Lecture 2 : Basics of Probability Theory

When an experiment is performed, the realization of the experiment is an outcome in the sample space. If the experiment is performed a number of times, different outcomes may occur each time or some outcomes may repeat. This "frequency of occurrence" of an outcome can be thought of as a probability. More probable outcomes occur more frequently. If the outcomes of an experiment can be described probabilistically, we are on our way to analyzing the experiment statistically.

1 Axiomatic Foundations

Definition 1.1 *A collection of subsets of S is called a sigma algebra (or Borel field), denoted by $\mathbb{B}$, if it satisfied the following three properties:*

a. $\emptyset \in \mathbb{B}$ *(the empty set is an element of* $\mathbb{B}$*).*

b. If $A \in \mathbb{B}$*, then* $A^c \in \mathbb{B}$ *(*$\mathbb{B}$ *is closed under complementation).*

c. If $A_1, A_2, \ldots \in \mathbb{B}$*, then* $\cup_{i=1}^{\infty} A_i \in \mathbb{B}$ *(*$\mathbb{B}$ *is closed under countable unions).*

Example 1.1 *(Sigma algebra-I) If S is finite or countable, then these technicalities really do not arise, for we define for a given sample space S,*

$$\mathbb{B} = \{\textit{all subsets of } S\textit{, including } S \textit{ itself}\}.$$

If S has n elements, there are 2^n sets in $\mathbb{B}$. For example, if $S = \{1, 2, 3\}$, then $\mathbb{B}$ is the following collection of $2^3 = 8$ sets: $\{1\}$, $\{1, 2\}$, $\{1, 2, 3\}$, $\{2\}$, $\{1, 3\}$, $\emptyset$, $\{3\}$, $\{2, 3\}$.

Example 1.2 *(Sigma algebra-II) Let $S = (-\infty, \infty)$, the real line. Then $\mathbb{B}$ is chosen to contain all sets of the form*

$$[a, b], \quad (a, b], \quad (a, b), \quad [a, b)$$

for all real numbers a and b. Also, from the properties of $\mathbb{B}$, it follows that $\mathbb{B}$ contains all sets that can be formed by taking (possibly countably infinite) unions and interactions of sets of the above varieties.

Definition 1.2 *Given a sample space S and an associated sigma algebra $\mathbb{B}$, a probability function is a function P with domain $\mathbb{B}$ that satisfies*

1. *$P(A) \geq 0$ for all $A \in \mathbb{B}$.*
2. *$P(S) = 1$.*
3. *If $A_1, A_2, \ldots \in \mathbb{B}$ are pairwise disjoint, then $P(\cup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} P(A_i)$.*

The three properties given in the above definition are usually referred to as the Axioms of Probability or the Kolmogorov Axioms. Any function P that satisfies the Axioms of Probability is called a probability function.

The following gives a common method of defining a legitimate probability function.

Theorem 1.1 *Let $S = \{s1, \ldots, s_n\}$ be a finite set. Let $\mathbb{B}$ be any sigma algebra of subsets of S. Let $p_1, \ldots, p_n$ be nonnegative numbers that sum to 1. For any $A \in \mathbb{B}$, define $P(A)$ by*

$$P(A) = \sum_{\{i: s_i \in A\}} p_i.$$

(The sum over an empty set is defined to be 0.) Then P is a probability function on $\mathbb{B}$. This remains true if $S = \{s_1, s_2, \ldots\}$ is a countable set.

Proof: We will give the proof for finite S. For any $A \in \mathbb{B}$, $P(A) = \sum_{i: s_i \in A} p_i \geq 0$, because every $p_i \geq 0$. Thus, Axiom 1 is true. Now,

$$P(S) = \sum_{i: s_i \in S} p_i = \sum_{i=1}^{n} p_i = 1.$$

Thus, Axiom 2 is true. Let $A_1, \ldots, A_k$ denote pairwise disjoint events. ($\mathbb{B}$ contains only a finite number of sets, so we need consider only finite disjoint unions.) Then,

$$P(\cup_{i=1}^{k} A_i) = \sum_{\{j: s_j \in \cup_{i=1}^{k} A_i\}} p_j = i \sum_{i=1}^{k} \sum_{i\{j: s_j \in A_i\}} p_j = \sum_{i=1}^{k} P(A_i).$$

The first and third equalities are true by the definition of $P(A)$. The disjointedness of the A_i's ensures that the second equality is true, because the same p_j's appear exactly once on each side of the equality. Thus, Axiom 3 is true and Kolmogorov's Axioms are satisfied. $\square$

Example 1.3 *(Defining probabilities-II) The game of darts is played by throwing a dart at a board and receiving a score corresponding to the number assigned to the region in which the dart lands. For a novice player, it seems reasonable to assume that the probability of the dart hitting a particular region is proportional to the area of the region. Thus, a bigger region has a higher probability of being hit.*

The dart board has radius r and the distance between rings is $r/5$. If we make the assumption that the board is always hit, then we have

$$P(\text{scoring } i \text{ points}) = \frac{\text{Area of region } i}{\text{Area of dart board}}.$$

For example,

$$P(scoring1point) = \frac{\pi r^2 - \pi(4r/5)^2}{\pi r^2} = 1 - (\frac{4}{5})^2.$$

It is easy to derive the general formula, and we find that

$$P(\text{scoring } i \text{ points}) = \frac{(6-i)^2 - (5-i)^2}{5^2}, \quad i = 1, \ldots, 5,$$

independent of π and r. The sum of the areas of the disjoint regions equals the area of the dart board. Thus, the probabilities that have been assigned to the five outcomes sum to 1, and, by Theorem 1.2.6, this is a probability function.

2 The calculus of Probabilities

Theorem 2.1 *If P is a probability function and A is any set in $\mathbb{B}$, then*

a. $P(\emptyset) = 0$, *where* $\emptyset$ *is the empty set.*

b. $P(A) \leq 1$.

c. $P(A^c) = 1 - P(A)$.

Theorem 2.2 *If P is a probability function and A and B are any sets in $\mathbb{B}$, then*

a. $P(B \cap A^c) = P(B) - P(A \cap B)$.

b. $P(A \cup B) = P(A) + P(B) - P(A \cap B)$.

c. *If* $A \subset B$, *then* $P(A) \leq P(B)$.

Formula (b) of Theorem 2.2 gives a useful inequality for the probability of an intersection. Since $P(A \cup B) \leq 1$, we have

$$P(A \cap B) = P(A) + P(B) - 1.$$

This inequality is a special case of what is known as *Bonferroni's inequality.*

Theorem 2.3 *If P is a probability function, then*

a. $P(A) = \sum_{i=1}^{\infty} P(A \cap C_i)$ *for any partition* $C_1, C_2, \ldots$;

b. $P(\cup_{i=1}^{\infty} A_i) \leq \sum_{i=1}^{\infty} P(A_i)$ *for any sets* $A_1, A_2, \ldots$. *(Boole's Inequality)*

PROOF: Since C_1, $C_2, \ldots$ form a partition, we have that $C_i \cap C_j = \emptyset$ for all $i \neq j$, and $S = \cup_{i=1}^{\infty} C_i$. Hence,

$$A = A \cap S = A \cap (\cup_{i=1}^{\infty} C_i) = \cup_{i=1}^{\infty} (A \cap C_i),$$

where the last equality follows from the Distributive Law. We therefore have

$$P(A) = P(\cup_{i=1}^{\infty} (A \cap C_i)).$$

Now, since the C_i are disjoint, the sets $A \cap C_i$ are also disjoint, and from the properties of a probability function we have

$$P(\sum_{i=1}^{\infty} (A \cap C_i)) = \sum_{i=1}^{\infty} P(A \cap C_i)s.$$

To establish (b) we first construct a disjoint collection $A_1^*, A_2^*, \ldots$, with the property that $\cup_{i=1}^{\infty} A_i^* = \cup_{i=1}^{\infty} A_i$. We define A_i^* by

$$A_1^* = A_1, \quad A_i^* = A_i \setminus (\sum_{j=1}^{i-1} A_j), \quad i = 2, 3, \ldots,$$

where the notation $A \setminus B$ denotes the part of A that does not intersect with B. In more familiar symbols, $A \setminus B = A \cap B^c$. It should be easy to see that $\cup_{i=1}^{\infty} A_i^* = \cup_{i=1}^{\infty} A_i$, and we therefore have

$$P(\cup_{i=1}^{\infty} A_i) = P(\cup_{i=1}^{\infty} A_i^*) = \sum_{i=1}^{\infty} P(A_i^*),$$

where the equality follows since the A_i^* are disjoint. To see this, we write

$$\begin{aligned} A_i^* \cap A_k^* &= \left\{A_i \setminus (\cup_{j=1}^{i-1} A_j)\right\} \cap \left\{A_k \setminus (\cup_{j=1}^{k-1} A_j)\right\} \quad \text{(definition of } A_i^*) \\ &= \left\{A_i \cap (\cup_{j=1}^{i-1} A_j)^c\right\} \cap \left\{A_k \cap (\cup_{j=1}^{k-1} A_j)^c\right\} \quad \text{(definition of } \setminus) \\ &= \left\{A_i \cap \bigcap_{j=1}^{i-1} A_j^c\right\} \cap \left\{A_k \cap \bigcap_{j=1}^{k-1} A_j^c\right\} \quad (DeMorgan'sLaws) \end{aligned}$$

Now if $i > k$, the first intersection above will be contained in the set A_k^c, which will have an empty intersection with A_k. If $k > i$, the argument is similar. Further, by construction $A_i^* \subset A_i$, so $P(A_i^*) \leq P(A_i)$ and we have

$$\sum_{i=1}^{\infty} P(A_i^*) \leq \sum_{i=1}^{\infty} P(A_i),$$

establishing (b). $\square$

There is a similarity between Boole's Inequality and Bonferroni's Inequality. If we apply Boole's Inequality to A^c, we have

$$P(\cup_{i=1}^n A_i^c) \leq \sum_{i=1}^{n} P(A_i^c),$$

and using the facts that $\cup A_i^c = (\cap A_i)^c$ and $P(A_i^c) = 1 - P(A_i)$, we obtain

$$1 - P(\cap_{i=1}^n A_i) \leq n - \sum_{i=1}^{n} P(A_i).$$

This becomes, on rearranging terms,

$$P(\cap_{i=1}^n A_i) \geq \sum_{i=1}^{n} P(A_i) - (n-1),$$

which is a more general version of the Bonferroni Inequality.

3 Counting

Methods of counting are often used in order to construct probability assignments on finite sample spaces, although they can be used to answer other questions also. The following theorem is sometimes known as the Fundamental Theorem of Counting.

Theorem 3.1 *If a job consists of k separate tasks, the i^{th} of which can be done in n_i ways, $i = 1, \ldots, k$, then the entire job can be done in $n_1 \times n_2 \times \cdots \times n_k$ ways.*

Example 3.1 *For a number of years the New York state lottery operated according to the following scheme. From the numbers 1,2, ..., 44, a person may pick any six for her ticket. The winning number is then decided by randomly selecting six numbers from the forty-four. So the first number can be chosen in 44 ways, and the second number in 43 ways, making a total of $44 \times 43 = 1892$ ways of choosing the first two numbers. However, if a person is allowed to choose the same number twice, then the first two numbers can be chosen in $44 \times 44 = 1936$ ways.*

The above example makes a distinction between counting with replacement and counting without replacement. The second crucial element in counting is whether or not the ordering of the tasks is important. Taking all of these considerations into account, we can construct a 2×2 table of possibilities.

Number of possible arrangements of size r from n objects

	Without replacement	With replacement
Ordered	$\frac{n!}{(n-r)!}$	n^r
Unordered	$\binom{n}{r}$	$\binom{n+r-1}{r}$

Let us consider counting all of the possible lottery tickets under each of these four cases.

Ordered, without replacement From the Fundamental Theorem of Counting, there are

$$44 \times 43 \times 42 \times 41 \times 40 \times 39 = \frac{44!}{38!} = 5,082,517,440$$

possible tickets.

Ordered, with replacement Since each number can now be selected in 44 ways, there are

$$44 \times 44 \times 44 \times 44 \times 44 \times 44 = 44^6 = 7,256,313,856$$

possible tickets.

Unordered, without replacement From the Fundamental Theorem, six numbers can be arranged in 6! ways, so the total number of unordered tickets is

$$\frac{44 \times 43 \times 42 \times 41 \times 40 \times 39}{6 \times 5 \times 4 \times 3 \times 2 \times 1} = \frac{44!}{6!38!} = 7,059,052.$$

Unordered, with replacement In this case, the total number of unordered tickets is

$$\frac{44 \times 45 \times 46 \times 47 \times 48 \times 49}{6 \times 5 \times 4 \times 3 \times 2 \times 1} = \frac{49!}{6!43!} = 13,983,816.$$

4 Enumerating outcomes

The counting techniques of the previous section are useful when the sample space S is a finite set and all the outcomes in S are equally likely. Then probabilities of events can be calculated by simply counting the number of outcomes in the event. Suppose that $S = \{s_1, \ldots, s_N\}$ is a finite sample space. Saying that all the outcomes are equally likely means that $P(\{s_i\}) = 1?N$ for every outcome s_i. Then, we have, for any event A,

$$P(A) = \sum_{s_i \in A} P(\{s_i\}) = \sum_{s_i \in A} \frac{1}{N} = \frac{\text{\# of elements in } A}{\text{\# of elements in } S}.$$

Example 4.1 *Consider choosing a five-card poker hand from a standard deck of 52 playing cards. What is the probability of having four aces? If we specify that four of the cards are aces, then there are 48 different ways of specifying the fifth card. Thus,*

$$P(\textit{four aces}) = \frac{48}{\binom{52}{5}} = \frac{48}{2,598,960}.$$

The probability of having four of a kind is

$$P(\textit{four of a kind}) = \frac{13 \times 48}{\binom{52}{5}} = \frac{624}{2,598,960}.$$

The probability of having exactly one pair is

$$P(\textit{exactly one pair}) = \frac{13\binom{4}{2}\binom{12}{3}4^3}{\binom{52}{5}} = \frac{1,098,240}{2,598,960}.$$