UNIT 6 PROBABILITY DISTRIBUTIONS

STRUCTURE

6.1 INTRODUCTION

In the previous unit you have learnt about sample space, theorems of probability, random variable, its expectation and variance, etc. Very often, the observations generated by different experiments have the same general type of behaviour. As a result, the associated discrete random variables can be described by a single formula. In fact, experience tells us that we need to know only a few important **probability distributions** to describe discrete random variables. We begin with the **binomial distribution -** the simplest of probability distributions. It applies to a wide variety of situations and in different disciplines like medicine, agriculture, biology, military and industrial engineering, etc. Though it arose out of simple coin tossing experiments, it can in particular be used for random walk problem, spin-systems, problems of heredity in genetics and mechanisms of quality control. This is discussed in Sec. 6.2.

You may now logically ask: Can there be any situation not covered by the binomial distribution? Definitely there are many situations which fall outside the domain of this distribution. For instance, the probability distribution describing random processes whose probability of occurrence is **small** and **constant** is known as **Poisson distribution.** The disintegration of radioactive nuclei, emission of photons by excited atoms and the ejection of electrons in photoelectric emission are examples of random processes in physics which obey Poisson distribution. Beyond physics, this distribution occurs in queuing problems such as incoming calls at a telephone exchange or the arrival times of customers at a booking office. In a sense, it is of immense value for problems of operations research and management. As we show, it is a limiting case of binomial distribution. The important features of this distribution are also discussed in Sec. 6.2.

The most important and very widely used continuous probability distribution is the **normal distribution.** It models many phenomena occurring in nature, industry and research. Physical measurements in areas such as meteorological experiments, rainfall studies, errors in scientific measurements etc. are explained with a normal distribution rather well. (It is also referred to as the Gaussian **distribution.)** It is particularly useful in computation. Under certain conditions it provides a good continuous approximation to the binomial distribution. You will learn very interesting facts about normal distribution in Sec. 6.3.

A brief discussion of other continuous distributions of particular use in physics is also included towards the end of this unit.

Objectives

After studying this unit you should be able to

- write down expressions for binomial, Poisson and normal distributions
- calculate the mean and variance of these distributions

- show that Poisson distribution is a limiting case of binomial distribution
- derive the properties of the normal distribution, and
- apply them to practical, problems.

6.2 THE BINOMIAL DISTRIBUTION

A statistical experiment often consists of repeated trials, each with two possible outcomes that may be labelled success and failure. The toss held in a cricket match io decide the batting/fielding first is a familiar example. The skipper calling right is labelled as success. You know that in the gulf war, Iraq fired several Scud missiles over Israel and Saudi Arabia. The desired targets were not hit and we can say that it was a failure. In pharmaceutical work, success and failure of a particular medicine can be identified depending on how a patient responds to it. Similarly, we may consider the random walk problem which manifests itself in Brownian motion, neutron diffusion, etc. A familiar example of random walk problem is the motion of a drunken man. Each step he takes is of unequal length and is so completely under the influence of alcohol that the direction of each step (whether forward of backward) is completely independent of the preceding step. So the characteristic feature of a Bernoulli process is that the trials are repeated and independent; the probability of success (p) remaining constant. The number X of successes inn Benioulli trials is called a binomial random variable. The probability distribution of this discrete random variable X is then called the binomial distribution. We denote its value by $b(x; n, p)$ where p is the constant probability of success.

To illustrate these ideas, let us toss a biased coin n times and calculate the probability of obtaining x heads, where $x = 0, 1, 2, \ldots, n$. Suppose that the probabilities of getting a head 'and a tail in a particular toss are p and q, respectively. Obviously, $p + q = 1$. And the number of heads is a random variable, X_n which takes values $0, 1, 2, \ldots, n$. Since tosses are independent, for $x = n$ we have by the theorem of compound probabilities

$$P(X_n = n) = pp \ldots p = p^n \qquad (6.1a)$$

Similarly,

$$P(X_n = 0) = qq \ldots q = q^n \qquad (6.1b)$$

For $X_n = x$, we must have a sequence of the form

$$\underbrace{HH \ldots H}_{x} \qquad \underbrace{TT \ldots T}_{n-x}$$

Thus, the total number of sequences for x successes and $n-x$ failures is equal to the number of partitions of n outcomes into two groups; x in one group and $n-x$ in the other. This obviously is given by $\binom{n}{x}$. Since each success occurs with probability p and each failure with probability $q = 1-p$, the probability for each sequence is $p^x q^{n-x}$. Hence for given values of n and p, the number of successes is given by the theorem of total probability:

$$b(x; n, p) = \binom{n}{x} p^x q^{n-x} \qquad x = 0, 1, 2, \ldots, n \qquad (6.2)$$

This distribution is known as binomial distribution. We may therefore conclude as follows:

> **Binomial** distribution: A Bernoulli trial can result in a success with probability p and a failure with probability $q = 1-p$. Then, the probability distribution of the binomial random variable X representing the **number** of Successes in n independent trials is given by
>
> $$b(x; n, p) = \binom{n}{x} p^x q^{n-x} \qquad x = 0, 1, 2, \ldots n$$

You would note that this distribution derives its name from the fact that $b(x; n, p)$ is the coefficient of s^x in the binomial expansion of $(q + ps)^n$, where s is an auxiliary variable. In other words, the $(n+1)$ th term in the binomial expansion of $(q+ps)^n$ corresponds to the various values of $b(x; n, p)$ for $x = 0, 1, 2, ..., n$. That is,

$$(q+ps)^n = \binom{n}{0} q^n + \binom{n}{1} ps\, q^{n-1} + \binom{n}{2} p^2 s^2 q^{n-2} + \dots$$
$$+ \binom{n}{x} p^x s^x q^{n-x} + \dots + \binom{n}{n} p^n s^n$$
$$= b(0; n, p) + b(1; n, p) + \dots b(x; n, p) + \dots + b(n; n, p)$$

Since $p + q = 1$, we find that $\sum_{x=0}^{\infty} b(x; n, p) = 1.$ (6.3)

We have illustrated these ideas in the following examples.

Example 1

Calculate the probability of getting x heads ($x = 0, 1, 2, 3, 4, 5, 6$) when an unbiased coin is tossed six times.

Solution

We have $p = q = \frac{1}{2}$, and $n = 6$. So Eq. (6.2) gives

$$b\left(x; 6, \frac{1}{2}\right) = \binom{6}{x}\left(\frac{1}{2}\right)^6$$

This yields $b\left(6; 6, \frac{1}{2}\right) = \frac{1}{64}, b\left(1; 6, \frac{1}{2}\right) = b\left(5; 6, \frac{1}{2}\right) = \frac{6}{64} = \frac{3}{32}$

$$b\left(2; 6, \frac{1}{2}\right) = b\left(4; 6, \frac{1}{2}\right) = \frac{15}{64}, \text{ and } b\left(3; 6, \frac{1}{2}\right) = \frac{20}{64} = \frac{5}{16}$$

The distribution is plotted in Fig. 6.1.

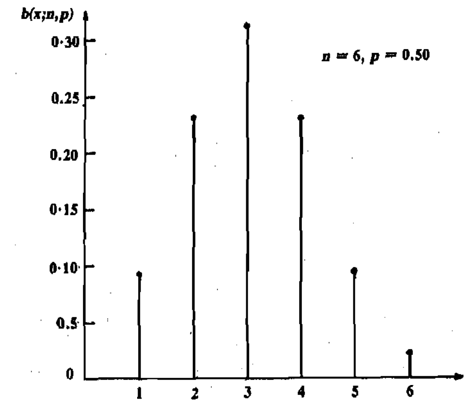


Fig. 6.1: The binomial distribution for $p = q = 0.5$ and $n = 6$

Example 2

In a precision-bombing attack, there is a 50% chance that any one bomb will strike the target At least two direct hits are needed to destroy the target completely. Calculate the minimum number of bombs required to completely destroy the target?

Solution

$$\text{Here, } p = \frac{1}{2}, \; q = \frac{1}{2}$$

Let the number of bombs required to completely destroy the target be n. Since at least two direct hits are needed to destroy the target completely, we have

$$\sum_{x=2}^{\infty} \binom{n}{x} \left(\frac{1}{2}\right)^n \geq \frac{99.9}{100}$$

Since

$$\sum_{x=0}^{\infty} \binom{n}{x} = 2^n = 1 + n + \sum_{n=2}^{\infty} \binom{n}{x}$$

the above inequality can be rewritten as

$$\frac{2^n - 1 - n}{2^n} \geq \frac{99.9}{100}$$

or

$$f(n) = [n \log 2 - \log (1 + n)] \geq 3$$

This inequality can be solved directly or graphically. A plot of this relation is shown in Fig. 6.2. You will note that the minimum value of n which satisfies the above inequality lies between 13 and 14. Hence, n = 14.

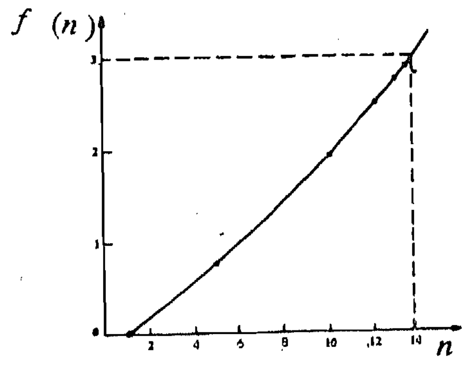


Fig. 6.2: Graph of [$n \log 2 - \log (1 + n)$] versus n

Example 3

A drunkard starts out from a lamp-post in the middle of a large city-square. Be is so completely under the influence of alcohol that the direction of each step (whether it is to the right or to the left) is completely independent of the preceding step. Suppose that each time the man takes a step, the probability of its being to the right is p, while the probability of its being to the left $q = 1 - p$. After taking N such steps the man could be at a distance ml $(-N \leq m \leq N)$ from the starting point. What is the probability W (m, N) that the man arrives at ml after N steps?

Solution

This is an example of random walk problem which arises very frequently in physics. You have studied it in connection with diffusion process and Brownian motion in the course PHE-06 on thermodynamics and Statistical Mechanics. Let us choose the x - axis along the road so that $x = 0$ represents the starting position (see Fig. 6.3). Since the successive steps are statistically independent of each other, they constitute a sequence of Bernoulli trials with probabilities of "success" (a step to the right) and "failure" (a step to the left) p and q, respectively, on each trial.

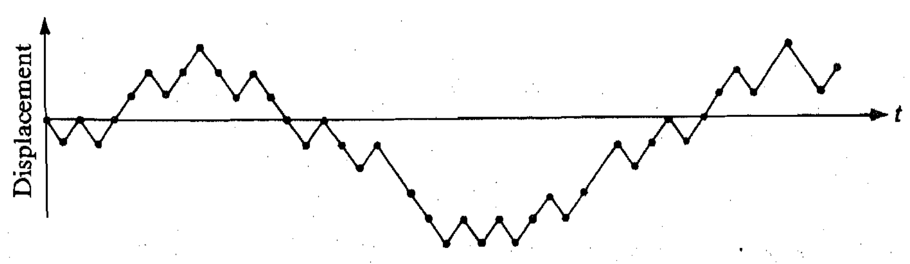


Fig. 6.3: Diagrammatic representation of one-dimeusionid random-walk

Let n denote the total number of steps to the right. Hence, the total number of steps to the left is $(N - n)$. Since the net displacement (measured to the right in units of l) is m, we have

$$m = n-(N-n) = 2n-N$$

Therefore,

$$n = \frac{m+N}{2}$$

Hence, the number of "**successes**" is $(m + N)/2$ which, of course, is always an integer. Then Eq. (6.2) gives

$$b\left(\frac{m+N}{2}; N, p\right) = W(m,N) = \binom{N}{\frac{m+N}{2}} p^{\frac{m+N}{2}} q^{\frac{N-m}{2}}$$

or

$$W(m,N) = \frac{N!}{\left(\frac{N+m}{2}\right)!\left(\frac{N-m}{2}\right)!} = p^{\frac{N+m}{2}} q^{\frac{N-m}{2}}$$

You can now test your understanding by attempting the following SAQ.

Spend 15 min

SAQ 1

(i) In an objective **type** examination for recruitment in a bank, *12* questions are true-false type. Calculate the probability of guessing at least 6 correct answers.

(ii) The probability that a certain component survives a given shock is **3/4.** Calculate the probability that 2 of the next *4* components tested survive.

(iii) The probability that a person recovers from a serious disease is ***0.40.*** What is the probability that **atleast** one of the eight persons admitted in a hospital survives?

From our discussion so far it should be clear to you that the probability distribution of a binomial random variable depends only on the values of n, p and q. It therefore seems reasonable to assume that the mean and variance of a binomial random variable also depend on the values of these parameters, To investigate whether or not this is true, let us calculate the mean and variance of the binomial distribution. We recall that

$$E(X) = \bar{X} = \sum_{x=0}^{n} x\, b(x;n,p) = \sum_{x=0}^{n} x \binom{n}{x} p^x q^{n-x}$$

where we have substituted for $b(x; n, p)$ from **Eq.(6.2).** You will note that **the term** $x = 0$ does not contribute in this case. In an expanded form, we can rewrite it as

$$E(X) = \binom{n}{1} p q^{n-1} + 2\binom{n}{2} p^2 q^{n-2} + 3\binom{n}{3} p^3 q^{n-3}$$

$$+ 4\binom{n}{4} p^4 q^{n-4} + \dots + n\binom{n}{n} p^n$$

$$= \frac{n!}{(n-1)!} p q^{n-1} + 2\frac{n!}{2!(n-2)!} p^2 q^{n-2} + 3\frac{n!}{(n-3)!\,3!} p^3 q^{n-3}$$

$$+ 4\frac{n!}{(n-4)!\,4!} p^4 q^{n-4} + \dots + np^n$$

$$= npq^{n-1} + n(n-1)p^2q^{n-2} + \frac{n(n-1)(n-2)}{2!} p^3 q^{n-3}$$

$$+ \frac{n(n-1)(n-2)(n-3)}{3!} p^4 q^{n-4} + \dots + np^n$$

$$= np\left[q^{n-1} + (n-1)\,p\,q^{n-2} + \frac{(n-1)(n-2)}{2!}p^2q^{n-3} + \frac{(n-1)(n-2)(n-3)}{3!}p^3q^{n-4} + \dots p^{n-1}\right]$$

$$= np\,(q+p)^{n-1}$$

$$= np \qquad (6.4)$$

since $p + q = 1$.

In the compact form, we can evaluate this sum by using the relations

$$n! = n\{(n-1)!\}$$

and

$$x/x! = 1/(x-1)!$$

Then

$$\overline{X} = np \sum_{x=1}^{n} \frac{(n-1)!}{(x-1)!\,(n-x)!} p^{x-1} q^{n-x}$$

$$= np \sum_{x=1}^{n} \frac{q^{n-1-(x-1)} p^{x-1} (n-1)!}{(x-1)!\,\{n-1-(x-1)\}!}$$

$$= np \sum_{r=0}^{n-1} \binom{n-1}{r} p^r q^{n-1-r}$$

where we have put $r = x - 1$. The summation term denotes binomial expansion of $(p+q)^{n-1}$, which is equal to one, so that

$$\overline{X} = np$$

The variance of X is given by

$$\text{Var}\,(X) = \sigma_X^2 = E\,(X^2) - [E\,(X)]^2$$

$$= \sum_{x=0} x^2\, b\,(x\,;n,p) - n^2p^2 \qquad (6.5)$$

To begin with the calculation, we consider the first term on RHS:

$$\sum_{x=0}^{n} x^2 b\,(x;n,p) = \sum_{x=0}^{n} x^2 \binom{n}{x} p^x q^{n-x} = \sum_{x=0}^{n} [\,x + x\,(x-1)\,] \binom{n}{x} p^x q^{n-x}$$

where we have written $x^2 = x + x\,(x-1)$. Proceeding further we note that

$$\sum_{x=0}^{\infty} x^2 \binom{n}{x} p^x q^{n-x} = \sum_{x=1}^{\infty} x \binom{n}{x} p^x q^{n-x} + \sum_{x=2}^{\infty} x\,(x-1) \binom{n}{x} p^x q^{n-x} \qquad (6.6)$$

since the term corresponding to $x = 0$ vanishes in the first summation, whereas the terms corresponding to $x = 0$, and 1 vanish in the second summation.

You would recognise that the first term on RHS signifies $E(X)$ and Eq. (6.4) tells us that it is equal to np. By expanding the second term on RHS and following the procedure used to arrive at Eq. (6.4) step by step, you can easily show that

$$\sum_{x=2}^{n} x\,(x-1) \binom{n}{x} p^2 q^{n-x} = n\,(n-1)\,p^2 \qquad (6.7)$$

SAQ 2

Spend 5 min

Prove the result contained in Eq. (6.7).

Inserting these results in Eq. (6.6), we find that

$$\sum_{x=0}^{n} x^2 \binom{n}{x} p^x q^{n-x} = np + n\,(n-1)\,p^2$$

$$= n^2p^2 + np\,(1-p)$$

$$= n^2p^2 + npq$$

On substituting this result in Eq. (6.5) we get the desired expression for Var (X)

$$\text{Var}\,(X) = n^2p^2 + npq - n^2p^2$$

$$= npq \qquad (6.8)$$

The standard deviation of a binomial distribution is therefore $\sqrt{npq}$. We may now conclude that

> The mean and variance of the binomial distribution $b(x\,;n,p)$ are
> $E(X) = np$ and $\text{Var}(X) = npq$

Let us now calculate mean and variance for the random variable of Example 1:

$$E(X) = (6)\times\left(\frac{1}{2}\right) = 3 \quad \text{and} \quad \mathrm{Var}(X) = (6)\times\left(\frac{1}{2}\right)\times\left(\frac{1}{2}\right) = \frac{3}{2}$$

We now illustrate it for a physical problem. You will get an opportunity to consider it in detail in Block 4 of **PHE-06** course.

Example 4

Consider a system of N noninteracting particles enclosed in a box of volume V. Imagine the box to be divided into two compartments of volume V_1 and V_2, as shown in Fig. 6.4. We now pose the question: What is the probability that x out of N particles are found in the volume V_1?

Solution

In the context of this problem, "success" means that a particle is found in the volume V_1 and "failure" means that it is found in V_2. Since, for a systern in thermal equilibrium, a particle is

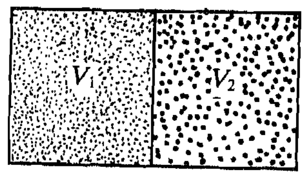


Fig. 6.4: Box of volume V divided into two compartments of volumes V_1 and V_2

equally likely to be anywhere in the volume V, we have

$$p = \frac{V_1}{V}, q = \frac{V_2}{V} = 1 - \frac{V_1}{V} \qquad \text{(i)}$$

According to Eq. (6.2) the required probability is given by

$$b(x; N, p) = \binom{N}{x} p^x q^{N-x} \qquad \text{(ii)}$$

Using Eqs. (6.4) and (6.5) you can easily verify that

$$\bar{x} = Np = N\frac{V_1}{V}$$

$$\sigma_x^2 = Npq = N\frac{V_1}{V}\left(1 - \frac{V_1}{V}\right) \qquad \text{(iii)}$$

and

$$\frac{\sigma_x}{\bar{x}} = \frac{1}{\sqrt{N}}\left(\frac{q}{p}\right)^{1/2} = \frac{1}{\sqrt{N}}\left(\frac{V}{V_1} - 1\right)^{1/2} \qquad \text{(iv)}$$

It means that σ_x/x is negligible for macroscopic systems. For the special case where $V_1 = V_2 = V/2$, we have

$$b(x; N, p) = \binom{N}{x}\left(\frac{1}{2}\right)^N.$$

You can readily show that the most probable distribution corresponds to $x = N/2$. The probability that there is no particle in the left half is

$$b\left(0; N, \frac{1}{2}\right) = \left(\frac{1}{2}\right)^N$$

Clearly, for macroscopic systems, this probability is vanishingly small. For example, $N \sim 10^{24}$ for a mole of a substance and

$$b\left(0; N, \frac{1}{2}\right) = \left(\frac{1}{2}\right)^{10^{24}} = (10)^{-3.10\times10^{23}}$$

This result brings out the statistical nature of thermodynamics in that events disallowed by the laws of thermodynamics (e.g., decrease in total entropy) are not physically impossible, but only highly improbable.

We now know that binomial distribution applies whenever each of the n repeated, independent trials has only two outcomes and the probability of success remains constant. But in many applications more than two outcomes are possible. In such cases binomial distribution becomes multinomial distribution, which we will not discuss here. However, a limiting case of binomial distribution for large n ($\to \infty$), small p ($\to 0$) and moderate $E(X)$ is the Poisson distribution. We now discuss it in some detail.

6.2.1 The Poisson Distribution

To obtain Poisson distribution from binomial distribution, we introduce a parameter $m = np$, which essentially is the mean of binomial distribution. Then Eq. (6.2) takes the form

$$b(x; n, p) = \frac{n!}{(n-x)!\,x!}\left(\frac{m}{n}\right)^x\left(1-\frac{m}{n}\right)^{n-x}$$

$$= \frac{m^x}{x!}\left(1-\frac{m}{n}\right)^n \frac{n!}{(n-x)!\,n^x\left(1-\frac{m}{n}\right)^x} \qquad (6.9)$$

You can easily check that

$$\lim_{n\to\infty}\left(1-\frac{m}{n}\right)^n = e^{-m}$$

Using this result in Eq. (6.9) we find that

$$\lim_{\substack{n\to\infty \\ np-\text{constant}}} b(n; x, p) = \frac{e^{-m}m^x}{x!}\left\{\frac{n!}{\lim\limits_{\substack{n\to\infty \\ np=\text{constant}}} (n-x)!\,n^x\left(1-\frac{m}{n}\right)^x}\right\} \qquad (6.10)$$

Using stirling's formula

$$n! = n^n e^{-n}\sqrt{2\pi n}$$

we can prove that the expression within the curly brackets is equal to one (TQ1). Thus in the limit $n\to\infty$, the binomial distribution takes the form

$$\lim_{\substack{n\to\infty \\ np=\text{constant}}} b(n; x, p) = p(x; m) = \frac{e^{-m}m^x}{x!} \qquad (6.11)$$

This distribution is called the Poisson distribution. The sum of probabilities

$$\sum_{x=0}^{\infty} p(x; m) = e^{-m}\sum_{x=0}^{\infty}\frac{m^x}{x!} = e^{-m}e^m = 1 \qquad (6.12)$$

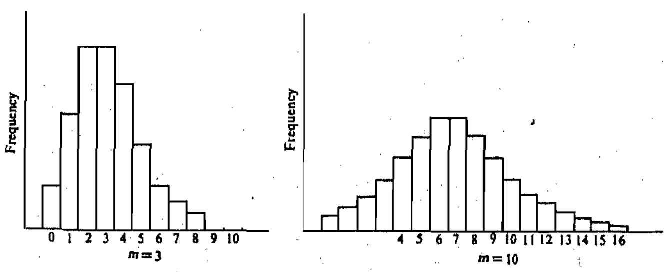


Fig. 6.5: Typical plots of Poisson distribution

For $m = 3$ and $m = 10$, the Poisson distributions are shown in Fig. 6.5.

Example 5

Between 2 p.m. and 4 **p.m.,** the average **number** of **phone** calls received at a telephone exchange per minute is 3. Calculate the probability that **during** one minute, chosen at **random,** there will be no incoming phone call.

Solution

From Eq. (6.11) we have

$$p(0; 3) = e^{-3} = \mathbf{0.0398}$$

Let us now calculate the mean **and** variance of this distribution. By definition,

$$E(X) = \sum_{x=0}^{\infty} x\,p(x;m) = \sum_{x=0}^{\infty} x \frac{e^{-m} m^x}{x!} = m e^{-m} \sum_{x=1}^{\infty} \frac{m^{x-1}}{(x-1)!}$$

since $x/x! = 1/(x-1)!$ and $x = 0$ does not yield an acceptable value. You will recall that the summation term simply defines exp *(m)*. (You can easily check this.) Substituting this **result** in the expression for $E(X)$, we get

$$E(X) = me^{-m}e^{m}$$

$$= m \qquad (6.13)$$

which proves our statement that parameter m signifies the mean of the distribution.

The variance of the Poisson **distribution can** be winputed easily by noting that

$$\mathrm{Var}(X) = \sum_{x=0}^{\infty} x^2 p(x;m) - [E(X)]^2 = \sum_{x=0}^{\infty} x^2 \frac{e^{-m} m^x}{x!} - m^2 \qquad (6.14)$$

The sum in the first **term** on the right hand side can be rewritten as

$$\sum_{x=0}^{\infty} [x + x(x-1)] \frac{e^{-m} m^x}{x!} = \sum_{x=0}^{\infty} \frac{x e^{-m} m^x}{x!} + e^{-m} \sum_{x=0}^{\infty} x(x-1) \frac{e^{-m} m^x}{x!}$$

$$= m + e^{-m} m^2 \sum_{x=2}^{\infty} \frac{m^{x-2}}{(x-2)!}$$

$$= m + m^2 e^{-m} e^{m}$$

$$= m(1+m) \qquad (6.15)$$

Hence

$$\mathrm{Var}(X) = m + m^2 - m^2 = m \qquad (6.16)$$

and **the** standard deviation is $\sqrt{m}$.

Example 6

During the second World war, the number of bomb-hits recorded in each of 576 small areas (of 0.25 sq. km. each) in the **south** of London are recorded below:

x_i	0	1	2	3	4	≥ 5
$f(x_i)$	229	211	93	35	7	1

Does it fit a Poisson distribution?

Solution

We have $m = \frac{537}{576} = \mathbf{0.9323}$

The theoretical Poisson frequency distribution is given by

$$np(x\,;m) = 576\,p\,(x\,;0.9323)$$

For $x = 0, 1, 2, 3, 4,$ and 5 the values are 226.74, 211.39, 98.54, 30.62, 7.14 and 1.57, respectively. As you can see, these are in excellent agreement with the observed values.

Example 7

Tests are made on electrical contacts to examine how often the contacts fail to interrupt the circuit due to their welding together. The results are tabulated below:

x_i (number of welds per test)	Frequency (f_i)	$np(x\,;1.8)$
0	10	8.3
1	13	14.9
2	13	13.4
3	8	8.0
4	4	3.6
5	2	1.3

How this data fits a Poisson distribution? Calculate its variance.

Solution

The mean number of welds per test = $m = \dfrac{89}{50} = 1.78 \approx 1.8$

The theoretical Poisson frequencies are given by

$$np(x;m) = 50\,p\,(m;1.8) = 50\,e^{-1.8}\,\frac{(1.8)^x}{x\,!}$$

These are listed in the last column of the above table.

To calculate the variance., we proceed as follows: Let us take the working mean a = 2. Next, we construct the following table:

x_i	$x_i - a = \eta_i$	f_i	$f_i\eta_i$	$f_i\eta_i^2$
0	−2	10	−20	40
1	−1	13	−13	13
2	0	13	0	0
3	1	8	8	8
4	2	4	8	16
5	3	2	6	18
		$\sum f_i = 50$	$\sum f_i\eta_i = -11$	$\sum f_i\eta_i^2 = 95$

Hence

$$\sigma_x^2 = \overline{X^2} - (\overline{X})^2 = \frac{95}{53} - \left(\frac{11}{50}\right)^2 = 1.85$$

Hence, the mean and variance are approximately equal.

A very interesting example of the Poisson distribution occurs in radioactive decay and cosmic radiations. You may have read about Lord Rutherford's pioneering experiments on a-particle scattering which led to the nuclear model of atom. The number of a-particles striking a counter, in a given interval of time, is a random variable, and follows a Poisson distribution (TQ2). Prof. Feller has so aptly summarised the importance of Poisson distribution:

"Stars in space, raisins in cake, weed seeds among grass seeds, flaws in materials, animal litters in fields are distributed in accordance with the Poisson law."

Before we close this discussion, it is instructive to compare the binomial and Poisson distributions. Form = 2, the plots for four different values of n and p are shown in Fig. 6.6. You will note that Poisson distribution fast approaches the binomial distribution as n increases or p decreases,

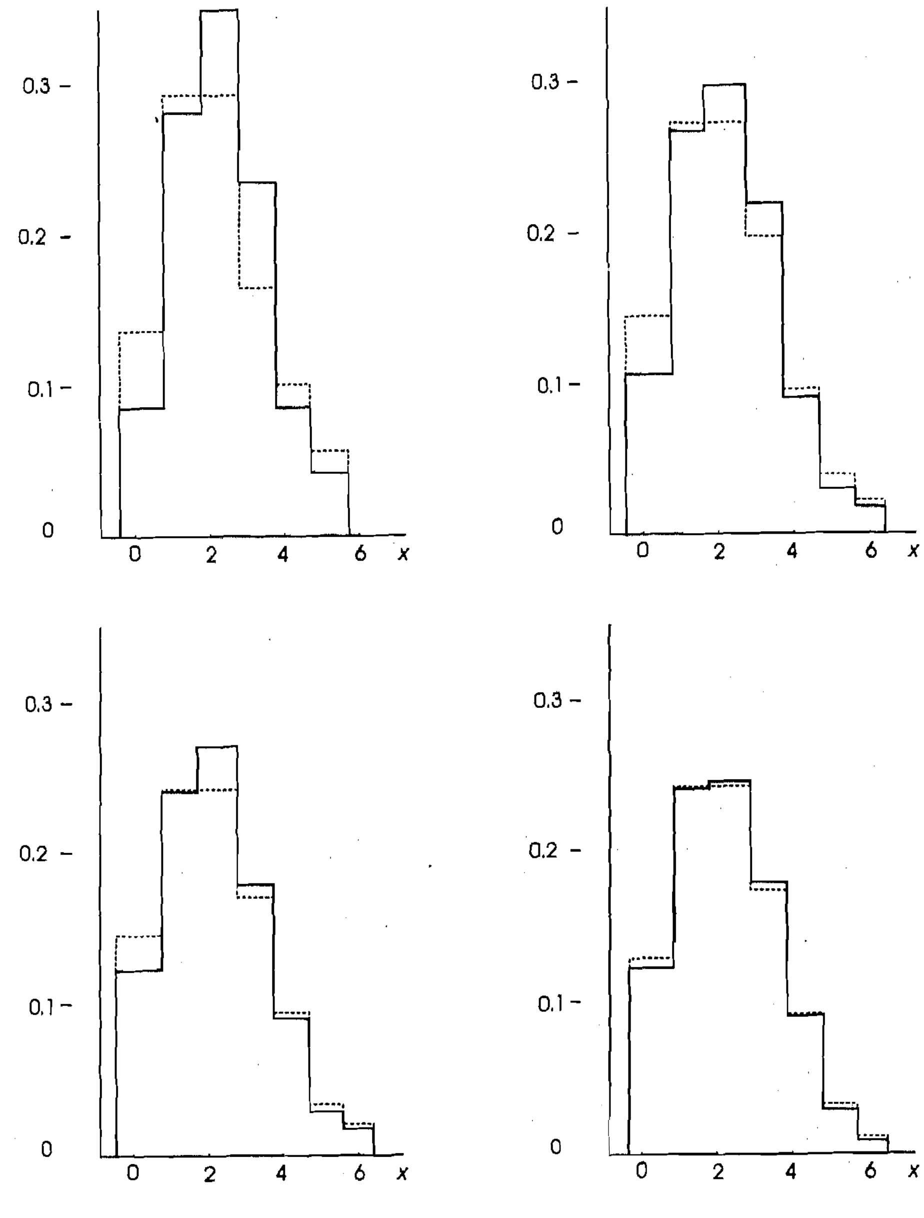


Fig. 6.6: Comparison of binomial and Pdssion probabilities form = 2, and four different sets of n and p. The continuous line plots show the binomial probability whereas the broken line plots show the Poisson distribution

So far we have discussed discrete probability distributions. Several natural and physical phenomena exhibit continuous distribution. For instance, Laplace and Gauss showed that errors in astronomical measurements are described by normal distribution. The I.Q. score9 and energy fluctuations in a statistical system are also normally distributed. Similarly fqr large samples, the sampling distributions are very nearly normal in many cases. This distribution occupies a unique position in statistical theory and you should understand it thoroughly. This is beautifully summed up in the observation made by A.C. Aitken:

"The role of the normal distribution is not unlike that of the straight line in geometry."

We will discuss it in detail in the following section.

6.3 THE NORMAL DISTRIBUTION

The probability density function for a continuous normal variable X, with mean μ and variance σ^2 is given by

$$n(x;\mu,\sigma) = \frac{1}{\sqrt{2\pi}\,\sigma}\exp\left[-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right] \qquad -\infty < x < \infty \qquad (6.17)$$

You will note that the normal distribution depends upon μ and σ, the mean and standard deviation of normal variable. In Fig. 6.7, we have sketched two normal curves having the same standard deviation but different means. Though these curves are identical in form, they are centred at different positions on the horizontal axis.

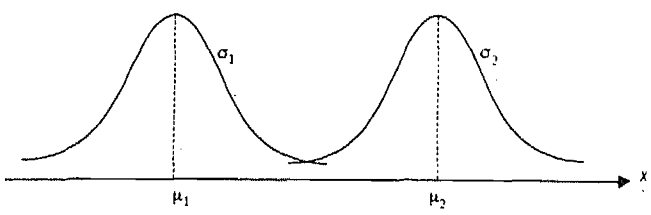


Fig. 6.7 : Normal curves with $\mu_1 < \mu_2$ and $\sigma_1 = \sigma_2$

Let us now sketch two normal curves with the same mean but different standard deviations. Will they be identical and separate? Certainly not. In this case, the two curves are centred at exactly the same position on the horizontal axis. But the curve with larger value of σ is shorter and more spread out (Fig. 6.8).

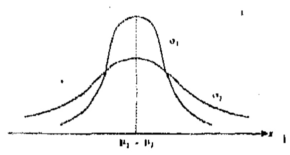


Fig. 6.8 : Normal curves with $\mu_1 = \mu_2$ and $\sigma_1 < \sigma_2$

The total area under a probability curve must be equal to one. Mathematically,

$$\int_{-\infty}^{\infty} n(x;\mu,\sigma)\,dx = 1$$

If you sketch two normal curves having different means and different standard deviations, you will find that they are centred at different positions on the horizontal axis and their shapes reflect different values of σ (Fig. 6.9).

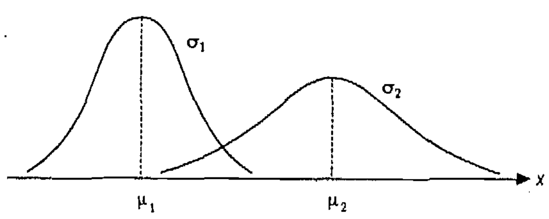


Fig. 6,9 : Normal curves with $\mu_1 < \mu_2$ and $\sigma_1 < \sigma_2$

By examining these figures you will note that (i) the curve is symmetric about the vertical axis through the mean μ, (ii) the curve approaches the horizontal axis asymptotically as we proceed in either direction away from the mean, and (iii) the total area under the curve and above the horizontal axis is equal to one. In practical calculations, it is convenient to eliminate μ and σ from Eq. (6.17). This can be achieved by introducing a variable

$$t = \frac{x-\mu}{\sigma} \qquad (6.18)$$

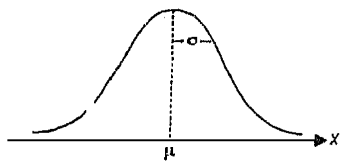
so that $\sigma\,dt = dx$. Now if we write

$$n(t)\,dt = n(x;\mu,\sigma)\,dx$$

it readily follows that Eq. (6.17) takes the form.

$$n(t) = \frac{1}{\sqrt{2\pi}}\exp\left(-\frac{t^2}{2}\right) \qquad -\infty < t < \infty \qquad (6.19)$$

This is the standard form of normal distribution wherein t is normally distributed with mean zero and variance 1. A plot of Eq. (6.19) is shown in Fig. 6.10.



Fig, 6.10: Standard normal curve

Now suppose that a random variable X obeys the normal distribution and we wish to calculate the probability that X lies in the interval $[a, b]$. If we let

$$\frac{a-\mu}{\sigma} = t_1 \quad \text{and} \quad \frac{b-\mu}{\sigma} = t_2$$

then

$$P(a \le x \le b) = P(t_1 \le t \le t_2)$$

$$= \int_{t_1}^{t_2} f(t) \qquad (6.20)$$

signifies the area between the ordinates at $t = t_1$ and $t = t_2$ with mean zero and variance one.

Let us now calculate the mean and variance of the normal distribution. To evaluate the mean, we have

$$E(X) = \frac{1}{\sqrt{2\pi}\,\sigma}\int_{-\infty}^{\infty} x \exp\left[-\frac{(x-\mu)^2}{2\sigma^2}\right] dx \qquad (6.21)$$

Setting $t = \frac{x-\mu}{\sigma}$ we find that $dx = \sigma\, dt$ and $x = \sigma t + \mu$. Then Eq. (6.21) takes the form

$$E(X) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} (\sigma t + \mu) \exp\left(-\frac{t^2}{2}\right) dt$$

$$= \frac{\sigma}{\sqrt{2\pi}}\int_{-\infty}^{\infty} t \exp\left(-\frac{t^2}{2}\right) dt + \frac{\mu}{\sqrt{2\pi}}\int_{-\infty}^{\infty} \exp\left(-\frac{t^2}{2}\right) dt$$

The first integral vanishes since t is an odd function and $\frac{t^2}{2}$ is an even function. So the expression for $E(X)$ assumes a compact form:

$$E(X) = \frac{\mu}{\sqrt{2\pi}}\int_{-\infty}^{\infty} \exp\left(-\frac{t^2}{2}\right) dt$$

The right hand side signifies μ times the area under a normal curve with mean zero and variance one, Hence

$$E(X) = \mu \qquad (6.22)$$

The variance of the normal distribution is given by

$$\mathrm{Var}(X) = \frac{1}{\sqrt{2\pi}\,\sigma}\int_{-\infty}^{\infty} (x-\mu)^2 \exp\left[-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right] dx \qquad (6.23)$$

Again setting $t = \frac{x-\mu}{\sigma}$, we obtain

$$\mathrm{Var}(X) = \frac{\sigma^2}{\sqrt{2\pi}}\int_{-\infty}^{\infty} t\,(t\, e^{-t^2/2})\, dt$$

You will note that $\frac{de^{-t^2/2}}{dt} = -t\, e^{-t^2/2}$. Therefore, while integrating by parts, we take t as differentiating function and $-t\, e^{-t^2/2}$ as integrating factor. This gives

$$\mathrm{Var}(X) = \frac{\sigma^2}{\sqrt{2\pi}}\left[-t\, e^{-t^2/2}\right]_{-\infty}^{\infty} + \int_{-\infty}^{\infty} \exp\left(-\frac{t^2}{2}\right) dt$$

$$= \frac{\sigma^2}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \exp\left(-\frac{t^2}{2}\right) dt$$

From Eq. (6.19) you will note that $\frac{1}{\sqrt{2\pi}} \exp\left(-\frac{t^2}{2}\right)$ is the standard form of the normal curve. Therefore, the integral $\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \exp\left(-\frac{t^2}{2}\right) dt$ represents the total area under the normal curve, which is equal to one. Hence

$$\text{Var}(X) = \sigma^2 \qquad (6.24)$$

In the above sections we have discussed three distributions which describe various natural and physical phenomena. Before we end this unit, we have a quick glance at some other continuous distributions that frequently arise in physical problems. Particular mention may be made about the Maxwell- Boltzmann distribution and the Cauchy distribution.

6.4 SOME CONTINUOUS DISTRIBUTIONS IN PHYSICS

The Maxwell-Boltzmann Distribution

This distribution plays a very important role in kinetic theory of gases. The distribution function for the speeds of gaseous molecules is given by

$$f(v)\,dv = 4\pi \left(\frac{m}{2\pi k_B T}\right)^{3/2} v^2 \cdot \exp\left(-\frac{m v^2}{2k_B T}\right) dv, \text{ for } 0 \le v \le \infty \qquad (6.25)$$

where v denotes the speed of a gas molecule, m its mass, T is the absolute temperature and k_B is the **Boltzmann constant** (1.38×10^{-23} J K^{-1}). It is plotted in Fig. 6.11.

The calculation of the mean and variance of this distribution is the subject of TQ3.

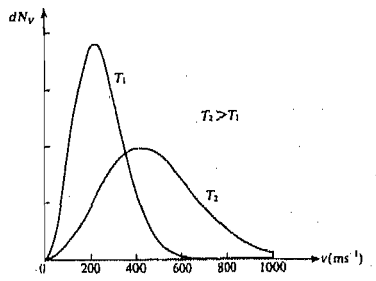


Fig. 6.11:The Maxwell-Boltzmann distribution

The Cauchy Distribution

In this case the distribution function is

$$f(x) = \frac{1}{\pi(1 + x^2)} \qquad -\infty < x < \infty \qquad (6.26)$$

This distribution, shown in Fig. 6.12, occurs in the configuration of atomic spectral lines in classical dispersion theory and NMR studies.

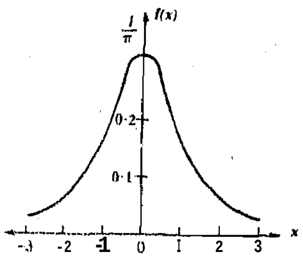


Fig. 6.12: A plot of Cauchy distribution

Let us now summarise what you have learnt in this unit.

6.5 SUMMARY

- The binomial distribution is given by

$$b(x; n, p) = \binom{n}{x} p^x q^{n-x}, \quad x = 0, 1, 2, ..., n$$

- The mean and variance of the binomial distribution are given by

$$E(X) = np \quad \text{and} \quad \text{Var}(X) = npq$$

- When n is large and p is very small, but np is a finite constant, the binomial distribution tends to the Poisson distribution:

$$p(x;m) = \frac{e^{-m} m^x}{x!}, \quad x = 0, 1, 2, \ldots$$

- The mean and variance of the Poisson distribution are given by

$$E(X) = m, \qquad \text{Var}(X) = m$$

- The normal distribution function is given by

$$n(x;\mu,\sigma) = \frac{1}{\sqrt{2\pi}\,\sigma} \exp\left[-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right], \qquad (-\infty \le x \le \infty;\ \sigma > 0)$$

with $E(X) = \mu$ and $\text{Var}(X) = \sigma^2$.

- The standard form of the normal distribution is

$$f(t) = \frac{1}{\sqrt{2\pi}} e^{-t^2/2} \qquad (-\infty \le t \le \infty)$$

where $t = \dfrac{x-\mu}{\sigma}$ is an $N(0, 1)$ variate.

Spend 40 min

6.6 TERMINAL QUESTIONS

1. Show that the expression

$$\frac{n!}{(n-x)!\,n^x\left(1-\dfrac{m}{n}\right)^x}$$

tends to unity in the limit $n \to \infty$, with $np = m$ being of moderate size.

2. Rutherford carried out some pioneering experiments on α-particles in the first two decades of this century. The following table is taken from one of his experiments. It gives the number of α-particles reaching a counter at intervals of 7.5 s each. Fit a Poisson distribution to this data.

Number of α-particles	Number of periods
0	57
1	203
2	383
3	525
4	532
5	408
6	273
7	139
8	45
9	27
10	10
11	4
12	2

3. Calculate $\overline{X}$, Var (X) and the most probable value of X for Maxwell-Boltzmann distribution.

6.7 SOLUTIONS AND ANSWERS

SAQs

1.(i) Here, $p = q = 0.5$, and $n = 12$

Therefore, the required probability is

$$p(6 \le x \le 12) = \sum_{x=6}^{12} \binom{12}{x} \left[\frac{1}{2}\right]^{12}$$

$$= \frac{2510}{4096} = \frac{1255}{2048} = 0.6128$$

(ii) Assuming that the tests are independent and p = 3/4 for each of the four tests, we obtain

$$b\left(2; 4, \frac{3}{4}\right) = \binom{4}{2}\left[\frac{3}{4}\right]^2\left[\frac{1}{4}\right]^2$$

$$= \frac{4!}{2!\,2!} \cdot \frac{3^2}{4^4}$$

$$= \frac{27}{128}$$

(iii) It is quite pertinent to assume that the recovery of patients is independent of each other. Thus we wish to calculate the probability

$$p(X \ge 1) = 1 - p(X = 0)$$

$$= 1 - \binom{8}{0}(0.40)^0(0.60)^8$$

$$= 1 - 0.017 = 0.983$$

2. $$\sum_{x=2}^{n} x(x-1)\binom{n}{x} p^x q^{n-x} = \sum_{x=2}^{n} x(x-1)\frac{n!}{x!(n-x)!} p^x q^{n-x}$$

$$= n(n-1)p^2 \sum_{x=2}^{n} \frac{(n-2)!}{(x-2)!(n-x)!} p^{x-2} q^{n-x}$$

$$= n(n-1)p^2 \sum_{x=2}^{n} \frac{(n-2)!}{(x-2)![(n-2)-(x-2)]!} p^{x-2} q^{n-2-(x-2)}$$

$$= n(n-1)p^2 \sum_{r=0}^{n-2} \binom{n-2}{r} p^r q^{n-2-r}$$

$$= n(n-1)p^2 (q+p)^{n-2}$$

$$= n(n-1)p^2$$

TQs

1. Let $A = \dfrac{n!}{(n-x)!\, n^x \left(1 - \dfrac{m}{n}\right)^x}$

Using Stirling's fonnula

$$n! = n^n e^{-n} \sqrt{2\pi n}$$

we obtain

$$A = \frac{n!}{e^x\left(1-\frac{x}{n}\right)^{n-x+1/2}\left(1-\frac{m}{n}\right)^x}$$

Since x is finite, we have

$$\lim_{n\to\infty} A = \frac{1}{e^x e^{-x}} = 1$$

Here, $p = 0.002$, $n = 2500$, and $m = np = 5$. Hence,

(a) $p(0; 5) = e^{-5} = 0.007$

(b) $$p(0 \le x \le 5) = \sum_{x=0}^{5} \frac{e^{-5} 5^x}{x!}$$

$$= e^{-5}\left[1 + 5 + \frac{25}{2} + \frac{125}{6} + \frac{625}{24} + \frac{3125}{120}\right]$$

$$= 0.007\,[6 + 12.5 + 20.8333 + 26.0417 + 26.0417]$$

$$= 0.007\,[91.3627] = 0.6395$$

2 The variable X (viz. the number of counts) takes on the values x_i with frequencies $f(x_i)$. Hence,

$$m = \frac{\sum_i x_i f(x_i)}{\sum_i f(x_i)} = \frac{10094}{2608} = 3.870$$

The theoretical Poisson frequencies can be computed from

$$np(x; m) = 2608\, e^{-3.87} \frac{(3.87)^x}{x!}$$

We obtain the values 54.4, 210.5, 407.4, 525.5, 508.4, 393.5, 253.8, 140.3, 67.9, 29.2, 11.3, 6.0 and 1.9, respectively. You can see the excellent agreement with the experimental values.

3. We have

$$\bar{v} = \int_0^\infty v f(v)\, dv = 4\pi\left(\frac{m}{2\pi k_B T}\right)^{3/2} \int_0^\infty v^3 \exp\left(-\frac{m v^2}{2 k_B T}\right) dv$$

Let

$$\frac{m v^2}{2k_B T} = y \quad \text{or} \quad v = \sqrt{\frac{2k_B T}{m}}\sqrt{y}$$

Then,

$$\bar{v} = 4\pi\left(\frac{m}{2\pi k_B T}\right)^{3/2} \int_0^\infty \frac{1}{2}\left(\frac{2k_B T}{m}\right)^2 e^{-y} y\, dy = \left(\frac{8k_B T}{\pi m}\right)^{1/2} \int_0^\infty e^{-y} y\, dy$$

Since the integral defines $\Gamma(2)$ which is equal to unity, we have

$$\bar{v} = \sqrt{\frac{8k_B T}{\pi m}}$$

Similarly,

$$\overline{v^2} = \int_0^\infty v^2 f(v)\, dv = \sqrt{\frac{8k_B T}{\pi m}}\sqrt{\frac{2k_B T}{m}}\,\Gamma(5/2)$$

$$= \frac{3k_B T}{m}$$

$$\therefore \quad \mathrm{Var}(v) = \frac{3k_B T}{m} - \frac{8k_B T}{\pi m} = \frac{k_B T}{m}\left(3 - \frac{8}{\pi}\right)$$

To calculate v_p, consider

$$f(v) = A\, v^2 e^{-Bv^2}$$

with

$$A = 4\pi\left(\frac{m}{2\pi k_B T}\right)^{3/2}$$

and

$$B = \frac{m}{2k_B T}$$

Hence,

$$f'(v) = A\,[\,2v\, e^{-Bv^2} + v^2 e^{-Bv^2}(-2Bv)\,]$$

The extrema are given by $f'(v) = 0$. Hence,

$$v_p = \frac{1}{\sqrt{B}} = \sqrt{\frac{2k_B T}{m}}$$

It is easy to check Bat $f''(v) < 0$ for this value of v. Hence, v_p gives the most probable speed.