INSTITUTE OF ACTUARIAL & QUANTITATIVE STUDIES

Class: FY BSc

Subject : Probability and Statistics -1

Chapter: Unit 1 Chapter 3

Chapter Name: Fundamentals of Probability

Topics to be covered

1. Origins and Uses of Probability
2. What is Probability?
3. Concepts of Probability
4. Sample Space
5. Event
6. Combination of Events
7. Mutually Exclusive Events
8. Probability of an Event
9. Theorem 1-7
10. Conditional Probability
11. Independence of Events
12. Bayes Theorem

1 Origins and Uses of Probability

- The theory of probability began with the study of games of chance such as poker. Predictions take the form of probabilities.
- To predict the likelihood of an earthquake, of rain, or whether you will get an A in this course, we use probabilities.
- Doctors use probability to determine the chance of a vaccination causing the disease the vaccination is supposed to prevent.
- A stockbroker uses probability to determine the rate of return on a client's investments.
- You might use probability to decide to buy a lottery ticket or not.
- In your study of statistics, you will use the power of mathematics through probability calculations to analyze and interpret your data.

2 What is Probability?

- The oldest way of defining probabilities, the classical probability concept, applies when all possible outcomes are equally likely, as is presumably the case in most games of chance.
- We can then say that if there are N equally likely possibilities, of which one must occur and n are regarded as favorable, or as a "success," then the probability of a "success" is given by the ratio $\frac{n}{N}$.

- Although equally likely possibilities are found mostly in games of chance, the classical probability concept applies also in a great variety of situations where gambling devices are used to make random selections—when office space is assigned to teaching assistants by lot, when some of the families in a township are chosen in such a way that each one has the same chance of being included in a sample study, when machine parts are chosen for inspection so that each part produced has the same chance of being selected, and so forth.

- A major shortcoming of the classical probability concept is its limited applicability, for there are many situations in which the possibilities that arise cannot all be regarded as equally likely. This would be the case, for instance, if we are concerned with the question whether it will rain on a given day, if we are concerned with the outcome of an election, or if we are concerned with a person's recovery from a disease.

Example

What is the probability of drawing an ace from an ordinary deck of 52 playing cards?

Solution:

Since there are n = 4 aces among the N = 52 cards, the probability of drawing an ace is $\frac{4}{52} = \frac{1}{13}$.

(It is assumed, of course, that each card has the same chance of being drawn.)

3 Concepts of Probability

- Since all probabilities pertain to the occurrence or nonoccurrence of events, let us explain first what we mean here by event and by the related terms experiment, outcome, and sample space.
- It is customary in statistics to refer to any process of observation or measurement as an experiment. In this sense, an experiment may consist of the simple process of checking whether a switch is turned on or off; it may consist of counting the imperfections in a piece of cloth; or it may consist of the very complicated process of determining the mass of an electron.
- The results one obtains from an experiment, whether they are instrument readings, counts, “yes” or “no” answers, or values obtained through extensive calculations, are called the outcomes of the experiment.

4 Sample Space

The set of all possible outcomes of an experiment is called the **sample space** and it is usually denoted by the letter *S*. Each outcome in a sample space is called an element of the sample space, or simply a sample point.

- If a sample space has a finite number of elements, we may list the elements in the usual set notation; for instance, the sample space for the possible outcomes of one flip of a coin may be written as : $S = \{H, T\}$ where H and T stand for head and tail.

- Sample spaces with a large or infinite number of elements are best described by a statement or rule; for example, if the possible outcomes of an experiment are the set of automobiles equipped with satellite radios, the sample space may be written $S = \{x|x \text{ is an automobile with a satellite radio}\}$
- This is read "S is the set of all x such that x is an automobile with a satellite radio." Similarly, if S is the set of odd positive integers, we write $S = \{2k+1|k = 0, 1, 2, \ldots\}$
- How we formulate the sample space for a given situation will depend on the problem at hand. If an experiment consists of one roll of a die and we are interested in which face is turned up, we would use the sample space : $S = \{1,2,3,4,5,6\}$

- It is desirable to use sample spaces whose elements cannot be divided (partitioned or separated) into more primitive or more elementary kinds of outcomes. In other words, it is preferable that an element of a sample space not represent two or more outcomes that are distinguishable in some way.

Example

Describe a sample space that might be appropriate for an experiment in which we roll a pair of dice, one red and one green. (The different colors are used to emphasize that the dice are distinct from one another.)

Solution:

The sample space that provides the most information consists of the 36 points given by

$$S_1 = \{(x, y) | x = 1, 2, \ldots, 6; y = 1, 2, \ldots, 6\}$$

where x represents the number turned up by the red die and y represents the number turned up by the green die. A second sample space, adequate for most purposes (though less desirable in general as it provides less information), is given by

$$S_2 = \{2, 3, 4, \ldots, 12\}$$

where the elements are the totals of the numbers turned up by the two dice.

5 Event

An event is a subset of a sample space.

- An event (outcome or result) can be identified with a collection of points, which constitute a subset of an appropriate sample space. Such a subset consists of all the elements of the sample space for which the event occurs, and in probability and statistics we identify the subset with the event.

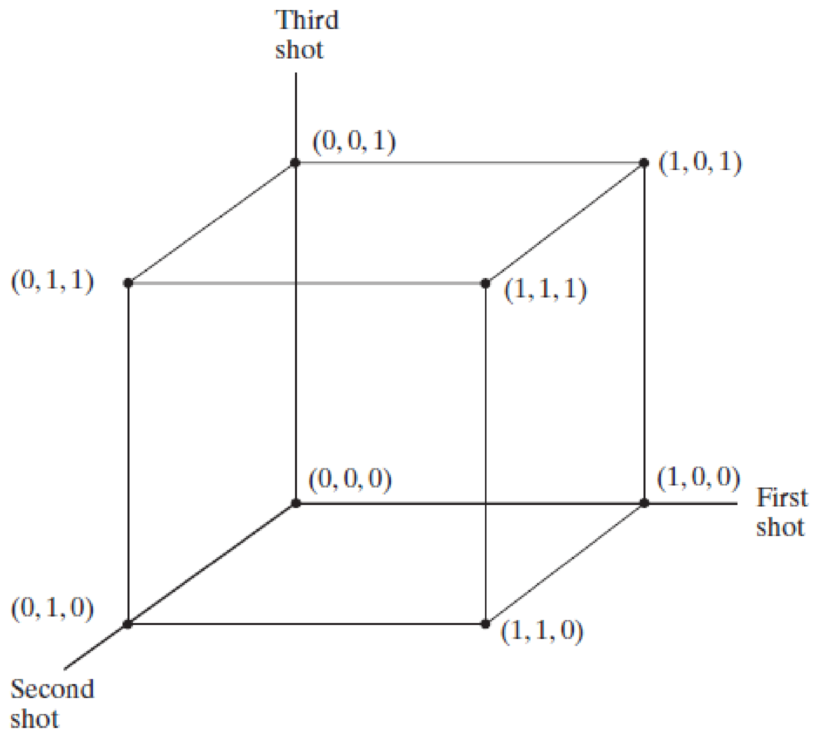
- According to our definition, any event is a subset of an appropriate sample space, but it should be observed that the converse is not necessarily true. For discrete sample spaces, all subsets are events, but in the continuous case some rather abstruse point sets must be excluded for mathematical reasons.

Example

If someone takes three shots at a target and we care only whether each shot is a hit or a miss, describe a suitable sample space, the elements of the sample space that constitute event M that the person will miss the target three times in a row, and the elements of event N that the person will hit the target once and miss it twice.

Solution:

If we let 0 and 1 represent a miss and a hit, respectively, the eight possibilities (0, 0, 0), (1, 0, 0), (0, 1, 0), (0, 0, 1), (1, 1, 0), (1, 0, 1), (0, 1, 1), and (1, 1, 1) may be displayed as in the adjoining figure.



Thus, it can be seen that

M = {(0, 0, 0)} and N = {(1, 0, 0), (0, 1, 0), (0, 0, 1)}

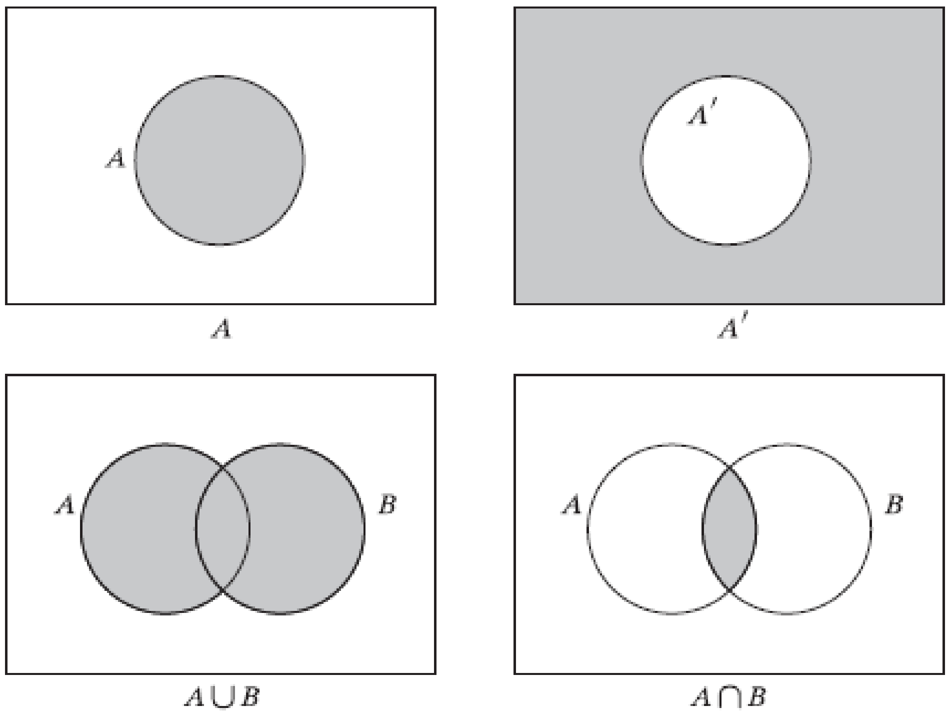
6 Combination of Events

- In many problems of probability we are interested in events that are actually combinations of two or more events, formed by taking **unions, intersections**, and **complements**.

- If A and B are any two subsets of a sample space S,
  - their union $A \cup B$ is the subset of S that contains all the elements that are either in A, in B, or in both;
  - their intersection $A \cap B$ is the subset of S that contains all the elements that are in both A and B;
  - and the complement A' of A is the subset of S that contains all the elements of S that are not in A.

6 Combination of Events

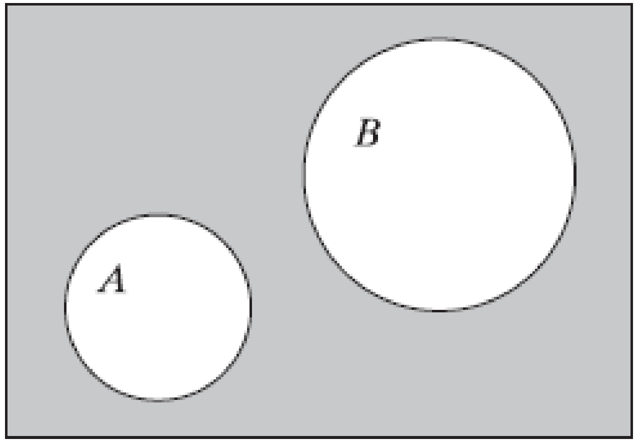
- Sample spaces and events, particularly relationships among events, are often depicted by means of Venn diagrams, in which the sample space is represented by a rectangle, while events are represented by regions within the rectangle, usually by circles or parts of circles. For instance, the shaded regions of the four Venn diagrams besides represent, respectively, event A, the complement of event A, the union of events A and B, and the intersection of events A and B.



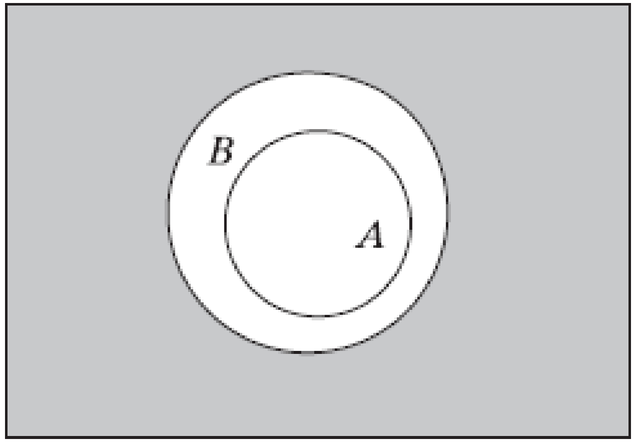
7 Mutually exclusive events

Two events having no elements in common are said to be mutually exclusive.

- When A and B are mutually exclusive, we write $A \cap B = \emptyset$, where $\emptyset$ denotes the empty set, which has no elements at all. The diagram on the right serves to indicate that A is contained in B, and symbolically we express this by writing $A \subset B$.



A and *B* mutually exclusive



A contained in *B*

8 Probability of an Event

- To formulate the postulates of probability, we shall follow the practice of denoting events by means of capital letters, and we shall write the probability of event A as P(A), the probability of event B as P(B), and so forth. The following postulates of probability apply only to discrete sample spaces, S.

POSTULATE 1: The probability of an event is a nonnegative real number; that is, $P(A) \geq 0$ for any subset A of *S*.
POSTULATE 2: $P(S) = 1$.
POSTULATE 3: If A1,A2,A3, . . ., is a finite or infinite sequence of mutually exclusive events of *S*, then

$$P(A_1 \cup A_2 \cup A_3 \cup \cdots) = P(A_1) + P(A_2) + P(A_3) + \cdots$$

8 Probability of an Event

- Since proportions are always positive or zero, the first postulate is in complete agreement with the frequency interpretation.
- The second postulate states indirectly that certainty is identified with a probability of 1; after all, it is always assumed that one of the possibilities in S must occur, and it is to this certain event that we assign a probability of 1. As far as the frequency interpretation is concerned, a probability of 1 implies that the event in question will occur 100 percent of the time or, in other words, that it is certain to occur.
- Taking the third postulate in the simplest case, that is, for two mutually exclusive events A_1 and A_2, it can easily be seen that it is satisfied by the frequency interpretation. If one event occurs, say, 28 percent of the time, another event occurs 39 percent of the time, and the two events cannot both occur at the same time (that is, they are mutually exclusive), then one or the other will occur 28+39 = 67 percent of the time. Thus, the third postulate is satisfied, and the same kind of argument applies when there are more than two mutually exclusive events.

- Note that the three postulates do not tell us how to assign probabilities to events; they merely restrict the ways in which it can be done.

Example

An experiment has four possible outcomes, A, B, C, and D, that are mutually exclusive. Explain why the following assignments of probabilities are not permissible:

(a) P(A) = 0.12,P(B) = 0.63,P(C) = 0.45,P(D) = −0.20;

(b) $\boldsymbol{P(A) = \frac{9}{120}, P(B) = \frac{45}{120}, P(C) = \frac{27}{120}, P(D) = \frac{46}{120}}$

Solution:

(a) P(D) = −0.20 violates Postulate 1;

(b) $\mathrm{P(S) = P(A \cup B \cup C \cup D)} = \frac{9}{120} + \frac{45}{120} + \frac{27}{120} + \frac{46}{120} = \frac{127}{120} \neq 1,$ and this violates Postulate 2.

9 Theorem 1

Theorem:
If A is an event in a discrete sample space S, then P(A) equals the sum of the probabilities of the individual outcomes comprising A.

Proof:
Let $O_1, O_2 O_3, . . .$, be the finite or infinite sequence of outcomes that comprise the event A. Thus,

$$A = O_1 \cup O_2 \cup O_3 \ldots$$

and since the individual outcomes, the O's, are mutually exclusive, the third postulate of probability yields

$$P(A) = P(O_1) + P(O_2) + P(O_3) + \cdots$$

This completes the proof.

Example

If we twice flip a balanced coin, what is the probability of getting at least one head?

Solution:

The sample space is S = {HH, HT, TH, TT}, where H and T denote head and tail. Since we assume that the coin is balanced, these outcomes are equally likely and we assign to each sample point the probability 14. Letting A denote the event that we will get at least one head, we get A = {HH, HT, TH} and

$$P(A) = P(HH) + P(HT) + P(TH)$$

$$= \frac{1}{4} + \frac{1}{4} + \frac{1}{4}$$

$$= \frac{3}{4}$$

9 Theorem 2

Theorem:
If an experiment can result in any one of N different equally likely outcomes, and if n of these outcomes together constitute event A, then the probability of event A is

$$P(A) = \frac{n}{N}$$

Proof:

Let O1,O2, . . . ,ON represent the individual outcomes in S, each with probability $\frac{\mathbf{1}}{\boldsymbol{N}}$**.** If A is the union of n of these mutually exclusive outcomes, and it does not matter which ones, then

$$P(A) = P(O_1 \cup O_2 \cup O_3 \ldots \cup O_n)$$

$$P(A) = P(O_1) + P(O_2) + \cdots P(O_n)$$

$$= \frac{1}{N} + \frac{1}{N} + \cdots \frac{1}{N} \Rightarrow n\ terms$$

$$= \frac{n}{N}$$

This completes the proof.

Example

A five-card poker hand dealt from a deck of 52 playing cards is said to be a full house if it consists of three of a kind and a pair. If all the five-card hands are equally likely, what is the probability of being dealt a full house?

Solution:

The number of ways in which we can be dealt a particular full house, say three kings and two aces, is $\binom{4}{3}\binom{4}{2}$.

Since there are 13 ways of selecting the face value for the three of a kind and for each of these there are 12 ways of selecting the face value for the pair, there are altogether

$$n = 13.12.\binom{4}{3}\binom{4}{2}$$

different full houses. Also, the total number of equally likely five-card poker hands is

$$N = \binom{52}{5}$$

and it follows by Theorem 2 that the probability of getting a full house is

$$P(A) = \frac{n}{N} = \frac{13.12\binom{4}{3}\binom{4}{2}}{\binom{52}{5}} = 0.0014$$

9 Theorem 3

Theorem:
If A and A' are complementary events in a sample space S, then $P(A') = 1 - P(A)$

Proof:
In the second and third steps of the proof that follows, we make use of the definition of a complement, according to which A and A' are mutually exclusive and $A \cup A' = S$. Thus, we write

$$\begin{aligned} 1 &= P(S) \quad \text{(by Postulate 2)} \\ &= P(A \cup A') \\ &= P(A) + P(A') \quad \text{(by Postulate 3)} \end{aligned}$$

and it follows that $P(A') = 1 - P(A)$.

9 Theorem 4

Theorem:
$P(\emptyset) = 0$ for any sample space S.

Proof:
Since S and $\emptyset$ are mutually exclusive and $S \cup \emptyset = S$ in accordance with the definition of the empty set $\emptyset$, it follows that

$$\begin{aligned} P(S) &= P(S \cup \emptyset) \\ &= P(S) + P(\emptyset) \quad \text{(by Postulate 3)} \end{aligned}$$

and, hence, that $P(\emptyset) = 0$.

9 Theorem 5

Theorem:
If A and B are events in a sample space S and $A \subset B$, then $P(A) \leq P(B)$.
Proof:
Since $A \subset B$, we can write

$$B = A \cup (A' \cap B)$$

as can easily be verified by means of a Venn diagram. Then, since A and $A' \cap B$ are mutually exclusive, we get

$$P(B) = P(A) + P(A' \cap B) \qquad \text{(by Postulate 3)}$$

$$\geq P(A) \qquad \text{(by Postulate 1)}$$

9 Theorem 6

Theorem:
$0 \leq P(A) \leq 1$ for any event A.

Proof:
Using Theorem 5 and the fact that ∅ ⊂ A ⊂ S for any event A in S, we have

$$P(\emptyset) \leq \mathrm{P}(A), +P(S)$$

Then, P(∅) = 0 and P(S) =1 leads to the result that

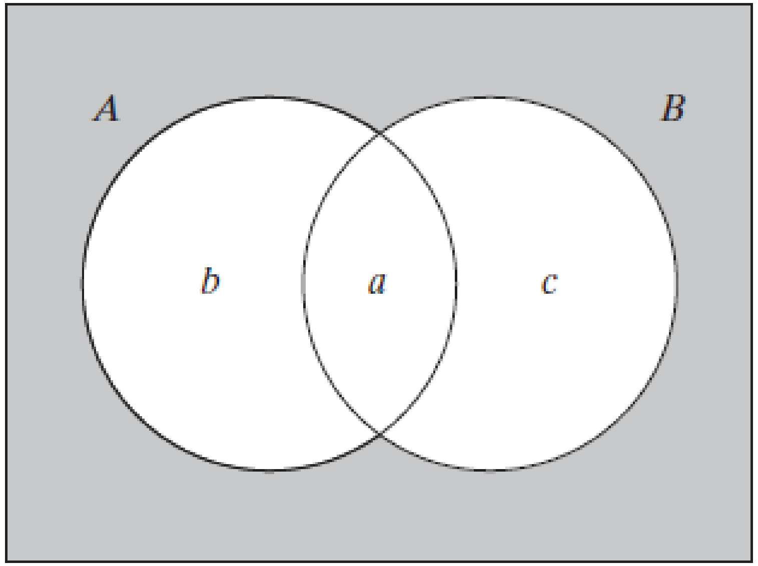
$$0 \leq P(A) \leq 1$$

9 Theorem 7

Theorem:
If A and B are any two events in a sample space S, then
$P(A \cup B) = P(A) + P(B) - P(A \cap B)$

Proof:
Assigning the probabilities a, b, and c to the mutually exclusive events $A \cap B, A \cap B'$, and $A' \cap B$ as in the adjoining Venn diagram, we find that

$$\begin{aligned} P(A \cup B) &= a+b+c \\ &= (a+b)+(c+a)-a \\ &= P(A)+P(B)-P(A \cap B) \end{aligned}$$



10 Conditional Probability

If A and B are any two events in a sample space S and $P(A) \neq 0$, the conditional probability of B given A is

$$P(B|A) = \frac{P(A \cap B)}{P(A)}$$

- Difficulties can easily arise when probabilities are quoted without specification of the sample space. For instance, if we ask for the probability that a lawyer makes more than $75,000 per year, we may well get several different answers, and they may all be correct.
- One of them might apply to all those who are engaged in the private practice of law, another might apply to lawyers employed by corporations, and so forth. Since the choice of the sample space (that is, the set of all possibilities under consideration) is by no means always self-evident, it often helps to use the symbol $P(A|S)$ to denote the conditional probability of event A relative to the sample space S or, as we also call it, "the probability of A given S." The symbol $P(A|S)$ makes it explicit that we are referring to a particular sample space S, and it is preferable to the abbreviated notation $P(A)$ unless the tacit choice of S is clearly understood.
- It is also preferable when we want to refer to several sample spaces in the same example.
- If A is the event that a person makes more than $75,000 per year, G is the event that a person is a law school graduate, L is the event that a person is licensed to practice law, and E is the event that a person is actively engaged in the practice of law, then $P(A|G)$ is the probability that a law school graduate makes more than $75,000 per year, $P(A|L)$ is the probability that a person licensed to practice law makes more than $75,000 per year, and $P(A|E)$ is the probability that a person actively engaged in the practice of law makes more than $75,000 per year.

Example

A manufacturer of airplane parts knows from past experience that the probability is 0.80 that an order will be ready for shipment on time, and it is 0.72 that an order will be ready for shipment on time and will also be delivered on time. What is the probability that such an order will be delivered on time given that it was ready for shipment on time?

Solution:

If we let R stand for the event that an order is ready for shipment on time and D be the event that it is delivered on time, we have $P(R) = 0.80$ and $P(R \cap D) = 0.72$, and it follows that

$$P(D|R) = \frac{P(R \cap D)}{P(R)} = \frac{0.72}{0.80} = 0.90$$

Thus, 90 percent of the shipments will be delivered on time provided they are shipped on time. Note that $P(R|D)$, the probability that a shipment that is delivered on time was also ready for shipment on time, cannot be determined without further information; for this purpose we would also have to know $P(D)$.

Example

Find the probabilities of randomly drawing two aces in succession from an ordinary deck of 52 playing cards if we sample

(a) without replacement;

(b) with replacement.

Solution:

(a) If the first card is not replaced before the second card is drawn, the probability of getting two aces in succession is

$$\frac{4}{52} * \frac{3}{51} = \frac{1}{221}$$

(b) If the first card is replaced before the second card is drawn, the corresponding probability is

$$\frac{4}{52} * \frac{4}{52} = \frac{1}{169}$$

11 Independence of events

Two events A and B are **independent** if and only if

$$\mathbf{P(A \cap B) = P(A) \cdot P(B)}$$

- Informally speaking, two events A and B are independent if the occurrence or nonoccurrence of either one does not affect the probability of the occurrence of the other.
- Symbolically, two events A and B are independent if $P(B|A) = P(B)$ and $P(A|B) = P(A)$, and it can be shown that either of these equalities implies the other when both of the conditional probabilities exist, that is, when neither $P(A)$ nor $P(B)$ equals zero.

- If two events are not independent, they are said to be dependent.

Example

A coin is tossed three times and the eight possible outcomes, HHH, HHT, HTH, THH, HTT, THT, TTH, and TTT, are assumed to be equally likely. If A is the event that a head occurs on each of the first two tosses, B is the event that a tail occurs on the third toss, and C is the event that exactly two tails occur in the three tosses, show that

(a) events A and B are independent;

(b) events B and C are dependent.

Solution:

Since

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

$$B = \{HHT, HTT, THT, TTT\}$$

$$C = \{HTT, THT, TTH\}$$

$$A \cap B = \{HHT\}$$

$$B \cap C = \{HTT, THT\}$$

Continued

the assumption that the eight possible outcomes are all equiprobable yields

$P(A) = \frac{1}{4}$, $P(B) = \frac{1}{2}$, $P(C) = \frac{3}{8}$,$P(A \cap B) = \frac{1}{8}$, and $P(B \cap C) = \frac{1}{4}$.

(a) Since $P(A) \cdot P(B) = \frac{1}{4} * \frac{1}{2} = \frac{1}{8} = P(A \cap B)$, events A and B are independent.

(b) Since $P(B) \cdot P(C) = \frac{1}{2} * \frac{3}{8} = \frac{3}{16} \neq P(B \cap C)$, events B and C are not independent.

11 Independence of events

Events $A_1, A_2, \ldots,$ and A_k are independent if and only if the probability of the intersections of any 2, 3, . . . , or k of these events equals the product of their respective probabilities.

- For three events A, B, and C, for example, independence requires that

$$P(A \cap B) = P(A) \cdot P(B)$$
$$P(A \cap C) = P(A) \cdot P(C)$$
$$P(B \cap C) = P(B) \cdot P(C)$$

- and

$$P(A \cap B \cap C) = P(A) \cdot P(B) \cdot P(C)$$

- It is of interest to note that three or more events can be pairwise independent without being independent.

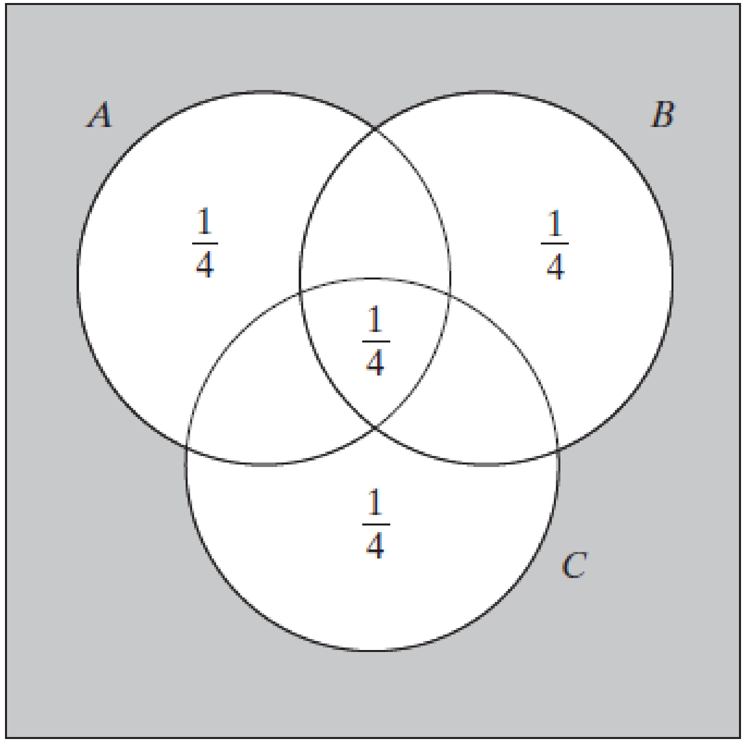
Example

Following figure shows a Venn diagram with probabilities assigned to its various regions. Verify that A and B are independent, A and C are independent, and B and C are independent, but A, B, and C are not independent.

Solution:

As can be seen from the diagram, $P(A) = P(B) = P(C) = \frac{1}{2}$

, $P(A \cap B) = P(A \cap C) = P(B \cap C) = \frac{1}{4}$, and $P(A \cap B \cap C) = \frac{1}{4}$.



Thus,

$$P(A) \cdot P(B) = \frac{1}{4} = P(A \cap B)$$

$$P(A) \cdot P(C) = \frac{1}{4} = P(A \cap C)$$

$$P(B) \cdot P(C) = \frac{1}{4} = P(B \cap C)$$

but

$$P(A) \cdot P(B) \cdot P(C) = \frac{1}{8} \neq (A \cap B \cap C)$$

12 Bayes Theorem

- Bayes' theorem (also known as Bayes' rule or Bayes' law) is a result in probabillity theory that relates conditional probabilities. If A and B denote two events, P(A|B) denotes the conditional probability of A occurring, given that B occurs.

- The two conditional probabilities P(A|B) and P(B|A) are in general different.

- Bayes theorem gives a relation between P(A|B) and P(B|A).

- An important application of Bayes' theorem is that it gives a rule how to update or revise the strengths of evidence-based beliefs in light of new evidence a posterior.

- As a formal theorem, Bayes' theorem is valid in all interpretations of probability. However, it plays a central role in the debate around the foundations of statistics: frequentist and Bayesian interpretations disagree about the kinds of things to which probabilities should be assigned in applications. Whereas frequentists assign probabilities to random events according to their frequencies of occurrence or to subsets of populations as proportions of the whole, Bayesians assign probabilities to propositions that are uncertain.
- A consequence is that Bayesians have more frequent occasion to use Bayes' theorem. The articles on Bayesian probability and frequentist probability discuss these debates at greater length.

12 Bayes Theorem

Bayes' theorem relates the conditional and marginal probabilities of stochastic events A and B:

$$\boldsymbol{P(A|B) = \frac{P(B|A).P(A)}{P(B)}}$$

Each term in Bayes' theorem has a conventional name:

- P(A) is the prior probability or marginal probability of A. It is "prior" in the sense that it does not take into account any information about B.
- P(A|B) is the conditional probability of A, given B. It is also called the posterior probability because it is derived from or depends upon the specified value of B.
- P(B|A) is the conditional probability of B given A.
- P(B) is the prior or marginal probability of B, and acts as a normalizing constant.

12 Bayes Theorem

Proof

- The probability of two events A and B happening, $P(A \cap B)$, is the probability of A, $P(A)$, times the probability of B given that A has occurred,

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

- On the other hand, the probability of A and B is also equal to the probability of B times the probability of A given B.

$$P(A \cap B) = P(B)P(A|B)$$

- Equating the two yields:

$$P(B)P(A|B) = P(A)P(B|A)$$

- and thus

$$P(A|B) = \frac{P(A)P(B|A)}{P(B)}$$

- This equation, known as Bayes Theorem is the basis of statistical inference

Example

A rare but serious disease, D, has been found in 0.01 percent of a certain population. A test has been developed that will be positive, p, for 98 percent of those who have the disease and be positive for only 3 percent of those who do not have the disease. Find the probability that a person tested as positive does not have the disease.

Solution:

Let $\bar{D}$ and $\bar{p}$ represent the events that a person randomly selected from the given population, respectively, does not have the disease and is found negative for the disease by the test. Substituting the given probabilities into the formula, we get

$$P(\bar{D}|p) = \frac{P(\bar{D})P(p|\bar{D})}{P(D)P(p|D) + P(\bar{D})P(p|\bar{D})} = \frac{0.9999 * 0.03}{0.0001 * 0.98 + 0.9999 * 0.03} = 0.997$$

This example demonstrates the near impossibility of finding a test for a rare disease that does not have an unacceptably high probability of false positives.