2

Chapter

Probability, Probability Distributions and Stochastic Processes

In Chapter 1, we have mentioned the various aspects of the theory of reliability along with some definitions. It is obvious from Chapter 1 that the study of the theory of reliability requires a background of probability. Therefore, before proceeding to the study of various systems, we give here an outline of probability theory reliable for the study of reliability as well as queueing problems. The theory of probability is concerned with the study of those methods of analysis that are common to the study of random phenomena in all the fields in which they arise.

2.1 PROBABILITY

Before proceeding to the definition of probability, we explain certain terms to be used at subsequent occasions.

An experiment resulting in any of the possible outcomes is called a *trial* and the possible outcomes are known as the events or cases. The events are said to be *equally likely* when we have no reason to expect any one rather than the other. Events are called *mutually exclusive* or incompatible if the occurrence of one of them precludes the occurrence of all the others. On the contrary, events are compatible if it is possible for them to happen simultaneously. Moreover, events are said to be exhaustive when they include all possible ones. Also, we call the cases to be favourable to an event if they entail its happening.

We now consider the definitions of probability:

(a) ***Mathematical or Priori Definition:*** If there are n exhaustive, mutually exclusive and equally likely cases and m are favourable to an event A, the probability of the happening of A is defined as the ratio $\frac{m}{n}$.

(b) ***Statistical or Empirical definition:*** If trials be repeated a great number of times under essentially the same conditions then the limit of the ratio of the number of times that an event happens to the total number of trials as the number of trials increases indefinitely is called the probability of the happening of that event. It is assumed that the ratio approaches a finite and a unique limit.

$$= \frac{M_1}{N} + \frac{m_2}{N} + \ldots + \frac{m_n}{N}$$

$$= P(A_1) + P(A_2) + \ldots + P(A_n)$$

which proves the theorem. This theorem is called the theorem of total probability.

Theorem 2: If the events $A_1, A_2, \ldots, A_n$ are not mutually exclusive, then probability of at least one of the n events is given by

$$P(A_1 + A_2 + \ldots + A_n) = \sum_{i=1}^{n} P(A_i) - \sum_{\substack{i,\, j=1 \\ i \neq j}}^{n} P(A_i A_j) + \ldots + (-1)^{n-1} P(A_1 A_2 \ldots A_n) \quad (2.3.2)$$

Proof: Consider the events A and B only.

Since, AB and $A\overline{B}$ are two exhaustive and mutually exclusive forms in which A can occur, we have

$$P(A) = P(AB) + P(A\overline{B})$$

Similarly, $P(B) = P(BA) + P(B\overline{A}) = P(AB) + P(\overline{A}B)$.

Adding, we get

$$P(A) + P(B) = P(AB) + [P(A\overline{B}) + P(\overline{A}B) + P(AB)]$$

By Theorem 1, the expression within the square brackets represents the probability $P(A + B)$ of the occurrence of at least one of the events A and B. Hence,

$$P(A + B) = P(A) + P(B) - P(AB). \quad (2.3.3)$$

Let B mean the occurrence of at least one of the events A_2 and A_3, then writing A_1 for A, (2.3.3) gives

$$P(A_1 + A_2 + A_3) = P(A_1) + P(A_2) + P(A_3) - P(A_1 A_2) - P(A_2 A_3) - P(A_3 A_1) + P(A_1 A_2 A_3)$$

$$= \sum_{i=1}^{3} P(A_i) - \sum_{\substack{i,\, j=1 \\ i \neq j}}^{3} P(A_i A_j) + P(A_1 A_2 A_3) \quad (2.3.4)$$

The general law for n events, which may be proved by mathematical induction, is

$$P(A_1 + A_2 + \ldots + A_n) = \sum_{i=1}^{n} P(A_i) - \sum_{\substack{i,\, j=1 \\ i \neq j}}^{n} P(A_i A_j) + \sum_{\substack{i,\, j,\, k=1 \\ i \neq j \neq k}}^{n} P(A_i A_j A_k) - \ldots + (-1)^{n-1} P(A_1 A_2 \ldots A_n), \quad (2.3.5)$$

where the second sum is overall combination of the numbers 1, 2, …, n taken two at a time, the third is overall combination of the numbers taken three at a time, and so forth.

Note: $\overline{A}$ and $\overline{B}$ denote the non-happening of events A and B, respectively.

Cor. If $A_1, A_2, \ldots, A_n$ are n mutually exclusive events, then $P(A_i A_j) = 0$, $P(A_i A_j A_k) = 0$, $P(A_1 A_2 \ldots A_n) = 0$.

Consequently, (2.3.5) reduces to

$$P(A_1 + A_2 + \ldots + A_n) = \sum_{i=1}^{n} P(A_i) = P(A_1) + P(A_2) + \ldots + P(A_n)$$

which is the theorem of total probability.

Theorem 3: The probability of the simultaneous occurrence of two events A and B is equal to the probability of A multiplied by the conditional probability of B, given that A has occurred (or it is equal to the probability of B multiplied by the conditional probability of A given that B has occurred), i.e.,

$$P(AB) = P(A)\ P(B|A) = P(B)\ P(A|\mathrm{B}). \qquad (2.3.6)$$

Proof: Let N denotes the total number of mutually exclusive and equally likely cases among which m cases are favourable to the event A. The cases favourable to both the events A and B are included in the m cases favourable to A. Let their number be m_1. Then the probability $P(AB)$ that both the events A and B will happen is given by

$$P(AB) = \frac{m_1}{N} = \frac{m}{N} \cdot \frac{m_1}{m}.$$

The ratio m/N is the probability of A, i.e., $m/N = P(A)$.

Assuming the occurrence of A there are only m equally likely cases left out of which m_1 are also favourable to B. Hence, the ratio m_1/m represents the conditional probability $P(B|A)$ of B, supposing that A has occurred. Hence,

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

Since, the compound event AB involves A and B symmetrically, we shall have

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

This theorem is called theorem of compound probability.

If A and B are independent events, then $P(B|A)$ is same as $P(B)$, and we have

$$P(AB) = P(A)\ P(B). \qquad (2.3.7)$$

In general, for independent events, we have

$$P(A_1 A_2 \ldots A_n) = P(A_1)\ P(A_2) \ldots P(A_n). \qquad (2.3.8)$$

Problems

1. Three groups of children contain respectively 3 girls and 1 boy; 2 girls and 2 boys; 1 girl and 3 boys. One child is selected at random from each group. Show that the chance that the three selected, consist of 1 girl and 2 boys is $\frac{13}{32}$.

   One girl and two boys may be selected in the following ways:

   (i) Girl from the I group, boy from II group, boy from III group.

The probability of this event $= \frac{3}{4} \cdot \frac{2}{4} \cdot \frac{3}{4} = \frac{9}{32}$ (Theorem 3)

(ii) Boy from I group, girl from II group, boy from III group. The probability of this event

$$\frac{1}{4} \cdot \frac{2}{4} \cdot \frac{3}{4} = \frac{3}{32}.$$

(iii) Boy from I group, boy from II group, girl from III group. The probability of this event

$$\frac{1}{4} \cdot \frac{2}{4} \cdot \frac{1}{4} = \frac{1}{32}$$

All these events are mutually exclusive, hence the chance that any of these events happen

$$= \frac{9}{32} + \frac{3}{32} + \frac{1}{32} = \frac{13}{32} \text{ (Theorem 1)}$$

2. A can hit a target in 4 times in 5 shots; B 3 times in 4 shots; C twice in 3 shots. They fire a volley. What is the probability that two shots at least hit?

Chance of A's hitting $= \frac{4}{5}$; Chance of B's hitting $= \frac{3}{4}$;

Chance of C's hitting $= \frac{2}{3}$.

For at least two hits, we may have,

(i) A, B, C all may hit with probability $= \frac{4}{5} \cdot \frac{3}{4} \cdot \frac{2}{3} = \frac{2}{5}$ (Th. 3)

(ii) B, C may hit and A may lose with probability $= \frac{1}{5} \cdot \frac{3}{4} \cdot \frac{2}{3} = \frac{1}{10}$

(iii) C, A may hit and B may lose with probability $= \frac{4}{5} \cdot \frac{1}{4} \cdot \frac{2}{3} = \frac{2}{15}$

(iv) A, B may hit and C may lose with probability $= \frac{4}{5} \cdot \frac{3}{4} \cdot \frac{1}{3} = \frac{1}{5}$

Since these are mutually exclusive events, the required probability $= \frac{2}{5} + \frac{1}{10} + \frac{2}{15} + \frac{1}{5}$

$$= \frac{5}{6} \text{ (Theorem 1)}$$

2.4 BAYE'S THEOREM

Statement: An event A can be explained by a set of exhaustive and mutually exclusive hypotheses B_1, B_2, …, B_n. Given

(i) 'a priori' probabilities $P(B_1)$, $P(B_2)$, …, $P(B_n)$ corresponding to a total absence of knowledge regarding the occurrence of A and

(ii) Conditional probabilities $P(A|B_1)$, $P(A|B_2)$, …, $P(A|B_n)$,

(a) It is required to form the 'a posterion' probabilities $P(B_1|A)$, $P(B_2|A)$, …, $P(B_n|A)$.

(b) Further, find the probabilities of materialisation of another event C, given the probabilities $P(C|AB_1)$, $P(C|AB_2)$, …, $P(C|AB_n)$.

Proof:

(i) By the theorem of compound probability (Sec. 2.3),

$$P(AB_i) = P(B_i)\ P(A|B_i) = P(A)\ P(B_i|A)$$

$$\therefore \qquad P(B_i|A) = \frac{P(B_i)\ P(A|B_i)}{P(A)} \tag{2.4.1}$$

Since the event A can materialise in the mutually exclusive forms AB_1, AB_2, …, AB_n, we have by the theorem of total probability (Theorem 1, Section 2.3)

$$P(A) = P(AB_1) + P(AB_2) + \ldots + P(AB_n)$$

$$= \sum_{i=1}^{n} P(B_i)\ P(A|B_i) \tag{2.4.2}$$

$$\therefore \qquad P(B_i|A) = \frac{P(B_i)\ P(A|B_i)}{\sum_{i=1}^{n} P(B_i)\ P(A|B_i)} \tag{2.4.3}$$

(ii)

$$P(C_1A) = \sum_{i=1}^{n} P(CB_i|A) = \sum_{i=1}^{n} P(B_i|A)\ P(C|AB_i)$$

$$= \frac{\sum_{i=1}^{n} P(B_i)\ P(A|B_i)\ P(C|AB_i)}{\sum_{i=1}^{n} P(B_i)\ P(A|B_i)} \quad \text{(using 2.4.3)} \tag{2.4.4}$$

Relation (2.4.3) is also known as the formula for 'a posterion' probability. This name is explained by the fact that this formula gives the probability of relation of the events B_i with respect to the occurrence of A. $P(B_i)$ is known as a priori probability.

EXAMPLE: There are five urns of the following compositions:

2 urns with 2 white and 3 black balls each,

2 urns with 1 white and 4 black balls each

1 urn with 4 white balls and 1 black ball.

A ball is chosen from one of the urns taken at random. It turned out to be white. What is the probability after the experiment (a posterion probability) that the ball was taken from the urn of the third composition

By hypothesis, we have

$$P(B_1) = \frac{2}{5},\ P(B_2) = \frac{2}{5},\ P(B_3) = \frac{1}{5}$$

$$P(A|B_1) = \frac{2}{5},\ P(A|B_2) = \frac{1}{5},\ P(A|B_3) = \frac{4}{5}$$

By Baye's formula, we have

$$P(B_3|A) = \frac{P(B_3)\,P(A|B_3)}{P(B_1)\,P(A|B_1) + P(B_2)\,P(A|B_2) + P(B_3)\,P(A|B_3)}$$

$$= \frac{\frac{1}{5}\cdot\frac{4}{5}}{\frac{2}{5}\cdot\frac{2}{5}+\frac{2}{5}\cdot\frac{1}{5}+\frac{1}{5}\cdot\frac{4}{5}} = \frac{4}{10} = \frac{2}{5}$$

In exactly the same way, we may find

$$P(B_1|A) = \frac{2}{5},\ P(B_2|A) = \frac{1}{5}.$$

2.5 RANDOM VARIABLE

Many applications of probability theory are based on the notion of a random variable. We use the term statistical experiment to describe any process by which several chance measurements are obtained. All the possible outcomes of a statistical experiment comprise a set which we call sample space. Often we are not interested in the details associated with each sample point but in some numerical description of the outcome for which we require the random variable concept. We define a random variable as follows.

A random variable is a variable quantity whose values depend on chance and for which a distribution function of probabilities has been defined or equivalently a quantity X is said to be a random variable (or, equivalently, X is said to be an observed value of a numerical valued random phenomenon) if for every real number x there exists a probability that X is less than or equal to x. Random variables are usually denoted by capital letters.

If the sample space contains a finite number of points or a sequence with as many elements as there are whole numbers, it is called a discrete sample space. On the other hand, if the elements of the sample space are infinite in number or as many as the number of points on a line segment, we say that we have continuous sample space. A random variable defined over a discrete sample space is called a discrete random variable while over a continuous sample space is called a continuous random variable. In most practical problems, continuous random variables represent measured data and discrete random variables represent count data such as the number of defectives in a sample of k items or the number of accidents per year.

2.6 SOME DISCRETE AND CONTINUOUS PROBABILITY DISTRIBUTIONS

The distribution function $F(.)$ of a numerical valued random phenomenon is defined as having its value, at any real number x, the probability that an observed value of the random phenomenon will be less than or equal to the number x. In symbols, for any real number x,

$$F(x) = P[\text{real numbers } x':x' \leq x].$$

If the probability function is specified by a probability mass function $p(.)$, then the distribution function $F(.)$ for any real number x, called discrete distribution function, is given by

$$F(x) = \Sigma\, p(x'), \text{ for any real number } x,$$

points $x' \leq x$ such that $p(x') > 0$.

If the probability function is specified by a probability density function $f(.)$, then the distribution function $F(.)$ for any real number x, called continuous distribution function, is given by

$$F(x) = \int_{-\infty}^{x} f(x')dx', \text{ for any real number } x.$$

Most of the distribution functions arising in practice are either discrete or continuous. Also, there are distribution functions which are neither discrete nor continuous. Such distribution functions are called *mixed*. A distribution function $F(.)$ is called mixed if it can be written as a linear combination of two distribution functions, namely, discrete and continuous denoted by $F_d(.)$ and $F_c(.)$ respectively in the following way:

$F(x) = C_1F_d(x) + C_2F_c(x)$, for any real number x in which C_1 and C_2 are constants between 0 and λ, whose sum is one.

It is to be noticed that $p(x)$ is a probability mass function for a discrete random variable X if, for any real number x,

(i) $p(x) = p(X = x)$,

(ii) $p(x) \geq 0$, and

(iii) $\sum_{x} p(x) = 1$

Similarly, $f(x)$ is a probability density function for a continuous random variable X, if for any real number x

(i) $P(a < x < b) = \int_{a}^{b} f(x)dx$,

(ii) $f(x) \geq 0$, and

(iii) $\int_{-\infty}^{\infty} f(x)dx = 1$.

If instead of one random variable X, we have two random variables X and Y, the probability mass function for their simultaneous occurrence is represented by $p(x, y)$ and is called joint probability mass function of X and Y. $p(x, y)$ is called a joint probability mass function for discrete random variables X and Y, if, for all real numbers x and y,

(i) $p(x, y) = P[X = x, Y = y]$,

(ii) $p(x, y) \geq 0$, and

(iii) $\sum_x \sum_y p(x, y) = 1$,

and is called a joint probability density function $f(x, y)$ for continuous random variables X and Y, if, for all real numbers x and y,

(i) $P[a < x < b, c < y < d] = \int_c^d \int_a^b f(x, y)dx\, dy$

(ii) $f(x, y) \geq 0$, and

(iii) $\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x, y)dx\, dy = 1$.

We now discuss some important probability distributions which can describe the behaviour of most of the random variables encountered in practice.

(a) ***Uniform distribution:*** The simplest of all discrete probability distributions is one where the random variable assumes all its values with equal probability. This distribution is called uniform probability distribution.

If a random variable X assumes the values $x_1, x_2, \ldots x_k$ with equal probability, then the discrete uniform distribution has the probability mass function given by

$$p(x, k) = 1/k,\ x = x_1, x_2, \ldots, x_k. \qquad (2.6.1)$$

(b) ***Binomial distribution:*** We call an experiment a binomial experiment if it has the following properties:

(i) It consists of n repeated trials.

(ii) Each trial results in an outcome that may be classified as a success or a failure.

(iii) The probability of success remains constant from trial to trial.

(iv) The repeated trials are independent.

The number X of successes in n trials of a binomial experiment is called a binomial random variable. The probability distribution of the binomial variable X is called binomial distribution, generally denoted by $b(x; n, p)$ since its value depends upon the number of trials and the probability of success on a given trial.

To derive an expression for probability mass function, we consider an experiment for exactly x successes and $n - x$ failures in a specified order. Each success occurs with probability p and each failure with probability $q = 1 - p$. Since all the trials are independent, the probability for x successes in a specified order from n trials is $p^x q^{n-x}$. The possible number of ways in which x successes can occur from n trials is $\binom{n}{x}$. Since all these possible ways of getting x successes are mutually exclusive, each having a probability $p^x q^{n-x}$, the probability of getting exactly x successes in a series of n independent trials is given by

$$b(x;\, n,\, p) = \begin{cases} \binom{n}{x}\, p^x\, q^{n-x}, & \text{for } x = 0, 1, \ldots, n \\ 0, & \text{otherwise} \end{cases} \tag{2.6.2}$$

$b(x;\, n,\, p)$ given by (2.6.2) is the probability mass function for binomial distribution. The sum of the probabilities

$$\sum_{x=0}^{n} b(x;\, n,\, p) = \sum_{x=0}^{n} \binom{n}{x}\, p^x\, q^{n-x} = (q + p)^n = 1.$$

The binomial distribution contains two independent constants, n and p(or q). These are called parameters of the binomial distribution. If $p = q = \frac{1}{2}$, the binomial distribution is symmetrical distribution and when $p \neq q$, it is a skew distribution.

(c) ***Poisson distribution:*** An experiment having the following properties is called a Poisson experiment.

(i) The number of successes occurring in one time interval or specified region are independent of those occurring in any other disjoint interval or region.

(ii) The probability of a single success occurring during a very short time interval or in a small region is proportional to the length of time interval or the size of the region and does not depend on the number of successes occurring outside this time interval or region.

(iii) The probability of more than one success occurring in such an interval or fall in such a small region is negligible.

The number X of successes in a Poisson experiment is called a Poisson random variable. The probability distribution of the Poisson variable X is called the Poisson distribution and is generally denoted by $p(x;\, \lambda)$.

Poisson distribution is the limiting form of binomial distribution when p (or q) is very small and n is large but the average number of successes $np($ or $nq) = \lambda$ is a finite constant. To derive probability mass function $p(x;\, \lambda)$ for Poisson distribution, we proceed as follows from binomial distribution.

The probability of x successes is given by

$$p(x) = \binom{n}{x}\, p^x\, q^{n-x}$$

which can be written as

$$p(x) = \frac{n!}{x!(n-x)!} \left(\frac{\lambda}{n}\right)^x \left(1 - \frac{\lambda}{n}\right)^{n-x} \quad [\because p = \frac{\lambda}{n}]$$

$$= \frac{\lambda^x}{x!}\left(1 - \frac{\lambda}{n}\right)^n \cdot \frac{\left(1 - \frac{\lambda}{n}\right)^{-x}}{n^x} \cdot \frac{n!}{(n-x)!}$$

Taking limit as $n \to \infty$, we have

$$\lim_{n\to\infty} p(x) = \frac{\lambda^x}{x!} e^{-\lambda} \lim_{n\to\infty} \left[\frac{1}{n^x} \cdot \frac{n!}{(n-x)!}\right]$$

Multiplying the two probabilities, we get

$$p(x = k) = {}^{x+k-1}C_{k-1}\, p^k q^x,\ x = 0, 1, 2, \ldots, \text{ and } x > 0 \tag{2.6.5}$$

Clearly,

$$\sum_{x=0}^{\infty} p(x = k) = p^k \sum_{x=0}^{\infty} {}^{x+k-1}C_{k-1}\, q^x\, p^k (1 - q)^{-k} = 1.$$

Since

$${}^{x+k-1}C_{k-1} = \frac{(k + x - 1)(k + x - 2)\ \ldots k}{x!} = (-1)^x \cdot {}^{-k}C_x$$

where,

$${}^{-k}C_x = \frac{(-k)(-k - 1)\ \ \ldots(-k - x + 1)}{x!} (-1)^x \cdot {}^{x+k-1}C_{k-1}$$

$p(x)$ can also be written as

$$p(x) = {}^{-k}C_x\, p^k (-q)^x,\ x = 0, 1, 2, \ldots \tag{2.6.6}$$

The probability distribution given by (2.6.6) is called *Pascal's distribution*. It has two parameters p and k. If $k = 1$ (2.6.6) reduces to the geometric distribution.

In the above distribution, k is an integer but the distribution remains meaningful even if k is not an integer but $k \geq 0$. In this case the distribution is called the *negative binomial distribution*. If, we put

$$k = \frac{1}{\beta},\ p = \frac{1}{1+\beta},\ q = \frac{\lambda}{1+\beta}$$

The distribution so obtained with two parameters β and λ is called Polya's distribution.

If we let $p \to 0$ and $k \to \infty$ in such a way that $\lim\limits_{k\to\infty} kp = \lambda$ (a finite constant), then (2.6.6) takes the form

$$p(x) = \lim_{\substack{k\to\infty \\ p\to 0}} (1 + p)^{-k} \frac{(k + x - 1)\ldots k}{x!} \left(\frac{p}{1+p}\right)^x$$

$$= \lim_{\substack{k\to\infty \\ p\to 0}} \left(1+\frac{\lambda}{k}\right)^{-k} \frac{1}{x!}\left(1+\frac{x-1}{k}\right) \ldots 1.\ \lambda^x \left(1+\frac{\lambda}{k}\right)^{-x}$$

$$= \frac{e^{-\lambda}\lambda^x}{x!}$$

which is the expression for Poisson distribution.

(f) ***Hypergeometric distribution:*** The distribution having

$$p(x) = \frac{{}^{m}C_x\ {}^{n}C_{r-x}}{{}^{m+n}C_r},\ x = 0, 1, 2, \ldots, r;\ r \leq m,\ r \leq n \tag{2.6.7}$$

as the probability law is known as Hypergeometric distribution.

(g) ***Multinomial distribution:*** The binomial distribution can easily be generalized to n repeated independent trials when each trial may result in one of the several outcomes say $E_1, E_2, \ldots, E_k$ with respective probabilities $p_1, p_2, \ldots, p_k$ in each trial where $p_1 + p_2 + \ldots + p_k = 1$.

The probability that n trials will result in E_1 occurring n_1 times, E_2 occurring n_2 times, …, E_k occurring n_k times in a fixed definite order is

$$p_1^{n_1}\, p_2^{n_2} \ldots p_k^{n_k}; \quad \sum_{i=1}^{k} n_i = n.$$

But we are interested in events occurring in any order. The number of mutually exclusive ways in which this can happen is $\dfrac{n!}{n_1, n_2 \ldots n_k}$

Hence, the required probability is

$$p(n_1, n_2, \ldots, n_k) = \frac{n!}{n_1, n_2 \ldots n_k} \cdot p_1^{n_1}\, p_2^{n_2} \ldots p_k^{n_k};$$

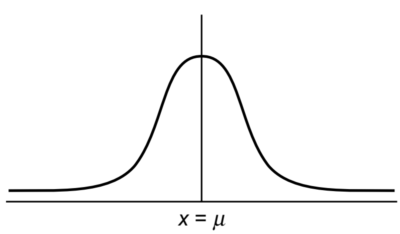
$$\sum_{i=1}^{k} n_i = n \tag{2.6.8}$$

This distribution is called multinomial probability distribution as the expression is the general term of the multinomial expansion of

$$(p_1 + p_2 + \ldots + p_k)^n.$$

This is not univariate distribution but is a multivariate distribution involving k variates $n_1, n_2, \ldots,$ n_{k-1}, n_k but only $(k - 1)$ variants are independent since $\sum_{i=1}^{k} n_i = n$.

(h) ***Normal distribution:*** Normal distribution is one of the most important continuous distributions in the field of statistics. Its graph, called the normal curve, is a bell-shaped curve (fig. below). The mathematical equation of this curve was developed by De Moivre in 1733. This distribution is also referred to as Gaussian distribution after the name of Gauss who also derived its equation for a study of errors in repeated measurements of the same quantity. *A* random variable x having the bell-shaped distribution is called a normal random variable.



Normal curve

The mathematical equation for the probability distribution of the continuous normal variable depends on two parameters μ and σ, its mean and standard deviation. This distribution can be is called a Cauchy variate with parameter λ.

(l) ***Exponential distribution:*** A continuous random variable X is said to be exponentially distributed if it has the probability density function.

$$f(x) = \begin{cases} \dfrac{1}{\beta} \cdot e^{-x/\beta}, x > 0, \beta > 0 \\ 0, \text{elsewhere.} \end{cases} \tag{2.6.14}$$

with β as the parameter.

This distribution plays a key role in the theory of queues and reliability.

(m) ***Erlang distribution:*** A two-parameter generalisation of the exponential distribution is given by the Erlang distribution with density function

$$f(x) = \frac{\lambda k(\lambda k x)^{k-1}}{(k-1)!}, \ e^{-k\lambda x}, \ 0 \le x \le \infty \tag{2.6.15}$$

k is a positive integer. For $k = 1$, we get the exponential distribution.

This distribution also plays an important role in the theory of queues and reliability.

(n) ***Weibull distribution:*** The continuous random variable X has a Weibull distribution with parameters α and β if its density function is given by

$$f(x) = \begin{cases} \alpha\beta\, x^{\beta-1} e^{-\alpha x^{\beta}}, x > 0; \alpha, \beta > 0 \\ 0, x < 0. \end{cases} \tag{2.6.16}$$

This distribution was used by Weibull in 1951 to describe experimentally observed variation in the fatigue resistance of steel, its elastic limits, etc. This distribution is widely used in reliability theory.

(o) ***Chi-Squared distribution:*** The continuous random variable X has a chi-squared distribution with v degrees of freedom if its density function is given by

$$f(x) = \begin{cases} \dfrac{1}{2^{v/2}\sqrt{v/2}}\, x^{\frac{v}{2}-1}\, e^{-\frac{x}{2}}, x > 0 \\ 0, \text{elsewhere} \end{cases} \tag{2.6.17}$$

v is a positive integer. This distribution can be obtained from Gamma distribution putting

$$\alpha = \frac{v}{2}, \ \beta = 2.$$

Before closing the section, we mention here some more results about probability distributions in the form of Appendix.

APPENDIX

(a) The rth moment about any point x' of the discrete probability distribution with probability mass function p_i is given by

(ii) the parameter is discrete or continuous. If the joint distribution of x_{t_1} and x_{t_2} depends on only $t_1 - t_2$, the process is called stationary, otherwise it may be termed evaluationary. A stochastic process is called Gaussian or normal if the joint distribution of any number of variables is a normal distribution.

We now discuss some processes which are useful in the theory of queues and reliability.

(a) ***Markov process:*** If the probability of the system being in a given state at the next trial depends just on its state at present and not upon the states it may have been in earlier times, we call the property of a process as the Markov property. Mathematically, we have

$$P[A_{k+1}|A_k, A_{k-1}, \ldots, A_1] = P[A_{k+1}|A_k]. \tag{2.7.1}$$

This states that at the kth trial the conditional probability of any event A_{k+1}, dependent on the next trial, will not depend on what has happened in past trials but only on what is happening at the present. Sometimes, it is said that the trials have no memory. A process having Markov property is called Markov Process.

Let us observe at n times the state of a system which has r possible states. We number the states 1, 2, …, r (or 0, 1, 2, …, $r - 1$) and let $A_k^{(j)}$ be the event that the system is in state j at time k. If

$$P[A_k^{(j_k)} \mid A_{k-1}^{(j_{k-1})}, \ldots, A_1^{(j_1)}] = P[A_k^{(j_k)} \mid A_{k-1}^{(j_{k-1})}] \tag{2.7.2}$$

holds, we say that the system is a *Markov chain* with r possible states. In other words, it states that at any time the conditional probability of transition from one's present state to any other state does not depend on how one arrived in one's present state.

If the conditional probability is independent of time, the Markov chain is said to be *homogeneous* (or time homogeneous). A state is said to be *absorbing state* if once the system reaches to it and stays there. In such a case the chain is said to be an *absorbing chain*. A chain in which every two states can communicate with each other and in which the system cannot be left is called an *ergodic chain*. A Markov chain with infinite number of states is said to be denumerable or denumerably infinite.

Transition Matrix: Let P_{jk} be the transition probability (probability of transition from state j at nth trial to the state k at $(n + 1)$st trial) satisfying

$$P_{jk} \geq 0, \sum_k p_{jk} = 1 \text{ for all } j.$$

The probabilities may be written in the matrix form

$$P = \begin{bmatrix} p_{11} & p_{12} & \cdot & \cdot & \cdot & \cdot & p_{1k} \\ p_{21} & p_{22} & \cdot & \cdot & \cdot & \cdot & p_{2k} \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ p_{k1} & p_{k2} & \cdot & \cdot & \cdot & \cdot & p_{kk} \end{bmatrix}$$

$$\lim_{h\to 0} \frac{r_0(h)}{h} = \lim_{h\to 0} \frac{r_1(h)}{h} = \lim_{h\to 0} \frac{r_2(h)}{h} = 0$$

Assume that $r_2(h)$ is the probability that in the time interval $(t, t + h)$ the population size will change by two or more. For $n \geq 1$, the event that $X_{t+h} = n$ (n members in population at time $t + h$) can then essentially happen in any one of following three mutually exclusive ways:

(i) the population size at time t is n and undergoes no change in the time interval $(t, t + h)$;
(ii) the population size at time t is $n - 1$ and increases by one in the time interval $(t, t + h)$;
(iii) the population size at time t is $n + 1$ and decreases by one in the time interval $(t, t + h)$.

Now let us define,

$\lambda_n h + r_1 h$ is the conditional probability that the population size will increase by one in the time interval $(t, t + h)$ for $h > 0$, and $\mu_n h + r_o(h)$ is the conditional probability that the population size will decrease by one in the time interval $(t, t + h)$ for $h > 0$, given that the population had size n at time t.

For $n = 0$, the event that $X_{t+h} = 0$ can happen only in ways (i) and (iii).

The events (i), (ii) and (iii) have probabilities

$p_n(t)(1 - \lambda_n h - \mu_n h)$, $p_{n-1}(t)\, \lambda_{n-1} h$ and $P_{n+1}(t)\, \mu_{n+1} h$ respectively.

Consequently, for $n \geq 1$, we get

$$p_n(t + h) = p_n(t)\,(1 - \lambda_n h - \mu_n h) + P_{n-1}(t)\, \lambda_{n-1} h + p_{n+1}(t)\, \mu_{n+1} h \tag{2.7.4}$$

For $n = 0$, we obtain

$$P_o(t + h) = p_o(t)(1 - \lambda_o h) + p_1(t)\, \mu_1 h. \tag{2.7.5}$$

Rearranging (2.7.4) and (2.7.5) and taking limits as h tends to zero, we obtain

$$\frac{d\, p_n(t)}{dt} = -(\lambda_n + \mu_n)p_n(t) + \lambda_{n-1}\, p_{n-1}(t) + \mu_{n+1}\, p_{n+1}(t) \tag{2.7.6}$$

$$\frac{d\, p_o(t)}{dt} = -\lambda_o p_o(t) + \mu_1 p_1(t) \tag{2.7.7}$$

If there is a maximum possible population size N then (2.7.4) holds only for $1 \leq n \leq N - 1$, whereas for $n = N$, we have

$$p_N(t + h) = p_N(t)(1 - \lambda_N h) + p_{N-1}(t)\, \lambda_{N-1} h. \tag{2.7.8}$$

Equations (2.7.6) and (2.7.7) can be solved by existing methods.

Due to its increasing and decreasing character, this process is called birth and death process. This process generally occurs in queueing and reliability problems.