§1 Lecture 1

We begin with the definition of probability.

Definition 1.1 (Probability). A mathematical model about random experiments.

In this class, we fix the notation $\Omega, \mathcal{F}, \mathbb{P}$. We say that Ω is our "sample space" or the set of all possible outcomes. For example, if our experiment was flipping a coin, then $\Omega = \{H, T\}$. It is also important to note that Ω can be countably or uncountably infinite. For example, our experiment might be picking a positive integer in which $\Omega = \mathbb{N}$. Or our experiment might be throwing a dart at a board in which Ω would be uncountably infinite.

We define $\mathcal{F}$ to be the set of all events, mathematically speaking it's a σ algebra. We can define an event to be a subset of Ω. For example, if our experiment is flipping two coins, then an event could be that the first coin is heads - the precise event would then be $\{HH, HT\}$ (notice that both HH and HT are in Ω).

Another interesting example might be where we roll a die as our experiment. Then we could say an event is rolling an *even* number that's greater than 4. The precise event would be $\{4, 6\}$. Notice, however, that we can actually write this event as the intersection of two events. That is

1. Rolling an even number : $\{2, 4, 6\}$.
2. Rolling a number greater than 4 : $\{4, 5, 6\}$.

Then the intersection of these two events is exactly our event of rolling an even number greater than 4.

Remark 1.2. When we say one event A happens, it means we get some outcome $\omega \in \Omega$ s.t. $\omega \in A$. If $\omega \notin A$, then A *doesn't* happen.

Remark 1.3. We can also do set operations on events. That is if A and B are events, then we can say that the event of A or B happening is exactly $A \cup B$.

Definition 1.4. We define the complement of event A to be $A^c = \{\omega \in \Omega : \omega \notin A\}$.

Remark 1.5. If Ω is finite, then one possible choice of $\mathcal{F}$ is $\mathcal{F} = 2^{\Omega}$.

Note that 2^Ω is defined to be the *power set* of Ω, or the "set of all subsets of Ω".

One more thing to note, but not terribly important for this class is that if Ω is infinite, then $\mathcal{F} \neq 2^\Omega$.

Finally, we define $\mathbb{P}$ to be a function from $\mathcal{F}$ to the closed interval $[0, 1]$. That is, $\mathbb{P} : \mathcal{F} \to [0, 1]$. Notice that this means we can only talk about the probability of one event.

If we want to talk about the probability of some outcome ω, then formally we would define the event $\{\omega\}$ and ask what is $\mathbb{P}(\{\omega\})$.

Recall that we said that $\mathcal{F}$ is a "σ algebra". Here we define it formally.

Definition 1.6. Let Ω be a set. We say that $\mathcal{F}$ is a σ algebra of Ω if it satisfies:

1. $\emptyset \in \mathcal{F}$
2. If $A_1, A_2, \dots \in \mathcal{F}$, then $\bigcup_{i=1}^{\infty} A_i \in \mathcal{F}$
3. If $A \in \mathcal{F}$, then $A^c \in F$ (recall that A^c is the *complement* of A)

§2 Lecture 2

Remark 2.1. $\mathcal{F}$ is formulated by our choice but it's usually the set of all events. For example, we might have $\mathcal{F} = \{\emptyset, \Omega\}$ or $\mathcal{F} = \{\emptyset, \Omega, A, A^c\}$.

Proposition 2.2. Let $\mathcal{F}$ be a σ-algebra and $A_1, A_2, \dots \in \mathcal{F}$, then $\bigcap_{i=1}^{\infty} A_i \in \mathcal{F}$.

Proof. Since $A_1, A_2, \dots \in \mathcal{F}$, then by Definition 1.6 we have that $A_1^c, A_2^c, \dots \in \mathcal{F}$ and again by definition $\bigcup_{i=1}^{\infty} A_i^c \in \mathcal{F}$. So by definition once again, $(\bigcup_{i=1}^{\infty} A_i^c)^c \in \mathcal{F}$ (since the complement of any event is also in $\mathcal{F}$). But this is equivalent to $\bigcap_{i=1}^{\infty} A_i$. $\square$

We now formally define what $\mathbb{P}$ is.

Definition 2.3. We define $\mathbb{P}$ as a probability measure where $\mathbb{P}$ is a function $\mathbb{P} : \mathcal{F} \to [0, 1]$ such that:

1. $\mathbb{P}(\emptyset) = 0$.
2. $\mathbb{P}(\Omega) = 1$.
3. If $A_1, A_2, \dots \in \mathcal{F}$ and $A_i \cap A_j = \emptyset, \forall i \neq j$ (or pairwise disjoint), then

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

Remark 2.4. The triple $(\Omega, \mathcal{F}, \mathbb{P})$ is called a probability space. This also uniquely characterizes a random experiment and vice versa.

For example, if we roll a die the experiment may change depending on how we define our probability space. We could define $\Omega = \{1, 2, 3, 4, 5, 6\}$ and let $\mathcal{F} = 2^{\Omega}$ (recall that this is the power set of Ω). This then would simulate simply rolling the die and recording what number we get. However, if we were to define $\mathcal{F} = \{\emptyset, \Omega, \{1, 3, 5\}, \{2, 4, 6\}\}$, then this would simulate rolling the die as a fair coin!

Thus, it is important to always specify what $\Omega, \mathcal{F}$ and $\mathbb{P}$ are with any random experiment.

We now prove a few basic properties of probability spaces.

Proposition 2.5. Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space.

1. If $A \in \mathcal{F}$, then $\mathbb{P}(A^c) = 1 - \mathbb{P}(A)$.
2. If $A, B \in \mathcal{F}$ and $A \subseteq B$ then $\mathbb{P}(B) = \mathbb{P}(A) + \mathbb{P}(B \setminus A) \geq \mathbb{P}(A)$. Recall that $B \setminus A = \{\omega \in \Omega : \omega \in B \wedge \omega \notin A\}$.
3. If $A_1, A_2, \cdots, A_n \in \mathcal{F}$, then
$$\mathbb{P}(\bigcup_{i=1}^{n} A_i) = \sum_{i=1}^{n} \mathbb{P}(A_i) - \sum_{i<j} \mathbb{P}(A_i \cap A_j) + \sum_{i<k<j} \mathbb{P}(A_i \cap A_j \cap A_k) - \cdots + (-1)^{n-1} \mathbb{P}(A_1 \cap \cdots \cap A_n)$$
4. If $A_1 \subseteq A_2 \subseteq \ldots$ then $\lim_{n\to\infty} \mathbb{P}(A_n) = \mathbb{P}(\bigcup_{i=1}^{\infty} A_n)$.
5. If $A_1 \subseteq A_2 \subseteq \ldots$ then $\lim_{n\to\infty} \mathbb{P}(A_n) = \mathbb{P}(\bigcap_{i=1}^{\infty} A_n)$.

Proof.

1. Notice that A and A^c are disjoint and $A \cup A^c = \Omega$. So $1 = \mathbb{P}(\Omega) = \mathbb{P}(A \cup A^c) = \mathbb{P}(A) + \mathbb{P}(A^c)$.
2. Since $A \subseteq B$, then $B = A \cup (B \setminus A)$ and $\mathbb{P}(B) = \mathbb{P}(A) + \mathbb{P}(B \setminus A)$. Since $\mathbb{P}(B \setminus A) \geq 0$ by Definition 2.3, then $\mathbb{P}(B) = \mathbb{P}(A) + \mathbb{P}(B \setminus A) \geq \mathbb{P}(A)$.
3. We will show by induction but we will only show the base case and leave the inductive step as an exercise for the reader.

   For the base case or $n = 2$, notice that $A_1 \cup A_2 = A_2 \cup (A_1 \setminus A_2)$ and that $A_1 = (A_1 \cap A_2) \cup (A_1 \setminus A_2)$. So $\mathbb{P}(A_1 \cup A_2) = \mathbb{P}(A_2 + \mathbb{P}(A_1 \setminus A_2)$ and $\mathbb{P}(A_1) = \mathbb{P}(A_1 \cap A_2) \cup \mathbb{P}(A_1 \setminus A_2)$. Combining these two equations gives us our base case.

   Hint: For the inductive step, maybe we can treat $\mathbb{P}(A_1 \cup A_2 \cup A_3) = \mathbb{P}((A_1 \cup A_2) \cup A_3)$.
4. Define $B_1 = A_1, B_2 = A_2 \setminus A_1$, and $B_i = A_i \setminus A_{i-1}$. Notice that this means that the B_i's are pairwise disjoint and $\bigcup_{i=1}^{\infty} B_i = \bigcup_{i=1}^{\infty} A_i$. So
$$\mathbb{P}(\bigcup_{i=1}^{\infty} A_i) = \mathbb{P}(\bigcup_{i=1}^{\infty} B_i) = \sum_{i=1}^{\infty} \mathbb{P}(B_i) = \lim_{n\to\infty} \sum_{i=1}^{n} \mathbb{P}(B_i)$$

But notice that

$$
\begin{aligned}
\mathbb{P}(B_1) &= \mathbb{P}(A_1) \\
\mathbb{P}(B_2) &= \mathbb{P}(A_2 \setminus A_1) = \mathbb{P}(A_2) - \mathbb{P}(A_1) \\
&\vdots \\
\mathbb{P}(B_n) &= \mathbb{P}(A_n) - \mathbb{P}(A_{n-1})
\end{aligned}
$$

This gives us a *telescoping* sum and so $\sum_{i=1}^{n} \mathbb{P}(B_i) = \mathbb{P}(A_n)$ and $\lim_{n\to\infty} \sum_{i=1}^{n} \mathbb{P}(B_i) = \lim_{n\to\infty} \mathbb{P}(A_n)$.

5. Similar to (4).

□

We now begin our discussion on *conditional probability*.

Definition 2.6. Let A and B be two events with $\mathbb{P}(B) > 0$. Then the conditional probability of A given B is

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