Lecture 1

Probability Space

In this lecture pay a close attention to the following concepts:

- **Random experiment:** An experiment whose outcome is unpredictable. Or, if we repeatedly perform the same experiment under the same (similar) circumstances, the outcome does not necessarily turn out to be the same. Each repetition of an experiment is called a trial. For instance if we roll a standard die to observe which face shows up, then the experiment is random because which face shows up varies unpredictably from trial to trial.
- **Sample Space:** The collection of all possible outcomes of the random experiment, denoted by S. For example, the sample space associated with rolling a standard die is $S = \{1, 2, 3, 4, 5, 6\}$.
- **Event:** A description (expressed in words or symbols) of those outcomes of the random experiment that we are interested in. An event is always a subset of the sample space, denoted by the capital letters A, B, C, etc. For example, for the rolling a standard die experiment, and event is *"an even face will show up"*. We can also write this event as the subset $A = \{2, 4, 6\}$ of the sample space. $B = \{5, 6\}$ is another example of an event for the same random experiment.
- **Making more events:** By using the "$\cup$" (union) or "$\cap$" (intersection) or "^{c}" (complement) operations of sets, we can make more events from the same sample space. For example, for the roll of a standard die experiment if $A = \{2, 4, 6\}$ is an event then $A^c = \{1, 3, 5\}$ is another event.
- **Event occurs:** The statement "A occurs" means that one of the elements of A was the outcome when the random experiment was performed. For example, if $A = \{2, 4, 6\}$ occurs then it simply means that either the face '2' showed up or face '4' showed up or face '6' showed up. So, the event "$A \cup B$ occurred" would mean that the outcome of the experiment happened to be either in A or in B or in both. On the other hand, "$A \cap B$ occurred" would mean that the outcome of the experiment was in fact both in A and in B.

Using the above concepts our main goal is to assign a "likelihood" (also called probability) to the occurrences of various events. If A is an event then $\mathbb{P}(A)$ will

denote the likelihood of occurrence of A. A few techniques of assigning this likelihood are:

- (i) The empirical approach, (a verification procedure).
- (ii) The counting approach, (proposed by Cardano in 1560).
- (iii) The measuring lengths/areas/volumes approach, (analogs of Cardano's).
- (iv) The independence approach.

1.1 The Empirical Approach

A simple and intuitive approach that approximates the probability of an event occurs involves gaining some experience by performing the experiment over and over again and recording its frequency. This is called the empirical approach.

Consider the experiment of tossing four coins and observing the faces that show up on the four coins. We want to estimate the probability of observing exactly three heads. To estimate this probability, we just toss the four coins a large number of times (say 200 times) and see how often exactly three heads occur! When I actually performed this experiment 200 times, on 56 occasions I observed exactly three heads. According to the frequency interpretation of probability, our estimate of the likelihood of observing three heads from four coins is $\frac{56}{200} = 0.28$. This ratio is called a relative frequency and it approximates the probability of the event.

When we repeat the experiment more often, the relative frequencies tend to settle down. For instance, when I programmed my computer to perform the same experiment 1000 times, the relative frequency was $\frac{236}{1000} = 0.236$. After 100,000 repetitions, the relative frequency of the event was $\frac{24{,}887}{100{,}000} = 0.24887$, which seems to be converging towards 0.25. Figure 1.1 provides four simulation runs using N number of trials. Our relative frequencies did not remain the same (due to the random nature). A remarkable fact, however, is that our relative frequencies will converge to a value[1] for sure. This limiting value is called the probability of the event. The inconvenient aspect of this (empirical) approach is that we have to repeat the experiment a large number of times. Would it not be great if we could find the limiting value without performing the random experiment even once? Well, sometimes we can!

Example - 1.1.1 - **(3 heads in 4 tosses — another perspective)** The above empirical approach of assigning probability to an event requires large number of repetitions. Here is a simpler way. The sample space for the experiment of tossing four coins has 16 elements as listed below.

$$S = \left\{ \begin{array}{cccc} HHHH, & HHHT, & HHTH, & HTHH, \\ THHH, & HHTT, & HTHT, & THHT, \\ TTHH, & THTH, & HTTH, & TTTH, \\ TTHT, & THTT, & HTTT, & TTTT \end{array} \right\}.$$

[1]This is due to a result known as the Law of Large Numbers. One of the "self-cleansing" or self-rejuvenating aspect of this course is that we will build enough (math) tools to prove the law.

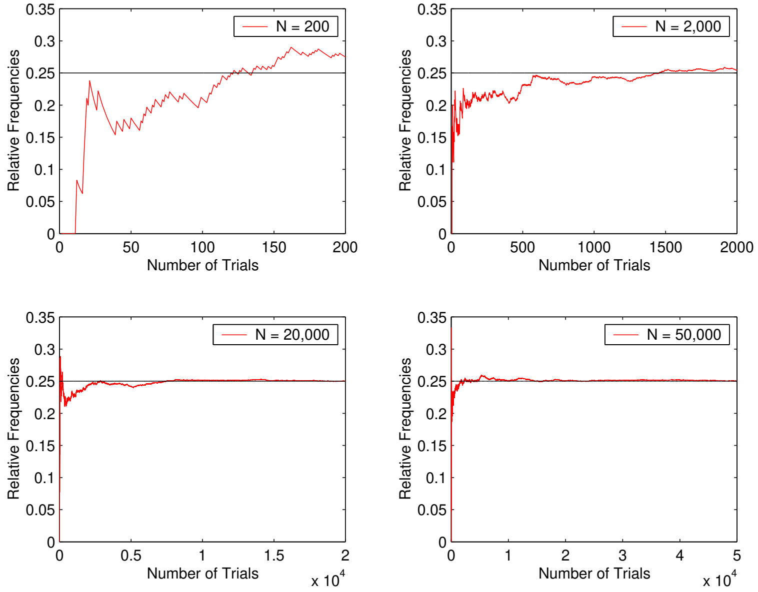

Figure 1.1: Relative Frequencies of 3 Heads in 4 Tosses.

The event, A, of observing exactly three heads in the four tosses is the following subset of S,

$$A = \{HHHT,\ HHTH,\ HTHH,\ THHH\}.$$

Just by shear luck we notice that the limiting value obtained by the empirical approach of the last example, namely 0.25, happens to be the same as

$$\mathbb{P}(A) \;=\; \frac{\text{number of elements of } A}{\text{number of elements of } S} \;=\; \frac{4}{16} \;=\; 0.25.$$

The issue is will the counting approach always give us the limiting value? Unfortunately, the answer is no! This counting approach works only sometimes. We will learn more about this in later lectures.

1.2 Axioms of Probability Theory

Regardless of how $\mathbb{P}(A)$ is calculated or approximated, it must follow the condition

$$\boxed{0 \le \mathbb{P}(A) \le 1} \tag{2.1}$$

since the probability is always some sort of a percentage and percentages lie between 0 and 1. Also, since the sample space, S, must always occur, we should always have

$$\boxed{\mathbb{P}(S) = 1.} \tag{2.2}$$

If we break up an event into disjoint pieces, the total probability should be the same as adding the individual pieces' probabilities. Symbolically, if the event E is broken up into disjoint parts, $A, B, C, \cdots$, then

$$\boxed{\mathbb{P}(E) = \mathbb{P}(A \cup B \cup C \cup \cdots) = \mathbb{P}(A) + \mathbb{P}(B) + \mathbb{P}(C) + \cdots} \tag{2.3}$$

The above three conditions constitute the official definition of a probability function $\mathbb{P}$ and form the fundamental axioms of probability theory. All thousands of books of probability theory that you see in the libraries are based on these three "innocent looking" axioms. It so happens that these axioms are not so innocent looking after all. They automatically give many more results. As an example the following theorem collects a few consequences of (2.1), (2.2), and (2.3).

Theorem - 1.2.1 - **(Workhorse)** *Let S be a sample space and let $\mathbb{P}$ be a probability function for the events of this sample space. If A, B are two events, then the following results hold:*

- *(i)* $\mathbb{P}(\emptyset) = 0$, *(impossibility rule)*
- *(ii)* $\mathbb{P}(A^c) = 1 - \mathbb{P}(A)$, *(complement rule)*
- *(iii)* $\mathbb{P}(A^c \cap B) = \mathbb{P}(B) - \mathbb{P}(A \cap B)$, *(general complement rule)*
- *(iv)* $\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B)$, *(union rule)*
- *(v) if $A \subseteq B$ then $\mathbb{P}(A) \le \mathbb{P}(B)$, (monotonicity).*

Proof: (i) Since $\emptyset \cup S = S$, and $\emptyset$ and S are disjoint, by axiom (2.3) of $\mathbb{P}$ we see that

$$\mathbb{P}(\emptyset) + \mathbb{P}(S) = \mathbb{P}(\emptyset \cup S) = \mathbb{P}(S) = 1,$$

the last statement from axiom (2.2) of $\mathbb{P}$. That is,

$$\mathbb{P}(\emptyset) + 1 = 1.$$

This gives $\mathbb{P}(\emptyset) = 0$. Now the reader should prove the rest of the parts. ♠

Remark - 1.2.1 - **(Sigma field)** A sharp observer should note that axioms (2.1), (2.2), and (2.3), and Theorem 1.2.1 require that the word "*event*" be defined a bit more carefully so that

(a) S should always be considered both as the sample space, and as an event.

(b) If A is an event then A^c should also be considered an event.

(c) If $A, B, C, \cdots$ are events then their union must also be considered as an event.

These requirements will be taken for granted through out this book. It will be helpful to imagine all the events to be put into a large box (i.e., a box full of events). This box full of events must obey the above conditions (a), (b), and (c). Any such box obeying these three conditions is called a sigma field of subsets of S. We will not emphasize this word in this book.

Example - 1.2.1 - If $\mathbb{P}(A \cup B) = 0.6$ and $\mathbb{P}(A \cup B^c) = 0.8$ find $\mathbb{P}(A)$.

Note that $(A \cup B) \cup (A \cup B^c) \supseteq B \cup B^c = S$, where S is the sample space. Also, note that $(A \cup B) \cap (A \cup B^c) = A$. Therefore, we have

$$\begin{aligned} 1 \;=\; \mathbb{P}(S) \;&=\; \mathbb{P}((A \cup B) \cup (A \cup B^c)) \\ &=\; \mathbb{P}(A \cup B) \;+\; \mathbb{P}(A \cup B^c) \;-\; \mathbb{P}((A \cup B) \cap (A \cup B^c)) \\ &=\; \mathbb{P}(A \cup B) \;+\; \mathbb{P}(A \cup B^c) \;-\; \mathbb{P}(A). \end{aligned}$$

Therefore, we see that

$$\mathbb{P}(A) \;=\; \mathbb{P}(A \cup B) \;+\; \mathbb{P}(A \cup B^c) \;-\; 1 \;=\; 0.6 + 0.8 - 1 \;=\; 0.4.$$

Example - 1.2.2 - Argue why

$$\mathbb{P}(A \cap B) \;\le\; \min\{\mathbb{P}(A), \mathbb{P}(B)\} \;\le\; \max\{\mathbb{P}(A), \mathbb{P}(B)\} \;\le\; \mathbb{P}(A \cup B).$$

The first inequality follows from the fact that $A \cap B \subseteq A$, as well as $A \cap B \subseteq B$. The monotonicity property of probability (c.f. Theorem 1.2.1, part (v)) gives that $\mathbb{P}(A \cap B) \le \mathbb{P}(A)$ as well as $\mathbb{P}(A \cap B) \le \mathbb{P}(B)$. Hence, $\mathbb{P}(A \cap B)$ is smaller (no larger) than both $\mathbb{P}(A)$ and $\mathbb{P}(B)$, giving the first inequality. The second inequality is trivial. The third inequality is a consequence of the facts that $A \subseteq A \cup B$, and $B \subseteq A \cup B$, and the monotonicity property of probability.

Example - 1.2.3 - A student assigned probabilities of events A, B as follows

$$\mathbb{P}(B) \;=\; 0.3, \qquad \mathbb{P}(A \cup B^c) \;=\; 0.6.$$

What is $\mathbb{P}(A)$?

Using properties (iv), (iii) and (ii) of the last theorem we see that

$$\begin{aligned} 0.6 \;=\; \mathbb{P}(A \cup B^c) \;&=\; \mathbb{P}(A) + \mathbb{P}(B^c) - \mathbb{P}(A \cap B^c), \qquad \text{by (iv)}, \\ &=\; \mathbb{P}(A) + \mathbb{P}(B^c) - \mathbb{P}(A) + \mathbb{P}(A \cap B), \qquad \text{by (iii)}, \\ &=\; \mathbb{P}(B^c) + \mathbb{P}(A \cap B) \;=\; (1 - 0.3) + \mathbb{P}(A \cap B), \qquad \text{by (ii)}. \end{aligned}$$

Therefore, $\mathbb{P}(A \cap B) = 0.6 - 0.7 = -0.1$. This contradicts axiom (2.1), meaning the assignment was flawed. This shows that one cannot assign probabilities to various events arbitrarily.

Remark - 1.2.2 - **(Summary)** In this lecture we have presented the definition of $\mathbb{P}(A)$, i.e., obeying three axioms, (2.1), (2.2), and (2.3). As their consequences these three axioms then led to a bunch of more axioms collected in Theorem 1.2.1. We noticed that the general probability statements that will always be true are

$$\mathbb{P}(S) = 1, \text{ (certainty rule)} \qquad \mathbb{P}(\emptyset) = 0, \text{ (impossibility rule)}.$$

What is still not fully settled is how does one go about and assign

$$\mathbb{P}(A) \text{ to any other } \underline{\text{arbitrary}} \text{ event } A?$$

This is a modeling issue since usually there are infinitely many ways such assignments can be made. The next three lectures present such assignment techniques that do not violate any of the axioms of probability theory.

1.3 Exercises

Exercise - 1.3.1 - *Use simulation to find an approximation of the probability of getting at least one double six in six rolls of a pair of fair dice.*

Exercise - 1.3.2 - *Prove parts (ii) through (v) of Theorem 1.2.1.*

Exercise - 1.3.3 - *Consider the experiment of tossing four fair coins and observing the outcome. List all the elements of the sample space. Give two examples of event spaces.*

Exercise - 1.3.4 - *For Exercise 1.3.3 define probability function $\mathbb{P}$ over the classes of events of your choice.*

Exercise - 1.3.5 - *Show that the probability of at least one of the three events, A, B, C, will occur is*

$$\mathbb{P}(A) + \mathbb{P}(B) + \mathbb{P}(C) - \mathbb{P}(A \cap B) - \mathbb{P}(A \cap C) - \mathbb{P}(B \cap C) + \mathbb{P}(A \cap B \cap C).$$

Exercise - 1.3.6 - *For any events $A_1, A_2, \cdots$, verify that*

$$1 - \sum_{i=1}^{\infty}(1 - \mathbb{P}(A_i)) \;\le\; \mathbb{P}\left(\cap_{i=1}^{\infty} A_i\right).$$

Exercise - 1.3.7 - *Let $S = \{a, b, c\}$ be a sample space with the power set as the class of events. Let $\mathbb{P}(\{a\}) = \frac{1}{4}$, $\mathbb{P}(\{b\}) = \frac{1}{3}$. Find $\mathbb{P}(\{c\})$ and $\mathbb{P}(\{a, c\})$.*

Exercise - 1.3.8 - *Let $S = \{a, b, c, d\}$ be a sample space such that $\mathbb{P}(\{a\}) = \mathbb{P}(\{b\}) = \frac{1}{4}$ and $\mathbb{P}(\{c\}) = 2\mathbb{P}(\{d\})$. Find the probability function $\mathbb{P}$.*

Exercise - 1.3.9 - *Is it possible to have an assignment of probabilities in some random experiment such that $\mathbb{P}(A) = \frac{1}{2}$, $\mathbb{P}(B) = \frac{1}{4}$ and $\mathbb{P}(A \cap B) = \frac{1}{3}$?*

Exercise - 1.3.10 - *What is the maximum possible value of $\mathbb{P}(A \cap B)$ when $\mathbb{P}(A)$ and $\mathbb{P}(B)$ are fixed?*

Exercise - 1.3.11 - *Prove that $\mathbb{P}(A \cap B) \ge \mathbb{P}(A) + \mathbb{P}(B) - 1$.*

Exercise - 1.3.12 - *If we know that $\mathbb{P}(A \cup B) = 2/3$ and $\mathbb{P}(A \cap B) = 1/3$, can we determine $\mathbb{P}(A)$ and $\mathbb{P}(B)$?*

Exercise - 1.3.13 - *Consider two events A and B such that $\mathbb{P}(A) = 1/3$ and $\mathbb{P}(B) = 1/2$. Determine the value of $\mathbb{P}(B \cap A^c)$ for each of the following conditions: (a) A and B are disjoint, (b) $A \subset B$, (c) $\mathbb{P}(A \cap B) = 1/8$.*

Exercise - 1.3.14 - *A die has been loaded so that the probability of a particular number coming up is proportional to that number. Compute (i) the probabilities of all the singleton events, (ii) the probability that an even number will occur, (iii) the probability that a number greater than 4 will occur.*