PERIYAR UNIVERSITY

NAAC 'A++' Grade with CGPA 3.61 (Cycle - 3)
State University - NIRF Rank 56 - State Public University Rank 25
Salem-636011, Tamilnadu, India.
SALEM - 636 011

CENTRE FOR DISTANCE AND ONLINE EDUCATION

(CDOE)

M.Sc. Mathematics

Semester - I

Core – I: Mathematical Statistics

(Candidates admitted from 2024 onwards)

Prepared by:

Centre for Distance and Online Education (CDOE)
Periyar University
Salem 636011

Contents

Syllabus

Mathematical Statistics

Objective: This course aims to teach the students about special distributions and random Process. To prepare students for lifelong learning and successful careers using their mathematical statistics skills.

Unit-I: Characteristic Function

Properties of characteristic functions-characteristic function and moments - semi invariants - the characteristic functions of sum of independent random variables determination of distribution function of the characteristic function - Probability generating function.

Unit-II: Some Probability Distributions

One-point and two-point distributions - The Bernoulli scheme: Binomial distribution The Poisson scheme: The generalized binomial distribution - The Polya and hypergeometric distributions - The Poisson distribution - The uniform distribution.

Unit-III: Some Probability Distributions

The normal distribution - The gamma distribution - The beta distribution - The Cauchy and Laplace distributions - The multinomial distribution - Compound distributions.

Unit-IV: Limit Theorems

Stochastic Convergence - Bernoulli's law of large numbers - the convergence of a sequence of distribution functions - The Levy-Cramér theorem - The De Moivre-Laplace theorem - The Lindeberg-Lévy theorem - The Lapunov theorem.

Unit-V: Markov Chains

Homogeneous Markov chains - The transition matrix - The Ergodic theorem - Random variables forming a homogeneous Markov Chain. Stochastic Processes: The Wiener Process - The Stationary Processes.

References

1. M. Fisz, Probability Theory and Mathematical Statistics, John wiley and sons, New Your, Third Edition, 1963.
2. T. Veerarajan, Fundamentals of Mathematical Statistics, Yesdee Publishing, 2017.
3. P. R. Vittal, Mathematical Statistics, Margham Publications, 2002.
4. T. Veerarajan, Probability, Statistics and Random Processes, Mc Graw Hill Education (India) Private Limited, Third Edition, 2015.
5. R.S.N. Pillai and V. Bagavathi, Statistics, Sultan Chand and Sons, 2010.
6. Singaravelu. A, S. Sivasubramanian, Probability and Random Processes, Meenakshi Agency, 2008.
7. DN Elhance, Veena Elhance and BM Aggarwal, Fundamentals of Statistics, Kitab Mahal.
8. S. C. Gupta and V. K. Kapoor, Fundamentals of Mathematical Statistics, Sultan Chand and Sons, 2008.

Unit 1

Characteristic Function

Objective

This course aims to teach the students about characteristic function and moments. Determination of distribution function of the characteristic function of probability generating function.

1.1 Properties of Characteristic Functions

In this section we investigate the expected value of a certain function of a random variable and obtain a method of investigation which is extremely useful in further work on probability theory and its application to statistics. Let X be a random variable and let $F(x)$ be its distribution function.

Definition 1.1.1 Characteristic Function
The function

$$\phi(t) = E\left(e^{itX}\right) \tag{1.1}$$

where t is a real number and i is the imaginary unit, is called the characteristic function of the random variable X or of the distribution function $F(x)$.

If X is a random variable of the discrete type with jump points x_k $(k = 1, 2, \ldots)$ and $P\left(X = x_k\right) = p_k$, the characteristic function of X has the form

$$\phi(t) = E\left(e^{itX}\right) = \sum_k p_k e^{itx_k} \tag{1.2}$$

Since $|e^{itx_k}| = 1$ and $\sum_k p_k = 1$, the series on the right-hand side of (1.2) is absolutely and uniformly convergent. Thus, the characteristic function $\phi(t)$, as the sum of a uniformly convergent series of continuous functions, is continuous for every real value of t.

Example 1.1.2 *The random variable X can take on the values $x_1 = -1$ and $x_2 = +1$ with probabilities $P(X = -1) = P(X = +1) = 0.5$. We shall determine the characteristic function of this random variable. By (1.2) we have,*

$$\phi(t) = 0.5e^{-it} + 0.5e^{it} = 0.5(\cos t - i \sin t) + 0.5(\cos t + i \sin t) = \cos t \tag{1.3}$$

If X is a random variable of the continuous type with density function $f(x)$, its characteristic function is given by the formula

$$\phi(t) = E\left(e^{itX}\right) = \int_{-\infty}^{+\infty} f(x)e^{itx}dx \int_{-\infty}^{+\infty} f(x)\left|e^{itx}\right| dx = \int_{-\infty}^{+\infty} f(x)dx = 1 \tag{1.4}$$

Since the integral in (1.4) is absolutely and uniformly convergent; hence $\phi(t)$ is a continuous function for every value of t.

Example 1.1.3 *The density $f(x)$ is defined as;*

$$f(x) = \begin{cases} 0, & \textit{for} \quad x < 0 \\ 1, & \textit{for} \quad 0 \leqslant x \leqslant 1 \\ 0, & \textit{for} \quad x > 1 \end{cases} \tag{1.5}$$

This distribution is called uniform or rectangular. Its characteristic function is;

$$\phi(t) = \int_{-\infty}^{+\infty} f(x)e^{itx}dx = \int_{0}^{1} f(x)e^{itx}dx = \left[\frac{e^{itx}}{it}\right]_0^1 = \frac{e^{it} - 1}{it} \tag{1.6}$$

We now investigate some of the properties of characteristic functions. We have;

$$\phi(0) = E\left(e^0\right) = E(1) = 1 \tag{1.7}$$

Since,

$$|\phi(t)| = \left|E\left(e^{itX}\right)\right| \leqslant E\left(\left|e^{itX}\right|\right) = 1$$

we have;

$$|\phi(t)| \leqslant 1 \tag{1.8}$$

We next have;

$$\phi(-t)=E\left(e^{-itX}\right)=E(\cos tX-i\sin tX)=E(\cos tX)-iE(\sin tX)$$

Since,

$$\phi(t)=E\left(e^{itX}\right)=E(\cos tX+i\sin tX)=E(\cos tX)+iE(\sin tX)$$

we obtain

$$\phi(-t)=\overline{\phi(t)} \tag{1.9}$$

where $\overline{\phi(t)}$ *denotes the complex number conjugate to* $\phi(t)$*. Every characteristic function must satisfy conditions (1.7), (1.8) and (1.9). These conditions are, however, not sufficient; thus not every function* $\phi(t)$ *satisfying these conditions is a characteristic function of some random variable. He has shown that a function* $\phi(t)$ *which is not identically constant and which, in a neighborhood of zero, can be represented in the form*

$$\phi(t)=1+0\left(t^{2+\alpha}\right)$$

with $\alpha>0$ *cannot be a characteristic function. It follows immediately that neither the function* $\phi(t)=\exp\left(-t^{4}\right)$ *nor the function* $\phi(t)=1/\left(1+t^{4}\right)$ *can be a characteristic function. Further giving necessary and sufficient conditions for a function* $\phi(t)$ *to be a characteristic function.*

Theorem 1.1.4 *Let the function* $\phi(t)$ *defined for* $-\infty<t<+\infty$ *satisfy condition (1.7). The function* $\phi(t)$ *is the characteristic function of some distribution function if and only if*

1. $\phi(t)$ *is continuous.*

2. *for* $n=1,2,3,\ldots$ *and every real* $t_1,\ldots,t_n$ *and complex* $a_1,\ldots,a_n$ *we have*

$$\sum_{j,k=1}^{n}\phi\left(t_j-t_k\right)a_j\bar{a}_k \geqslant 0$$

Let us recall that a function satisfying second condition of theorem 1.1 is called positive definite. Another necessary and sufficient condition for the function $\phi(t)$ *to be a characteristic function.*

Let Us Sum Up

Learners, in this section we have seen that definition of characteristic function and properties of characteristic function and also given standard theorems.

Check Your Progress

1. A feedback system is stable if the number of zeros (z) of a characteristic equation in the right half of the s- plane is:
A. $Z=1$
B. $Z=0$
C. $Z=2$
D. None of these
2. The function $\phi(t)$ is:
A. $E\left(e^{itX}\right)$
B. $\left(e^{itX}\right)$
C. $E\left(e^{tX}\right)$
D. $E\left(e^{it}\right)$

1.2 Characteristic Function and Moments

Consider a random variable X and suppose that its l th moment $m_l=E\left(X^l\right)$ exists. Suppose that X is a random variable of the discrete type with jump points x_k. Then we can differentiate (1.2) l times with respect to t. In fact, the l th derivative with respect to t of the expression under the summation sign in (1.2) equals $p_k l^i {x_k}^l e^{itx_k}$. On the other hand, from the existence of the l th moment there follows the existence of the absolute l th moment. Since

$$\sum_k \left|i^l p_k x_k e^i e^{itx_k}\right| = \sum_k \left|p_k {x_k}^l\right| = \beta_l$$

we can differentiate (1.2) l times under the summation sign. Hence we have,

$$\phi^{(l)}(t) = \sum_k p_k l^l {x_k}^l e^{itx_k} = E\left(i^l X^l e^{itX}\right)$$

Suppose now that $f(x)$ is the density function of a random variable X of the continuous type. Then we can differentiate (1.4) l times. Indeed, the l th derivative with respect to t of the expression under the integral sign in (1.4) equals $i^l x^l f(x) e^{itx}$. We have

$$\int_{-\infty}^{+\infty} \left| i^l x^l f(x) e^{itx} \right| dx = \int_{-\infty}^{+\infty} \left| x^l f(x) \right| dx = \beta_l.$$

By assumption, the absolute moment β_l is finite. Thus we can differentiate the formula for $\phi(t) l$ times under the integral sign. We obtain

$$\phi^{(l)}(t) = \int_{-\infty}^{+\infty} i^l x^l f(x) e^{itx} dx = E\left(i^l X^l e^{itX}\right) \tag{1.10}$$

Thus we have obtained the same result as for a random variable of the discrete type

$$\phi^{(l)}(t) = E\left(i^l X^l e^{itx}\right) \tag{1.11}$$

Let us compute $\phi^{(l)}(0)$ from relation (1.11). We have

$$\phi^{(l)}(0) = i^l E\left(X^l\right) = i^l m_l \tag{1.12}$$

Hence

$$m_l = \frac{\phi^{(l)}(0)}{i^l} \tag{1.13}$$

Thus we have proved the following theorem.

Theorem 1.2.1 *If the lth moment m_l of a random variable exists, it is expressed by formula (1.13), where $\phi^{(l)}(0)$ is the lth derivative of the characteristic function $\phi(t)$ of this random variable at $t = 0$.*

Example 1.2.2 *Suppose that the random variable X has a Poisson distribution, that is, it can take on the values $x_k = k$, where k is any non-negative integer, and the probability function is given by the formula*

$$P(X = k) = \frac{\lambda^k}{k!} e^{-\lambda} \tag{1.14}$$

where λ is a positive constant. We shall find the characteristic function of X. From (1.2) *we obtain*

$$\phi(t) = \sum_{k=0}^{\infty} e^{itk} \frac{\lambda^k}{k!} e^{-\lambda} = e^{-\lambda} \sum_{k=0}^{\infty} \frac{\left(\lambda e^{it}\right)^k}{k!} = \exp(-\lambda) \exp\left(\lambda e^{it}\right) = \exp\left[\lambda\left(e^{it} - 1\right)\right]. \tag{1.15}$$

Furthermore,

$$\phi'(t) = \lambda i \exp(it) \exp\left[\lambda\left(e^{it} - 1\right)\right] \tag{1.16}$$

From (1.13) we obtain

$$m_1 = \frac{\phi'(0)}{i} = \frac{\lambda i}{i} = \lambda \tag{1.17}$$

Similarly,

$$\phi''(t) = \lambda i^2 \exp(it) \exp\left[\lambda\left(e^{it} - 1\right)\right]\left[\lambda \exp(it) + 1\right] \tag{1.18}$$

Hence

$$m_2 = \frac{\phi''(0)}{i^2} = \frac{i^2 \lambda \cdot (\lambda + 1)}{i^2} = \lambda(\lambda + 1) \tag{1.19}$$

Thus the central moment of the second order is

$$\mu_2 = \lambda(\lambda + 1) - \lambda^2 = \lambda \tag{4.2.9}$$

In a similar manner we can obtain the moments of higher orders.

Example 1.2.3 *We shall find the characteristic function and the moments of a normal distribution. We have*

$$f(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2} \tag{1.20}$$

Hence

$$\phi(t) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} \exp(itx) \exp\left(-\frac{x^2}{2}\right) dx \tag{1.21}$$

$$= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} \exp\left[-\frac{(x - it)^2}{2}\right] \exp\left(-\frac{t^2}{2}\right) dx = \exp\left(-\frac{t^2}{2}\right) \tag{1.22}$$

Since $\phi'(t) = -t \exp\left(-t^2/2\right)$, *we have*

$$m_1 = \frac{\phi'(0)}{i} = 0 \tag{1.22 a}$$

Next, we have $\phi''(t) = (t^2 - 1) \exp\left(-t^2/2\right)$; *hence*

$$m_2 = \frac{\phi''(0)}{i^2} = 1 \tag{1.23}$$

We have already obtained the same values m_1 and m_2 in examples. The reader can verify that all the odd order moments equal zero and that the even order moments are expressed by the formula

$$m_{2l} = 1 \cdot 3 \cdot 5 \cdot \ldots \cdot (2l-1) \tag{1.24}$$

We notice that the converse of theorem is not true. An example of a random variable, whose expectation does not exist and whose characteristic function is differentiable at $t = 0$. But if the characteristic function $\phi(t)$ has a finite derivative of an even order $2k$ at $t' = 0$, then the moment of order $2k$ of the corresponding random variable exists. As we know, in this case all the moments of orders smaller than $2k$ also exist.

Let Us Sum Up

Learners, in this section we have seen that the characteristic function and moments. Also given theorem and examples.

Check Your Progress

1. Which of the following statements is true about the relationship between the characteristic function and the moment generating function?
A. They are identical and can be used interchangeably.
B. The characteristic function is defined for all real t, while the MGF is defined only for t in a neighborhood around zero.
C. The MGF is defined for all real t, while the characteristic function is defined only for t in a neighborhood around zero.
D. The characteristic function and the MGF are not related in any way.
2. The moments of a random variable X can be obtained from the characteristic function $\phi_X(t)$ by:
A. Differentiating $\phi_X(t)$ with respect to t and then evaluating at $t = 0$.
B. Integrating $\phi_X(t)$ with respect to t and then evaluating at $t = 0$.
C. Differentiating $\phi_X(t)$ with respect to t and then evaluating at $t = 1$.
D. Integrating $\phi_X(t)$ with respect to t and then evaluating at $t = 1$.

1.3 Semi-Invariants

Now consider the characteristic function of a linear transformation of the random variable X. First consider the translation

$$Y = X + b$$

Denoting by $\phi_1(t)$ the characteristic function of the random variable Y, we obtain

$$\phi_1(t) = E\left(e^{itY}\right) = E\left(e^{it(X+b)}\right) = E\left(e^{itX}\right)e^{itb} = e^{itb}\phi(t) \tag{1.25}$$

We see that when the random variable is translated by a constant b, its characteristic function is multiplied by the factor e^{itb}. Now let

$$Y = aX$$

We have,

$$\phi_1(t) = E\left(e^{itY^Y}\right) = E\left(e^{itaX}\right) = \phi(at) \tag{1.26}$$

Thus, the characteristic function of the random variable aX equals the characteristic function of the random variable X at the point at. In particular, if $a = -1$, we obtain

$$\phi_1(t) = \phi(-t) = \overline{\phi(t)}$$

Now let us consider the transformation

$$Y = aX + b$$

Denoting the characteristic functions of the random variables X and Y by $\phi(t)$ and $\phi_1(t)$ respectively, we obtain from equations (1.25) and (1.26)

$$\phi_1(t) = e^{ibt}\phi(at) \tag{1.27}$$

In particular, let

$$Y = \frac{X - m_1}{\sigma}$$

where m_1 and σ denote respectively the expected value and the standard deviation of X. Then

$$\phi_1(t) = \exp\left(-\frac{m_1 it}{\sigma}\right)\phi\left(\frac{t}{\sigma}\right) \tag{1.28}$$

Sometimes it is conveniel. to deal with a set of parameters other than the set of moments. We obtain such parameters by considering the function

$$\psi(t) = \log\phi(t) \tag{1.29}$$

where $\phi(t)$ is the characteristic function of the random variable under consideration. Let us formally expand the function $\phi(t)$ in a power series in a neighborhood of $t = 0$,

$$\phi(t) = 1 + \sum_{s=1}^{\infty}\frac{m_s}{s!}(it)^s \tag{1.30}$$

Let us denote by z the series on the right-hand side of (1.30) and let us formally expand the function $\psi(t)$ into a power series

$$\psi(t) = \log\phi(t) = \log(1+z) = \frac{z}{1} - \frac{z^2}{2} + \frac{z^3}{3} - \ldots = \sum_{s=1}^{\infty}\frac{\kappa_s}{s!}(it)^s \tag{1.31}$$

From (1.30) and (1.31) we obtain the formal equation

$$\phi(t) = 1+\sum_{s=1}^{\infty}\frac{m_s}{s!}(it)^s = \exp\left[\sum_{s=1}^{\infty}\frac{\kappa_s}{s!}(it)^s\right] = 1+\sum_{s=1}^{\infty}\frac{\kappa_s}{s!}(it)^s+\frac{1}{2!}\left[\sum_{s=1}^{\infty}\frac{\kappa_s}{s!}(it)^s\right]^2+\frac{1}{3!}\left[\sum_{s=1}^{\infty}\frac{\kappa_s}{s!}(it)^s\right]^3+\ldots \tag{1.32}$$

Definition 1.3.1 *The coefficients κ_s in (1.32) are called semi-invariants. To express the semi-invariants in terms of the moments or the moments in terms of the semi-invariants, we compare successively the coefficients of $(it)^s$ for particular values of s in equation (1.32). In this way we obtain*

$$\begin{aligned}
\kappa_1 &= m_1\\
\kappa_2 &= m_2 - m_1^2 = \sigma^2\\
\kappa_3 &= m_3 - 3m_1m_2 + 2m_1^3\\
\kappa_4 &= m_4 - 3m_2^2 - 4m_1m_3 + 12m_1^2m_2 - 6m_1^4
\end{aligned} \tag{1.32 a}$$

and also

$$m_1 = \kappa_1$$
$$m_2 = \kappa_2 + \kappa_1^2 \tag{1.32 b}$$
$$m_3 = \kappa_3 + 3\kappa_1\kappa_2 + \kappa_1^3$$
$$m_4 = \kappa_4 + 3\kappa_2^2 + 4\kappa_1\kappa_3 + 6\kappa_1^2\kappa_2 + \kappa_1^4$$

The semi-invariants can also be expressed in terms of the central moments,

$$\kappa_1 = m_1$$
$$\kappa_2 = \mu_2 = \sigma^2 \tag{1.32 c}$$
$$\kappa_3 = \mu_3$$
$$\kappa_4 = \mu_4 - 3\mu_2^2$$

where $\phi(t)$ is the characteristic function of the random variable under consideration. Let us formally expand the function $\phi(t)$ in a power series in a neighborhood of $t = 0$,

$$\phi(t) = 1 + \sum_{s=1}^{\infty} \frac{m_s}{s!}(it)^s \tag{1.33}$$

Let us denote by z the series on the right-hand side of (1.3.3) and let us formally expand the function $\psi(t)$ into a power series

$$\psi(t) = \log \phi(t) = \log(1 + z) = \frac{z}{1} - \frac{z^2}{2} + \frac{z^3}{3} - \ldots = \sum_{s=1}^{\infty} \frac{\kappa_s}{s!}(it)^s \tag{1.34}$$

From (1.33) and (1.34) we obtain the formal equation

$$\begin{aligned} \phi(t) = 1 + \sum_{s=1}^{\infty} \frac{m_s}{s!}(it)^s &= \exp\left[\sum_{s=1}^{\infty} \frac{\kappa_s}{s!}(it)^s\right] \\ &= 1 + \sum_{s=1}^{\infty} \frac{\kappa_s}{s!}(it)^s + \frac{1}{2!}\left[\sum_{s=1}^{\infty} \frac{\kappa_s}{s!}(it)^s\right]^2 + \frac{1}{3!}\left[\sum_{s=1}^{\infty} \frac{\kappa_s}{s!}(it)^s\right]^3 + \ldots \end{aligned} \tag{1.34 a}$$

Definition 1.3.2 *The coefficients κ_s in (1.34) are called semi-invariants. To express the* *semi-invariants in terms of the moments or the moments in terms of the semi-invariants, we compare successively the coefficients of $(it)^s$ for particular values of s in equation (1.34 a). In this way we obtain*

$$\begin{aligned}
\kappa_1 &= m_1 \\
\kappa_2 &= m_2 - m_1^2 = \sigma^2 \\
\kappa_3 &= m_3 - 3m_1m_2 + 2m_1^3 \\
\kappa_4 &= m_4 - 3m_2^2 - 4m_1m_3 + 12m_1^2m_2 - 6m_1^4
\end{aligned} \tag{1.34 b}$$

and also

$$\begin{aligned}
m_1 &= \kappa_1 \\
m_2 &= \kappa_2 + \kappa_1^2 \\
m_3 &= \kappa_3 + 3\kappa_1\kappa_2 + \kappa_1^3 \\
m_4 &= \kappa_4 + 3\kappa_2^2 + 4\kappa_1\kappa_3 + 6\kappa_1^2\kappa_2 + \kappa_1^4
\end{aligned} \tag{1.34 c}$$

The semi-invariants can also be expressed in terms of the central moments,

$$\begin{aligned}
\kappa_1 &= m_1 \\
\kappa_2 &= \mu_2 = \sigma^2 \\
\kappa_3 &= \mu_3 \\
\kappa_4 &= \mu_4 - 3\mu_2^2
\end{aligned} \tag{1.34 d}$$

Let Us Sum Up

Learners, in this section we have seen that definitions of semi-invariants of characteristic functions and also given theorems and applications.

Check Your Progress

1. A semi-invariant of a random variable X is defined as:

A. The expectation of X^k for some integer k.

B. The expectation of e^{tX} for some real number t.

C. The characteristic function $\mathbb{E}[e^{itX}]$.

D. The moment generating function $\mathbb{E}[e^{tX}]$.

2. The semi-invariants of a random variable X are related to:

A. The higher moments of X.

B.The Fourier transform of the probability density function of X.

C. The cumulants of X.

D. The characteristic function of X.

1.4 Characteristic Function and Independent Random Variables

From (1.34) and (1.34 a) it follows that if the moment of the l th order exists, all the semi-invariants of order not greater than l also exist. The name semi-invariants comes from the fact that under a translation, that is, under a transformation $Y = X + b$, all semi-invariants except κ_1 remain unchanged. If we denote by $\phi(t)$ and $\phi_1(t)$ the characteristic functions of the random variables X and Y, respectively, we have, by equation (1.34 b)

$$\log \phi_1(t) = bit + \log \phi(t) \tag{1.35}$$

Thus the translation changes only the coefficient of the term with it to the first power in the expansion (1.35); hence it changes only the semiinvariant of the first order.

Example 1.4.1 *We shall compute the semi-invariants of the Poisson distribution discussed. The characteristic function of the Poisson distribution is*

$$\phi(t) = \exp\left[\lambda\left(e^{it} - 1\right)\right] \tag{1.36}$$

Hence we obtain

$$\psi(t) = \log \phi(t) = \lambda \left(e^{it} - 1\right) = \lambda \left(\sum_{k=0}^{\infty} \frac{(it)^k}{k!} - 1\right) = \lambda \sum_{k=1}^{\infty} \frac{(it)^k}{k!} \tag{1.37}$$

From formula (1.37), we obtain

$$\kappa_k = \lambda \quad (k = 1, 2, \ldots) \tag{1.38}$$

Using the formulas for the relations between semi-invariants and moments we can obtain from formula (1.38) the moments of arbitrary order of the Poisson distribution. Let X and Y be two independent random variables. From the considerations of random variables e^{itX} and e^{itY} are also independent. We shall find the characteristic function of the sum

$$Z = X + Y$$

Let $\phi(t), \phi_1(t)$ and $\phi_2(t)$ denote respectively the characteristic functions of the random variables Z, X, and Y. We have

$$\phi(t) = E\left(e^{itZ}\right) = E\left(e^{it(X+Y)}\right) = E\left(e^{itX} e^{itY}\right) \tag{1.39}$$

By the independence of the random variables e^{itX} and e^{itY}.

$$\phi(t) = E\left(e^{ilX}\right) E\left(e^{itY}\right) = \phi_1(t)\phi_2(t) \tag{1.40}$$

This result can be generalized to an arbitrary finite number of independent random variables.

Theorem 1.4.2 *The characteristic function of the sum of an arbitrary finite number of independent random variables equals the product of their characteristic functions. Thus, if Z is the sum of n independent random variables,*

$$Z = X_1 + X_2 + \ldots + X_n \tag{1.41}$$

and $\phi(t), \phi_1(t), \phi_2(t), \ldots, \phi_n(t)$ denote the characteristic functions of $Z, X_1, X_2, \ldots, X_n$, respectively, then

$$\phi(t) = \phi_1(t)\phi_2(t)\ldots\phi_n(t) \tag{1.42}$$

Example 1.4.3 *Suppose two independent random variables X_1 and X_2 have Poisson distributions*

$$P\left(X_1=r\right)=\frac{\lambda_1^r}{r!}e^{-\lambda_1}, \quad P\left(X_2=r\right)=\frac{\lambda_2^r}{r!}e^{-\lambda_2}(r=0,1,\ldots) \tag{1.43}$$

Consider the random variable

$$Z=X_1-X_2 \tag{1.44}$$

We shall determine the characteristic function and the semi-invariants of Z. By equation (1.44) the characteristic functions $\phi_1(t)$ and $\phi_2(t)$ of X_1 and X_2 have the form

$$\phi_1(t)=\exp\left[\lambda_1\left(e^{it}-1\right)\right], \quad \phi_2(t)=\exp\left[\lambda_2\left(e^{it}-1\right)\right] \tag{1.45}$$

By (1.45), the characteristic function of $-X_2$ is

$$\phi_2(-t)=\exp\left[\lambda_2\left(e^{-it}-1\right)\right] \tag{1.46}$$

Since X_1 and $-X_2$ are independent, we obtain by (1.46) for the characteristic function of the random variable Z

$$\phi(t)=\exp\left[\lambda_1\left(e^{it}-1\right)\right]\exp\left[\lambda_2\left(e^{-it}-1\right)\right]=\exp\left(\lambda_1e^{it}+\lambda_2e^{-it}-\lambda_1-\lambda_2\right)$$

Expanding the exponents e^{it} and e^{-it} into power series, we obtain

$$\phi(t)=\exp\left[\left(\lambda_1-\lambda_2\right)(it)+\left(\lambda_1+\lambda_2\right)\frac{(it)^2}{2!}+\left(\lambda_1-\lambda_2\right)\frac{(it)^3}{3!},+\ldots\right]$$
$$\psi(t)=\log\phi(t)=\left(\lambda_1-\lambda_2\right)\frac{(it)}{1!}+\left(\lambda_1+\lambda_2\right)\frac{(it)^2}{2!}+\left(\lambda_1-\lambda_2\right)\frac{(it)^3}{3!}+\ldots$$

From (1.46) it follows that all the semi-invariants of odd order of Z equal $\lambda_1-\lambda_2$, and all the semi-invariants of even order equal $\lambda_1+\lambda_2$. The expected value and the variance of Z can be obtained from (1.46),

$$m_1=\kappa_1=\lambda_1-\lambda_2, \quad \sigma^2=\kappa_2=\lambda_1+\lambda_2$$

We notice that the converse of theorem 1.1.13 is not true; that is, the characteristic function of the sum of dependent random variables may equal the product of their characteristic *functions.*

Example 1.4.4 *The joint distribution of the random variable (X,Y) is given by the density*

$$f(x,y)=\begin{cases}\frac{1}{4}\left[1+xy\left(x^2-y^2\right)\right] & \text{for} \quad |x|\leqslant 1 \quad \text{and} \quad |y|\leqslant 1\\ 0 & \text{for all other points.}\end{cases}$$

We first show that the random variables X and Y are dependent. The marginal distributions in the domains $|x|\leqslant 1$ and $|y|\leqslant 1$ are, respectively, of the form

$$f_1(x)=\int_{-1}^{+1}\frac{1}{4}\left[1+xy\left(x^2-y^2\right)\right]dy=\frac{1}{4}\left(y+\frac{1}{2}x^3y^2-\frac{1}{4}xy^4\right)_{-1}^{+1}=\frac{1}{2}$$
$$f_2(y)=\int_{-1}^{+1}\frac{1}{4}\left[1+xy\left(x^2-y^2\right)\right]dx=\frac{1}{4}\left(x+\frac{1}{4}x^4y-\frac{1}{2}x^2y^3\right)_{-1}^{+1}=\frac{1}{2}$$

We then obtain $f_1(x)f_2(y)=\frac{1}{4}\neq f(x,y)$; hence the random variables X and Y are not independent. We now find the density of the sum $Z=X+Y$. Then,

$$f_3(z)=\int_{-\infty}^{+\infty}f(x,z-x)dx$$

The end points of the intervals of x values, for which $f(x,z-x)>0$, depend on z. To find them, observe that by introducing the variables x,z instead of x,y we transform the square $|x|\leqslant 1,|y|\leqslant 1$ into the domain defined by the inequalities;

$$|x|\leqslant 1, \quad x-1\leqslant z\leqslant x+1 \tag{1.47}$$

The shaded area represents the domain in the (x,y) plane defined by the inequalities $|x|\leqslant 1$, $|y|\leqslant 1$, and the corresponding domain in the (x,z) plane. Let us write inequalities (1.47) in the form

$$|x|\leqslant 1, \quad z-1\leqslant x\leqslant z+1$$

Furthermore, we notice that for $z\leqslant 0$ we have

$$z-1\leqslant -1, \quad z+1\leqslant 1$$

Thus for $z\leqslant 0$ we integrate the function $f(x,z-x)$ from -1 to $z+1$, and for $z>0$ from $z-1$ to 1.

After simple computations we obtain;

$$f_3(z)=\begin{cases}\int_{-1}^{z+1}\frac{1}{4}\left(1+3z^2x^2-2zx^3-z^3x\right)dx=\frac{1}{4}(2+z) & \textit{for } -2\leqslant z\leqslant 0,\\ \int_{z-1}^{1}\frac{1}{4}\left(1+3z^2x^2-2zx^3-z^3x\right)dx=\frac{1}{4}(2-z) & \textit{for } 0<z\leqslant 2,\\ 0 & \textit{for } |z|>2.\end{cases}$$

A distribution such as that of Z is called a triangular distribution. The graph of the function $f_3(z)$ is represented in the figure. We now determine the characteristic functions of X, Y and $Z = X + Y$. We have

$$\phi_1(t)=\frac{1}{2}\int_{-1}^{+1}e^{itx}dx=\frac{1}{2}\left[\frac{e^{itx}}{it}\right]_{-1}^{+1}=\frac{e^{it}-e^{-it}}{2it}=\frac{\sin t}{t}$$

Similarly,

$$\phi_2(t)=\frac{\sin t}{t}$$

Since the variable z takes on the values from the interval $[-2,+2]$, we find

$$\begin{aligned}\phi_3(t)&=\frac{1}{4}\int_{-2}^{0}(2+z)e^{itz}dz+\frac{1}{4}\int_{0}^{2}(2-z)e^{itz}dz=\frac{1}{4}\left(\frac{2-e^{2it}-e^{-2it}}{t^2}\right)\\ &=\frac{1}{2t^2}\left(1-\frac{e^{2it}+e^{-2it}}{2}\right)=\frac{1}{2t^2}(1-\cos 2t)=\left(\frac{\sin t}{t}\right)^2\end{aligned}$$

It follows that the equality $\phi_3(t)=\phi_1(t)\phi_2(t)$ holds; nevertheless X and Y are dependent.

Let Us Sum Up

Learners, in this section we have seen that the defintions characteristic function of the sum of independent random variables and also given theorems and applications.

Check Your Progress

1. Let X and Y be independent random variables with characteristic functions $\phi_X(t)$ and $\phi_Y(t)$, respectively. The characteristic function of the sum $Z = X + Y$ is:

A. $\phi_X(t)\cdot\phi_Y(t)$

B. $\phi_X(t)+\phi_Y(t)$

C. $\phi_X(t)\cdot\phi_Y(-t)$

D. $\phi_X(t)+\phi_Y(-t)$

2. If $X_1, X_2, \ldots, X_n$ are independent random variables with characteristic functions $\phi_{X_i}(t)$, the characteristic function of their sum $S_n = X_1 + X_2 + \cdots + X_n$ is:

A. $\phi_{X_1}(t) \cdot \phi_{X_2}(t) \cdot \cdots \cdot \phi_{X_n}(t)$

B. $\phi_{X_1}(t) + \phi_{X_2}(t) + \cdots + \phi_{X_n}(t)$

C. $\phi_{X_1}(t) \cdot \phi_{X_2}(t) \cdot \cdots \cdot \phi_{X_n}(t)$

D. $\phi_{X_1}(t) \cdot \phi_{X_2}(t) \cdots \phi_{X_n}(t)$

1.5 Distribution Function and Characteristic Function

We know that uniquely determines the characteristic function of a given distribution function. We shall prove the theorem of Lévy that the converse is also true: from the characteristic function we can uniquely determine the distribution function. Let $F(x)$ and $\phi(t)$ denote respectively the distribution function and the characteristic function of the random variable X. If $a+h$ and $a-h(h>0)$ are continuity points of the distribution function $F(x)$,

$$F(a+h)-F(a-h)=\lim_{T \to \infty} \frac{1}{\pi} \int_{-T}^{T} \frac{\sin ht}{t} e^{-ita} \phi(t) dt \tag{1.48}$$

Before proving it we shall show how to apply theorem. Since the numbers a and h are arbitrary, formula (1.48) gives the difference $F\left(x_2\right)-F\left(x_1\right)$ for arbitrary continuity points x_1 and x_2. By the relation

$$F\left(x_2\right)-F\left(x_1\right)=P\left(x_1 \leqslant X<x_2\right)$$

if we know the characteristic function $\phi(t)$, we obtain from theorem the probability that the value of X belongs to an arbitrary. Let $x=x_2$ be a given continuity point and let $x_1 \rightarrow -\infty$, where the passage to the limit is performed over the set of continuity points. Here the sequence of differences $F(x)-F\left(x_1\right)$ is determined by the characteristic function and is convergent to $F(x)$; thus the distribution function $F(x)$ is determined at every continuity point; hence it is determined everywhere. We now give the proof of theorem.

Proof: We give the proof only for a random variable of the continuous type with density function $f(x)$. Denote

$$J = \frac{1}{\pi} \int_{-T}^{+T} \frac{\sin ht}{t} e^{-ita} \phi(t) dt \tag{1.49}$$

From the definition of the characteristic function we obtain

$$\begin{aligned} J &= \frac{1}{\pi} \int_{-T}^{+T} \left[\int_{-\infty}^{+\infty} \frac{\sin ht}{t} e^{-ita} e^{itx} f(x) dx \right] dt \\ &= \frac{1}{\pi} \int_{-T}^{T} \left[\int_{-\infty}^{+\infty} \frac{\sin ht}{t} e^{it(x-a)} f(x) dx \right] dt \end{aligned}$$

We notice that we can interchange the order of integration since the limits of integration with respect to t are finite and the integral is absolutely convergent with respect to x. Thus $\int_{-\infty}^{+\infty} \left| \frac{\sin ht}{t} e^{it(x-a)} \right| f(x) dx = \int_{-\infty}^{+\infty} \left| \frac{\sin ht}{t} \right| f(x) dx \leqslant h \int_{-\infty}^{+\infty} f(x) dx = h$. We obtain

$$\begin{aligned} J &= \frac{1}{\pi} \int_{-\infty}^{+\infty} \left[\int_{-T}^{T} \frac{\sin ht}{t} e^{it(x-a)} f(x) dt \right] dx \\ &= \frac{1}{\pi} \int_{-\infty}^{+\infty} \left[\int_{-T}^{T} \frac{\sin ht}{t} \{ \cos[(x-a)t] + i \sin[(x-a)t] \} f(x) dt \right] dx \\ &= \frac{2}{\pi} \int_{-\infty}^{+\infty} \left\{ \int_{0}^{T} \frac{\sin ht}{t} \cos[(x-a)t] f(x) dt \right\} dx \end{aligned}$$

By the formula

$$\sin A \cos B = \frac{1}{2} [\sin(A+B) + \sin(A-B)]$$

and the substitution $A = ht, B = xt - at$, we obtain

$$\begin{aligned} J &= \int_{-\infty}^{+\infty} \left\{ \frac{1}{\pi} \int_{0}^{T} \frac{\sin[(x-a+h)t]}{t} dt \right. \\ &= \int_{-\infty}^{+\infty} g(x,T) f(x) dx \end{aligned}$$

where $g(x,T)$ denotes the expression in the braces. It is known from mathematical analysis that the integral $\int_0^T (\sin x/x) dx$ is bounded for all $T > 0$ and converges to $\frac{1}{2}\pi$ as $T \to +\infty$. It follows that the expression $|g(x,T)|$ is bounded and

$$\lim_{T \to \infty} \frac{1}{\pi} \int_{0}^{T} \frac{\sin \alpha t}{t} dt = \begin{cases} \frac{1}{2} & \text{for } \alpha > 0 \\ -\frac{1}{2} & \text{for } \alpha < 0 \end{cases}$$

Here the convergence is uniform with respect to α where $|\alpha| = |x - a \pm h| > \delta > 0$. From this fact we obtain

$$\lim_{T\to\infty} g(x,T) = \begin{cases} 0 & \text{for } x < a-h \\ \frac{1}{2} & \text{for } x = a-h \\ 1 & \text{for } a-h < x < a+h \\ \frac{1}{2} & \text{for } x = a+h \\ 0 & \text{for } x > a+h \end{cases}$$

It follows that in computing $\lim_{T\to\infty} J$ we can pass to the limit under the integral sign on the right-hand side. Hence we obtain

$$\begin{aligned} \lim_{T\to\infty} J &= \int_{-\infty}^{+\infty} \lim_{T\to\infty} g(x,T) f(x) dx \\ &= \int_{a-h}^{a+h} f(x)dx = F(a+h) - F(a-h) \end{aligned}$$

From above equation and we obtained. Thus the theorem is proved for a random variable of the continuous type. For a random variable of the discrete type the proof is similar; it is only necessary to replace the integrals by series. If the characteristic function $\phi(t)$ is absolutely integrable over the interval $(-\infty, +\infty)$, then the corresponding density function $f(x)$ can be determined [1] by $\phi(t)$. In fact, from the absolute integrability of the function $\phi(t)$ it follows that the improper integral (above eqaution) exists. Dividing both sides of equation by $2h$, we then have

$$\frac{F(x+h) - F(x-h)}{2h} = \frac{1}{2\pi} \int_{-\infty}^{+\infty} \frac{\sin ht}{ht} e^{-itx} \phi(t) dt \qquad (1.50)$$

where $x + h$ and $x - h$ are continuity points of $F(x)$. When $h \to 0$, the expression under the integral sign tends to $e^{-itx}\phi(t)$. Moreover, the expression under the integral sign is, in absolute value, not greater than $|\phi(t)|$, which by assumption is integrable. It follows that we can pass to the limit with $h \to 0$ under the integral sign in expression (1.50). Then we obtain

$$\lim_{h\to 0} \frac{F(x+h) - F(x-h)}{2h} = \frac{1}{2\pi} \int_{-\infty}^{+\infty} e^{itx} \phi(t) dt.$$

Since the right-hand side of this equation is a continuous function of x, we obtain

$$F'(x) = f(x) = \frac{1}{2\pi} \int_{-\infty}^{+\infty} e^{-itx} \phi(t) dt \tag{1.51}$$

From the absolute and uniform convergence of the last integral it follows that the density $F'(x)$ exists and is a continuous function. Thus formula (1.50) allows us to determine the density $f(x)$ from the characteristic function $\phi(t)$, under the assumption that $\phi(t)$ is absolutely integrable.

Example 1.5.1 *The characteristic function of the random variable X is given by the formula*

$$\phi(t) = \exp\left(-\frac{t^2}{2}\right) \tag{1.52}$$

From the above equation, we obtain

$$\begin{aligned} f(x) &= \frac{1}{2\pi} \int_{-\infty}^{+\infty} \exp(-itx) \exp\left(-\frac{t^2}{2}\right) dt \\ &= \frac{1}{2\pi} \int_{-\infty}^{+\infty} \exp\left[-\frac{(t+ix)^2}{2}\right] \exp\frac{(ix)^2}{2} dt \\ &= \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right) \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} \exp\left[-\frac{(t+ix)^2}{2}\right] dt = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right) \end{aligned}$$

If the random variable X is of the discrete type and can take on only integer values, then its probability function can easily be obtained from the characteristic function For every integer k, let

$$p_k = P(X = k)$$

where, of course, not all p_k must be positive. We have

$$\phi(t) = \sum_{k=-\infty}^{\infty} p_k e^{ikt}$$

Let k' be a fixed integer. Then we have

$$e^{-itk'} \phi(t) = \sum_{\substack{k=-\infty \\ k \neq k'}}^{+\infty} e^{-it(k'-k)} p_k + p_{k'}$$

Integrating both sides of this equation from $-\pi$ to $+\pi$ and using the fact that for every

$k \neq k'$, we have

$$\int_{-\pi}^{\pi} e^{-it(k'-k)}dt = 0$$

we obtain, replacing k' by k,

$$p_k = \frac{1}{2\pi}\int_{-\pi}^{\pi} e^{-itk}\phi(t)dt \tag{1.53}$$

Example 1.5.2 *Let us find the density function of the random variable X, whose characteristic function is*

$$\phi_1(t) = \begin{cases} 1-|t| & \textit{for} \quad |t| \leqslant 1 \\ 0 & \textit{for} \quad |t| > 1 \end{cases} \tag{4.5.8}$$

It is obvious that the function $\phi_1(t)$ is absolutely integrable over the interval ($-\infty < t < +\infty$). From formula, we obtain

$$f(x) = \frac{1}{2\pi}\int_{-\infty}^{+\infty} e^{-itx}\phi_1(t)dt = \frac{1}{2\pi}\int_{-1}^{0}(1+t)e^{-itx}dt + \frac{1}{2\pi}\int_{0}^{1}(1-t)e^{-itx}dt$$

$$\begin{aligned}\int_{-1}^{0}(1+t)e^{-itx}dt &= \left[\frac{e^{-itx}}{-ix}(1+t)\right]_{-1}^{0} - \frac{1}{-ix}\int_{-1}^{0} e^{-itx}dt \\ &= -\frac{1}{ix} + \frac{1}{ix}\left[\frac{e^{-itx}}{-ix}\right]_{-1}^{0} \\ &= -\frac{1}{ix} - \frac{1}{(ix)^2}\left(1-e^{ix}\right) \\ \int_{0}^{1}(1-t)e^{-itx}dt &= \left[\frac{e^{-itx}}{-ix}(1-t)\right]_{0}^{1} + \frac{1}{-ix}\int_{0}^{1} e^{-itx}dt \\ &= \frac{1}{ix} - \frac{1}{ix}\left[\frac{e^{-itx}}{-ix}\right]_{0}^{1} = \frac{1}{ix} + \frac{1}{(ix)^2}\left(e^{-ix}-1\right)\end{aligned}$$

We then have

$$f(x) = \frac{1}{2\pi x^2}\left(2-e^{ix}-e^{-ix}\right) = \frac{1}{\pi x^2}\left(1-\frac{e^{ix}+e^{-ix}}{2}\right) = \frac{1-\cos x}{\pi x^2} \tag{1.54}$$

Let us now consider the random variable Y of the discrete type, with the probability *function defined by the formulas*

$$P(Y=0)=\frac{1}{2} \tag{1.54 a}$$
$$P[Y=(2k-1)\pi]=\frac{2}{(2k-1)^2\pi^2} \quad (k=0,\pm1,\pm2,\ldots)$$

The characteristic function of this random variable is

$$\begin{aligned}\phi_2(t) &= \frac{1}{2}+\sum_{k=-\infty}^{\infty}\frac{2}{(2k-1)^2\pi^2}e^{it(2k-1)\pi} \\ &= \frac{1}{2}+\frac{2}{\pi^2}\sum_{k=-\infty}^{+\infty}\frac{\cos(2k-1)t\pi+i\sin(2k-1)t\pi}{(2k-1)^2} \\ &= \frac{1}{2}+\frac{4}{\pi^2}\sum_{k=1}^{\infty}\frac{\cos(2k-1)t\pi}{(2k-1)^2}\end{aligned} \tag{1.54 b}$$

We shall show that for $|t|\leqslant 1$ *we have*

$$\phi_1(t)=\phi_2(t)$$

Let Us Sum Up

Learners, in this section we have seen that determination of the distribution function by the characteristic function and also given theorems and applications.

Check Your Progress

1.To recover the distribution function $F_X(x)$ from its characteristic function $\phi_X(t)$, one must use:
A. The fourier transform of $\phi_X(t)$.
B. The inverse Fourier transform of $\phi_X(t)$.
C. The Laplace transform of $\phi_X(t)$.
D. The moment generating function of $\phi_X(t)$.
2. If the characteristic function $\phi_X(t)$ of a random variable X is given, the distribution function $F_X(x)$ can be computed by:
A. Differentiating $\phi_X(t)$ with respect to t and evaluating at $t=0$.

B. Integrating $\phi_X(t)$ with respect to t and evaluating at $t = 0$.
C. Using the inverse Fourier transform of $\phi_X(t)$.
D. Computing $\mathbb{E}[e^{itX}]$ for various values of t.

1.6 Characteristic Function of Multidimensional Random Vectors

Expanding the function $\psi(t) = |t|$ in the interval $|t| \leqslant 1$ in a Fourier series, we have

$$\psi(t) = \frac{a_0}{2} + \sum_{n=1}^{\infty} a_n \cos n\pi t$$

We compute the coefficients of this expansion from the formulas

$$\begin{aligned}
\frac{a_0}{2} &= \int_0^1 t dt = \frac{1}{2} \\
a_n &= 2\int_0^1 t \cos n\pi t dt = \left[\frac{2t \sin n\pi t}{n\pi}\right]_0^1 - \frac{2}{n\pi}\int_0^1 \sin n\pi t dt \\
&= -\frac{2}{n\pi}\left[\frac{-\cos n\pi t}{n\pi}\right]_0^1 = 2\frac{\cos n\pi - 1}{\pi^2 n^2}
\end{aligned}$$

For even n we have $a_n = 0$, and for odd n, that is, for $n = 2k-1$, we have

$$a_{2k-1} = -\frac{4}{(2k-1)^2\pi^2}$$

Finally we obtain

$$\psi(t) = |t| = \frac{1}{2} - \frac{4}{\pi^2}\sum_{k=1}^{\infty}\frac{\cos(2k-1)\pi t}{(2k-1)^2} \tag{1.55}$$

From above formulas and we have $\phi_2(t) = 1 - |t| = \phi_1(t)$, in spite of the fact that $\phi_1(t)$ and $\phi_2(t)$ are the characteristic functions of two different distributions. We observe that for $|t| > 1$ the characteristic functions $\phi_1(t)$ and $\phi_2(t)$ are not equal. In fact, from the definition we then have $\phi_1(t) \equiv 0$, whereas the function $\phi_2(t)$ is not identically zero since the values taken by this function in the interval $|t| \leqslant 1$ repeat periodically. The notion of the characteristic function of a one-dimensional random variable can be generalized to a random variable with an arbitrary finite number of dimensions. We restrict ourselves to two-dimensional random variables. Let (X, Y) be a two-dimensional random variable and let $F(x, y)$ be its distribution function. Let t and u be two arbitrary real numbers. The characteristic function of the random variable (X, Y) or of the distribution function $F(x, y)$ is defined by the formula

$$\phi(t, u) = E\left[e^{i(tX+uY)}\right] \tag{1.56}$$

Example 1.6.1 *The two-dimensional random variable can take on four pairs of values:* $(+1, +1), (+1, -1), (-1, +1)$*, and* $(-1, -1)$ *with the probabilities*

$$P(X = 1, Y = 1) = \frac{1}{3}, P(\dot{X} = 1, Y = -1) = \frac{1}{3}$$
$$P(X = -1, Y = 1) = \frac{1}{6}, P(X = -1, Y = -1) = \frac{1}{6}$$

The reader can verify that X *and* Y *are independent. For the characteristic function of the random variable (* X, Y *), we obtain from the equation* $\phi(t, u) = E\left(e^{i(tX+uY)}\right) = \frac{1}{3}e^{i(t+u)} + \frac{1}{3}e^{i(t-u)} + \frac{1}{6}e^{i(-t+u)} + \frac{1}{6}e^{i(-t-u)}$

$$= \frac{1}{3}e^{it}\left(e^{iu} + e^{-iu}\right) + \frac{1}{6}e^{-it}\left(e^{iu} + e^{-iu}\right) = \frac{1}{6}\left(e^{iu} + e^{-iu}\right)\left(2e^{it} + e^{-it}\right)$$
$$= \frac{1}{3}\cos u(3\cos t + i\sin t)$$

We shall investigate some of the properties of characteristic functions of multidimensional random variables. We have

$$\phi(0, 0) = E\left(e^{i(0X+0Y)}\right) = 1|\phi(t, u)| = \left|E\left(e^{i(tX+uY)}\right)\right| \leqslant E\left(\left|e^{i(tX+uY)}\right|\right) = 1 \tag{1.57}$$

Hence

$$|\phi(t, u)| \leqslant 1 \tag{1.58}$$

$$\phi(-t, -u) = E\left(e^{-i(tX+uY)}\right) = \overline{\phi(t, u)}$$

It can be shown that, as in the one-dimensional case, if all the moments of order k *of a multidimensional random variable exist, then the derivatives*

$$\frac{\partial^k \phi(t, u)}{\partial t^{k-l} \partial u^l} \quad \textit{for } l = 0, 1, 2, \ldots, k \tag{1.59}$$

exist and can be obtained from the formula

$$\frac{\partial^k \phi(t,u)}{\partial t^{k-l}\partial u^l} = i^k E\left(X^{k-l}Y^l e^{i(tX+uY)}\right) \tag{1.60}$$

From the above equation, we see that the moment $m_{k-l,l}$ can be obtained from the formula

$$m_{k-l,l} = E\left(X^{k-l}Y^l\right) = \frac{1}{i^k}\left[\frac{\partial^k \phi(t,u)}{\partial t^{k-l}\partial u^l}\right]_{\substack{t=0\\u=0}} \tag{1.61}$$

For the moments of the first and second order we obtain the expressions $m_{10} = \frac{1}{i}\left[\frac{\partial\phi(t,u)}{\partial t}\right]_{\substack{t=0\\u=0}}$, $m_{01} = \frac{1}{i}\left[\frac{\partial\phi(t,u)}{\partial u}\right]_{\substack{t=0\\u=0}}$, $m_{20} = \frac{1}{i^2}\left[\frac{\partial^2\phi(t,u)}{\partial t^2}\right]_{\substack{t=0\\u=0}}$, $m_{11} = \frac{1}{i^2}\left[\frac{\partial^2\phi(t,u)}{\partial t\partial u}\right]_{\substack{t=0\\u=0}}$, $m_{02} = \frac{1}{i^2}\left[\frac{\partial^2\phi(t,u)}{\partial u^2}\right]_{\substack{t=0\\u=0}}$. *We obtain the characteristic functions of the marginal distributions of the random variables X and Y from formula of equation by putting $u = 0$ or $t = 0$, respectively. Thus*

$$\phi(t,0) = E\left(e^{itX}\right) = \phi_1(t) \tag{1.62}$$

This is simply the characteristic function of X. Similarly,

$$\phi(0,u) = E\left(e^{itY}\right) = \phi_2(u) \tag{1.63}$$

is the characteristic function of Y. We shall now give without proof the generalization of theorem to two-dimensional random vectors. The proof is similar to that for a one-dimensional random variable.

Theorem 1.6.2 *Let $\phi(t,u)$ be the characteristic function of the random variable (X,Y). If the rectangle $(a-h \leqslant X < a+h, b-g \leqslant Y < b+g$) is a continuity rectangle, then*

$$P(a-h \leqslant X < a+h, b-g \leqslant Y < b+g)$$

$$= \lim_{T\to\infty}\frac{1}{\pi^2}\int_{-T}^{+T}\int_{-T}^{+T}\frac{\sin ht}{t}\frac{\sin gu}{u}\exp[-i(at+bu)]\phi(t,u)dtdu$$

Thus, if we know $\phi(t,u)$, formula (1.63) allows us to determine the probability

$$P\left(x_1 \leqslant X < x_2, y_1 \leqslant Y < y_2\right) \tag{1.64}$$

for an arbitrary continuity rectangle. However, the probabilities (4.6.12) for continuity *rectangles completely determine the probability distribution in the plane* (x, y).

Theorem 1.6.3 *Let* $F(x,y), F_1(x), F_2(y), \phi(t,u), \phi_1(t)$*, and* $\phi_2(u)$ *denote the distribution functions and the characteristic functions of the random variables* $(X,Y), X$ *and* Y*, respectively. The random variables* X *and* Y *are then independent if and only if the equation*

$$\phi(t,u) = \phi_1(t)\phi_2(u) \tag{1.65}$$

holds for all real t *and* u*.*

Proof: *Suppose that* X *and* Y *are independent. From theorem we have, for any real* t *and* u

$$\phi(t,u) = E\left(e^{i(tX+uY)}\right) = E\left(e^{itX}e^{iuY}\right) = E\left(e^{itX}\right)E\left(e^{iuY}\right) = \phi_1(t)\phi_2(u)$$

We obtain the equation

$$P\left(X_1 \leqslant X < x_2, y_1 \leqslant Y < y_2\right) = P\left(x_1 \leqslant X < x_2\right)P\left(y_1 \leqslant Y < y_2\right)$$

which is valid for arbitrary continuity rectangles. From the above equation we obtain, for arbitrary x *and* y

$$F(x,y) = F_1(x)F_2(y)$$

Thus the theorem is proved. The following Cramer-Wold theorem is useful in the theory of random vectors.

Theorem 1.6.4 *Prove that distribution function* $F(x,y)$ *of a two-dimensional random variables* (X,Y) *is uniquely determined by the class of all one-dimensional distribution functions of* $tX+uY$*, where* t *and* u *run over all possible real values.*

Proof: *Suppose we are given for all real* t *and* u *the characteristic functions* $\phi_z(v)$ *of* $Z = tX+uY, \quad \phi_z(v) = E\{\exp[iv(tX+uY)]\} = E\{\exp[i(vtX+vuY)]\}$*. Putting* $v=1$*, we obtain for the right-hand side of the expression*

$$E\{\exp[i(tX+uY)]\}$$

which is the characteristic function $\phi(t,u)$ *of the distribution function* $F(x,y)$*. According to the function* $\phi(t,u)$ *uniquely determines* $F(x,y)$*. Thus the theorem is proved. Let us write*

$$P(tX+uY<z) = P(X\cos\alpha + Y\sin\alpha < w)$$

where $\cos\alpha = \frac{t}{\sqrt{t^2+u^2}}, \quad \sin\alpha = \frac{u}{\sqrt{t^2+u^2}}, \quad w = \frac{z}{\sqrt{t^2+u^2}}(0 \leqslant \alpha \leqslant 2\pi)$. *The Cramer-Wold theorem can now be formulated in the following way: The distribution function* $F(x, y)$ *is uniquely determined by the distribution functions of the projections of* (X, Y) *on all straight lines passing through the origin. With probability* $1, (X, Y)$ *satisfies the inequality*

$$X^2 + Y^2 \leqslant R^2 < \infty$$

then the distribution function $F(x, y)$ *is uniquely determined by the class of distribution functions of* $X \cos\alpha_1 + Y \sin\alpha$*, where* α *runs over an arbitrary countable set of different values from the interval* $[0, 2\pi]$.

Let Us Sum Up

Learners, in this section we have seen that he characteristic function of multidimensional random vectors and applications.

Check Your Progress

1. The characteristic function of a multidimensional random vector $\mathbf{X} = (X_1, X_2, \ldots, X_d)^T$ is defined as:
A. $\mathbb{E}[e^{i\mathbf{t}^T\mathbf{X}}]$
B. $\mathbb{E}[e^{\mathbf{t}^T\mathbf{X}}]$
C. $\mathbb{E}[e^{i\mathbf{t}\mathbf{X}}]$
D. $\mathbb{E}[e^{\mathbf{t}\mathbf{X}}]$
2. If $\mathbf{X}$ and $\mathbf{Y}$ are independent d-dimensional random vectors with characteristic functions $\phi_{\mathbf{X}}(\mathbf{t})$ and $\phi_{\mathbf{Y}}(\mathbf{t})$, respectively, the characteristic function of $\mathbf{X} + \mathbf{Y}$ is:
A. $\phi_{\mathbf{X}}(\mathbf{t}) \cdot \phi_{\mathbf{Y}}(\mathbf{t})$
B. $\phi_{\mathbf{X}}(\mathbf{t}) + \phi_{\mathbf{Y}}(\mathbf{t})$
C. $\phi_{\mathbf{X}}(\mathbf{t}) \cdot \phi_{\mathbf{Y}}(-\mathbf{t})$
D. $\phi_{\mathbf{X}}(\mathbf{t}) + \phi_{\mathbf{Y}}(-\mathbf{t})$

1.7 Probability-Generating Functions

When investigating random variables which take on only the integers $k = 0, 1, 2, \ldots$ it is simpler to deal with probability generating functions than with characteristic functions. Let X be a random variable and let

$$p_k = P(X = k) \quad (k = 0, 1, 2, \ldots)$$

where $\sum_k p_k = 1$.

Definition 1.7.1 *The function defined by the formula*

$$\psi(s) = \sum_k p_k s^k \tag{1.66}$$

where $-1 \leqslant s \leqslant 1$, is called the probability generating function of X. We notice that $\psi(1) = \sum_k p_k = 1$. Hence the series on the right-hand side of above eqaution is absolutely and uniformly convergent in the interval $|s| \leqslant 1$. Thus the generating function is continuous. It determines the probability function uniquely, since $\psi(s)$ can be represented in a unique way as a power series of the form the above eqaution.

Example 1.7.2 *The random variable X has a binomial distribution, that is,*

$$p_k = \binom{n}{k} p^k (1-p)^{n-k} \quad (k = 0, 1, \ldots, n)$$

Therefore

$$\psi(s) = \sum_{k=0}^{n} \binom{n}{k} (ps)^k (1-p)^{n-k} = (ps + q)^n$$

Example 1.7.3 *The random variable X has a Poisson distribution, that is*

$$p_k = e^{-\lambda} \frac{\lambda^k}{k!} \quad (k = 0, 1, 2, \ldots)$$

Therefore

$$\psi(s) = \sum_{k=0}^{\infty} e^{-\lambda} \frac{(\lambda s)^k}{k!} = e^{-\lambda} e^{\lambda s} = e^{-\lambda(1-s)} \tag{1.67}$$

The moments of the random variable X can be determined by the derivatives at the point 1 of the generating function. Let us for example, determine the moments of the first and *second order. We have*

$$\psi'(s) = \sum_k k p_k s^{k-1}$$
$$\psi''(s) = \sum_k k(k-1) p_k s^{k-2}$$

1.8 Let Us Sum Up

Learners, in this section we have seen that probability-generating functions and applications.

Check Your Progress

1. If X is a discrete random variable with probability mass function $p_X(x)$, the probability generating function $G_X(t)$ is given by:
A. $G_X(t) = \sum_x e^{tx} p_X(x)$
B. $G_X(t) = \sum_x t^x p_X(x)$
C. $G_X(t) = \sum_x p_X(x) e^{tx}$
D. $G_X(t) = \sum_x e^{itx} p_X(x)$
2. Which of the following properties is true about the probability generating function $G_X(t)$ of a discrete random variable X?
A. $G_X(t)$ is always a real-valued function.
B. $G_X(1) = 1$.
C. $G_X(t)$ is always a polynomial.
D. $G_X(t)$ can be used to compute the mean and variance of X directly.

Glossary

1. The $\phi(t) = E\left(e^{itX}\right)$ is characteristic function of t.
2. The function $f(x)$ is density function of x.
3. The function $M_X(t)$ is moment generating function of t.
4. The μ_r is rth order moment.

1.9 Unit Summary

The first unit content on properties of characteristic functions, characteristic function and moments, semi-invariants, the characteristic functions of sum of independent random variables, determination of distribution function of the characteristic function and probability generating function.

Self-Assessment Questions

Short Answers: (5 Marks)

1. Prove that the function $\varphi(t) = \exp(-|t|^r)$ with $r > 2$ is not the characteristic function of any random variable.

2. Prove that the characteristic function of a random variable X is real if and only if X has a symmetric distribution about 0.

3. Let $(\varphi(t)$ be the characteristic function of the random variable X. Prove that, if X is of the continuous type, then $\lim_{|t|\to\infty} (\varphi(t) = 0$.

4. Let $(\varphi(t)$ be the characteristic function of the random variable X then prove that if X is of the discrete type, then $\lim_{|t|\to\infty}$ sup $(\varphi(t) = 1$.

5. Prove that if $(\varphi(t)$ is the characteristic function of a random variable then so is $|(\varphi(t)|^2$.

Long Answers: (8 Marks)

1. Let $(\varphi(t)$ be the characteristic function of the random variable X. Prove that
   (a) if X is of the continuous type, then $\lim_{|t|\to\infty} (\varphi(t) =$ O
   (b) if X is of the Discrete type, then $\lim_{|t|\to\infty}$ sup $(\varphi(t)$=1.

2. Show that the characteristic functions $(\varphi_1(t)$, $(\varphi_2(t)$, and $(\varphi_3(t)$may satisfy the relation $(\varphi_1(t)(\varphi_2(t) = (\varphi_1(t)(\varphi_3(t)$
   in spite of the fact that $(\varphi_2(t)$, and $(\varphi_3(t)$are not identically equal.

3. Find the characteristic function and the moments of a normal distribution

$$f(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}.$$

4. The characteristic function of the random variable X is given by:

$$\phi(t) = \exp\left(-\frac{t^2}{2}\right).$$

Exercises

1. Find the characteristic functions of the random variables whose densities are (a)

$$f(x) = \begin{cases} 0 & \text{for} \quad |x| \geqslant a > 0 \\ \frac{a-1}{a^2} & \text{for} \quad |x| < a \end{cases}$$

(b)

$$f(x) = \frac{2\sin^2(ax/2)}{\pi a x^2}$$

2. Prove that the function

$$\phi(t) = \exp\left(-|t|^r\right)$$

with $r > 2$ is not the characteristic function of any random variable.

3. Let $\phi(t)$ be the characteristic function of the random variable X. Prove that (a) if X is of the continuous type, then $\lim_{|t|\to\infty} \phi(t) = 0$, (b) if X is of the discrete type, then $\lim_{|t|\to\infty} \sup |\phi(t)| = 1$.

4. Prove that (a) if $\phi(t)$ is the characteristic function of a random variable, then so is $|\phi(t)|^2$.

5. Prove that the characteristic function of a random variable X is real if and only if X has a symmetric distribution about 0.

Answers to Check Your Progress

Section (Modulo) 1.1

1. B. $Z = 0$

2. A. $E\left(e^{itX}\right)$

Section (Modulo) 1.2

1. B. The characteristic function is defined for all real t, while the MGF is defined only for t in a neighborhood around zero.

2. A. Differentiating $\phi_X(t)$ with respect to t and then evaluating at $t = 0$.

Section (Modulo) 1.3

1. A. The expectation of X^k for some integer k.

2. D. The characteristic function of X.

Section (Modulo) 1.4

1. A.$\phi_X(t) \cdot \phi_Y(t)$

2. D. $\phi_{X_1}(t) \cdot \phi_{X_2}(t) \cdots \phi_{X_n}(t)$

Section (Modulo) 1.5

1. B. The inverse Fourier transform of $\phi_X(t)$.

2. C. Using the inverse Fourier transform of $\phi_X(t)$.

Section (Modulo) 1.6

1. A.$\mathbb{E}[e^{i\mathbf{t}^T\mathbf{X}}]$

2. A.$\phi_{\mathbf{X}}(\mathbf{t}) \cdot \phi_{\mathbf{Y}}(\mathbf{t})$

Section (Modulo) 1.7

1. B. $G_X(t) = \sum_x t^x p_X(x)$

2. B. $G_X(1) = 1$.

References

1. M. Fisz, Probability Theory and Mathematical Statistics, John wiley and sons, New Your, Third Edition, 1963.

Suggested Readings

1. T. Veerarajan, Fundamentals of Mathematical Statistics, Yesdee Publishing, 2017.

2. P. R. Vittal, Mathematical Statistics, Margham Publications, 2002.

3. R.S.N. Pillai and V. Bagavathi, Statistics, Sultan Chand and Sons, 2010.

4. S. C. Gupta and V. K. Kapoor, Fundamentals of Mathematical Statistics, Sultan Chand and Sons, 2008.

Unit 2

Probability Discrete Distributions

Objective

This course aims to teach the students about some discrete probability distributions with one-point and two-point distributions. The Binomial distribution, Poisson and generalized binomial distribution and Polya and hyper-geometric distributions. Also Poisson distribution and uniform distribution.

2.1 One-Point and Two-Point Distributions

In this section we investigate to more closely some probability distributions of special importance in either theory or practice. We begin with the one-point distribution.

Definition 2.1.1 *The random variable X has a one-point distribution if there exists a point x_0 such that*

$$P\left(X = x_0\right) = 1 \tag{2.1}$$

We also say that the probability mass is concentrated at one point. It is clear that a random variable with a one-point distribution has a degenerate distribution. Formula (2.1) gives us the probability function. The distribution function of this probability distribution is given by the formula

$$F(x) = \begin{cases} 0 & \text{for } x \leqslant x_0 \\ 1 & \text{for } x > x_0 \end{cases} \tag{2.2}$$

We obtain the characteristic function of this distribution from the formula

$$\phi(t) = e^{itx_0} \tag{2.3}$$

We know that see that $m_1 = x_0$ and, more generally, we have $m_k = {x_0}^k$ for every k. Hence we obtain

$$D^2(X) = m_2 - m_1^2 = x_0^2 - x_0^2 = 0$$

Conversely, if the variance of a random variable X equals zero, then X has a one-point distribution. To prove this suppose that

$$D^2(X) = E[X - E(X)]^2 = 0 \tag{2.4}$$

Let Us Sum Up

Learners, in this section we have seen that definitions of one point and two point distributions.

Check Your Progress

1. Which of the following is a property of the binomial distribution?
A. The number of trials is infinite
B. The probability of success is constant across trials
C. The trials are not independent
D. The distribution is continuous
2. What is the mean of a normal distribution with parameters μ and σ^2?
A. σ
B. σ^2
C. μ

D. $\mu + \sigma^2$

2.2 Bernoulli Scheme of Binomial Distribution

Since the expression $[X - E(X)]^2$ is non-negative, equation (2.4) is satisfied only if;

$$P[X - E(X) = 0] = 1, \quad \text{or} \quad P[X = E(X)] = 1$$

From (2.1) we find that the random variable X has a one-point distribution.

Definition 2.2.1 *The random variable X has a two-point distribution if there exist two values x_1 and x_2, such that;*

$$P(X = x_1) = p, \quad P(X = x_2) = 1 - p \quad (0 < p < 1) \tag{2.5}$$

We often put $x_1 = 1$ and $x_2 = 0$. In place of (2.5) we then have

$$P(X = 1) = p, \quad P(X = 0) = 1 - p \quad (0 < p < 1) \tag{2.6}$$

This distribution is called the zero-one distribution. The characteristic function of distribution (2.6) is given by the formula;

$$\phi(t) = pe^{it \cdot 1} + (1 - p)e^{it \cdot 0} = pe^{it} + 1 - p = 1 + p\left(e^{it} - 1\right) \tag{2.7}$$

We obtain for every k

$$m_k = p \tag{2.8}$$

Hence

$$D^2(X) = m_2 - m_1^2 = p - p^2 = p(1 - p) \tag{2.9}$$

Then,

$$\mu_3 = m_3 - 3m_1m_2 + 2m_1^3 = p - 3p^2 + 2p^3 = p(1 - p)(1 - 2p)$$
$$\gamma = \frac{\mu_3}{\mu_2^{3/2}} = \frac{p(1 - p)(1 - 2p)}{p^{3/2}(1 - p)^{3/2}} = \frac{1 - 2p}{\sqrt{p(1 - p)}}$$

We see that if $p = 0.5$ then $\gamma = 0$ since here X has a symmetric distribution. From the following scheme of trials, called the Bernoulli scheme, we obtain a random variable X with binomial distribution. We perform n random experiments. Through an experiment we can obtain the event A, which we designate a success, with probability p, or the complementary A, which we designate a failure, with probability $q = 1 - p$. The results of the n experiments are independent. As a result of n random experiments, event A may occur k times ($k = 0, 1, 2, \ldots, n$). The number of occurrences of A is a random variable X that can take on the values $k = 0, 1, \ldots, n$, where the equality $X = k$ means that in n experiments the event A has occurred k times. It is that X has the binomial probability function given by the formula

$$P(X = k) = \binom{n}{k} p^k (1 - p)^{n-k} \tag{2.10}$$

The distribution function of the binomial distribution is given by the formula

$$F(x) = P(X < x) = \sum_{k<x} \binom{n}{k} p^k (1 - p)^{n-k}$$

where the summation extends over all non-negative integers less than x. We notice that for $n = 1$ the binomial distribution is reduced to the zero-one distribution. For $n \geqslant 2$ the binomial distribution can also be obtained from the zero-one distribution as follows. Let $X_r(r = 1, 2, \ldots, n)$ be independent random variables with the same zero-one distribution. The probability function of every X_r has the form

$$P\left(X_r = 1\right) = p, \quad P\left(X_r = 0\right) = 1 - p$$

Consider the random variable equal to the sum of the X_r,

$$X = X_1 + X_2 + \ldots + X_n \tag{2.11}$$

The random variable X can take on the values $k = 0, 1, 2, \ldots, n$. The event $X = k$ occurs if and only if k of the n random variables X_r take on the value one and $n - k$ of them take on the value zero. For a given k this may happen in $\binom{n}{k}$ different ways. By the independence of the random variables X_r we obtain formulas (2.10). From (2.7) and (2.2), for the characteristic function $\phi(t)$ of X, we obtain

$$\phi(t) = \left[1 + p\left(e^{it} - 1\right)\right]^n \tag{2.12}$$

We obtain the moments of this distribution. In particular,

$$m_1 = np, \qquad m_2 = np + n(n-1)p^2 \tag{2.13}$$

$$\mu_2 = np(1-p), \qquad \mu_3 = np(1-p)(1-2p)$$

We then have

$$\gamma = \frac{1-2p}{\sqrt{np(1-p)}} \tag{2.14}$$

We have already obtained the formula for μ_2 by another method.

Example 2.2.2 *The above table gives the binomial distributions for the values $p_1 = 0.1, p_2 = 0.3$, and $p_3 = 0.5$ for $n = 20$. The first column gives the values $k = 0, 1, \dots, 20$ and the remaining columns the probabilities that the random variable takes on the value k. These probabilities are given with a maximum error of 0.00005 .*

Table

	$P(X=k)$				$P(X=k)$		
k	$p_1 = 0.1$	$p_2 = 0.3$	$p_3 = 0.5$	k	$p_1 = 0.1$	$p_2 = 0.3$	$p_3 = 0.5$
0	0.1216	0.0008	-	11	-	0.0120	0.1602
1	0.2702	0.0068	-	12	-	0.0039	0.1201
2	0.2852	0.0278	0.0002	13	-	0.0010	0.0739
3	0.1901	0.0716	0.0011	14	-	0.0002	0.0370
4	0.0898	0.1304	0.0046	15	-	-	0.0148
5	0.0319	0.1789	0.0148	16	-	-	0.0046
6	0.0089	0.1916	0.0370	17	-	-	0.0011
7	0.0020	0.1643	0.0739	18	-	-	0.0002
8	0.0004	0.1144	0.1201	19	-	-	-
9	0.0001	0.0654	0.1602	20	-	-	-
10	-	0.0308	0.1762				

We know that p is to 0.5 , the more symmetric is the distribution and the greater its dispersion. We might have expected these results from comparing the values of the *parameters μ_2 and γ for the considered values of p with $n = 20$. These values, computed from (2.4) and (2.5) are given in below table.*

	$p_1 = 0.1$	$p_2 = 0.3$	$p_3 = 0.5$
$\sigma = \sqrt{\mu_2}$	*1.34*	*2.05*	*2.24*
γ	*0.597*	*0.195*	*0.000*

Let X and Y be two independent random variables with binomial distributions and let the characteristic functions of X and Y be, respectively

$$\phi_1(t) = \left[1 + p\left(e^{it} - 1\right)\right]^{n_1}$$

$$\phi_2(t) = \left[1 + p\left(e^{it} - 1\right)\right]^{n_2}$$

Consider the random variable

$$Z = X + Y$$

Because of the independence of X and Y, the characteristic function of Z is;

$$\phi(t) = \left[1 + p\left(e^{it} - 1\right)\right]^{n_1+n_2} \tag{2.6}$$

As we see from (2.6), Z has the binomial distribution with $n = n_1 + n_2$. This is the addition theorem for the binomial distribution. In applications we often deal with the distribution of

$$Y = \frac{X}{n}$$

where the random variable X has the binomial distribution. The random variable Y can take on the values

$$\frac{k}{n} = 0, \frac{1}{n}, \ldots, \frac{n-1}{n}, 1$$

Since the probability that $Y = k/n$ is equal to the probability that $X = k$, the probability function of Y is given by (2.1)

$$P\left(Y = \frac{k}{n}\right) = P(X = k) = \binom{n}{k} p^k (1-p)^{n-k}$$

Let Us Sum Up

Learners, in this section we have seen that the definition of Bernoulli scheme of the binomial distribution and example.

Check Your Progress

1. What are the parameters of a binomial distribution?
A. Mean and variance
B. Number of trials and probability of success
C. Mean and standard deviation
D. Rate and shape
2. What is the probability mass function (pmf) of the binomial distribution:
A. $P(X=k)=\frac{n!}{k!(n-k)!}p^k(1-p)^{n-k}$
B. $P(X=k)=\frac{n!}{k!(n-k)!}(1-p)^k p^{n-k}$
C. $P(X=k)=\frac{n!}{k!(n-k)!}p^{n-k}(1-p)^k$
D. $P(X=k)=\frac{n!}{k!(n-k)!}p^k(1-p)^k$

2.3 Poisson Scheme of Generalized Binomial Distribution

We know that the characteristic function of Y is;

$$\phi(t)=\left[1+p\left(e^{it/n}-1\right)\right]^n \tag{2.7}$$

From the characteristic function we obtain the moments. In particular,

$$m_1=p, \quad m_2=\frac{p}{n}+\frac{n-1}{n}p^2, \quad \mu_2=\frac{p(1-p)}{n} \tag{2.8}$$

Poisson considered the following scheme of experiments. We perform n random trials. As a result of the k th trial $(k=1,2,\ldots,n)$, the event A (or a success) may occur with probability p_k; thus the probability of the complementary event, $q_k=1-p_k$. The results of the n experiments are independent. Unlike Bernoulli's scheme, here the probabilities of the occurrence of event A in individual trials are not necessarily equal. The number of occurrences of A in n trials is a random variable. We say that this random variable has a generalized binomial distribution. The random variable Z with the generalized binomial distribution can also be represented as the sum

$$Z = Z_1 + \ldots + Z_n \tag{2.9}$$

where the random variables $Z_k (k = 1, 2, \ldots, n)$ are independent and have the zero-one distribution with the probability functions;

$$P(Z_k = 1) = p_k, \quad P(Z_k = 0) = 1 - p_k$$

The formula for the probability function of the random variable Z is not as simple as that for the probability function of the binomial distribution. The probability that $Z = r$ can be found by the summation of the probabilities of each possible combination of r 1's and $(n - r)$0's.

Example 2.3.1 *We have three lots of oranges. The fraction of rotten oranges in the first lot is $p_1 = 0.02$, in the second, $p_2 = 0.05$, and in the third, $p_3 = 0.01$. We choose one orange at random from each lot. We assign the number one to the appearance of a good orange, and the number zero to the appearance of a rotten one. Here Z_1, Z_2, and Z_3 are random variables which take on the value 0 or 1, according to whether we have obtained a rotten or a good orange from the first, second, or third lot. These random variables are independent, and we have;*

$$P(Z_1 = 1) = 0.98, P(Z_2 = 1) = 0.95, P(Z_3 = 1) = 0.99$$

Consider the random variable $Z = Z_1 + Z_2 + Z_3$. This random variable can take on the values $r = 0, 1, 2$, or 3.

Let Us Sum Up

Learners, in this section we have seen that the defintion of Poisson scheme and the generalized Binomial distributions with examples.

Check Your Progress

1. If a binomial random variable X has parameters $n = 10$ and $p = 0.3$, what is the mean of the distribution?
A. $10 \times 0.3 = 3$
B. $10 \times 0.7 = 7$
C. $0.3 \times 0.7 = 0.21$
D. $0.3 \times 10 = 30$
2. What is the variance of a binomial distribution with parameters n and p?
A. $np(1-p)$
B. np
C. $n^2p(1-p)$
D. $n(1-p)$

2.4 Polya and Hypergeometric Distributions

As a result of the independence of Z_1, Z_2, Z_3 we obtain the probabilities

$$\begin{aligned}
P(Z=0) =& P(Z_1=0)\, P(Z_2=0)\, P(Z_3=0) = 0.00001 \\
P(Z=1) =& P(Z_1=0)\, P(Z_2=0)\, P(Z_3=1) \\
&+ P(Z_1=0)\, P(Z_2=1)\, P(Z_3=0) \\
&+ P(Z_1=1)\, P(Z_2=0)\, P(Z_3=0) = 0.00167 \\
P(Z=2) =& P(Z_1=0)\, P(Z_2=1)\, P(Z_3=1) \\
&+ P(Z_1=1)\, P(Z_2=0)\, P(Z_3=1) \\
&+ P(Z_1=1)\, P(Z_2=1)\, P(Z_3=0) = 0.07663 \\
P(Z=3) =& P(Z_1=1)\, P(Z_2=1)\, P(Z_3=1) = 0.92169
\end{aligned}$$

then

$$P(Z=0) + P(Z=1) + P(Z=2) + P(Z=3) = 1$$

The characteristic function of Z, defined by (2.1) to have a generalized binomial distribution, can be obtained from (1.7) using the independence of the Z_k, namely

$$\phi(t) = \prod_{k=1}^{n} \left[1 + p_k \left(e^{it} - 1\right)\right] \tag{2.10}$$

We compute the first two moments of Z. We have

$$m_1 = \sum_{k=1}^{n} p_k, \quad m_2 = \sum_{k=1}^{n} p_k + \sum_{l=1}^{n}\sum_{k=1}^{n} p_l p_k, \quad \mu_2 = \sum_{k=1}^{n} p_k \left(1 - p_k\right) \tag{2.11}$$

As we see, formulas (2.4) for m_1, m_2, and μ_2 are particular cases of the corresponding formulas (2.3).

Example 2.4.1 *We compute the expected value and the standard deviation of the random variable Z of example. We have*

$$E(Z) = m_1 = (1 - p_1) + (1 - p_2) + (1 - p_3) = 2.92$$
$$\sigma = \sqrt{\mu_2} = \sqrt{0.0196 + 0.0475 + 0.0099} = \sqrt{0.0770} = 0.28$$

In practice we often deal with distributions which can be reduced to a scheme called the Polya scheme. Imagine that we have b white and c black balls in an urn. Let $b + c = N$. We draw one ball at random, and before drawing the next ball we replace the one we have drawn and add s balls of the same color. This procedure is repeated n times. Denote by X the random variable which takes on the value $k (k = 0, 1, \ldots, n)$ if as a result of n drawings we draw a white ball k times. We shall find the probability function of X. We notice that the probability of the successive drawing of k white balls is;

$$\frac{b(b + s) \ldots [b + (k - 1)s]}{N(N + s) \ldots [N + (k - 1)s]}$$

Similarly, the probability of drawing k white balls in turn and then $n - k$ black balls is;

$$\frac{b(b + s) \ldots [b + (k - 1)s] c(c + s) \ldots [c + (n - k - 1)s]}{N(N + s) \ldots [N + (n - 1)s]}$$

We notice that the last expression also gives the probability of drawing k white and $n - k$ black balls in any given order. The order of drawing affects only the order of the terms in *the numerator of this expression. Since k white and $n-k$ black balls can be drawn in $\binom{n}{k}$ different ways, we have (2.1)* $\quad P(X=k)$

$$= \binom{n}{k} \frac{b(b+s)\dots[b+(k-1)s]c(c+s)\dots[c+(n-k-1)s]}{N(N+s)\dots[N+(n-1)s]}$$

Definition 2.4.2 *The random variable X with the probability distribution given by (2.1) has a Polya distribution. Denote*

$$Np = b, Nq = c, N\alpha = s$$

As we see, p and q are the probabilities of drawing a white and a black ball, respectively, on the first drawing. Formula (2.1) takes the form (2.2) is;

$$P(X=k) = \binom{n}{k} \frac{p(p+\alpha)\dots[p+(k-1)\alpha]q(q+\alpha)\dots[q+(n-k-1)\alpha]}{1(1+\alpha)\dots[1+(n-1)\alpha]}$$

It is obvious that $\sum_{k=0}^{n} \binom{n}{k} \frac{p(p+\alpha)\dots[p+(k-1)\alpha]q(q+\alpha)\dots[q+(n-k-1)\alpha]}{1(1+\alpha)\dots[1+(n-1)\alpha]} = 1$. We shall compute the first and second moments of X. The first moment is given by the formula

$$E(X) = \sum_{k=0}^{n} kP(X=k) = pn \sum_{k=1}^{n} \binom{n-1}{k-1} \times \frac{(p+\alpha)\dots[p+(k-1)\alpha]q(q+\alpha)\dots[q+(n-k-1)\alpha]}{(1+\alpha)\dots[1+(n-1)\alpha]}$$

Putting $l = k-1$, we obtain

$$E(X) = pn \sum_{l=0}^{n-1} \binom{n-1}{l} \frac{(p+\alpha)\dots(p+l\alpha)q(q+\alpha)\dots[q+(n-l-2)\alpha]}{(1+\alpha)\dots[1+(n-1)\alpha]} \tag{2.12}$$

It is easy to verify that the term under the summation sign in the last formula represents the probability of obtaining l white and $n-l-1$ black balls in $n-1$ drawings according to a Pólya scheme where at the beginning the urn contains $N+s$ balls, including $b+s$ white and c black ones. From (2.3) it follows that the sum on the right-hand side of (2.4) equals 1; hence

$$E(X) = np \tag{2.13}$$

For the second moment we obtain

$$E\left(X^2\right) = \sum_{k=0}^{n} k^2 P(X=k) = np\sum_{k=1}^{n} k\binom{n-1}{k-1}$$
$$\times \frac{(p+\alpha)\dots[p+(k-1)\alpha]q(q+\alpha)\dots[q+(n-k-1)\alpha]}{(1+\alpha)\dots[1+(n-1)\alpha]}$$

Putting $l = k-1$ we obtain $E\left(X^2\right)$

$$\begin{aligned}
=&np\sum_{l=0}^{n-1}(l+1)\binom{n-1}{l}\frac{(p+\alpha)\dots(p+l\alpha)q(q+\alpha)\dots[q+(n-l-2)\alpha]}{(1+\alpha)\dots[1+(n-1)\alpha]}\\
=&np\left(\sum_{l=0}^{n-1} l\binom{n-1}{l}\frac{(p+\alpha)\dots(p+l\alpha)q(q+\alpha)\dots[q+(n-l-2)\alpha]}{(1+\alpha)\dots[1+(n-1)\alpha]}\right.\\
&\left.+\sum_{l=0}^{n-1}\binom{n-1}{l}\frac{(p+\alpha)\dots(p+l\alpha)q(q+\alpha)\dots[q+(n-l-2)\alpha]}{(1+\alpha)\dots[1+(n-1)\alpha]}\right\}\\
=&np(A+B).
\end{aligned}$$

After some simple transformations we have

$$A = \frac{(p+\alpha)(n-1)}{1+\alpha}\sum_{r=0}^{n-2}\binom{n-2}{r} \tag{2.14}$$

$$\times\frac{(p+2\alpha)\dots[p+(r+1)\alpha]q\dots[q+(n-r-3)\alpha]}{(1+2\alpha)\dots[1+(n-1)\alpha]}$$

Expression B is identical with the sum in (2.14); hence $B=1$. We notice further that the term under the summation sign in (2.6) is the probability of drawing r white and $n-r-2$ black balls in $n-2$ drawings according to a Polya scheme where the urn contains $N+2s$ balls at the beginning, among which $b+2s$ are white and c are black. Thus we obtain from (2.3)

$$A = \frac{(p+\alpha)(n-1)}{1+\alpha}$$

Finally,

$$E\left(X^2\right) = np\left[\frac{(p+\alpha)(n-1)}{1+\alpha}+1\right] = np\frac{np+q+n\alpha}{1+\alpha} \tag{2.15}$$

Using (2.5), we obtain

$$D^2(X) = npq\frac{1+n\alpha}{1+\alpha} \tag{2.16}$$

The Polya scheme can be applied to such phenomena as infectious diseases where the realization of an event (appearance of the disease) causes an increase in the probability of being infected with the disease. In the Polya scheme s may also be negative. Since the inequalities

$$b + (k-1)s \geqslant 1 \text{ and } c + (n-k-1)s \geqslant 1$$

must hold, k must then satisfy the double inequality

$$\max\left(0, n-1+\frac{c-1}{s}\right) \leqslant k \leqslant \min\left(n, \frac{1-b}{s}+1\right)$$

Let $N, b,$ and c tend to infinity so that

$$p = \frac{b}{N} = \text{ constant.} \tag{2.17}$$

Here, of course, $q = 1 - p$ is also constant. Suppose that $\lim \alpha = 0$. $N \to \infty$ this condition will be satisfied, in particular, if s is constant and N tends to infinity. It follows from (2.1) and (2.2) that

$$\lim_{N\to\infty} P(X=k) = \binom{n}{k} p^k q^{n-k} \tag{2.18}$$

We have proved the following theorem.

Theorem 2.4.3 *If for $N = 1, 2, \ldots$ equality (2.10) is satisfied and lim $\alpha = 0$, then the probability function of the random variable X with the $N \to \infty$ Polya distribution tends to the probability function of the binomial distribution as $N \to \infty$. A particular case of the Polya distribution is the hypergeometric distribution. In this distribution $s = -1$, which simply means that we do not replace the ball which has been drawn before drawing the next ball. The probability function of the hypergeometric distribution can be obtained from (2.2) by putting $\alpha = -1/N$. We obtain for k satisfying the double inequality* $\max(0, n - Nq) \leqslant k \leqslant \min(n, Np)$ $P(X = k)$

$$= \binom{n}{k} \frac{Np(Np-1)\ldots(Np-k+1)Nq\ldots(Nq-n+k+1)}{N(N-1)\ldots(N-n+1)}$$
$$= \frac{\binom{Np}{k}\binom{Nq}{n-k}}{\binom{N}{n}}$$

The expected value $E(X)$ equals np, and formula (2.9) for the variance takes the form

$$D^2(X) = \frac{N-n}{N-.1}npq \tag{2.20}$$

The hypergeometric distribution is often applied in statistical quality control of mass production. For example, let the lot under control consist of b good items, and $N-b=c$ defective items. Here a good item plays the role of a white ball; a defective item, the role of a black ball; and the lot under control, the role of the urn. From the lot we draw n items at random to determine their quality; usually the chosen items are not returned to the lot. If the numbers b and c are known, by using the formulas obtained previously we can compute the probability that among n chosen items there are k good ones.

Let Us Sum Up

Learners, in this section we have seen that the definition of Polya and Hypergeometric Distributions and also given theorems and examples.

Check Your Progress

1. What is the expected value of a hypergeometric distribution with parameters N, K, and n?

A. $\frac{nK}{N}$

B. $\frac{N-nK}{N}$

C. $\frac{n}{N}$

D. $\frac{K}{N}$

2. What is the variance of a hypergeometric distribution with parameters N, K, and n?

A. $\frac{nK(N-K)(N-n)}{N^2(N-1)}$

B. $\frac{nK(N-K)}{N}$

C. $\frac{K(N-K)(N-n)}{N^2}$

D. $\frac{nK}{N}$

2.5 Poisson Distribution

If b and c áre unknown, and the investigation of the quality of some number of items may serve to estimate these numbers. In example 2.1 we considered a random variable X with a Poisson distribution. Let us summarize the most important properties of such a random variable. Such a random variable can take on the values $r = 0, 1, 2, \ldots$ Its probability function is given by the formula

$$P(X = r) = \frac{\lambda^r}{r!} e^{-\lambda} \tag{2.21}$$

where λ is a positive constant. According to (2.6) its characteristic function has the form

$$\phi(t) = e^{\lambda\left(e^{it} - 1\right)}$$

From (2.7) to (2.9), we obtain

$$m_1 = \lambda, \quad m_2 = \lambda(\lambda + 1), \quad \mu_2 = \lambda$$

The probability function (2.6) can be obtained as the limit of a sequence of probability functions of the binomial distribution. We shall prove Poisson's theorem.

Theorem 2.5.1 *Let the random variable X_n have a binomial distribution defined by the formula*

$$P\left(X_n = r\right) = \frac{n!}{r!(n-r)!} p^r (1-p)^{n-r} \tag{2.22}$$

where r takes on the values $0, 1, 2, \ldots, n$. If for $n = 1, 2, \ldots$ the relation

$$p = \frac{\lambda}{n} \tag{2.23}$$

holds, [1] where $\lambda > 0$ is a constant, then

$$\lim_{n \to \infty} P\left(X_n = r\right) = \frac{\lambda^r}{r!} e^{-\lambda} \tag{2.24}$$

Since the expected value of X_n is np, condition (2.24) means that as n increases the expected value of X_n remains constant. [1] The assertion of this theorem will still hold if relation (2.24) is replaced by

$$\lim_{n \to \infty} np = \lambda$$

Proof: *Let us transform formula (2.24) in the following way:*

$$\begin{aligned}P\left(X_{.n}=r\right) &= \frac{n!}{r!(n-r)!}\left(\frac{\lambda}{n}\right)^{r}\left(1-\frac{\lambda}{n}\right)^{n-r} \\ &= \frac{\lambda^{r}}{r!}\left(1-\frac{\lambda}{n}\right)^{n}\frac{n(n-1)\ldots(n-r+1)}{n^{r}}\cdot\frac{1}{\left(1-\frac{\lambda}{n}\right)^{r}} \\ &= \frac{\lambda^{r}}{r!}\left(1-\frac{\lambda}{n}\right)^{n}\frac{1\cdot\left(1-\frac{1}{n}\right)\ldots\left(1-\frac{r-1}{n}\right)}{\left(1-\frac{\lambda}{n}\right)^{r}}\end{aligned}$$

Using the fact that

$$\lim_{n\to\infty}\left(1-\frac{\lambda}{n}\right)^{n}=e^{-\lambda} \quad \textit{and} \quad \lim_{n\to\infty}\frac{1\cdot\left(1-\frac{1}{n}\right)}{\text{we obtain formula (2.24).}}$$

One of the binomial distribution with $n = 5$ and $p = 0.3$, hence $\lambda = np = 1.5$, and one of the Poisson distribution with the same expected value $\lambda = 1.5$. And two such graphs for $n = 10$ and $p = 0.15$; hence again $\lambda = 1.5$. For larger values of n, for instance, $n = 100$, the graphs of the binomial and Poisson distributions will almost coincide. Often the Poisson distribution is interpreted as a distribution of a random variable which can take on many different values (the number n is large) but with small probabilities (the probability $p = \lambda/n$ is small). That is why the Poisson distribution is sometimes called the law of small numbers. However, as is shown by the next two examples, this name is not justified. Bortkiewicz, who has investigated the Poisson distribution, has given some empirical examples of random events to which this distribution can be applied.

Example 2.5.2 *Computed the number of soldiers in ten cavalry corps who died within a period of twenty years from a kick by a horse. We consider as a random variable the number $r(r = 0, 1, 2, \ldots)$ of men in one corps killed in one year by a kick from a horse. The number of observations was $10 \times 20 = 200$, that is, the observations concerned ten army corps over a period of twenty years.*

Table

The following frequencies of appearance of values of r. The frequencies of death from a kick by a Horse.

r	0	1	2	3	4
Frequency	0.545	0.325	0.110	0.015	0.005
Probability	0.544	0.331	0.101	0.021	0.003

From the central row of this table we compute the mean

$$E(X) = 0 \cdot 0.545 + 1 \cdot 0.325 + 2 \cdot 0.110 + 3 \cdot 0.015 + 4 \cdot 0.005 = 0.61$$

Let us compute the corresponding probabilities $P(X = r)$ for the Poisson distribution with $\lambda = 0.61$. Usually we find these probabilities from Poisson distribution tables, but here we compute them directly. We have

$$P(X = 0) = e^{-0.61} = 0.544$$

$$P(X = 1) = 0.61e^{-0.61} = 0.331$$

$$P(X = 2) = \frac{0.61^2 e^{-0.61}}{2!} = 0.101$$

$$P(X = 3) = \frac{0.61^3 e^{-0.61}}{3!} = 0.021$$

$$P(X = 4) = \frac{0.61^4 e^{-0.61}}{4!} = 0.003$$

These values are presented in the lower row of above table. As we see, these probabilities differ but little from the corresponding frequencies. In many physical and technical problems we deal with distributions close to the Poisson distribution. Here we give an example from physics.

Example 2.5.3 *We present here the results of the famous experiments. They observed the numbers of α particles emitted by a radioactive substance in $n = 2608$ periods of $7.5\mathrm{sec}$ each. These data are presented in table. In this table n_i denotes the number of periods in which the number of emitted particles was equal to i. The average number λ of particles emitted during a period of $7.5\mathrm{sec}$ is*

$$\lambda = \frac{\sum n_i i}{n} = 3.87$$

and

$$p_i = \frac{3.87^i}{i!} e^{-3.87}$$

Table

The reader will notice the striking closeness of the second and third columns in above table

i	n_i	np_i
0	57	54.399
1	203	210.523
2	383	407.361
3	525	525.496
4	532	508.418
5	408	393.515
6	273	253.817
7	139	140.325
8	45	67.882
9	27	29.189
10	16	17.075
	2608	$2,608.000$

Just as for the binomial distribution we can prove the addition theorem for independent random variables with Poisson distributions. Let the independent random variables X_1 and X_2 have the respective Poisson distributions

$$P(X_1 = r) = \frac{\lambda_1^r}{r!} e^{-\lambda_1}, \quad P(X_2 = r) = \frac{\lambda_2^r}{r!} e^{-\lambda_2} \quad (r = 0, 1, 2, \ldots)$$

Consider the random variable

$$X = X_1 + X_2$$

According to (2.6) the characteristic functions of X_1 and X_2 are

$$\phi_1(t) = \exp\left[\lambda_1 \left(e^{it} - 1\right)\right], \quad \phi_2(t) = \exp\left[\lambda_2 \left(e^{it} - 1\right)\right]$$

By the independence of X_1 and X_2 the characteristic function of X has the form

$$\phi(t) = \exp\left[(\lambda_1 + \lambda_2)\left(e^{it} - 1\right)\right] \tag{2.25}$$

Formula (2.25) represents the characteristic function of a random variable with the Poisson distribution having the expected value $\lambda_1 + \lambda_2$. This proves the addition theorem for independent random variables with Poisson distributions. Raikov has proved that the converse theorem is also true: if X_1 and X_2 are independent and $X = X_1 + X_2$ has a Poisson distribution, then each of the random variables X_1 and X_2 has a Poisson distribution.

Let Us Sum Up

Learners, in this section we have seen that the definition of Poisson distribution distribution with Illustrations.

Check Your Progress

1. What is the parameter of the Poisson distribution?
A. Mean
B. Variance
C. Rate or average number of occurrences
D. Standard deviation
2. Which of the following scenarios is best modeled by a Poisson distribution?
A. Number of emails received in an hour
B. Height of individuals in a population
C. The time between arrivals of buses
D. The number of successes in a series of trials

2.6 Uniform Distribution

Raikov's theorem is true for an arbitrary finite number of independent random variables $X_1, \ldots, X_n$. The simplest example of a random variable of the continuous type is a random variable with the uniform distribution. In above example we considered a particular case of the uniform distribution. The general definition is as follows.

Definition 2.6.1 *The random variable X has a uniform, or rectangular distribution if its density function $f(x)$ is given by the formula*

$$f(x) = \begin{cases} \frac{1}{2h} & \text{for } a-h \leqslant x \leqslant a+h, \text{ where } a \text{ and } h>0 \text{ are constants} \\ 0 & \text{otherwise} \end{cases} \tag{2.25 a}$$

The distribution function $F(x)$ of this random variable is given by the formula

$$F(x) = \begin{cases} 0 & \text{for } x < a-h, \\ \frac{1}{2h}\int_{a-h}^{x} dx = \frac{x-(a-h)}{2h} & \text{for } a-h \leqslant x \leqslant a+h, \\ 1 & \text{for } x > a+h. \end{cases}$$

The characteristic function of X is

$$\begin{aligned} \phi(t) &= \frac{1}{2h}\int_{a-h}^{a+h} e^{itx} dx = \frac{1}{2h}\left(\frac{e^{itx}}{it}\right)_{a-h}^{a+h} \\ &= \frac{1}{2h} \cdot \frac{e^{it(a+h)} - e^{it(a-h)}}{it} = e^{ita}\frac{\sin th}{th} \end{aligned} \tag{2.25.b}$$

We obtain the moments directly from the formula

$$m_k = \frac{1}{2h}\int_{a-h}^{a+h} x^k dx = \frac{1}{2h} \cdot \frac{(a+h)^{k+1} - (a-h)^{k+1}}{k+1} \tag{2.26}$$

In particular,

$$m_1 = a, \quad m_2 = \frac{1}{3}\left(3a^2 + h^2\right)$$

Hence

$$\mu_2 = m_2 - m_1^2 = \frac{1}{3}h^2 \tag{2.27}$$

By a linear transformation of X we can obtain a random variable with a uniform distribution in the interval $[0,1]$. To do this we write

$$Y = \frac{X-(a-h)}{2h}$$

The density of Y, which we shall denote by $f_1(y)$, is given by the following formula:

$$f_1(y) = \begin{cases} 1 & \text{in the interval } [0,1] \\ 0 & \text{otherwise} \end{cases} \tag{2.28}$$

This is the rectangular distribution which we considered in above example. The density of this distribution. In statistical problems we often deal with rectangular distributions. It is worthwhile to mention that if the distribution function $F(x)$ of the random variable X is continuous, then the random variable $Y = F(X)$ has the uniform distribution given by (2.6). In fact, to every infinite interval $-\infty < X \leqslant x$ of values of the random variable X there corresponds the set of values of the random variable Y contained in the interval $0 \leqslant Y \leqslant y = F(x)$. On the other hand, by the assumption that the distribution function $F(x)$ is continuous, to every $y(0 \leqslant y \leqslant 1)$ there corresponds at least one x satisfying the relation

$$y = F(x) = P(X < x) \tag{2.29}$$

However, transformation (2.7) may not be one-to-one since the inverse image $F^{-1}(y)$ of some values of y may be an interval in which the function $F(x)$ is constant. Here, for a given y we can take for $x = F^{-1}(y)$ any of the values of x from the interval in which the distribution function $F(x)$ is constant, and for every such value of x we shall have $F\left[F^{-1}(y)\right] = y$; in particular we can take as $x = F^{-4}(y)$ the least x for which this equality holds. If we denote by $F_1(y)$ the distribution function of the random variable Y, we obtain

$$\begin{aligned} F_1(y) &= P(Y < y) = P[F(X) < y] \\ &= \begin{cases} 0 & \text{for } y \leqslant 0 \\ P\left[X < F^{-1}(y)\right] = F\left[F^{-1}(y)\right] = y & \text{for } 0 < y < 1, \\ 1 & \text{for } y \geqslant 1 \end{cases} \end{aligned}$$

2.7 Let Us Sum Up

Learners, in this section we have seen that definition of uniform distribution and its *pdf*.

Check Your Progress

1. What is the mean of a continuous uniform distribution over the interval $[a, b]$?

A. $\frac{a+b}{2}$

B. $\frac{a-b}{2}$

C. $\frac{a+b}{b-a}$

D. $\frac{b-a}{2}$

2. For a discrete uniform distribution where X can take integer values from 1 to 10, what is the probability of any specific value?

A. $\frac{1}{10}$

B. $\frac{1}{5}$

C. $\frac{1}{20}$

D. $\frac{1}{15}$

Glossary

1. The function $F(x)$ is the distribution function of X.
2. The n and p is the parameters of Binomial distribution.
3. The μ is mean of normal distribution.
4. The k is number of observations.
5. The Z is random variable with generalized binomial distribution.

2.8 Unit Summary

The second unit content on one-point and two-point distributions, the Bernoulli, Binomial distribution and Poisson distribution. Also the generalized binomial distribution, Polya, hyper-geometric distributions, Poisson distribution and uniform distribution.

Self-Assessment Questions

Short Answers: (5 Marks)

1. Prove the equality

$$\sum_{l=0}^{k} \binom{n_1}{l} \binom{n_2}{k-l} = \binom{n_1+n_2}{k}$$

2. Prove that

$$\sum_{m=k}^{n} \binom{n}{m} p^m (1-p)^{n-m} = \frac{n!}{(k-1)!(n-k)!} \int_0^p t^{k-1}(1-t)^{n-k} dt.$$

Long Answers: (8 Marks)

1. Show that if the random variables X_1 and X_2 have zero-one distributions and are uncorrelated, they are independent. Also check whether this property holds for all two-point random variables.

Exercises

1. Let the random variable Z have a generalized binomial distribution. Show that if the p_k are functions of n such that $\sum_{k=1}^{n} p_k = \lambda$ is fixed, and $\alpha_n = \max(p_1, \ldots, p_n)$ tends to zero as $n \to \infty$, then prove that

$$\lim_{n\to\infty} P(Z = r) = e^{-\lambda} \frac{\lambda^r}{r!} \quad (r = 0, 1, 2, \ldots)$$

2. $F(x)$ is the distribution function of a random variable X with the zeroone distribution. Find the distribution function of the random variable $Y = F(X)$.
   (a) Do the same for a random variable X with the binomial distribution.
   (b) Do the same when X has the Poisson distribution.

Answers to check your progress

Section 2.1

1. B. The probability of success is constant across trials.

2. C. μ

Section 2.2

1. B. Number of trials and probability of success

2. A. $P(X = k) = \frac{n!}{k!(n-k)!}p^k(1-p)^{n-k}$

Section 2.3

1. A. $10 \times 0.3 = 3$

2. A. $np(1-p)$

Section 2.4

1. A. $\frac{nK}{N}$

2. A. $\frac{nK(N-K)(N-n)}{N^2(N-1)}$ Section 2.5

1. C. Rate or average number of occurrences

2. A. Number of emails received in an hour

Section 2.6

1. A. $\frac{a+b}{2}$

2. A. $\frac{1}{10}$

References

1. M. Fisz, Probability Theory and Mathematical Statistics, John Wiley and sons, New Your, Third Edition, 1963.

Suggested Readings

1. T. Veerarajan, Fundamentals of Mathematical Statistics, Yesdee Publishing, 2017.

2. P.R. Vittal, Mathematical Statistics, Margham Publications, 2002.

3. R.S.N. Pillai and V. Bagavathi, Statistics, Sultan Chand and Sons, 2010.

4. S. C. Gupta and V. K. Kapoor, Fundamentals of Mathematical Statistics, Sultan Chand and Sons, 2008.

Unit 3

Probability Continuous Distributions

Objective

This course aims to teach the students about some continuous probability distributions are normal distribution, gamma distribution, beta distribution, Cauchy and laplace distributions, multinomial distribution and compound distributions.

3.1 Normal Distribution

In this section we discuss the probability continuous distributions. We know that the formulas;

$$F_1'(y) = f_1(y) = \begin{cases} 1 & \text{for } 0 \leqslant y \leqslant 1 \\ 0 & \text{for the remaining } y \end{cases}$$

In the examples we have often considered random variables with normal distributions. We now investigate the general form of the normal distribution.

Definition 3.1.1 *The random variable X has a normal distribution if its density function is given by the formula*

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(x-m)^2}{2\sigma^2}\right) \tag{3.1}$$

where $\sigma > 0$. We first verify that (3.1) is a density. To see this let us denote

$$Y = \frac{X-m}{\sigma} \tag{3.2}$$

We obtain

$$f(y) = \frac{1}{\sqrt{2\pi}} e^{-y^2/2} \tag{3.3}$$

Since the function $f(y)$ given by (3.3) is a density, we have the equation

$$\frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{+\infty} \exp\left(-\frac{(x-m)^2}{2\sigma^2}\right) dx = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} e^{-y^2/2} dy = 1$$

The characteristic function $\phi(t)$ of the random variable Y has already been obtained in example we have $\phi(t) = e^{-t^2/2}$. Using equations (2.14), (2.15), and (3.2), we obtain the expression

$$\phi_1(t) = \exp\left(itm - \frac{1}{2}\sigma^2 t^2\right) \tag{3.4}$$

for the characteristic function of X. From (3.4) and (32.4) we obtain the moments

$$m_1 = m, \quad m_2 = \sigma^2 + m^2, \quad \mu_2 = \sigma^2 \tag{3.5}$$

As we can see from equalities (3.5), the constants m and σ which appear in (3.1) may be easily interpreted; m is the expected value of X and σ is its standard deviation. The shape of the curve of the density of the normal distribution depends on the parameter σ. The normal curve is representing three normal distributions with the same expected value $m = 0$ and different standard deviations: $\sigma = 1, \sigma = 0.5$ and $\sigma = 0.25$. The normal distribution with expected value m and standard deviation σ is often denoted by $N(m; \sigma)$. By the symmetry of the normal curve with respect to the expected value m all the central moments of odd order vanish,

$$\mu_{2k+1} = 0 \quad \text{for every } k \tag{3.6}$$

It can be easily shown that

$$\mu_{2k} = 1 \cdot 3 \cdot \ldots \cdot (2k-1)\sigma^{2k} \tag{3.7}$$

Formula (2.13) is a particular case of formula (3.7) for $\sigma = 1$. There are very exact tables of the normal distribution which are used in computation. Usually we are interested in the probability that the random variable X with a normal distribution differs in absolute value from the expected value $m = E(X)$ by more than $\lambda\sigma(\lambda > 0)$, that is, more than a given multiple of the standard deviation. We find this probability, expressed as a function of λ, in the tables of the normal distribution giving the value of the integral

$$P(|X-m|>\lambda\sigma)=\frac{2}{\sqrt{2\pi}}\int_{\lambda}^{\infty}e^{-y^2/2}dy$$

In fact,

$$P(|X-m|>\lambda\sigma)=P\left(\frac{|X-m|}{\sigma}>\lambda\right)=P(|Y|>\lambda)$$

where $Y=(X-m)/\sigma$. We may also ask for the probability that X exceeds the expected value by more than a given multiple of the standard deviation $\lambda\sigma$, that is, the probability $P(X>m+\lambda\sigma)$. We have

$$P(X>m+\lambda\sigma)=P(Y>\lambda)=\frac{1}{\sqrt{2\pi}}\int_{\lambda}^{+\infty}e^{-y^2/2}dy$$

Example 3.1.2 *The random variable X has the distribution $N(1;2)$. Find the probability that X is greater than 3 in absolute value. Let us introduce the standardized random variable $Y=(X-1)/2$. We have*

$$\begin{aligned}P(|X|>3)&=P(|2Y+1|>3)=P\left(\left|Y+\frac{1}{2}\right|>\frac{3}{2}\right)\\&=P\left(Y+\frac{1}{2}<-\frac{3}{2}\right)+P\left(Y+\frac{1}{2}>\frac{3}{2}\right)=P(Y<-2)+P(Y>1)\end{aligned}$$

By definition of Y, we have

$$\begin{aligned}P(Y.<-2)&=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{-2}e^{-t^2/2}dt=\frac{1}{\sqrt{2\pi}}\int_{2}^{+\infty}e^{-t^2/2}dt\cong 0.023\\P(Y>1)&=\frac{1}{\sqrt{2\pi}}\int_{1}^{+\infty}e^{-t^2/2}dt\cong 0.159\end{aligned}$$

The values of these integrals are obtained from tables of the normal distribution. Finally, we have

$$P(|X|>3)=0.182$$

From tables of the normal distribution we see that, for a random variable X with the *normal distribution $N(m;\sigma)$, the following equalities are satisfied:*

$$P(|X-m|>\sigma)\cong 0.3173$$

$$P(|X-m|>2\sigma)\cong 0.0455$$

$$P(|X-m|>3\sigma)\cong 0.0027$$

We see thus that the normal distribution is highly concentrated around its expected value. The probability that the value of X differs from the expected value by more than 3σ is smaller than 0.01 . This property of the normal distribution has led mány statisticians to apply the three-sigma rule, according to which for an arbitrary distribution there is small probability that the random variable differs from the expected value by more than 3σ. This rule should be applied very carefully. In fact, from the Chebyshev inequality follows only the fact that for an arbitrary random variable X whose variance exists

$$P(|X-m|\geqslant 3\sigma)\leqslant \frac{1}{9}$$

The three-sigma rule can be applied only to distributions which do not differ much from the normal distribution. Thus they must be almost symmetric distributions, having only one maximum point in the neighborhood of the center of symmetry. The addition theorem also holds for the normal distribution. Let X and Y be two independent random variables, and let X have the distribution $N\left(m_1;\sigma_1\right)$ and Y the distribution $N\left(m_2;\sigma_2\right)$. The respective characteristic functions of these distributions are

$$\phi_1(t)=\exp\left(m_1 it-\frac{1}{2}t^2\sigma_1^2\right)$$
$$\phi_2(t)=\exp\left(m_2 it-\frac{1}{2}t^2\sigma_2^2\right)$$

Because of the independence of X and Y, the random variable $Z=X+Y$ has the characteristic function

$$\phi(t)=\exp\left[\left(m_1+m_2\right)it-\frac{1}{2}\left(\sigma_1^2+\sigma_2^2\right)t^2\right] \tag{3.8}$$

Expression (3.8) is the characteristic function of the normal distribution $N\left(m_1+m_2;\sqrt{{\sigma_1}^2+{\sigma_2}^2}\right)$, which was to be proved. Cramér proved that the converse theorem is also true: if X_1 and

X_2 are independent and the random variable $X = X_1 + X_2$ has a normal distribution, then each of the random variables X_1 and X_2 has a normal distribution. Cramér's theorem is true for an arbitrary finite number of independent random variables. Besides Cramer's theorem, many others are known which characterize the normal distribution. We present here the theorem of Skitowitch. Let $X_1, \ldots, X_n$ be independent and have the same nondegenerate distribution. Then the independence of the random variables L_1 and L_2, defined by

$$L_1 = a_1 X_1 + \ldots + a_n X_n$$

$$L_2 = b_1 X_1 + \ldots + b_n X_n$$

with $\sum_{j=1}^{n} a_j b_j = 0$ and $\sum_{j=1}^{n} (a_j b_j)^2 \neq 0$, is a necessary and sufficient condition for the distributions of the random variables $X_1, \ldots, X_n$ to be normal. For $n = 2$ this theorem has been proved by Bernstein, Darmois, and Gnedenko.

Let Us Sum Up

Learners, in this section we have seen that definition of normal distribution and also given theorems and Illustrations.

Check Your Progress

1. What is the shape of the probability density function pdf of a normal distribution?
A. Symmetric bell curve
B. Skewed to the right
C. Skewed to the left
D. Uniform
2. What is the total area under the probability density function pdf of a normal distribution?
A. 1
B. $\sqrt{2\pi}$
C. σ
D. σ^2

3.2 Gamma Distribution

The normal distribution is of great importance in probability theory and statistics. In nature and technology we very often deal with distributions that are close to normal. This phenomenon is an object of investigation of the theory of stochastic processes. Moreover, under rather general assumptions the normal distribution is the limiting distribution for sums of independent random variables when the number of terms increases to infinity. This question is discussed in the next chapter. In applications we often use a distribution associated with the gamma function, defined for $p > 0$ by the formula

$$\Gamma(p) = \int_0^{\infty} x^{p-1} e^{-x} dx \tag{3.1}$$

It is known that integral (3.1) is uniformly convergent with respect to p and thus $\Gamma(p)$ is a continuous function. Integrating (3.1) by parts, we obtain

$$\Gamma(p+1) = \int_0^{\infty} x^p e^{-x} dx = \left[-e^{-x} x^p\right]_0^{\infty} + p \int_0^{\infty} x^{p-1} e^{-x} dx$$

Hence (3.2)

$$\Gamma(p+1) = p\Gamma(p)$$

In particular, if $p = n$, where n is an integer, we obtain from (3.2)

$$\Gamma(n+1) = n\Gamma(n)$$

$$\Gamma(n) = (n-1)\Gamma(n-1) \tag{3.3}$$

$$\Gamma(2) = 1\Gamma(1).$$

Since

$$\Gamma(1) = \int_0^{\infty} e^{-x} dx = -\left[e^{-x}\right]_0^{\infty} = 1$$

we obtain from equalities (3.3)

$$\Gamma(n+1) = n! \tag{3.4}$$

Substituting $y = x/a (a > 0)$ in (3.1), we have

$$\frac{\Gamma(p)}{a^p} = \int_0^{\infty} y^{p-1} e^{-ay} dy \tag{3.5}$$

Equation (3.5) is also valid when a is a complex number $a = b + ic$, where $b > 0$. We shall not give the proof of (3.5) for this case. Let X be a random variable with the density defined by the formula

$$f(x) = \begin{cases} 0 & \text{for } x \leqslant 0 \\ \frac{b^p}{\Gamma(p)} x^{p-1} e^{-bx} & \text{for } x > 0 \end{cases} \tag{3.6}$$

where $b > 0$ and $p > 0$. The fact that (3.6) defines a density follows directly from (3.5), since

$$\int_{-\infty}^{+\infty} f(x)dx = \int_0^{+\infty} \frac{b^p}{\Gamma(p)} x^{p-1} e^{-bx} dx = 1$$

and $f(x)$ is a non-negative function.

Definition 3.2.1 *If a random variable X has the density given by (3.6) we shall say that X has a gamma distribution is represents such a density for $p = 1$ and $b = 0.5$. We now find the characteristic function of this distribution. We have*

$$\phi(t) = \int_{-\infty}^{+\infty} e^{itx} f(x)dx = \frac{b^p}{\Gamma(p)} \int_0^{+\infty} x^{p-1} e^{-(b-it)x} dx \tag{3.7}$$

Since, as has already been stated, equation (3.5) is valid when $a = b + ic$ a nd $b > 0$, then

$$\phi(t) = \frac{b^p}{\Gamma(p)} \cdot \frac{\Gamma(p)}{(b - it)^p} = \frac{1}{(1 - it/b)^p} \tag{3.8}$$

The function $\phi(t)$ can be differentiated an arbitrary number of times. Its k th derivative is expressed by the formula

$$\phi^{(k)}(t) = \frac{p(p+1)\dots(p+k-1)}{b^k} i^k \frac{1}{(1 - it/b)^{p+k}} \qquad \text{for } k = 1, 2, \dots$$

From (2.4) we obtain

$$m_k = \frac{\phi^{(k)}(0)}{i^k} = \frac{p(p+1)\dots(p+k-1)}{b^k} \tag{3.9}$$

In particular, we have,

$$m_1 = \frac{p}{b}, \quad m_2 = \frac{p(p+1)}{b^2}, \quad \mu_2 = \frac{p}{b^2} \tag{3.10}$$

Example 3.2.2 *The random variable X has the gamma distribution with the density* *given by the formula*

$$f(x) = \begin{cases} 0 & \textit{for} \quad x \leqslant 0 \\ 2e^{-2x} & \textit{for} \quad x > 0 \end{cases}$$

The reader may verify that if we substitute $p = 1$ and $b = 2$ in (3.6) we obtain the distribution considered in this example. What is the probability that X is not smaller than two?, We have

$$P(X \geqslant 2) = 2\int_{2}^{\infty} e^{-2x}dx = -\left[e^{-2x}\right]_2^{\infty} = e^{-4} \cong 0.0183$$

In more complicated cases we can make use of the tables by K. Pearson to compute probabilities of the gamma distribution. The probability distribution considered in example is a particular case of the exponential distribution.

Definition 3.2.3 *The random variable with density $f(x)$, defined by the formula*

$$f(x) = \begin{cases} 0 & \text{for } x \leqslant 0 \\ \lambda e^{-\lambda x} & \text{for } x > 0 \end{cases} \tag{3.11}$$

where $\lambda > 0$, has an exponential distribution. We now show that the addition theorem is valid for random variables with gamma distributions. Let X_1 and X_2 be two independent random variables with gamma distributions and with the respective characteristic functions

$$\phi_k(t) = \frac{1}{(1 - it/b)^{p_k}} \quad (k = 1, 2)$$

Let Us Sum Up

Learners, in this section we have seen that definition of gamma distribution and also given theorems and Illustrations.

Check Your Progress

1. What are the two parameters of the gamma distribution?

A. Hape parameter (α) and scale parameter (β)

B. Mean (μ) and standard deviation (σ)

C. Rate parameter (λ) and variance (σ^2)
D. Mean (μ) and variance (σ^2)
2. Which of the following distributions is a special case of the gamma distribution with integer shape parameter α?
A. Normal distribution
B. Poisson distribution
C. Uniform distribution
D. Chi-squared distribution

3.3 Beta Distribution

Consider the sum of these random variables, $X = X_1 + X_2$. From the independence of X_1 and X_2 it follows that the characteristic function $\phi(t)$ of X equals

$$\phi(t) = \frac{1}{(1 - it/b)^{p_1}} \cdot \frac{1}{(1 - it/b)^{p_2}} = \frac{1}{(1 - it/b)^{p_1+p_2}}$$

We see that X also has the gamma distribution, which proves the theorem. Laha and Lukacs have given theorems characterizing the gamma distribution. We mention here the following quite simple theorem of Lukacs. Let the independent random variables X and Y with nondegenerate distributions take on only positive values. Then X and Y have the gamma distribution with the same parameter b if and only if the random variables U and V, where

$$U = X + Y, \quad V = \frac{X}{Y}$$

are independent. In the applications we also deal with a distribution associated with the function defined by the formula

$$B(p, q) = \int_0^1 x^{p-1}(1 - x)^{q-1} dx, \quad \text{where } p > 0, q > 0 \tag{3.12}$$

In the monograph of Saks and Zygmund the reader will find a proof of the following equation connecting the function $B(p, q)$ with the function Γ defined by (3.13):

$$B(p, q) = \frac{\Gamma(p)\Gamma(q)}{\Gamma(p + q)} \tag{3.13}$$

Definition 3.3.1 *We say that the random variable X has a beta distribution if its density* *is given by the formula*

$$f(x) = \begin{cases} \frac{1}{B(p,q)} x^{p-1}(1-x)^{q-1} & \text{for } 0 < x < 1 \\ 0 & \text{for } x \leqslant 0 \quad \text{and} \quad x \geqslant 1 \end{cases} \tag{3.14}$$

where $p > 0, q > 0$. That the function $f(x)$ given by (3.14) is a density follows from formula (3.11) and the fact that it is non-negative. It is convenient to obtain the moments of the beta distribution directly from the formula

$$\begin{aligned} m_k &= \frac{\Gamma(p+q)}{\Gamma(p)\Gamma(q)} \int_0^1 x^{p+k-1}(1-x)^{q-1} dx = \frac{\Gamma(p+q)}{\Gamma(p)\Gamma(q)} \cdot B(p+k, q) \\ &= \frac{\Gamma(p+q)\Gamma(p+k)}{\Gamma(p)\Gamma(p+q+k)} = \frac{p(p+1)\dots(p+k-1)}{(p+q)(p+q+1)\dots(p+q+k-1)} \end{aligned} \tag{3.15}$$

In particular,

$$m_1 = \frac{p}{p+q}, \quad m_2 = \frac{p(p+1)}{(p+q)(p+q+1)} \tag{3.15}$$

$$\mu_2 = \frac{pq}{(p+q)^2(p+q+1)} \tag{3.16}$$

The density of the beta distribution with $p = q = 2$.

Example 3.3.2 *The random variable X has the beta distribution with $p = q = 2$; hence its density $f(y)$ has the form*

$$f(y) = \begin{cases} 0 & \textit{for} \quad y \leqslant 0 \textit{ and } y \geqslant 1 \\ \frac{\Gamma(4)}{\Gamma'(2)\Gamma'(2)} y(1-y) = 6y(1-y) & \textit{for} \quad 0 < y < 1 \end{cases}$$

What is the probability that X is not greater than 0.2?, We have

$$P(Y \leqslant 0.2) = 6 \int_0^{0.2} y(1-y) dy = 6 \left[\frac{y^2}{2} - \frac{y^3}{3} \right]_0^{0.2} = 0.104$$

For computing the probabilities of the beta distribution we can use Pearson's tables [4].

Definition 3.3.3 *The random variable X has a Cauchy distribution if its density is given by the formula is;*

$$f(x) = \frac{1}{\pi} \cdot \frac{\lambda}{\lambda^2 + (x-\mu)^2}, \quad \text{where} \lambda > 0 \tag{3.17}$$

The function $f(x)$ is non-negative. By substituting

$$y = \frac{x-\mu}{\lambda}, \tag{3.18}$$

we obtain

$$\int_{-\infty}^{+\infty} f(x)dx = \frac{1}{\pi}\int_{-\infty}^{+\infty} \frac{dy}{1+y^2} = \frac{1}{\pi}[\arctan y]_{-\infty}^{+\infty} = 1$$

To find the characteristic function of the random variable X let us first find the characteristic function of the random variable Y which is the linear transformation of X given by (3.2). Thus Y has the density

$$f(y) = \frac{1}{\pi}\cdot\frac{1}{1+y^2} \tag{3.19}$$

and the characteristic function

$$\phi(t) = \frac{1}{\pi}\int_{-\infty}^{+\infty} e^{ity}\frac{1}{1+y^2}dy \tag{3.20}$$

To find $\phi(t)$ consider first the density

$$f_1(y) = \frac{1}{2}e^{-|y|} \tag{3.21}$$

The reader may verify that expression (3.10) is a density. The characteristic function of the random variable with the density (3.10) is

$$\begin{aligned}\phi_1(t) &= \frac{1}{2}\int_{-\infty}^{+\infty} e^{ity}e^{-|y|}dy = \frac{1}{2}\int_{-\infty}^{+\infty}(\cos ty + i\sin ty)e^{-|y|}dy \\ &= \int_0^{\infty}\cos tye^{-y}dy\end{aligned}$$

Integrating by parts, we obtain

$$\begin{aligned}\int_0^{\infty}\cos tye^{-y}dy &= \left[-e^{-y}\cos ty\right]_0^{\infty} - t\int_0^{\infty}\sin tye^{-y}dy \\ &= 1 - t\int_0^{\infty}\sin tye^{-y}dy\end{aligned}$$

Similarly,

$$\int_0^{\infty} \sin tye^{-y}dy = \left[-e^{-y}\sin ty\right]_0^{\infty} + t\int_0^{\infty} e^{-y}\cos tydy$$
$$= t\int_0^{\infty} e^{-y}\cos tydy$$

Hence we obtain

$$\int_0^{\infty} e^{-y}\cos tydy = 1 - t^2\int_0^{\infty} e^{-y}\cos tydy$$

Finally we have

$$\phi_1(t) = \int_0^{\infty} e^{-y}\cos tydy = \frac{1}{1+t^2} \tag{3.22}$$

The characteristic function (3.17) is absolutely integrable over the interval $(-\infty, +\infty)$; by (3.6) its corresponding density is

$$f_1(y) = \frac{1}{2\pi}\int_{-\infty}^{+\infty} \frac{e^{-ity}}{1+t^2}dt \tag{3.23}$$

From (3.5) we obtain

$$e^{-|y|} = \frac{1}{\pi}\int_{-\infty}^{+\infty} \frac{e^{-ity}}{1+t^2}dt$$

Changing e^{-ity} into e^{ity} under the integral sign (this does not affect the value of the integral) and changing the roles of t and y, we obtain

$$e^{-|t|} = \frac{1}{\pi}\int_{-\infty}^{+\infty} \frac{e^{ity}}{1+y^2}dy \tag{3.24}$$

The right-hand side of (3.8) is identical with that of (3.4); thus we finally obtain

$$\phi(t) = e^{-|t|} \tag{3.25}$$

Since X is a linear transformation of Y, for the characteristic function $\phi_2(t)$ of X we obtain the formula

$$\phi_2(t) = \exp(i\mu t - \lambda|t|) \tag{3.26}$$

Since, as can easily be seen, the function $\phi_2(t)$ is not differentiable at $t = 0$, none of the moments of the Cauchy distribution exist. The addition theorem is valid for the Cauchy distribution. In fact, let X_1 and X_2 be two independent random variables with densities $g_1(x) = \frac{1}{\pi}\cdot\frac{\lambda_1}{\lambda_1^2+(x-\mu_1)^2}, \quad g_2(x) = \frac{1}{\pi}\cdot\frac{\lambda_2}{\lambda_2^2+(x-\mu_2)^2} \quad (\lambda_1, \lambda_2 > 0)$. The function $\phi(t)$ can be differentiated an arbitrary number of times. Its k th derivative is expressed by the formula is;

$$\phi^{(k)}(t) = \frac{p(p+1)\ldots(p+k-1)}{b^k} i^k \frac{1}{(1-it/b)^{p+k}} \qquad \text{for } k = 1, 2, \ldots$$

From (2.4) we obtain

$$m_k = \frac{\phi^{(k)}(0)}{i^k} = \frac{p(p+1)\ldots(p+k-1)}{b^k} \tag{3.27}$$

In particular, we have,

$$m_1 = \frac{p}{b}, \quad m_2 = \frac{p(p+1)}{b^2}, \quad \mu_2 = \frac{p}{b^2} \tag{3.28}$$

Example 3.3.4 *The random variable X has the gamma distribution with the density given by the formula*

$$f(x) = \begin{cases} 0 & \textit{for} \quad x \leqslant 0 \\ 2e^{-2x} & \textit{for} \quad x > 0 \end{cases}$$

The reader may verify that if we substitute $p = 1$ and $b = 2$ in (3.6) we obtain the distribution considered in this example. What is the probability that X is not smaller than two?, We have

$$P(X \geqslant 2) = 2\int_2^{\infty} e^{-2x} dx = -\left[e^{-2x}\right]_2^{\infty} = e^{-4} \cong 0.0183$$

In more complicated cases we can make use of the tables by K. Pearson to compute probabilities of the gamma distribution. The probability distribution considered in example is a particular case of the exponential distribution.

Definition 3.3.5 *The random variable with density $f(x)$, defined by the formula*

$$f(x) = \begin{cases} 0 & \textit{for } x \leqslant 0 \\ \lambda e^{-\lambda x} & \textit{for } x > 0 \end{cases} \tag{3.29}$$

where $\lambda > 0$, has an exponential distribution. We now show that the addition theorem is valid for random variables with gamma distributions. Let X_1 and X_2 be two independent *random variables with gamma distributions and with the respective characteristic functions*

$$\phi_k(t) = \frac{1}{(1 - it/b)^{p_k}} \quad (k = 1, 2)$$

Let Us Sum Up

Learners, in this section we have seen that the definition of Beta distributionand also given theorems and Illustrations.

Check Your Progress

1. What are the two shape parameters of the beta distribution?

A. α and β

B. μ and σ

C. λ and θ

D. α and λ

2. What is the mean of a beta distribution with shape parameters α and β?

A. $\frac{\beta}{\alpha+\beta}$

B. $\frac{\alpha+\beta}{\alpha}$

C. $\frac{\alpha}{\alpha+\beta}$

D. $\frac{\alpha}{\beta}$

3.4 Multinomial Distribution

Let us consider the following generalized Bernoulli scheme. We perform n random experiments. As a result of each experiment one of the pairwise exclusive events $A_j(j = 1, 2, \ldots, r, r + 1)$ occurs. Let $p_j = P(A_j)$, where $p_1 + \ldots + p_r + p_{r+1} = 1$. The results of the n experiments are independent. Consider the random variable ($X_1, \ldots, X_r, X_{r+1}$), where $X_j = k_j$ means that event A_j has occurred k_j times, $k_j = 0, 1, \ldots, n$.

$$P(X_1 = k_1, \ldots, X_r = k_r,\ X_{r+1} = k_{r+1}) \\ = \frac{n!}{k_1! \ldots k_r! k_{r+1}!} p_1^{k_1} \ldots p_r^{k_r} p_{r+1}^{k_{r+1}} \tag{3.30}$$

where $k_1 + \ldots + k_r + k_{r+1} = n$. This formula gives us the probability that A_1 occurs k_1 times, A_2 occurs k_2 times, $\ldots, A_{r+1}$ occurs k_{r+1} times. We notice that the random variables $X_1, \ldots, X_r, X_{r+1}$ satisfy the linear relation $X_1 + \ldots + X_r + X_{r+1} = n$. Let us express one of the random variables, say X_{r+1}, in terms of the remaining ones, that is, $X_{r+1} = n - X_1 \ldots - X_r$. Then formula (3.1) can be written in the form

$$P\left(X_1 = k_1, \ldots, X_r = k_r\right) = \frac{n!}{k_1! \ldots k_r!(n-K)!} p_1^{k_1} \ldots p_r^{k_r} q^{n-K}$$

where $K = k_1 + \ldots + k_r$ and $q = 1 - p_1 - \ldots - p_r$.

Definition 3.4.1 *The random variable $(X_1, \ldots, X_r)$ with the probability function given by formula (3.1) is said to have a multinomial distribution.*

Let $\left(Y_1^{(1)}, Y_2^{(1)}, \ldots, Y_r^{(1)}\right)$ and $\left(Y_1^{(2)}, Y_2^{(2)}, \ldots, Y_r^{(2)}\right)$ be two random variables. Addition of two multidimensional random variables will be understood in the vector sense, that is, we say that the random variable $(X_1, X_2, \ldots, X_r)$ is the sum of the random variables $\left(Y_1^{(1)}, Y_2^{(1)}, \ldots, Y_r^{(1)}\right)$ and $\left(Y_1^{(2)}, Y_2^{(2)}, \ldots, Y_r^{(2)}\right)$, and we write

$$(X_1, X_2, \ldots, X_r) = \left(Y_1^{(1)}, Y_2^{(1)}, \ldots, Y_r^{(1)}\right) + \left(Y_1^{(2)}, Y_2^{(2)}, \ldots, Y_r^{(2)}\right)$$

if $X_j = Y_j^{(1)} + Y_j^{(2)} (j = 1, 2, \ldots, r)$. Now let $\left(Y_1^{(m)}, Y_2^{(m)}, \ldots, Y_r^{(m)}\right)$, where $m = 1, 2, \ldots, n$, be independent random vectors with the same distribution, having at most one coordinate different from zero, where for $m = 1, \ldots, n$ and $j = 1, 2, \ldots, r$

$$P\left(Y_j^{(m)} = 1\right) = p_j, \quad P\left(Y_j^{(m)} = 0\right) = 1 - p_j \tag{3.31}$$

$$P\left(Y_1^{(m)} = 0, \ldots, Y_r^{(m)} = 0\right) = q = 1 - p_1 - \ldots - p_r$$

Let Us Sum Up

Learners, in this section we have seen that definition of Multinomial Distribution and also given theorems and examples.

Check Your Progress

1. What is the multinomial distribution a generalization?

A. Binomial distribution

B. Poisson distribution

C. Uniform distribution

D. Normal distribution

2. What is the mean of X_i in a multinomial distribution with parameters n and probabilities $p_1, p_2, \ldots, p_k$?

A. $n(1-p_i)$

B. np_i

C. p_i

D. n

3.5 Compound Distributions

It is easy to verify that the random variable $(X_1, X_2, \ldots, X_r)$ with a multinomial distribution satisfies the relation

$$(X_1, X_2, \ldots, X_r) = \sum_{m=1}^{n} \left(Y_1^{(m)}, Y_2^{(m)}, \ldots, Y_r^{(m)}\right)$$

By (2.1) we find that the characteristic function $\phi_m(t_1, t_2, \ldots, t_r)$ of $\left(Y_1^{(m)}, Y_2^{(m)}, \ldots, Y_r^{(m)}\right)$, for $m = 1, 2, \ldots, n$, is of the form

$$\phi_m(t_1, t_2, \ldots, t_r) = \sum_{j=1}^{r} p_j e^{it_j} + q$$

Hence, the characteristic function $\phi(t_1, t_2, \ldots, t_r)$ of $(X_1, X_2, \ldots, X_r)$ we obtain the formula

$$\phi(t_1, t_2, \ldots, t_r) = \prod_{m=1}^{n} \phi_m(t_1, t_2, \ldots, t_r) = \left(\sum_{j=1}^{r} p_j e^{itj} + q\right)^n \tag{3.32}$$

From the last formula and from formulas analogous to (2.8) we obtain for $j = 1, 2, \ldots, r$

$$E(X_j) = np_j, \quad \lambda_{jj} = D^2(X_j) = np_j(1-p_j) \tag{3.33}$$

and for $j, k = 1, 2, \ldots, r$ and $j \neq k$

$$\lambda_{jk} = E\left[\left(X_j - np_j\right)\left(X_k - np_k\right)\right] = -np_j p_k \tag{3.34}$$

In applications we often deal with a random variable X whose distribution depends on a parameter α which is a random variable with a specified distribution. Then, we say that the random variable X has a compound distribution. We shall investigate more closely two compound distributions, namely, a compound binomial distribution and a compound Poisson distribution. Let the random variables $X_k (k = 1, 2, \ldots)$ be independent and have the zero-one distribution defined by the probabilities $P\left(X_k = 1\right) = p$ and $P\left(X_k = 0\right) = 1 - p$. Consider the random variable $X = X_1 + X_2 + \ldots + X_N$. For a fixed N, X has the binomial distribution

$$P(X = s) = \binom{N}{s} p^s (1 - p)^{N-s} \quad (s = 0, 1, \ldots, N) \tag{3.35}$$

Let N be a random variable independent of $X_k (k = 1, 2, \ldots)$ with the Poisson distribution

$$P(N = n) = \frac{\lambda^n}{n!} e^{-\lambda} \quad (n = 0, 1, 2, \ldots) \tag{3.36}$$

As we see, here N plays the role of the parameter α mentioned previously. Consider the two-dimensional random variable (X, N). We have

$$P(X = s, N = n) = P(X = s \mid N = n) P(N = n)$$

We are interested in the probability of the event $X = s$ for every s; in other words, we want to find the marginal distribution of X. This distribution is given by the formula

$$P(X = s) = \sum_{n=0}^{\infty} P(X = s \mid N = n) P(N = n) \quad (s = 0, 1, 2, \ldots)$$

Considering (3.1), (3.2), and the fact that $\binom{n}{s} = 0$ for $n < s$, we obtain

$$P(X = s) = \sum_{n=0}^{\infty} e^{-\lambda} \frac{\lambda^n}{n!} \binom{n}{s} p^s (1-p)^{n-s} \tag{3.37}$$
$$= \frac{e^{-\lambda} p^s}{s!} \sum_{n=s}^{\infty} \frac{\lambda^n (1-p)^{n-s}}{(n-s)!} = \frac{e^{-\lambda} p^s \lambda^s}{s!} \sum_{n=s}^{\infty} \frac{[\lambda(1-p)]^{n-s}}{(n-s)!}$$
$$= \frac{e^{-\lambda} p^s \lambda^s}{s!} e^{\lambda(1-p)} = \frac{e^{-\lambda p} (\lambda p)^s}{s!}$$

We have obtained the Poisson distribution with expected value equal to $p\lambda$. This distribution is called a compound binomial distribution.

Example 3.5.1 *The probability that a newborn baby will be a boy is $p = 0.517$. The number X of boys in a family with N children (N constant) is a random variable with the binomial distribution*

$$P(X = s) = \binom{N}{s} 0.517^s \cdot 0.483^{v-s} \quad (s = 0, 1, \ldots, N)$$

We might want the probability that $X = s$ for all possible values of N, that is, the probability that there will be s boys in a family with an arbitrary number of children. Here N is a random variable with a certain distribution which can be determined empirically for a given population by establishing the fraction of families with no children, with one child, and so on. If we know the distribution of N, we can calculate the probability $P(X = s)$ for every s in a manner similar to the derivation of (3.3). We now investigate a compound Poisson distribution. Let the random variable X have the Poisson distribution given by the formula

$$P(X = k) = \frac{\lambda^k}{k!} e^{-\lambda} \quad (k = 0, 1, 2, \ldots) \tag{3.38}$$

and let λ be a random variable of the continuous type with the density

$$f(\lambda) = \begin{cases} \frac{a^v}{\Gamma(v)} \lambda^{v-1} e^{-a\lambda} & \textit{for } \lambda > 0 \\ 0 & \textit{for } \lambda \leqslant 0 \end{cases} \tag{3.39}$$

where $v > 0, a > 0$. Now consider the two-dimensional random variable (X, λ). Here one of the random variables is discrete and the other is continuous. For every $h > 0$ and

$\lambda_1 > 0$ *we have*

$$P(X = k, \lambda_1 \leqslant \lambda \leqslant \lambda_1 + h) =$$
$$P(X = k \mid \lambda_1 \leqslant \lambda \leqslant \lambda_1 + h)\, P(\lambda_1 \leqslant \lambda \leqslant \lambda_1 + h)$$

Leı us divide both sides of this equality by h *and pass to the limit as* $h \to 0$*. From (3.8) and (3.5) we obtain* $\lim_{h\to 0} \frac{1}{h} P(X = k, \lambda_1 \leqslant \lambda \leqslant \lambda_1 + h) = \frac{\lambda_1^k}{k!} e^{-\lambda_1} \frac{a^v}{\Gamma(v)} \lambda_1^{v-1} e^{-a\lambda_1}$.

3.5.1 Two-Dimensional Distribution

Determines the two-dimensional distribution of (X, λ). Writing on the right-hand side of (3.7) λ in place of λ_1, we obtain the marginal distribution of X from the formula

$$P(X = k) = \int_0^\infty \frac{\lambda^k}{k!} e^{-\lambda} \frac{a^v}{\Gamma(v)} \lambda^{v-1} e^{-a\lambda} d\lambda$$

From (3.5) we obtain

$$\begin{aligned} P(X = k) &= \frac{a^v}{\Gamma(v)} \int_0^\infty \frac{\lambda^{k+v-1} e^{-(a+1)\lambda}}{k!} d\lambda = \frac{a^v}{\Gamma(v)} \cdot \frac{1}{k!} \cdot \frac{\Gamma(k+v)}{(a+1)^{k+v}} \\ &= \left(\frac{a}{1+a}\right)^v \frac{v(v+1)\dots(v+k-1)}{(1+a)^k k!} \end{aligned} \tag{3.40}$$

For simplicity in notation we generalize the symbol $\binom{n}{r}$, which has been used only for positive integer values of n. For every x and every positive integer r we denote

$$\binom{x}{r} = \frac{x(x-1)\dots(x-r+1)}{r!} = \frac{x^{(r)}}{r!}$$

Further, let $p = 1/(1+a)$ and $q = 1 - p = a/(1+a)$. By assumption, we have $a > 0$, and hence the inequalities $0 < p < 1$ and $0 < q < 1$. Using this notation, we can write (3.7) in the form

$$P(X = k) = (-1)^k \binom{-v}{k} p^k q^v \quad (k = 0, 1, 2, \dots) \tag{3.41}$$

Definition 3.5.2 *The compound Poisson distribution whose probability function is defined by formula (3.41), is called the negative binomial distribution. Let us compute the* *characteristic function of this distribution.*

$$\phi(t) = \sum_{k=0}^{\infty} P(X = k)e^{itk} = q^v \sum_{k=0}^{\infty} (-1)^k \binom{-v}{k} p^k e^{itk}$$

Using Maclaurin's expansion,

$$(1-p)^{-v} = \sum_{k=0}^{\infty} (-1)^k \binom{-v}{k} p^k, \quad |p| < 1$$

we obtain

$$\phi(t) = q^v \left(1 - pe^{it}\right)^{-v} \tag{3.42}$$

It follows that $m_1 = \frac{\phi'(0)}{i} = v\frac{p}{q}, \quad m_2 = \frac{\phi''(0)}{i^2} = \left(\frac{p}{q}\right)^2 v(v+1) + \frac{p}{q}v$,

$$\mu_2 = m_2 - m_1^2 = v\frac{p}{q}\left(1 + \frac{p}{q}\right)$$

For the ordinary moment of order r we obtain the formula $m_r = \sum_{l=0}^{r-1} (-1)^{r-l} \binom{r-1}{l} \left(\frac{p}{q}\right)^{r-l} (-v)_{r-l}$ $(r = 1, 2, \ldots)$. Greenwood and Yule gave some examples of applications of the negative binomial distribution, of which one follows.

Example 3.5.3 *The number of accidents among 414 machine operators was investigated for three successive months. The data are presented in Table. The symbol k in the first column denotes the number of accidents which happened.*

Table

Observed k Frequency Probability

0	0.715	0.722
1	0.179	0.167
2	0.063	0.063
3	0.019	0.027
4	0.010	0.012
5	0.010	0.005
6	0.002	
7	0.000	
8	0.002	

to the same operator during the period under investigation. In the second column are given the observed frequencies for the operators who had k accidents in the period under investigation, and in the third column, the corresponding probabilities calculated from formula (3.8). The unknown parameters v and p appearing in this formula were found in the following way: the expected value and the variance of the observed distribution were computed, and then it was assumed that they coincide with the values of $E(X)$ and μ_2 defined by (3.10). In this way two equations were obtained which make it possible to determine the unknown parameters. As we see, the observed frequencies differ little from the computed probabilities. This can be explained as follows. The probability that a machine operator will have k accidents during the period under investigation is determined by the Poisson distribution with parameter λ. The value of this parameter is influenced by many factors depending on time, such as the extent of the protective measures taken and the atmospheric conditions. We can regard λ as a random variable. Assuming that λ has a gamma distribution, it has been established that the observed and predicted frequencies are close to each other.

Definition 3.5.4 *The distribution of a random variable X given by (3.8), where v is an integer, is called the Pascal distribution. This distribution can also be obtained from other considerations, not as a compound but as a simple distribution. Consider a sequence of experiments. Suppose that as a result of an experiment either the event A or the event $\bar{A}$ may occur, and suppose that the results of the experiments are independent. We say there is a success if the event A occurs, and a failure if not. Suppose that $P(A) = p$; thus $P(\bar{A}) = 1 - p = q$. Denote by X_r the number of successes following the $(r-1)$ th failure and preceding the r th failure. Thus, for instance, X_1 is the number of successive successes* *preceding the first failure, X_2 the number of successive successes after the first failure and before the second failure. Let us consider the random variable $X = X_1 + X_2 + \ldots + X_v$. The event $X = k$ is the product of two events; the event that the $(k+v)$ th experiment will lead to a failure and the event that among the remaining $k+v-1$ experiments k will lead to successes. The probability of the first event is q and the probability of the second is;*

$$\binom{k+v-1}{k} p^k q^{v-1}$$

Hence

$$\begin{aligned} P(X=k) &= \binom{k+v-1}{k} p^k q^v = \frac{v(v+1)\ldots(v+k-1)}{k!} p^k q^v \\ &= (-1)^k \binom{-v}{k} p^k q^v \quad (k = 0, 1, 2, \ldots) \end{aligned}$$

This formula is identical with (3.8). We now prove a theorem for the negative binomial distribution, which is analogous to theorem for the binomial distribution. If the equation

$$E(X) = v\frac{p}{q} = c$$

where c is a positive constant, is satisfied for every v, then the probability function of the negative binomial distribution tends to the corresponding function of the Poisson distribution as $v \to \infty$.

Proof: From (3.8) we have

$$\begin{aligned} \lim_{v\to\infty} P(X=k) &= \lim_{v\to\infty} \frac{v(v+1)\ldots(v+k-1)}{k!} p^k (1-p)^v \\ &= \frac{c^k}{k!} \lim_{v\to\infty} \frac{v(v+1)\ldots(v+k-1)}{(v+c)^k} \left(1 - \frac{c}{v+c}\right)^v = \frac{e^{-c}c^k}{k!} \end{aligned} \quad (3.43)$$

Formula (3.13) allows us to apply tables of the Poisson distribution to a negative binomial distribution. Consider now the random variable Y defined by the formula

$$Y = \sum_{k=1}^{N} X_k \quad (3.44)$$

where $X_k (k = 1, 2, \ldots)$ and N are random variables and N takes on only positive integer values.

Theorem 3.5.5 *Let the random variable N be independent of the random variables*

$X_1, X_2, \ldots$ *If the inequality*

$$\sum_{k=1}^{\infty} P(N \geqslant k) E(|X_k|) < \infty \qquad (3.45)$$

is satisfied, the expected value of the random variable Y exists and

$$E(Y) = \sum_{k=1}^{\infty} P(N \geqslant k) E(X_k) \qquad (3.46)$$

Proof: *From (3.14) and (3.22) we obtain*

$$\begin{aligned} E(Y) &= \sum_{n=1}^{\infty} P(N = n) E(Y \mid N = n) \\ &= \sum_{n=1}^{\infty} P(N = n) \sum_{k=1}^{n} E(X_k) = \sum_{k=1}^{\infty} E(X_k) \cdot \sum_{n=k}^{\infty} P(N = n) \\ &= \sum_{k=1}^{\infty} E(X_k) P(N \geqslant k) \end{aligned}$$

From (3.15) the theorem follows. Suppose in addition that the random variables X_k have the same distribution. Then, assumption (3.15) is satisfied if $E(N)$ and the expected value $E(X)$ of X_k exist. Here, formula (3.16) has the form

$$E(Y) = E(X) \sum_{k=1}^{\infty} P(N \geqslant k) = E(X) \sum_{k=1}^{\infty} k P(N = k) = E(X) E(N) \qquad (3.48)$$

The reader will notice that the expected value of the compound binomial distribution satisfies relation (3.17).

Check Your Progress

1. What is a compound distribution?

A. A distribution resulting from combining two or more distributions

B. A distribution resulting from scaling a single distribution

C. A distribution resulting from shifting a single distribution

D. A distribution resulting from adding two or more identical distributions

2. Which of the following distributions can be used to model the number of successes

in a sequence of independent Bernoulli trials with a random number of trials?
A. Negative binomial distribution
B. Poisson distribution
C. Uniform distribution
D. Exponential distribution

Glossary

1. The $F_1'(y)$ is pdf of normal distribution.
2. The σ is standard deviation of normal distribution.
3. The $\Gamma(p)$ is gamma distribution of p.
4. The μ_r is rth order moment of beta distribution.

3.6 Let Us Sum Up

Learners, in this section we have seen that definition of compound distribution and two dimensional distribution with Illustrations.

3.7 Unit Summary

The third unit content on the normal distribution, gamma distribution, beta distribution, cauchy distribution, laplace distributions, multinomial distribution and compound distributions. Also given theorems and examples.

Self-Assessment Questions

Short Answers: (5 Marks)

1. Show that if the random variables X_1 and X_2 have zero-one distributions and are uncorrelated, they are independent. (b) Check whether this property holds

for all two-point random variables. The random variable X has the binomial distribution given by (2.1). Let

$$A_{nk} = \frac{P(X = k+1)}{P(X = k)} \quad (k = 0, 1, \ldots, n-1)$$

2. Prove that (a) the expressions $A_{nk} - 1$ and $(n+1)p - 1 - k$ are either both equal to zero, both positive, or both negative. (b) $P(X = k)$ takes on its maximum value either at one point k_0, which satisfies the inequality

$$(n+1)p - 1 < k_0 < (n+1)p$$

if $(n+1)p$ is not an integer, or at two points $(n+1)p - 1$ and $(n+1)p$ if $(n+1)p$ is an integer. (c) for $k > (n+1)p$

$$A_{nk} < \exp\left[-\frac{k + (n+1)p}{n}\right]$$

$$\sum_{l=0}^{k} \binom{n_1}{l}\binom{n_2}{k-l} = \binom{n_1 + n_2}{k}$$

3. Prove that

$$\sum_{m=k}^{n} \binom{n}{m} p^m (1-p)^{n-m} = \frac{n!}{(k-1)!(n-k)!} \int_0^p t^{k-1}(1-t)^{n-k} dt.$$

4. Prove that for arbitrary $\lambda_1 > 0, \lambda_2 > 0$, and non-negative integer k

$$\sum_{l=0}^{k} \frac{\lambda_1^l \lambda_2^{k-l}}{l!(k-l)!} = \frac{(\lambda_1 + \lambda_2)^k}{k!}$$

Long Answers: (8 Marks)

1. If $F(x)$ is the distribution function of a random variable X with the zero one distribution. Find the distribution function of the random variable $Y = F(X)$.

   (a) Do the same for a random variable X with the binomial distribution.

   (b) Do the same when X has the Poisson distribution.

2. The random variable X is said to have a $\log$ normal distribution if its density is of the form

$$f(x) = \begin{cases} 0 \\ \frac{1}{(x-c)\sigma\sqrt{2\pi}} \exp\left\{-\frac{[\log(x-c)-m]^2}{2\sigma^2}\right\} & (x \leqslant c) \\ (x < c) \end{cases}$$

where c is a constant. Find $E(X)$ and $D^2(X)$.

3. (a). Prove that for every $x > 0$

$$\frac{1}{\sqrt{2\pi}} e^{-x^2/2}\left(\frac{1}{x} - \frac{1}{x^3}\right) < 1 - \Phi(x) < \frac{1}{x\sqrt{2\pi}} e^{-x^2/2}$$

where $\Phi(x)$ is the distribution function of the normal distribution $N(0;1)$.
(b) Find the analogous inequality for $x < 0$.

Exercises

1. The random variables $X_i (i = 1, 2, 3)$ are independent and have the same distribution $N(0;1)$. Find the distribution function of the random variable $Y = \max_{1 \leqslant i \leqslant 3} |X_i|$.

2. Let $X_1, \dots, X_n$ be independent and have the same distribution function $F(x)$ and moments of arbitrary order. Let $a_1, \dots, a_n$ and $b_1, \dots, b_n$ satisfy the relations

$$\sum_{j=1}^{n} a_j = \sum_{j=1}^{n} b_j; \sum_{j=1}^{n} {a_j}^2 = \sum_{j=1}^{n_1} {b_j}^2$$

where the sequence $a_1, \dots, a_n$ is not a permutation of the sequence $b_1, \dots, b_n$. The distribution function $F(x)$ is normal if and only if the random variables L_1 and L_2 defined as

$$L_1 = \sum_{j=1}^{n} a_j X_j, L_2 = \sum_{j=1}^{n} b_j X_j$$

have the same distribution.

3. Let $X_1, \dots, X_n$ be independent and have the same nondegenerate distribution function $F(x)$. Let

$$U = X_1 + \ldots + X_n, \quad V = \sum_{r=1}^{n} \sum_{s=1}^{n} a_{rs} X_r X_s$$

$$B_1 = \sum_{r=1}^{n} a_{rr}, B_2 = \sum_{r=1}^{n} \sum_{s=1}^{n} a_{rs}.$$

(a) If $B_1 \neq 0$ and $B_2 = 0$, then $F(x)$ is the normal distribution function if and only if V has constant regression (of the first type) on U, i.e., $E(V \mid u) = E(V)$, with probability 1.

(b) If $E(V) = 0, B_1 \neq 0$, and $B_2 \neq 0$, then $F(x)$ is the gamma distribution function if and only if V has constant regression on U.

Answers to Check Your Progress

Session (Modulo) 3.1

1. A. Symmetric bell curve
2. A. 1

Session (Modulo) 3.2

1. A. hape parameter (α) and scale parameter (β)
2. D. Chi-squared distribution

Session (Modulo) 3.3

1. A. α and β
2. C. $\frac{\alpha}{\alpha+\beta}$

Session (Modulo) 3.4

1. A.Binomial distribution
2. B. np_i

Session (Modulo) 3.5

1. A. A distribution resulting from combining two or more distributions
2. A. Negative binomial distribution

References

1. M. Fisz, Probability Theory and Mathematical Statistics, John Wiley and sons, New Your, Third Edition, 1963.

Suggested Readings

1. T. Veerarajan, Fundamentals of Mathematical Statistics, Yesdee Publishing, 2017.

2. P.R. Vittal, Mathematical Statistics, Margham Publications, 2002.

3. T. Veerarajan, Probability, Statistics and Random Processes, Mc Graw Hill Education (India) Private Limited, Third Edition, 2015.

4. R.S.N. Pillai and V. Bagavathi, Statistics, Sultan Chand and Co., 2010.

5. Singaravelu.A, S. Sivasubramanian, Probability and Random Processes , Meenakshi Agency, 2008,

6. DN Elhance, Veena Elhance and BM Aggarwal, Fundamentals of Statistics, Kitab Mahal.

7. S. C. Gupta and V. K. Kapoor, Fundamentals of Mathematical Statistics, Sultan Chand and Sons, 2008.

Unit 4

Limit Theorems

Objective

This course aims to teach the students about limit theorems of stochastic convergence, Bernoulli's law of large numbers with the convergence of a sequence of distribution functions and Levy-Cramer theorem, De Moivre-Laplace theorem, Lindeberg-Levy theorem and Lapunov theorem.

4.1 Stochastic Converges

In this section we discuss the modern theory of limit distributions for sums of independent random variables has developed greatly during the last thirty years, due mainly to Khintchin, Gnedenko, Kolmogorov, and Lévy. A uniform general theory has been developed, in which the limit theorems presented in this section are only particular cases of general theorems which give conditions for the convergence of a sequence of distribution functions of sums (much more general than the sums considered here) to a limit distribution function and establish the set of all possible limit distributions. The reader will find a detailed discussion of this theory in the books by Lévy Gnedenko and Kólmogorov, and Loève. The question of the convergence of a sequence of distribution functions for dependent random variables is also extremely interesting. Investigations in this domain were originated by Markov. Bernstein has obtained some important results. Consider the following example.

Example 4.1.1 *The random variable Y_n can take on the values*

$$0, \frac{1}{n}, \frac{2}{n}, \ldots, \frac{n-1}{n}, 1$$

and its probability function is given by the formula

$$P\left(Y_n = \frac{r}{n}\right) = \binom{n}{r}\frac{1}{2^n} \quad (r = 0, 1, \ldots, n) \tag{4.1}$$

Consider the random variable X_n defined by the formula

$$X_n = Y_n - \frac{1}{2} \tag{4.2}$$

Thus X_n can take on the values

$$-\frac{1}{2}, \frac{2-n}{2n}, \frac{4-n}{2n}, \ldots, \frac{n-4}{2n}, \frac{n-2}{2n}, \frac{1}{2}$$

The probability function of X_n is given by the formula

$$P\left(X_n = \frac{2r-n}{2n}\right) = \binom{n}{r}\frac{1}{2^n}$$

Let $n = 2$. The random variable X_2 can take on the values

$$-0.5, 0, 0.5$$

with the respective probabilities $\frac{1}{4}, \frac{1}{2}, \frac{1}{4}$. Let ε be a positive number, say $\varepsilon = 0.3$. We see that

$$P\left(|X_2| > 0.3\right) = P\left(X_2 = -\frac{1}{2}\right) + P\left(X_2 = \frac{1}{2}\right) = 0.5$$

Now let $n = 5$. The random variable X_5 can take on the values

$$-0.5, -0.3, -0.1, 0.1, 0.3, 0.5$$

with the respective probabilities

$$\frac{1}{32}, \frac{5}{32}, \frac{10}{3}\frac{1}{2}, \frac{10}{3}, \frac{5}{32}, \frac{1}{32}$$

Hence

$$P\left(|X_5| > 0.3\right) = 0.0625$$

Now let $n = 10$. The random variable X_{10} can take on the values

$$-0.5, -0.4, -0.3, -0.2, -0.1, 0.0, 0.1, 0.2, 0.3, 0.4, 0.5$$

with the respective probabilities $\overline{1}\frac{1}{1024}, \frac{10}{1024}, \frac{45}{1024}, \frac{120}{1024}, \frac{210}{1024}, \frac{252}{1024}, \frac{210}{1024}, \frac{120}{1024}, \frac{45}{1024}, \frac{10}{1024}, \frac{-1}{1024}$. *Hence*

$$P\left(|X_{10}| > 0.3\right) \cong 0.02$$

We see that for $n = 10$ the probability that X'_n will exceed $\varepsilon = 0.3$ in absolute value is very small. From the theorem, it follows that in our example

$$\lim_{n\to\infty} P\left(|X_n| > 0.3\right) = 0 \tag{4.3}$$

and, moreover, that for the sequence of random variables X_n defined by formula (4.2), relation (4.3) is satisfied for every $\varepsilon > 0$. Before we present the theorem just mentioned, we define the notion of stochastic convergence.

Definition 4.1.2 *The sequence $\{X_n\}$ of random variables is called stochastically convergent* [1] *to zero if for every $\varepsilon > 0$ the relation*

$$\lim_{n\to\infty} P\left(|X_n| > \varepsilon\right) = 0 \tag{4.4}$$

is satisfied. We notice that in this definition we say nothing about the convergence of the random variables X_n to zero in the sense which is understood in analysis. Thus, if the sequence $\{\mathrm{X}_n\}$ is stochastically convergent to zero, it does not follow that for every $\varepsilon > 0$ we can find a finite n_0 such that for all $n > n_0$ the relation $|X_n| < \varepsilon$ will be satisfied. From the definition of stochastic convergence it follows only that the probability of the event $|X_n| \geqslant \varepsilon$ tends to zero as $n \to \infty$.

Theorem 4.1.3 *Let $F_n(x) (n = 1, 2, \ldots)$ be the distribution function of the random variable X_n. The sequence $\{X_n\}$ is stochastically convergent to zero if and only if the sequence $\{F_n(x)\}$ satisfies the relation*

$$\lim_{n\to\infty} F_n(x) = \begin{cases} 0 & \textit{for } x \leqslant 0 \\ 1 & \textit{for } x > 0 \end{cases} \tag{6.2.5}$$

Proof: *Suppose that the sequence $\{X_n\}$ is stochastically convergent to zero. From relation (4.4) it follows that for an arbitrary $\varepsilon > 0$ as $n \to \infty$ we have*

$$P(X_n < -\varepsilon) = F_n(-\varepsilon) \to 0 \tag{4.6}$$

$$P(X_n > \varepsilon) = 1 - F_n(\varepsilon) - P(X_n = \varepsilon) \to 0$$

Since for every $\varepsilon > 0$ we can find an ε_1 such that $0 < \varepsilon_1 < \varepsilon$, it follows from relation (4.4) that for an arbitrary $\varepsilon > 0$ we have $P(X_n = \varepsilon) \to 0$. Hence from (4.6) it follows that

$$1 - F_n(\varepsilon) \to 0 \tag{4.7}$$

Replacing ε by $-x$ in formula (4.6) and by x in formula (4.7), where $x > 0$, we obtain (4.5). Suppose now that (4.5) is satisfied. Then for arbitrary $\varepsilon > 0$ we have

$$\lim_{n\to\infty} P(X_n < -\varepsilon) = \lim_{n\to\infty} F_n(-\varepsilon) = 0$$

$$\lim_{n\to\infty} P(X_n > \varepsilon) \leqslant \lim_{n\to\infty} P(X_n \geqslant \varepsilon) = \lim_{n\to\infty} [1 - F_n(\varepsilon)] = 0$$

Relation (4.4) follows immediately from the last two relations, which proves the theorem. We remind the reader that the random variable X with a one-point distribution such that $P(X = 0) = 1$, has the distribution function

$$F(x) = \begin{cases} 0 & \textit{for } x \leqslant 0 \\ 1 & \textit{for } x > 0 \end{cases} \tag{4.8}$$

and this distribution function is continuous at every point $x \neq 0$. From relations (4.6) and (4.7) it follows that for every point $x \neq 0$ the sequence of distribution functions $F_n(x)$ converges (in the usual sense) to the distribution function $F(x)$ defined by formula (4.8). We conclude that the sequence of distribution functions $F_n(x)$ of random variables, convergent stochastically to zero, converges to the distribution function of the one-point distribution at every point $x \neq 0$. Since the points $x \neq 0$ are continuity points of this distribution function, we can formulate the preceding result in the following way. The sequence $\{X_n\}$ of random variables is stochastically convergent to zero if and only if the sequence $\{F_n(x)\}$ of distribution functions of these random variables is convergent to the distribution function $F(x)$ given by (4.8) at every continuity point of the latter. We stress *the fact that at the point of discontinuity of $F(x)$, that is, at the point $x = 0$, the sequence $\{F_n(0)\}$ may not converge to $F(0)$. We can also consider the stochastic convergence of a sequence of random variables $\{X_n\}$ to a constant $c \neq 0$. This will mean that the sequence of random variables $\{Y_n\} = \{X_n - c\}$ is stochastically convergent to zero. Similarly, we can define the stochastic convergence of a sequence of random variables $\{X_n\}$ to a random variable X. This will mean that the sequence of random variables $\{Z_n\} = \{X_n - X\}$ is stochastically convergent to zero. We now prove the theorem stated in Section 6.2, of which formula (4.3) is a particular case. Denote by $\{Y_n\}$ the sequence of random variables with probability functions given by the formula*

$$P\left(Y_n = \frac{r}{n}\right) = \binom{n}{r} p^r (1-p)^{n-r} \tag{4.9}$$

where $0 < p < 1$ and r can take on the values $0, 1, 2, \ldots, n$. Further denote

$$X_n = Y_n - p \tag{4.10}$$

Theorem 4.1.4 *The sequence of random variables $\{X_n\}$ given by (4.7) and (4.8) is stochastically convergent to 0, that is, for any $\varepsilon > 0$ we hare*

$$\lim_{n \to \infty} P\left(|X_n| > \varepsilon\right) = 0 \tag{4.11}$$

Proof: *We shall use the Chebyshev inequality in the proof. By equalities (3.8) we have*

$$E\left(X_n\right) = 0 \tag{4.12}$$

$$\sigma_n = \sqrt{D^2\left(X_n\right)} = \sqrt{p(1-p)/n} \tag{4.13}$$

Substituting (4.4) and (4.5) into the Chebyshev inequality, we obtain

$$P\left(|X_n| > k\sqrt{p(1-p)/n}\right) \leqslant \frac{1}{k^2} \tag{4.14}$$

where k is an arbitrary positive number. Set

$$k = \varepsilon\sqrt{n/p(1-p)}$$

Let Us Sum Up

Learners, in this section we have seen that definition of limit theorems and stochastic converges and also given theorems and with Illustrations.

Check Your Progress

1. Let $\{X_n\}$ be a sequence of random variables. If $X_n \xrightarrow{p} X$ and $X_n \xrightarrow{d} X$ (converges in distribution), which of the following statements is true?
A. Convergence in probability implies convergence in distribution.
B. Convergence in distribution implies convergence in probability.
C. Convergence in probability implies almost sure convergence.
D. Convergence in distribution implies almost sure convergence.
2. If $X_n \xrightarrow{a.s.} X$ which of the following statements about X_n and X is true?
A. X_n converges in L^2 if $X_n \xrightarrow{a.s.} X$
B. $X_n \xrightarrow{p} X$
C. $X_n \xrightarrow{d} X$
D. X_n converges in mean if $X_n \xrightarrow{a.s.} X$

4.2 Convergence of A Sequence of Distribution Functions

The inequality (4.6). We then obtain the inequality

$$P\left(|X_n| > \varepsilon\right) \leqslant \frac{p(1-p)}{n\varepsilon^2} < \frac{1}{n\varepsilon^2} \tag{4.15}$$

From inequality (4.15) it follows that for every $\varepsilon > 0$ we have (4.13), which was to be proved. The theorem just proved is called the Bernoulli law of large numbers. This law can be interpreted in practice as follows. We perform n experiments according to the Bernoulli scheme, where the probability of the event A is equal to p. The law of large numbers states that, for large values of n, the probability that the observed frequency of A will differ little from p is close to one. In the following sections we investigate other laws of large numbers. In Section 4.2 we considered a sequence of distribution functions which converges to the distribution function (4.8) of the one-point distribution at every continuity point of this distribution. We now investigate the question of convergence of sequences of distribution functions generally.

Definition 4.2.1 *The sequence $\{F_n(x)\}$ of distribution functions of the random variables $\{X_n\}$ is called convergent, if there exists a distribution function $F(x)$ such that, at every continuity point of $F(x)$, the relation*

$$\lim_{n\to\infty} F_n(x) = F(x) \tag{4.16}$$

is satisfied. The distribution function $F(x)$ is called the limit distribution function. As we see, it is not required that the sequence $\{F_n(x)\}$ converge to $F(x)$ at the discontinuity points of $F(x)$. The sequence $\{X_n\}$ of random variables defined by formula (4.2) is stochastically convergent to zero; thus the sequence $\{F_n(x)\}$ of their distribution functions converges to the distribution function $F(x)$ defined by formula (4.8). This distribution function is discontinuous at $x = 0$. It is easy to verify that the sequence of numbers $\{F_n(0)\}$ is not convergent to $F(0)$. Consider the subsequence of the sequence $\{F_n(0)\}$ containing only terms with the odd indices $n = 2k + 1$. The random variable X_{2k+1} can take on the values

$$-\frac{1}{2}, \frac{2-(2k+1)}{2(2k+1)}, \frac{4-(2k+1)}{2(2k+1)}, \ldots, \frac{2k+1-4}{2(2k+1)}, \frac{2k+1-2}{2(2k+1)}, \frac{1}{2}$$

For every k, half of these terms are each less than zero, the other half greater than zero. The probability that X_{2k+1} will take on a value less than zero equals 0.5 . Thus, for every k we have $P(X_{2k+1} < 0) = F_{2k+1}(0) =$ 0.5 . Since $F(0) = 0$, we have

$$\lim_{k\to\infty} F_{2k+1}(0) = 0.5 \neq F(0)$$

From (4.16) it follows that $\lim F_n(0) \neq F(0)$. Nevertheless, by the definition of the convergence of a sequence of distribution functions, the sequence of distribution functions of example 6.2.1 is convergent to the distribution function given by formula (4.8). It is important to note that we speak about the convergence of a sequence of distribution functions only when it is convergent to a distribution function. This remark is important since it may happen that a sequence of distribution functions converges to a function that is not a distribution function.

Example 4.2.2 *Let us consider the sequence $\{X_n\}$ of random variables with the one-point* *distributions given by the formula*

$$P\left(X_n = n\right) = 1 \quad (n = 1, 2, \ldots)$$

The distribution function $F_n(x)$ *of* X_n *is of the form*

$$F_n(x) = \begin{cases} 0 & \text{for } x \leqslant n \\ 1 & \text{for } x > n \end{cases}$$

We have the relation

$$\lim_{n\to\infty} F_n(x) = 0 \quad (-\infty < x < \infty)$$

Thus the sequence $\{F_n(x)\}$ *is not convergent to a distribution function. Let the sequence* $\{F_n(x)\}$ *be convergent to the distribution function* $F(x)$*. Let* a *and* b*, where* $a < b$*, be two arbitrary continuity points of the limit distribution function* $F(x)$*. Then we have*

$$\lim_{n\to\infty} P\left(a \leqslant X_n < b\right) = F(b) - F(a) \tag{4.17}$$

In fact,

$$P\left(a \leqslant X_n < b\right) = F_n(b) - F_n(a) \tag{4.18}$$

From the assumption that a *and* b *are continuity points of the distribution function* $F(x)$ *it follows that*

$$F_n(b) \to F(b), \quad F_n(a) \to F(a) \tag{4.19}$$

From (4.18) and (4.19) follows (4.17). Let the sequence $\{F_n(x)\}$ *be convergent to the distribution function* $F(x)$*. Let* $P_n(S)$ *and* $P(S)$ *denote the probability functions corresponding respectively to the distribution functions* $F_n(x)$ *and* $F(x)$*. It can be shown that for an arbitrary Borel set* S *on the real line* R *such that* $P(\bar{S} \cap \overline{R - S}) = 0$ *(here* $\bar{A}$ *denotes the closure of the set* A *) we have the relation*

$$\lim_{n\to\infty} P_n(S) = P(S) \tag{4.20}$$

We observe that even when the limit distribution function is everywhere continuous, Borel sets S *may exist, for which (4.20) is not satisfied. This will happen if* $P(\bar{S} \cap \overline{R - S}) > 0$*. The following example is due to Robbins.*

Example 4.2.3 *The random variable $X_n (n = 1, 2, \ldots)$ has the density $f_n(x)$ given by*

$$f_n(x) = \begin{cases} \frac{2^n}{\varepsilon} & \textit{if } \frac{i}{n} - \frac{\varepsilon}{(n2^n)} < x < \frac{i}{n} \quad (i = 1, \ldots, n) \\ 0 & \textit{otherwise} \end{cases}$$

where $0 < \varepsilon < 1$. The distribution function $F_n(x)$ of X_n is then, for $i = 1, \ldots, n$, of the form

$$F_n(x) = \begin{cases} 0 & \textit{if} \quad x \leqslant 0 \\ \frac{i-1}{n} & \textit{if} \quad \frac{i-1}{n} \leqslant x \leqslant \frac{i}{n} - \frac{\varepsilon}{(n2^n)} \\ \frac{i-1}{n} + \frac{2^n\left(x - \frac{i}{n} + \frac{\varepsilon}{n2^n}\right)}{\varepsilon} & \\ 1 & \textit{if } \frac{i}{n} - \frac{\varepsilon}{n2^n} < x < \frac{i}{n} \\ \textit{if } x \geqslant 1 & \end{cases} \tag{4.21}$$

Thus for every x in the interval $I = [0, 1]$ we have

$$0 \leqslant x - F_n(x) \leqslant \frac{1}{n}$$

By considering the values taken by $F_n(x)$ outside the interval I, we obtain for every real x

$$\lim_{n \to \infty} F_n(x) = F(x) = \begin{cases} 0 & \textit{if } x \leqslant 0 \\ x & \textit{if } 0 < x < 1 \\ 1 & \textit{if } x \geqslant 1 \end{cases} \tag{4.22}$$

Let us denote by S_n the set on which $f_n(x) > 0$, and by S_∞ the Borel set defined as

$$S_\infty = \sum_{n=1}^{\infty} S_n$$

Let $P_n(S)$ and $P(S)$ denote the probability functions which correspond to the distribution functions $F_n(x)$ and $F(x)$, respectively. We have for $n = 1, 2, \ldots$

$$P_n(S_n) = \int_{S_n} f_n(x) dx = 1$$

Since $S_\infty > S_n$ we obtain $P_n(S_\infty) = 1$; hence,

$$\lim_{n\to\infty} P_n(S_\infty) = 1$$

On the other hand,

$$P(S_\infty) = \int_{S_\infty} dx \leqslant \sum_{n=1}^{\infty} \left(n \frac{\varepsilon}{n2^n} \right) = \varepsilon < 1 \tag{4.24}$$

thus

$$\lim_{n\to\infty} P_n(S_\infty) \neq P(S_\infty)$$

despite the fact that the distribution function $F(x)$ given by (4.22) is every where continuous. This is because $P\left(\bar{S}_\infty \cap \overline{R - S_\infty}\right) > 0$. Indeed [1]*, we have $P\left(\bar{S}_\infty \cap \overline{R - S_\infty}\right) = P\left(\bar{S}_\infty \cap \overline{I - S_\infty}\right)$, and the set $I - S_\infty$ is perfect and nowhere dense in I; hence $I - S_\infty = \overline{I - S_\infty}$ and $\overline{I - (I - S_\infty)} = \bar{S}_\infty = I$. Thus we obtain, using (4.24),*

$$P\left(\bar{S}_\infty \cap \overline{I - S_\infty}\right) = P(I \cap I - S_\infty) = P(I - S_\infty) \geqslant 1 - \varepsilon$$

We now give the generalization of definition to random vectors.

Definition 4.2.4 *The sequence of distribution functions $\{F_n(x_1, \ldots, x_k)\}$ of random vectors $(X_{n1}, X_{n2}, \ldots, X_{nk})$ is convergent if there exists a distribution function $F(x_1, \ldots, x_k)$ such that at every one of its continuity points.*

$$\lim_{n\to\infty} F_n(x_1, x_2, \ldots, x_k) = F(x_1, x_2, \ldots, x_k) \tag{4.25}$$

It is not difficult to show that if (6.25) holds, and $P_n(S)$ and $P(S)$ denote the respective probability functions, then for every Borel set in k-dimensional Euclidean space R^k such that $P\left(\bar{S} \cap \overline{R^k - S}\right) = 0$ relation (4.20) holds. This relation holds, in particular, for continuity intervals. The following theorem has important applications. We present it without proof.

Theorem 4.2.5 *Let $\{F_n(x_1, \ldots, x_k)\}$ $(n = 1, 2, \ldots)$ be a sequence of distribution functions of random vectors $(X_{n1}, \ldots, X_{nk})$ and let $F(x_1, \ldots, x_k)$ and $P(S)$ be the distribution function and probability function of a random vector $(X_1, \ldots, X_k)$, respectively. Relation (6.25) holds if and only if for* [1] *Information concerning the notions introduced here can be found in this section.*

Let Us Sum Up

Learners, in this section we have seen that the definition of convergence of a sequence of distribution functions and also given theorems and examples.

Check Your Progress

1. Which of the following is a necessary condition for the convergence of $\{F_n\}$ to F in distribution?
A. $F_n(x) \to F(x)$ uniformly for all $x \in \mathbb{R}$.
B. $F_n(x) \to F(x)$ at every point $x \in \mathbb{R}$.
C. $F_n(x) \to F(x)$ at all continuity points of F.
D. $F_n(x) \to F(x)$ for all x such that F is not continuous.
2. If $\{F_n\}$ converges to F uniformly, then $\{F_n\}$ also converges is:
A. Almost surely to F
B. In probability to F
C. In distribution to F
D. In mean to F

4.3 The Riemann-Stieltjes Integral and Law of Large Numbers

Every function $g(x_1, \ldots, x_k)$ continuous on a set S satisfying the relation $P(S) = 1$, the equality

$$\lim_{n\to\infty} H_n(\alpha) = H(\alpha)$$

holds at every continuity point α of $H(\alpha)$, where $H_n(\alpha)$ and $H(\alpha)$ are the distribution functions of $g(X_{n1}, \ldots, X_{nk})$ and $g(X_1, \ldots, X_k)$, respectively. In further considerations we use the theorem proved by Levy and Cramér which makes it possible to investigate the convergence of a sequence of distribution functions $\{F_n(x)\}$ of random variables $\{X_n\}$ to a distribution function $F(x)$ by investigating the convergence of the sequence of characteristic functions $\phi_n(t)$ of the random variables X_n. This theorem plays an important role in probability theory. The proof of this theorem requires the notion of the Riemann-Stieltjes integral, which for simplicity is called the Stieltjes integral.

It will be seen that distributions of random variables of the continuous and discrete types, considered separately, can be treated together by means of the Stieltjes integral. We first introduce the notion of a function of bounded variation. Let $F(x)$ be a function defined in the interval $[a, b]$, which can be either finite or infinite. Let us take a partition of the interval $[a, b]$ with the points

$$a = x_0 < x_1 < x_2 < \ldots < x_n = b$$

and form the sum

$$T = \sum_{k=0}^{n-1} \left| F\left(x_{k+1}\right) - F\left(x_k\right) \right|$$

The value of T may depend, of course, on the number n and on the partition into subintervals.

Definition 4.3.1 *The least upper bound of the values of T is called the total absolute variation of the function $F(x)$ in the interval $[a, b]$.*

Definition 4.3.2 *If the total absolute variation of the function $F(x)$ in the interval $[a, b]$ is finite, we shall say that F is a function of bounded variation on the interval $[a, b]$. It is easy to verify that every nondecreasing bounded function is of bounded variation. Indeed, here the expression $F\left(x_{k+1}\right) - F\left(x_k\right)$ is nonnegative for arbitrary k and arbitrary partition of the interval $[a, b]$, hence*

$$T = \sum_{k=0}^{n-1} \left[F\left(x_{k+1}\right) - F\left(x_k\right) \right] = F(b) - F(a)$$

and our assertion then follows from the assumption that $F(b)$ and $F(a)$ are finite. It also follows that every distribution function $F(x)$.

Let Us Sum Up

Learners, in this section we have seen that the definition of Riemann-Stieltjes integral and law of large numbers and also given theorems and examples.

Check Your Progress

1. Which of the following conditions is necessary for the existence of the Riemann-Stieltjes integral $\int_a^b f(x)\,dg(x)$?

A. f must be continuous on $[a, b]$ and g must be of bounded variation on $[a, b]$.

B. f must be integrable over $[a, b]$ and g must be continuous on $[a, b]$.

C. f must be differentiable on $[a, b]$ and g must be of bounded variation.

D. f must be continuous on $[a, b]$ and g must be differentiable on $[a, b]$.

2. The formula for integration by parts in the Riemann-Stieltjes integral is:

A. $\int_a^b f(x)\,dg(x) = f(b)g(b) - f(a)g(a) - \int_a^b g(x)\,df(x)$

B. $\int_a^b f(x)\,dg(x) = f(a)g(b) - f(b)g(a) + \int_a^b g(x)\,df(x)$

C. $\int_a^b f(x)\,dg(x) = f(a)g(b) - f(a)g(a) - \int_a^b f(x)\,dg(x)$

D. $\int_a^b f(x)\,dg(x) = \int_a^b g(x)\,df(x)$

4.4 Levy-Cramer Theorem

If $Z = X/Y$ and $P(Y = 0) = 0$,

$$F(z) = \int_{-\infty}^{0} [1 - F_1(zy)]\,dF_2(y) + \int_{0}^{\infty} F_1(zy)dF_2(y) \tag{4.26}$$

We first present the Levy-Cramer theorem in the form of two theorems.

Theorem 4.4.1 *If the sequence $\{F_n(x)\}$ $(n = 1, 2, \ldots)$ of distribution functions is convergent to the distribution function $F(x)$, then the corresponding sequence of characteristic functions $\{\phi_n(t)\}$ converges at every point t $(-\infty < t < +\infty)$ to the function $\phi(t)$ which is the characteristic function of the limit distribution function $F(x)$, and the convergence to $\phi(t)$ is uniform with respect to t in every finite interval on the t-axis.*

Proof: *From the definition of a characteristic function we have*

$$\phi_n(t) = \int_{-\infty}^{\infty} e^{itx} dF_n(x), \quad \phi(t) = \int_{-\infty}^{\infty} e^{itx} dF(x)$$

Let $a < 0$ and $b > 0$ be continuity points of the distribution function $F(x)$. We have

$$\begin{aligned}\phi_n(t) &= \int_{-\infty}^{a} e^{itx} dF_n(x) + \int_{a}^{b} e^{itx} dF_n(x) + \int_{b}^{\infty} e^{itx} dF_n(x) \\ &= I_{n1} + I_{n2} + I_{n3} \\ \phi(t) &= \int_{-\infty}^{a} e^{itx} dF(x) + \int_{a}^{b} e^{itx} dF(x) + \int_{b}^{\infty} e^{itx} dF(x) \\ &= I_1 + I_2 + I_3.\end{aligned} \tag{4.27}$$

Consider the difference

$$I_{n2} - I_2 = \int_{a}^{b} e^{itx} dF_n(x) - \int_{a}^{b} e^{itx} dF(x)$$

Integrating by parts, we obtain

$$I_{n2} - I_2 = e^{itx} \left\{ [F_n(x)]_a^b - [F(x)]_a^b \right\} - it \int_{a}^{b} [F_n(x) - F(x)]\, e^{itx} dx$$

hence

$$|I_{n2} - I_2| \leqslant |F_n(b) - F(b)| + |F_n(a) - F(a)| + |t| \int_{a}^{b} |F_n(x) - F(x)|\, dx$$

Let $\varepsilon > 0$ be arbitrary. By the assumption of the theorem and by the fact that a and b are continuity points of $F(x)$, we obtain, for sufficiently large n,

$$|F_n(b) - F(b)| < \frac{\varepsilon}{9}, \quad |F_n(a) - F(a)| < \frac{\varepsilon}{9}$$

Furthermore, by the Lebesgue theorem on passage to the limit under the integral sign by the assumption of the theorem, and by the fact that $|F_n(x) - F(x)|$ is uniformly bounded in every interval, we obtain

$$\lim_{n\to\infty} \int_{a}^{b} |F_n(x) - F(x)|\, dx = \int_{a}^{b} \lim_{n\to\infty} |F_n(x) - F(x)|\, dx$$

Since the function under the integral sign on the right-hand side of the last formula is equal to zero except at most at a countable number of points, the integral under *consideration is equal to zero. Suppose now that t satisfies the inequality $T_1 < t < T_2$, where T_1 and T_2 are arbitrary fixed numbers, and let K be the greater of the numbers $|T_1|$ and $|T_2|$, that is, $K = \max(|T_1|, |T_2|)$. Then, for sufficiently large n and all t under consideration, we have*

$$|t| \int_a^b |F_n(x) - F(x)|\, dx \leqslant K \int_a^b |F_n(x) - F(x)|\, dx < \frac{\varepsilon}{9}$$

Thus we obtain

$$|I_{n2} - I_2| < \frac{\varepsilon}{3} \tag{4.28}$$

Now consider the difference

$$I_{n1} - I_1 = \int_{-\infty}^{a} e^{itx} dF_n(x) - \int_{-\infty}^{a} e^{itx} dF(x)$$

We have

$$|I_{n1} - I_1| \leqslant \int_{-\infty}^{a} dF_n(x) + \int_{-\infty}^{a} dF(x) = F_n(a) + F(a).$$

Thus, if a is sufficiently large in absolute value, then, by the assumption of the theorem and the continuity of $F(x)$ at a, we have, for sufficiently large n,

$$F_n(a) < \frac{\varepsilon}{6}, \quad F(a) < \frac{\varepsilon}{6}$$

Hence for all t and sufficiently large n,

$$|I_{n1} - I_1| < \frac{\varepsilon}{3} \tag{4.29}$$

Similarly, we obtain that

$$|I_{n3} - I_3| < \frac{\varepsilon}{3} \tag{4.30}$$

The theorem follows from formulas (4.27) to (4.30).

Theorem 4.4.2 *If the sequence of characteristic functions $\{\phi_n(t)\}$ converges at every point $t(-\infty < t < +\infty)$ to a function $\phi(t)$ continuous in some interval $|t| < \tau$, then the sequence $\{F_n(x)\}$ of corresponding distribution functions converges to the distribution function $F(x)$ which corresponds to the characteristic function $\phi(t)$.*

Proof: *In the proof we use the Helly theorem, which states that every sequence of distribution functions $\{F_n(x)\}$ contains a subsequence $\{F_{n_k}(x)\}$ convergent to some nondecreasing* *function $F(x)$. The function $F(x)$ can be changed at its discontinuity points so that it becomes continuous from the left. It does not, however, follow from the Helly theorem that $F(x)$ is a distribution function. Since $F(x)$ is the limit of distribution functions, we have $0 \leqslant F(x) \leqslant 1$, but we do not know whether $F(-\infty) = 0$ and $F(+\infty) = 1$. We show that the last relations are satisfied. Suppose that*

$$\alpha = F(+\infty) - F(-\infty) < 1 \tag{4.31}$$

Since $\phi_n(t) \to \phi(t)$ and $\phi_n(0) = 1$, we have $\phi(0) = 1$. By the assumption that the function $\phi(t)$ is continuous, it follows that in some neighborhood of the origin $t = 0$ it will differ little from 1 ; thus for sufficiently small τ we have the inequality

$$\frac{1}{2\tau}\left|\int_{-\tau}^{\tau}\phi(t)dt\right| > 1 - \frac{\varepsilon}{2} > \alpha + \frac{\varepsilon}{2} \tag{4.32}$$

where the number ε is chosen in such a way that $\alpha + \varepsilon < 1$. Since the subsequence $\{F_{n_k}(x)\}$ converges to $F(x)$, it follows from relation (4.31) that we can choose $a > 4/\varepsilon\tau$ such that a and $-a$ are continuity points of the limit distribution function, and a number K such that for $k > K$

$$\alpha_k = F_{n_k}(a) - F_{n_k}(-a) < \alpha + \frac{\varepsilon}{4}$$

On the other hand, since $\phi_n(t) \to \phi(t)$, it follows from relation (4.32) that for sufficiently large k the inequality

$$\frac{1}{2\tau}\left|\int_{-\tau}^{\tau}\phi_{n_k}(t)dt\right| > \alpha + \frac{\varepsilon}{2} \tag{4.33}$$

is satisfied. We show that this inequality is not satisfied. Indeed, we have

$$\int_{-\tau}^{\tau}\phi_{n_k}(t)dt = \int_{-\tau}^{\tau}\left[\int_{-\infty}^{+\infty}e^{itx}dF_{n_k}(x)\right]dt = \int_{-\infty}^{+\infty}\left[\int_{-\tau}^{\tau}e^{itx}dt\right]dF_{n_k}(x)$$

Since $|e^{itx}| = 1$, we obtain

$$\left|\int_{-\tau}^{\tau}e^{itx}dt\right| \leqslant 2\tau \tag{4.34}$$

Moreover,

$$\left|\int_{-\tau}^{\tau}e^{itx}dt\right| = \left|\left[\frac{e^{itx}}{ix}\right]_{-\tau}^{\tau}\right| = \frac{2}{|x|}|\sin\tau x| \leqslant \frac{2}{|x|} < \frac{2}{a} \quad \textit{for } |x| > a \tag{4.35}$$

Divide the whole axis into two parts, namely, into the interval $|x| \leqslant a$ and the complement *of this interval. We have*

$$
\begin{aligned}
&\left|\int_{-\infty}^{+\infty}\left(\int_{-\tau}^{\tau} e^{itx}dt\right) dF_{n_k}(x)\right| \\
&\quad \leqslant \left|\int_{|x|\leqslant a}\left(\int_{-\tau}^{\tau} e^{itx}dt\right) dF_{n_k}(x)\right| + \left|\int_{|x|>a}\left(\int_{-\tau}^{\tau} e^{itx}dt\right) dF_{n_k}(x)\right|
\end{aligned}
$$

Using inequality. (4.34) for $|x| \leqslant a$ and inequality (4.35)) for $|x| > a$, we obtain

$$
\begin{aligned}
\frac{1}{2\tau}\left| \int_{-\tau}^{\tau} \phi_{n_k}(t)dt \leqslant \right|\int_{|x|\leqslant a} dF_{n_k}(x)\right| + \frac{1}{a\tau}\left|\int_{|x|>a} \dot{d}F_{n_k}(x)\right| \\
\leqslant \alpha_k + \frac{1}{a\tau} \leqslant \alpha_k + \frac{\varepsilon}{4} < \alpha + \frac{\varepsilon}{4} + \frac{\varepsilon}{4} = \alpha + \frac{\varepsilon}{2}
\end{aligned}
\tag{4.36}
$$

The last inequality contradicts inequality (4.33). Hence the function $F(x)$ is a distribution function. From above theorem it follows that $\phi(t)$ is its characteristic function. We now prove that not only the subsequence $\{F_{n_k}(x)\}$, but the whole sequence $\{F_n(x)\}$ converges to $F(x)$. If this were not so there would be another subsequence $\{F_n(x)\}$ convergent to a limit function $\widetilde{F}(x)$ different from $F(x)$. The previous reasoning implies that $\tilde{F}(x)$ is a distribution function and from theorem it follows that $\widetilde{F}(x)$ has the same characteristic function as $F(x)$. Hence, by theorem, $\tilde{F}(x) \equiv F(x)$. Thus every subsequence of the sequence $\{F_n(x)\}$ contains a subsequence convergent to the same distribution function $F(x)$; hence the sequence $\{F_n(x)\}$ converges to $F(x)$. From theorems a and $6.6.1b$ we obtain immediately:

Theorem 4.4.3 *Levy-Cramer: Let $\{X_n\}$ $(n = 1, 2, \ldots)$ be a sequence of random variables and let $F_n(x)$ and $\phi_n(t)$ be respectively the distribution function and the characteristic function of X_n. Then the sequence $\{F_n(x)\}$ is convergent to a distribution function $F(x)$ if and only if the sequence $\{\phi_n(t)\}$ is convergent at every point $t(-\infty < t < +\infty)$ to a function $\phi(t)$ continuous in some neighborhood $|t| < \tau$ of the origin. The limit function - $\phi(t)$ is then the characteristic function of the limit distribution function $F(x)$ and the convergence $\phi_n(t) \to \phi(t)$ is uniform in every finite interval on the t-axis. We observe that theorem 4.1 remains true if we assume the continuity of the limit function $\phi(t)$ only at the point $t = 0$. We also observe that in the general case of theorem 4.1 we cannot replace the convergence at every point t in the interval $(-\infty, +\infty)$ by convergence in some interval on the t-axis containing the origin. If, however, all the random variables X_n are uniformly bounded from above (or below), then for the sequence $\{F_n(x)\}$ of distribution functions to converge to a distribution function $F(x)$, it is sufficient that in some interval $|t| < \tau$ the sequence $\{\phi_n(t)\}$ is convergent to a function $\phi(t)$ continuous at the origin. This theorem* *was proved by Zygmund. A We use theorem in proving the de Moivre-Laplace theorem. Denote by* $\{X_n\}$ *a sequence of random variables with the binomial distribution. For every* n *the random variable* X_n *can take on the values* $0, 1, \ldots, n$ *and its probability function is given by the formula*

$$P\left(X_n = r\right) = \binom{n}{r} p^r q^{n-r} \tag{4.37}$$

where $0 < p < 1$ *and* $q = 1 - p$*. As we know from formulas (4.4), we have*

$$E\left(X_n\right) = np, \quad D^2\left(X_n\right) = npq$$

Consider the sequence $\{Y_n\}$ *of standardized random variables*

$$Y_n = \frac{X_n - np}{\sqrt{npq}} \tag{4.38}$$

We shall prove a limit theorem called the de Moivre-Laplace theorem.

Theorem 4.4.4 *Let* $\{F_n(y)\}$ *be the sequence of distribution functions of the random variables* Y_n *defined, where the* X_n *have the binomial distribution given by formula (4.37). If* $0 < p < 1$*, then for every* y *we have the relation*

$$\lim_{n\to\infty} F_n(y) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{y} e^{-y^2/2} dy \tag{4.39}$$

Proof: *According to formula (3.3), the characteristic function* $\phi_x(t)$ *of* X_n *has the form*

$$\phi_x(t) = \left(q + pe^{it}\right)^n \tag{4.40}$$

Thus by equality (2.17) the characteristic function $\phi_y(t)$ *of the random variable* Y_n *is given by the formula*

$$\begin{aligned}\phi_y(t) &= \exp\left(-\frac{npit}{\sqrt{npq}}\right)\left[q + p\exp\left(\frac{it}{\sqrt{npq}}\right)\right]^n \\ &= \left[q\exp\left(-\frac{pit}{\sqrt{npq}}\right) + p\exp\left(\frac{qit}{\sqrt{npq}}\right)\right]^n\end{aligned} \tag{4.41}$$

Let us expand the function e^{iz} *in the neighborhood of* $z = 0$ *according to the Taylor* *formula for k terms with the remainder in the Peano form,*

$$e^{iz} = \sum_{j=0}^{k} \frac{(iz)^j}{j!} + o\left(z^k\right)$$

We obtain

$$p \exp\left(\frac{qit}{\sqrt{npq}}\right) = p + it\sqrt{pq/n} - \frac{qt^2}{2n} + o\left(\frac{t^2}{n}\right)$$
$$q \exp\left(-\frac{pit}{\sqrt{npq}}\right) = q - it\sqrt{pq/n} - \frac{pt^2}{2n} + o\left(\frac{t^2}{n}\right)$$

where for every t we have

$$\lim_{n\to\infty} no\left(\frac{t^2}{n}\right) = 0 \tag{4.42}$$

Substituting these expressions in formula (4.41) and considering the fact that $p + q = 1$, we obtain

$$\phi_y(t) = \left[1 - \frac{t^2}{2n} + o\left(\frac{t^2}{n}\right)\right]^n$$

Thus

$$\log \phi_y(t) = n \log\left[1 - \frac{t^2}{2n} + o\left(\frac{t^2}{n}\right)\right] = n\log(1+z)$$

We observe that for every fixed t for sufficiently large n, we have $|z| < 1$. Thus we can write

$$\log \phi_y(t) = -\frac{t^2}{2} + no\left(\frac{t^2}{n}\right)$$

By (4.42) we obtain

$$\lim_{n\to\infty} \log \phi_y(t) = -\frac{t^2}{2}$$

Hence

$$\lim_{n\to\infty} \phi_y(t) = e^{-t^2/2}$$

We have thus established that the sequence of characteristic functions $\phi_y(t)$ of the standardized random variables Y_n given by formula (4.38) onverges as $n \to \infty$ to the characteristic function of a random variable with a normal distribution whose distribution function is given by the right-hand side of formula (4.39). By theorem $6.6.1b$ we immediately obtain formula (4.39). We observe that the convergence in formula (4.39) holds for every y, since the distribution function of the normal distribution has no discontinuity points. The de Moivre-Laplace theorem is proved. Let y_1 and y_2 be two arbitrary points with $y_1 < y_2$.

From relation (4.39) it follows that

$$\lim_{n\to\infty} P\left(y_1 < Y_n < y_2\right) = \lim_{n\to\infty}\left[F_n\left(y_2\right) - F_n\left(y_1\right)\right] = \frac{1}{\sqrt{2\pi}}\int_{y_1}^{y_2} e^{-y^2/2}dy. \tag{4.43}$$

We shall rewrite the de Moivre-Laplace theorem in another form. By formula (4.38) we have

$$P\left(y_1 < Y_n < y_2\right) = P\left(y_1 < \frac{X_n - np}{\sqrt{npq}} < y_2\right)$$
$$= P\left(y_1\sqrt{npq} + np < X_n < y_2\sqrt{npq} + np\right)$$

Thus we obtain

$$\lim_{n\to\infty} P\left(y_1\sqrt{npq} + np < X_n < y_2\sqrt{npq} + np\right) = \frac{1}{\sqrt{2\pi}}\int_{y_1}^{y_2} e^{-y^2/2}dy$$

Let

$$x_1 = y_1\sqrt{npq} + np, \quad x_2 = y_2\sqrt{npq} + np \tag{4.44}$$

We can write formula (4.43) in the asymptotic form

$$P\left(x_1 < X_n < x_2\right) \cong \frac{1}{\sqrt{2\pi}}\int_{y_1}^{y_2} e^{-y^2/2}dy$$

where y_1 and y_2 are determined by (4.44). We say that the random variable X_n has an asymptotically normal distribution $N(np; \sqrt{npq})$. Replacing y_1 and y_2 with

$$y_1 + \frac{1}{2\sqrt{npq}} \quad \textit{and} \quad y_2 - \frac{1}{2\sqrt{npq}}$$

respectively, we get a somewhat better approximation.

Example 4.4.5 *We throw a coin $n = 100$ times. We assign the number 1 to the appearance of heads and the number 0 to the appearance of tails. The probability of each of these events is equal to $p = q = 0.5$. What is the probability that heads will appear more than 50 times and less than 60 times?. The random variable X_n can here take on values from*

0 to 100. We have

$$E\left(X_n\right)=50, \quad D^2\left(X_n\right)=25$$
$$P\left(50<X_n<60\right)=P\left(\frac{50-50}{5}<\frac{X_n-50}{5}<\frac{60-50}{5}\right)$$
$$=P\left(0<\frac{X_n-50}{5}<2\right) \cong \frac{1}{\sqrt{2\pi}}\int_{0.1}^{1.9} e^{-t^2/2}dt$$

From tables of the normal distribution we obtain that the value of this integral is 0.4315.
From the de Moivre-Laplace limit theorem we obtain an analogous theorem for the sequence of random variables

$$U_n=\frac{X_n}{n}$$

where X_n has the binomial distribution given by formula (4.37). Indeed, since $E\left(U_n\right)=p$ and $D^2\left(U_n\right)=pq/n$, we obtain the relation

$$Z_n=\frac{U_n-p}{\sqrt{pq/n}}=\frac{X_n-np}{\sqrt{npq}}=Y_n$$

where the random variables Y_n are defined by formula (4.38). Since the sequence $\{F_n(y)\}$ of distribution functions of Y_n satisfies formula (4.39), we obtain for the sequence $\{F_n(z)\}$ of the distribution functions of Z_n

$$\lim_{n\to\infty} F_n(z)=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{z} e^{-z^2/2}dz$$

Similarly, for every pair of constants z_1 and z_2, where $z_1<z_2$, we obtain the relation

$$\lim_{n\to\infty} P\left(z_1<\sqrt{n/pq}\left(U_n-p\right)<z_2\right)=\frac{1}{\sqrt{2\pi}}\int_{z_1}^{z_2} e^{-z^2/2}dz. \tag{4.45}$$

Letting

$$u_1=z_1\sqrt{pq/n}+p, \quad u_2=z_2\sqrt{pq/n}+p \tag{4.46}$$

We can rewrite formula (4.45) in the asymptotic form

$$P\left(u_1<U_n<u_2\right)\cong\frac{1}{\sqrt{2\pi}}\int_{z_1}^{z_2} e^{-z^2/2}dz \tag{4.47}$$

where z_1 and z_2 are determined. We say that the random variable U_n satisfying relation

(4.47) has an asymptotically normal distribution $N(p; \sqrt{pq/n})$.

Let Us Sum Up

Learners, in this section we have seen that definition of Levy-Cramer theorem and also given some theorems with Illustrations.

Check Your Progress

1. Levy's Cramér theorem applies to:
A. Dependent random variables with finite variance.
B. Independent and identically distributed random variables with finite mean and variance.
C. Independent and identically distributed random variables with finite mean only.
D. Dependent random variables with finite mean and variance.
2. The distribution of the sample mean $\overline{X}_n$ converges to a normal distribution according to Lévy's Cramér theorem if:
A. The variance of X_i is zero.
B. The moment-generating function of X_i is finite for all $t \in \mathbb{R}$.
C. The moment-generating function of X_i is finite for some interval $|t| \leq t_0$.
D. The random variables X_i are not identically distributed.

4.5 Lindeberg-Levy Theorem

The Bernoulli law of large numbers, proved in allows us to state only that for every $\varepsilon > 0$ the probability of the inequality

$$\left| \frac{X_n}{n} - p \right| > \varepsilon$$

tends to zero as $n \to \infty$. The limit theorem, which we have just proved, allows us (for large n) to compute approximately the probability that the random variable $X_n/n - p$

is contained in the interval

$$\left(z_1\sqrt{\frac{p(1-p)}{n}}, z_2\sqrt{\frac{p(1-p)}{n}}\right)$$

for arbitrary z_1 and z_2 $(z_1 < z_2)$.

Example 4.5.1 *A box contains a collection of IBM cards corresponding to the workers from some branch of industry. Of the workers* 20% *are minors and* 80% *adults. We select one IBM card in a random way and mark the age given on this card. Before choosing the next card, we return the first one to the box, so that the probability of selecting the card corresponding to a minor remains 0.2 . We observe* n *cards in this manner. What value should* n *have in order that the probability will be 0.95 that the frequency of cards corresponding to minors lies between 0.18 and 0.22 ? Denote the frequency of the appearance of the card corresponding to a minor by* U_n*. We then have*

$$E\left(U_n\right) = 0.2, \quad D^2\left(U_n\right) = \frac{0.16}{n}, \quad \sqrt{D^2\left(U_n\right)} = \frac{0.4}{\sqrt{n}}$$

Consider the probability

$$\begin{aligned} P\left(0.18 < U_n < 0.22\right) &= P\left(\frac{-0.02}{0.4/\sqrt{n}} < \frac{U_n - 0.2}{0.4/\sqrt{n}} < \frac{0.02}{0.4/\sqrt{n}}\right) \\ &= P\left(-0.05\sqrt{n} < \frac{U_n - 0.2}{0.4}\sqrt{n} < 0.05\sqrt{n}\right) \cong 0.95 \end{aligned}$$

By formula (4.47) we obtain

$$0.95 \cong \frac{1}{\sqrt{2\pi}}\int_{-0.05\sqrt{n}}^{0.05\sqrt{n}} e^{-z^2/2}dz$$

From tables of the normal distribution we obtain $0.05\sqrt{n} \cong 1.96$*; consequently* $n \cong 1537$*. The De Moivre-Laplace theorem is, as we shall see later, a particular case of a more general limit theorem, namely, the Lindeberg-Lévy theorem. Consider a sequence* $\{X_k\}$ $(k = 1, 2, \ldots)$ *of equally distributed, independent random variables whose moment of the second order exists. For every* k *denote*

$$E\left(X_k\right) = m, \quad D^2\left(X_k\right) = \sigma^2 \tag{4.48}$$

Consider the random variable Y_n defined by the formula

$$Y_n = X_1 + X_2 + \ldots + X_n \tag{4.49}$$

We have $E(Y_n) = nm$ and, by the independence of the X_n,

$$D^2(Y_n) = n\sigma^2$$

Let

$$Z_n = \frac{Y_n - mn}{\sigma\sqrt{n}} \tag{4.50}$$

We shall prove the following theorem.

Theorem 4.5.2 *If $X_1, X_2, \ldots$ are independent random variables with the same distribution, whose standard deviation $\sigma \neq 0$ exists, then the sequence $\{F_n(z)\}$ of distribution functions of the random variables Z_n, given by formulas (4.50) and (4.49), satisfies, for every z, the equality*

$$\lim_{n\to\infty} F_n(z) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{z} e^{-z^2/2} dz \tag{4.51}$$

Proof: *Let us write equality (4.50) in the form*

$$Z_n = \frac{1}{\sigma\sqrt{n}} \sum_{k=1}^{n} (X_k - m)$$

All the random variables $X_k - m$ have the same distribution, hence the same characteristic function $\phi_x(t)$. According to formulas (2.15) and (2.3) the characteristic function $\phi_z(t)$ of Z_n has the form

$$\phi_z(t) = \left[\phi_x\left(\frac{t}{\sigma\sqrt{n}}\right)\right]^n \tag{4.52}$$

We have assumed the existence of the first and second moments, and we have

$$E(X_k - m) = 0 \quad \textit{and} \quad D^2(X_k - m) = \sigma^2$$

Hence we can expand the function $\phi_x(t)$ in a neighborhood of the point $t = 0$ according to the MacLaurin formula as follows:

$$\phi_x(t) = 1 - \frac{1}{2}\sigma^2 t^2 + o\left(t^2\right) \tag{4.53}$$

Substituting expression (4.53) in formula (4.52), we obtain

$$\phi_z(t) = \left[1 - \frac{t^2}{2n} + o\left(\frac{t^2}{n}\right)\right]^n$$

where for every t we have

$$\lim_{n\to\infty} no\left(\frac{t^2}{n}\right) = 0 \tag{4.54}$$

Let

$$u = -\frac{t^2}{2n} + o\left(\frac{t^2}{n}\right)$$

We obtain

$$\log\phi_z(t) = n\log(1+u) = n\left[-\frac{t^2}{2n} + o\left(\frac{t^2}{n}\right)\right] = -\frac{t^2}{2} + no\left(\frac{t^2}{n}\right)$$

By relation (4.54) we obtain $\lim\log\phi_z(t) = -t^2/2$. *Hence*

$$\overset{n\to\infty}{\lim_{n\to\infty}} \phi_z(t) = e^{-t^2/2}$$

The expression $e^{-t^2/2}$ is the characteristic function of a random variable with the normal distribution. By theorem 6.6.1b *we obtain relation (4.51), which proves the theorem of Lindeberg-Lévy. et z_1 and z_2 be two arbitrary numbers with $z_1 < z_2$. By relation (4.51) we obtain*

$$(6.8.8) \lim_{n\to\infty} P\left(z_1 < Z_n < z_2\right) = \lim_{n\to\infty}\left[F_n\left(z_2\right) - F_n\left(z_1\right)\right] = \frac{1}{\sqrt{2\pi}}\int_{z_1}^{z_2} e^{-z^2/2}dz. \tag{4.55}$$

From formula (4.50) we obtain

$$\begin{aligned} P\left(z_1 < Z_n < z_2\right) &= P\left(z_1 < \frac{Y_n - nm}{\sigma\sqrt{n}} < z_2\right) \\ &= P\left(z_1\sigma\sqrt{n} + nm < Y_n < z_2\sigma\sqrt{n} + nm\right) \end{aligned}$$

Thus, we obtain from formula (4.55)

$$\lim_{n\to\infty} P\left(z_1\sigma\sqrt{n} + nm < Y_n < z_2\sigma\sqrt{n} + nm\right) = \frac{1}{\sqrt{2\pi}}\int_{z_1}^{z_2} e^{-z^2/2}dz \tag{4.56}$$

Let

$$y_1 = z_1\sigma\sqrt{n} + nm, \quad y_2 = z_2\sigma\sqrt{n} + nm \tag{4.57}$$

Now we can write formula (4.56) in the asymptotic form

$$P\left(y_1 < Y_n < y_2\right) \cong \frac{1}{\sqrt{2\pi}} \int_{z_1}^{z_2} e^{-z^2/2} dz$$

where z_1 and z_2 are determined by relations (4.57). Thus the random variable Y_n defined by formula (4.49) has an asymptotically normal distribution $N(mn; \sigma\sqrt{n})$. When a sum of random variables has an asymptotically normal distribution, we say that it satisfies the central limit theorem. Thus, for the sum Y_n under consideration, the central limit theorem holds.

Example 4.5.3 *Suppose that the random variables $\{X_k\}$ $(k = 1, 2, \ldots)$ are independent and each of them has the same two-point distribution, that is, for every k we have*

$$P\left(X_k = 1\right) = p, \quad P\left(X_k = 0\right) = 1 - p, \quad \textit{where} \quad 0 < p < 1.$$

Consider the random variable $Y_n = X_1 + X_2 + \ldots + X_n$. From the fact that $E\left(X_k\right) = p$ and $D^2\left(X_k\right) = pq$, we obtain by theorem that Y_n has an asymptotically normal distribution $N(np; \sqrt{npq})$. Since the random variable Y_n has the binomial distribution, this example is, strictly speaking, a new proof of the de Moivre-Laplace limit. theorem, which, as we see, is a particular case of the Lindeberg-Lévy theorem.

Example 4.5.4 *The random variables $X_n (n = 1, 2, \ldots)$ are independent and each of them has the Poisson distribution given by the formula*

$$P\left(X_n = r\right) = \frac{2^r}{r!} e^{-2} \quad (r = 0, 1, 2, \ldots)$$

Let us find the probability that the sum $Y_{100} = X_1 + X_2 + \ldots + X_{100}$ is greater than 190 and less than 210. The random variable Y_{100} has approximately the normal distribution $N(100;\ 10\sqrt{2}\ )$, since each of the random variables X_n has standard deviation $\sigma = \sqrt{2}$ and expected value $m = 2$. Thus we have

$$P\left(190 < Y_{100} < 210\right) = P\left(-0.707 < \frac{Y_{100} - 200}{10\sqrt{2}} < 0.707\right)$$

From the normal distribution tables we find that the required probability is 0.52. From *the Lindeberg-Levy theorem we obtain the following:*

Theorem 4.5.5 *Suppose that the random variables $X_1, X_2, \ldots$ are independent and have the same distribution with standard deviation $\sigma \neq 0$. Let the random variable U_n be defined by the formula*

$$U_n = \frac{X_1 + X_2 + \ldots + X_n}{n}$$

Furthermore, let $F_n(v)$ be the distribution function of the random variable V_n defined as

$$V_n = \frac{U_n - E\left(U_n\right)}{\sqrt{D^2\left(U_n\right)}}$$

Then the sequence $\{F_n(v)\}$ satisfies the relation

$$\lim_{n\to\infty} F_n(v) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{v} e^{-v^2/2} dv \tag{4.58}$$

Proof: *We have $E\left(U_n\right) = m$ and $D^2\left(U_n\right) = \sigma^2/n$. Hence*

$$V_n = \frac{\frac{1}{n}\sum_{k=1}^{n} X_k - m}{\sigma/\sqrt{n}} = \frac{\sum_{k=1}^{n} X_k - nm}{\sigma\sqrt{n}} = Z_n$$

where the random variables Z_n are defined by formula (4.50). Since the sequence $\{F_n(z)\}$ satisfies relation (4.51), the sequence $\{F_n(v)\}$ satisfies (4.58). Now let v_1 and v_2 be two arbitrary numbers with $v_1 < v_2$.

$$\lim_{n\to\infty} P\left(v_1 < V_n < v_2\right) = \frac{1}{\sqrt{2\pi}} \int_{v_1}^{v_2} e^{-v^2/2} dv \tag{4.59}$$

Let

$$u_1 = \frac{v_1\sigma}{\sqrt{n}} + m, \quad u_2 = \frac{v_2\sigma}{\sqrt{n}} + m \tag{4.60}$$

Formula (4.59) can be written in the asymptotic form

$$P\left(u_1 < U_n < u_2\right) \cong \frac{1}{\sqrt{2\pi}} \int_{v_1}^{v_2} e^{-v^2/2} dv$$

where v_1 and v_2 are determined from relations (4.60). Thus the random variable U_n has an asymptotically normal distribution $N(m; \sigma/\sqrt{n})$. In other words, the arithmetic mean of n independent random variables with the same, although arbitrary, distribution, where it is only assumed that the moment of the second order exists, has, for large n, an *asymptotically normal distribution.*

Example 4.5.6 *The random variables $X_1, X_2, \ldots$ are independent and have the uniform distribution defined by the density*

$$f(x) = \begin{cases} 1 & \textit{for } x \textit{ in the interval } [0,1] \\ 0 & \textit{for } x < 0 \textit{ and } x > 1 \end{cases}$$

By formulas (3.4) and (3.5) we have

$$m = \frac{1}{2}, \quad \sigma = \frac{1}{\sqrt{12}}$$

Consider the random variable

$$Y_n = \frac{X_1 + X_2 + \ldots + X_n}{n}$$

By theorem, the random variable Y_n has the asymptotically normal distribution $N\left(\frac{1}{2}; 1/\sqrt{12n}\right)$. For $n = 48$ compute the probability that Y_n will be smaller than 0.4 . We have [1]

$$\begin{aligned} P\left(Y_n < 0.4\right) &= P\left(\frac{Y_n - \frac{1}{2}}{1/\sqrt{576}} < \frac{0.4 - \frac{1}{2}}{1/\sqrt{576}}\right) \\ &= P\left(\frac{Y_n - \frac{1}{2}}{\frac{1}{24}} < -2.4\right) \cong \Phi(-2.4) \cong 0.0082 \end{aligned}$$

As we see, although the random variables $X_k (k = 1, 2, \ldots)$ have a uniform distribution in the interval $[0, 1]$, their arithmetic mean has, for large n, approximately a distribution in which values that are less than $m = 0.5$ by more than 0.1 appear extremely rarely.

Example 4.5.7 *The random variables $X_r (r = 1, 2, \ldots)$ are independent and have the same distribution. Each of them can take on the values $k = 0, 1, 2, \ldots, 9$ with the probabilities $P\left(X_r = k\right) = 0.1$ for every k. We have*

$$m = E\left(X_r\right) = \frac{1}{10} \sum_{k=0}^{9} k = 4.5$$

$$\sigma^2 = D^2\left(X_r\right) = \frac{1}{10} \sum_{k=0}^{9} (k - m)^2 = \frac{1}{10} \sum_{k=0}^{9} k^2 - m^2 = 28.50 - 20.25 = 8.25$$

Thus $\sigma = 2.87$. Consider the random variable

$$Y_{100} = \frac{X_1 + X_2 + \ldots + X_{100}}{100}$$

What is the probability that Y_{100} will exceed 5?. By theorem we know that Y_{100} has approximately the normal distribution $N(4.5; 2.87/\sqrt{100}$.) We obtain

$$P\left(Y_{100} > 5\right) = P\left(\frac{Y_{100} - 4.5}{0.287} > \frac{5 - 4.5}{0.287}\right) = P\left(\frac{Y_{100} - 4.5}{0.287} > 1.74\right)$$
$$\cong 1 - \Phi(1.74) \cong 0.041$$

We now show by an example that the arithmetic mean of n random variables with the same distribution may not have an asymptotically normal distribution, if their moment of the second order does not exist.

Example 4.5.8 *The random variables $X_k(k = 1, 2, \ldots)$ are independent and have the Cauchy distribution given by formula (3.3). Since the characteristic function of X_k has, for every k, the form*

$$\phi_k(t) = e^{-|t|}$$

The distribution function of the normal distribution $N(0; 1)$ is denoted by $\Phi(x)$. the characteristic function $\phi(t)$ of the random variable

$$Y_n = \frac{X_1 + X_2 + \ldots + X_n}{n}$$

takes the form

$$\phi(t) = e^{-n|t|/n} = e^{-|t|}$$

Hence for an arbitrary n the random variable Y_n has the Cauchy distribution. Thus Y_n does not have an asymptotically normal distribution. We notice, however, that a random variable with a Cauchy distribution does not have a standard deviation. Let the random variables $X_k(k = 1, 2, \therefore)$ satisfy the assumptions of theorem and let $E\left(X_k\right) = 0$. Consider for every n the partial sums

$$S_j = \sum_{k=1}^{j} X_k \quad (j = 1, 2, \ldots, n)$$

Erdos and $\mathrm{Kac}[1,2]$ *have found the limit distributions for the sequences of random variables*

$$\left\{\max_{1\leqslant j\leqslant n}\frac{S_j}{\sqrt{n}}\right\},\left\{\max_{1\leqslant j\leqslant n}\frac{|S_j|}{\sqrt{n}}\right\},\left\{\frac{1}{n^2}\sum_{j=1}^{n}S_j^2\right\},\left\{\frac{1}{n^{3/2}}\sum_{j=1}^{n}|S_j|\right\}$$

These definitions began a series of fruitful investigations concerning the limit distributions of a large class of functionals defined on the vectors $(S_1,\ldots,S_n)$*, even with much more general assumptions concerning the random variables* X_k *than those considered here. In the preceding section we discussed the limit distribution of the sum of independent random variables with the same distribution, and we established that if the variance of these random variables exists, their sum has an asymptotically normal distribution. However, the distribution of a sum of independent random variables may not converge to the normal distribution if the terms do not have the same distribution, even if all the random variables have standard deviations. We now prove the Lapunov theorem, which gives a sufficient condition for a sum of independent random variables to have a limiting normal distribution. Consider a sequence* $\{X_k\}$ *of independent random variables whose moments of the third order exist.*

Let Us Sum Up

Learners, in this section we have seen that definition of Lindeberg-Levy theorem with examples.

Check Your Progress

1. Which of the following conditions is necessary for the Lindeberg-Levy Central Limit Theorem to hold?
A. The random variables must be identically distributed.
B. The random variables must be independent.
C. The random variables must have finite mean and variance.
D. The random variables must be uniformly distributed.
2. In the context of the Lindeberg-Levy theorem, what does the notation $Z_n=\frac{S_n-n\mu}{\sqrt{n\sigma^2}}$ represent?
A. The sum of the random variables.
B. The standardized sum of the random variables.

C. The variance of the sum of the random variables.

D. The mean of the sum of the random variables.

4.6 Lapunov -Theorem

Let $\{X_k\}$ $(k = 1, 2, \ldots)$ be a sequence of independent random variables whose moments of the third order exist, and let $m_k, \sigma_k \neq 0, a_k$, and b_k denote the expected value, standard deviation, central moment of the third order, and the absolute central moment of the third order of X_k, respectively. Furthermore, let

$$B_n = \sqrt[3]{\sum_{k=1}^{n} b_k}, \quad C_n = \sqrt{\sum_{k=1}^{n} {\sigma_k}^2}$$

If the relation

$$\lim_{n\to\infty} \frac{B_n}{C_n} = 0$$

is satisfied, the sequence $\{F_n(z)\}$ of the distribution functions of the random variables Z_n, defined as

$$Z_n = \frac{\sum_{k=1}^{n} (X_k - m_k)}{C_n} \tag{4.61}$$

satisfies, for every z, the relation

$$\lim_{n\to\infty} F_n(z) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{z} e^{-z^2/2} dz \tag{4.62}$$

Proof: Let

$$Y_k = \frac{X_k - m_k}{C_n}$$

Let $\phi_{x_k}(t)$ denote the characteristic function of the random variable $X_k - m_k$. From the fact that $E(X_k - m_k) = 0$ and that the moments ${\sigma_k}^2$ and a_k exist, we obtain by formula (1.2) the expansion of $\phi_{x_k}(t)$ into the sum

$$\phi_{x_k}(t) = 1 - \frac{1}{2}\sigma_k^2 t^2 + \frac{1}{6} a_k (it)^3 + o\left(a_k t^3\right)$$

By formula (1.15), the characteristic function $\phi_{y_k}(t)$ of Y_k equals

$$\phi_{y_k}(t) = \phi_{x_k}\left(\frac{t}{C_n}\right) = 1 - \frac{{\sigma_k}^2 t^2}{2{C_n}^2} + \frac{a_k}{6{C_n}^3}(it)^3 + o\left(\frac{a_k t^3}{{C_n}^3}\right) = 1 + u_k$$

For every t we have

$$\lim_{n\to\infty}\left[o\left(\frac{a_k t^3}{C_n^3}\right) : \frac{a_k t^3}{{C_n}^3}\right] = 0 \tag{4.63}$$

Since by the Lapunov inequality we have $\sigma_k < \sqrt[3]{b_k}$, condition implies, for every t,

$$\lim_{n\to\infty}\left|\frac{-{\sigma_k}^2 t^2}{2{C_n}^2}\right| \leqslant \lim_{n\to\infty}\frac{\sqrt[3]{{b_k}^2}}{2{C_n}^2}t^2 \leqslant \lim_{n\to\infty}\frac{1}{2}\cdot\frac{{B_n}^2}{{C_n}^2}t^2 = 0 \tag{4.64}$$

Furthermore, we have

$$\lim_{n\to\infty}\left|\frac{a_k}{6C_n^3}(it)^3\right| \leqslant \lim_{n\to\infty}\frac{b_k}{6{C_n}^3}|t|^3 \leqslant \lim_{n\to\infty}\frac{{B_n}^3|t|^3}{6{C_n}^3} = 0 \tag{4.65}$$

It follows from relations (4.64) and (4.65) that

$$\lim_{n\to\infty} u_k = 0$$

and the convergence is uniform with respect to k. Hence for every t there exists a number $N = N(t)$ such that for $n > N$ and all $k \leqslant n$ we have the inequality $|u_k| < \frac{1}{2}$. Thus

$$\begin{aligned}\log\phi_{y_k}(t) &= \log(1+u_k) = u_k - \frac{1}{2}u_k^2 + \frac{1}{3}u_k^3 - \dots \\ &= u_k - \frac{1}{2}u_k^2\left(1 - \frac{2}{3}u_k + \frac{2}{4}u_k^2 - \dots\right) = u_k - \frac{1}{2}u_k^2 v_k\end{aligned}$$

We notice that

$$\begin{aligned}|v_k| \leqslant 1 + \tfrac{2}{3}|u_k| + \tfrac{2}{4}|u_k|^2 + \dots &< 1 + |u_k| + |u_k|^2 + \dots \\ &< 1 + \tfrac{1}{2} + \tfrac{1}{4} + \dots = 2\end{aligned}$$

Thus we can write

$$\log\phi_{y_k}(t) = u_k + \vartheta_k u_k^2 \tag{4.66}$$

where $\vartheta_k = -\frac{1}{2}v_k$, and $|\vartheta_k| < 1$. Denote by $\phi_z(t)$ the characteristic function of the random variable Z_n. By formula (1.3), we have $\phi_z(t) = \prod_{k=1}^{n} \phi_{y_k}(t)$. Hence

$$\log \phi_z(t) = \sum_{k=1}^{n} \log \phi_{y_k}(t)$$

By equality (4.66) we obtain

$$\log \phi_z(t) = \sum_{k=1}^{n} \left(u_k + \vartheta_k u_k^2\right) \tag{4.67}$$

Next, we have

$$\sum_{k=1}^{n} u_k = -\frac{t^2}{2} + \sum_{k=1}^{n} \frac{a_k (it)^3}{6C_n^3} + \sum_{k=1}^{n} o\left(\frac{a_k t^3}{C_n^3}\right) \tag{4.68}$$

We notice that for every t

$$\lim_{n\to\infty} \left|\sum_{k=1}^{n} \frac{a_k (it)^3}{6{C_n}^3}\right| \leqslant \lim_{n\to\infty} \sum_{k=1}^{n} \frac{b_k |t|^3}{6{C_n}^3} = \lim_{n\to\infty} \frac{{B_n}^3 |t|^3}{6{C_n}^3} = 0. \tag{4.69}$$

Hence, by formula (4.63) we obtain

$$\lim_{n\to\infty} \sum_{k=1}^{n} o\left(\frac{a_k t^3}{C_n^3}\right) = \lim_{n\to\infty} \sum_{k=1}^{n} \left\{\frac{a_k t^3}{6C_n^3} \cdot \left[o\left(\frac{a_k t^3}{C_n^3}\right) : \frac{a_k t^3}{6C_n^3}\right]\right\} = 0 \tag{4.70}$$

From formula (4.68), because of formulas (4.69) and (4.70), it follows that for every t we have

$$\lim_{n\to\infty} \sum_{k=1}^{n} u_k = -\frac{t^2}{2} \tag{4.71}$$

We now find the limit of the sum

$$\sum_{k=1}^{n} u_k^2 = \sum_{k=1}^{n} \left[\frac{-{\sigma_k}^2 t^2}{2{C_n}^2} + \frac{a_k (it)^3}{6{C_n}^3} + o\left(\frac{a_k t}{{C_n}^3}\right)\right]^2$$

By the Lapunov inequality and condition we obtain

$$\lim_{n\to\infty} \sum_{k=1}^{n} \frac{{\sigma_k}^4 t^4}{4{C_n}^4} \leqslant \lim_{n\to\infty} \sum_{k=1}^{n} \frac{\sqrt[3]{{b_k}^4}}{4{C_n}^4} t^4 = \lim_{n\to\infty} \sum_{k=1}^{n} \frac{b_k \sqrt[3]{b_k}}{4{C_n}^3 C_n} t^4$$

$$\leqslant \lim_{n\to\infty} \sum_{k=1}^{n} \frac{b_k}{4{C_n}^3} t^4 = \lim_{n\to\infty} \frac{{B_n}^3}{4{C_n}^3} t^4 = 0, \tag{4.72}$$

$\lim_{n\to\infty} \left| \sum_{k=1}^{n} \left[\frac{a_k (it)^3}{6C_n^3} \right]^2 \right| \leqslant \lim_{n\to\infty} \sum_{k=1}^{n} \left| \frac{a_k t^3}{6C_n^3} \right|^2$

$$\leqslant \lim_{n\to\infty} \sum_{k=1}^{n} \frac{{b_k}^2 t^6}{36{C_n}^6} \leqslant \lim_{n\to\infty} \frac{{B_n}^6 t^6}{36{C_n}^6} = 0 \tag{4.73}$$

Taking formula (4.63) into consideration, we obtain

$$\lim_{n\to\infty} \sum_{k=1}^{n} \left[o\left(\frac{a_k t^3}{6C_n^3} \right) \right]^3 = 0 \tag{4.74}$$

Similarly, we obtain

$$\begin{aligned} \lim_{n\to\infty} \left| \sum_{k=1}^{n-1} \sum_{j=k+1}^{n} \frac{\sigma_k^2 a_j i^3 t^5}{6{C_n}^5} \right| &= 0 \\ \lim_{n\to\infty} \left| \sum_{k=1}^{n-1} \sum_{j=k+1}^{n} \frac{t^2 \sigma_k^2}{C_n^2} o\left(\frac{a_j t^3}{C_n^3} \right) \right| &= 0 \\ \lim_{n\to\infty} \left| \sum_{k=1}^{n-1} \sum_{j=k+1}^{n} \frac{a_k i^3 t^3}{3C_n^3} o\left(\frac{a_j t^3}{C_n^3} \right) \right| &= 0 \end{aligned} \tag{4.75}$$

From formulas (4.72) to (4.75) and the fact that for every $k \leqslant n$ we have $|\vartheta_k| < 1$, we obtain

$$\lim_{n\to\infty} \sum_{k=1}^{n} \vartheta_k u_k^2 = 0 \tag{4.76}$$

Using (4.71) and (4.76) we obtain from formula (4.67) that for every t the relation

$$\lim_{n\to\infty} \log \phi_z(t) = -\frac{t^2}{2}$$

holds. Hence

$$\lim_{n\to\infty} \phi_x(t) = e^{-t^2/2}$$

By the last relation and by theorem, we obtain formula (4.62), which proves the Lapunov theorem. For another proof of Lapunov's theorem. The Lapunov theorem only gives a sufficient condition for relation (4.62). We shall now present without proof the theorem of LindebergFeller, giving a necessary and sufficient condition. T

Let Us Sum Up

Learners, in this section we have seen that definition of Lapunov theorem and also given theorems and Illustrations.

Check Your Progress

1. Lapinov's theorem provides conditions under which:
A. The sum of random variables converges to a Poisson distribution.
B. The sum of random variables converges to a normal distribution.
C. The sum of random variables converges to a uniform distribution.
D. The sum of random variables does not converge to any distribution.
2. In Lapinov's theorem, the moment-generating function $M(t)$ is bounded by:
A. e^{Kt^2}
B. e^{Kt}
C. e^{Kt^2} where K is a constant.
D. e^t

4.7 Lindeberg-Feller Theorem

Let $\{X_k\}$ $(k = 1, 2, \ldots)$ be a sequence of independent random variables whose variances exist, and let $G_k(x), m_k$, and $\sigma_k \neq 0$ denote, respectively, the distribution function, the expected value and the standard deviation of the random variable X_k, and let $F_n(z)$ denote the distribution function of the standardized random variable Z_n given by formula (4.61). Then the relations

$$\lim_{n\to\infty} \max_{1\leq k\leq n} \frac{\sigma_k}{C_n} = 0, \quad \lim_{n\to\infty} F_n(z) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{z} e^{-z^2/2} dz$$

hold if and only if, for every $\varepsilon > 0$,

$$\lim_{n\to\infty} \frac{1}{C_n^2} \sum_{k=1}^{n} \int_{|x-m_k|>\varepsilon C_n} (x - m_k)^2 \, dG_k(x) = 0 \tag{6.9.20}$$

If all the X_k are of the continuous type and $g_k(x)$ is the density of X_k, then condition (4.77) takes the form

$$\lim_{n\to\infty} \frac{1}{{C_n}^2} \sum_{k=1}^{n} \int_{|x-m_k|>\varepsilon C_n} (x-m_k)^2 g_k(x)dx = 0 \tag{4.78}$$

If, however, all the X_k are of the discrete type with jump points x_{kl} and jumps $p_{kl}(l = 1, 2, \ldots)$, formula (4.77) takes the form

$$\lim_{n\to\infty} \frac{1}{C_n^2} \sum_{k=1}^{n} \sum_{|x_{kl}-m_k|>\varepsilon C_n} (x_{kl}-m_k)^2 p_{kl} = 0 \tag{4.79}$$

From the theorem of Lindeberg-Feller follows this theorem.

Theorem 4.7.1 *Let $\{X_k\}$ $(k = 1, 2, \ldots)$ be a sequence of independent, uniformly bounded random variables, that is, there exists a constant $a > 0$ such that for every k*

$$P(|X_k| \leqslant a) = 1 \tag{4.80}$$

and suppose that $D^2(X_k) \neq 0$ for every k. Then a necessary and sufficient condition for relation (4.62) to hold is

$$\lim_{n\to\infty} C_n^2 = \infty \tag{4.81}$$

Proof: *Suppose that (4.81) is satisfied. From formula (4.80) it follows that the random variables $X_k - m_k$ are uniformly bounded. Hence for every $\varepsilon > 0$ we can find an N such that for $n > N$ we have*

$$P(|X_k - m_k| < \varepsilon C_n; k = 1, 2, \ldots, n) = 1$$

Formula (4.77) follows immediately. Suppose now that (4.62) holds, and (4.81) does not. Then there exists a $C < \infty$ such that $\lim {C_n}^2 = C^2$. From the last relation, and from formulas (4.62) and (4.61), it follows that $\sum_{k=1}^{\infty}(X_k - m_k$) has the normal distribution $N(0; C)$. Let

$$U = (X_2 - m_2) + (X_3 - m_3) + \ldots$$

The random variables $X_1 - m_1$ and U are independent, and their sum has a normal distribution. By the Cramér theorem both $X_1 - m_1$ and U have normal distributions. However, by hypothesis (4.80), the random variable $X_1 - m_1$ does not have a normal *distribution. Hence (4.61) is not satisfied, and the theorem is proved. In particular, it follows from this theorem that if the random variable* $Y_n = \sum_{k=1}^{n} X_k$ *has the generalized binomial distribution, that is, if the probability function of* X_k *is given by the formulas* $P(X_k = 1) = p_k, P(X_k = 0) = q_k = 1 - p_k (k = 1, 2, \ldots)$*, then the divergence of the series* $\Sigma p_k q_k$ *is a necessary and sufficient condition for* Y_k *to have the asymptotically normal distribution*

$$N\left(\sum_{k=1}^{n} p_k; \sqrt{\sum_{k=1}^{n} p_k q_k}\right)$$

Example 4.7.2 *At a construction site there are lots of bricks from five different factories. Judging by previous experience, the quality of bricks from different factories differs and the fraction of defective items is not the same for all lots. The production of the* i *th factory is characterized by the number* p_i*, giving the fraction of good bricks. The values of* p_i *are the following:*

$$p_1 = 0.95, \quad p_2 = 0.90, \quad p_3 = 0.98, \quad p_4 = 0.92, \quad p_5 = 0.96$$

Since the lots are very large, we assume it is certain that the defectiveness of a lot produced by the i *th factory is exactly* $1 - p_i (i = 1, \ldots, 5)$. *The probability of choosing a good brick from a given lot is thus* p_i. *We select 20 bricks at random from each lot. Since each lot contains many bricks, and the drawing of 20 bricks does not change practically the probability of selecting a good brick, we may assume that this probability is constant while drawing bricks and hence equals* p_i. *After checking the quality of all 100 selected bricks, it turned out that 11 of them were defective. This result created some doubts as to whether the assumptions about the numbers* p_i *were not too optimistic. The mathematical model of this example is the following. We have 100 independent random variables* X_k *and each of them can take on two values; 1 when a good brick is selected and 0 when a defective one is selected. These random variables are divided into five groups. The* i *th group consists of those random variables which take on the value 1 with probability* p_i. *Let us form the random variable*

$$Y_{100} = X_1 + \ldots + X_{20} + X_{21} + \ldots + X_{40} + X_{41} + \ldots + X_{60} + X_{61} + \ldots + X_{80} + X_{81} + \ldots + X_{100}$$

This is a random variable with a generalized binomial distribution. We have

$$\begin{aligned} E\left(Y_{100}\right) = & 20 \cdot 0.95 + 20 \cdot 0.90 + 20 \cdot 0.98 + 20 \cdot 0.92 + 20 \cdot 0.96 = 94.20 \\ D^2\left(Y_{100}\right) = & 20 \cdot 0.05 \cdot 0.95 + 20 \cdot 0.10 \cdot 0.90 + 20 \cdot 0.02 \cdot 0.98 \\ & + 20 \cdot 0.08 \cdot 0.92 + 20 \cdot 0.04 \cdot 0.96 = 5.382 \\ \sigma = & 2.32 \end{aligned}$$

Before we apply the central limit theorem, we must examine the result obtained above which gives the divergence of the series $\Sigma p_k q_k$ *as a necessary and sufficient condition for the convergence of the generalized binomial distribution to the normal distribution. If, however, this series is convergent, then* $p_k q_k \to 0$ *as* $k \to \infty$. *Hence* $\min\left(p_k; 1 - p_k\right) \to 0$. *Thus the sequence* $\{p_k\}$ *must contain a subsequence convergent either to zero or to one. In the language of this example, this would mean that the series* $\Sigma p_k q_k$ *would converge if the bricks produced contained very often (theoretically an infinite number of times) either only good or only defective items. However, many years of practice in the production of bricks show that this is not true and thus the series* $\Sigma p_k q_k$ *is not convergent. Thus we can apply the central limit theorem. According to this theorem, the random variable* Y_{100} *has approximately the normal distribution* $N(94.2; 2.32)$. *Thus we have*

$$P\left(Y_{100} \leqslant 89\right) = P\left(\frac{Y_{100} - 94.2}{2.32} \leqslant -2.25\right) \cong \Phi(-2.25)$$

From tables of the normal distribution we find that $\Phi(-2.25)$ *is rather small, about 0.01 . In such cases we are inclined to accept the conclusion that our assumptions about the* p_i *were too optimistic. In this example we have touched on questions which will be systematically and exhaustively considered. This example was given to show that the central limit theorem is not only a beautiful mathematical achievement but can also be applied to the solution of many practical problems. We see how important a role the normal distribution plays in probability theory and its applications. However, the theorem which we now present shows that under rather general assumptions a sequence of distribution function of sums of independent random variables may converge to a limit distribution function different from the normal. Consider a sequence* $\{Y_n\}$ $(n = 1, 2, \ldots)$ *of random variables, where for every* n, Y_n *is the sum of* n *independent random variables*

X_{nk} $(k = 1, 2, \ldots, n)$,

$$Y_n = \sum_{k=1}^{n} X_{nk} \tag{4.82}$$

These sums are more general than the sums considered in this section, where we have

$$X_{nk} = X_k \quad (n = 1, 2, \ldots; k = 1, 2, \ldots, n)$$

We restrict ourselves to the case when, for every n*, the random variables* $X_{nk}(k = 1, 2, \ldots, n)$ *have the same distribution* [1] *given by the probability function*

$$P\left(X_{nk} = x_l\right) = p_{nl} \quad (l = 1, 2, \ldots, r) \tag{4.83}$$

where

$$0 < p_{nl} < 1, \sum_{l=1}^{r} p_{nl} = 1$$

and $r(r \geqslant 2)$ *is some natural number.*

Theorem 4.7.3 *Let* Y_n *be defined by formula (4.82) and let* $X_{nk}(k = 1, 2, \ldots, n)$ *be independent and have the distribution defined by formula (4.83). Let* $F_n(z)$ *be the distribution function of the random cariable* Z_n *defined as*

$$Z_n = \frac{Y_n - E\left(Y_n\right)}{\sqrt{D^2\left(Y_n\right)}}$$

Then: I. If

$$\lim_{n \rightarrow \infty} n\left(p_{n1} p_{n2} + p_{n1} p_{n3} + \ldots + p_{n, r-1} p_{nr}\right) = \infty \tag{4.84}$$

[1] *The case when the* X_{nk} *do not have the same distribution for all* k *was considered by Kubik. the sequence* $\{F_n(z)\}$ *satisfies the relation*

$$\lim_{n \rightarrow \infty} F_n(z) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{z} e^{-z^2/2} dz$$

II. If the limits (finite or infinite)

$$\lim_{n \rightarrow \infty} p_{nl} \quad \textit{and} \quad \lim_{n \rightarrow \infty} n p_{nl} \quad (l = 1, 2, \ldots, r)$$

exist, and the relation

$$\lim_{n\to\infty} n\left(p_{n1}p_{n2} + p_{n1}p_{n3} + \ldots + p_{n,r-1}p_{nr}\right) = \lambda \tag{4.85}$$

where $\lambda > 0$, holds, then the sequence $\{F_n(z)\}$ converges to the distribution function of a random variable which is a linear combination of $s(1 \leqslant s \leqslant r-1)$ independent random variables, each having a Poisson distribution. We notice that in theorem we have dealt with sums of the form (4.82). Indeed, let Y_n be the number of successes in n trials in the Bernoulli scheme and let the probability of success p_n be a function of n satisfying the relation

$$\lim_{n\to\infty} np_n = \lambda \tag{4.86}$$

where $\lambda > 0$. Then we can write

$$Y_n = \sum_{k=1}^{n} X_{nk}$$

where X_{nk} is the number of successes (equal to 0 or 1) in the k th trial $(k = 1, 2, \ldots, n)$; thus the X_{nk} are independent and have the same distribution given by the formulas

$$P\left(X_{nk} = 1\right) = p_{n1} = p_n, \quad P\left(X_{nk} = 0\right) = p_{n2} = 1 - p_n$$

and we have $r = 2$. From formula (4.86) follow the relations

$$\lim_{n\to\infty} p_{n1} = 0, \quad \lim_{n\to\infty} p_{n2} = 1$$

$$\lim_{n\to\infty} np_{n1} = \lambda, \quad \lim_{n\to\infty} np_{n2} = \infty, \quad \lim_{n\to\infty} np_{n1}p_{n2} = \lambda$$

where $\lambda > 0$. All the assumptions of assertion II of theorem are satisfied; thus the sequence $\{F_n(z)\}$, where $F_n(z)$ is the distribution function of the random variable

$$Z_n = \frac{Y_n - np_n}{\sqrt{np_n\left(1 - p_n\right)}}$$

converges as $n \to \infty$ to the distribution function of a Poisson random variable with the parameter λ. This is the integral Poisson theorem, whereas theorem is the local Poisson theorem.

4.8 Let Us Sum Up

Learners, in this section we have seen that definition of Lindeberg-Feller theorem and also given examples.

Check Your Progress

1. In the Lindeberg-Feller theorem, the Lindeberg condition involves:
A. The boundedness of the moment-generating function.
B. The convergence of the sample variance to zero.
C. The contribution of large deviations to the variance of the sum.
D. The identically distributed nature of the random variables.
2. The sample variance S_n^2 in the Lindeberg-Feller theorem is given by:
A. $\frac{1}{n}\sum_{i=1}^{n}(X_i - \mathbb{E}[X_i])^2$
B. $\frac{1}{n}\sum_{i=1}^{n}\sigma_i^2$
C. $\frac{1}{n}\sum_{i=1}^{n}(X_i - \mu_i)^2$
D. $\frac{1}{n}\sum_{i=1}^{n}\text{Var}(X_i)$

4.9 Unit Summary

The fourth unit content on limit theorems are stochastic convergence, Bernoulli's law of large numbers, the convergence of a sequence of distribution functions, The Levy-Cramer theorem, De Moivre-Laplace theorem, Lindeberg-Levy theorem and Lapunov theorem.

Glossary

1. The $X_n \xrightarrow{p} X$ is converges in probability p.
2. The $X_n \xrightarrow{d} X$ is diverges in probability X.
3. If $X_n \xrightarrow{a.s.} X$ is converges almost surely X.
4. $F_n(x) \to F(x)$ uniformly for all $x \in \mathbb{R}$.

5. $F_n(x) \to F(x)$ at every point $x \in \mathbb{R}$.

6. $F_n(x) \to F(x)$ at all continuity points of F.

Self-Assessment Questions

Short Answers: (5 Marks)

1. Prove that the sequence $\{F_n(x)\}$ converges to the distribution function $F(x)$ if and only if the relation

$$\lim_{n \to \infty} F_n(x) = F(x)$$

holds for all points x in a set S which is everywhere dense in the interval ($-\infty$, $+\infty$).

2. Show that if the sequence of characteristic functions $\{\phi_n(t)\}$ converges to the characteristic function $\phi(t)$ and $t_n \to t_0$, then $\phi_n(t_n) \to \phi(t_0)$.

3. Prove that if $X_1, X_2, \ldots$ are independent random variables with the same distribution, whose standard deviation $\sigma \neq 0$ exists, then the sequence $\{F_n(z)\}$ of distribution functions of the random variables Z_n, given by formulas and satisfies, for every z, the equality

$$\lim_{n \to \infty} F_n(z) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{z} e^{-z^2/2} dz$$

Long Answers: (8 Marks)

1. Let $Y_n = \sum_{k=1}^{k_n} X_{nk}$, where the random variables X_{nk} ($k = 1, \ldots, k_n$) are independent for each n and have the probability functions given by the formula $P(X_{nk} = x_{nkl}) = p_{nkl}$, where $\sum_{l=1}^{n} p_{nkl} = 1 (n = 1, 2, \ldots,\ k = 1, 2, \ldots, k_n, l = 1, 2, \ldots, r, r \geqslant 2$). Assume that (a) the X_{nk} are asymptotically constant, that is, for every $\varepsilon > 0$ we have
$\lim_{n \to \infty} \max_{1 \leq k \leq k_n} P(|X_{nk} - m_{nk}| > \varepsilon) = 0$, where m_{nk} is the median of X_{nk},
(b) $\lim_{n \to \infty} \max_{1 \leq k \leq k_n} z_{nkl} = \lim_{n \to \infty} \min_{1 \leq k \leq k_n} z_{nkl} = x_l (l = 1, \ldots, r)$,
where $z_{nkl} = x_{nk,l+1} - x_{nkl}$. Find the class of all possible limit distribution functions of sequences $\{F_n(y)\}$ of distribution functions of $Y_n - A_n$ for arbitrary sequences of constants $\{A_n\}$.

2. Let us denote by ${m_k}^{(n)}$ the moment of order k of the random variable X_n with the distribution function $F_n(x)$. Prove that if, for $k = 1, 2, \ldots$, the finite limits

$$m_k = \lim_{n \to \infty} m_k^{(n)}$$

exist and, moreover, these limits uniquely determine a distribution function $F(x)$ the sequence $\{F_n(x)\}$ converges to $F(x)$.

3. Let the random variables $X_1, X_2, X_3, \ldots$ satisfy all the assumptions of the Lindeberg-Levy theorem, and suppose that the moment $E\left|X_i\right|^3$ exists. Then the relation

$$\left|F_n(z) - \Phi(z)\right| \leqslant c \frac{E\left|X_i\right|^3}{\sigma^3} \frac{1}{\sqrt{n}}$$

holds, where c is a constant.

4. The random variables $X_i (i = 1, 2, \ldots)$ are independent and have the same probability distribution, given by the formulas

$$P\left(X_i = 0\right) = P\left(X_i = 3\right) = P\left(X_i = 7\right) = P\left(X_i = 12\right) = \frac{1}{4}.$$

Check whether for this sequence the local limit theorem of Gnedenko holds.

5. The random variable X has the Poisson distribution with the parameter λ. Let $u_r = P(X = r)(r = 0, 1, \ldots), t = (r - \lambda)/\sqrt{\lambda}$ and

$$v_r = \frac{1}{\sqrt{2\pi\lambda}} \exp\left(-\frac{t^2}{2}\right)$$
$$w_r = v_r \left(1 - \frac{t}{2\sqrt{\lambda}} + \frac{t^2}{6\sqrt{\lambda}}\right)$$

Applying Stirling's formula, show that if $\lambda \to \infty$ and $r \to \infty$ in such a way that t remains bounded in absolute value, then for any $\varepsilon > 0$

$$\lim_{\lambda \to \infty} \left[\lambda^{1-\varepsilon}\left(u_r - v_r\right)\right] = 0,$$

$$\lim_{\lambda \to \infty} \left[\lambda^{3/2-\varepsilon}\left(u_r - w_r\right)\right] = 0.$$

Exercises

1. The random variable X has the Poisson distribution with the parameter λ. Let $u_r = P(X = r)(r = 0, 1, \ldots), t = (r - \lambda)/\sqrt{\lambda}$ and

$$v_r = \frac{1}{\sqrt{2\pi\lambda}} \exp\left(-\frac{t^2}{2}\right)$$
$$w_r = v_r \left(1 - \frac{t}{2\sqrt{\lambda}} + \frac{t^2}{6\sqrt{\lambda}}\right)$$

Applying Stirling's formula, show that if $\lambda \to \infty$ and $r \to \infty$ in such a way that t remains bounded in absolute value, then for any $\varepsilon > 0$

$$\lim_{\lambda \to \infty} \left[\lambda^{1-\varepsilon}\left(u_r - v_r\right)\right] = 0,$$
$$\lim_{\lambda \to \infty} \left[\lambda^{3/2-\varepsilon}\left(u_r - w_r\right)\right] = 0.$$

2. The distribution functions $F(x)$ and $G(x)$ are said to be of the same type if there exist constants $a > 0$ and b such that for every x

$$G(x) = F(ax + b).$$

Prove that if the sequence of distribution functions $\{F_n(x)\}$ converges as $n \to \infty$ to a nondegenerate distribution function $F(x)$, and if $F_n\left(a_n x + b_n\right)$ converges to a nondegenerate distribution function $G(x)$, then $G(x)$ is of the same type as $F(x)$.

3. Prove that $\{X_k\}\,(k = 1, 2, 3, \ldots)$ be a sequence of independent and identically distributed random variables. If, for some constants a and $A_n (n = 1, 2, 3, \ldots)$, the relation

$$\lim_{n \to \infty} P\left(\frac{1}{a\sqrt{n}} \sum_{k=1}^{n} X_k - A_n < z\right) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{z} \exp\left(-\frac{z^2}{2}\right) dz$$

holds for any z, then the variance σ^2 of X_k exists. If this is so, then $a = \sigma$ and A_n may be chosen to equal $\frac{\sqrt{n}}{\sigma} E\left(X_{li}\ \right)$

Answers to Check Your Progress

Session (Modulo) 4.1

1. A. Convergence in probability implies convergence in distribution.

2. B. $X_n \xrightarrow{p} X$

Session (Modulo) 4.2

1. C. $F_n(x) \to F(x)$ at all continuity points of F.

2. C. In distribution to F

Session (Modulo) 4.3

1. A. f must be continuous on $[a, b]$ and g must be of bounded variation on $[a, b]$.

2. A. $\int_a^b f(x)\,dg(x) = f(b)g(b) - f(a)g(a) - \int_a^b g(x)\,df(x)$

Session (Modulo) 4.4

1. B. Independent and identically distributed random variables with finite mean and variance.

2. C. The moment-generating function of X_i is finite for some interval $|t| \leq t_0$.

Session (Modulo) 4.5

1. B. The random variables must be independent.

2. B. The standardized sum of the random variables.

Session (Modulo) 4.6

1. B. The sum of random variables converges to a normal distribution.

2. C. e^{Kt^2} where K is a constant.

Session (Modulo) 4.7

1. C. The contribution of large deviations to the variance of the sum.

2. B. $\frac{1}{n}\sum_{i=1}^{n} \sigma_i^2$

References

1. M. Fisz, Probability Theory and Mathematical Statistics, John Wiley and sons, New Your, Third Edition, 1963.

Suggested Readings

1. T. Veerarajan, Fundamentals of Mathematical Statistics, Yesdee Publishing, 2017.

2. P.R. Vittal, Mathematical Statistics, Margham Publications, 2002.

3. T. Veerarajan, Probability, Statistics and Random Processes, Mc Graw Hill Education (India) Private Limited, Third Edition, 2015.
4. R.S.N. Pillai and V. Bagavathi, Statistics, Sultan Chand and Sons, 2010.
5. Singaravelu.A, S. Sivasubramanian, Probability and Random Processes , Meenakshi Agency 2008,
6. DN Elhance, Veena Elhance and BM Aggarwal, Fundamentals of Statistics, Kitab Mahal.
7. S. C. Gupta and V. K. Kapoor, Fundamentals of Mathematical Statistics, Sultan Chand and Sons, 2008.

Unit 5

Markov Chain

Objective

This course aims to teach the students about Markov chain with homogeneous Markov chains and transition matrix, Ergodic theorem and random variables forming a homogeneous Markov chain.

5.1 Introduction of Markov Chain

In this section we have mainly considered independent random events and independent random variables. In fact, in the applications of probability theory we can often assume that the random events or random variables under consideration are independent. However, there are many problems in physics, engineering, and other areas of applications of probability theory where the assumption of independence is not satisfied, not even approximately. Therefore, the investigation of dependent random events and dependent random variables is an important problem in probability theory. But to abandon the assumption of independence creates serious complications in the reasoning and in the proofs. It is a great achievement of Markov that in the investigation of dependent events he distinguished a scheme of experiments, now called the scheme of events forming a Markov chain, which can be considered as the simplest generalization of the scheme of independent trials. Markov's investigations have become the starting point for the development of a new and important branch of probability theory, the theory of Markov stochastic processes.

Let Us Sum Up

Learners, in this section we have seen that introduction of Markov chains with example.

Check Your Progress

1. In a Markov chain, the transition matrix P represents:
A. The probability of transitioning between different states in one step.
B. The probability distribution of the initial state.
C. The probability of transitioning from the current state to the next state over multiple steps.
D. The probability distribution of the stationary state.
2. A Markov chain is said to be irreducible if:
A. It is possible to return to the starting state in a finite number of steps.
B. There is a positive probability of reaching any state from any other state.
C. The chain has a stationary distribution.
D. The chain is periodic.

5.2 Homogeneous Markov Chains

We assume that all the conditional probabilities appearing in this and the following chapters are defined. Imagine that we are given a sequence of experiments and as a result of each experiment there can be one and only one event from a finite or countable set of pairwise exclusive events $E_1, E_2, E_3, \ldots$ We call these events states. When the event E_j occurs we say that the system passes into the state E_j. We use the symbol $E_j^{(n)}$ to denote that at the nth trial the system passes into the state E_j; the symbol $E_j^{(0)}$ denotes that the initial state was E_j. Next we denote by $p_{ij}^{(n)}$ the conditional probability that at the nth trial the system passes into the state E_j, provided that after the $(n-1)$-st trial it was in the state E_j, that is,

$$p_{ij}^{(n)} = P\left(E_j^{(n)} \mid E_i^{(n-1)}\right)$$

Definition 5.2.1 *We say that a sequence of trials forms a Markov chain if for any* $i, j, n = 1, 2, 3, \ldots$ *the equalities*

$$p_{ij}^{(n)} = P\left(E_j^{(n)} \mid E_i^{(n-1)}\right) \tag{5.1}$$
$$= P\left(E_j^{(n)} \mid E_i^{(n-1)} E_{i_{n-2}}^{(n-2)} \ldots E_{i_1}^{(1)} E_{i_0}^{(0)}\right)$$

are satisfied for arbitrary $E_{i_{n-2}}^{(n-2)}, \ldots, E_{i_1}^{(1)}, E_{i_0}^{(0)}$.

Definition 5.2.2 *We say that a sequence of trials forms a homogeneous Markov chain, if for $i, j = 1, 2, 3, \ldots$ the probability $p_{ij}^{(n)}$ is independent of n, that is,*

$$p_{ij}^{(n)} = p_{ij} \quad (n = 1, 2, \ldots) \tag{5.2}$$

The probability p_{ij} is called the transition probability from the state E_i to the state E_j in one trial. We also use the time terminology, that is, we consider the trials as performed at every unit of.time and, instead of saying that at the nth trial the system passes from the state E_i to the state E_j, we say that this transition is performed at the moment $t = n$. Besides this, we shall assume that at the initial moment, that is, at $t = 0$, the system may be in the state E_i with probability $P(E_i)$. In this terminology p_{ij} is the transition probability from the state E_i to the state E_j in a unit of time. By formulas (5.1), (5.2), and (1.7) we obtain the following formula for the probability of the product of states ($E_{i_0} E_{i_1} \ldots E_{i_n}$) in n successive trials of a homogeneous Markov chain:

$$P\left(E_{i_0} E_{i_1} \ldots E_{i_n}\right) = P\left(E_{i_0}\right) P\left(E_{i_1} \mid E_{i_0}\right) \ldots P\left(E_{i_n} \mid E_{i_{n-1}}\right) \tag{5.3}$$
$$= P\left(E_{i_0}\right) p_{i_0 i_1} \ldots p_{i_{n-1} i_n}.$$

The reader will notice an essential difference between the last formula and formula (1.7). It follows from formula (5.3) that the probability of every product of states is given if we know all the transition probabilities p_{ij} and all the probabilities $P(E_{i_0})$ of the initial states.

Let Us Sum Up

Learners, in this section we have seen that the definition of homogeneous Markov chains.

Check Your Progress

1. What characterizes a Markov chain as homogeneous?

A. The transition probabilities are constant over time.

B. The transition probabilities vary with time but are stationary.

C. The chain has a finite number of states.

D. The chain exhibits periodic behavior.

2. For a homogeneous Markov chain with transition matrix P, what does the matrix P^n denote?

A. The matrix of state probabilities after n steps.

B. The matrix of initial state distributions after n steps.

C. The matrix of transition probabilities after n steps.

D. The matrix of cumulative transition probabilities up to n steps.

5.3 Transition Matrix

The matrix with the transition probabilities p_{ij} as elements is called the transition matrix. This matrix is denoted by $\mathbf{M}_1$,

$$\mathbf{M}_1 = \begin{bmatrix} p_{11} & p_{12} & p_{13} \cdots \\ p_{21} & p_{22} & p_{23} \cdots \\ p_{31} & p_{32} & p_{33} \cdots \\ \cdots & \cdots & \cdots \end{bmatrix}$$

We observe that all the elements p_{ij}, being probabilities, are non-negative. Suppose that the system is in the state E_i. The event that as a result of the experiment the system either remains in the state E_i or passes to any of the states E_j, where $i \neq j$, is the sure event. Since the events E_j are pairwise exclusive, for $i = 1, 2, 3, \ldots$, we obtain the formula

$$P\left[\left(\sum_j E_j\right) \mid E_i\right] = \sum_j p_{ij} = 1 \tag{5.4}$$

Thus the sum of the terms in each row of the matrix $\mathbf{M}_1$ equals one. However, the sum of the terms in a column need not be one.

Example 5.3.1 *Consider a sequence of trials in the Bernoulli scheme. Here we have two* *states E_1 and E_2, and in each experiment*

$$p_{11} = p_{21} = p, \quad p_{12} = p_{22} = q$$

Thus the transition matrix is of the form

$$\mathbf{M}_1 = \begin{bmatrix} p & q \\ p & q \end{bmatrix}$$

It is easy to verify that in an independent sequence of trials the rows of the transition matrix are always identical.

Example 5.3.2 *Here we consider the random walk with absorbing barriers. It is a model of certain phenomena which often appear in physics. A particle may be at one of the points $1, 2, 3, \ldots, s$ on the x-axis. It will remain forever, with probability one, at the point $x = 1$ if it arrives there at some moment t. The same is true for the point $x = s$. The points 1 and s are called absorbing barriers. If at the moment t the particle comes to the point $x = i$, where $2 \leqslant i \leqslant s-1$, then during the next unit of time the particle will pass to the point $i+1$ with probability p and to the point $i-1$ with probability $q = 1-p$. Here we have a homogeneous Markov chain with s states, where the state E_i occurs if the particle has the coordinate $x = i$. In fact, the probability of passing from the state E_i to the state E_j at the moment t does not depend on the previous path of the particle and does not depend on t but only on the state at the moment t. The transition probabilities are*

$$p_{11} = p_{ss} = 1,$$

and for $2 \leqslant i \leqslant s-1$

$$p_{ij} = \begin{cases} p & \text{for } j = i+1 \\ q = 1-p & \text{for } j = i-1 \\ 0 & \text{otherwise} \end{cases}$$

Thus the transition matrix has the form

$$\mathbf{M}_1 = \begin{bmatrix} 1 & 0 & 0 & 0 & \dots & 0 & 0 & 0 \\ q & 0 & p & 0 & \dots & 0 & 0 & 0 \\ 0 & q & 0 & p & \dots & 0 & 0 & 0 \\ \dots & \dots & . & \dots & \dots & \dots & . & \\ 0 & 0 & 0 & 0 & \dots & q & 0 & p \\ 0 & 0 & 0 & 0 & \dots & 0 & 0 & 1 \end{bmatrix}$$

We now give an example due to Malecot of the application of Markov chains to genetics.

Example 5.3.3 *In the genetics based on Mendel's laws we assume that inherited characteristics depend on the genes. Genes always appear in pairs. In the simplest case, which we consider here, every gene may be of one of two forms, A or a. If both genes of the organism being considered are of type A, we say that the organism is of genotype AA; if both genes are of type a we say that it is of genotype aa; finally, if one gene is of type A and the other of type a we say that the organism is of genotype Aa. Furthermore, we assume that the reproductive cells, or gametes, have only one gene; thus the gametes of an organism of genotype AA or aa have the gene A or a, respectively, whereas the gametes of an organism of genotype Aa may have the gene A or a with equal probability. An offspring receives one gene from each parent under the conditions of the Bernoulli scheme. This should be understood as follows: consider the set of all genes of all organisms belonging to the generation of parents of a given offspring as the population from which two genes are drawn at randomunder the conditions of the Bernoulli scheme. Similarly, the genotype structure of N offspring is a result of $2N$ such drawings from the set of genes under consideration. Suppose, now, that the population under consideration consists of N elements in each generation. This may be achieved by an appropriate selection of organisms in each generation. Thus we have $2N$ genes in each generation. If in some generation $i (0 \leqslant i \leqslant 2N)$ of the genes are of the form A, we say that the generation is in the state E_i. From the assumed reproduction scheme it follows that we have here a homogeneous Markov chain with $2N+1$ possible states: $E_0, E_1, \dots E_{2N}$. The probability of passing from the state E_i in some generation to the state E_j in the next generation is given by the formula*

$$p_{ij} = \binom{2N}{j} \left(\frac{i}{2N}\right)^j \left(1 - \frac{i}{2N}\right)^{2N-j}$$

We observe that the states E_0 and $E_{2,V}$ are the absorbing barriers. Indeed, if in some generation the population is in one of these states it will remain there forever; if, for instance, all the organisms are of the genotype AA, no offspring can have the gene a.

Example 5.3.4 *Here we consider a model of a random walk without absorbing barriers, having a countable number of states. The set of states is the set of all non-negative integers and the transition probabilities are given by the formulas*

$$p_{ij} = \begin{cases} p & p_{11} = q = 1 - p, \\ q & \text{for } i = 1, 2, 3, \ldots; \qquad\qquad j = i + 1, \\ 0 & \text{for the remaining pairs } (i, j). \end{cases}$$

The number 0 is a reflecting barrier. The transition matrix $\mathbf{M}_1$ is of the form

$$\mathbf{M}_1 = \begin{bmatrix} q & p & 0 & 0 & 0 & 0 & \ldots \\ q & 0 & p & 0 & 0 & 0 & \ldots \\ 0 & q & 0 & p & 0 & 0 & \ldots \\ \ldots & \ldots & \ldots & \ldots & \ldots & & \end{bmatrix}$$

Example 5.3.5 *Let us now return to the Polya scheme. We use the notation. We have two states, state E_1 consists of drawing a white ball as state E_2 consists of drawing a black ball, and the initial probabilities are p_1 b/N and $p_2 = 1 - p_1 = c/N$, respectively. The probability of passing from t state E_1 in the first drawing to the state E_1 in the second drawing is (b+. $(N + s)$. However, the probability of choosing a white ball in the third drawi if in the second drawing a white ball was drawn, equals $(b + 2s)/(N + 2$ provided that in the first drawing we obtained the state E_1, and it equals (b+($N + 2s$)) provided in the first drawing we obtained the state E_2. Thus t sequence of trials in the Polya scheme is not a Markov chain. We can, however, obtain a Markov chain in the Pólya scheme if we define t states in another way, namely, if we agree to say that after n drawings the syste is in the state $E_i(i = 0, 1, 2, \ldots, n)$, if i is the number of white balls obtained n drawings. Then at the $(n + 1)$ st trial the system may remain in the state E_i pass to the state E_{i+1}, according to whether in the $(n + 1)$ st trial a black or white ball was drawn. These transition probabilities depend only on the state the system after the nth trial and are independent of the results of the first n trials. However, these probabilities depend on the number of trials and we ha here a nonhomogeneous Markov chain with the transition* *probabilities $p_{ij}^{(n-)}$ given by the formula*

$$p_{ij}^{(n+1)} = \begin{cases} \frac{c+(n-i)s}{N+ns} & \textit{for } j = i \\ \frac{b+is}{N+ns} & \textit{for } j \neq i+1 \\ 0 & \textit{for } j \neq i, i+1 \end{cases}$$

We denote by $p_{ij}(n)$ the probability of passing in n trials from t *state E_i to the state E_j in a homogeneous Markov chain. Sometimes call it the probability of transition in n steps. We show how to compu the probabilities $p_{ij}(n)$ from the probabilities p_{ij}. Let us start by compt ing $p_{ij}(2)$. We observe that the event A of passing from the state E_i the state E_j in two trials is the union of the pairwise exclusive events where A_k occurs if and only if the system passes from the state E_i to E_k the first step and from E_k to E_j in the second step. Thus for every pi (i, j) we have*

$$p_{ij}(2) = \sum_k p_{ik} p_{kj} \tag{5.5}$$

where the summation is extended over all possible states.

Let Us Sum Up

Learners, in this section we have seen that definition of transition matrix also given theorems and Illustrations.

Check Your Progress

1.Which of the following is true for a transition matrix P?

A. Each entry p_{ij} can be negative.

B. Each row of P must sum to 1.

C. Each column of P must sum to 1.

D. The matrix P is not required to be square.

2. The Chapman-Kolmogorov equation is used to:

A. Calculate the stationary distribution of the Markov chain.

B. Relate the n-step and m-step transition probabilities.

C. Find the eigenvalues of the transition matrix.

D. Compute the long-term behavior of the Markov chain.

5.4 The Ergodic Theorem

In an analogous way we find the formulas

$$p_{ij}(n) = \sum_k p_{ik}(m) p_{kj}(n-m) \tag{5.6}$$

where $n = 2, 3, 4, \dots$ and m is an integer satisfying the condition $1 \leqslant m < n$. Equation (5.6) plays a basic role in the theory of homogeneous Markov chains and is called the Markov equation. The matrix whose elements are the transition probabilities $p_{ij}(n)$ is called the matrix of transition in n steps and is denoted by the symbol $\mathbf{M}_n$. It is easy to find the relation between the matrices $\mathbf{M}_n$ and $\mathbf{M}_1$. Let us first find the relation between the matrices $\mathbf{M}_1$ and $\mathbf{M}_2$. From formula it follows that the element of matrix $\mathbf{M}_2$ at the intersection of the i th row and j th column is the sum of products of the elements of the i th row by the j th column of $\mathbf{M}_1$. Thus, according to the rule of multiplication of matrices, we obtain

$$\mathbf{M}_2 = \mathbf{M}_1^2$$

By induction and formula (5.6), we have

$$\mathbf{M}_n = \mathbf{M_1}^n \tag{5.7}$$

We start this section with a classification of states of Markov chains; this will allow us to interpret the assumptions of the ergodic theorem. This classification was introduced by Kolmogorov.

Definition 5.4.1 *The state E_i is called unintrinsic if there exists a state E_j and an integer k such that $p_{ij}(k) > 0$ and $p_{ji}(m) = 0$ for $m = 1, 2, 3, \dots$*

Definition 5.4.2 *The state E_i is called intrinsic if, for every state E_j, the existence of an integer k_j such that $p_{ij}(k_j) > 0$ implies the existence of an integer m_i such that $p_{j2}(m_i) > 0$.*

Definition 5.4.3 *The intrinsic state E_i is called periodic if there exists an integer $d > 1$ such that $p_{i2}(n) = 0$ for n not a multiple of d. We observe, however, that we cannot pass from the state E_1 to E_s, nor can we pass from E_s^- to E_1, despite the fact that both states are intrinsic. This remark gives rise to the following definition.*

Definition 5.4.4 *The set W of intrinsic states forms one class if for every pair of intrinsic states E_i and E_j of W there exists an integer m_{ij} such that $p_{ij}\,(m_{ij}) > 0$. We now discuss the ergodic theorem. This theorem tells how the probabilities $p_{ij}(n)$ behave as $n \to \infty$. In other words, it explains what influence the initial state E_i has on the probability $p_{ij}(n)$ after a large number of steps n. We know that a condition for the convergence $p_{ij}(n) \to p_j$ for a homogeneous Markov chain, where the limits p_j are independent of i, that is, are independent of the initial state E_i. The theorem given here does not give a complete solution to this problem; in particular, it does not consider Markov chains with a countable number of states.*

Theorem 5.4.5 *Let $\mathrm{M}_1 = [p_{ij}]$ be the matrix of one step transition probabilities in a homogeneous Markov chain with a finite number of states $E_1, \dots, E_s$. If there exists an integer r such that the terms $p_{ij}(r)$ of the matrix M_r satisfy the relation*

$$\min_{1 \leqslant i \leqslant s} p_{ij}(r) = \delta > 0 \tag{5.8}$$

in $s_1\,(s_1 \geqslant 1)$ columns, then the equalities

$$\lim_{n \to \infty} p_{ij}(n) = p_j \quad (j = 1, 2, \dots, s) \tag{5.9}$$

are satisfied, and $p_j \geqslant \delta$ for those j for which relation (5.8) holds. Moreover, $\sum_j p_j = 1$ and

$$|p_{ij}(n) - p_j| \leqslant (1 - s_1\delta)^{n/r-1}\,. \tag{5.10}$$

As we see, one of the assumptions of this theorem requires that the elements $p_{ij}(r)$ of at least one column of the matrix M_r be positive. The above theorem is a modification of the theorem of Markov, which requires that for some integer r all the elements of the matrix M_r be positive. Then in the assertion of the theorem we have $p_j > 0 (j = 1,\ 2, \dots, s)$. Theorem is called the ergodic theorem and the limit probabilities p_j are called the ergodic probabilities. The explanation of this name is given at the end of this section.

Example 5.4.6 *Let us return to example and suppose, for simplicity, that $s = 3$. Then* *we have. Let us compute*

$$\mathbf{M}_1 = \begin{bmatrix} 1 & 0 & 0 \\ q & 0 & p \\ 0 & 0 & 1 \end{bmatrix}$$

$$\mathbf{M}_2 = {\mathbf{M}_1}^2 = \begin{bmatrix} 1 & 0 & 0 \\ q & 0 & p \\ 0 & 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} 1 & 0 & 0 \\ q & 0 & p \\ 0 & 0 & 1 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ q & 0 & p \\ 0 & 0 & 1 \end{bmatrix} = \mathbf{M}_1$$

In general, we have

$$\mathbf{M}_n = \mathbf{M}_1$$

Thus the assumptions of the ergodic theorem are not satisfied. There does not exist an r *such that the matrix* $\mathbf{M}_r$ *has at least one column of positive elements* $p_{ij}(r)$*. It is obvious that the assertion of above theorem is not satisfied either, since* $p_{11}(n) = 1$ *so that* $\lim_{n\to\infty} p_{11}(n) = 1$*, while* $\lim_{n\to\infty} p_{21}(n) = q$ *and* $\lim_{n\to\infty} p_{31}(n) = 0$*. The irregularity of this Markov chain is caused by the existence of two intrinsic states* E_1 *and* E_3 *such that passage from one to the other is impossible; thus the set of states does not form one class.*

Example 5.4.7 *Consider a homogeneous Markov chain with four states* E_1*,* E_2, E_3, E_4 *and the transition matrix We obtain here*

$$\mathbf{M}_1 = \begin{bmatrix} 0 & 0 & \frac{1}{2} & \frac{1}{2} \\ 0 & 0 & \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} & 0 & 0 \\ \frac{1}{2} & \frac{1}{2} & 0 & 0 \end{bmatrix}$$

$$\mathbf{M}_2 = \begin{bmatrix} \frac{1}{2} & \frac{1}{2} & 0 & 0 \\ \frac{1}{2} & \frac{1}{2} & 0 & 0 \\ 0 & 0 & \frac{1}{2} & \frac{1}{2} \\ 0 & 0 & \frac{1}{2} & \frac{1}{2} \end{bmatrix}$$

Generally, for $k = 1, 2, 3, \ldots$*, we have*

$$\mathbf{M}_{2k+1} = \mathbf{M}_1, \quad \mathbf{M}_{2k} = \mathbf{M}_2$$

Thus neither the assumption nor the assertion of theorem is satisfied. The reader will notice the periodicity of this Markov chain. All the states are intrinsic; but they are periodic, so that, for instance, the system may return from the state E_1 to the state E_1 only in an even number of steps. This periodicity causes the observed irregularity, as a result of which the ergodic theorem is not satisfied.

Example 5.4.8 *Let us return to example and suppose that the number of states is 3 and the matrix $\mathbf{M}_1$ has the form*

$$\mathbf{M}_1 = \begin{bmatrix} q & p & 0 \\ q & 0 & p \\ 0 & q & p \end{bmatrix}$$

Then

$$\mathbf{M}_2 = \begin{bmatrix} q^2 + pq & qp & p^2 \\ q^2 & 2qp & p^2 \\ q^2 & pq & qp + p^2 \end{bmatrix}$$

Thus the assumptions of theorem are satisfied. We observe that all three states are intrinsic, nonperiodic, and form one class. We show later how to compute the ergodic probabilities.

Example 5.4.9 *Let us modify example, such a way that the transition matrix takes the form*

$$\mathbf{M}_1 = \begin{bmatrix} 1 & 0 & 0 \\ q & 0 & p \\ 0 & q & p \end{bmatrix}$$

In this example, the state E_1 is an absorbing barrier, and the state E_3 is a reflecting barrier. We have Thus the assumptions of theorem are satisfied. It is easy to verify that the state E_1 is intrinsic and not periodic and the remaining two states are unintrinsic. Later we show that the limit probabilities p_2 and p_3 are zero. These examples suggest, and it can be shown that this is true, that if in a homogeneous Markov chain with a finite number of states all the intrinsic states are nonperiodic and form one class, then the assumptions of theorem are satisfied. However, the possibility that some states are unintrinsic is not excluded. But if all the states are intrinsic, nonperiodic, and form one class, then there exists an r such that all the elements $p_{ij}(r)$ of the matrix $\mathbf{M}_r$ are positive, hence greater than some $\delta > 0$, since there are only a finite number of them. Let us mention here that

Kaucky and Konêcny have given necessary and sufficient conditions for the ergodicity of homogeneous Markov chains; their conditions are expressed in terms of eigenvalues of the matrix $\mathbf{M}_1$. *We now give the proof of theorem.*

Proof: *For* $v = 1, 2, 3, \ldots$, *denote*

$$b_j(v) = \min_{1 \leqslant i \leqslant s} p_{ij}(v), \quad B_j(v) = \max_{1 \leqslant i \leqslant s} p_{ij}(v) \tag{5.11}$$

Considering formula (5.6) for $v = 1, 2, 3, \ldots$, *we obtain*

$$\begin{aligned} b_j(v+1) = \min_{1 \leqslant i \leqslant s} p_{ij}(v+1) &= \min_{1 \leqslant i \leqslant s} \sum_{k=1}^{s} p_{ik} p_{kj}(v) \\ &\geqslant \min_{1 \leqslant i \leqslant s} \sum_{k=1}^{s} p_{ik} b_j(v) = b_j(v) \end{aligned}$$

Hence

$$b_j(v+1) \geqslant b_j(v) \tag{5.12}$$

Similarly,

$$B_j(v+1) \leqslant B_j(v) \tag{5.13}$$

From formulas (5.12) and (5.13) we obtain

$$b_j(1) \leqslant b_j(2) \leqslant \ldots \leqslant B_j(2) \leqslant B_j(1) \tag{5.14}$$

Let r *and* $\sum_k^+$ *and* $\sum_k^-$ *denote, respectively, the sums extended over those* k *for which* $p_{ik}(r) \geqslant p_{mk}(r)$ *and* $p_{ik}(r) < p_{mk}(r)$. *Then*

$$\sum_k^{+} \left[p_{ik}(r) - p_{mk}(r)\right] + \sum_k^{-} \left[p_{ik}(r) - p_{mk}(r)\right] = 0 \tag{5.15}$$

Suppose that $n > r$. *Consider the difference*

$$\begin{aligned} B_j(n) - b_j(n) &= \max_{1 \leqslant i \leqslant s} p_{ij}(n) - \min_{1 \leqslant m \leqslant s} p_{mj}(n) \\ &= \max_{1 \leqslant i \leqslant s} \sum_{k=1}^{s} p_{ik}(r) p_{kj}(n-r) - \min_{1 \leqslant m \leqslant s} \sum_{k=1}^{s} p_{mk}(r) p_{kj}(n-r) \\ &= \max_{1 \leqslant i, m \leqslant s} \sum_{k=1}^{s} \left[p_{ik}(r) - p_{mk}(r)\right] p_{kj}(n-r) \end{aligned}$$

$\leqslant \max_{1\leqslant i,m\leqslant s}\left\{\sum_k^+ [p_{ik}(r)-p_{mk}(r)]\,B_j(n-r)+\sum_k^- [p_{ik}(r)-p_{mk}(r)]\,b_j(n-r)\right\}$. *Hence by formula (5.15) and (5.16)*

$$B_j(n)-b_j(n)\leqslant \max_{1\leqslant i,m\leqslant s}\sum_k^+ [p_{ik}(r)-p_{mk}(r)]\,[B_j(n-r)-b_j(n-r)]$$
$$=[B_j(n-r)-b_j(n-r)]\max_{1\leqslant i,m\leqslant s}\sum_k^+ [p_{ik}(r)-p_{mk}(r)]$$

Suppose that relation (5.8) holds for w terms of the sum $\sum_k^+$. Obviously, $w\leqslant s_1$, where s_1 is the number of columns for which (5.8) is satisfied. Thus

$$-\sum_k^+ p_{mk}(r)\leqslant -w\delta$$

Next, since for s_1-w terms of the sum $\sum_k -$ relation (5.8) is also satisfied, we have

$$\sum_k^+ p_{ik}(r)+(s_1-w)\,\delta\leqslant 1$$

Finally,

$$\sum_k^+ [p_{ik}(r)-p_{mk}(r)]\leqslant 1-(s_1-w)\,\delta-w\delta=1-s_1\delta \tag{5.16}$$

From formulas (5.15) and (5.16) follows the inequality

$$B_j(n)-b_j(n)\leqslant (1-s_1\delta)\,[B_j(n-r)-b_j(n-r)]$$

Similarly, for $n>2r$

$$B_j(n)-b_j(n)\leqslant (1-s_1\delta)^2\,[B_j(n-2r)-b_j(n-2r)]$$

Repeating this procedure $[n/r]^1$ times, we obtain

$$B_j(n)-b_j(n)\leqslant (1-s_1\delta)^{[n/r]}\left\{B_j\left(n-\left[\frac{n}{r}\right]r\right)-b_j\left(n-\left[\frac{n}{r}\right]r\right)\right\} \tag{5.17}$$

We observe that from (5.17) and from the fact that $\delta > 0, s_1 \geqslant 1$ follows the inequality

$$0 \leqslant 1 - s_1\delta < 1$$

Thus from formula (5.14) follows the existence of the limits of $\{b_j(n)\}$ and $\{B_j(n)\}$, and limits are equal. Therefore,

$$\lim_{n\to\infty} \max_{1\leqslant i\leqslant s} p_{ij}(n) = \lim_{n\to\infty} \min_{1\leqslant i\leqslant s} p_{ij}(n) = p_j \tag{5.18}$$

which proves formula (5.9). Next, it is obvious that for those j for which relation (5.8) is satisfied, we have $p_j \geqslant \delta$. The equality $\sum_j p_j = 1$
is also obvious. It remains to prove relation (5.10). In fact, by formulas (5.14) and (5.15) we obtain

$$|p_{ij}(n) - p_j| \leqslant B_j(n) - b_j(n) \leqslant (1 - s_1\delta)^{n/r-1}$$

which completes the proof of theorem. We now show how to calculate the ergodic probabilities p_j if they are known to exist. By formula (5.6) we obtain

$$p_{ij}(n) = \sum_{k=1}^{s} p_{ik}(n-1)p_{kj}$$

Thus, if the ergodic probabilities p_j exist, then after passage to the limit as $n \to \infty$ on both sides of the last inequality we have

$$p_j = \sum_{k=1}^{s} p_k p_{kj} \quad (j = 1, 2, \ldots, s) \tag{5.19}$$

[1] *The symbol $[A]$ denotes here the greatest integer not exceeding A. From these equations and from the relation*

$$\sum_{j=1}^{s} p_j = 1$$

we can determine the probabilities p_j.

Example 5.4.10 *Let us return to example and calculate the ergodic probabilities p_j.*

Formula (5.19) gives us three linear equations

$$p_1 = p_1 q + p_2 q$$

$$p_2 = p_1 p + p_3 q$$

$$p_3 = p_2 p + p_3 p$$

Hence

$$p_2 = \frac{p}{q} p_1$$
$$p_3 = \left(\frac{p}{q}\right)^2 p_1$$

Since $p_1 + p_2 + p_3 = 1$*, we obtain*

$$p_1 \left[1 + \frac{p}{q} + \left(\frac{p}{q}\right)^2\right] = 1$$

Thus if $p = q = \frac{1}{2}$*, then* $p_1 = p_2 = p_3 = \frac{1}{3}$*, and thus in the limit each state has the same probability. If* $p \neq q$*, then*

$$p_j = \frac{1 - (p/q)}{1 - (p/q)^3} \left(\frac{p}{q}\right)^{j-1} \qquad (j = 1, 2, 3)$$

If $p > q$*, then the probabilities* p_j *increase with the number* j *of the state; if* $p < q$ *they decrease. These results agree with our intuition. Thus if* $p/q = 2$ *we have and if* $p/q = \frac{1}{2}$

$$p_1 = \frac{1}{7}, \quad p_2 = \frac{2}{7}, \quad p_3 = \frac{4}{7},$$

$$p_1 = \frac{4}{7}, \quad p_2 = \frac{2}{7}, \quad p_3 = \frac{1}{7}$$

We observe that all three ergodic probabilities are positive. This is because all states are intrinsic. In a Markov chain with a countable number of states ergodic probabilities of intrinsic states may be equal zero.

Example 5.4.11 *Let us calculate the limit probabilities. By formula (5.19) we have*

$$p_1 = p_1 + p_2 q$$

$$p_2 = p_3 q$$

$$p_3 = (p_2 + p_3)\, p$$

Since $p_1 + p_2 + p_3 = 1$*, we obtain* $p_1 = 1, p_2 = p_3 = 0$*. As has been mentioned, this is because the states* E_2 *and* E_3 *are unintrinsic. The notion of ergodicity and conditions for the validity of the ergodic theorem for nonhomogeneous Markov chains can be found Kolmogorov, Sarymsakov, and Hajnal. We now find the relations between the ergodic probabilities and the absolute probabilities in a homogeneous Markov chain. Let us compute the absolute probability of the event that after* n *steps the system passes into the state* E_j*. Denote this probability by* $c_j(n)$*. We have*

$$c_j(n) = \sum_k P(E_k)\, p_{kj}(n) = \sum_k c_k(n-1) p_{kj} \tag{5.20}$$

where $P(E_k)$ *is the initial probability of the state* E_k*.*

Definition 5.4.12 *A homogeneous Markov chain for which the equalities*

$$P(E_j) = c_j(1) \quad (j = 1, 2, \ldots)$$

are satisfied is called a stationary chain and the probabilities $c_j(n)$ are called stationary absolute probabilities. We observe that from the last equalities and from formula (5.20), for $j = 1, 2, \ldots$ and $n = 1, 2, 3, \ldots$, it follows that

$$c_j(1) = c_j(2) = \ldots = c_j(n) = c_j$$

Thus from formula (5.20) we obtain the equalities

$$c_j = \sum_{k:} c_k p_{kj} \quad (j = 1, 2, \ldots) \tag{5.21}$$

Suppose that the number of states is finite and equal to s. Suppose that the assumptions of theorem are satisfied; thus the ergodic probabilities p_i exist. By comparing formulas is easy to verify that $c_j = p_j (j = 1, 2, 3, \ldots, s)$. Thus; if the initial probabilities $P(E_j)$ are equal, for $j = 1, \ldots, s$, to the ergodic probabilities p_j, then $c_j(n) = p_j$ will be constant for $n = 1, 2, 3, \ldots$; hence the chain will be stationary. We shall have an equilibrium in the sense of the invariance of absolute probabilities. This explains the name "ergodic theorem." However, we observe that for an arbitrary Markov chain with a finite number of states we have the following theorem.

Theorem 5.4.13 *The limits of the absolute probabilities*

$$\lim_{n \to \infty} c_j(n) = c_j \quad (j = 1, 2, \ldots, s) \tag{5.22}$$

for a homogenous Markov chain with a finite number of states exist independently of the initial distribution if and only if the ergodic probabilities p_j exist. We then have $c_j = p_j (j = 1, 2, \ldots, s)$.

Let Us Sum Up

Learners, in this section we have seen that definition of Ergodic theorem and also given theorems and Illustrations.

Check Your Progress

1. In the context of the ergodic theorem, what does it mean if a dynamical system is ergodic?
A. Every invariant set under the system's evolution has measure zero or one.
B. Every function is invariant under the system's evolution.
C. The system exhibits chaotic behavior.
D. The system has a periodic orbit.
2. For a system satisfying the ergodic theorem, if X is a measure-preserving dynamical system and f is an integrable function, what does the ergodic theorem state about the function f?
A. The time average of f converges to its space average almost everywhere.
B. The time average of f converges to the mean of the function over time.
C. The space average of f is constant over time.
D. The function f must be periodic.

5.5 Random Variables Forming a Homogeneous Markov Chain

Suppose that (5.22) is satisfied and c_j does not depend on the initial distribution. Then we may put $P(E_i) = 1$ and $P(E_j) = 0 (i \neq j)$. Hence by formula (5.20) we have

$$c_j(n) = p_{i,}(n)$$

Therefore, by (5.22)

$$p_j = \lim_{n\to\infty} p_{ij}(n) = c_j \quad (i, j = 1, 2, \ldots, s)$$

Conversely, suppose that the ergodic probabilities p_j exist. Then by formula (5.20) for an arbitrary initial distribution we obtain

$$\lim_{n\to\infty} c_j(n) = \lim_{n\to\infty} \sum_{k=1}^{s} P(E_k) p_{kj}(n) = p_j \sum_{k=1}^{s} P(E_k) = p_j$$

The considerations of the previous sections may be applied to random variables. Let $\{X_n\} (n = 0, 1, 2, \ldots)$ be random variables that can take on the values $x_i (i = 1, 2, 3, \ldots)$. The values x_i correspond to the states E_i previously discussed. We now give definitions analogous to the definitions.

Definition 5.5.1 *We say that the sequence* $\{X_n\} (n = 0, 1, 2, \ldots)$ *of random variables with possible values* $x_i (i = 1, 2, 3, \ldots)$ *forms a Markov chain if for* $i, j, n = 1, 2, 3, \ldots$ *the equalities.*

$$\begin{aligned} p_{ij}^{(n)} &= P(X_n = x_j \mid X_{n-1} = x_i) \\ =& P\left(X_n = x_j \mid X_{n-1} = x_i, X_{n-2} = x_{i_{n-2}}, \ldots, X_1 = x_{i_1}, X_0 = x_{i_0}\right) \end{aligned} \tag{5.23}$$

are satisfied for arbitrary $x_{i_{n-2}}, \ldots, x_{i_1}, x_{i_0}$.

Definition 5.5.2 *We say that the sequence* $\{X_n\} (n = 0, 1, 2, \ldots)$ *of random variables with possible values* $x_i (i = 1, 2, 3, \ldots)$*, forms a homogeneous Markov chain if for* $i, j, n =$

$1, 2, 3, \ldots$ *theconditional probabilities* $p_{ij}^{(n)}$ *are independent of* n*, that is,*

$$p_{ij}^{(n)} = p_{ij} \tag{5.24}$$

In the terminology of random variables the probability $p_{ij}(n)$ *of transition from the state* E_i *to* E_j *in* n *steps is the probability that* $X_n = x_j$ *provided* $X_0 = x_i$*, which means*

$$p_{ij}(n) = P\left(X_n = x_j \mid X_0 = x_i\right)$$

Formula (5.6) takes the form

$$p_{ij}(n) = \sum_k P\left(X_m = x_k \mid X_0 = x_i\right) P\left(X_n = x_j \mid X_m = x_k\right) \tag{5.23}$$

where $1 \leqslant m < n$*. The absolute probabilities* $c_j(n)$ *expressed by formula (5.20) take the form*

$$c_j(n) = P\left(X_n = x_j\right) = \sum_k P\left(X_0 = x_k\right) P\left(X_n = x_j \mid X_0 = x_k\right) \tag{7.5.4}$$

Definition 5.5.3 *A sequence* $\{X_n\}$ $(n = 0, 1, 2, 3, \ldots)$ *of random variables with possible values* $x_i (i = 1, 2, \ldots)$ *forming a homogeneous Markov chain is stationary if for* $j = 1, 2, 3, \ldots$

$$c_j(0) = P\left(X_0 = x_j\right) = P\left(X_1 = x_j\right) = c_j(1) = c_j$$

If follows from the last equality that for $n = 0, 1, 2, \ldots$ and $j = 1, 2, 3, \ldots$

$$P\left(X_n = x_j\right) = c_j$$

Thus a stationary sequence of random variables is a sequence of identically distributed random variables. It is also easy to formulate the classification of states and the theorems proved previously in the terminology of random variables. We leave this to the reader. It should be stated that the theory of limit distributions for random variables forming a homogeneous Markov chain is less advanced than the same theory for independent random variables. Conditions for the validity of the central limit theorem for Markov chains with three states were found by Markov, and for chains with an arbitrary finite number of states by Romanovsky Fréchet, and Onicescu and

Mihoc. Doeblin showed that for a certain class of Markov chains with a countable number of states, the question of limit theorems can be reduced to the analogous question for independent random variables. For chains with an arbitrary number of states, some results were obtained by Doeblin, Doob, Dynkin, and Chung $[2, 3]$. A quite advanced, result, which is essentially a generalization of the Lindeberg-Lévy theorem to a large class of random variables forming a homogeneous Markov chain, was recently obtained by Nagayev. The local limit theorem for Markov chains with a finite number of states was given by Kolmogorov. Sirazhdinov obtained some results concerning the rate of convergence to the limit distribution in the local and integral limit theorems for Markov chains with a finite number of states. In the paper quoted, Nagayev obtained the local central limit theorem for Markov chains with a countable number of states and estimated the rate of convergence to the normal distribution. Some theorems concerning the laws of large numbers for random variables forming a Markov chain can be found in the book by Doo and the paper of Chung. Breiman recently obtained a general result in this field. We merely state here without proof the law of large numbers and the central limit theorem for random variables forming a homogeneous Markov chain with a finite number of states.

Theorem 5.5.4 *Let $\{X_k\}$ $(k = 0, 1, 2, \ldots)$ be a stationary sequence of random variables forming a homogeneous Markov chain with a finite number of states. If all the intrinsic states are nonperiodic and form one class, then*

$$P\left[\lim_{n\to\infty} \frac{1}{n+1} \sum_{k=0}^{n} X_k = E\left(X_0\right)\right] = 1 \tag{5.25}$$

Thus if the assumptions of this theorem are satisfied, then the sequence $\{X_k\}$ obeys the strong law of large numbers. Compare this theorem with the theorem of Kolmogorov.

Example 5.5.5 *Consider a stationary sequence of random variables $X_{kc}(k = 0, 1, 2, \ldots$) which can take on only two values x_1 and x_2 and form a homogeneous Markov chain. The number x_1 is the state E_1 and the number x_2 is the state E_2. Suppose that the transition matrix is*

$$\mathbf{M}_1 = \begin{bmatrix} p_{11} & p_{12} \\ p_{21} & p_{22} \end{bmatrix}$$

where $0 < p_{12} < 1, 0 < p_{21} < 1$. The assumptions of theorem are satisfied for $r = 1$. Both states are intrinsic and nonperiodic and they form one class. By formula (5.19) and the *relation* $p_1 + p_2 = 1$ *we obtain the ergodic probabilities*

$$p_1 = \frac{p_{21}}{p_{12} + p_{21}}, \quad p_2 = \frac{p_{12}}{p_{12} + p_{21}} \tag{5.26}$$

and $0 < p_1 < 1, 0 < p_2 < 1$. *By the assumption that* $\{X_k\}$ *is stationary we have* $P(X_k = x_1) = p_1$ *and* $P(X_k = x_2) = p_2 (k = 0, 1, 2, \ldots)$. *By theorem the relation*

$$\lim_{n=\infty} \frac{1}{n+1} \sum_{k=0}^{n} X_k = \frac{p_{21}}{p_{12} + p_{21}} x_1 + \frac{p_{12}}{p_{12} + p_{21}} x_2 \tag{5.27}$$

holds with probability one. In particular, let $x_1 = 1$ *and* $x_2 = 0$. *Then the event* $(\sum_{k=0}^{n} X_k = m)$ *occurs if* m *times among the possible* $n + 1$ *times the system is in the state* E_1. *Relation (7.5.7) states that, with probability one, the average number of times that the system is in the state* E_1 *tends to* $p_{21}/(p_{12} + p_{21})$. *If we treat the appearance of the value* $x_1 = 1$ *as a success, we see that the number of successes in a sequence of trials forming a homogeneous Markov chain obeys the strong law of large numbers. The weak law of large numbers for this example was obtained by Markov. We now present the central limit theorem. Let*

$$Y_n = \sum_{k=0}^{n} [X_k - E(X_k)]$$

Theorem 5.5.6 *Let* $\{X_k\} (k = 0, 1, 2, \ldots)$ *be a sequence of random variables forming a homogeneous Markov chain with a finite number of states. If all the intrinsic states are nonperiodic and form one class, and if the variance* $D^2(Y_n)$, *when the sequence* $\{X_k\}$ *is stationary, satisfies the relation*

$$\lim_{n\to\infty} \frac{D^2(Y_n)}{n+1} = \sigma^2 > 0 \tag{5.28}$$

then for an arbitrary initial distribution of the random variable X_0, *the relation*

$$\lim_{n\to\infty} P\left(\frac{Y_n}{\sigma\sqrt{n+1}} < y\right) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{y} e^{-y^2/2} dy \tag{5.29}$$

is satisfied.

Example 5.5.7 *Let us return to example, in which the sequence* $\{X_k\}$ *is stationary, and set* $x_1 = 1, x_2 = 0$. *Let* $Z_k = X_k - E(X_k)$. *Since* $\{X_k\}$ *is stationary, we have* $E(X_k) = p_1 (k = 0, 1, 2, \ldots)$. *Let us find* $D^2(Y_n)$ *and verify that relation (7.5.8) is satisfied. Since* *the sequence $\{X_k\}$ is stationary, we obtain*

$$D^2(Y_n) = \sum_{k=0}^{n} D^2(Z_k) + 2\sum_{k=0}^{n-1}\sum_{m=k+1}^{n} E(Z_k Z_m) \tag{5.30}$$
$$= (n+1)p_1 p_2 + 2\sum_{k=0}^{n-1}\sum_{m=k+1}^{n} E(Z_k Z_m)$$

To find $E(Z_k Z_m)$, we observe that the random variable $Z_k Z_n$ can take on the following values with the respective probabilities:

$$P\left[Z_k Z_m = (1-p_1)^2\right] = P(X_k = 1)\, P(X_m = 1 \mid X_{kc} = 1) = p_1 p_{11}(m-k)$$
$$P\left[Z_k Z_m = (1-p_1)(-p_1)\right] = P(X_k = 1)\, P(X_m = 0 \mid X_k = 1)$$
$$+ P(X_k = 0)\, P(X_m = 1 \mid X_k = 0)$$
$$= p_1 p_{12}(m-k) + p_2 p_{21}(m-k)$$
$$P\left(Z_k Z_m = {p_1}^2\right) = P(X_k = 0)\, P(X_m = 0 \mid X_k = 0) = p_2 p_{22}(m-k).$$

Hence, after some simple computations,

$$E(Z_k Z_m) = p_1 {p_2}^2 p_{11}(\dot{m}-k) - {p_1}^2 p_2 p_{12}(m-k) \tag{5.31}$$
$$- p_1 p_2^2 p_{21}(m-k) + p_1^2 p_2 p_{22}(m-k)$$
$$= p_1 p_2 \left[p_{11}(m-k) - p_{21}(m-k)\right] = p_1 p_2 \left(p_{11} - p_{21}\right)^{m-k}$$

Therefore, using formula (5.15) and letting $\delta = p_{11} - p_{21}$, we obtain

$$D^2(Y_n) = p_1 p_2 \left\{n + 1 + 2\left[n\delta + (n-1)\delta^2 + \ldots + \delta^n\right]\right\}$$
$$= p_1 p_2 \left[n + 1 + 2\left(\sum_{j=1}^{n} \delta^j + \sum_{j=1}^{n-1} \delta^j + \ldots + \delta\right)\right]$$
$$= p_1 p_2 \left[n + 1 + \frac{2\delta n}{1-\delta} - \frac{2\delta^2(1-\delta^n)}{(1-\delta)^2}\right]$$

5.6 Let Us Sum Up

Learners, in this section we have seen that the defintions of random variables forming a homogeneous Markov chain and also given theorems and applications.

Check Your Progress

1. Which of the following best defines a sequence of random variables $\{X_n\}$ as forming a homogeneous Markov chain?
A. The probability distribution of X_{n+1} depends only on X_n and not on $X_{n-1}, X_{n-2}, \ldots, X_0$.
B. The probability distribution of X_{n+1} depends on X_n and X_{n-1}.
C. The sequence $\{X_n\}$ is independent and identically distributed.
D. The probability distribution of X_n is uniform across all n.
2. In a homogeneous Markov chain, what does it mean if the transition probabilities are time-invariant?
A. The transition probabilities vary with time but are stationary.
B. The probability distribution of the chain at each time step is the same.
C. The chain is always in the same state.
D. The transition probabilities from state i to state j do not change over time.

5.7 Unit Summary

The fifth unit content on homogeneous Markov chains, transition matrix, Ergodic theorem, random variables forming a homogeneous Markov Chain.

Glossary

1. The $p_{ij}^{(n)}$ is The probability p_{ij} is called the transition probability from the state E_i to the state $E_,$ in one trial.

2. The M_1 is a matrix with the transition probabilities p_{ij} as elements is called the transition matrix. This matrix is denoted by M_1.

3. If $p_{ij}^{(n+1)}$ is $p_{ij}(n)$ the probability of passing in n trials from t state E_i to the state E_j in a homogeneous Markov chain.

4. The $c_j(n)$ the absolute probability of the event that after n steps the system passes into the state E_j.

Self-Assessment Questions

Short Answers: (5 Marks)

1. If the transition matrix of a homogeneous Markov chain with four states has the form

$$\mathbf{M}_1 = \begin{bmatrix} \frac{1}{4} & \frac{1}{2} & 0 & \frac{1}{4} \\ \frac{1}{5} & 0 & \frac{1}{3} & \frac{7}{15} \\ 0 & \frac{2}{3} & \frac{1}{3} & 0 \\ \frac{1}{4} & \frac{1}{4} & \frac{1}{4} & \frac{1}{4} \end{bmatrix}$$

(a) Calculate all states.
(b) Check whether the ergodic theorem holds.
(c) If so, find the ergodic probabilities.

2. If the transition matrix of a homogeneous Markov chain with four states has the form

$$\mathbf{M}_1 = \begin{bmatrix} \frac{1}{3} & \frac{2}{3} & 0 & 0 \\ \frac{3}{4} & \frac{1}{8} & \frac{1}{8} & 0 \\ 0 & 0 & \frac{1}{2} & \frac{1}{2} \\ 0 & 0 & \frac{1}{3} & \frac{2}{3} \end{bmatrix}$$

(a) Classify all states.
(b) Check whether the ergodic theorem holds.
(c) If so, find the ergodic probabilities.

3. (a) Prove that for an arbitrary homogeneous Markov chain with a finite number of states the limits exist.

$$\lim_{n\to\infty} \frac{1}{n} \sum_{k=1}^{n} p_{ij}(k) = q_{ij}$$

4. Let $K_{ij}(n)$ denote the probability of passing from the state E_i to the state E_j for the first time on the nth step and let

$$L_{ij} = \sum_{n=1}^{\infty} K_{ij}(n), \quad R_{ij} = \sum_{n=1}^{\infty} nK_{ij}(n)$$

The expression R_{jj} is called the mean recurrence time of the state E_j. The state E_j is called recurrent if $L_{jj} = 1$, transient if $L_{jj} < 1$. A recurrent state with $R_{jj} = \infty$ is called a null state. A recurrent state which is neither a null state nor periodic is called an ergodic state. Show that (a) $K_{ij}(n) = p_{ij}(n) - K_{ij}(1)p_{jj}(n-1) - \ldots - K_{ij}(n-1)p_{jj}$.

5. Let $K_{ij}(n)$ denote the probability of passing from the state E_i to the state E_j for the first time on the nth step and let

$$L_{ij} = \sum_{n=1}^{\infty} K_{ij}(n), \quad R_{ij} = \sum_{n=1}^{\infty} nK_{ij}(n)$$

The expression R_{jj} is called the mean recurrence time of the state E_j. The state E_j is called recurrent if $L_{jj} = 1$, transient if $L_{jj} < 1$. A recurrent state with $R_{jj} = \infty$ is called a null state. A recurrent state which is neither a null state nor periodic is called an ergodic state. Show that homogeneous Markov chain with a finite number of states, the state E_j is recurrent if and only if it is intrinsic.

Long Answers: (8 Marks)

1. Let us consider a homogeneous Markov chain with a countable number of states with the transition matrix

$$\mathbf{M}_1 = \begin{bmatrix} p_1 & 1-p_1 & 0 & 0 & 0 & \ldots \\ p_2 & 0 & 1-p_2 & 0 & 0 & \ldots \\ p_3 & 0 & 0 & 1-p_3 & 0 & \ldots \\ \ldots & \ldots & \ldots & \ldots & \ldots & \ldots \end{bmatrix}$$

Show that if $\sum_{j=1}^{\infty} \dot{p}_j < \infty$, then all states are transient and if $\sum_{j=1}^{\infty} p_j = \infty$, then all states are recurrent. Deduce that there may exist states which are at the same time transient and intrinsic.

2. Let us denote by Ω_{ij} the probability that the system will return an infinite number of times to the state E_j if at the initial moment it was in the state E_i. Prove that (a) if $L_{ij} = 1$, then $\Omega_{ij} = 1$. (b) for a set of intrinsic states which form one class either all $\Omega_{ij} < 1$ or all $\Omega_{ij} = 1$.

3. Prove that in a set of intrinsic states which form one class, either all $L_{ij} < 1$ or all $L_{ij} = 1$.

4. Let $\mathbf{M}_1 = [p_{ij}]$ denote the transition matrix of a homogeneous Markov chain with a countable number of states $E_1, E_2, E_3, \ldots$

   I. If all states are recurrent, non-null and nonperiodic and form one class, then for $i, j = 1, 2, \ldots$,

$$\lim_{n \to \infty} p_{ij}(n) = p_j = 1/R_{jj}$$

   where $p_1 + p_2 + \ldots = 1, p_j > 0$, and $p_j = c_j$, where c_j is the stationary probability. II. If E_j is a transient or a recurrent null state, then for all i we have $\lim_{n\to\infty} p_{ij}(n) = 0$. III. If E_j is a recurrent, non-null state and has period $d > 1$, then prove that

$$\lim_{n \to \infty} p_{ij}(n) = d/R_{jj}$$

5. Let us consider a homogeneous Markov chain with a countable number of states and with the transition matrix

$$\mathbf{M}_1 = \begin{bmatrix} \frac{1}{2} & \frac{1}{2} & 0 & 0 & 0 & 0 & \cdots \\ \frac{2}{3} & 0 & \frac{1}{3} & 0 & 0 & 0 & \cdots \\ \frac{3}{4} & 0 & 0 & \frac{1}{4} & 0 & 0 & \cdots \\ \frac{4}{5} & 0 & 0 & 0 & \frac{1}{5} & 0 & \cdots \\ \cdots & & \cdots & \cdots & \cdots & \cdots & \cdots \end{bmatrix}$$

   Show that $\lim_{n\to\infty} p_{ij}(n) = p_j = e^{-1}/j (i, j, = 1, 2, \ldots)$.

Exercises

1. Prove that

$$\frac{D^2(Y_n)}{n+1} = p_1 p_2 \left(1 + 2\frac{\delta}{1-\delta} \cdot \frac{n}{n+1}\right) - \frac{p_1 p_2}{n+1} \cdot \frac{2\delta^2(1-\delta^n)}{(1-\delta)^2}.$$

2. The transition matrix of a homogeneous Markov chain with four states has the form

$$\mathbf{M}_1 = \begin{bmatrix} \frac{1}{4} & \frac{1}{2} & 0 & \frac{1}{4} \\ \frac{1}{5} & 0 & \frac{1}{3} & \frac{7}{15} \\ 0 & \frac{2}{3} & \frac{1}{3} & 0 \\ \frac{1}{4} & \frac{1}{4} & \frac{1}{4} & \frac{1}{4} \end{bmatrix}$$

(a) Classify all states. (b) Check whether the ergodic theorem holds. (c) If so, find the ergodic probabilities.

3. The transition matrix of a homogeneous Markov chain with four states has the form

$$\mathbf{M}_1 = \begin{bmatrix} \frac{1}{3} & \frac{2}{3} & 0 & 0 \\ \frac{3}{4} & \frac{1}{8} & \frac{1}{8} & 0 \\ 0 & 0 & \frac{1}{2} & \frac{1}{2} \\ 0 & 0 & \frac{1}{3} & \frac{2}{3} \end{bmatrix}$$

(a) Classify all states. (b) Check whether the ergodic theorem holds. (c) If so, find the ergodic probabilities.

Answers to check your progress

Session (Modulo) 5.1

1. A. The probability of transitioning between different states in one step.

2. B. There is a positive probability of reaching any state from any other state.

Session (Modulo) 5.2

1. A. The transition probabilities are constant over time.

2. C. The matrix of transition probabilities after n steps.

Session (Modulo) 5.3

1. B. Each row of P must sum to 1.

2. B. Relate the n-step and m-step transition probabilities.

Session (Modulo) 5.4

1. A. Every invariant set under the system's evolution has measure zero or one.

2. A. The time average of f converges to its space average almost everywhere.

Session (Modulo) 5.5

1. A. The probability distribution of X_{n+1} depends only on X_n and not on $X_{n-1}, X_{n-2}, \ldots, X_0$.
2. D. The transition probabilities from state i to state j do not change over time.

References

1. M. Fisz, Probability Theory and Mathematical Statistics, John Wiley and sons, New Your, Third Edition, 1963.

Suggested Readings

1. T. Veerarajan, Fundamentals of Mathematical Statistics, Yesdee Publishing, 2017.
2. P. R. Vittal, Mathematical Statistics, Margham Publications, 2002.
3. T. Veerarajan, Probability, Statistics and Random Processes, Mc Graw Hill Education (India) Private Limited, Third Edition, 2015.
4. R.S.N. Pillai and V. Bagavathi, Statistics, Sultan Chand and Sons, 2010.
5. Singaravelu.A, S. Sivasubramanian, Probability and Random Processes , Meenakshi Agency, 2008.
6. DN Elhance, Veena Elhance and BM Aggarwal, Fundamentals of Statistics, Kitab Mahal.
7. S. C. Gupta and V. K. Kapoor, Fundamentals of Mathematical Statistics, Sultan Chand and Sons, 2008.