

# Class Notes

Math 525 - Probability Theory  
Professor Tao Chen  
University of Michigan



**Jinyan Miao**  
and his friends from UMich Honors Math

Fall 2022



# Contents





## Probability Space

Probability is a study of mathematical models of random events.

### Definition 1.0.0.0.1

*For a mathematical model of an experiment of random events, let  $\Omega$  be the set of all possible outcomes,  $\Omega$  is called the sample space. An event is some subset of  $\Omega$ , but not all subsets of  $\Omega$  are considered to be events.*

**Example:** For flipping a coin,  $\Omega = \{H, T\}$ , where  $H$  is getting a head and  $T$  is getting a tail. For flipping two coins one by one,  $\Omega = \{HH, HT, TH, TT\}$ . For rolling a die,  $\Omega = \{1, 2, 3, 4, 5, 6\}$ , each number corresponds to the event of getting that number from the die rolling. For randomly picking a ball from a box that has one red ball and two blue balls,  $\Omega = \{R, B\}$ . Notice that for the last example, there are three balls in the box, but the sample space has only two elements.

**Example:** The process of randomly picking a positive integer, the sample space is a countable infinite set  $\Omega = \mathbb{N}$ .

**Example:** The process of randomly throwing a dart at a dart board, and assuming that the dart will always land on the board, then  $\Omega$  is an uncountable infinite set.

### Definition 1.0.0.0.2

*For a mathematical model of an experiment of random events,  $\mathcal{F}$  is the set of events, and mathematically,  $\mathcal{F}$  is a  $\sigma$ -algebra on the sample space  $\Omega$ .*

**Example:** Consider the experiment of flipping two coins one by one. If the event is *first coin is head*, then the corresponding mathematical representation of such event is  $\{HH, HT\}$ . If the event is *getting at least one head*, then the mathematical representation of such event is  $\{HH, HT, TH\}$ . If the event is *at least one head or at least one tail*, then the mathematical representation of such event is  $\{HH, HT, TH, TT\} = \Omega$ . The event of *getting three heads* is represented by  $\emptyset$  because it is clearly impossible to get such outcome. Notice that, the union of the representation of the event *getting at least one head*, with that of the event *getting at least one tail*, is  $\Omega$ :

$$\{HH, HT, TH\} \cup \{HT, TH, TT\} = \{HH, HT, TH, TT\} = \Omega$$

**Example:** Consider the experiment of rolling a die.

Event	Representation
<i>Getting even number</i>	$\{2, 4, 6\}$
<i>Getting even number that is greater than or equal to 4</i>	$\{4, 6\}$
<i>Getting a number that is greater or equal to 4</i>	$\{4, 5, 6\}$

Note that, for the first and the last events, the intersection of their representation gets us the second event.

**Example:** Consider throwing a dart at a dart board, and consider the dart board to have 1 meter radius. Event examples of such experiment includes *the dart lands on the board less than or equal to 0.3 meter distance from the center*, and *the dart lands on the board more than or equal to 0.7 meter distance from the center*.

**Note:** When we say an event  $A$  happens, that means that we get some outcome  $\omega \in \Omega$  such that  $\omega \in A$ . On the other hand,  $A$  does not happen if  $\omega \notin A$ .

**Remark:** We can do set operation of the representations of two events  $A, B$  as follows:

$$\begin{aligned}
 A \cup B & \quad \text{(at least one of } A \text{ and } B \text{ happens)} \\
 A \cap B & \quad \text{(both } A \text{ and } B \text{ happens)} \\
 A^c := \Omega - A & \quad \text{(} A \text{ does not happen)}
 \end{aligned}$$

**Remark:** For a mathematical model of experiment outcomes, if the sample space  $\Omega$  is finite, then one possible choice of the set of events  $\mathcal{F}$  is  $\mathcal{F} := 2^\Omega$ , where  $2^\Omega$  is the set of all subsets of  $\Omega$  called the power set of  $\Omega$ .

**Example:** For the experiment of flipping a coin,  $\Omega = \{H, T\}$  defines the sample space, and hence  $2^\Omega = \{\emptyset, \Omega, \{H\}, \{T\}\}$  can be chosen as  $\mathcal{F}$ .

**Note:** If the sample space  $\Omega$  for a mathematical model is infinite or uncountable, then  $\mathcal{F}$  cannot be chosen as  $2^\Omega$ .

#### Definition 1.0.0.0.3

For a mathematical model of an experiment of random events,  $\mathbb{P} : \mathcal{F} \rightarrow [0, 1]$  defines a measure on the sample space and it measures the likelihood of an event to occur. In other words, any input of  $\mathbb{P}$  should be an event, we can only talk about probability of an event.

**Note:** For sample space  $\Omega$ , event  $A \subseteq \Omega$ , and outcome  $\omega \in \Omega$ , we note here  $\mathbb{P}(\omega)$  is not well-defined, while  $\mathbb{P}(A)$  and  $\mathbb{P}(\{\omega\})$  might be well-defined.

**Example:** Now consider flipping two fair coins one by one, and consider the event of *getting at least one head*, denote such event as  $A$ . Here  $A = \{HH, HT, TH\}$ , and the sample space  $\Omega = \{HH, HT, TH, TT\}$ .  $\mathbb{P}$  can be defined as the following:

$$\mathbb{P}(A) := \frac{\#(E)}{\#(\Omega)} \text{ for any event } E \subseteq \Omega \quad (1.1)$$

in which case we get  $\mathbb{P}(A) = 3/4$ . For the event of *getting two heads*, denoted as  $B$ , we write  $\mathbb{P}(B) = 1/4$ .

**Note:** The definition of  $\mathbb{P}$  proposed by (1.1) in fact does not solve all problems in all models. We consider a simple example of flipping an unfair coin, which has 60% of getting a head. The event  $A = \{H\}$  of *getting a head* should have probability of  $\mathbb{P}(A) = 0.6 \neq 1/2$ . Moreover, we see that (1.1) only defines  $\mathbb{P}$  when both  $\Omega$  and  $E$  are finite, that is, (1.1) fails for infinite sample space  $\Omega$ .

In the following, we revise Definition 1.0.0.0.2 in a general way:

**Definition 1.0.0.0.4**

Let  $\Omega$  be a set,  $\mathcal{F}$  is called a  $\sigma$ -algebra on  $\Omega$  provided that it satisfies:

1.  $\emptyset \in \mathcal{F}$
2. If  $A_n \in \mathcal{F}$  for all  $n \in \mathbb{N}$ , then  $\bigcup_{n \in \mathbb{N}} A_n \in \mathcal{F}$
3. If  $A \in \mathcal{F}$ , then  $A^c \in \mathcal{F}$

**Remark:** For a sample space  $\Omega$ , the  $\sigma$ -algebra  $\mathcal{F}$  on  $\Omega$  is usually chosen as the set of all possible events.

**Example:** Consider a sample space  $\Omega$ , one can choose  $\mathcal{F} = \{\emptyset, \Omega\}$ . For a subset  $A \subseteq \Omega$ , one can also choose  $\mathcal{F} = \{\emptyset, A, A^c, \Omega\}$ . If  $\Omega$  is finite, then one can take  $\mathcal{F} = 2^\Omega$ .

**Proposition 1.0.1**

Let  $\mathcal{F}$  be a  $\sigma$ -algebra on a sample space  $X$ , and let  $A_n \in \mathcal{F}$  for all  $n \in \mathbb{N}$ , then we have:

$$\bigcap_{n=1}^{\infty} A_n \in \mathcal{F}$$

*Proof.* Since  $A_n \in \mathcal{F}$  for all  $n \in \mathbb{N}$ , then by definition of  $\sigma$ -algebra, we have  $A_n^c \in \mathcal{F}$  for all  $n \in \mathbb{N}$ , and we can write the following according to De Morgan's Law:

$$\bigcap_{n=1}^{\infty} A_n = \left( \left( \bigcap_{n=1}^{\infty} A_n \right)^c \right)^c = \left( \bigcup_{n=1}^{\infty} A_n^c \right)^c \in \mathcal{F}$$

This completes the proof. □

**Definition 1.0.1.0.1**

A probability measure on a sample space  $\Omega$  equipped with the  $\sigma$ -algebra  $\mathcal{F}$  is a function  $\mathbb{P} : \mathcal{F} \rightarrow [0, 1]$  that satisfies the followings:

1.  $\mathbb{P}(\emptyset) = 0$ ,  $\mathbb{P}(\Omega) = 1$
2. If  $A_n \in \mathcal{F}$  for all  $n \in \mathbb{N}$  with  $A_i \cap A_j = \emptyset$  for all  $i \neq j$ , then we have:

$$\sum_{n=1}^{\infty} \mathbb{P}(A_n) = \mathbb{P} \left( \bigcup_{n=1}^{\infty} A_n \right)$$

**Definition 1.0.1.0.2**

For sample space  $\Omega$ ,  $\sigma$ -algebra  $\mathcal{F}$  on  $\Omega$ , and probability function  $\mathbb{P} : \mathcal{F} \rightarrow [0, 1]$ , the triple  $(\Omega, \mathcal{F}, \mathbb{P})$  is called a probability space.



**Remark:** Any random event is uniquely described by a triple  $(\Omega, \mathcal{F}, \mathbb{P})$ .

**Example:** Consider an experiment of flipping a fair coin. Here we can write  $\Omega = \{H, T\}$ ,  $\mathcal{F} = \{\emptyset, \{H\}, \{T\}, \Omega\}$ , and  $\mathbb{P}$  be defined by:

$$\mathbb{P}(\{H\}) = \mathbb{P}(\{T\}) = \frac{1}{2}$$

On the other hand, if the coin is not fair, we can instead make  $\mathbb{P}$  be defined by:

$$\mathbb{P}(\{H\}) = 0.4 \qquad \mathbb{P}(\{T\}) = 0.6$$

**Proposition 1.0.2**

Let  $(\Omega, \mathcal{F}, \mathbb{P})$  be a probability space, and let  $A, B \in \mathcal{F}$ . We have the followings:

1.  $\mathbb{P}(A^c) = 1 - \mathbb{P}(A)$ .
2. If  $A \subseteq B$ , then we have  $\mathbb{P}(B) = \mathbb{P}(A) + \mathbb{P}(B \setminus A) \geq \mathbb{P}(A)$ .
3. If  $A_1, A_2, \dots, A_n \in \mathcal{F}$ , then we can write:

$$\mathbb{P}\left(\bigcup_{i=1}^n A_i\right) = \sum_{i=1}^n \mathbb{P}(A_i) - \sum_{i < j} \mathbb{P}(A_i \cap A_j) + \sum_{i < j < k} \mathbb{P}(A_i \cap A_j \cap A_k) - \dots + (-1)^{n-1} \mathbb{P}\left(\bigcap_{i=1}^n A_i\right)$$

4. If  $A_i \in \mathcal{F}$  for all  $i \in \mathbb{N}$  with  $A_i \subseteq A_{i+1}$ , then we have:

$$\lim_{i \rightarrow \infty} \mathbb{P}(A_i) = \mathbb{P}\left(\bigcup_{i=1}^{\infty} A_i\right)$$

5. If  $A_i \in \mathcal{F}$  for all  $i \in \mathbb{N}$  with  $A_i \supseteq A_{i+1}$ , then we have:

$$\lim_{i \rightarrow \infty} \mathbb{P}(A_i) = \mathbb{P}\left(\bigcap_{i=1}^{\infty} A_i\right)$$

*Proof.* Let  $A \in \mathcal{F}$ , then  $A$  and  $A^c$  are disjoint and  $A \cup A^c = \Omega$ , then we can write:

$$1 = \mathbb{P}(\Omega) = \mathbb{P}(A) + \mathbb{P}(A^c) \quad \Rightarrow \quad \mathbb{P}(A^c) = 1 - \mathbb{P}(A)$$

This proves (1). For (2), since  $A \subseteq B$ , then we can write  $B = A \cup (B \setminus A)$ , and hence we can write the following with the fact that  $\mathbb{P}(B \setminus A) \geq 0$ :

$$\mathbb{P}(B) = \mathbb{P}(A) + \mathbb{P}(B \setminus A) \geq \mathbb{P}(A)$$

For (3), one can proceed by induction, here we do the base case where  $n = 2$ , note that we can write the following for the case where  $n = 2$ :

$$A_1 \cup A_2 = A_2 \cup (A_1 \setminus A_2) \qquad A_1 = (A_1 \cap A_2) \cup (A_1 \setminus A_2)$$

Then we can write:

$$\mathbb{P}(A_1 \cup A_2) = \mathbb{P}(A_2) + \mathbb{P}(A_1 \setminus A_2) \qquad \mathbb{P}(A_1) = \mathbb{P}(A_1 \cap A_2) + \mathbb{P}(A_1 \setminus A_2)$$

and hence combining we see that we have:

$$\mathbb{P}(A_1 \cup A_2) = \mathbb{P}(A_2) + \mathbb{P}(A_1) - \mathbb{P}(A_1 \cap A_2)$$

For (4), we define  $B_1 = A_1$ ,  $B_2 = A_2 \setminus A_1$ , and  $B_i = A_i \setminus A_{i-1}$  for  $i \geq 2$ . Then we can write the following as  $B_i$  are pairwise disjoint and  $\bigcup_{i=1}^{\infty} B_i = \bigcup_{i=1}^{\infty} A_i$ :

$$\begin{aligned}
 \mathbb{P}\left(\bigcup_{i=1}^{\infty} A_i\right) &= \mathbb{P}\left(\bigcup_{i=1}^{\infty} B_i\right) = \sum_{i=1}^{\infty} \mathbb{P}(B_i) \\
 &= \lim_{n \rightarrow \infty} \sum_{i=1}^n \mathbb{P}(B_i) \\
 &= \lim_{n \rightarrow \infty} \left( \mathbb{P}(A_1) + \sum_{i=2}^n (\mathbb{P}(A_i) - \mathbb{P}(A_{i-1})) \right) \\
 &= \lim_{n \rightarrow \infty} \mathbb{P}(A_n)
 \end{aligned}$$

One can show (5) with similar argument, and that completes the proof of the proposition.  $\square$

## Conditional Probability

Conditional probability is the probability with given information. For example, consider the experiment of flipping a fair coin 3 times independently. Consider the event of *getting exactly two heads*, denoted as  $A = \{HHT, HTH, THH\}$ , and consider the event of *first is head*, denoted as  $B = \{HHT, HTT, HTH, HHH\}$ . Given that the first is head, the probability of having exactly 2 heads is an example of conditional probability:

$$\text{probability of having exactly 2 heads given that the first is head} = \frac{2}{4} = \frac{1}{2}$$

### Definition 2.0.0.1

Let  $A$  and  $B$  be two events such that  $\mathbb{P}(B) > 0$ , then the conditional probability of  $A$  given  $B$  is defined by the following:

$$\mathbb{P}(A|B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)}$$

For the example of flipping a fair coin 3 times given above, we have  $A \cap B = \{HHT, HTH\}$ , and hence we can write:

$$\mathbb{P}(A|B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)} = \frac{\frac{\#(A \cap B)}{\#(\Omega)}}{\frac{\#(B)}{\#(\Omega)}} = \frac{\#(A \cap B)}{\#(B)} = \frac{2}{4} = \frac{1}{2}$$

### Theorem 2.1 (Law of Total Probability)

Let  $B_1, B_2, \dots, B_n$  be a partition of a set  $\Omega$ , that is, we write:

$$\bigcup_{i=1}^n B_i = \Omega \quad B_i \cap B_j = \emptyset \text{ for } i \neq j, \quad \text{and } \mathbb{P}(B_i) > 0 \text{ for all } 1 \leq i \leq n$$

Then we have the following holds:

$$\mathbb{P}(A) = \sum_{i=1}^n \mathbb{P}(A|B_i) \cdot \mathbb{P}(B_i) = \sum_{i=1}^n \mathbb{P}(A \cap B_i)$$

From Theorem 2.1, in particular, let  $B$  be such that  $0 < \mathbb{P}(B) < 1$ , then we can write:

$$\mathbb{P}(A) = \mathbb{P}(A|B) \cdot \mathbb{P}(B) + \mathbb{P}(A|B^c) \cdot \mathbb{P}(B^c)$$

where  $B$  and  $B^c$  together form a partition of  $\Omega$ .

**Example:** Suppose we have 3 urns. The first urn has three red balls and a blue ball, the second urn has one blue ball and one red ball, the third urn has three blue balls and two red balls. Here we pick an urn at random, then pick a ball from that urn at random. Define event  $A$  as *picking a red ball*, here we want to find  $\mathbb{P}(A)$ :

$$\begin{aligned} \mathbb{P}(A) &= \mathbb{P}(A|U_1) \cdot \mathbb{P}(U_1) + \mathbb{P}(A|U_2) \cdot \mathbb{P}(U_2) + \mathbb{P}(A|U_3) \cdot \mathbb{P}(U_3) \\ &= \frac{3}{4} \cdot \frac{1}{3} + \frac{1}{2} \cdot \frac{1}{3} + \frac{2}{5} \cdot \frac{1}{3} = \frac{11}{20} \end{aligned}$$

where  $U_i$  denote the event of *picking the  $i$ -th urn*.

**Theorem 2.2** (Bayes Formula)

Let  $A$  and  $B$  be events such that  $\mathbb{P}(A) > 0$  and  $\mathbb{P}(B) > 0$ , then we can write the following:

$$\mathbb{P}(A|B) = \frac{\mathbb{P}(B|A) \cdot \mathbb{P}(A)}{\mathbb{P}(B)}$$

*Proof.* Here we can write the following:

$$\mathbb{P}(A|B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)} = \frac{\mathbb{P}(B \cap A)}{\mathbb{P}(B)} = \frac{\mathbb{P}(B|A) \cdot \mathbb{P}(A)}{\mathbb{P}(B)}$$

and that completes the proof.  $\square$

**Example:** Suppose that 0.5% of the population has some disease  $D$ , and there is a test that detects  $D$ . 96% of the time, the test will detect  $D$  given that a person gets  $D$ . The test might also give false positives 2% of the time. Suppose someone takes the test and gets positive result. Here we denote  $B$  as the event *the person has the disease*, and denote  $A$  as the event *the test is positive*. The probability that such person has the disease  $D$  for given information is given by the following:

$$\begin{aligned} \mathbb{P}(B|A) &= \frac{\mathbb{P}(A|B) \cdot \mathbb{P}(B)}{\mathbb{P}(A)} = \frac{\mathbb{P}(A|B) \cdot \mathbb{P}(B)}{\mathbb{P}(A|D) \cdot \mathbb{P}(D) + \mathbb{P}(A|D^c) \cdot \mathbb{P}(D^c)} \\ &= \frac{0.96 \cdot 0.005}{0.005 \cdot 0.96 + 0.995 \cdot 0.02} = 0.1943 \end{aligned}$$

## Independence

*Independence says that one event happens or not does not effect other events.*

### **Definition 3.0.0.0.1**

*Event  $A$  and event  $B$  are independent provided that we have:*

$$\mathbb{P}(A \cap B) = \mathbb{P}(A) \cdot \mathbb{P}(B)$$

### **Definition 3.0.0.0.2**

*For a set of events  $\mathcal{A} = \{A_i \mid i \in \mathcal{J}\}$  where  $\mathcal{J} = \{1, 2, \dots, n\}$  is a finite index set. Events in  $\mathcal{A}$  are said to be independent of each other provided that we have pairwise independence, that is, we have  $A_i$  being independent of  $A_j$  for all  $i, j \in \mathcal{J}$  that satisfy  $i \neq j$ . Equivalently, events in  $\mathcal{A}$  are said to be independent provided that we have mutual independence, that is the following holds:*

$$\mathbb{P}\left(\bigcap_{i=1}^n A_i\right) = \prod_{i=1}^n \mathbb{P}(A_i)$$

**Note:** Event  $A$  and event  $B$  are independent does not mean  $A \cap B = \emptyset$ .

**Note:** If event  $B$  satisfies  $\mathbb{P}(B) > 0$ , then from definition, we have:

$$\mathbb{P}(A \cap B) = \mathbb{P}(A|B) \cdot \mathbb{P}(B)$$

If in addition,  $A$  and  $B$  are independent, then we can write:

$$\mathbb{P}(A|B) = \mathbb{P}(A)$$