Lectures 1-2

In these lectures, we introduce the basic notions *random experiment, sample space, events* and *probability of event.*

We begin with a snap shot of the history of probability. Prehistory of probability theory is about the calculations of probabilities of outcomes of games of chances with dice or cards. One of the earliest book on this is *Liber de ludo aleae (The book on games of chances)* by Gerolamo Cardano (1501 - 1576) published posthumosly in 1663. In the Chapter 14 of this 15 page work, the first definition of classical probability is given which reads as:

So there is one general rule, that we should consider the whole circuit (all possible outcomes) and the number of those casts which represents in how many ways the favorable result can occur and compare the number to the rest of the circuit and according to that proportion should the mutual wagers be laid so that one may contend on equal terms.

To get a glimpse of how the definition of probability evolved over the years, let us look at the definitions given by Leibniz, De Moivre and Laplace. *The definition given by Gottfried Wilhelm Leibniz (1646–1716) in 1710: If a situation can lead to different advantageous results ruling out each other, the estimation of the expectation will be the sum of the possible advantages for the set of all these results, divided into the total number of results.*

Abraham de Moivre's (1667–1754) definition from the 1711 De Mensura Sortis: *If p is the number of chances by which a certain event may happen, and q is the number of chances by which it may fail, the happenings as much as the failings have their degree of probability; but if all the chances by which the event may happen or fail were equally easy, the probability of happening will be to the prob- ability of failing as p to q.*

The definition given in 1774 by Pierre– Simon Laplace (1749–1827), with whom the formal definition of classi cal probability is usually associated. In his first probability paper, Laplace states: *The probability of an event is the ratio of the number of cases favorable to it, to the number of possible cases, when there is nothing to make us believe that one case should occur rather than any other, so that these cases for us equally possible.*

The mathematical methods of probability is widely believed to began in the correspondence Fermat and Pascal on a question posed in 1654 by the Gambler Antoine Gombaud famously known as the problem of points. Statement of the problem of points is the following.

Suppose two players I and II stake equal money on being the first to win n points in a game in which the winner of each point is decided by the toss *of a fair coin, with head for I and tail for II. If the game is interrupted when I still lacks m points and II lacks l, how should the total stake be divided between them?*

Problem is been solved by both using multiple ways. The method which essentially opened the way into the mathematical methods of probability is the one using combinatorial arguments. In fact the method put the problem into a general frame work and used combinatorial techniques (here using Pascal's triangle) to solve it.

During the begining of 18th century, Jacob Bernoulli (in *Ars Conjectandi*) and de Moivre (in *Doctrines of chances*) put probability theory on a solid mathematical framework. In fact Bernoulli proved a version of Law of large numbers.

Second half of the 19th century saw another revolution in theory of probability through the birth of Statistical Mechanics (term coined by Gibbs). James Maxwell Clark used probability theory to explain kinetic theory of gases, according to which a gas consists of a large number of molecules moving about in an arbitrary fashion. Infact, Maxwell realized that one need to compute how molecules are distributed with respect to their speed, as opposed to the early kinetic theory by Clausius which is based on assuming that all molecules are with same speed. Maxwell used the concept of distribution function, a concept from probability and a gave a heuristic derivation of the velocity distribution of molecules. This was later refined by Boltzman to arrive at the Maxwell-Boltzman distribution.

In 1900, Hibert in his sixth problem proposed to put probability in the axiomatic framework. In 1919, von Mises first proposed an axiomatic treatment of probability theory through *Kollektivs*, mathematical abstraction representing an infinite series of independent trials. Andrei Kolmogorov in 1933 came up a more elegant and simple axiomatisation which we use today. Let us begin to understand probability theory starting from this axiomatic approach.

By a random experiment, we mean an experiment which has multiple outcomes and one don't know in advance which outcome is going to occur. We call this an experiment with 'random' outcome. We assume that the set of all possible outcomes of the experiment is known.

Definition 0.1 *Sample space of a random experiment is the set of all possible outcomes of the random experiment.*

Unless specified otherwise, we denote sample space by Ω.

Example 0.1 *Toss a coin and note down the face. This is a random experiment, since there are multiple outcomes and outcome is not known before the toss, in other words, outcome occur randomly. More over, the sample space is* $\{H, T\}$

Example 0.2 *Toss two coins and note down the number of heads obtained. Here* $\Omega = \{0, 1, 2\}$.

Example 0.3 *(Urn problem) Two balls say 'R' and 'B' are distributed 'at random' in three urns labeled* 1, 2, 3. *Here the order of occupancy in an urn is irrelvent. The sample space is*

$$\begin{aligned}\Omega = \{&(RB, -, -), (-, RB, -), (-, -, RB), (R, B, -), (R, -, B),\\ &(B, R, -), (B, -, R), (-, R, B), (-, B, R)\}.\end{aligned}$$

We will give examples with infinte number of outcomes.

Example 0.4 *A coin is tossed untill two heads or tails come in succession. The sample space is*

$$\Omega = \{HH, TT, HTT, THH, HTHH, THTT, \cdots\}.$$

Example 0.5 *Pick a point 'at random' from the interval* $(0, 1]$. *'At random' means there is no bias in picking any point. Sample space is* $(0, 1]$.

Definition 0.2 *(Event) Any subset of a sample space is said to be an event.*

Example 0.6 $\{H\}$ *is an event corresponding to the sample space in Example 0.1.*

Definition 0.3 *(mutually exclusive events) Two events* A, B *are said to be mutually exclusive if* $A \cap B = \emptyset$.

If A and B are mutually exclusive, then occurrence of A implies non occurrence of B and vice versa. Note that non occurrence of A need not imply occurrence of B, since in general $A^c \neq B$.

Example 0.7 *The events* $\{H\}$, $\{T\}$ *of the sample space in Example 0.1 are mutually exclusive. But the events* $\{H, T\}$, $\{T\}$ *are not mutually exclusive.*

Now we introduce the concept of probability of events (in other words probability measure). Intuitively probability quantifies the chance of the occurrence of an event. We say that an event has occurred, if the outcome belongs to the event. In general it is not possible to assign probabilities to all events from the sample space. For the experiment given in Example 0.5, it is not possible to assign probabilities to all subsets of $(0,\ 1]$, for example not possible to assign probability to a Vitali set[1]. So one need to restrict to a smaller class of subsets of the sample space. For the random experiment given in Example 0.5, it turns out that one can assign probability to each sub interval in $(0,\ 1]$ as its length. Therefore, one can assign probability to any finite union of intervals in $(0,\ 1]$, by representing the finite union of intervals as a finite disjoint union of intervals and assign the probability as the sum of the length of these disjoint intervals. In fact one can assign probability to any countable union intervals in $(0,\ 1]$ by preserving the desirable property "probability of a countable disjoint union is the sum of the probabilities". Also note that if one can assign probability to an event, then one can assign probability to its complement, since occurrence of the event is same as the non-occurrence of its complement. Thus one seek to define probability on those class of events which satisfies "closed under complementation" and "closed under countable union". This leads to the following special family of events where one can assign probabilities.

Definition 0.4 *A family of subsets $\mathcal{F}$ of a nonempty set Ω is said to be a σ-field if it satisfies the following.*
(i) $\Omega \in \mathcal{F}$
(ii) if $A \in \mathcal{F}$, *then* $A^c \in \mathcal{F}$
(iii) if $A_1, A_2, A_3, \cdots \in \mathcal{F}$, *then* $\bigcup_{n=1}^{\infty} A_n \in \mathcal{F}$.

Example 0.8 *Let Ω be a nonempty set. Define*

$$\mathcal{F}_0 \ = \ \{\emptyset,\ \Omega\},\ \mathcal{P}(\Omega) \ = \ \{A \,|\, A \subseteq \Omega\}\,.$$

Then $\mathcal{F}_0$, $\mathcal{P}(\Omega)$ are σ-fields. Moreover, if $\mathcal{F}$ is a σ-field of subsets of Ω, then

$$\mathcal{F}_0 \ \subseteq \ \mathcal{F} \ \subseteq \ \mathcal{P}(\Omega)\,,$$

i.e., $\mathcal{F}_0$ is the smallest and $\mathcal{P}(\Omega)$ is the largest σ-field of subsets of Ω.

[1]Vitali set is a subset of $\mathbb{R}$ which is not (Lebesque) measurable, named after its inventor Giuseppe Vitali. Curious to know how it is defined, consult "Real Analysis' by Royden

Example 0.9 *Let Ω be a nonempty set and $A \subseteq \Omega$. Define*

$$\sigma(A) \;=\; \{\emptyset,\ A,\ A^c,\ \Omega\}\,.$$

Then $\sigma(A)$ is a σ-field and is the smallest σ-field containing the set A. $\sigma(A)$ is called the σ-field generated by A.

Lemma 0.1 *Let I be an index set and $\{\mathcal{F}_i \,|\, i \in I\}$ be a family of σ-fields of subsets of Ω. Then $\mathcal{F} = \cap_{i\in I}\mathcal{F}_i$ is a σ-field.*

Proof. Since $\Omega \in \mathcal{F}_i$ for all $i \in I$, we have $\Omega \in \mathcal{F}$. Now,

$$\begin{aligned} A \in \mathcal{F} \;&\Rightarrow\; A^c \in \mathcal{F}_i \text{ for all } i \in I \\ &\Rightarrow\; A^c \in \mathcal{F}\,. \end{aligned}$$

Similarly it follows that

$$\begin{aligned} A_1, A_2 \cdots \in \mathcal{F} \;&\Rightarrow\; A_1, A_2, \cdots \in \mathcal{F}_i \text{ for all } i \in I \\ &\Rightarrow\; \cup_{n=1}^{\infty} A_n \in \mathcal{F}_i,\ \text{ for all } i \\ &\Rightarrow\; \cup_{n=1}^{\infty} A_n \in \mathcal{F}\,. \end{aligned}$$

Hence $\mathcal{F}$ is a σ-field.

Example 0.10 *Let $\mathcal{A} \subseteq \mathcal{P}(\Omega)$. Then*

$$\cap\{\mathcal{F} \,|\, \mathcal{F} \text{ is a } \sigma - \text{field and } \mathcal{A} \subseteq \mathcal{F}\}$$

is a σ-field and is the smallest σ-field containing $\mathcal{A}$. We denote it by $\sigma(\mathcal{A})$

This can be seen as follows. From Lemma 0.1, $\sigma(\mathcal{A})$ is a σ-field. From the definition of $\sigma(\mathcal{A})$, it follows that $\mathcal{A} \subseteq \sigma(\mathcal{A})$. If $\mathcal{F}$ is a σ-field containing $\mathcal{A}$, then $\sigma(\mathcal{A}) \subseteq \mathcal{F}$. Hence, $\sigma(\mathcal{A})$ is the smallest σ-field containing $\mathcal{A}$.

The following notion of field is useful for constructing probabilities of events as we will be describing in the last part of this chapter.

Definition 0.5 *A family $\mathcal{F}$ of subsets of a non empty set Ω is said to be a field if $\mathcal{F}$ satisfies*

(i) $\Omega \in \mathcal{F}$

(ii) if $A \in \mathcal{F}$, then $A^c \in \mathcal{F}$

(iii) if $A_1,\, A_2 \in \mathcal{F}$, then $A_1 \cup A_2 \in \mathcal{F}$.

Example 0.11 *Any σ-field is a field. In particular, $\mathcal{F}_0, \sigma(A), \mathcal{P}(\Omega)$ are fields.*

Example 0.12 *Let* $\Omega = \{1, 2, \dots\}$. *Define*

$$\mathcal{F} = \{A \subseteq \Omega \,|\, \text{either } A \text{ is finite or } A^c \text{ is finite } \}.$$

Then $\mathcal{F}$ *is a field but not a* σ*-field.*

Note that (i) and (ii) in the definition of field follows easily. To see (iii), for $A_1, A_2 \in \mathcal{F}$*, if both* A_1, A_2 *are finite so is* $A_1 \cup A_2$ *and if either* A_1 *or* A_2 *is not finite, then* $(A_1 \cup A_2)^c$ *is finite. Hence (iii) follows. i.e.,* $\mathcal{F}$ *is a field.*

To see that $\mathcal{F}$ *is not a* σ*-field, take*

$$A_n = \{2n+1\}, \; n = 1, 2, \dots .$$

Now

$$\cup_{n=1}^{\infty} A_n = \{3, 5, \dots\} \notin \mathcal{F}.$$

Definition 0.6 *(Probability measure)*

Let Ω *be a nonempty set and* $\mathcal{F}$ *be a* σ*-field of subsets of* Ω*. A map* $P : \mathcal{F} \to [0,\ 1]$ *is said to be a probability measure if* P *satisfies*

(i) (Total mass 1) $P(\Omega) = 1$,

(ii) (Countable additivity) if $A_1, A_2, \dots \in \mathcal{F}$ *and are pairwise disjoint, then*

$$P\Big(\cup_{n=1}^{\infty} A_n \Big) = \sum_{n=1}^{\infty} P(A_n).$$

Definition 0.7 *(Probability space).*

The triplet $(\Omega, \mathcal{F}, P)$*; where* Ω *is a nonempty set (sample space),* $\mathcal{F}$ *is a* σ*-field and* P *is a probability measure; is said to be a probability space.*