Lecture Notes 1: Probability and Statistics

Welcome to the course MSO-201 (Probability and Statistics). This is the first probability and statistics course at IIT Kanpur at the undergraduate level. I am sure all of you have seen exposed to some elementary probability and statistics at your high school. The main aim of this course is to introduce probability in a more formal way. You will see that the way the probability has been introduced to you has some limitations. First we will introduce probability in a more formal way which was introduced by the Russian probabilist Academician Kolmogorov. We will discuss about random variables, both discrete and continuous, transformation of random variables etc. The second part of this course deals with the modern tools of statistics. We will mainly discuss about population, sample, estimation of the unknown parameters of the population based on a sample, and we will discuss about different testing of hypothesis problems. There are several good books available in this topic. I had mentioned about two books in the first course handout (FCH). I have already mentioned about the grading schemes also in the FCH. I will be giving regular assignments and the only way to learn this course is to do these problems. If you have problems in solving those problems please discuss those in the tutorial session.

Let us remember the way the probability has been introduced usually in the high school level. Suppose we want to define the probability of an event A, then it is defined as

$$P(A) = \frac{\text{no. of favorable evenets in } A}{\text{total number of events}}.$$

Let us look at a given example. Suppose we toss a fair coin three times, then what is the probability that exactly three heads will appear. Clearly in this case

$$A = \{HHT, HTH, THH\}$$

and the possible cases are

$$\{HHH, HHT, HTH, HTT, THH, THT, THT, TTT\}.$$

Therefore,

$$P(A) = \frac{3}{8}.$$

Similarly, if we want is the probability that we will be observing a 'H' before a tail, then in this case

$$A = \{HHH, HHT, HTT, HTH\}$$

Hence,

$$P(A) = \frac{4}{8} = \frac{1}{2}.$$

Clearly, one problem we can see that this definition cannot be used if the total number of events is not finite or if the coin is biased. Hence, we need a more general definition of probability. Before defining the probability function (measure) we need to define certain quantities. First we will define what is a random experiment. A random experiment is an operation or an experiment which can produce different outcomes even it is being performed under the same environmental conditions. Some of the simple toy examples we can think of as random experiments are the following.

Example 1: Suppose we toss a coin 5 times. Clearly, even if it is being conducted under the same set of environmental conditions, we might get different results at different outcomes.

Example 2: Suppose we draw three cards without replacement from a well-shuffled deck of cards. It is clearly a random experiment.

Example 3: Suppose a box has three red balls, 5 green balls and two white balls. We draw three balls with replacement. It is a random experiment.

The all possible outcomes of a random experiment is called the **Sample Space**. We will denote the Sample Space as Ω. For example 1, Ω has 2^5 elements and Ω is of the form:

$$\Omega = \{(a_1 a_2 a_3 a_4 a_5); a_i = \text{H or T}\}.$$

Exercise: Find the sample spaces for Example 2 and Example 3.

Now we are ready to define the probability function (measure). The probability function is a set function defined on the subsets of Ω and it satisfies the following properties:

Properties:

(1) If $A \subset \Omega$, then $P(A) \in [0,1]$.

(2) $P(\Omega) = 1$.

(3) If $A_1, A_2, \ldots$ are disjoint subsets of Ω, then

$$P\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} P(A_i).$$

Therefore, any set function which satisfies the above three properties is called a probability function or a probability measure. It can be easily seen that the usual definition of the probabilities can be obtained as a special case of this. Let us look at some of the examples.

Example 1: Suppose we toss an unbiased coin three times. What is the probability that we will get at least two heads. Now in this case all possible outcomes are the following:

$$\Omega = \{HHH, HHT, HTH, HTT, THH, THT, TTH, TTT\}.$$

Now if A denotes the event that at least two head will appear, then

$$A = \{HHT, HHT, HTH, THH\}.$$

Hence, according to the classical definition of probability $P(A) = 4/8 = 1/2$. Now if we want to use the above definition, then we can define a probability function $P(\cdot)$ as follows:

$$P(HHH) = P(HHT) = P(HTH) = P(HTT) = \frac{1}{8},$$
$$P(THH) = P(THT) = P(TTH) = P(TTT) = \frac{1}{8},$$

and it can be extended for any subset of Ω uniquely, as it has to satisfy the above properties of a probability function. Now in this case also it can be easily seen that $P(A) = 1/2$.

Example 2: Suppose we want to address the same question as in Example 1, but here it is assumed that the coin is biased and $P(H) = 2/3$ and $P(T) = 1/3$. I cannot use the classical definition to answer this question. Now under the current definition we can define a new probability function as follows:

$$P(HHH) = \frac{8}{27}, \quad P(HHT) = P(HTH) = P(THH) = \frac{4}{27},$$
$$P(HTT) = P(THT) = P(TTH) = \frac{2}{27}, \quad P(TTT) = \frac{1}{27}.$$

It can be easily verified that it is a proper probability function, and in this case $P(A) = 20/27$.

Example 3: Suppose we toss an unbiased coin until we get the first head. What is the probability that we need at most 4 tosses. In this case

$$\Omega = \{H, TH, TTH, TTTH, TTTTH, \ldots\}$$

and

$$P(H) = \frac{1}{2}, P(TH) = \frac{1}{2^2}, \ldots, P(T \ldots TH) = \frac{1}{2^{k+1}} \text{ if there are k-tails before a head}, \ldots$$

We have

$$A = \{H, TH, TTH, TTTH\},$$

hence

$$P(A) = \frac{1}{2} + \frac{1}{2^2} + \frac{1}{2^3} + \frac{1}{2^4}.$$

In general if Ω has finite or countable number of points, we can easily define the probability function on any subsets of Ω as follows. Suppose

$$\Omega = \{\omega_1, \omega_2, \ldots\}$$

and $p_1, p_2, \ldots$ are non-negative real numbers such that $p_i \geq 0$, for $i = 1, 2, \ldots$, and

$$\sum_{i=1}^{\infty} p_i = 1.$$

We can define the probability function $P(\cdot)$ as follows:

$$P(\omega_1) = p_1, P(\omega_2) = p_2, \ldots,$$

and it is clear that it can be extended for any subset of Ω uniquely as follows. Suppose

$$A = \{\omega_{i_1}, \omega_{i_2}, \ldots\}$$

then

$$P(A) = \sum_{j=1}^{\infty} P(\omega_{i_j}) = \sum_{j=1}^{\infty} p_{i_j}.$$

It can be easily seen that it satisfies all the three properties of a probability function. Hence, this is the most general description of a probability function when the number of outcomes is countable.

Now the question is can we follow the same procedure when the number of elements in Ω is not countable. For example, suppose we want to choose a number at random between $[0, 1]$. In this case $\Omega = [0, 1]$, which is not a countable set. Unfortunately we cannot extend easily in this situation. We need to define some more quantities. We will do it in the next lecture note.