UNIT 12   GEOMETRIC AND NEGATIVE BINOMIAL DISTRIBUTIONS

Structure

12.1   INTRODUCTION

In Units 9 and 11, we have studied the discrete distributions – Bernoulli, Binomial, Discrete Uniform and Hypergeometric. In each of these distributions, the random variable takes finite number of values. There may also be situations where the discrete random variable assumes countably infinite values. Poisson distribution, wherein discrete random variable takes an indefinite number of values with very low probability of occurrence of event, has already been discussed in Unit 10. Dealing with some more situations where discrete random variable assumes countably infinite values, we, in the present unit, discuss geometric and negative binomial distributions. It is pertinent to mention here that negative binomial distribution is a generalization of geometric distribution. Some instances where these distributions can be applied are "deaths of insects", "number of insect bites".

Like binomial distribution, geometric and negative binomial distributions also have independent trials with constant probability of success in each trial. But, in binomial distribution, the number of trials (n) is fixed whereas in geometric distribution, trials are performed till first success and in negative binomial distribution trials are performed till a certain number of successes.

Secs. 12.2 and 12.3 of this unit discuss geometric and negative binomial distribution, respectively along with their properties.

Objectives

After studing this unit, you would be able to:

- define the geometric and negative binomial distributions;

- calculate the mean and variance of these distributions;

- compute probabilities of events associated with these distributions;

- identify the situations where these distributions can be applied; and

- know about distinguishing features of these distributions like memoryless property of geometric distribution.

12.2 GEOMETRIC DISTRIBUTION

Let us consider Bernoulli trials i.e. independent trials having the constant probability 'p' of success in each trial. Each trial has two possible outcomes – success or failure. Now, suppose the trial is performed repeatedly till we get the success. Let X be the number of failures preceding the first success. Example of such a situation is "tossing a coin until head turns up". X defined above may take the values 0, 1, 2, …. Letting q be the probability of failure in each trial, we have

$P[X = 0] = P$[Zero failure preceding the first success]

$$= P(S)$$

$$= p,$$

$P[X = 1] = P$[One failure preceding the first success]

$$= P[F \cap S]$$

$$= P(F) \, P(S) \ [\because \text{trials are independent}]$$

$$= qp$$

$P[X = 2] = P$[Two failures preceding the first success]

$$= P[F \cap F \cap S]$$

$$= P(F) \, P(F) \, P(S)$$

$$= qqp$$

$$= q^2 p$$

and so on.

Therefore, in general, probability of x failures preceding the first success is

$P[X = x] = q^x p; \ x = 0, 1, 2, 3, ...$

Notice that for $x = 0, 1, 2, 3,...$ the respective probabilities $p, qp, q^2p, q^3p,...$ are the terms of geometric progression series with common ratio q. That is why, the above probability distribution is known as geometric distribution [see Unit 3 of MST-001].

Hence, the above discussion leads to the following definition:

Definition: A random variable X is said to follow geometric distribution if it assumes non-negative integer values and its probability mass function is given by
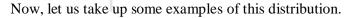
$$P[X = x] = \begin{cases} q^x p \text{ for } x = 0, 1, 2, ... \\ 0, \ \text{otherwise} \end{cases}$$

Notice that

$$\sum_{x=0}^{\infty} q^x p = p + q\, p + q^2 p + q^3 p + ...$$

$$= p[1 + q + q^2 + q^3 + ...]$$

$$= p\left(\frac{1}{1-q}\right) = \frac{p}{p} = 1$$

$\left[\because \text{sum of infinite terms of G.P.} = \frac{a}{1-r} = \frac{1}{1-q} \text{ (see Unit 3 of MST-001)}\right]$

Now, let us take up some examples of this distribution.

Example 1: An unbiased die is cast until 6 appear. What is the probability that it must be cast more than five times?

Solution: Let p be the probability of a success i.e. getting 6 in a throw of the die

$\therefore \ p = \frac{1}{6}$ and $q = 1 - p = \frac{5}{6}$

Let X be the number of failures preceding the first success.

$\therefore$ by geometric distribution,

$P[X = x] = q^x p; \ x = 0, 1, 2, 3, \ldots$

$$= \left(\frac{5}{6}\right)^x \left(\frac{1}{6}\right) \text{ for } x = 0, 1, 2, 3, \ldots$$

Thus, the desired probability = P[The die is to be cast more than five times]

$$= P \text{ [The number of throws is at least 6]}$$

$$= P\left[\begin{array}{l} \text{The number of failures preceding} \\ \text{the first success is at least 5} \end{array}\right]$$

$$= P[X \geq 5]$$

$$= P[X = 5] + P[X = 6] + P[X = 7] + \ldots$$

$$= \left(\frac{5}{6}\right)^5 \left(\frac{1}{6}\right) + \left(\frac{5}{6}\right)^6 \left(\frac{1}{6}\right) + \left(\frac{5}{6}\right)^7 \left(\frac{1}{6}\right) + \ldots$$

$$= \left(\frac{5}{6}\right)^5 \left(\frac{1}{6}\right)\left[1 + \frac{5}{6} + \left(\frac{5}{6}\right)^2 + \ldots\right]$$

$$= \left(\frac{5}{6}\right)^5 \left(\frac{1}{6}\right)\left[\frac{1}{1 - \frac{5}{6}}\right] = \left(\frac{5}{6}\right)^5$$

Let us now discuss some properties of geometric distribution.

Mean and Variance

Mean of the geometric distribution is given as

$\text{Mean} = E(X) = \sum_{x=0}^{\infty} x\, q^x p \ = p \sum_{x=1}^{\infty} x\, q^x \ = p \sum_{x=1}^{\infty} x\, q^{x-1} . q$

$$= pq\sum_{x=1}^{\infty} x\, q^{x-1} = pq\sum_{x=1}^{\infty} \frac{d}{dq}(q^x)$$

> $\dfrac{d}{dq}(q^x)$ is the differentiation of q^x w.r.t. q where x is kept as constant
>
> $[\because \dfrac{d}{dx}(x^m) = mx^{m-1}$, where m is constant (see Unit 6 of MST-001)]

$$= pq\frac{d}{dq}\left[\sum_{x=1}^{\infty} q^x\right] \qquad \left[\begin{array}{l}\because \text{ sum of the derivatives is} \\ \text{the derivatives of the sums}\end{array}\right]$$

$$= pq\frac{d}{dq}\left[\sum_{x=1}^{\infty} q^x\right]$$

$$= pq\frac{d}{dq}\left[q + q^2 + q^3 + ...\right]$$

$$= pq\frac{d}{dq}\left[\frac{q}{1-q}\right]$$

$$= pq\left[\frac{(1-q)-q(-1)}{(1-q)^2}\right] \qquad \left[\begin{array}{l}\text{Applying quotient rule} \\ \text{of differentiation}\end{array}\right]$$

$$= pq\left[\frac{1-q+q}{p^2}\right]$$

$$= \frac{q}{p}. \qquad\qquad ...(1)$$

Variance of the geometric distribution is

$$V(X) = E(X^2) - [E(X)]^2,$$

where

$$E(X^2) = \sum_{x=\infty}^{\infty} x^2 p(x)$$

$$= \sum_{x=0}^{\infty} \left[x(x-1) + x\right] p(x)$$

$[\because x^2 = x(x-1) + x$ (it has already been discussed in Unit 9)]

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

$$= \sum_{x=2}^{\infty} x(x-1)q^x p + \left(\frac{q}{p}\right) \qquad \text{[Using (1) in second term]}$$

$$= pq^2\sum_{x=2}^{\infty} x(x-1)q^{x-2} + \left(\frac{q}{p}\right) \qquad\qquad [\because q^x = q^{x-2}.q^2]$$

$$= pq^2 \sum_{x=2}^{\infty} \frac{d^2}{dq^2}(q^x) + \frac{q}{p} \begin{bmatrix} \because \frac{d^2}{dq^2}(q^x) = \frac{d}{dq}\left(\frac{d}{dq} q^x\right) \\ = \frac{d}{dq}(xq^{x-1}) = x(x-1)q^{x-2} \\ \text{treating x as constant} \end{bmatrix}$$

$$= pq^2 \frac{d^2}{dq^2}\left(\sum_{x=2}^{\infty} q^x\right) + \frac{q}{p}$$

$$= pq^2 \frac{d^2}{dq^2}\left[q^2 + q^3 + q^4 + ...\right] + \frac{q}{p}$$

$$= pq^2 \frac{d^2}{dq^2}\left[\frac{q^2}{1-q}\right] + \frac{q}{p}$$

$$= pq^2 \frac{d}{dq}\left[\frac{(1-q)2q - q^2(-1)}{(1-q)^2}\right] + \frac{q}{p}$$

$$= pq^2 \frac{d}{dq}\left[\frac{2q - 2q^2 + q^2}{(1-q)^2}\right] + \frac{q}{p}$$

$$= pq^2 \frac{d}{dq}\left[\frac{2q - q^2}{(1-q)^2}\right] + \frac{q}{p}$$

$$= pq^2 \left[\frac{(1-q)^2(2-2q) - (2q-q^2).2(1-q)^1(-1)}{(1-q)^4}\right] + \frac{q}{p}$$

$$= pq^2 \left[\frac{(1-q)\{2(1-q)^2 + 2(2q-q^2)\}}{(1-q)^4}\right] + \frac{q}{p}$$

$$= pq^2 . \frac{p\left[2p^2 + 2q(2-q)\right]}{p^4} + \frac{q}{p} \qquad \text{as } p = 1 - q$$

$$= 2\left(\frac{q}{p}\right)^2 \left[p^2 + q(2-q)\right] + \frac{q}{p}$$

$$= 2\left(\frac{q}{p}\right)^2 \left[(1-q)^2 + q(2-q)\right] + \frac{q}{p}$$

$$= 2\left(\frac{q}{p}\right)^2 \left[1 + q^2 - 2q + 2q - q^2\right] + \frac{q}{p}$$

$$= 2\frac{q^2}{p^2}(1) + \frac{q}{p} = \frac{2q^2}{p^2} + \frac{q}{p}$$

$$\therefore \quad V(X) = E(X^2) - [E(X)]^2$$

$$= \frac{2q^2}{p^2} + \frac{q}{p} - \left(\frac{q}{p}\right)^2$$

$$= \frac{q^2}{p^2} + \frac{q}{p} = \frac{q}{p}\left(\frac{q}{p}+1\right) = \frac{q}{p}\left(\frac{1-p}{p}+1\right) = \frac{q}{p}\left(\frac{1}{p}-1+1\right)$$

$$= \frac{q}{p}\cdot\frac{1}{p} = \frac{q}{p^2}$$

Remark 1: Variance $= \dfrac{q}{p^2} = \dfrac{q}{p \cdot p} = \dfrac{\text{Mean}}{p}$

$\Rightarrow$ Variance > Mean $\qquad [\because p < 1 \Rightarrow \dfrac{\text{Mean}}{p} > \text{Mean}\,]$

Hence, unlike binomial distribution, variance of the geometric distribution is greater than mean.

Example 2: Comment on the following:

The mean and variance of geometric distribution are 4 and 3 respectively.

Solution: If the given geometric distribution has parameter p(probability of success in each trial).

Then,

Mean $= \dfrac{q}{p} = 4$ and Variance $= \dfrac{q}{p^2} = 3$

$\Rightarrow \dfrac{1}{p} = \dfrac{3}{4}$

$\Rightarrow p = \dfrac{4}{3}$, which is impossible, since probability can never exceed unity.

Hence, the given statement is wrong.

Now, you can try the following exercises.

E1) Probability of hitting a target in any attempt is 0.6, what is the probability that it would be hit on fifth attempt?

E2) Determine the geometric distribution for which the mean is 3 and variance is 4.

Lack of Memory Property

Now, let us discuss the distinguishing property of the geometric distribution i.e. the 'lack of memory' property or 'forgetfulness property'. For example, in a random experiment satisfying geometric distribution the wait up to 3 trials (say) for the first success does not affect the probability that one will have to wait for a further 5 trials if it is given that the first two trials are failures. The geometric distribution is the only discrete distribution which has the forgetfulness (memoryless) property. However, there is one continuous distribution which also has the memoryless property and that is the exponential distribution which we will study in Unit 15 of MST-003. The exponential distribution is also the only continuous distribution having this property. It is pertinent to mention here that in several aspects, the geometric distribution is discrete analogs of the exponential distribution.

Let us now give mathematical/statistical discussion on 'memoryless property' of geometric distribution.

Suppose an event occurs at one of the trials 1, 2, 3, 4,… and the occurrence time X has a geometric distribution with probability p. Let X be the number of trials preceding to which one has to wait for successful attempt.

Thus, $P[X \geq j] = P[X = j] + P[X = j+1] + ...$

$$= q^j p + q^{j+1} p + q^{j+2} p + ...$$

$$= q^j p [1 + q + q^2 + ...]$$

$$= q^j p \left[\frac{1}{1-q} \right] = q^j p \left(\frac{1}{p} \right) = q^j$$

Now, let us consider the event $[X > j+k]$

Now, $P[X \geq j+k \mid X \geq j]$ means the conditional probability of waiting for at least j + k unsuccessful trials given that we waited for at least j unsuccessful attempts; and is given by

$$P[X \geq j+k \mid X \geq j] = \frac{P[X \geq j+k \mid X \geq j]}{P[X \geq j]}$$

$$= \frac{P[(X \geq j+k) \cap (X \geq j)]}{P[X \geq j]}$$

$$= \frac{P[X \geq j+k]}{P[X \geq j]} \quad [\because X \geq j+k \text{ implies that} \geq j]$$

$$= \frac{q^{j+k}}{q^j} = q^k$$

$$= P[X \geq k] \quad \begin{bmatrix} \because P[X \geq j] = q^j \text{ already} \\ \text{obtained in this section} \end{bmatrix}$$

So, $P[X \geq j+k \mid X \geq j] = P[X \geq k]$

The above result reveals that the conditional probability of at least first j+k trials are unsuccessful before the first success given that at least first j trial were unsuccessful, is the same as the probability that the first k trials were unsuccessful. So, the probability to get first success remains same if we start counting of k unsuccessful trials from anywhere provided all the trials preceding to it are unsuccessful i.e. the future does not depend on past, it depends only on the present. So, the geometric distribution forgets the preceding trials and hence this property is given the name "forgetfulness property" or "Memoryless property" or "lack of memory" property.

12.3 NEGATIVE BINOMIAL DISTRIBUTION

Negative binomial distribution is a generalisation of geometric distribution. Like geometric distribution, variance of this distribution is also greater than its mean. There are many instances including 'deaths of insects' and 'number of insect bites' where negative binomial distribution is employed.

Negative binomial distribution is a generalisation of geometric distribution in the sense that geometric distribution is the distribution of 'number of failures preceding the first success' whereas the negative binomial distribution is the distribution of 'number of failures preceding the rth success'.

Let X be the random variable which denote the number of failures preceding the rth success. Let p be the probability of a success and let x failures are there preceding the rth success and hence for this the number of trials is $x + r$.

Now, $(x + r)^{th}$ trial is success, but the remaining $(r - 1)$ successes in the $x + r - 1$ trials can happen in any $r - 1$ trials out of the $(x + r - 1)$ trials. Thus, happening of first $(r - 1)$ successes in $(x + r - 1)$ trials follow binomial distribution with 'p' as the probability of success in each trial and thus is given by

$$^{x+r-1}C_{r-1} \ p^{r-1} \ q^{(x+r-1)-(r-1)} = {}^{x+r-1}C_{r-1} \ p^{r-1}q^x \text{, where q = 1 - p}$$

[$\because$ by binomial distribution, the probability of x successes in n trials with p as the probability of success is $^{n}C_{x} \ p^x \ q^{n-x}$.]

Therefore,

P[x failures preceding the rth success]

=P[{First $(r - 1)$ successes in $(x + r - 1)$ trials} $\cap$ {success in $(x + r)^{th}$ trial}]

=P[First$(r - 1)$ successes in $(x + r - 1)$ trails]. P[success in $(x + r)^{th}$ trial]

$$= \left({}^{x+r-1}C_{r-1} p^{r-1} q^x \right) p$$

$$= {}^{x+r-1}C_{r-1} p^r q^x$$

The above discussion leads to the following definition:

Definition: A random variable X is said to follow a negative binomial distribution with parameters r (a positive integer) and p $(0 < p < 1)$ if its probability mass function is given by:

$$P[X = x] = \begin{cases} {}^{x+r-1}C_{r-1} \, p^r \, q^x & \text{for } x = 0, 1, 2, 3, ... \\ 0, & \text{otherwise} \end{cases}$$

Now, as we know that

$${}^{n}C_r = {}^{n}C_{n-r}, \qquad \text{[See 'combination' in Unit 4 of MST-001]}$$

$\therefore \ {}^{x+r-1}C_{r-1}$ can be written as

$${}^{x+r-1}C_{(x+r-1)-(r-1)} = {}^{x+r-1}C_x$$

$$= \frac{\lfloor x+r-1}{\lfloor x \lfloor x+r-1-x}} = \frac{\lfloor r+x-1}{\lfloor x \lfloor r-1}$$

$$= \frac{\big(r+(x-1)\big)\big(r+(x-2)\big)...(r+1)(r)\lfloor r-1}{\lfloor x \ \lfloor r-1}$$

$$= \frac{(r+x-1)(r+x-2)...(r+1)r}{\lfloor x}$$

$$= \frac{\{-(-r-x+1)\}\{-(-r-x+2)\}...\{-(-r-1)\}\{-(-r)\}}{\lfloor x}$$

[∵ Numerator is product of x terms from $r + 0$ to $r+(x-1)$ and we have taken
common (-1) from each of these x terms in the product.]

$$= (-1)^x \frac{(-r-x+1)(-r-x+2)...(-r-1)(-r)}{\lfloor x}$$

$$= \frac{(-1)^x (-r)(-r-1)(-r-2)...\{-r-(x-1)\}}{\lfloor x}$$

[Writing the terms in the numerator in reverse order]

$$= (-1)^x \binom{-r}{x}$$

Note: The symbol $\binom{n}{x}$ stands for ${}^{n}C_x$ if n is positive integer and is equal to

$\dfrac{n(n-1)(n-2)...\{n-(x-1)\}}{\lfloor x}$. We may also use the symbol $\binom{n}{x}$ if n is any

real but in this case though it does not stand for ${}^{n}C_x$, yet it is equal to

$\dfrac{n(n-1)(n-2)...\{n-(n-x)\}}{\lfloor x}$.

Hence, the probability distribution of negative binomnial distribution can be
expressed in the following form:

$$P[X = x] = (-1)^x \binom{-r}{x} p^r q^x$$

$$= \binom{-r}{x} (-q)^x p^r \text{ for } x = 0, 1, 2, 3, \ldots$$

$$= \binom{-r}{x} (-q)^x (1)^{-r-x} p^r \text{ for } x = 0, 1, 2, \ldots$$

Here, the expression $\binom{-r}{x} (-q)^x (1)^{-r-x}$ is similar to the binomial distribution

$$\binom{n}{x} p^x q^{n-x}$$

$\therefore \binom{-r}{x} (-q)^x (1)^{-r-x}$ is the general term of $\left[1 + (-q) \right]^{-r} = (1-q)^{-r}$

You have already studied in Unit 9 of this Course that $\binom{n}{x} p^x q^{n-x}$ is the general term of $[q + p]^n$.

and hence

$$P[X = x] = \binom{-r}{x} (-q)^x (1)^{-r-x} \cdot p^r \text{ is the general term of } (1-q)^{-r} p^r$$

$\therefore P[X = 0], P[X = 1], P[X = 2],\ldots$ are the successive terms of the binomial expansion $(1-q)^{-r} p^r$ and hence the sum of these probabilities

$$= (1-q)^{-r} p^r$$

$$= p^{-r} p^r \ [\because \ 1 - q = p]$$

$$= 1,$$

which must be, being a probability distribution.

Also, as the probabilities of the negative binomial distribution for $X = 0, 1, 2, \ldots$ are the successive terms of

$$(1-q)^{-r} p^r = (1-q)^{-r} \left(\frac{1}{p} \right)^{-r} = \left[(1-q) \frac{1}{p} \right]^{-r} = \left[\frac{1}{p} + \left(-\frac{q}{p} \right) \right]^{-r}, \text{ which is a binomial}$$

expansion with negative index $(-r)$, it is for this reason the probability distribution given above is called the negative binomial distribution.

Mean and Variance

Mean and variance of the negative binomial distribution can be obtained on observing the form of this distribution and comparing it with the binomial distribution as follows:

The probabilities of binomial distribution for $X = 0, 1, 2, \ldots$ are the successive terms of the binomial expansion of $(q + p)^n$ and the mean and variance obtained for the distribution are

Mean $= np = (n)(p)$ i.e.   Product of index and second term in $(q + p)$

Variance $= npq = (n)(p)(q)$ i.e. Product of index, second term in $(q + p)$ and first term in $(q + p)$

Similarly, the probabilities of negative binomial distribution for $X = 0, 1, 2, \ldots$ are the successive term of the expansion of $\left[\dfrac{1}{p} + \left(-\dfrac{q}{p}\right)\right]^{-r}$ and thus, its mean and variance are:

Mean $= $ (index) $\left[\text{second term in} \left\{\dfrac{1}{p} + \left(\dfrac{-q}{p}\right)\right\}\right] = (-r)\left[\left(-\dfrac{q}{p}\right)\right] = \dfrac{rq}{p}$, and

Variance $= $ (index) $\left[\text{second term in} \left\{\dfrac{1}{p} + \left(\dfrac{-q}{p}\right)\right\}\right]$ [First term in $\left\{\dfrac{1}{p} + \left(-\dfrac{q}{p}\right)\right\}$]

$$= (-r)\left(-\dfrac{q}{p}\right)\left(\dfrac{1}{p}\right)$$

$$= \dfrac{rq}{p^2}.$$

Remark 2

i) If we take $r = 1$, we have $P[X = x] = pq^x$ for $x = 0, 1, 2, \ldots$ which is geometric probability distribution.

Hence, geometric distribution is a particular case of negative binomial distribution and the latter may be regarded as the generalisation of the former.

ii) Putting $r = 1$ in the formulas of mean and variance of negative binomial distribution, we have

Mean $= \dfrac{(1)q}{p} = \dfrac{q}{p}$, and

Variance $= \dfrac{(1)(q)}{p^2} = \dfrac{q}{p^2}$,

which are the mean and variance of geometric distribution.

Example 3: Find the probability that third head turns up in 5 tosses of an unbiased coin.

Solution: It is a negative binomial situation with $p = \dfrac{1}{2}$, $r = 3$,

$x + r = 5 \Rightarrow x = 2$.

$\therefore$ by negative binomial distribution, we have

$$P[X = 2] = {}^{x+r-1}C_{r-1}\, p^r q^x$$

$$= {}^{2+3-1}C_{3-1}\left(\frac{1}{2}\right)^3\left(\frac{1}{2}\right)^2 = {}^4C_2\left(\frac{1}{2}\right)^5 = \frac{4\times 3}{2}\times\frac{1}{32} = \frac{3}{16}$$

Example 4: Find the probability that a third child in a family is the family's second daughter, assuming the male and female are equally probable.

Solution: It is a negative binomial situation with

$$p = \frac{1}{2} \qquad\qquad [\because \text{male and female are equally probable}]$$

$$r = 2,\ x + r = 3$$

$$\Rightarrow\ x = 1$$

$\therefore$ by negative binomial distribution,

$$P[X = 1] = {}^{x+r-1}C_{r-1}\, p^r\, q^x = {}^{1+2-1}C_{2-1}\left(\frac{1}{2}\right)^2\left(\frac{1}{2}\right)^1 = {}^2C_1\left(\frac{1}{2}\right)^3 = 2\times\frac{1}{8} = \frac{1}{4}.$$

Example 5: A proof-reader catches a misprint in a document with probability 0.8. Find the expected number of misprints in the document in which the proof-reader stops after catching the 20th misprint.

Solution: Let X be the number of misprints not caught by the proof-reader and r be the number of misprints caught by him/her. It is a negative binomial situation where we are to obtain the expected (mean) number of misprints in the document i.e. $E(X + r)$. We will first obtain mean number of misprints which could not be caught by the proof-reader i.e. $E(X)$.

Here, p = 0.8 and hence q = 0.2, r = 20.

Now, by negative binomial distribution,

$$E(X) = \frac{rq}{p} = \frac{(20)(0.2)}{(0.8)} = 5$$

Therefore, $E(X + r) = E(X) + r = 5 + 20 = 25$.

Hence, the expected number of misprints in the document till he catches the 20th misprint is 25.

Now, we are sure that you will be able to solve the following exercises:

E3) Find the probability that fourth five is obtained on the tenth throw of an unbiased die.

E4) An item is produced by a machine in large numbers. The machine is known to produce 10 per cent defectives. A quality control engineer is testing the item randomly. What is the probability that at least 3 items are examined in order to get 2 defectives?

E5) Find the expected number of children in a family which stops producing children after having the second daughter. Assume, the male and female births are equally probable.

We now conclude this unit by giving a summary of what we have covered in it.

12.4 SUMMARY

The following main points have been covered in this unit:

1) A random variable X is said to follow **geometric distribution** if it assumes non-negative integer values and its probability mass function is given by

$$P[X = x] = \begin{cases} q^x p & \text{for } x = 0, 1, 2, ... \\ 0, & \text{otherwise} \end{cases}$$

2) For **geometric** distribution, **mean** $= \dfrac{q}{p}$ and **variance** $= \dfrac{q}{p^2}$.

3) A random variable X is said to follow a **negative binomial distribution** with parameters r (a positive integer) and p $(0 < p < 1)$ if its probability mass function is given by:

$$P(X = x) = \begin{cases} {}^{x+r-1}C_{r-1}\, p^r\, q^x & \text{for } x = 0, 1, 2, 3, ... \\ 0, & \text{otherwise} \end{cases}$$

4) For **negative binomial** distribution, **mean** $= \dfrac{rq}{p}$ and **variance** $= \dfrac{rq}{p^2}$.

5) For both these distributions, **variance > mean.**

12.5 SOLUTIONS/ANSWERS

E1) Let p be the probability of success i.e. hitting the target in an attempt.

∴ $p = 0.6$, $q = 1 - p = 0.4$.

Let X be the number of unsuccessful attempts preceding the first successful attempt.

∴ by geometric distribution,

$P[X = x] = q^x p$ for $x = 0, 1, 2, ...$

$\qquad = (0.4)^x (0.6)$ for $x = 0, 1, 2, ...$

Thus, the desired probability = P[hitting the target in fifth attempt]

$\qquad\qquad\qquad$ = P[The number of unsuccessful attempts before the first success is 4]

$\qquad\qquad\qquad = P[X = 4]$

$\qquad\qquad\qquad = (0.4)^4 (0.6) = (0.0256)(0.6) = 0.01536.$

E2) Let p be the probability of success in an attempt, and $q = 1 - p$

Now, mean $= \dfrac{q}{p} = 3$ and Variance $= \dfrac{q}{p^2} = 4$

$$\Rightarrow \frac{1}{p} = \frac{4}{3} \Rightarrow p = \frac{3}{4} \text{ and hence } q = \frac{1}{4}.$$

Now, let X be the number of failures preceding the first success,

$$\therefore P[X = x] = q^x p$$

$$= \left(\frac{1}{4}\right)^x \left(\frac{3}{4}\right) \text{ for } x = 0, 1, 2, \dots$$

This is the desired probability distribution.

E3) It is a negative binomial situation with

$$r = 4, \ x + r = 10 \Rightarrow x = 6, \ p = \frac{1}{6} \text{ and hence } q = \frac{5}{6}$$

$$\therefore P[X = 6] = {}^{x+r-1}C_{r-1} p^r q^x$$

$$= {}^{6+4-1}C_{4-1} \left(\frac{1}{6}\right)^4 \left(\frac{5}{6}\right)^6$$

$$= {}^9C_3 \cdot \frac{625 \times 25}{36 \times 36 \times 36 \times 36 \times 36}$$

$$= \frac{9 \times 8 \times 7 \times 625 \times 25}{6 \times 36 \times 36 \times 36 \times 36 \times 36} = 0.0217$$

E4) It is a negative binomial situation with $r = 2$, $x + r = 3 \Rightarrow x = 1$, $p = 0.1$ and hence $q = 0.9$.

Now, the required probability $= P[X + r \geq 3]$

$$= P[X \geq 1]$$

$$= 1 - P[X = 0]$$

$$= 1 - \left[{}^{0+r-1}C_{r-1} p^r q^0 \right]$$

$$= 1 - \left[{}^{0+2-1}C_{2-1} (0.1)^2 (0.9)^0 \right]$$

$$= 1 - (1)(0.01) = 0.99.$$

E5) It is a negative binomial situation with $p = \frac{1}{2}$, $q = \frac{1}{2}$, $r = 2$.

Let X be the number of boys in the family

$$\therefore E(X) = \frac{rq}{p} = \frac{(2)\left(\frac{1}{2}\right)}{\frac{1}{2}} = 2.$$

$$\Rightarrow E(X + r) = E(X) + r = 2 + 2 = 4$$

$\therefore$ the required expected value = 4.