and the associated notation, in a somewhat formal mathematical way.

Think of this as learning the syntax and semantics of a new programming language. There is nothing very deep here, it is just a framework through which deeper ideas can be expressed precisely and succinctly.

3.1 Sets

Sets are collections of *elements*. We will mostly consider sets of numbers, but elements can be shapes, colors, people, almost anything.

A *set* can be specified by listing its elements within braces, as in

$$A = \{1,2,3,4,5,6\} \text{ (the possible outcomes of the roll of a die)}$$
$$B = \{Red, Yellow, Blue\}$$
$$C = \{H,T\} \text{(the possible outcomes of the flip of a coin)}$$

We say that 5 is an element of A (or simply "5 is in A") and denote it by $5 \in A$. Sets are *unordered* collections, in other words $\{1,2,3,4,5,6\} = \{5,2,1,3,4,6\}$. The number of times an element can appear in a set is either 0 or 1, an element cannot appear multiple times in the set (for that there is a different construct called *bags*).

Instead of listing the elements of a set, one can define a set by specifying the precise conditions for an element to be in the set. For example:

$$\mathbb{Z}^+ = \{x : x \text{ is a positive integer}\}$$
$$\mathbb{R} = \{x : x \text{ is a real number}\}$$
$$A = \{x : x \in \mathbb{Z}^+ \text{ and } 1 \leq x \leq 6\}$$
$$I = \{x : x \in \mathbb{R} \text{ and } 0 \leq x \leq 1\}$$

In this notation, the part before the colon defines the form of the element, and the part after the colon defines the conditions that this element must obey to belong to the set.

Note that $\mathbb{Z}^+$, $\mathbb{R}$ and I are *infinite sets*, infinite sets can only be defined using a conditional definition. Finite sets such as A can be defined either by enumeration or by using a conditional definition.

We use either $\{\}$ or $\emptyset$ to denote the *empty set*. The empty set contains no elements.

3.1.1 The size of a set

The *size* of a set A is the number of elements in it and is denoted by $|A|$. Thus $|\{H, T\}| = 2$, $|\emptyset| = 0$ and $|\{x : x \in \mathbb{Z}^+ \text{ and } 50 \leq x \leq 100\}| = 51$.

Sets can be infinite, thus $\mathbb{Z}^+, \mathbb{R}$ and I are infinite sets. In fact even though both $\mathbb{Z}^-$ and I are infinite, I is "more infinite" than $\mathbb{Z}^+$. We say that $\mathbb{Z}^+$ is *countably infinite* while I is *uncountably infinite* which is much larger. [1]

3.1.2 Supersets and subsets

When set S is contained in set T (that is, $x \in S \Rightarrow x \in T$), we write $S \subseteq T$. For instance, $\mathbb{Z}^+ \subseteq \mathbb{R}$. The *empty set* contains no elements and is denoted by $\{\}$ or $\emptyset$.

Suppose A, B are two sets. The *intersection* of the two sets, denoted $A \cap B$ contains the elements that are in *both* sets. Thus

$$\{1, 2, 3, 4, 5, 6\} \cap \{2, 4, 6, 8, 10\} = \{2, 4, 6\}$$

The *union* of two sets contains those elements that are *either* A or B (or both), thus

$$\{1, 2, 3, 4, 5, 6\} \cup \{2, 4, 6, 8, 10\} = \{1, 2, 3, 4, 5, 6, 8, 10\}$$

We say that two sets A, B are *disjoint* if $A \cap B = \emptyset$. In other words, if A and B have not element in common.

3.1.3 Spaces and complements

In probability we call the set of all possible outcomes (for a particular experiment) as *the outcome space* and denote it by Ω. Subsets of Ω are called *events*. The *complement* of an event A is the set of outcomes that are in Ω but not in A. The complement of the set A is denoted A^c. For example, if $\Omega = \{1, 2, 3, 4\}$ is the outcome space and $A = \{2, 4\}$ is an event then $A^c = \{1, 3\}$.

It is not hard to see that A and A^c are disjoint, i.e. $A \cap A^c = \emptyset$. It is also clear that $A \cup A^c = \Omega$. We say that the pair of sets $\{A, A^c\}$ is a *partition* of Ω. You can think of a partition as associating a binary label such as $\{0, 1\}$ with each element of Ω thereby dividing Ω into two complementary sets.

Partitions can divide into more than two parts. A partition of Ω into k parts consists of a set of k non-empty sets $A_1, A_2, \ldots, A_k$ that are pairwise disjoint, i.e. for any $1 \leq i < j \leq k$ $A_i \cap A_j = \emptyset$ and $A_1 \cup A_2 \cup \cdots \cup A_k = \Omega$.

3.1.4 Tuples, and products of sets

Suppose we toss a coin three times. We can represent the outcome by a 3-tuple like (H, H, T) (where H means heads and T means tails). The set of all such tuples is

$$\{(H, H, H), (H, H, T), (H, T, H), (H, T, T), (T, H, H), (T, H, T), (T, T, H), (T, T, T)\}.$$

[1] The difference between countably and uncountably infinite sets is known in math as "totally deep, dude!" and is not for the faint of heart. If you feel up to the task, check out the Utube video "Vsouce: The Banach Tarski paradox" `https://www.youtube.com/watch?v=s86-Z-CbaHA`

It can also be written as $\{H,T\} \times \{H,T\} \times \{H,T\}$, or even more simply, $\{H,T\}^3$.

More generally, if $S_1, S_2, \ldots, S_k$ are sets, then $S_1 \times S_2 \times \cdots \times S_k$ is the set of all k-tuples in which the first entry is from S_1, the second entry is from S_2, and so on. For instance,

$$\mathbb{R}^2 = \{\text{all points in the plane}\}$$
$$\mathbb{R}^d = \{\text{all points in } d\text{-dimensional space}\}.$$

Note that, unlike sets, *order* is significant in tuples. Also note that the same element can appear multiple times in a tuple, but not in a set.

We are often also interested in sets of tuples that cannot be expressed as products of individual sets. For instance, the points in the unit circle in $\mathbb{R}^2$ are

$$\{(x,y) : x^2 + y^2 \leq 1\},$$

a subset of $\mathbb{R}^2$ that cannot be written as a product $S_1 \times S_2$.

When S is product set of the form $S_1 \times \cdots \times S_k$, then the size of S is the product of the sizes of $S_1, S_2, \ldots, S_k$, i.e. $|S| = |S_1| \cdot S_2 \cdots |S_k|$.

For example, suppose there is a group of n concert-goers, each of whom is selecting a band T-shirt. The available colors are red, yellow, and black. How many possible outcomes are there? Well, let $C = \{\text{red}, \text{yellow}, \text{black}\}$ be the set of possible colors, and represent each outcome as an n-tuple in which the ith entry is the color of the ith person's T-shirt. Then the possible outcomes are C^n, a set of size $|C|^n = 3^n$.

3.2 Permutations and combinations

Armed with the concepts of sets, tuples and set size we can now tackle some more interesting combinatorial questions.

3.2.1 Sampling with and without replacement when the order matters

Suppose there are four children Alice, Bill, Christie, and Doug at an animal shelter, checking out the current pool of n dogs. Each child writes down the name of the dog he or she likes most. How many possible outcomes are there?

We can represent each outcome as a 4-tuple (Alice's choice, Bill's choice, Christie's choice, Doug's choice) in which each entry is the name of a dog. So the number of outcomes is n^4.

Now suppose that these same children are actually picking out dogs. First Alice chooses a dog to adopt, then Bill chooses a dog to adopt, and so on. How many outcomes are there now?

In this situation, Alice has n choices, but Bill has only $n-1$ choices, Christie has $n-2$ choices, and Doug has $n-3$ choices. So there are $n(n-1)(n-2)(n-3)$ possible outcomes.

The first situation is called *sampling with replacement*: the outcomes are tuples in which the same element (dog) can occur more than once. The number of such k-tuples, chosen from n elements, is n^k. In the example, $k = 4$. The second situation is *sampling without replacement*: the outcomes are tuples in which no element can be repeated. The number of such k-tuples, chosen from n elements, is $n(n-1)(n-2)\cdots(n-k+1)$.

Here's a related question: how many ways are there to order (shuffle) a deck of 52 cards? (Each such ordering is called a *permutation* of the cards.) Well, the result is a 52-tuple, drawn from a set of size 52, in which no card is repeated. Therefore, the number of permutations is $52 \cdot 51 \cdot 50 \cdots 1$, which is called 52 *factorial* and denoted as 52!. More generally, the number of permutations of n elements is n factorial or $n!$.