2. Probability

2.1 Concept of Probability

Consider an experiment where the outcomes are random. Examples of these types of experiments include the roll of a die, a coin toss, measurement of the voltage drop across a resistor etc. Thee outcomes may be discrete as is the case for a coin toss (heads or tails) or roll of a die (face showing 1, 2, 3, 4, 5, or 6) or continuous as is the case with measurement of the voltage across a resistor (voltage lying between 1 and 1.2 V or 1.0 and 1.001 V etc.). All the possible outcomes of any one of these experiments form a sample space with the outcomes being identified as elements of the sample space. We can refer to the outcomes of the experiments as events belonging to the sample space. For example, in the roll of a die, observation of a 5 is an event. In the experiment where voltage developed across a resistor is measured, an event may be the observation of a voltage between 3V and 4V. The reasons for the choice of a range of values in the case of a voltage will become obvious as we go a discussion of random variables.

For each of the events in the sample space, we may be able to provide in a quantitative fashion the likelihood of occurrence of that particular experiment. Consider for example that we roll a die a number of times. We note down how many times a 5 was observed. If this experiment, i.e., the roll of a die was repeated N times and a 5 was observed m times, we can safely say that the frequency \( f_5 \) of a 5 occurring is given by

\[
f_5 = \frac{m}{N} \quad (2.1)
\]

If we now assume that the number of times the experiment is repeated is increased to infinity, we say that the probability, \( p \) of a 5 showing up during the roll of a die can be expressed as

\[
p = \frac{m}{N} , N \rightarrow \infty \quad (2.2)
\]

If we identify an event by A, the probability of the event A occurring is written as \( P(A) \). Another interesting observation can be made if we examine the experiment of a coin toss. Let A define the event of observing heads, and B define the event of observing tails. In this case the sample space S contains only two events, A and B. If out of the N tosses, we observe heads \( h \) times, the probabilities of the events A and B can be written as

\[
P(A) = \frac{h}{N} ; \quad P(B) = \frac{N-h}{N} = 1 - \frac{h}{N} \quad (2.3),
\]

making

\[
P(A) + P(B) = 1 \quad (2.4).
\]
The outcomes that are not part of the event can be identified as belonging to the complimentary event, \( A^c \). In the case of a coin toss, \( B = A^c \). This is illustrated in Figure 2.1.

![Figure 2.1](image)

Figure 2.1 The event \( A \) and its complimentary event \( A^c \) are shown. The rectangular box contains the sample space consisting of all the possible outcomes.

Based on the discussion in the previous paragraph, we can now state the axioms of probability:

1. Probability of any event is always a nonnegative number, i.e., \( P(A) \geq 0 \)
2. Probability of all events in the sample space, \( P(S) = 1 \).
3. If the sample space contains a number of events (\( A_1, A_2, A_3, \ldots \)) that are mutually exclusive, i.e., the occurrence of one event precludes the occurrence of the others (example: roll of a dice), the probability of observing the outcomes of (for example) \( A_1, A_2, \) and \( A_3 \) will be

\[
P(A_1 + A_2 + A_3) = P(A_1) + P(A_2) + P(A_3)
\]  

(2.5)

Examples:

1. Two coins are tossed simultaneously. What is the sample space?

Answer: \{HH\} \{HT\} \{TH\} \{TT\}

2. Two fair dice are rolled. What is the probability of observing 8?

Answer: The sample space for this experiment consists of 36 possibilities and is given below:

\[
\begin{align*}
(1\ 1) & \quad (1\ 2) & \quad (1\ 3) & \quad (1\ 4) & \quad (1\ 5) & \quad (1\ 6) \\
(2\ 1) & \quad (2\ 2) & \quad (2\ 3) & \quad (2\ 4) & \quad (2\ 5) & \quad (2\ 6) \\
(3\ 1) & \quad (3\ 2) & \quad (3\ 3) & \quad (3\ 4) & \quad (3\ 5) & \quad (3\ 6) \\
(4\ 1) & \quad (4\ 2) & \quad (4\ 3) & \quad (4\ 4) & \quad (4\ 5) & \quad (4\ 6) \\
(5\ 1) & \quad (5\ 2) & \quad (5\ 3) & \quad (5\ 4) & \quad (5\ 5) & \quad (5\ 6) \\
(6\ 1) & \quad (6\ 2) & \quad (6\ 3) & \quad (6\ 4) & \quad (6\ 5) & \quad (6\ 6)
\end{align*}
\]
Of these 36 possibilities, only 5 can lead to 8. Hence the probability is given by 5/36.

3. Numbers 1 thro' 6 are given. Find the number of ways of picking two numbers (a) without replacement where order does not matter, (b) without replacement where order does matter (c) repetition is allowed where order does matter.

(a)  
\[
\begin{array}{cccccc}
12 & 23 & 34 & 45 & 56 \\
13 & 24 & 35 & 46 \\
14 & 25 & 36 \\
15 & 26 \\
16
\end{array}
\]
Fifteen ways \( C^N_k \) \( N=6; k=2; \) \( \frac{N!}{k!(N-k)!} = 15 \)

(b)  
\[
\begin{array}{cccccc}
12 & 21 & 31 & 41 & 51 & 61 \\
13 & 23 & 32 & 42 & 52 & 62 \\
14 & 24 & 34 & 43 & 53 & 63 \\
15 & 25 & 35 & 45 & 54 & 64 \\
16 & 26 & 36 & 46 & 56 & 65
\end{array}
\]
Thirty Ways \( P^N_k \); \( \frac{N!}{(N-k)!} = 30 \) Permutations

(c)  
\[
\begin{array}{cccccc}
11 & 21 & 31 & 41 & 51 & 61 \\
12 & 22 & 32 & 42 & 52 & 62 \\
13 & 23 & 33 & 43 & 53 & 63 \\
14 & 24 & 34 & 44 & 54 & 64 \\
15 & 25 & 35 & 45 & 55 & 65 \\
16 & 26 & 36 & 46 & 56 & 66
\end{array}
\]
Thirtysix Ways \( N^k \); \( 6^2 = 36 \)

4. Of the 5000 computer chips manufactured, 12 turned out to be defective. Based on this observation, estimate the probability of failure of chips manufactured at this plant.

The probability of failure is 12/5000.
2.2 Joint Probability

Consider a case of two experiments going on at the same time. For example, we are tossing a coin and rolling a die simultaneously. Let the event A be identified as the observation of \{\text{heads}\} for the coin and B be identified as the observation of \{4\} for the die. We can now create a new event C as the joint observation of A and B,

$$\{C\} = \{A, B\} = \{A \cap B\} = \{\text{heads, 4}\} = \{4, \text{heads}\}$$

The probability of the event \{C\} is referred to as the joint probability of A and B. We can now extend the concept to a number of experiments or outcomes. If we have a number of events \(A_1, A_2, A_3, \ldots, A_N\), we can express the joint probability of all these events as

$$P(A_1, A_2, A_3, \ldots, A_N).$$

Two events, A and B, are said to be statistically independent if the joint probability \(P(A, B)\) can be expressed as the product of the two marginal probabilities, \(P(A)\) and \(P(B)\),

$$P(A, B) = P(A \cap B) = P(A)P(B).$$

In general, if the events \(A_1, A_2, A_3, \ldots, A_N\) are statistically independent, their joint probability can be expressed as the product of the marginal probabilities,

$$P(A_1, A_2, A_3, \ldots, A_N) = P(A_1)P(A_2)P(A_3)\ldots P(A_N).$$

Two events, A, and B, are said to be mutually exclusive, if the joint probability is zero, i.e.,
\[ P(A, B) = 0 \; ; \; A \text{ and } B \text{ mutually exclusive.} \]

Note that \( P(A, B) = P(B, A) = P(AB) = P(BA) \).

The total probability, \( P(A+B) \) is the probability that either \( A \), \( B \) both occur, and is given by

\[ P(A \cup B) = P(A + B) = P(A) + P(B) - P(AB). \quad (2.9) \]

If \( A \) and \( B \) are mutually exclusive,

\[ P(A + B) = P(A \cup B) = P(A) + P(B) \quad (2.10) \]

and if \( A \) and \( B \) are independent

\[ P(A + B) = P(A) + P(B) - P(A)P(B) \quad (2.11) \]

**Example**

If we assume that we have a fair coin, i.e., \( P\{\text{heads}\} = P\{\text{tails}\} = 0.5 \) and we have a fair or unbiased die, i.e., \( P\{4\} = P\{3\} = (1/6) \) and so on, in a joint experiment the probability of observing *heads and a 4* is \( 0.5 \times (1/6) = (1/12) \).

2.3 Conditional Probability and Bayes’ Rule

In some of the experiments, the likelihood of an event depends (A) on another event (B). This will require that we find the conditional probability of one of the events given that the other event A has taken place. For example, consider that we have two manufacturers (I) and (II) supplying processors for computers. Supplier (I) manufactures processors with a very degree of tolerance and only 2% of the processors fail. Supplier (II) manufactures processors with a tolerance such that 3% of the processors fail. The question now is the following. If we pick a computer at random, what is the likelihood that it breaks down due to a faulty processor?

If we the event \( F \) is processor, we are given that \( P(F \text{ given that we have a processor from I}) = 0.02 \) and \( P(F \text{ given that we have a processor from I}) = 0.03 \). These are referred to as conditional probabilities, \( P(F|I) \) and \( P(F|II) \) respectively.

In general, for any two events \( A \) and \( B \), we can express the conditional probability \( P(A|B) \) as
\[ P(A \mid B) = \frac{P(AB)}{P(B)}. \]  \hfill (2.12)

Rewriting,
\[ P(AB) = P(A \mid B)P(B). \]  \hfill (2.13)

Reversing the roles of A and B, we can also write
\[ P(B \mid A) = \frac{P(AB)}{P(A)}. \]  \hfill (2.14)

Combining equations (1.13) and (1.14), we can write
\[ P(A \mid B)P(B) = P(B \mid A)P(A) = P(AB). \]  \hfill (2.15)

It is clearly seen from the above, if the events A and B are statistically independent
\[ P(A \mid B) = P(A) \]  \hfill (2.16)

and
\[ P(B \mid A) = P(B). \]  \hfill (2.17)

Consider now the case of a sample space \( S \) consisting of five events \( A_1, A_2, \ldots, A_5 \) as shown in Figure 2.3. An arbitrary event B is also shown. The event B can be expressed as
\[ \{B\} = \{BA_1\} + \{BA_2\} + \{BA_3\} + \{BA_4\} + \{BA_5\}. \]  \hfill (2.18)

The probability of the event B can now be expressed as

![Figure 2.3 Concept of Bayes’ Theorem](image-url)
\[ P(B) = P(BA_1) + P(BA_2) + P(BA_3) \]
\[ + P(BA_4) + P(BA_5) = \sum_{k=1}^{5} P(BA_k). \]  
\[ \text{(2.19)} \]

Using the relationship between conditional probability and joint probability, we can rewrite the equation (1.19) as

\[ P(B) = \sum_{k=1}^{5} P(B \mid A_k)P(A_k). \]  
\[ \text{(2.20)} \]

The equation (1.20) can be generalized to

\[ P(B) = \sum_{k=1}^{M} P(B \mid A_k)P(A_k). \]  
\[ \text{(2.21)} \]

This is known as the total probability theorem. Note that \( P(B \mid A_k) \) is the conditional probability and \( P(A_k \mid B) \) is known as the *a posteriori* probability. The probability of the event \( A_k \), \( P(A_k) \) is known as the *a priori* probability. Using equation (1.14), we can obtain an expression for the *a posteriori* probability as

\[ P(A_k \mid B) = \frac{P(A_k B)}{P(B)} = \frac{P(B \mid A_k)P(A_k)}{P(B)}. \]  
\[ \text{(2.22)} \]

This relationship is known as the Bayes’ Theorem.

**Examples**

1. There are two computers (A and B) in an office. The secretary makes the following observation:

   The probability that both will boot up is 0.1; the probability that the B boots and the A does not boot up is 0.2; probability that neither boots up is 0.4.

   (a) Find the probability that A will boot up.
   (b) Find the probability that A will boot up, given that the car B will boot up.
   (c) Find the probability that car B will boot up, given that car A does not boot up.

Example: Let us examine what has been given. If A and B represent the events the computers boot up, \( A^c \) and \( B^c \) represent the corresponding complimentary events that they do not boot up.

i.e. \( P(A) + P(A^c) = 1 \) and \( P(B) + P(B^c) = 1 \);
We are given the following probabilities: \( P(AB) = 0.1 \); \( P(B|A^c) = 0.2 \); \( P(A^cB^c) = 0.4 \);

2. In a multiple-choice examination, the fraction students who know the answers is 20%. They write the correct answers in the test. The rest of the students guess the answers. If there are 30 questions in the test, the probability of getting the right answer by guessing is 1/30. If a student got all of them right, what is the probability that she/he new all the answers.

Answer:

3. A computer manufacturer uses chips from three different manufacturers, A, B, and C. The percentages of defective chips from these manufacturers respectively are 0.002, 0.001, and 0.00015. One chip is randomly picked and was found to be defective. What is the probability that it came from the manufacturer B?

Answer:

2.4 Bernoulli Trials and Repeated Experiments

In many of the engineering and medical research, the same experiment or trial may be repeated a number of times. For example, a particular medicine is given to 1000 people and if we know that the probability of curative effects of medicine is 0.7, we would like to know the probability of 600 successful cures. A particular bolt or rivet used in a machine assembly has 0.5% chance of failure. If that rivet is used in 10000 machines, what is the probability that 9999 machines will be fine? To understand these, let us take the analogy of a coin toss. Instead of tossing the coin once, the coin is now tossed a number of times say \( N \). Our interest is in finding out, the probability of observing \( k \) heads at the end of this ‘repeated trial’. Let the probability of seeing heads in a single toss is \( p \). One of the possible sequences of seeing \( k \) heads is seeing heads in the first \( k \) tosses followed by tails in the remaining tosses \( (N-k) \)

\[
\text{HHHHHHHHH.....k times TTTTTTTTT.. (N-k) times}
\]

Taking advantage of the fact that each of these out comes (heads or tails) is independent, the probability of observing this sequence will be \( p^k(1-p)^{N-k} \) where we have also made use of the fact that heads and tails form a mutually exclusive set and the probability of seeing tails hence is \( (1-p) \). However, the sequence given in eqn. (1) is not the only one in which we can see \( k \) heads in \( N \) tosses. There are other combinations of occurrences of heads and tails that will give rise to \( k \) heads. In other words, we have \( \binom{N}{k} \) ways of observing \( k \) heads. (Note that the order is not important here). Each of these possible sequences of having \( k \) heads still has the same probability. Making use of the fact that each of these
sequences are mutually exclusive of others, we can now express the probability of observing \( k \) heads in \( N \) tosses as

\[
P(\text{k Heads}) = P(\text{k Heads occur in any order})
\]

\[
= p^k (1 - p)^{N-k} + p^k (1 - p)^{N-k} .... C_k^N \text{ times} \quad (2.23)
\]

\[
= C_k^N p^k (1 - p)^{N-k}
\]

These repeated trial experiments where each attempt is independent of the other are referred to as the ‘Bernoulli Trials”. The probability expression given in eqn. () is also known as the Bionomial probability.

**Examples of Bernoulli Trials and Bionomial probability**

1. The probability of failure of a chip is 0.001. If this chip has been installed in 8 computers, what is the probability that 7 of them do not break down?

   **Answer**

   Prob of failure =0.01 \((p)\). The number of computers that break down = 1.
   Probability that 7 of them do not break down = Prob. that one breaks down =

   \[
   C_1^8 p^1 (1 - p)^7 = C_1^8 (0.01)^1 (0.99)^7 = 0.0746 = C_7^8 (0.99)^7 (0.01)^1
   \]

2. A student is in the habit of coming to the class late 30% of the time. If the student has five classes in a week, what is the probability that, the student will be late for at least 4 classes?

   **Answer**

   \[
P(\text{late for at least 4 classes}) = P(\text{late for 4 classes}) + P(\text{late for 5 classes})
   \]

   \[
   = C_4^5 (0.3)^4 (0.7)^1 + C_5^5 (0.3)^5 (0.7)^0
   \]