# CIS7 Unit 6 – Chapter 9: Probability Notes

The mathematical theory of **probability** is a way of formally **representing and reasoning about uncertain events.**

In 1654, mathematicians Blaise Pascal and Pierre de Fermat corresponded about the odds for gambling outcomes and thus became co-founders of the theory of probability.

In 1774, Pierre-Simon Laplace defined the **probability of an event as the ratio of the number of favorable outcomes to the total number of possible outcomes**.

In this chapter, we will primarily follow **Laplace and mostly consider events where all outcomes are equally likely**. This is a pretty common occurrence and when it happens, the task of computing probabilities reduces to a task of counting.

An **experiment or random experiment** **yields one of a possible set of outcomes.**

We will **assume that all outcomes are equally likely.**

The probability of an event E that is a subset of a finite sample space S is

Probability of event E in sample space S.

Let E1, E2, andE3 be events in sample space S.

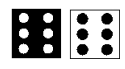
Probability of events E1, E2, E3 in sample space.

The Inclusion-Exclusion Principle tells us that

Inclusion-exclusion principle with events E1, E2, E3.

## Example 9.2 Dice

Experiment: Roll **a pair** of fair dice once.



1. What is the size of the sample space?

There are **6 possibilities for the black die** and **6 for the white one**. By **the product rule**, there are a total of **6 · 6 = 36 possible tosses**.

1. What is the probability of each of these events?

The **total is 6.**

There are exactly **5 ways that the two dice that result in a sum of 6**.

5 ways to obtain result of sum of 6 in a pair of dice.

So the **probability of the sum being 6 is 5/36**.

1. The total is 9.

There are exactly **4 ways that the two dice that result in a sum of 9**.

4 ways to obtain sum of 9 in a pair of dice.

So **the probability of the sum being 9 is 4/36 = 1/9**.

1. The total is greater than 8.

There are exactly **4 ways that the two dice that result in a sum of 9**,

4 ways to obtain sum of 9 in a pair of dice

Exactly **3 ways that the two dice that result in a sum of 10.**

3 ways to obtain sum of 10 in a pair of dice.

Exactly **2 ways that the two dice that result in a sum of 11**

2 ways to obtain sum of 11 in a pair of dice.

Exactly **1 ways that the** **two dice that result in a sum of 12**,

1 way to obtain sum of 12 in a pair of dice.

So the probability of a **sum greater than 8 is (4 + 3 + 2 + 1)/36 = 10/36 = 5/18**

1. The total is seven or eleven.

There are exactly **2 ways that the two dice result in a sum of 11**.

2 ways to obtain sum of 11 in a pair of dice

and **6 ways that the sum can be 7**

6 ways to obtain sum of 7 in pair of dice.

So the probability a **7 or 11 is (2 + 6)/36 = 8/36 = 2/9**.

1. Both **dice have the same number.**

There are **6 ways that the numbers can be the same**.

6 ways to get same numbers on a pair of dice.

**So the probability that the numbers are the same is 6/36 = 1/6**.

1. The **numbers are 4 and 3**.

There are exactly **2 ways this can happen so the probability** is **2/36 = 1/18**.

2 ways to obtain numbers 4 and 3 on a pair of dice.

1. Snake-eyes.

There is **only one way this can happen so the probability is 1/36**.

1 way to obtain a pair of 1s on a pair of dice.

## Exercise 9.2.2 Cards

Experiment: Draw a single card from a normal 52 card deck.

1. What is the size of the sample space?

Each card is a possible event. The **size of the sample space is 52**.

What is the probability of each of these events?

1. The card is a face card

There are **12 face cards** (**4 jacks, 4 queens, 4 kings**) so the **probability of a face card is 12/52** = **3/13**.

1. The card is black.

Half the cards are black so **the probability of drawing a black card is 1/2**.

1. The card is a heart.

There are **13 hearts** so the probability is **13/52 = 1/4**.

1. The card is a queen.

There are **4 queens** so the probability is **4/52 = 1/13**.

1. The card is a number (**2 through 10**).

There are **4 · 9 = 36** **number cards** so the probability is **36/52 = 9/13**

1. The card is the Ace of Spades.

There is **one Ace of Spades** so the probability is **1/52.**

1. The card is a joker.

The **probability is 0**. There are **no jokers in a standard deck of 52 cards**

## Exercise 9.2.3: Urns

Experiment: An urn contains 15 red balls and 10 blue balls. A single ball is drawn.

1. What is the size of the sample space?

There are **25 balls so the size of the sample space is 25**. You might want to think of the red balls as numbered from 1 to 15 and the black ones as numbered from 1 to

10.

What is the probability of each of these events?

1. The ball drawn is red.

There are **15 red balls** so the probability is **15/25 = 3/5**.

1. The ball drawn is blue.

There are 10 blue balls so the probability is **10/25 = 2/5**. Notice that **3/5+ 2/5 = 1**; **drawing a red ball and drawing a blue ball are complementary events**.

Experiment: An urn contains 15 red balls and 10 blue balls. Three balls are drawn at once.

1. What is the size of the sample space?

The number of possible draws is

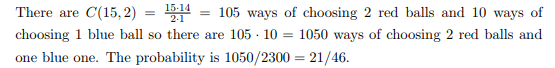
Choose 3 of 25 outputs 2300

What is the probability of each of these events?

1. All three balls are red.

Choose 3 of 15 outputs 455, probability of red balls being chosen is 91/460

1. Two balls are red and one is blue.



Experiment: An urn contains 15 red balls and 10 blue balls. Three balls are drawn sequentially and each is returned to the urn before the next ball is drawn.

1. What is the size of the sample space?

There are 25 choices for each of the balls so **there are possible 253 = 15625 sequences of three balls (with replacement).**

What is the probability of each of these events?

1. All three balls are red.

There are 153 = 3375 sequences of three red balls (with replacement) so the probability is

Choose 3 balls that are red, 27/125 probability.

1. Two balls are red and one is blue.

There are 3 positions for the blue ball.

There are 10 balls that can go in this position.

There are 152 = 225 sequences of two red balls for the remaining places.

In all, there are 3 · 10 · 152 = 6750 outcomes with two red balls and one blue ball. Choose 2 red balls and 1 blue ball, probability is 36/125

## Example 9.2.4 Bytes

Experiment: Toss a fair coin 8 times, 1 for heads, 0 for tails to generate a byte.

1. What is the size of the sample space?

There are **28 = 256 bytes**.

What is the probability of each of these events?

1. The byte has exactly four ones.

Each byte with exactly four ones corresponds to a subset of size 4 of the 8 positions in the byte. 70 subsets of 4 ones in byte position. 

Therefore, the probability of generating a byte with exactly four 1s is **70/256 = 35/128**.

1. The byte starts and ends with a 1.

The first and last positions in the byte are fixed but the other six places can be anything. The number of such bytes is 26 = 64 and the **probability is 26/28 =**

**1/4.**

1. The byte starts and ends with the same bit.

There are **twice as many bytes in the event than in part C**. The **probability is 1/2**.

1. The byte contains the substring 111111

Those strings with exactly 6 ones: 11111100, 01111110, 00111111

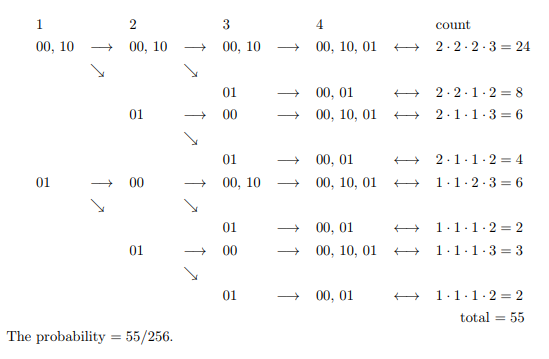
Those strings with exactly 7 ones: 11111110, 11111101, 01111111, 10111111

and one string with exactly 8 ones: 11111111.

The probability that the byte contains the **substring 111111 is 8/256 = 1/64.**

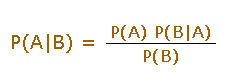
1. The byte does not contain 2 consecutive ones.

Think of making up the string from the **2-bit blocks, 00, 10, 01** with the constraint **that 01 can only be followed by 00 or 01**.



## Conditional Probability Bayes Theorem

Bayes Theorem a way of **finding a probability when we know certain other probabilities**.



Which tells us:

1. how often A happens given that B happens, written P(A|B),

When we know:

1. how often B happens given that A happens, written P(B|A)
2. and how likely A is on its own, written P(A)
3. and how likely B is on its own, written P(B)

**Example**: If dangerous fires are rare (1%) but smoke is fairly common (10%) due to barbecues, and 90% of dangerous fires make smoke then:

P(Fire|Smoke) =

= = 9%

Suppose you are a contestant in a game show. The host shows you **three closed doors**. **One door leads** to a **brand new car**, and the other **two doors each lead to a lemon**. If you are able to find the door leading to the car, you get the car! You are asked to select one of the doors, but not open it. You go ahead and select one of the doors. At this point, the host opens one of the other two doors that reveals a lemon, and asks: “Do you want to stay with your choice, or switch?” What should you do? Does it matter?

* If the car is behind A, then the host can open any one of the other two doors since both lead to lemons. And in this case, your correct response (in hindsight) is to stay.
* If the car is behind B, then the host will open door C. In this case, your correct response is to switch.
* Similarly, if the car is behind C, then the host will open door B, and your correct response is to switch.
* So in 2 out of the 3 equally likely cases, you should switch. This means you are more likely to win the car if you switch!

### Conditional Probability

The **conditional probability**: a measure of the probability of an event (some particular situation occurring) given that another event has occurred.

* Events can be **"Independent"**, meaning each **event is not affected by any other events**.

Example: Tossing a coin.

Each toss of a coin is a perfect isolated thing. ***What it did in the past will not affect the current toss?*** The chance is simply 1-in-2, or 50%, just like ANY toss of the coin. So each toss is an Independent Event.

* **Events can also be "dependent,” which means they can be affected by previous events.**

**Example:**

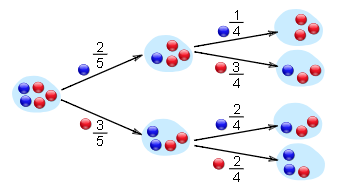
There are 2blue and 3 red marbles are in a bag.What are the chances of getting a blue marble?

The chance is 2 in 5**.** But after taking one out the chances change!

If we got a red marble before, then the chance of a blue marble next is 2 in 4.

If we got a blue marble before, then the chance of a blue marble next is 1 in 4.

**The next event depends on what happened in the previous event, and is called dependent.**



An **event E1 given event E2** is the probability **that event E1 occurs, given that event E2 occurs.** It is denoted by Pr[E1|E2] and can be defined as:

Probability formula given E1 intersects E2.

Example:

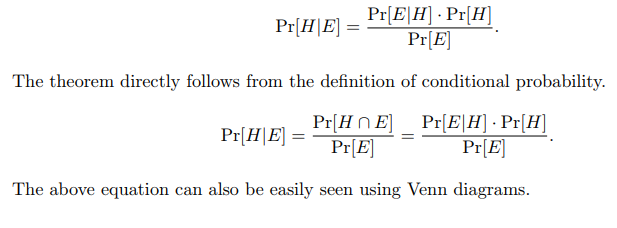
Consider the roll of two fair dice. A **priori**, the probability that the roll of the first die yields a 5 is 1/6. But if we are told that the sum of the two dice is 9, then the probability of obtaining a 5 in the first roll, given this new information changes

Formula for rolling dice to obtain a 5 in the first roll, 1/4 probability.

### Bayes Theorem

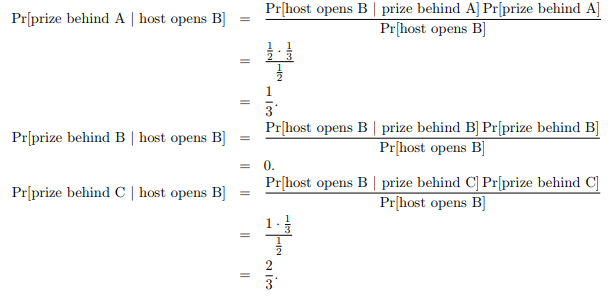
**Bayes Theorem** relates one **conditional probability** (e.g., the probability of a hypothesis H given an observation E) with its inverse (the probability of an observation given a hypothesis).

Bayes Theorem is used heavily in statistics, analysis of data sets, machine learning, information retrieval, and several diverse applications in science and engineering.



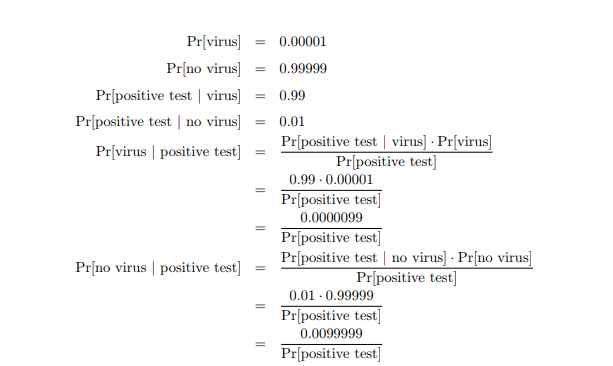
### Explaining Monty Hall paradox using Bayes Theorem

When the host asks you whether you would like to switch, the calculation you should do is to determine the probability that you will win the prize given the information provided to you.

As before, suppose without loss of generality that the door you select is labeled A. Also, let us label the door that the host opens as B. Now, consider the following calculations.

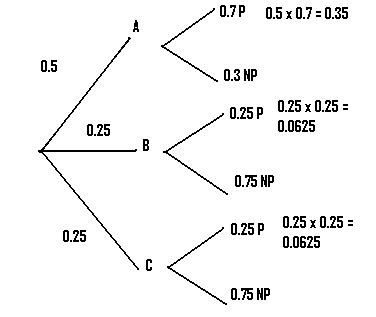
You would be better off switching from door A to door C, i.e., from the door you have selected to the other unopened door.

Another application of Bayes TheoremSuppose the a priori probability of you being infected with the H1N1 virus is 10−5. Further suppose that a blood test is 99% accurate and you test positive. How likely is it that you actually have the virus? Let us do the calculations.

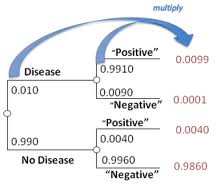


### Tree-diagram

A probability tree has two main parts: **the branches and the ends**. The probability of each branch is generally written on the branches, while the outcome is written on the ends of the branches.



Probability Trees make the question of whether to multiply or add probabilities: **multiply along the branches and add probabilities down the columns**.



## Markov Chains

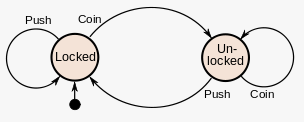
Markov Chains: a mathematical system that experiences t**ransitions from one state to another according to certain probabilistic rules**.

Markov chains are often used to model **probabilistic systems when the chance of a future event is only dependent on the present and not dependent on the past**.

**Finite State Machine**: also known as **finite state automaton** is a computation model that can be implemented with hardware or software and can be used to simulate **sequential logic** and some computer programs. Finite state automata generate regular languages. Finite state machines can be used to model problems in many fields including mathematics, artificial intelligence, games, and linguistics.

**A system where particular inputs cause particular changes in state can be represented using finite state machines.**

This example describes the various states of a turnstile. Inserting a coin into a turnstile will unlock it, and after the turnstile has been pushed, it locks again. Inserting a coin into an unlocked turnstile, or pushing against a locked turnstile will not change its state.

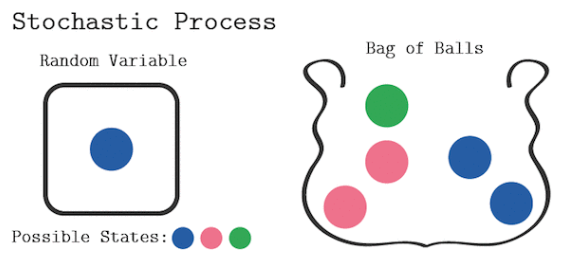


There are **two types of finite state machines** (FSMs): **deterministic finite state machines**, often called **deterministic finite automata**, and **non-deterministic finite state machines**, often called **non-deterministic finite automata**. There are slight variations in ways that state machines are represented visually, but the ideas behind them stem from the same computational ideas. By definition, **deterministic finite automata recognize, or accept, regular languages, and a language is regular if a deterministic finite automaton accepts it**. FSMs are usually taught using languages made up of **binary strings that follow a particular pattern**. Both regular and non-regular languages can be made out of binary strings.

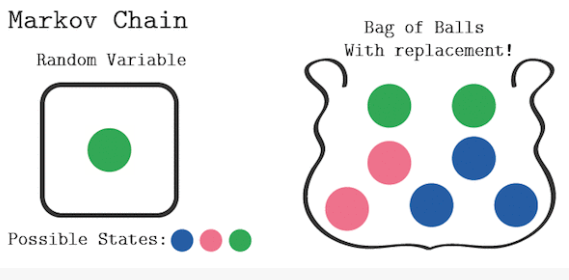
An example of a binary string language is: the language of all strings that have a 0 as the first character. In this language, 001, 010, 0, and 01111 are valid strings (along with many others), but strings like 111, 10000, 1, and 11001100 (along with many others) are not in this language.

A Markov chain is a **stochastic process**, **the process of some values changing randomly over time**, but it differs from a general stochastic process in that a Markov chain must be **"memory-less"**. That is, (**the probability of) future actions are not dependent upon the steps that led up to the present state**. This is called the **Markov property**. While the theory of Markov chains is important precisely because so many "everyday" processes satisfy the Markov property, there are many common examples of stochastic properties that do not satisfy the Markov property.

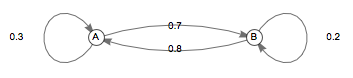
A common probability question ask: What is the probability of getting a certain color ball, when selecting uniformly and at random from a bag of multicolored balls? It could also ask what the probability of the next ball is, and so on. In such a way, **a stochastic process begins to exist with color for the random variable, and it does not satisfy the Markov property**. Depending upon which balls are removed, the probability of getting a certain color ball later may be drastically different.



A variant of the same question asks once again for ball color, but **it allows replacement each time a ball is drawn.** Once again, this **creates a stochastic process** with color for the random variable. This process, however, does satisfy the Markov property.



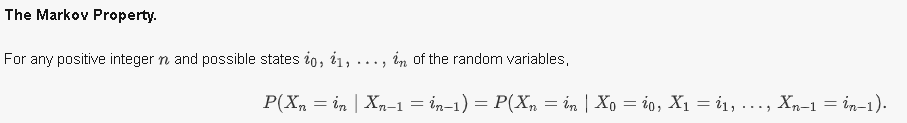
A (time-homogeneous) Markov chain built on states A and B is depicted in the diagram below. What is the probability that a process beginning on A will be on B after 2 moves?



In order to move from A to B, the process must either stay on A the first move, then move to B the second move; or move to B the first move, then stay on B the second move. According to the diagram, the probability of that is: 0.3 \* 0.7 + 0.7 \* 0.2 = 0.35

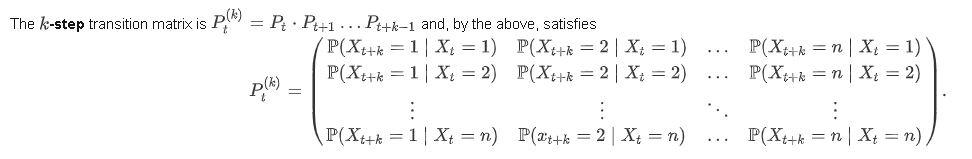
Alternatively, the probability that the process will be on A after 2 moves is: 0.3 \* 0.3 + 0.7 \* 0.8 = 0.35

Since there are only two states in the chain, the process must be on B if it is not on A, and therefore, the probability that the process will be on B after 2 moves is: 1 – 0.65 = 0.35



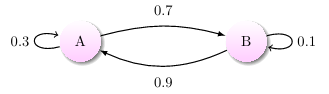
**The sum of the (conditional) probabilities associated with all transition arcs outgoing from a state must be 1.**

**Transition matrix**: is a matrix containing information on the probability of transitioning between states.



Example:

For the time-independent Markov chain described by the picture below, what is its -**step transition matrix**?



Transition Matrix: P

2-step Transition Matrix:

P2 = \* = =

Note that the **sum of each row must be 1**, corresponding to the fact that the sum of the (conditional) probabilities associated with **all transition arcs outgoing from a state must be 1**.

Any matrix whose rows sum to 1 and whose elements are values in the range [0, 1] (and thus can be interpreted as probabilities) is said to be **stochastic**. Every stochastic matrix corresponds to a Markov chain and can be interpreted as such.