U.P. Rajarshi Tandon
Open University, Prayagraj

MScSTAT–203(N)/ MASTAT–203(N) Stochastic Process

MScSTAT–203(N)/MASTAT–203(N)
Stochastic Process

Course Design Committee

Dr. Ashutosh Gupta — **Chairman**
Director, School of Sciences
U. P. Rajarshi Tandon Open University, Prayagraj

Prof. Anoop Chaturvedi — **Member**
Department of Statistics
University of Allahabad, Prayagraj

Prof. S. Lalitha — **Member**
Department of Statistics
University of Allahabad, Prayagraj

Prof. Himanshu Pandey — **Member**
Department of Statistics, D. D. U. Gorakhpur University, Gorakhpur.

Prof. Shruti — **Member-Secretary**
School of Sciences
U.P. Rajarshi Tandon Open University, Prayagraj

Course Preparation Committee

Dr. Anupma Singh — **Writer**
Department of Statistics, Ewing Chistian College
University of Allahabad, Prayagraj
(Unit - 1, 2, 3, 4, 5, 6, 7 & 8)

Dr. Pratyasha Tripathi — **Writer**
Department of Statistics
Tilka Manjhi Bhagalpur University, Bhagalpur, Bihar
(Unit – 9, 10, 11, 12, 13, 14 & 15)

Prof. Anoop Chaturvedi — **Editor**
Department of Statistics
University of Allahabad, Prayagraj

Prof. Shruti — **Course / SLM Coordinator**
School of Sciences
U. P. Rajarshi Tandon Open University, Prayagraj

First Edition: ***July 2024***
ISBN : 978-93-48270-94-8

MScSTAT–203(N)/ MASTAT–203(N)
STOCHASTIC PROCESS

Printed and Published by Col. Vinay Kumar, Registrar, Uttar Pradesh Rajarshi Tandon Open University Prayagraj, 2024
Printed by: Cygnus Information Solution Pvt.Ltd., Lodha Supremus Saki Vihar Road Andheri East, Mumbai.

Blocks & Units Introduction

The present SLM on ***Stochastic Process*** consists of sixteen Units with four Blocks.

The ***Block 1-Types of Processes*** is the first block of said SLM, which is divided into four units.

The ***Unit-1 - Poisson Process***, is the first unit of present self-learning material, which describes Poisson (point) process, Brownian motion process, thermal noise, Markov short noise, two valued processes, Model for system reliability, mean value function and covariance kernel of Poisson process, Increment process of a Poisson process, Stationary and evolutionary process.

In ***Unit–2 - Branching Process***, introduction to Simple Branching Process, Probability Generating Function, Moments, define simple Branching process and give definitions of various terms. The probability-generating function of the process and its moments are derived.

In ***Unit–3 - Wiener Process***, discusses the mean value function and covariance kernel of Wiener process, Arc-sine law, Martingales, Stopping time, and optional sampling theorem.

The ***Unit–4 Renewal Process***, deals with the distribution and asymptotic distribution of renewal process, elementary renewal theorem, and the delayed and equilibrium renewal process.

The ***Block – 2 - Markov Chains and Markov Process*** is the second block of said SLM, which is divided into four units.

In ***Unit-5 - Markov Dependent Trials***, discuss Two state Markov sequences and Markov chains. Also explain chain recurrent events and delayed recurrent event.

The ***Unit-6 - Transition Probabilities***, Deal with determination of n-step transition probabilities, Chapman-Kolmogorov equations, first return and first passage probabilities, fundamental theorem of probability of extinction, higher transition probabilities in Markov classification of states and chain.

The ***Unit-7 - Classification of States***, explain Classification of states, communication states, periodicity, stationary probability distributions, limit theorems, Ergodic chains and Irreducible Ergodic chains.

The ***Unit-8 - Continuous Time Markov Processes***, Markov processes in Continuous time. Interval arrival time, stopping time, optional stopping theorem, wald's equation, forward and backward equations for homogeneous case, random variable technique.

The ***Block – 3 - Random Walk and Queuing Process*** is the third block of said SLM, which is divided into four units.

The ***Unit – 9 - Random Walk and Gambler's Ruin Problem*** deals with the Random walk, Brownian motion as a random walk, one-dimensional, two-dimensional and three-dimensional random walks, duality in random walk and gambler's ruin problem.

The ***Unit – 10 - Queuing Process*** deals with the Birth and death processes, renewal process, Queuing models- Specification & Effectiveness, Measures, the $E_k/M/1$, $M/E_k/1$; M/M/1; M/M/k & M/G/1 queuing process.

The ***Unit – 11- Distributions*** deals with the Compound distribution, Machine Interference Problem, Waiting Time Distribution for M/M/1 and M/M/k models,

The last unit of this Block is ***Unit – 12 – Martingales*** discussed about the Martingales, Boob – Decomposition, Martingale convergence theorems.

The ***Block – 4 - Applied Stochastic Process*** is the last fourth block of said SLM, which is divided into three units.

The ***Unit – 13 - Homogeneous Process*** deals with the random variable technique, homogeneous birth and death process, divergent birth process, the general birth and death process, multiplicative process, effect of immigration for homogeneous process.

The ***Unit – 14 - Non-Homogeneous Process*** is discussed about Simple non homogeneous process, Polya process, effect of immigration for non-homogeneous process, Diffusion, Backward Kolmogorov diffusion equation, Fokker-Planck equation.

The unit of this SLM is ***Unit – 15 – Non-Markovian Process*** is discussed about Some multi-dimensional prey and predator, Non-Markovian Process, Embedded Markov Process, Application to population growth, epidemic and counter models.

At the end of every block/unit the summary, self-assessment questions are given.

U.P. Rajarshi Tandon Open University, Prayagraj

MScSTAT – 203(N)/ MASTAT – 203(N) Stochastic Process

Block 1: Types of Processes

Course Design Committee

Course Preparation Committee

©UPRTOU
First Edition: ***July 2024***
ISBN : 978-93-48270-94-8

MScSTAT–203(N)/ MASTAT–203(N)
STOCHASTIC PROCESS

Printed and Published by Col. Vinay Kumar, Registrar, Uttar Pradesh Rajarshi Tandon Open University Prayagraj, 2024
Printed by: Cygnus Information Solution Pvt.Ltd., Lodha Supremus Saki Vihar Road Andheri East, Mumbai.

Blocks & Units Introduction

The present SLM on ***Stochastic Process*** consists of sixteen Units with four Blocks.

The ***Block 1-Types of Processes*** is the first block of said SLM, which is divided into four units.

The ***Unit-1 - Poisson Process***, is the first unit of present self-learning material, which describes Poisson (point) process, Brownian motion process, thermal noise, Markov short noise, two valued processes, Model for system reliability, mean value function and covariance kernel of Poisson process, Increment process of a Poisson process, Stationary and evolutionary process.

In ***Unit–2 - Branching Process***, introduction to Simple Branching Process, Probability Generating Function, Moments, define simple Branching process and give definitions of various terms. The probability-generating function of the process and its moments are derived.

In ***Unit–3 - Wiener Process***, discusses the mean value function and covariance kernel of Wiener process, Arc-sine law, Martingales, Stopping time, and optional sampling theorem.

The ***Unit–4 Renewal Process***, deals with the distribution and asymptotic distribution of renewal process, elementary renewal theorem, and the delayed and equilibrium renewal process.

At the end of every block/unit the summary, self-assessment questions are given.

UNIT – 1 POISSON PROCESS

Structure

1.1 Introduction

In various fields of physical and life we encounter with a random process running along in time. In such processes we study about the phenomenon changing with time (or some other parameter). We consider families of random variables (random variable), which are functions of time parameter, say t, i.e., families of r.v.'s of the type $\{X_t, t \in T\}$, where T is some index set of possible values of t.

Thus, we define a stochastic process as the family of random variables $\{X_t, t \in T\}$. The set of all possible values of X_t, say S, is called the State Space of the stochastic process. The index set T is called the parameter space.

The elements t $(\in T)$ are referred as the time parameter. However, it is not necessary that t is always a time parameter.

If T is a singleton set, we have a single random variable If T is a finite set, say, $T = \{1,2,\ldots\ldots,n\}$, then we have a random vector the study of which pertains to the multivariate statistical analysis.

In stochastic processes we usually consider processes with T an infinite set (countable infinite or uncountable). Also, the state space S can be countable or uncountable. Hence, the following four situations may arise:

T countable, S countable

T countable, S uncountable

T uncountable, S countable

T uncountable, S uncountable

Examples:

X_t: outcome of the t^{th} throw in throning a die, $t \geq 1$. Then $\{X_t, t \geq 1\}$ constitutes a stochastic process. Here $S = \{1,2,\ldots\ldots,6\}; T = \{1,2,3,\ldots\ldots\}$. Both S and T are countable.

X_t is the number of telephone calls received at a switchboard during the period $(0,t), t \in (0,\infty)$. Then $\{X_t;\ t \in (0,\infty)\}$ is a stochastic process Here $S = \{1,2,3,\ldots\}$. Hence S is countable while $T = (0,\infty)$ is uncountable.

X_l : number of weak spots in a textile fiber in a length $(0,l)$ of the fiber. Then $\{X_l; l \in L\}$ is a stochastic process for some index set L.

$\{N_v; v \in V\}$, where N_v is the number of insects in volume v of the soil.

X_t: number of radioactive emissions recorded in a counter in the period $(0,t)$.

$\{N_t, t \in T\}$ here N_t is no of flowers in a plant at time t.

$\{X_t, t \in T\}$, where X_t is magnitude of the signal in an ECG at time t.

$\{X_n, n \in N\}$, where X_n is price of the share of some company on the nth day.

Brownian motion $\{(X_t, Y_t, Z_t);\ t \in T\}$, where (X_t, Y_t, Z_t) is the position of a particle (in three-dimensional space) at time t.

$\{N_t, t \in T\}$, where N_t is size of the population of a country at time t.

Definition: A stochastic process is an indexed family of random variables $\{X_t, t \in T\}$, so that we can write x(t) = X (t, w) in terms of a probability space $\{\Omega, \mathcal{F}, P\}, \omega \in \Omega$. Here Ω is the sample space, $\mathcal{F}$ is a field and P is a probability measure.

In some cases, the members of the family are mutually independent; see example (i), but in general, we come across processes whose members are mutually dependent. Different stochastic processes are described according to the nature of dependence among the members of the family.

A Poisson process is a model for a series of discrete events where the average time between events is known, but the exact timing of events is random. The arrival of an event is independent of the event before.

We know the average time between events, but the events are randomly spaced in time (Stochastic). We might have back-to-back failures, but we could also go years between failures because the process is stochastic.

Criteria:

- Events are independent of each other. The occurrence of one event does not affect the probability another event will occur.
- The average rate (events per time period) is constant.
- Two events cannot occur at the same time.

For example, Customers calling a help centre, visitors to a website, radioactive decay in atoms, photons arriving at a space telescope, and movements in a stock price. Poisson processes are generally associated with time, but they don't have to be. In the case of stock prices, we might know the average movements per day (events per time), but we could also have a Poisson process for the number of trees in an acre (events per area).

1.2 Objectives

After studying this unit, you should be able to

- Define and derive the Poisson process
- Discuss Brownian Motion Process, thermal Noise, and Markov Short Noise
- Explain the two-valued Process and Model for System Reliability
- Evaluate mean value Function and Covariance Kernel of Poisson Process and Increment Process of a Poisson Process
- Discuss Stationary and Evolutionary Process

1.3 Poisson (Point) Process

Let $N(t)$ be the number of occurrences of an event E in an interval $(0, t]$. Let

$$P_n(t) = P[N\,(t)\ =\ n]$$

This probability is a function of the time t. The possible values of n are $n = 0,1,2, \ldots$. Thus

$$\sum_{n=0}^{\infty} P_n(t) = 1.$$

The family of random variables $\{N(t), t \geq 0\}$ is a stochastic process. Here the time t is continuous and the state space of $N(t)$ is discrete and interval-valued. Such a process is called a counting process. In interval $(0, \mathrm{t}]$ the points at which the event occurs are distributed randomly.

Definition: Let $t_1 < t_2, < \cdots t_n < \cdots$ represent the time points at which the event occurs. The random variables $T_1 = t_1, t_2 = t_2 - t_1 \ldots T_n = t_n - t_{n-1}$ are called interarrival times.

The stochastic process $\{N(t), t \geq 0\}$ is a continuous time parameter stochastic process with state space $\{0,1,2, \ldots\}$.

Now we shall show that under certain conditions $N(t)$ follows a Poisson distribution.

Conditions for Poisson Process:

(i) **Stationarity:** The probability of n occurrences (of event E) in an interval of length t depends only on the length t of the interval and n and is independent of where the interval is situated. Thus $P_n(t)$ gives the number of occurrences (of E) in the interval $(T, T + t)\ \forall\ T \geq 0$.

(ii) **Independence:** The probability of n occurrences (of E) in interval $(\mathrm{T}, \mathrm{T} + \mathrm{t})$ is independent of the number of occurrences (of E) before T. This implies the independence of various number of events occurring during non-overlapping time intervals. Thus, for

given n and $t_1 < t_2 \ldots t_n$, $N_{t_1}, N_{t_2} - N_{t_1}, \ldots, N_{t_n} - N_{t_{n-1}}$ are independent random variables.

(iii) **Orderliness:** The occurrence of two or more-point events at a single point of time is impossible. Let $P_{>1}(h)$ be the probability of more than one occurrence (of E) in a time interval of length h. then

$$\lim_{n \to 0} \frac{P_{>1}(h)}{h} = 0,$$

$$i.e.\ P_{>1}(h) = o(h).$$

Note: Here o(h) represents a function $g(h)$ defined for $h > 0$ with the property that

$$\lim_{n \to 0} \frac{g(h)}{h} = 0.$$

$$or \sum_{k=2}^{\infty} P_k(h) = o(h)$$

Here $P_k(h)$ denotes the probability of k occurrences (of E) in a time interval width h.

(iv) $P_1(h) = \lambda h + o(h)$ where $\lambda(> 0)$ is a constant.

We shall see later that (i), (ii) and (iii) imply (iv).

Theorem 1.3.1: Under the conditions (i), (ii), (iii) and (iv), $N(t)$ follows a Poisson distribution with mean λt, *i.e*, $P_n(t)$ is given by.

$$P_n(t) = \frac{e^{-\lambda t}(\lambda t)^n}{n!}; n = 0,1,2, \ldots \qquad (1.1)$$

Proof: For $n \geq 0$ consider $P_n(t + h)$. The n events can happen in time interval $(0, t + h]$ in the following $n + 1$ mutually exclusive ways:

$A_1, A_2, \ldots, A_{n+1}$

A_1: n events in interval $(0, t]$ and no event between $(t, t + h]$

A_2: $n-1$ events in interval $(0,t]$ and one event between $(t,t+h]$

A_3: $n-2$ events in interval $(0,t]$ and two event between $(t,t+h]$

⋮

A_{n+1}: no event in interval $(0,t]$ and n event between $(t,t\ +\ h]$

Now

$P(A_1)$

$= P[N(t) = n\ ]\ P\ [N(h) = 0|N\ (t) = n]$

$= P_n(t)P_0(h)$ $\qquad\qquad$ (from condition (ii))

$P(A_2)$

$= P[N(t) = n-1]P\ [N(h) = 1|N\ (t) = n-1]$

$= P_{n-1}(t)P_1(h)$

⋮

$P(A_{n+1}) = P_0(t)P_n(h)$

Then

$$P_n(t+h) = \sum_{k=0}^{n} P_{n-k}(t)P_k(h)$$

$$= \sum_{k=0}^{1} P_{n-k}(t)\ P_k(h) + \sum_{k=2}^{n} P_{n-k}\ (t)\ P_k(h)$$

$$= \sum_{k=0}^{1} P_{n-k}\ (t)\ P_k(h) + R_k$$

Now

$$R_k = \sum_{k=2}^{n} P_{n-k}\,(t)\,P_k(h)$$

$$\leq \sum_{k=2}^{n} P_k(h)$$

$$\leq \sum_{k=2}^{\infty} P_k(h)$$

$$= P_{>1}(h)$$

$$= o(h) \quad \big(\text{By condition (iii)}\big)$$

Hence

$$P_n(t+h) = P_n(t)P_0(h) + P_{n-1}(t)P_1(h) + o(h) \qquad (1.2)$$

Again from (iv)

$$P_1(h) = \lambda h + o(h)$$

and

$$\sum_{n=0}^{\infty} P_n(h) = 1$$

Therefore

$$P_0\,(h) = 1 - \sum_{n=1}^{\infty} P_n(h)$$

$$= 1 - P_1(h) - P_{>1}(h)$$

$$= 1 - \lambda h + o(h).$$

Thus, from (1.2), we have

$P_n(t+h)$

$= P_n(t)[1-\lambda h+o(h)]+P_{n-1}(t)\,[\lambda h+o(h)]$

$= P_n(t)(1-\lambda h)+P_{n-1}(t)\lambda h+o(h).$

Hence

$$\frac{P_n(t+h)-P_n(t)}{h}=\lambda[P_{n-1}(t)-P_n(t)]+\frac{o(h)}{h}$$

Taking limit as $h \to 0$, we have

$$\frac{d}{dt}P_n(t) = P_n'(t)$$

$$= \lambda[P_{n-1}(t)-P_n(t)]; n \geq 1 \qquad (1.3)$$

Which is a differential-difference equation. For $n = 0$ we get

$P_0(t+h)$

$= P_0(t)P_0(h)$

$= P_0(t)\,[1-\lambda h]+o(h)$

or

$$\frac{P_0(t+h)-P_0(t)}{h}=\lambda P_0(t)+o(h) \qquad (1.4)$$

As $h \to 0$, (1.4) reduces to

$P_n'(t) = -\lambda P_0(t)$

or

$$\frac{d}{dt}\log P_0(t) = -\lambda \qquad (1.5)$$

or $logP_0(t) = -\lambda t + K$ (1.6)

K is a constant. Writing $C = e^K$, (1.6) gives

$P_0(t) = Ce^{-\lambda t}$

Since the occurrence of no event in an interval of zero width is a sure event, we have $P_0(0) = 1$.

Hence, we obtain $C = 1$. Therefore

$P_0(t) = e^{-\lambda t}$

For $n = 1$

$P_n'(t) = \lambda[P_0(t) - P_1(t)]$

or

$\frac{d}{dt}P_1(t) + \lambda P_1 = \lambda e^{-\lambda t}$

$e^{\lambda t}\left[\frac{d}{dt}P_1(t) + \lambda P_1(t)\right] = \lambda$

or $\frac{d}{dt}\left[e^{\lambda t}P_1(t)\right] = \lambda$

Hence

$e^{\lambda t}P_1(t) = \lambda t + C.$

Since $P_1(0) = 0,$ we obtain $C = 0$. Therefore

$P_1(t) = \lambda t\, e^{-\lambda t}$

$= \frac{(\lambda t)^1 e^{-\lambda t}}{1!}$

Hence theorem holds for $n = 0$ and $n = 1$. Suppose the result holds for $n = k - 1$, so that

$$P_{k-1}(t) = \frac{(\lambda t)^{k-1} e^{-\lambda t}}{(k-1)!}$$

Then, for $n = k$, the equation (1.3) becomes

$$\frac{d}{dt} P_k(t) + \lambda P_k(t) = \frac{(\lambda t)^{k-1} e^{-\lambda t}}{(k-1)!}$$

or $$e^{\lambda t} \frac{d}{dt} P_k(t) + e^{\lambda t} \lambda P_k(t) = \frac{(\lambda t)^{k-1}}{(k-1)!}$$

or $$\frac{d}{dt}\left[e^{\lambda t} P_k(t)\right] = \frac{(\lambda)^k t^{k-1}}{(k-1)!}$$

or

$$e^{\lambda t} P_k(t)$$

$$= \frac{(\lambda)^k}{(k-1)!} \int t^{k-1}\, dt + C$$

$$= \frac{\lambda^k t^k}{(k-1)!\, k} + C$$

$$= \frac{(\lambda t)^k}{k!} + C$$

For $k \geq 2$, $P_k(0) = 0$, we have $C = 0$. Hence

$$P_k(t) = \frac{(\lambda t)^k e^{-\lambda t}}{k!}$$

Therefore, by induction we get the result of the theorem for all n.

Result 1.3.1.: The assumptions (i), (ii) and (iii) imply assumption (iv).

Proof: For proving this result, let us consider a time interval of unit length and let

$$p = P_0(1)$$

Divide this time interval in n equal parts, so that

$$p = \left[P_0\left(\frac{1}{n}\right)\right]^n \Rightarrow P_0\left(\frac{1}{n}\right) = p^{\frac{1}{n}}$$

Hence, for positive integer k

$$P_0\left(\frac{k}{n}\right) = p^{\frac{k}{n}}$$

For any positive number t and positive integer n, $\exists$ an integer k such that

$$\frac{k-1}{n} \leq t \leq \frac{k}{n}$$

Here, k is the smallest integer greater than nt.

Since $P_0(t)$ is a non-increasing function of t

$$P_0\left(\frac{k-1}{n}\right) \geq P_0(t) \geq P_0\left(\frac{k}{n}\right)$$

or

$$p^{\frac{k-1}{n}} \geq P_0(t) \geq p^{\frac{k}{n}}$$

Let $n \to \infty$ so that

$$\lim_{n\to\infty} \frac{k}{n} = \lim_{n\to\infty} \frac{k-1}{n} = t$$

and we obtain

$$P_0(t) = p^t \qquad (0 \leq p^t \leq 1)$$

Case I: Let $p = 0$. Hence $P_0(t) = 0\ \forall\ t$, *i.e.,* the probability of at least one point event occurring in any time interval of length t is 1. In other words, in an arbitrary length of time infinitely many events will occur with probability 1. This case is of no interest.

Case II: $p = 1$ hence $P_0(t) = 1 \forall t$. Thus, there is no stream to be studied.

Case III: $0 < p < 1$ is of real interest. Here, substituting $p = e^{-\lambda}$ for some $\lambda > 0$, we have

$P_0(t) = [P_0(1)]^t$

$= p^t$

$= e^{-\lambda t}$

Now, for any time interval t

$P_0(t) + P_1(t) + P_{>1}(t) = 1$

or

$P_1(t) = 1 - P_0(t) - P_{>1}(t)$

$= 1 - e^{-\lambda t} + o(t)$ {by assumption (iii)}

$= 1 - \left\{1 - \lambda t + \frac{(\lambda t)^2}{2!} - \cdots..\right\} + o(t)$

$= 1 - \{1 - \lambda t + o(t)\} + o(t)$

$= \lambda t + o(t).$

Thus (i), (ii), (iii) imply (iv).

Theorem 1.3.2.: The interval between two successive occurrences of a Poisson process $\{N(t), t \geq 0\}$ with parameter λ has an exponential distribution with mean 1/λ.

Proof: Let X be the random variable representing the time interval between two successive occurrences of $\{N(t), t \geq 0\}$ and let $F(x) = P(X \leq x)$ be its distribution function.

Suppose E_i and E_{i+1} are two successive events and E_i occurred at time t_i. Then

$P\{X > x\}$

$= P\{E_{i+1}$ did not occur in $(t_i, t_i + x) | E_i$ occured at time $t_i\}$

$= P\{$no event occurs in interval $(t_i, t_i + x) | N(t_i) = i\}$

$= P\{N(x) = 0 | N(t_i) = i\}$

$= P_0(x) = e^{-\lambda t};\ x > 0.$

Hence

$F(x)$

$= P\{X \leq x\}$

$= 1 - P\{X > x\}$

$= 1 - e^{-\lambda x}; x > 0.$

The pdf of X is

$f(x) = \lambda e^{-\lambda x} \quad x > 0.$

which is the pdf of an exponential with mean $1/\lambda$. Hence the theorem follows.

If X_i denotes the interval between E_i and E_{i+1}; $i = 1,2,$ ... then X_1, X_2 ... are independently distributed. We state this result in the following theorem without proof.

Theorem 1.3.3.: The interarrival times (the interval between successive occurrences) of a Poisson process with mean λt are identically independently distributed random variables following the exponential distribution with mean 1/λ.

The following theorem states that the converse of the above theorem is also true.

Theorem 1.3.4.: If the intervals between successive occurrences of an event E are iid with common exponential distribution with mean $1/\lambda$. Then the events E form a Poisson process with mean λt.

Proof: Let Z_n be the interval between $(n-1)^{th}$ and n^{th} occurrences of a process $\{N(t)\}$ having an exponential distribution with mean $1/\lambda$ and let $Z_1, Z_2,$ ... be iid random variables having an exponential distribution with mean $1/\lambda$. Then sum $W_n = \sum_{i=1}^{n} Z_i$ is the waiting time up to the n^{th} occurrence, i.e., the time from origin to the n^{th} subsequent occurrence. Them W_n follows a gamma distribution with parameters λn. the pdf of W_n is given by

$$g(x) = \frac{\lambda^n x^{n-1} e^{-\lambda x}}{\Gamma(n)};\ \ x > 0.$$

$$P\{N(t) < n\} = P\{W_n = Z_1 + \cdots + Z_n > t\}$$

$$= 1 - P\{W_n \leq t\}.$$

Therefore

$$P\{N(t) = n\} = P\{N(t) < n+1\} - P\{N(t) < n\}$$

$$= P\{W_n \leq t\} - P\{W_{n+1} \leq t\}$$

Since

$$P\{W_n \leq t\} = \int_0^t \frac{\lambda^n x^{n-1} e^{-\lambda x}}{\Gamma(n)} dx$$

$$= \frac{1}{\Gamma(n)} \int_0^{\lambda t} y^{n-1} e^{-y} dy$$

$$= 1 - \frac{1}{\Gamma(n)} \int_{\lambda t}^{\infty} y^{n-1} e^{-y} dy$$

Integrating by parts we obtain

$$\int_{\lambda t}^{\infty} y^{n-1} e^{-y} dy$$

$$= (n-1)! \sum_{j=0}^{n-1} \frac{e^{-\lambda t} (\lambda t)^j}{j!}$$

$$= \Gamma(n) \sum_{j=0}^{n-1} \frac{e^{-\lambda t} (\lambda t)^j}{j!}.$$

Hence

$$P\{W_n \le t\} = 1 - \sum_{j=0}^{n-1} \frac{e^{-\lambda t}(\lambda t)^j}{j!}$$

Thus, the probability distribution of $N(t)$ is

$$p_n(t) = P\{N(t) = n\}$$

$$= P\{W_n \le t\} - P\{ W_{n+1} \le t\}$$

$$= \left(1 - \sum_{j=0}^{n-1} \frac{e^{-\lambda t}(\lambda t)^j}{j!}\right) - \left(1 - \sum_{j=0}^{n} \frac{e^{-\lambda t}(\lambda t)^j}{j!}\right)$$

$$= \frac{e^{-\lambda t}(\lambda t)^n}{n!};\ \ n = 0,1,2, \ldots$$

Thus, the process $\{N(t)\}$is a Poisson process with mean λt.

Note: $W_n = W_n(t)$is the waiting time for the n^{th} arrival. The distribution function of $W_n(t)$is given by

$$P\{W_n \le t\} = F_n(t) \quad \text{(say)}$$

$$= 1 - \sum_{j=0}^{n-1} \frac{e^{-\lambda t}(\lambda t)^j}{j!}$$

For obtaining the pdf of $W_n(t)$ we, have

$$F_n(t) = \frac{d}{dt}\, F_n(t)$$

$$= \lambda\, e^{-\lambda t} \left\{ \sum_{j=0}^{n-1} \frac{(\lambda t)^j}{j!} - \sum_{j=0}^{n-1} \frac{(\lambda t)^{j-1}}{j!} \right\}$$

$$= \frac{\lambda(\lambda t)^{n-1}\, e^{-\lambda t}}{\acute{\Gamma}(n)}; \qquad (0 < t < \infty),$$

which is the pdf of a gamma distribution with parameters (λ, n). $f_n(t)$ is called the n^{th} Erlang density in the context of queueing theory.

Theorem 1.3.5.: Given only one occurrence of a Poisson process $\{N(t)\}$ by the time T, the distribution of time points γ in $[0, T]$ at which it occurred is uniform in $[0, T]$.

Proof: We have

$P[\gamma \leq t] = P$ [The event occurs one time before the time t]

$= P\ [N(t) = 1]$

$= e^{-\lambda t}\ \lambda t$

$P[N(T) = 1] = e^{-\lambda T}\ \lambda T$

and

$P[N(T) = 1 | \gamma \leq t]$

$= P$[event does not occur in interval (t, T)]

$= e^{-\lambda(T-t)}$

Therefore,

$P[|\gamma \leq t | N(T) = 1]$

$$= \frac{P\ [\gamma \leq t]\ [N(T) = 1 | \gamma \leq t]}{P[N(T) = 1]}$$

$$= \frac{t}{T};\ 0 < t \leq T$$

$= G_\gamma[t | N(T) = 1]$ (say)

$G_\gamma[t | N(T) = 1]$ is the conditional cdf of γ given $\{N(T) = 1\}$. Then the conditional pdf of γ given $\{N(T) = 1\}$ is

$$g_r[t|N(T) = 1] = \frac{1}{T};\ 0 < t \leq T$$

which is the pdf of a uniform distribution in $[0, T]$. Hence the theorem follows.

1.4 Brownian Motion Process

When a particle of microscopic size is immersed in a fluid, it is subjected to a great number of random independent impulses owing to collisions with molecules. The resulting vector function $\{X(t), Y(t), Z(t)\}$ representing the position of the particle as a function of time is known as Brownian motion.

1.5 Thermal Noise

Thermal noise Consider a resistor in an electric network. Because of the random motions of the conduction electrons in the resistor, there will occur small random fluctuations in the voltage $X(t)$ across the ends of the resistor. The fluctuating voltage $X(t)$ is called thermal noise.

1.6 Markov Shot Noise

Let $N(t)$ be the number of shocks by t, X_i be the value of the i^{th} shock and S_i be the time of the i^{th} shock.

The Total shock value at time t is

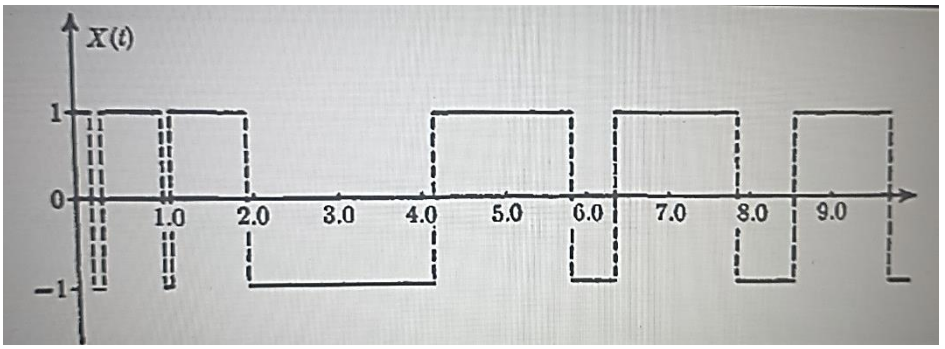
$$X(t) = \sum_{i=1}^{N(t)} X_i\, exp\{-\alpha(i - S_i\}$$

Where α is a constant that determines the exponential rate of decrease. When $X_i, i \geq 1$, are assumed to be i.i.d. and $\{X_i, i \geq 1\}$ is independent of the Poisson process $\{N(t), t \geq 0\}$, is called a shot noise process.

A Shot noise process possesses the Markovian property that given the present state the future is conditionally independent of the past.

1.7 Two Valued Process

Class of stochastic processes are taken only two values, which may be taken to be real numbers A and B. Such a stochastic process will be called a two-valued process. A typical sample function of a two-valued process is graphed in below figure



A two-valued process $\{X(t), t > 0\}$ whose possible values are 1 and (-1) will be called a one-minus-one process. If $X(t)$ is a one-minus-one process, then

$$Y(t) = \frac{B + A}{2} + \left(\frac{B - A}{2}\right) X(t)$$

is a two-valued process, its possible values are A and B.

Case I: It represents a one-minus-one process $\{X(t), t > 0\}$ is as follows. Define $N(0) = 0$ for $t > 0$, let $N(t)$ be the number of times in the interval $(0, t]$ that the one-minus-one process $X(\bullet)$ has changed its value. Then $\{N(t), t > 0\}$ the counting process of the one-minus-one process. It writes

$$X(t) = X(0)(-1)^{N(t)}$$

where $X(0)$ is the initial value of the one-minus-one process.

If $\{N(t), t > 0\}$ is a Poisson process, then $\{X(t), t > 0\}$ a random telegraph, signal. A stochastic process $\{X(t), t > 0\}$ is called a random telegraph signal if

(i) Its values are 1 and (-1) successively

(ii) The times at which the values change is distributed according to a Poisson process $\{N(t), t > 0\}$ at mean rate ν

(iii) $X(0)$ is a random variable, independent of the Poisson process $\{N(t), t > 0\}$, such that

$$P[X(0) = 1] = P[X(0) = -1] = \frac{1}{2}$$

Case II: It represents a two-valued process in terms of the time between changes of values. Let $\{X(t), t > 0\}$ be a zero-one process (that is, a stochastic process, which takes only the values 0 and 1).

1.8 Model for System Reliability

Consider a system that can be in one of two states, "ON" or "OFF". If it is "ON" it serves for a random time before breakdown. If it is "OFF", it is off for a random time before being repaired. Let $X(t)$ be equal to 1 or 0 depending on whether the system is "ON" or "OFF" at time t.

1.9 Mean Value Function and Covariance Kernel of Poisson Process

For a single random variable, the role of its mean and variance is played for a stochastic process by its mean value function and its covariance kernel.

Let $\{X(t), t \in T\}$ be a stochastic process with finite second moments. Its mean value function is defined for all t in T by

$$m(t) = E[X(t)]$$

and its covariance kernel defined for all s and t in T by

$$K(s,t) = Cov[X(s), X(t)]$$

Example: Suppose $X(t)$ represents the position of a particle in motion with a constant velocity. One may assume that $X(t)$ is of the form

$$X(t) = X_0 + Vt$$

where X_0 is a random variable for its initial position and V is random variables for the velocity respectively. The mean value function of $\{X(t), t > 0\}$ is given by

$$m(t) = E[X(t)] = E[X_0] + tE[V]$$

and covariance kernel of $\{X(t), t > 0\}$ is given by

$$K(s,t) = Cov[X(s), X(t)] = Var[X_0] + (s+t)Cov[X_0, V] + st\, Var\,[V]$$

Note: Obtain the mean value function and covariance kernel of $\{X(t), t > 0\}$, one does not need to know the joint probability law of X_0 and V, but only their means, variances, and covariance.

1.10 Increment Process of a Poisson Process

Let $\{N(t), t > 0\}$ be a Poisson process of intensity ν, and let L be a positive constant. A new stochastic process $\{X(t), t > 0\}$ can be defined by

$$X(t) = N(t + L) - N(t) \qquad (1.7)$$

For example, if $N(t)$ represents the number of events of a certain kind occurring in the interval 0 to t, then X(t) represents the number of events occurring in a time interval of length L beginning at t. In principle, one could determine the joint probability law of $X(t_1), X(t_2), \dots, X(t_n)$ for any choice of n time points $t_1, t_2, \dots, t_n$. However, it is usually more convenient to begin one's study of a stochastic process by computing its mean value function and covariance kernel.

For the process $\{X\{t), t > 0\}$ defined in (1.7), the mean value function is

$m(t) = E[X(t)] = E[N(t + L) - N(t)] = \nu L$

The covariance kernels

$K(s,t) = Cov[X(s), X(t)]$

Assume that $s \leq t$, two cases are

(i) $t < s + L$

(ii) $t > s + L$. In case (ii), $X(s)$ and $X\{t)$ are independent random

variables and consequently have zero covariance. In case (i), we write

$$K(s,t) = Cov[N(s+L) - N(s), N(t+L) - N(t)]$$

$$= Cov[N(s+L) - N(t) + N(t) - N(s), N(t+L) - N(t)]$$

$$= Cov[N(s+L) - N(t), N(t+L) - N(t)] \qquad (1.8)$$

The last step follows as $N(t) - N(s)$ and $N(t+L) - N(t)$ have zero covariance.

Consider $N(t+L) - N(t) = N(t+L) - N(s+L) + N(s+L) - N(t)$. Putting this value in (1.8), we have

$$K(s,t) = Var[N(s+L) - N(t)]$$
$$= \nu(s+L-t)$$
$$= \nu\{L - (t-s)\}$$

Thus, the covariance kernel of the process $\{X\{t), t > 0\}$ defined by (1.7) is

$$K(s,t) = \begin{cases} \nu(L - |t-s|) & if\ |t-s| \le L\ for\ all\ s,t \ge 0 \\ 0 & if\ |t-s| > L\ for\ all\ s,t \ge 0 \end{cases}$$

1.11 Stationary and Evolutionary Process

A stationary process is one whose distribution remains the same as time progresses, because the random mechanism producing the process is not changing as time progresses. An evolutionary process is one, which is not stationary.

A Poisson process $\{N(t), t > 0\}$ is evolutionary since the distribution of $N(t)$ is functionally dependent on t. On the other hand, the process $\{X(t), t > 0\}$ defined by (1.7) appears as if it might be stationary, because the distribution of the number of events occurring in a time interval of fixed length L does not functionally depend on the time t at which the interval begins.

An index set T is said to be a linear index set if it has the property that the sum $(t + h)$ is a member of T, of any two members t and h of T.

Examples of such index sets are $T = \{1,2, \dots\}, T = \{0, \pm 1, \pm 2, \dots\}, T = \{t: t \geq 0\}$ and $T = \{t: -\infty < t < \infty\}$.

A stochastic process $\{X(t), t \in T\}$, whose index set T is linear, is said to be

(i) Strictly stationary of order k, where k is a given positive integer, if, for any k points $t_1, t_2, \dots, t_n$ in T, and any h in T, the k-dimensional random vectors $\{X(t_1), X(t_2), \dots, X(t_n)\}$ and $\{X(t_1 + h), X(t_2 + h), \dots, X(t_n + h)\}$ are identically distributed,

(ii) Strictly stationary if for any integer k it is strictly stationary of order k.

For example, consider the stochastic process $\{X(t), t > 0\}$ defined by (1.7). One easily verifies that it is strictly stationary of order 1. With a little more effort, one could prove that it is strictly stationary of order 2.

Covariance Stationarity Process:

A stochastic process $\{X(t), t \in T\}$ is said to be covariance stationary if it possesses finite second moments, if its index set T is linear, and if its covariance kernel $K(s,t)$ is a function only of the absolute difference $|s - t|$, in the sense that there exists a function $R(v)$ such that for all s and t in T

$$K(s,t) = R(s - t)$$

or, more precisely, $R(v)$ has the property that for every t and v in T

$$Cov\,[X(t), X(t + v)] = R(v),$$

where $R(v)$ is the covariance function of the covariance stationary time series $\{X(t), t \in T\}$.

1.12 Summary

In this unit, we have covered the following points:

- We have Defined Poisson process and explain the conditions of Poisson process.
- We have Defined Brownian Motion Process, thermal Noise and Markov Short Noise.

- We have explained the two Valued Process and Model for System Reliability
- We have described Mean value Function and Covariance Kernel of Poisson Process and Increment Process of a Poisson Process
- We have discussed Stationary and Evolutionary Process

1.13 Self-Assessment Exercise

1. The stochastic processes {X(t), t > 0} is defined $X(t) = A + Bt$, in which A and B are independent random variables, each uniformly distributed on the unit interval. Compute

(i) The mean value function $m(t) = E[X(t)]$

(ii) The covariance kernel $K(s,t) = Cov[X(s), X(t)]$

(iii) The covariance function $R(v)$ if the process is covariance stationary.

2. The stochastic processes {X(t), t > 0} is defined $X(t) = A + Bt + Ct^2$, in which A B and C are independent random variables, each with mean 1 and variance 1. Compute

(i) The mean value function $m(t) = E[X(t)]$

(ii) The covariance kernel $K(s,t) = Cov[X(s), X(t)]$

(iii) The covariance function $R(v)$ if the process is covariance stationary.

3. The stochastic processes {X(t), t > 0} is defined $X(t) = A\cos wt + B\sin wt$, in which w is a positive constant, and A and B are uncorrelated random variables with means 0 and variances σ^2. Compute

(i) The mean value function $m(t) = E[X(t)]$

(ii) The covariance kernel $K(s,t) = Cov[X(s), X(t)]$

(iii) The covariance function $R(v)$ if the process is covariance stationary.

4. Show that if $\{X(t), t \in T\}$ is strictly stationary of order k, then $\{X(t), t \in T\}$ is strictly stationary of order k', for any integer $k' < k$.

5. Let $X(t) = \sin wt$, where w is uniformly distributed on 0 to 2π.

(i) Show that $\{X(t),\ t = 1,2, \dots\}$ is covariance stationary but is not strictly stationary

(ii) Show that $\{X(t), t \geq 0\}$ is neither covariance stationary nor strictly stationary.

1.14 References

- Medhi, J. (2002): Stochastic Process, 2nd edition, New Age International Publishers.
- Osborne, M. F. M. (1959): Reply to comments on Brownian motion in the stock market, Oper. Res., Vol. 7, pp. 807-811.
- Wonham, W. M., and A. T. Fuller. (1958): Probability densities of the smoothed random telegraph signal. Jour, of Electronics and Control, Vol. 4, pp. 567-576.
- Wax, N. (1954): Noise and Stochastic Processes. New York: Dover.
- Wold, H. (1954): A Study in the Analysis of Stationary Time Series, 2nd ed. (with an appendix by Peter Whittle). Uppsala, Swedeu.
- Fortet, R. (1951): Random functions from a Poisson process, Proc. 2nd Berkeley Symp.
- Kac, M. (1947): Random Walk and the theory of Brownian motion, Amer. Math. Monthly, Vol. 54, p. 369.
- Wang, Ming Chen, and G. E. Uhlenbeck. (1945): On the theory of the Brownian motion II, Rev. Mod. Phys., Vol. 17, pp. 323-342.

1.15 Further Readings

- Ross, S. M. (2014). Stochastic Processes. Wiley.
- Levin, D. A., Peres, Y., & Wilmer, E. L. (2009). Markov Chains and Mixing Times. American Mathematical Society.
- Lefebvre, M. (2006). Applied Stochastic Processes. Springer.
- Grimmett, G., & Stirzaker, D. (2001). Probability and Random Processes. Oxford University Press.

- Taylor, H. M., & Karlin, S. (1998). An Introduction to Stochastic Modelling. Academic Press.
- Resnick, S. I. (1992). Adventures in Stochastic Processes. Birkhäuser.
- Karatzas, I., & Shreve, S. E. (1991). Brownian Motion and Stochastic Calculus. Springer-Verlag.

UNIT – 2: BRANCHING PROCESS

Structure

2.1 Introduction

A simple branching process is a mathematical model that describes the evolution of a population where each individual can produce a random number of offspring in the next generation. This model is often used in biology, genetics, and other fields to study phenomena such as population growth, species extinction, and the spread of genes.

In a simple branching process, the population evolves in discrete steps called generations. The individuals in one generation give rise to the individuals in the next. Each individual in the population produces a random number of offspring according to a specified probability distribution. A common choice is the Poisson distribution, but other distributions can be used as well. The process starts with an initial population, typically consisting of a single individual, though more complex initial conditions can be considered. The process can lead to extinction, where the population dies out, or survival, where the population grows indefinitely or stabilizes.

The history of the study of branching processes dates back to 1874, when a mathematical model was formulated by Galton and Watson for the problem of 'extinction of family'. Later, the model has not attracted much attention for a long time. However, the situation gradually changed during the last 60 years and much attention has been devoted to it. This is because of the development of interest in the applications of probability theory, in general, and its applications in a variety of biological, physical, and other problems where one is concerned with objects that can generate similar objects. Such objects may be biological entities, such as human beings, animals, genes, bacteria, neutrons yielding new neutrons under a nuclear chain reaction or in the process of nuclear fission.

This model provides a foundation for understanding more complex stochastic processes and can be extended in various ways, such as considering different types of individuals, varying offspring distributions, or adding immigration and emigration.

2.2 Objectives

After reading this unit, you should be able to:

- Describe the branching process with examples.
- Obtain the probability-generating function, expectations, and variance for a branching process.
- Derive the probabilities of extinction.
- Express various limiting properties for the probability of extinction.

2.3. Simple Branching Process

Let p_i be the probability that a man produces i sons. Let each son have the same probability distribution for sons of his own and so on. One may be interested in the probability that the male line extinct in a generation, or the probability distribution of the number of descendants in the n^{th} generation.

2.3.1 Assumptions of the Simple Branching Process

Suppose we start with a population of X_0 individuals (or objects), which form the 0^{th} generation. These objects are called ancestors. The offspring reproduced or the object generated by the objects of the 0^{th} generation are the direct descendants of the ancestors and are said to form the 1^{st} generation; the objects generated by those of the 1^{st} generation form the 2^{nd} generation, and so on. Let X_n be the number of individuals in the n^{th} generation. These are composed of the descendent of the $(n-1)^{th}$ generation.

The model proposed by Watson was based on the following assumptions:

(i) The objects reproduce independently of other objects, *i.e.*, there is no interference.

(ii) The number X of individuals produced by an individual has the probability distribution

$$P(X = k) = p_k;\ k = 0,1,2, \ldots;\ \sum p_k = 1.$$

(iii) The probability distribution $\{p_k\}$ remains the same from generation to generation.

The sequence of random variables $\{X_n; n = 0,1,2, \ldots\}$ constitutes a Galton-Watson (G.W.) branching process with offspring distribution $\{p_k; k = 0,1,2, \ldots\}$

2.4 Probability Generating Function of the Branching Process

Let

$$g(s) = \sum_{k=0}^{\infty} p_k\, s^k;\quad 0 \le s \le 1$$

be the pgf of X and $g_n(s)$ be the pgf of X_n; i.e.

$$g_n(s) = \sum_{k} P\{X_n = k\} s^k; 0 \le s \le 1.$$

Without loss of generality, we assume that $X_0 = 1$, *i.e.,* the process starts with one individual. Then

$$g_0(s) = s$$
$$g_1(s) = g(s).$$

Theorem 2.4.1.: We have

$$g_n(s) = g_{n-1}[g(S)] \qquad (2.1)$$

$$g_n(s) = g[g_{n-1}(s)] \qquad (2.2)$$

Proof: We can write

$$X_n = \sum_{r=1}^{X_{n-1}} \xi_r$$

Where ξ_r are iid random variables with probability distribution $\{p_k\}$. Now

$$P\{X_n = k\}$$

$$= \sum_{j=0}^{\infty} P\{x_n = k | x_{n-1} = j\} P\{X_{n-1} = j\}$$

$$= \sum_{j=0}^{\infty} P\left\{\sum_{r=1}^{\infty} \xi_r = k\right\} P\{X_{n-1} = j\}$$

Therefore

$$g_n(s)$$

$$= \sum_{k=0}^{\infty} P\{x_n = k\} s^k$$

$$= \sum_{k=0}^{\infty} s^k \left[\sum_{j=0}^{\infty} P\left\{\sum_{r=1}^{j} \xi_r = k\right\} P\{X_{n-1} = j\}\right]$$

$$= \sum_{j=0}^{\infty} P\{X_{n-1} = j\} \left[\sum_{k=1}^{\infty} P\left\{\sum_{r=1}^{j} \xi_r = k\right\} s^k\right]$$

Since $\xi_1, \xi_2, \ldots$ are iid random variables each with pgf $g(s)$, the pgf of $\sum_{r=1}^{j} \xi_r$ is given by

$$\sum_{k=1}^{\infty} P\left\{\sum_{r=1}^{j} \xi_r = k\right\} s^k = E\left[s^{\sum_{r=1}^{j} \xi_r}\right]$$
$$= [g(s)]^j.$$

Thus

$$g_n(s)$$
$$= \sum_{j=0}^{\infty} P\{x_{n-1} = j\} [g(s)]^j$$
$$= g_{n-1}\big(g(s)\big)$$

which gives (2.1).

Substituting $n = 2,3, \ldots$ in (2.1) we get

$$g_2(s) = g_1(g(s) = g\big(g(s)\big)$$
$$g_3(s) = g_2(g(s) = g(g\big(g(s)\big) = g(g_2(s))$$
$$g_4(s) = g_3(g(s) = g\big(g_3(s)\big)$$

In general

$$g_n(s) = g_{n-1}\big(g(s)\big) = g_{n-2}\big[g\big(g(s)\big)\big] = g_{n-2}\big(g_2(s)\big) = g_{n-3}\Big(g\big(g_2(s)\big)\Big) = g_{n-3}\big(g_3(s)\big)$$
$$= \cdots = g_{n-k}\big(g_k(s)\big) \qquad (k = 0,1,2, \ldots, n)$$

For $k = n - 1$

$$g_n(s) = g_1[g_{n-1}(s)] = g[g_{n-1}(s)].$$

This proves result (2.2) of the theorem.

2.5 Moments of X_n

Theorem 2.5.1.: If we assume that $E(X_1) = \sum_{k=0}^{\infty} k\, p_k = \mu$ and $Var\ (X_1) = \sigma^2$, then

$$E(X_n) = \mu^n \qquad (2.3)$$

$$Var\ (X_n) = \begin{cases} \dfrac{\mu^{n-1}(\mu^n - 1)}{\mu - 1}\sigma^2 & if\ \mu \neq 1 \\ n\sigma^2 & if \quad \mu = 1 \end{cases} \qquad (2.4)$$

Proof: We have

$$g_n(s) = g_{n-1}\big(g(s)\big) \qquad (2.5)$$

Differentiating (2.5) with respect to s we get

$$g_n'(s) = g_{n-1}'\big(g(1)\big)g'(s)$$

So that

$$g_n'(1) = g_{n-1}'\big(g(1)\big)g'(1) = g_{n-1}'(1)(\mu)$$

On iterating, we get

$$g_n'(1) = g_{n-2}'(1)\mu^2 = g_{n-3}'(1)\mu^3 = \cdots = g_1'(1)\mu^{n-1} = \mu^n.$$

Again

$Var\ (X_n)$

$= E[X_n(X_n - 1)] + E(X_n) - [E(X_n)]^2$

$= g_n''(1) + g_n'(1) - [g_n'(1)]^2$

Now

$g_n''(s)$

$= g''_{n-1}\big(g(s)\big)[g'(s)]^2 + g'_{n-1}\big(g(s)\big)\, g''(s)$

So that

$g''_n(1)$

$= \ g''_{n-1}(1)\big(g(s)\big)[g'(s)]^2 + g'_{n-1}\big(g(s)\big)\, g''(s)$

$= \ g'_{n-1}(1)\, \mu^2 + \mu^{n-1} m$

where

$m = g''(1)$

$= E[X_1(X_1 - 1)]$

$= \ \sigma^2 + \ \mu^2 - \mu$

$= \sigma^2 + \mu(\mu - 1).$

On iterating we obtain

$g''_n(1)$

$= m\mu^{n-1} + \mu^2[m\mu^{n-2} + \mu^2 g''_{n-2}(1)]$

$= m\ (\mu^{n-1} + \mu^n) + \mu^4 g''_{n-2}(1)$

$= \cdots$

$= m(\mu^{n-1} + \mu^n + \cdots \ldots + \mu^{n-2}) + \mu^{2n-2} g''_1(1)$

$= m\ \mu^{n-1}\ (1 + \mu + \cdots \ldots + \mu^{n-2}) + \mu^{2n-2}\ m$

$= m\ \mu^{n-1}\ (1 + \mu + \cdots \ldots + \mu^{n-2} + \mu^{n-1})$

$$= \begin{cases} m.n & if\ \mu = 1 \\ m\mu^{n-1}\ \dfrac{(\mu^n - 1)}{\mu - 1} & if\ \mu \neq 1 \end{cases}$$

Hence, for $\mu \neq 1$

$Var\ (X_n)$

$= m\mu^{n-1}\ \frac{(\mu^n - 1)}{\mu - 1} + \mu^n - \mu^{2n}$

$= \left(\sigma^2 + \mu(\mu - 1)\right)\mu^{n-1}\ \frac{(\mu^n - 1)}{\mu - 1} + \mu^n - \mu^{2n}$

$= \sigma^2\mu^{n-1}\ \frac{(\mu^n - 1)}{\mu - 1} + \mu^{2n} - \mu^n + \mu^n - \mu^{2n}$

$= \sigma^2\mu^{n-1}\ \frac{(\mu^n - 1)}{\mu - 1}$

and for $\mu = 1$

$Var\ (X_n) = \sigma^2 n.$

Hence the theorem follows.

Example 2.5.1.: A branching process $\{X_n : n = 0,1,2, \ldots\}$ with offspring distribution given by $P_0 = \frac{1}{6}\ , p_1 = \frac{1}{2}\ , p_2 = \frac{1}{3}$

Determine Expectation and variance of the population at generation 8.

Solution: Let X be the number of offspring produced by one individual.

$E[X] = \sum_k k\ p_k$

$= 1 \times \frac{1}{2} + 3 \times \frac{1}{3}$

$= \frac{1}{2} + 1$

$= 1.5 = \mu$

$$Var[X] = \sum_x (k-\mu)^2 \, p_k$$

$$= (0-1.5)^2 \times \frac{1}{6} + (1-1.5)^2 \times \frac{1}{2} + (3-1.5)^2 \times \frac{1}{3}$$

$$= \frac{2.25}{6} + \frac{0.25}{2} + \frac{2.25}{3}$$

$$= \frac{2.25 + 0.75 + 4.50}{6}$$

$$= 1.25 = \sigma^2$$

We know

$$E[X_n] = \mu^n$$

$$= (1.5)^8$$

$$= 25.63$$

$$Var\,(X_n) = \sigma^2 \mu^{n-1} \, \frac{(\mu^n - 1)}{\mu - 1} \qquad if\ \mu \neq 1$$

$$= (1.25)^2 (1.5)^{8-1} \frac{\{(1.5)^{8-1} - 1\}}{(1.5-1)}$$

$$= (1.56) \times (17.09) \times \frac{\{17.09 - 1\}}{(1.5-1)}$$

$$= 26.66 \times \frac{16.09}{0.5} = \frac{428.96}{0.5} = 857.92$$

If $X_n = 0$, the population is extinct by the n^{th} generation. Obviously, if $X_n = 0$ for $n = m$ then $X_n = 0$ for $n > m$. Thus $P\{X_{n+1} = 0 \,|X_n = 0\} = 1$. The extinction of the process occurs when the random sequence $\{X_n\}$ is consist of zero for all except a finite number of values of n.

Let, $T = min\{n: X_n = 0\}$: time of extinction

If $T < \infty$, the population is extinct after a finite number of generations.

Theorem 2.5.2. (Fundamental Theorem of Probability of Extinction):

If,$\mu\ (= \sum_{k=0}^{\infty} k\, p_k) \leq 1$, the probability of ultimate extinction is 1. If μ>1, the probability of ultimate extinction is the positive root less than unity of the equation

$$g(s) = s \qquad (2.6)$$

Proof: Let

q_n
$= P$robability that the population starting with one ancestor dies out before the n^{th} generation

Therefore

$$q_n = P\{X_n = 0\}$$

The *pgf* of X_n is

$$g_n(s) = \sum_{k=0}^{\infty} P\{X_n = k\}s^k; 0 \leq s \leq 1$$

Hence

$g_n(0)$
$= P\ \{X_n = 0\} = q_n$

Now, if

$p_0 = P\{X = 0\} = 0,$ then

$X_0 \leq X_1 \leq X_2 \leq \cdots$

and $T = \infty$ almost surely, *i.e.,* extinction can never occur.

If $p_0 = 1$ then the population extinct just after the zeroth generation.

We exclude these trivial cases and assume that $0 < p_0 < 1$.

If $p_0 > 0$ and $p_0 + p_1 = 1$, then

$P\{T < n+1 | X_0 = 1\}$

$= p_0 + p_1 p_0 + p_1^2 p_0 + \cdots + p_1^n p_0$

$= p_0 \dfrac{1 - p_1^n}{1 - p_1} = 1 - p_1^n$

$\to 1$ as $n \to \infty$

Hence $T < \infty$ almost surely.

We exclude these trivial cases and assume that $0 < p_0 < p_0 + p_1 < 1$.

Now

$g(s) = p_0 + p_1, s + p_2 s^2 + \cdots \ldots; \quad 0 \le s \le 1$

$g(0) = p_0 > 0$

and for $0 < s \le 1$

$g'(s) > 0$

$g''(s) > 0,$

i.e., for $0 < s \le 1$, $g(s)$ is a continuous, strictly increasing convex function of s.

Since $g(s)$ is convex, the line $y = s$ can intersect the curve $y = g(s)$ in at most two points for $s > 0$. One of these points is $(1,1)$. Thus, there may or may not be another point of intersection. The two possibilities are shown in Figure I and II:

Figure I

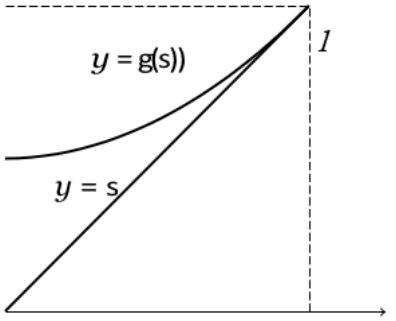
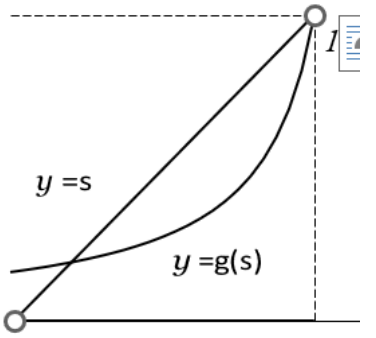


Figure II



Now

$g_{n+1}(s) = g\,(g_n(s))$

substituting $s = 0$, we get

$g_{n+1}(0) = g\big(g_n(0)\big)$

or $q_{n+1} = g(q_n)$ (2.7)

Substituting $n = 0,1,2,\ldots$ respectively in (2.7), we get

$q_1 = g\,(0)$

$= p_0 > 0$

$= q_1 > 0$

$q_2 = g\,(q_1)$

$> g\,(0)$

$= q_1$ {since g(s) is an increasing funciton of s}

$\Rightarrow q_1 > q_2$

Assuming that $q_n > q_{n-1}$

We have

$q_{n+1} = g(q_n)$

$> g\,(q_{n-1}) = \ q_n$

Hence by induction

$q_{n+1} > \ q_n \ \forall \ n = 0,1,2, \ldots$

i.e., the sequence $\{q_0, q_1, \ldots \ldots . q_n, q_{n-1} \ldots\}$ is an increasing sequence bounded above by unity. Hence q_n must have a limit

$\lim_{n\to\infty} q_n = q$ (say), $0 \leq q \leq 1$

q is the probability of ultimate extinction. From (2.2) it follows that q satisfies

$$q \ = \ g(q) \qquad (2.8)$$

Thus, the probability of ultimate extinction is a solution of (2.8).

Let λ be an arbitrary positive root of (2.8). At least one such root exists which is λ=1.

Then

$q_1 = g(0)$

$< g\,(\lambda) = \lambda$ (λ is positive)

$i.e. \ q_1 < \ \lambda$

$q_2 = g(q_1)$

$< g\,(\lambda) = \lambda$

$\Rightarrow \ q_2 < \lambda$

By induction $q_n < \lambda \ \forall \ n = 1,2, \ldots$. Letting $n \to \infty$, we observe that $q < \lambda$.

Since λ is an arbitrary positive root of (2.8), it follows that q is the smallest positive root of (2.8). Thus, we examine the roots of the equation $s = g(s)$ in $(0,1]$. The roots are intersection points of $y = s$ and $y = g(s)$.

If $g'(1) = \mu > 1$ figure II prevails and ∃ a unique positive root q<1.

Thus, if $\mu > 1$, the probability of extinction is <1.

If $g'(1) = \mu \leq 1$ then there is no root <1 and we have $q = 1$.

This proves the theorem.

Example 2.5.2.: Let

$$p_0 = \frac{1}{4}, p_1 = \frac{1}{3}, p_2 = \frac{5}{12}$$

i.e.,

$$P[X_1 = 0|X_0 = 1] = \frac{1}{4},$$

$$P[X_1 = 1|X_0 = 1] = \frac{1}{3}$$

and $P[X_1 = 2|X_0 = 1] = \frac{5}{12}$

Such that

$$P[X_1 = 0|X_0 = 1] + P[X_1 = 1|X_0 = 1] + P[X_1 = 2|X_0 = 1]$$

$$= \frac{1}{4} + \frac{1}{3} + \frac{5}{12}$$

$$= \frac{3 + 4 + 5}{12}$$

$$= \frac{12}{12} = 1$$

Then

$$\mu = \sum_{k=0}^{\infty} k\ p_k$$

$$= 0 \times \frac{1}{4} + 1 \times \frac{1}{3} + 2 \times \frac{5}{12}$$

$$= \frac{2+5}{6}$$

$$= \frac{7}{6} > 1$$

The probability of extinction is

$$q\ = \sum_{s=0}^{\infty} q^s\, p_s$$

$$= q^0 \times \frac{1}{4} + q^1 \times \frac{1}{3} + q^2 \times \frac{5}{12}$$

or $5q^2 + 4q + 3 = 12q$

or

$5q^2 - 8q + 3 = 0$

or $(5q - 3)(q - 1) = 0$

or $q = \frac{3}{5}$ and $q = 1$

Hence the probability of ultimate extinction is 0.6.

Example 2.5.3.: Consider a branching process originating from a single element, where each individual element either replaces itself with probability p_1, or fails to replace itself (i.e., dies out without producing an offspring) with probability $p_0 = 1 - p_1, 0 < p_0 < 1$. Obviously, $\mu = 0 \times (p_0) = 1 \times (p_1) = p_1 < 1$.

$E(X_n) = \mu^n \to 0$ as $n \to \infty$

Here

$g(s) = E(s^X)$

$= s^0(p_0) + s^1(p_1)$

$= p_0 + sp_1$

So that $s = 1$ is the only solution of the equation $g(s) = s$ as $p_0 + p_1 = 1$. As such, the probability q that the process will be extinct is 1.

Example 2.5.4.: Consider a branching chain originated by a single element, where each individual element either replaces itself by k offspring with probability p_k, or fails to replace itself (i.e., dies out without producing an offspring) with probability $p_0 = 1 - p_k, 0 < p_0 < 1$. Obviously,

$\mu = 0 \times (p_0) + k \times (p_k)$

$= kp_k$

Hence

$E(X_n) = \mu^n$

$= (k\, p_k)^n$

$\to 0\,, 1\ or\ \infty\ \ as\ n \to \infty$, depending on whether $k\, p_k < 1$, or =1, or > 1.

Here,

$g(s) = E(s^X)$

$= s^0(p_0) + s^k(p_k)$

$= p_0 + p_k s^k$

Now the equation $g(s) = s$ would imply that we need to solve $s - p_k s^k - p_0 = 0$, i.e., $s\left(1 - p_k s^{(k-1)}\right) = p_0$ for s. For $k > 1$, from fundamental theorem, as $p_0 < 1$ (since $p_1 = 0$), the probability q of extinction of the process is 1, if and only, if $\mu = k\, p_k \leq 1$.

Example 2.5.5.: Consider the probability distribution of the number of offspring in a branching chain originated by a single individual, where each individual element generates three offspring, following the Poisson law, with an average $\lambda > 0$, which is given by

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

Thus,

$$g(s) = E(s^X)$$

$$= s^0(p_0) + s(p_1) + s^2(p_2) + \cdots$$

$$= \left\{s^0\left(\frac{\lambda^0}{0!}\right) + s\left(\frac{\lambda^1}{1!}\right) + s^2\left(\frac{\lambda^2}{2!}\right) + \cdots\right\} e^{-\lambda}$$

$$= e^{s\lambda} e^{-\lambda}$$

$$= exp\{-\lambda(1-s)\}$$

And

$$\mu = \lambda . E(X_n)$$

$$= \lambda^n$$

$\rightarrow 0$ or 1, or ∞ according as $\lambda < 1$, or $= 1$, or > 1.

Here

$$g(s) = exp\{-\lambda(1-s)\}$$

Notice that

$$p_0 + p_1 = e^{-\lambda} + \lambda e^{-\lambda}$$

$$= (1+\lambda)e^{-\lambda}$$

$$\leq 1$$

So that from fundamental theorem, the probability q of extinction of the process is 1 provided $\lambda \leq 1$; otherwise $q < 1$.

Example 2.5.6.: A branching process starts from 10 individuals, and each reproduces according to the probability distribution $(p_0, p_1, p_2, \dots)$, where

$$p_0 = \frac{1}{4}, p_1 = \frac{1}{4}, p_2 = \frac{1}{2}, p_n = 0 \; for \; n > 2$$

Find the extinction probability for the whole population.

Solution: The extinction probability for the population starting from each of the 10 initial individuals is given by the smallest solution of

$$\frac{1}{4} + \frac{1}{4}s + \frac{1}{2}s^2 = s$$

which is $s = \frac{1}{2}$. Since the population starts with 10 individuals, the extinction probability for the whole population is $\left(\frac{1}{2}\right)^{10} = \frac{1}{1024}$.

Example 2.5.7.: Let the distribution of the number of offspring be geometric with $p_k = p(1-p)^k, k = 0,1,2, \dots \quad ; 0 < p < 1$. Find the probability of extinction of the process.

Solution: We know that

$$g(s) = \sum_{k=0}^{\infty} p_k \, s^k$$

$$= \sum_{k=0}^{\infty} p(1-p)^k \, s^k$$

$$= p \sum_{k=0}^{\infty} [(1-p)s]^k$$

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

Consider,

$g(s) = s$

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

$\Rightarrow p = s - (1-p)s^2$

$\Rightarrow (1-p)s^2 - s + p = 0$

$\Rightarrow s = \frac{-(-1) \pm \sqrt{(-1)^2 - 4(1-p)p}}{2(1-p)} \quad \left\{\because ax^2 + bx + c = 0 \,; \Rightarrow x = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a}\right\}$

$\Rightarrow s = \frac{1 \pm \sqrt{1 - 4(1-p)p}}{2(1-p)}$

$\Rightarrow s = \frac{1 \pm \sqrt{\{p + (1-p)\}^2 - 4(1-p)p}}{2(1-p)}$

$\Rightarrow s = \frac{1 \pm \sqrt{p^2 + (1-p)^2 + 2p(1-p) - 4(1-p)p}}{2(1-p)}$

$\Rightarrow s = \frac{1 \pm \sqrt{p^2 + (1-p)^2 - 2p(1-p)}}{2(1-p)}$

$\Rightarrow s = \frac{1 \pm \sqrt{\{p - (1-p)\}^2}}{2(1-p)}$

$\Rightarrow s = \frac{1 \pm \{p - (1-p)\}}{2(1-p)}$

$\Rightarrow s = \frac{1 \pm \{2p - 1\}}{2(1-p)}$

$\Rightarrow s = \frac{1 + \{2p - 1\}}{2(1-p)}, \frac{1 - \{2p - 1\}}{2(1-p)}$

$\Rightarrow s = \frac{p}{(1-p)}, 1$

Hence the probability of extinction $= min\left\{\frac{p}{(1-p)}, 1\right\}$

$$= \begin{cases} \frac{p}{(1-p)} & if \quad 0 \leq p < \frac{1}{2} \\ 1 & if \quad \frac{1}{2} \leq p \leq 1 \end{cases}$$c

2.6 Total Progeny in Branching Process

Let X_n be the size of n^{th} generation, $n = 0,1,2,...$ and $X_0 = 1$.

The random variables $Z_n, n \geq 1$ defined as

$Z_n = 1 + X_1 + X_2 + \cdots + X_n$

where (X_n) is a branching process.

Here Z_n is the total number of descendants up to and including the n^{th} generation. Let $Y = \sum_{n=0}^{\infty} X_n = \lim_{n\to\infty} Z_n$ be the total number of individuals ever born.

The probability generating function (PGF) of the total progeny in a simple branching process is an essential tool for understanding the distribution of the total number of individuals ever born in the process. Here, we prove the PGF of the total progeny.

PGF of Total Progeny:

Let $R_n(s)$ be the PGF of Z_n and $g(s)$ be the PGF of the offspring distribution.

Theorem 2.6.1.: The probability generating function $R_n(s)$ of Z_n satisfies the recurrence relation

$R_n(s) = s\, g\{R_{n-1}(s)\}$

Proof: Let $Y_n = X_1 + X_2 + \cdots + X_n$ and let $G_n(s)$ be its p.g.f., then

$$R_n(s) = \sum_{j=0}^{\infty} P[Z_n = j]s^j$$

$$= \sum_{j=0}^{\infty} P[1 + Y_n = j]s^j$$

$$= \sum_{j=1}^{\infty} P[Y_n = j - 1]s^j$$

$$= s\sum_{j=1}^{\infty} P[Y_n = j - 1]s^{j-1}$$

$$= s\sum_{j=0}^{\infty} P[Y_n = j]s^j$$

$$\because G_n(s) = \sum_{j=0}^{\infty} P[Y_n = j]s^j$$

$$\Longrightarrow R_n(s) = sG_n(s) \qquad (2.9)$$

Now by conditioning on the size X_1 of the 1^{st} generation, we get

$$P[Y_n = k] = P(X_1 + X_2 + \cdots + X_n = k)$$

$$= \sum_{i=0}^{\infty} P[X_1 + X_2 + \cdots + X_n = k | X_1 = i] \times P[X_1 = i]]$$

$$= \sum_{i=0}^{\infty} P[\text{total number of descendants in the succeeding } (n - 1) \text{ generations following the } 1^{\text{st}} \text{ generation is } (k - i) | X_1 = i] \times P[X_1 = i]$$

i.e.

$$P[Y_n = k] = \sum_{i=0}^{\infty} P[X_2 + \cdots + X_n = k - i | X_1 = i] \times P[X_1 = i]]$$

If the process starts with one ancestor, then the probability of having r descendants in succeeding m generations is the coefficient of $s^r in\ G_m(s)$; and if it starts with i ancestors then the probability of having r descendants in succeeding m generations will be the coefficient of $s^r in\ [G_m(s)]^i$. Thus

$$P[Y_n = k] = \sum_{i=0}^{\infty} \left[Coefficient\ of\ s^{(k-i)} in\ \{G_{n-1}(s)\}^i\right] \times p_i$$

$$= \sum_{i=0}^{\infty} \left[Coefficient\ of\ s^{(k)} in\ \{s\ G_{n-1}(s)\}^i\right] \times p_i$$

$$= Coefficient\ of\ s^{(k)} in \sum_{i=0}^{\infty} p_i\ \{G_{n-1}(s)\}^i$$

$$\because g(s) = \sum_{i=0}^{\infty} p_i s^i \Rightarrow g\{sG_{n-1}(s)\} = \sum_{i=0}^{\infty} p_i \{sG_{n-1}(s)\}^i$$

Hence $P[Y_n = k] = Coefficient\ of\ s^{(k)} in\ \ g\{sG_{n-1}(s)\}$

$$\therefore G_n(s) = \sum_{k=0}^{\infty} P[Y_n = k] s^k$$

$$= g\{sG_{n-1}(s)\}$$

From (2.9), we have

$$R_n(s) = sG_n(s)$$

$$= s\ g\left[R_{(n-1)}(s)\right]$$

Hence proved the theorem.

For example, if $n = 1$, then

$$R_1(s) = s\, g[R_0(s)] \qquad (2.10)$$

where

$$R_0(s) = \sum_{k=0}^{\infty} P[Z_0 = k]s^k$$

$$= \sum_{k=0}^{\infty} P[X_0 = k]s^k$$

$$= s$$

Using (2.10), we get

$$R_1(s) = s\, g(s)$$

$$= s\sum_{j=0}^{\infty} p_j s^j$$

$$= \sum_{j=0}^{\infty} p_j s^{j+1}$$

$$\Rightarrow \sum_{j=0}^{\infty} P[Z_1 = j]s^j$$

$$= \sum_{j=0}^{\infty} p_j s^{j+1}$$

$$= p_0 s + p_1 s^2 + \cdots$$

$$\Rightarrow P[Z_1 = 0] + P[Z_1 = 1]s + P[Z_1 = 2]s^2 + \cdots = p_0 s + p_1 s^2 + \cdots$$

$$\left.\begin{array}{l} P[Z_1 = 0] = 0 \\ \Rightarrow P[Z_1 = 1] = p_0 \\ P[Z_1 = 2] = p_1 \end{array}\right\}$$

$$\Rightarrow P[Z_1 = j] = \begin{cases} p_{j-1}\,, & j = 1,2 \\ 0 & j = 0 \end{cases}$$

Example 2.6.2.: Let

$$p_k = \begin{cases} \frac{1}{4} & if \;\; k = 0 \\ \frac{k}{4} & if \;\; k = 1,2 \end{cases}$$

$$g(t) = \sum_{k=0}^{\infty} p_k \, t^k$$

$$= p_0 + p_1 t + p_2 t^2$$

$$= \frac{1}{4} + \frac{1}{4} t + \frac{1}{2} t^2$$

Consider,

$$s \, g(t) = t$$

$$\Rightarrow s\left\{\frac{1}{4} + \frac{1}{4} t + \frac{1}{2} t^2\right\} - t = 0$$

$$\Rightarrow s + st + 2st^2 = 4t$$

$$\Rightarrow 2st^2 + (s-4)t + s = 0$$

$$\Rightarrow t = \frac{-(s-4) \pm \sqrt{(s-4)^2 - 4 \times 2s \times s}}{2 \times 2s}$$

$$\Rightarrow t = \frac{(4-s) \pm \sqrt{(s-4)^2 - 4 \times 2s \times s}}{2 \times 2s}$$

$$\Rightarrow t = \frac{(4-s) \pm \sqrt{-7s^2 - 8s + 16}}{4s}$$

Hence the probability of extinction $= \begin{cases} \frac{(4-s) + \sqrt{-7s^2 - 8s + 16}}{4s} \\ \frac{(4-s) - \sqrt{-7s^2 - 8s + 16}}{4s} \end{cases}$

2.7 Summary

In this unit, we have discussed the following points:

- The branching process is a Markov chain with a transition probability matrix $((P_{ij}))$.
- The probability generating function of the offspring distribution is given by

$$g(s) = \sum_{k=0}^{\infty} p_k \, s^k; \quad 0 \leq s \leq 1$$

- The expectations and variance for the branching process are proved.
- The extinction probability is given by $q = g(q)$
- The fundamental theorem for the probability of extinction given byon is state and proved.

2.8 Self-Assessment Exercise

1. If X_n denotes the number of individuals in the n^{th} generation of a simple branching process, then find the probability generating function of X_n. Using the probability generating function, obtain the mean and variance of X_n.

2. State and prove the fundamental theorem of probability of extinction of a simple branching process.

3. Let $X = (X_n \, n \in N)$ be the Galton-Watson branching process with U having the distribution $\{p_n\}n \in N$. For the following $\{p_n\}n \in N$ and the probability of ultimate extinction and the expected number of members of the n^{th} generation:

   i. $p_0 = p_1 = \frac{1}{5}, \; p_2 = \frac{3}{5}, p_k = 0 \; \forall k = 3,4, \ldots$

   ii. $p_0 = \frac{1}{12}, \; p_1 = \frac{5}{12} \; p_2 = \frac{1}{2}, p_k = 0 \; \forall k = 3,4, \ldots$

   iii. $p_0 = \frac{1}{2}, \; p_1 = 0 \; p_2 = 0 \; p_3 = \frac{1}{2}, p_k = 0 \; \forall k = 4,5 \ldots$

iv. $p_0 = \frac{1}{10},\ p_1 = \frac{2}{5}\ p_2 = \frac{1}{2}, p_k = 0\ \forall k = 3,4, \ldots$

v. $p_k = \left(\frac{1}{2}\right)^{k+1}; k = 0,1, \ldots$

vi. $p_k = pq^k\ \forall k = 0,1, \ldots; p + q = 1.$

4. Suppose that the family sizes X have geometric distribution on $\{0, 1, 2, \ldots\}$ with parameter $p \in (0, 1)$. Find the distribution of X_n, *i.e.*, the probabilities $P(X_n = j) \forall\ j = 0, 1, 2, \ldots$
5. Suppose in a branching process, the offspring distribution is as follows:

$p_k = {}^{N}C_k\ p^k q^{(N-k)}, q = 1 - p, 0 < p < 1, k = 0,1,2, \ldots.$

Discuss the probability of extinction of this branching process.

2.9 References

- Cinlar, E. (1975): "Introduction to Stochastic Processes", Prentice Hall.
- Feller, W. (1968): "Introduction to Probability and Applications", New Age India International.
- Fortet, R. (1951): "Random functions from a Poisson process", Proc. 2nd Berkeley Symp.
- Harris, T. E. (1963): "The Theory of Branching Processes", Springer Verlag.
- Hoel, P. G., Port, S. C. and Stone, C. J. (1991): "Introduction to Stochastic Processes", University Book Stall.
- Kac, M. (1947): "Random Walk and the theory of Brownian motion", Amer. Math. Monthly, Vol. 54, p. 369.
- Karlin, S. and Taylor, H. M. (1995): "A First Course in Stochastic Processes", Academic Press.
- Kennedy, D. P. (1975): "The Galton-Watson process conditioned on the total progeny", J, AppL Prob. 12, 800-806. 16.Medhi, J. (2002): Stochastic Process, New Age International Publishers.
- Osborne, M. F. M. (1959): "Reply to comments on Brownian motion in the stock market", Oper. Res., Vol. 7, pp. 807-811.

- Wax, N. (1954): “Noise and Stochastic Processes”, New York: Dover.
- Wold, H. (1954): “A Study in the Analysis of Stationary Time Series”, 2nd ed. (with an appendix by Peter Whittle). Uppsala, Swedeu.
- Wonham, W. M., and A. T. Fuller. (1958): “Probability densities of the smoothed random telegraph signal. Jour”, of Electronics and Control, Vol. 4, pp. 567-576.

2.10 Further Reading

- Grimmett, G., & Stirzaker, D. (2001): “Probability and Random Processes”, Oxford University Press.
- Karatzas, I., & Shreve, S. E. (1991): “Brownian Motion and Stochastic Calculus”, Springer- Verlag.
- Lefebvre, M. (2006): “Applied Stochastic Processes”, Springer.
- Levin, D. A., Peres, Y., & Wilmer, E. L. (2009): “Markov Chains and Mixing Times”, American Mathematical Society.
- Resnick, S. I. (1992): “Adventures in Stochastic Processes”, Birkhäuser.
- Ross, S. M. (2014): “Stochastic Processes”, Wiley.
- Taylor, H. M., & Karlin, S. (1998): “An Introduction to Stochastic Modelling”, Academic Press.

UNIT– 3 WEINER PROCESS

Structure

3.1 Introduction

The Wiener process can be considered a continuous version of the simple random walk. This continuous-time stochastic process is a highly studied and used object. It plays a key role in fields related to stochastic processes such as stochastic calculus and the theories of Markov processes, martingales, Gaussian processes etc.

The Wiener process is named after mathematician Norbert Wiener. It is a specific kind of Brownian Motion used in finance and mathematics. It is characterized by a mean of zero and a variance that increases linearly with time (often standardized to one for each time interval), making it a standardized model for the random walk hypothesis.

Brownian Motion was named after the botanist Robert Brown, who in the early 19th century, observed the erratic movement of pollen particles when suspended in water under a microscope. This random motion, known as “Brownian Motion,” became a fundamental concept in the world of physics and was later used to provide empirical evidence of the existence of atoms.

While Brownian Motion was initially an observational phenomenon, the Wiener Process was its formal mathematical representation.

3.2 Objectives

After reading this unit, you should be able to:

- Follow the concept of Wiener process
- Understand the mean value function and covariance kernel of the Wiener process,
- Explain Martingales with examples,
- Define Stopping times and proposition.
- State and prove the Optional sampling theorem.

3.3 Wiener Process

A real-valued stochastic process $\{W(t): t \geq 0\}$ is said to be a Wiener process if

(i) $\{W(t): t \geq 0\}$ has stationary independent increments,

(ii) $W(t)\ \forall\ t > 0$ is normally distributed,

(iii) $E[W(t)] = 0$ for all $t > 0$,

(iv) $W(0) = 0$.

Consider $W(t) - W(s)$ for any s, t. Since $W(t) - W(s)$ is normal, its probability law is determined by its mean and variance.

We have, $E[W(t) - W(s)] = 0$.

Therefore, the characteristic function of $\{W(t) - W(s)\}$ is

$$\phi_{\{W(t)-W(s)\}}(U) = \exp\left[-\frac{1}{2}U^2 Var\{W(t) - W(s)\}\right] \qquad (3.1)$$

Using the conditions (i), there is some positive constant σ^2, such that for $t \geq s \geq 0$,

$$Var\{W(t) - W(s)\} = \sigma^2|t - s| \qquad (3.2)$$

The probability law of a Wiener process is determined by the conditions (i)-(iv) up to a parameter σ^2.

Remark: The conditional probability density function of $W(t)$, given that $W(t_1) = w_1$ and $W(t_2) = w_2$, is normal with

$$mean = w_1 + \frac{w_2 - w_1}{t_2 - t_1}(t - t_1) \qquad for \quad t_1 < t < t_2, \qquad (3.3)$$

and

$$Variance = \sigma^2 \frac{(t_2 - t)(t - t_1)}{t_2 - t_1} \qquad (3.4)$$

3.3.1 Application of Wiener Process

1. The Black-Scholes Model, a ground-breaking formula for option pricing, relies heavily on the principles of Brownian Motion. The future price of the stock is uncertain and can be thought of as moving randomly—much like a pollen particle suspended in water.

2. The Wiener Process, with its mathematical rigor, provides the necessary foundation to model this randomness in a way that is suitable for financial calculations.

3. The Wiener Process is instrumental in risk management by offering a structured approach to model asset price dynamics. By quantifying the stochastic behaviour of asset prices, risk managers can devise robust risk mitigation strategies and asset allocation frameworks, ensuring the resilience of financial portfolios against market volatility.

4. In fixed income, the Wiener Process is applied in modelling the evolution of interest rates over time. Through stochastic differential equations driven by the Wiener Process, analysts can better anticipate interest rate trends, which are critical for bond pricing, interest rate derivatives, and overall fixed-income portfolio management.

5. Underpinning the random walk hypothesis, the Wiener Process aids in modeling asset price dynamics. By recognizing the stochastic nature of asset prices, financial analysts can develop forecasting models, enhancing the accuracy and reliability of market predictions.

6. Incorporating the Wiener Process in portfolio optimization helps in understanding the probable price paths of constituent assets. This understanding is vital for optimizing asset allocation, managing portfolio risk, and ultimately enhancing portfolio performance over time.

Example 3.3.1.: Suppose that liquid in a container is placed in a coordinate system, and at time 0, a pollen particle suspended in the liquid is at $(0, 0, 0)$, the origin. Let $W(t)$ be the x-coordinate of the position of the pollen after t minutes. Suppose that $\{W(t): t \geq 0\}$ is a Brownian motion with variance parameter 4 and, after one minute, the x-coordinate of the pollen's position is 2.

(i) What is the probability that after 2 minutes it is between 0 and 1?

(ii) What is the expected value and variance of the x-coordinate of the position of pollen after 30 seconds?

Solution: (i) The probability is

$$P[0 < W(2), 1 | W(1) = 2]$$
$$= P[-2 < W(2) - W(1) < -1 | W(1) = 2]$$
$$= P[-2 < W(2) - W(1) < -1]$$

The last step follows by the independent increments property of the Wiener process. Since $\{W(2) - W(1)\}$ is normal with mean 0 and variance $(t - s)\sigma^2 = (2 - 1) \times 4 = 4$.

Letting $Z \sim N(0,1)$, we get

$$P[-2 < W(2) - W(1) < -1] = P\left[\frac{-2 - 0}{2} < Z < \frac{-1 - 0}{2}\right]$$

$$= P[-1 < Z < -0.5]$$

$$= \Phi(-0.5) - \Phi(-1)$$

$$= 0.1523$$

(ii) Using (3.3) and (3.4) and putting the $t = \frac{1}{2},\ t_1 = 0, t_2 = 1, w_1 = 0, w_2 = 2$, we get

$$E[W(t)|W(t_1) = w_1 \ and \ W(t_2) = w_2]$$
$$= w_1 + \frac{w_2 - w_1}{t_2 - t_1}(t - t_1) \qquad \text{for} \quad t_1 < t < t_2,$$

$$E\left[W\left(\frac{1}{2}\right)\middle|W(0) = 0 \ \ and \ W(1) = 2\right] = 0 + \frac{2-0}{1-0}\left(\frac{1}{2} - 0\right)$$

$$= \frac{2}{2} = 1$$

$$Variance[W(t)|W(t_1) = w_1 \ and \ W(t_2) = w_2]$$
$$= \sigma^2 \frac{(t_2 - t)(t - t_1)}{t_2 - t_1}$$

$$Var\left[W\left(\frac{1}{2}\right)\middle|W(0) = 0 \ \ and \ W(1) = 2\right]$$
$$= 4 \times \frac{\left(1 - \frac{1}{2}\right)\left(\frac{1}{2} - 0\right)}{1 - 0}$$

$$= 4 \times \left(\frac{1}{2} \times \frac{1}{2}\right) = 1$$

3.4 Mean Value Function and Covariance Kernel of the Wiener Process

Let $\{W(t): t \geq 0\}$ be the Wiener process with parameter σ^2. Using (3.1) and (3.2) for all $t \geq 0$, we have

$$m(t) = E[W(t)] = 0,$$

$$Var[W(t)] = \sigma^2 t$$

The covariance kernel $K(s,t)$ for $s < t$:

$$Cov[W(s), W(t)] = Cov[W(s), W(t) - W(s) + W(s)]$$

$$= Cov[W(s), W(t) - W(s)] + Cov[W(s), W(s)]$$

$$= Var[W(s)]$$

$$= \sigma^2 s$$

We have utilized the result that $W(s)$ and $\{W(t) - W(s)\}$ are independent, and therefore have zero co-variance.

The covariance kernel of the Wiener process with parameter σ^2 is given by

$$K(s,t) = \sigma^2 \min(s,t) \qquad for\ all\ s,t \geq 0$$

3.5 Martingales

A stochastic process $\{M_t : t \in T\}$ is said to be Martingales if

(i) $E(|M_t|) < \infty, \forall\ t \in T$

(ii) $E\left(M_{n+1} \middle| M_{t_1} = a_1, M_{t_2} = a_2, \dots, M_{t_n} = a_n\right) = a_n$

For all values of $a_i's$ and $t_1 < t_2 < \cdots < t_n < t_{n+1}$.

Example 3.5.1.: Assume that a player plays against an infinitely rich opponent. He takes to one unit of money with probability p and loses one unit of money with probability $q = (1-p)$. Let Z_n be player gain in the n^{th} game and $M_n = Z_1 + Z_2 + \cdots + Z_n$ be a player cumulative gain in the first n games. Show that $\{M_n\}$ is a Martingale iff $p = q = \frac{1}{2}$.

Proof: We have

$$E(M_{n+1}|M_1, M_2, \dots, M_n) = E(M_n + Z_{n+1}|M_1, M_2, \dots, M_n)$$

$$(\because M_{n+1} = Z_1 + Z_2 + \cdots + Z_n + Z_{n+1} = M_n + Z_{n+1})$$

$$\Rightarrow E(M_{n+1}|M_1, M_2, \dots, M_n)$$

$= E(M_n|M_1, M_2, \dots, M_n) + E(Z_{n+1}|M_1, M_2, \dots, M_n)$

$\Rightarrow E(M_{n+1}|M_1, M_2, \dots, M_n)$

$= M_n + E(Z_{n+1})$

$= M_n + \{1 \times p + (-1) \times q\}$

$= M_n + p - q$

Hence $\{M_n : n \geq 1\}$ is martingale iff $p = q = \frac{1}{2}$.

Example3.5.2.: Let $\{Z_i : i = 1,2, \dots\}$ be iid random variables with mean 1 i.e. $E[Z_i] = 1$, $i = 1,2,$ ... and let

$$M_n = \prod_{i=1}^{n} Z_i$$

Then $\{M_n : n \geq 1\}$ is a martingale.

Proof: We have

$E(M_{n+1}|M_1, M_2, \dots, M_n)$

$= E(M_n \times Z_{n+1}|M_1, M_2, \dots, M_n)$

$(\because M_{n+1} = \prod_{i=1}^{n+1} Z_i = (\prod_{i=1}^{n} Z_i) Z_{n+1} = M_n \times Z_{n+1})$

$\Rightarrow E(M_{n+1}|M_1, M_2, \dots, M_n)$

$= E(M_n|M_1, M_2, \dots, M_n) \times E(Z_{n+1}|M_1, M_2, \dots, M_n)$

$(\because Z_{n+1}\ is\ independent\ of\ M_n = \prod_{i=1}^{n} Z_i\ )$

$\Rightarrow E(M_{n+1}|M_1, M_2, \dots, M_n)$

$= M_n \times E(Z_{n+1})$

$= M_n \times 1$

$= M_n$

Hence $\{M_n : n \geq 1\}$ is martingale.

Example 3.5.3.: Consider an urn with a red ball and a blue ball. Suppose initially the urn contains one ball of each color. At each time step, a ball is chosen at random from the urn. If a red ball is chosen put back and another red ball is added to the urn. Similarly, if a green ball is chosen, return it to the urn along with another green ball. Let X_n be the number of red balls in the urn after n draws. Then $X_0 = 1$ and X_n is a Markov chain with transitions

$$P[X_{n+1} = k+1 | X_n = k] = \frac{k}{n+2}$$

$$P[X_{n+1} = k | X_n = k] = \frac{n+2-k}{n+2}.$$

Proof: Let, $M_n = \frac{X_n}{n+2}$

be the fraction of red balls after n draws. Then M_n is a Martingale.

Notice that, $P[X_{n+1} | X_n] = X_n + \frac{X_n}{n+2}$

Let $\mathcal{F}_n$ is the information obtained from $M_0, M_1, \ldots, M_n$. i.e.

$E[M_{n+1} | \mathcal{F}_n] = M_n$

Using conditional expectation

$E[M_{n+2} | \mathcal{F}_n] = E[E\{M_{n+2} | \mathcal{F}_{n+1}\} | \mathcal{F}_n]$

$= E[M_{n+1} | \mathcal{F}_n]$

$= M_n$

Therefore

$$E[M_{n+1}|\mathcal{F}_n] = E[(n+3)^{-1}X_{n+1}|X_n]$$

$$= \frac{1}{n+3}\left[X_n + \frac{X_n}{n+2}\right]$$

$$= \frac{X_n}{n+2}$$

$$= M_n$$

1. A process M_n with $E(|M_n|) < \infty$ is called a sub-martingale with respect to $X_0, X_1, \ldots$ if $\forall m < n, E(M_n|\mathcal{F}_n) \geq M_n$.

2. A process M_n with $E(|M_n|) < \infty$ is called a super martingale with respect to $X_0, X_1, \ldots$ if $\forall m < n, E(M_n|\mathcal{F}_n) \leq M_n$.

3.6 Stopping Times

A positive integer-valued, random variable T is said to be a random time for the process

$\{M_n, n \geq 1\}$ if the event $\{T = n\}$ is determined by the random variables $M_1, M_2, \ldots, M_n$. If $P[T < \infty] = 1$, then the random time T is said to be a stopping time.

Let T be a random time for the process $\{M_n, n \geq 1\}$ and let

$$\bar{M}_n = \begin{cases} M_n & if\ n \leq T \\ M_T & if\ n > T \end{cases}$$

$\{\bar{M}_n, n \geq 1\}$ is called the stopped process.

Let T be a stopping time and F_T be the class of sets, i.e.,

$$F_T = \{A \in F_\infty : A \cap (T \leq t) \in F_t\ \forall\ t \in R\}$$

Proposition: If T is a stopping time for the martingale $\{M_n\}$, then the stopped process $\{\bar{M}_n\}$ is a martingale.

Proof: Let

$$\delta_n = \begin{cases} 1 & if\ T \geq n \\ 0 & if\ T < n \end{cases}$$

If we have not yet stopped after observing $M_1, M_2, \dots, M_{n-1}$. Insist that

$$\bar{M}_n = \bar{M}_{n-1} + \delta_n(M_n - M_{n-1})$$

Case I: If $T \geq n$, then $\bar{M}_n = M_n$ and $\bar{M}_{n-1} = M_{n-1}$.

Case I: If $T < n$, then $\bar{M}_n = \bar{M}_{n-1}$.

Consider

$$E[\bar{M}_n | M_1, M_2, \dots, M_{n-1}] = E[\bar{M}_{n-1} + \delta_n(M_n - M_{n-1}) | M_1, M_2, \dots, M_{n-1}]$$

$$= \bar{M}_{n-1} + \delta_n E[(M_n - M_{n-1}) | M_1, M_2, \dots, M_{n-1}]$$

$$= \bar{M}_{n-1}.$$

where $\{M_n\}$ is a martingale.

Since the stopped process is a martingale, and since $\bar{M}_1 = M_1$, then

$$E[\bar{M}_n] = E[M_1] \quad for\ all\ n$$

Suppose that T is a stopping time, i.e., $P[T < \infty] = 1$.

We have

$$\bar{M}_n = \begin{cases} M_n & if\ n \leq T \\ M_T & if\ n > T \end{cases}$$

It follows that $\bar{M}_n$ equals M_T when n is sufficiently large i.e.,

$$\bar{M}_n \to M_T \quad \text{as } n \to \infty, \text{with probability 1.}$$

or

$$E[\bar{M}_n] \to E[M_T] \quad \text{as } n \to \infty \qquad (3.5)$$

Further, $E[\bar{M}_n] = E[M_1]$ for all n, then $E[M_T] = E[M_1]$ for all n.

Remark: If either:

(i) $\bar{M}_n$ are uniformly bounded, or

(ii) T is bounded, or,

(iii) $E[T] < \infty$, and there is an $M < \infty$ such that

$E[|M_{n+1} - M_n||M_1, M_2, \dots, M_n] < M,$

Then (3.5) is valid. Thus $E[M_n] = E[M_1]$.

Theorem 3.6.1.: Suppose T is a stopping time. If $A \in \mathcal{F}_T$ then

$A \cap (T \leq t) \in \mathcal{F}_{n_t} \; \forall \; t \in R$ (3.6)

Conversely, if $A \in \mathcal{F}_{n_\infty}$ and (8.6) is valid, then $A \in \mathcal{F}_T$.

Proof: Let $A \in \mathcal{F}_T$, $t \in R$, then

$$A \cap (T < t) = \bigcup_{k=1}^{\infty} \{A \cap (T < t - 1/k)\} \in \mathcal{F}_t$$

Conversely, if $A \in F_\infty$ and (8.5) is valid, then for any $t \in R$ and for any natural m.

$$A \cap (T \leq t) = \bigcap_{k=m}^{\infty} \{A \cap (T < t + 1/k)\} \in \mathcal{F}_{t+1/m}$$

Therefore

$$A \cap (T \leq t) \in \bigcap_{m=1}^{\infty} \mathcal{F}_{t+\frac{1}{m}} = \mathcal{F}_t$$

3.7 Optional Sampling Theorem

Theorem 3.7.1.: Suppose $M_0, M_1, \cdots$ is a martingale with respect to $X_0, X_1, \cdots$ and T is a stopping time satisfying $P(T < \infty) = 1$,

$$E(|M_n| < \infty) \tag{3.7}$$

and

$$\lim_{n\to\infty} E\left(|M_n| I(T > n)\right) = 0 \tag{3.8}$$

Then, $E(M_T) = E(M_0)$. Here $I(\cdot)$ is an indicator function.

Proof: Let $\mathcal{F}_n$ be the information contained in $X_0, X_1, \cdots, X_n$ and $I(T > n)$ be the indicator function of event $\{T > n\}$, which is measurable with respect to $\mathcal{F}_n$ (Since we need only the information up to time n to determine if we have stopped by time n). M_T is the random variable which equals M_j if $T = j$. We can write

$$M_T = \sum_{j=0}^{K} M_j I(T = j)$$

$$E(M_T|\mathcal{F}_{K-1}) = E(M_K I(T = K)|\mathcal{F}_{K-1}) + \sum_{j=0}^{K} E\left(M_j I(T = j)|\mathcal{F}_{K-1}\right)$$

For $j \leq (K-1)$, $M_j I(T = j)$ $\mathcal{F}_{K-1}$ is measurable. Hence

$$E\left(M_j I(T = j)|\mathcal{F}_{K-1}\right) = M_j I(T = j)$$

Since T is known to be no more than K, then event $\{T = K\}$ is the same as the event $\{T > K-1\}$. The latter event is measurable with respect to $\mathcal{F}_{K-1}$. Hence

$$E(M_K I(T = K)|\mathcal{F}_{K-1})$$

$$= E(M_K I(T > K-1)|\mathcal{F}_{K-1})$$

$$= I(T > K-1)E(M_K|\mathcal{F}_{K-1})$$

$$= I(T > K-1)E(M_{K-1})$$

Therefore

$$E(M_T|\mathcal{F}_{K-1})$$

$$= I(T > K-1)E(M_{K-1}) + \sum_{j=0}^{K-1} E\left(M_j I(T=j)\right)$$

$$= I(T > K-2)E(M_{K-2}) + \sum_{j=0}^{K-2} E\left(M_j I(T=j)\right)$$

$$E(M_T|\mathcal{F}_{K-2})$$

$$= E(E(M_K|\mathcal{F}_{K-1})|\mathcal{F}_{K-2})$$

$$= I(T > K-3)E(M_{K-1}) + \sum_{j=0}^{K-3} E\left(M_j I(T=j)\right)$$

We continue this process until we get $E(M_T|\mathcal{F}_0) = M_0$. Now, consider the stopping time $T_n = min(T, n)$. Then

$$M_T = M_{T_n} + M_T I(T > n) - M_n I(T > n)$$

$$E(M_T) = E\left(M_{T_n}\right) + E\left(M_T I(T > n)\right) - E\left(M_n I(T > n)\right)$$

Since T_n is a bounded stopping time, hence $E\left(M_{T_n}\right) = M_0$. Then $P(T > n) \to 0$ as $n \to \infty$. If $E|M_T| < \infty$ then $E\left(|M_T| I(T > n)\right) \to 0$. If M_n and T are given so that

$\lim_{n\to\infty} E\left(|M_T| I(T > n)\right) = 0$, then, $E(M_T) = E(M_0)$. Hence the theorem follows■

The third term $E\left(M_T I(T > n)\right)$ in $E(M_T)$ is troublesome. There are many examples of interest where the stopping time T is not bounded.

Consider the Example. Let $\{T > n\}$ be the event that the first n tosses are tails and has probability 2^{-n}. If this event occurs, the bettor has lost a total $(2^n - 1)$ rupees, *i.e.*, $M_n = 1 - 2^n$. Hence

$$E\left(M_T I(T > n)\right) = 2^{-n}(1 - 2^n)$$

which does not go to 0 as $n \to \infty$.

Example 3.7.1.: (Gambler's Ruin Problem Revisited)

Let X_n be a simple random walk $p = \frac{1}{2}$ on $\{0,1,2, \ldots\}$ with absorbing barriers. Suppose $X_0 = a$ and $M_n \equiv X_n$. Then, X_n is a martingale. Let stopping time $T = min\{j : X_j = 0 \text{ or } N\}$ and since X_n is bounded, we have

$E(M_T) = E(M_0) = a.$

But in this case

$E(M_T)$
$= 0P(X_T = 0) + NP(X_T = N)$
$= NP(XT = N)$

Therefore,

$$P(X_T = N) = \frac{a}{N}$$

This gives another derivation of gambler's ruin result for simple random walk.

Example 3.7.2.: Let X_n be as in Example 8 and $M_n = X_n^2 - n$. Then, M_n is a martingale with respect to X_n.

$E(M_{n+1}|\mathcal{F}_n) = E(X_{n+1}^2 - (n+1)|\mathcal{F}_n)$
$= X_n^2 + 1 - (n+1)$
$= M_n.$

Let stopping time $T = min\{j: X_j = 0 \text{ or } N\}$ and since M_n is not a bounded martingale so it is not immediate that (3.7) and (3.8) hold. However there exists $C < \infty$ and $\rho < 1$ such that

$P(T > n) \leq C\rho^n.$

Since $|M_n| \leq N^2 + n$,

$E(|M_n|) < \infty$

and

$E\big(|M_n| I(T > n)\big)$
$\leq C\rho^n(N^2 + n) \to 0$

Hence, optional sampling theorem holds and $E(M_T) = E(M_0) = a^2$.

$E(M_T) = E(X_T^2) - E(T)$
$= N^2 P(X_T = N) - E(T)$
$= aN - E(T)$

Hence

$E(T) = aN - a^2 = a(N - a)$.

3.8 Summary

In this unit, we have covered the following points:

- We have explained Wiener process with examples.
- We have discussed the Mean value function and covariance kernel of Wiener process.
- We have explained Martingales.
- We have explained Stopping times with example.
- We have stated and proved the Optional sampling theorem.

3.9 Self-Assessment Exercise

1. State and prove Optional Sampling Theorem.
2. Consider the Polya urn problem. Let M_n be the proportion of red balls after n draws (starting with one red and one green ball). Prove by induction on n that

$$P\left(M_n = \frac{k}{n+2}\right) = \frac{1}{n+1} \quad , \quad k = 1,2, \ldots,$$

3. Consider a biased random walk on the integers with probability $p < \frac{1}{2}$ of moving to the right and probability $(1-p)$ of moving to the left. Let Z_n be the value at time n and assume that $Z_0 = a$, where $0 < a < N$. Show that

$M_n = \left\{\frac{1-p}{p}\right\}^{Z_n}$ is a martingale.

4. Let X_n be the number of individuals in the n^{th} generation of a branching process in which each individual produces offspring from a distribution with mean μ and variance σ^2. We have seen $M - n = \mu^{-n} X_n$ is a martingale. Let $\mathcal{F}_n$ denote the information contained in $X_0, X_1, \ldots, X_n$. Show that

$E(X_{n+1}^2 | \mathcal{F}_n) = \mu^2 X_n^2 + \sigma^2 X_n$.

5. Let $\{W(t): t > 0\}$ be the Wiener process with parameter σ^2. $X(t) = At + W(t)$, in which A is a positive constant. Compute:

(i) The mean value function $m(t) = E[X(t)]$

(ii) The covariance kernel $K(s,t) = Cov[X(s), X(t)]$.

3.10 References

- Medhi, J. (2002): Stochastic Process, 2nd edition, New Age International Publishers.
- Osborne, M. F. M. (1959): Reply to comments on Brownian motion in the stock market, Oper. Res., Vol. 7, pp. 807-811.
- Wonham, W. M., and A. T. Fuller. (1958): Probability densities of the smoothed random telegraph signal. Jour, of Electronics and Control, Vol. 4, pp. 567-576.
- Wax, N. (1954): Noise and Stochastic Processes. New York: Dover.
- Wold, H. (1954): A Study in the Analysis of Stationary Time Series, 2nd ed. (with an appendix by Peter Whittle). Uppsala, Swedeu.
- Fortet, R. (1951): Random functions from a Poisson process, Proc. 2nd Berkeley Symp.

- Kac, M. (1947): Random walk and the theory of Brownian motion, Amer. Math. Monthly, Vol. 54, p. 369.
- Wang, Ming Chen, and G. E. Uhlenbeck. (1945): On the theory of the Brownian motion II, Rev. Mod. Phys., Vol. 17, pp. 323-342.

3.11 Further Reading

- Ross, S. M. (2014). Stochastic Processes. Wiley.
- Levin, D. A., Peres, Y., & Wilmer, E. L. (2009). Markov Chains and Mixing Times. American Mathematical Society.
- Lefebvre, M. (2006). Applied Stochastic Processes. Springer.
- Grimmett, G., & Stirzaker, D. (2001). Probability and Random Processes. Oxford University Press.
- Taylor, H. M., & Karlin, S. (1998). An Introduction to Stochastic Modelling. Academic Press.
- Resnick, S. I. (1992). Adventures in Stochastic Processes. Birkhäuser.
- Karatzas, I., & Shreve, S. E. (1991). Brownian Motion and Stochastic Calculus. Springer-Verlag.

UNIT– 4 RENEWAL PROCESS

Structure

4.1 Introduction

Renewal processes are a fundamental tool in stochastic modeling, providing a way to analyze and predict the behavior of systems that experience recurring events over time.

A renewal process describes the times at which a certain event, such as a system reset or renewal, occurs. This process is widely used in fields like operations research, queuing theory, and reliability engineering to model events that occur randomly over time. This process is a special case of a counting process, which tracks the number of events that have occurred by a certain time. Specifically, in a renewal process, the times between successive events (also called interarrival times or interrenewal times) are independent and identically distributed (i.i.d.) random variables. The process "renews" itself every time an event occurs, meaning that the statistical properties of the process reset after each event.

The renewal process has been used in many areas, such as physics, biology, engineering, reliability, queueing, finance, and others. For example, in reliability modelling, a renewal process models the failure process of a system that either, at failure, is replaced with a new one, or, after repair, is “as good as new”.

4.2 Objectives

After studying this unit, you should be able to:

- Identify Renewal process and define Distribution of renewal process with example.
- Explain asymptotic distribution of renewal process.
- State and prove Elementary renewal theorem with example.
- Describe equilibrium renewal process.

4.3 Distribution and Asymptotic Distribution of Renewal Process

Let $\{X_n, n \geq 1\}$ be a sequence of non-negative i.i.d. random variables with

$P[X_1 = 0] < 1, \mu = E[X_1]$ and $F(x) = P[X_1 \leq x]$.

Suppose that $S_n = X_1 + X_2 + \cdots + X_n$ for $n \geq 1$ with distribution function

$F_n(x) = P[S_n \leq x]$.

Let $N(t) = \max\{n: S_n \leq t\}$ be the number of occurrences in the interval $(0, t)$. Here S_n is the time epoch of n^{th} renewal.

Then the process $\{N(t): t \geq 0\}$ is called a renewal process. $N(t)$ gives the number of renewals in $(0, t)$.

X_1 is called the inter renewal time or inter arrival time.

X_n is called the inter renewal time between the $(n-1)^{th}$ and n^{th} renewal.

We know that $S_n = X_1 + X_2 + \cdots + X_n$, where $X_n's$ are i.i.d. random variables.

$$\frac{S_n - E(S_n)}{n} \to 0$$

$$\Rightarrow \frac{S_n}{n} \to \frac{E(S_n)}{n}$$

$$= \frac{nE(X_1)}{n}$$

$$= \mu \qquad \text{as } n \to \infty$$

Since $\mu > 0$, it implies that $S_n \to \infty$ as $n \to \infty$.

Thus S_n can be less than t for almost a finite number of values of n. Thus,

$N(t) = \max\{n: S_n \leq t\}$

Example: Consider the Poisson process with rate parameter λ. The waiting times are independent, exponential random variables with parameter λ and $N(t)$ is the Poisson process. In this case $\mu = 1/\lambda$.

Example: Let Y_n be an irreducible, positive recurrent, discrete-time Markov chain staring in state x. Let

$X_1 = \min\{n > 0: Y_n = x\}$

And

$S_n = \min\{n > 0 : Y_{X_1+X_2+\cdots+X_n} = x\}$.

Example: Let Y_t be an irreducible, positive recurrent, continuous time Markov chain staring in state x. Define

$R_1 = \inf\{t > 0: Y_t \neq x\}$

$Q_1 = \inf\{t > R_1: Y_t = x\}$

Then $X_1 = R_1 + Q_1$ and in general

$R_i = \inf\{t > 0: Y_{X_1+X_2+\cdots+X_{i-1}+t} \neq x\}$

$Q_i = \inf\{t > 0: Y_{X_1+X_2+\cdots+X_{i-1}+R_i+t} = x\}$

Then $T_i = R_i + Q_i$

The random variables R_i are exponential with parameter $\lambda(x)$, the rate at which the chain is changing from state x.

Example: **Distribution of Renewal Process**

We have

$N(t) \geq n \quad \Leftrightarrow S_n \leq t$

Hence

$$P[N(t) = n] = P[N(t) \geq n] - P[N(t) \geq n+1]$$

$$= P[S_n \leq t] - P[S_{n+1} \leq t]$$

$$= P\left[\sum_{i=1}^{n} X_i \leq t\right] - P\left[\sum_{i=1}^{n+1} X_i \leq t\right]$$

$$= F_n(t) - F_{n+1}(t)$$

{Since $F_n(x) = P(X_1 + X_2 + \cdots + X_n \leq x)$}

Let $m(t) = E[N(t)]$ be the renewal function.

Renewal Equation:

The renewal equation is a fundamental concept in renewal theory. It is used to describe and analyze systems that reset or renew themselves at random points in time, such as machinery that undergoes repair or replacement, populations that reproduce, or events that recur at random intervals.

Theorem 4.3.1: **(Renewal Equation)**

The renewal function $m(t)$ is given by

$$m(t) = \sum_{n=1}^{\infty} F_n(t)$$

Proof: We know that

$$m(t) = E[N(t)]$$

$$= \sum_{n=0}^{\infty} n\, P[N(t) = n]$$

$$= \sum_{n=1}^{\infty} n\, P[F_n(t) - F_{n+1}(t)]$$

$$= F_1(t) - F_2(t) + 2F_2(t) - 2F_3(t) + 3F_3(t) - 3F_4(t) + \cdots$$

$$= \sum_{n=1}^{\infty} F_n\,(t)$$

$\Rightarrow m(t) = \sum_{n=1}^{\infty} F_n$ (t)

Laplace Transformation of $\boldsymbol{m(t)}$

Let $f(x) = F'(x)$ be the pdf of X_1 and $g^*(s)$ be the Laplace transformation of $g(t)$, i.e.

$$g^*(s) = \int_0^{\infty} \exp(-st)\ g(t)dt$$

So,

$$m^*(s) = \int_0^{\infty} \exp(-st)\ m(t)dt$$

$$= \int_0^{\infty} \exp(-st) \sum_{n=1}^{\infty} F_n\,(t)dt$$

$$= \sum_{n=1}^{\infty} \int_0^{\infty} \exp(-st)\ F_n(t)dt$$

$$= \sum_{n=1}^{\infty} F_n^*(s)$$

$\{F_n^*(s)$ is the Laplace transformation of the distribution function $F_n(t)$ of $X_1 + X_2 + \cdots + X_n$.

$f_n^*(s)$ is the Laplace transformation of density function of $X_1 + X_2 + \cdots + X_n$. i.e.

$f_n^*(s) = s\ F_n^*(s)\}$

$$\Rightarrow m^*(s) = \frac{1}{s}\sum_{n=1}^{\infty} f_n^*(s)$$

$$= \frac{1}{s}\sum_{n=1}^{\infty}\{f^*(s)\}^n = \frac{1}{s}\left\{\frac{f^*(s)}{1-f^*(s)}\right\}$$

Example 4.3.1.: If $X_1 \sim Gamma(\alpha, \beta)$, the pdf is

$$f(x) = \frac{\beta^{\alpha}}{\Gamma\alpha}\exp(-\beta x)\, x^{(\alpha-1)}$$

The Laplace transformation of $f(t)$ is

$$f^*(s) = \int_0^{\infty} \exp(-st)\ f(t)dt$$

$$= \int_0^{\infty} \exp(-st)\ \frac{\beta^{\alpha}}{\Gamma\alpha}\exp(-\beta t)\, t^{(\alpha-1)}dt$$

$$= \frac{\beta^{\alpha}}{\Gamma\alpha}\int_0^{\infty} \exp\{-(s+\beta)t\}\, t^{(\alpha-1)}dt$$

Since $\int_0^\infty \frac{\beta^\alpha}{\Gamma\alpha} \exp\{-\beta t\}\, t^{(\alpha-1)} dt = 1$

we have $\int_0^\infty \frac{(s+\beta)^\alpha}{\Gamma\alpha} \exp\{-(s+\beta)t\}\, t^{(\alpha-1)} dt = 1$

Hence

$$\int_0^\infty \exp\{-(s+\beta)t\}\, t^{(\alpha-1)} dt = \frac{\Gamma\alpha}{(s+\beta)^\alpha}$$

Therefore

$$f^*(s) = \frac{\beta^\alpha}{\Gamma\alpha}\left\{\frac{\Gamma\alpha}{(s+\beta)^\alpha}\right\}$$

$$= \left(\frac{\beta}{s+\beta}\right)^\alpha$$

If $\alpha = 1$, we get

$$f^*(s) = \frac{\beta}{s+\beta}$$

We know that

$m^*(s)$

$$= \frac{1}{s}\left\{\frac{f^*(s)}{1-f^*(s)}\right\}$$

$$= \frac{1}{s}\left\{\frac{\frac{\beta}{s+\beta}}{1-\frac{\beta}{s+\beta}}\right\}$$

Thus

$$m^*(s) = \frac{1}{s}\left\{\frac{\beta}{s}\right\}$$

$$= \frac{\beta}{s^2}$$

If $\alpha = 2$, then

$$f^*(s) = \left(\frac{\beta}{s+\beta}\right)^2$$

We know that

$$m^*(s)$$

$$= \frac{1}{s}\left\{\frac{f^*(s)}{1-f^*(s)}\right\}$$

$$= \frac{1}{s}\left\{\frac{\left(\frac{\beta}{s+\beta}\right)^2}{1-\left(\frac{\beta}{s+\beta}\right)^2}\right\}$$

Therefore

$$m^*(s)$$

$$= \frac{1}{s}\left(\frac{\beta}{s}\right)^2$$

$$= \frac{\beta^2}{s^3}.$$

Theorem 4.3.2.: Every renewal function $m(t)$ satisfies the renewal equation

$$\mathrm{m(t)} = \mathrm{F(t)} + \int_0^{\mathrm{t}} \mathrm{m(t-x)dF(x)}$$

Proof: $\mathrm{m(t)} = \mathrm{E[N(t)]}$

$$= \int_0^{\infty} E[N(t)|X_1 = x]\, f(x)dx$$

Consider

$$E[N(t)|X_1 = x] = \begin{cases} 0 & \text{if } x > t \\ 1 + E[N(t-x)] & \text{if } x \leq t \end{cases}$$

Hence

$$m(t) = \int_0^{t} [1 + E\{N(t-x)\}]\, f(x)dx$$

$$= \int_0^{t} f(x)dx + \int_0^{t} [E\{N(t-x)\}]\, f(x)dx$$

$$\Rightarrow m(t) = F(t) + \int_0^{t} m(t-x)dF(x)$$

Example 4.3.2.: If $X_1 \sim \text{Exp}(\theta)$

The pdf is

$$f(x) = \begin{cases} \theta \exp(-\theta x) & 0 \leq x < \infty \\ 0 & \text{otherwise} \end{cases}$$

The distribution function is

$$F(x) = 1 - \exp(-\theta x) \quad ; \quad 0 < x < \infty$$

The renewal function is

$$m(t) = 1 - \exp(-\theta t) + \int_0^{t} m(t-x)\theta \exp(-\theta x)\, dx$$

$$= 1 - \exp(-\theta t) + \theta \exp(-\theta t) \int_0^t m(t-x) \exp(-\theta x)\, dx$$

Let $y = t - x$, then $x = t - y$ and Jacobian is $dx = dy$.

We get

$$m(t) = 1 - \exp(-\theta t) + \theta \exp(-\theta t) \int_0^t m(y) \exp(\theta y)\, dy$$

Differentiating with respect to t, we get

$$\frac{d}{dt}\{m(t)\} = \frac{d}{dt}\left\{1 - \exp(-\theta t) + \theta \exp(-\theta t) \int_0^t m(y) \exp(\theta y)\, dy\right\}$$

$$= \theta \exp(-\theta t) - \theta^2 \exp(-\theta t) \int_0^t m(y) \exp(\theta y)\, dy + \theta \exp(-\theta t)\{m(t) \exp(\theta t)\}$$

$$= \theta \exp(-\theta t) - \theta^2 \exp(-\theta t) \int_0^t m(y) \exp(\theta y)\, dy + \theta\, m(t)$$

$$= \theta\, m(t) + \theta \exp(-\theta t) - \theta\{m(t) - 1 + \exp(-\theta t)$$

$$= \theta\, m(t) + \theta \exp(-\theta t) - \theta m(t) + \theta - \theta \exp(-\theta t)$$

$$= \theta$$

Therefore

$$m'(t) = \theta$$

$$m(t) = \theta t + C$$

At $t = 0, m(0) = 0 = C$

Hence

$m(t) = \theta t$

Example 4.3.3.: Let $X_1 \sim U(0,1)$

The pdf is

$$f(x) = \begin{cases} 1 & 0 \le x < 1 \\ 0 & \text{otherwise} \end{cases}$$

The distribution function is

$F(x) = x \;\; ; \quad 0 \le x \le 1$

The renewal function is

$$m(t) = F(t) + \int_0^t m(t-x)dF(x)$$

$$= t + \int_0^t m(t-x)dx$$

Let $y = t - x$, then $x = t - y$ and Jacobian is $dx = dy$, we get

$$m(t) = t + \int_0^t m(y)dy$$

Differentiating with respect to t, we get

$$\frac{d}{dt}\{m(t)\} = \frac{d}{dt}\left\{t + \int_0^t m(y)dy\right\}$$

$$= 1 + m(t)$$

$$\Rightarrow m'(t) = 1 + m(t)$$

$\Rightarrow m'(t) - m(t) = 1$

Multiplying $\exp(-t)$ on both sides, we get

$\{m'(t) - m(t)\}\exp(-t) = \exp(-t)$

$\Rightarrow \frac{d}{dt}[m(t)\exp(-t)]$

$= \exp(-t)$

$= \frac{d}{dt}[C - \exp(-t)]$

Therefore

$m(t)\exp(-t) = C - \exp(-t)$

At $t = 0$,

$m(t) = 0$

$\Rightarrow C = 1$

Hence

$m(t) = C\exp(t) - 1$

$= \exp(t) - 1$

Theorem 4.3.3.: Show that $N(t)$ has finite expectation. i.e.

$m(t) < \infty$ for all $0 \leq t < \infty$

Proof: We have $P[X_n = 0] < 1$.

By the continuity property of probabilities $\exists\ \alpha > 0$ such that $P[X_n \geq \alpha] > 0$.

Define a related renewal process $\{\overline{X}_n : n \geq 1\}$ by

$$\bar{X}_n = \begin{cases} 0 & \text{if } X_n < \alpha \\ \alpha & \text{if } X_n \geq \alpha \end{cases}$$

Let, $\bar{N}(t) = \max\{n: \bar{X}_1 + \bar{X}_2 + \cdots + \bar{X}_n \leq t\}$

Related renewal process can only take place at times $t = n\alpha\,, n = 0,1,2,\ldots$ Also the number of renewals at each of these times are independent geometric random variables with mean

$$\frac{1}{P\{X_n \geq \alpha\}}$$

Thus,

$$E[\bar{N}(t)]$$
$$\leq \frac{\frac{t}{\alpha} + 1}{P\{X_n \geq \alpha\}}$$
$$< \infty$$

Since $\bar{X}_n \leq X_n \quad \Rightarrow \bar{N}(t) \geq N(t)$.

Some Limit Theorems

Claim: Let $N(\infty) = \lim_{t\to\infty} N(t)$ be the total number of renewals that occurs, then $N(\infty) = \infty$ with probability 1.

Proof: Suppose, if possible, $N(\infty) < \infty$. This implies that

$$N(\infty) = n_0 < \infty$$

Therefore

$$P[N(\infty) < \infty]$$
$$= P[X_n = \infty \text{ for some } n]$$
$$= P\left[\bigcup_{n=1}^{\infty}\{X_n = \infty\}\right]$$

$$\leq \sum_{n=1}^{\infty} P\{X_n = \infty\}$$

$$= 0$$

$$\Rightarrow P[N(\infty) < \infty] \leq 0$$

$$\Rightarrow P[N(\infty) < \infty] = 0$$

$$\Rightarrow P[N(\infty) = \infty] = 1$$

Theorem 4.3.4.: Show that

$$\lim_{t\to\infty} \frac{N(t)}{t} = \frac{1}{\mu} \quad \text{with probability 1.}$$

Proof: Suppose $S_{N(t)}$ is the time of the last renewal prior to or at time t and $S_{N(t)+1}$ is the time of the first renewal after time t. Then

$$S_{N(t)} \leq t \leq S_{N(t)+1}$$

$$\Rightarrow \frac{S_{N(t)}}{N(t)} \leq \frac{t}{N(t)} \leq \frac{S_{N(t)+1}}{N(t)}$$

Using the Strong law of large numbers, then

$$\frac{S_{N(t)}}{N(t)} \to \mu \text{ as } N(t) \to \infty$$

But $N(t) \to \infty$ as $t \to \infty$

Therefore

$$\frac{S_{N(t)}}{N(t)} \to \mu \text{ as } t \to \infty$$

Consider

$$\frac{S_{N(t)+1}}{N(t)} = \frac{S_{N(t)+1}}{N(t)+1} \times \frac{N(t)+1}{N(t)}$$

$$= \mu \times 1 \qquad \left\{ \because \frac{N(t)+1}{N(t)} = 1 + \frac{1}{N(t)} \to 1 + 0 \quad \text{as } N(t) \to \infty \right\}$$

By Squeeze Theorem

$$\frac{t}{N(t)} \to \mu \qquad \text{as } N(t) \to \infty$$

But $t \to \infty \Rightarrow N(t) \to \infty$. Thus,

$$\frac{t}{N(t)} \to \mu \qquad \text{as } \ t \to \infty$$

$$\Leftrightarrow \frac{N(t)}{t} \to \frac{1}{\mu} \qquad \text{as } t \to \infty$$

Here $1/\mu$ is the rate of the renewal process.

Note: Sandwich Theorem or Squeeze Theorem:

Let f, g and h be real functions such that $f(x) \leq g(x) \leq h(x)$ for all x in the common domain of definition. For some real number a, if

$$\lim_{x \to a} f(x) = \lim_{x \to a} h(x) = L$$

Then

$$\lim_{x \to a} g(x) = L$$

Example 4.3.5.: A container contains an infinite collection of coins. Each coin has its own probability of landing heads and these probabilities are the values of independent random variables that are uniformly distributed over $(0,1)$. i.e. $P[H] = p \ , \ p \sim U(0,1)$.

(For different coins, p is different). Suppose we are to flip coins sequentially, at any time either flipping a new coin or one that had previously been used. Our objective is to maximize the long-run proportion of flips that lands on heads. How should we proceed?

Solution: Let $N(n)$ be the number of tails in the first n flips and so the long-run proportion of heads, called P_h, is given by

$$P_h = \lim_{n\to\infty} \frac{n - N(n)}{n} = 1 - \lim_{n\to\infty} \frac{N(n)}{n}$$

Using the proportion,

$$\lim_{n\to\infty} \frac{N(n)}{n} = \frac{1}{E(\text{number of flips between successive tails})}$$

The number of flips of a coin until it lands tails is geometric with mean $\frac{1}{1-p}$. Hence conditioning gives

$$E(\text{number of flips between successive tails}) = \int_0^1 \left(\frac{1}{1-p}\right) dp = \infty$$

$$\Rightarrow \lim_{n\to\infty} \frac{N(n)}{n} = \frac{1}{\infty} = 0$$

$$\Rightarrow P_h = 1$$

4.4 Elementary Renewal Theorem

For a renewal process,

$$\frac{m(t)}{t} \to \frac{1}{\mu} \qquad \text{as } t \to \infty$$

Proof: Clearly, $t < S_{N(t)+1}$

Taking expectation on both sides, we get

$t < E[S_{N(t)+1}]$

$$= E\left[\sum_{i=1}^{N(t)+1} X_i\right]$$

$= E(X_1)E\{N(t)+1\}$ by Wald's equation

$= \mu\{m(t)+1\}$

$$\Rightarrow \frac{m(t)}{t} > \frac{1}{\mu} - \frac{1}{t} \qquad (4.1)$$

Define

$$\bar{X}_i = \begin{cases} X_i & X_i \le a \quad \text{for } a > 0 \\ a & X_i > a \end{cases}$$

$= \min\{X_i, a\}$

Let $\{\bar{N}(t), t \ge 0\}$ be the corresponding renewal process with $\bar{S}_n = \bar{X}_1 + \bar{X}_2 + \cdots + \bar{X}_n$.

If $\bar{\mu} = E(\bar{X}_1)$ then $\bar{\mu} \le \mu$ and $\bar{m}(t) \ge m(t)$.

Consider

$\bar{\mu}\{m(t)+1\} \le \bar{\mu}\{\bar{m}(t)+1\}$

$= E[\bar{S}_{\bar{N}(t)+1}]$

$$= E\left[\sum_{i=1}^{\bar{N}(t)+1} \bar{X}_i\right]$$

$\le (t + a)$

Since

$$E\left[\bar{S}_{\bar{N}(t)+1}\right] = E[\bar{S}_{\bar{N}(t)} + \bar{X}_{\bar{N}(t)+1}]$$

$$\Rightarrow \frac{m(t)}{t} \leq \frac{1}{t\,\bar{\mu}}(t+a) - \frac{1}{t}$$

$$\Rightarrow \frac{m(t)}{t} = \frac{1}{\bar{\mu}} + \frac{a}{t\,\bar{\mu}} - \frac{1}{t} \qquad (4.2)$$

Combining (4.1) and (4.2), we have

$$\frac{1}{\mu} - \frac{1}{t} < \frac{m(t)}{t} \leq \frac{1}{\bar{\mu}} + \frac{a}{t\,\bar{\mu}} - \frac{1}{t}$$

Now putting $a = \sqrt{t}$, we get

$$\frac{1}{\mu} - \frac{1}{t} < \frac{m(t)}{t} \leq \frac{1}{\bar{\mu}} + \frac{1}{\bar{\mu}\sqrt{t}} - \frac{1}{t}$$

$$\Rightarrow \frac{1}{\mu} < \frac{m(t)}{t} \leq \frac{1}{\bar{\mu}} \qquad \text{as } t \to \infty$$

Hence

$$\lim_{t\to\infty} \frac{m(t)}{t} = 1/\mu$$

Example 4.4.1.: Suppose that potential customers arrive at a single-server bank in accordance with $P(\lambda)$. However, the potential customer enters the bank only if the server is free. If we assume that the amount of time spent in the bank by an entering customer is a random variable having distribution G, then

(i) Find the rate at which customers enter the bank.

(ii) What is the proportion of potential customers that cannot join the bank?

(iii) What is the utilization of the bank?

Proof: (i) We know that

Mean service time= μG

Rate of customer arriving= λ.

Mean inter arrival time$(\mu) = \frac{1}{\lambda} + \mu G$

Therefore,

Rate of entering customer $= \frac{1}{\mu}$

$$= \frac{1}{\frac{1}{\lambda} + \mu G}$$

$$= \frac{\lambda}{1 + \lambda \mu G}$$

(ii) Proportion of last customer= $1 -$Proportion of entering customers

$$= 1 - \frac{\frac{1}{\mu}}{\lambda}$$

$$= 1 - \frac{\frac{\lambda}{1 + \lambda \mu G}}{\lambda}$$

$$= \frac{\lambda(1 + \lambda \mu G) - \lambda}{\lambda(1 + \lambda \mu G)}$$

$$= \frac{\lambda \mu G}{(1 + \lambda \mu G)}$$

(iii) Utilization $= \frac{\text{Mean Service time}}{\text{Mean inter arrival time}}$

$$= \frac{\mu G}{\frac{1}{\lambda} + \mu G}$$

$$= \frac{\lambda\mu G}{1 + \lambda\mu G}$$

4.5 Equilibrium Renewal Process

Let $\{N^e(t): t \geq 0\}$ be a delayed renewal process with cumulative distribution function G given as

$$G(x) = \lim_{t\to\infty} P[\gamma_1 \leq x]$$

$$= \frac{1}{\mu}\int_0^x \{1 - F(s)\}\, ds$$

Then $\{N^e(t): t \geq 0\}$ is called an equilibrium renewal process. As before we define the renewal function as the mean number of renewals, i.e.

$$m^e(t) = E[N^e(t)] \qquad t \geq 0$$

Theorem 4.5.1.: The renewal function $m^e(t)$ satisfies

$$m^e(t) = \frac{t}{\mu} \quad \text{for all } t \geq 0$$

Proof: we know that

$$G(x) = \frac{1}{\mu}\int_0^x \{1 - F(s)\}\, ds$$

Taking Laplace transform on both sides of above equation, we get

$$\overline{G}(t) = \int_0^\infty \exp(-ts)\ dG(s)$$

$$= \frac{1}{\mu}\int_0^\infty \exp(-ts)\,\{1 - F(s)\}\, ds$$

$$= \frac{1}{\mu}\int_0^{\infty} \exp(-ts) \int_s^{\infty} dF(r)\, ds$$

$$= \frac{1}{\mu}\int_0^{\infty} \int_0^{r} \exp(-ts)\, dF(r)\, ds$$

$$= \frac{1}{\mu}\int_0^{\infty} \frac{\{1 - \exp(-tr)\}}{t} dF(r)$$

$$= \frac{1}{\mu}\left\{\frac{1 - \bar{F}(t)}{t}\right\}$$

Then the renewal function is

$$\bar{m}^e(t) = \frac{\bar{G}(t)}{1 - \bar{F}(t)}$$

$$= \frac{1}{\mu t}$$

4.6 Summary

In this unit, we have discussed the following points:

- We have identified Renewal process and defined Distribution of renewal process
- We have explained asymptotic distribution of renewal process
- We have state and proved Elementary renewal theorem
- We have discussed equilibrium renewal process

4.7 Self-Assessment Exercise

1. Let $\{X_n : n \geq 1\}$ be an i.i.d, sequence of interoccurrence times with common probability density function given by

$$f(x) = \begin{cases} \exp\{-(x-1)\} & \text{if } x > 1 \\ 0 & \text{otherwise} \end{cases}$$

Let $\{N(t), t \geq 0\}$ be the corresponding renewal process. Find the Laplace transform $m^*(s)$, of the renewal function $m(s)$.

2. Let $\{X_n : n \geq 1\}$ be an i.i.d. sequence of interoccurrence times with common probability mass function given by

$P(X_n = 0) = 0.2,$
$P(X_n = 1) = 0.3,$
$P(X_n = 2) = 0.5$

Let $\{N(t), t \geq 0\}$ be the corresponding renewal process. Find the Laplace transform $m^*(s)$, of the renewal function $m(s)$.

3. Show that the equilibrium renewal process $\{N^e(t) : t \geq 0\}$ has stationary increments.

4. State and prove elementary renewal theorem.

5. Consider a system with two components which are arranged in series. The system fails if either of the two components fail. On failure, a component is replaced instantaneously. Suppose each component works independently and has life and exponential life time distribution with parameter λ. Let $N(t)$ denotes the number of failures for the system in the time interval $[0, t)$. Find the distribution of N(t), and also the renewal function.

4.8 References

- Anderson, T. W., and L. A. Goodman. (1957): "Statistical inference about Markov chains", Ann. Math. Stat., Vol. 28, pp. 89-109.
- Bartlett, M. S. (1955): “An Introduction to Stochastic Processes”, London: Cambridge University Press.
- Biflingsley, P. (1961): "Statistical methods in Markov chains," Ann. Math. Stat., VoL 32, pp. 12-40.

- Cinlar, E. (1975): “Introduction to Stochastic Processes”, Prentice Hall.
- Dover. Gardiner, C. (2009): “Stochastic methods: A handbook for the natural sciences”. Springer, 4th edition.
- Feller, W. (1968): “Introduction to Probability and Applications”, New Age India International.
- Friedman, A. (2004): “Stochastic differential equations and applications”.
- Hoel, P. G., Port, S. C. and Stone, C. J. (1991): “Introduction to Stochastic Processes”, University Book Stall.
- John, F. (1978): “Partial differential equations”. Spring, 3rd edition.
- Karlin, S. and Taylor, H. M. (1995): “A First Course in Stochastic Processes”, Academic Press.
- Koralov, L. B. and Sinai, Y. G. (2010): “Theory of Probability and Random Processes”. Springer.
- Medhi, J. (2002): “Stochastic Process", New Age International Publishers.
- Oksendal, B. (2005): “Stochastic Differential Equations”. Springer, 6 edition.
- Pavliotis, G. A. (2014): “Stochastic Processes and Applications”. Springer.
- Varadhan, S. R. S. (2007): “Stochastic Processes”, volume 16 of Courant Lecture Notes. American Mathematical Society.

4.9 Further Reading

- Gikhman, I. and Skorokhod, A. V. (2004): “The theory of stochastic processes II. Classics in Mathematics”. Springer-Verlag.
- Grimmett, G., & Stirzaker, D. (2001): “Probability and Random Processes”, Oxford University Press.
- Karlin, S. and Taylor, H. M. (1981): “A second course in stochastic processes”. Academic Press.
- Lefebvre, M. (2006): “Applied Stochastic Processes”, Springer.
- Levin, D. A., Peres, Y., & Wilmer, E. L. (2009): “Markov Chains and Mixing Times”, American Mathematical Society.
- Resnick, S. I. (1992): “Adventures in Stochastic Processes”, Birkhäuser.

- Ross, S. M. (2014): “Stochastic Processes”, Wiley.
- Taylor, H. M., & Karlin, S. (1998): “An Introduction to Stochastic Modelling”, Academic Press.

U.P. Rajarshi Tandon Open University, Prayagraj

MScSTAT–203 (N)/ MASTAT–203 (N) Stochastic Process

Block: 2 Markov Chains and Markov Process

Course Design Committee

Course Preparation Committee

©UPRTOU
First Edition: ***July 2024***
ISBN : 978-93-48270-94-8

MScSTAT – 203N/ MASTAT – 203N
STOCHASTIC PROCESS

Printed and Published by Col. Vinay Kumar, Registrar, Uttar Pradesh Rajarshi Tandon Open University Prayagraj, 2024
Printed by: Cygnus Information Solution Pvt.Ltd., Lodha Supremus Saki Vihar Road Andheri East, Mumbai.

Block & Units Introduction

The present SLM on ***Stochastic Process*** consists of sixteen Units with four Blocks.

The ***Block – 2 - Markov Chains and Markov Process*** is the second block of said SLM, which is divided into four units.

In ***Unit-5 - Markov Dependent Trials***, discuss Two state Markov sequences and Markov chains. Also explain chain recurrent events and delayed recurrent event.

The ***Unit-6 - Transition Probabilities***, Deal with determination of n-step transition probabilities, Chapman-Kolmogorov equations, first return and first passage probabilities, fundamental theorem of probability of extinction, higher transition probabilities in Markov classification of states and chain.

The ***Unit-7 - Classification of States***, explain Classification of states, communication states, periodicity, stationary probability distributions, limit theorems, Ergodic chains and Irreducible Ergodic chains.

The ***Unit-8 - Continuous Time Markov Processes***, Markov processes in Continuous time. Interval arrival time, stopping time, optional stopping theorem, wald's equation, forward and backward equations for homogeneous case, random variable technique.

At the end of every block/unit the summary, self-assessment questions are given.

UNIT–5 MARKOV DEPENDENT TRIALS

Structure

5.1 Introduction

The Markov Chain is named after Andrey Markov (1856-1922), a Russian mathematician. Andrey Markov produced the first results in 1906 for this process having finite space.

A stochastic process is a family of random variables, which varies with inspect to time (the parameter) and take specific values in a state space. Real time space may be either discrete or continuous. In the ensuring section, we elucidate the concept of Markov Chain and its transition matrix. We also discuss the order of the MC. Higher order transition probabilities are also computed from Kolmogorov equation as well as transition probability matrix.

5.2 Objectives

After studying this unit, you should be able to:

- Explain the concept of a Two state Markov chain,
- Discuss the chain recurrent events,
- Explain the delayed recurrent events and application to the theory of success runs

5.3 Two State Markov Sequences

Example 5.3.1.: Consider a sequence of mutually independent Bernoulli trails with

$\Omega = \{S, F\}$ and $P(S) = p, P(F) = q(= 1 - p)$ in each trail. Define

$$X_n = \begin{cases} 1 \; if \; outcome \; of \; the \; n \; th \; trail \; is \; S \\ 2 \; if \; outcome \; of \; the \; nth \; trail \; is \; F \end{cases} \qquad (5.1)$$

Then $\{X_n, n = 1,2, \dots\}$ is a stochastic process.

Further

$P\{X_{n+1} = j_{n+1} | X_1 = j_1, \dots, X_n = j_n\}$

$= P\{X_{n+1} = j_{n+1}\}$, (because differnt trails are independent).

$j_r = 1,2; \; r = 1, \dots, n$. The trials are independent and the outcome of the $(n + 1)$ trials does not depend on the outcomes of the previous n trials.

Now we assume some kind of dependence between different Bernoulli trials.

Definition: Consider a sequence of Bernoulli random variable's $\{X_n, n = 0,1,2, \dots\}$, such that $P(X_n = 1) = p \; and \; P(X_n = 0) = q \; (= 1 - p), \forall \; n = 0,1,2, \dots$ Further $n = 0,1,2 \dots$ and for each possible value of $j_0, j_1, \dots j_n, j_{n+1}$, we have

$$P(X_{n+1} = j_{n+1} | X_0 = j_0, X_1 = j_1, \dots \dots \dots, X_n = j_n,)$$

$$= P(X_{n+1} = j_{n+1} | X_n = j_n) \qquad (5.2)$$

Then $\{X_n, n = 0,1,2, \dots\}$ is called a two-state Market Chain or Markov development trails.

In Markov dependent trails, the outcome of the $(n + 1)^{th}$ trail depends on the outcome of the n^{th} trial and, given the outcome of the n^{th} trial, it does not depend on the outcomes of the first $(n - 1)$ trials.

If we call outcome of the n^{th} trial as "PRESENT", outcome of the $(n+1)^{th}$trial as "FUTURE", outcomes of the first $(n-1)$ trials as "PAST", then the property (5.1) implies that the "FUTURE" depends only on "PRESENT" and not on the PAST.

This is called the Markov property, memoryless property, forgetfulness property or loss of memory property.

The Russian mathematician Markov considered such trials for the first time.

The sequence of independent Bernoulli trials (see Example 5.3.1.) is a trivial example of Markov dependent trials.

Let

$p_{ij} = P(X_{n+1} = j | X_n = i); i = 1,2, j = 1,2.; n = 0,1,2, \ldots$

The independent of p_{ij} from n is referred as the Markov sequence is (time or temporally) homogeneous.

If $X_n = i$, we say that the state of the process or the system at time n is i.

If $X_n = i$ and $X_{n+1} = j$, we say that there is a transition from the state i to the state j at time n+1, $(i, j = 1,2)$. Symbolically $i \rightarrow j$ at time (n+1); its probability is p_{ij}.

The four probabilities p_{11}, p_{12}, p_{21} and p_{22} are called the transition probabilities. However, $p_{12} = 1 - p_{11}$ and $p_{21} = 1 - p_{22}$. Hence only two of the four probabilities are the independent parameters. We may write these transition probabilities in matrix from as

$$P = \begin{bmatrix} p_{11} & p_{12} \\ p_{21} & p_{22} \end{bmatrix} = \begin{bmatrix} p_{11} & 1 - p_{11} \\ 1 - p_{22} & p_{22} \end{bmatrix}$$

P is called the matrix of transition probabilities or Transition Probability Matrix (TPM). The $(i,j)^{th}$ element of P denotes the conditional probability of a transition to state j at time $(n+1)$ given that the system is in state i at time n. Note that we are assuming that the transition probabilities are independent of time (n).

Given P we should be able to study the behavior of the process over a passage of time provided that the initial condition is given, *i.e.,* how the process started.

Let

$p_1^{(0)} = \text{prob of } S \text{ at the initial trial } = P(X_0 = 1)$

$$p_2^{(0)} = \text{prob of } F \text{ at the initial trial } = P(X_0 = 2)$$
$$= 1 - p_1^{(0)}$$

Thus, the initial probabilities vector is given by

$p^{(0)} = \left(p_1^{(0)}, p_2^{(0)}\right)$

The probability of S at the n^{th} trial is

$$p_n(S) = p_1^{(n)} = P(X_n = 1)$$

The probability of F at the n^{th} trial is

$p_n(F) = p_2^{(n)} = P(X_n = 2) = 1 - p_1^{(n)}$.

$p_1^{(n)}$ and $p_2^{(n)}$ are the state occupancy probabilities at the n^{th} trial.

The state occupancy probabilities vector at the n^{th} trial is

$p^{(n)} = \left(p_1^{(n)}, p_2^{(n)}\right)$.

Suppose

$p_{11}^{(n)} = P(X_n = 1|X_0 = 1)$

$$p_{12}^{(n)} = P(X_n = 2|X_0 = 1)$$

$$= 1 - p_{11}^{(n)}$$

$$p_{22}^{(n)} = P(X_n = 2|X_0 = 2)$$

$$p_{21}^{(n)} = P(X_n = 2|X_0 = 1)$$

$$= 1 - p_{22}^{(n)}$$

Then $p_{11}^{(n)}, p_{12}^{(n)}, p_{21}^{(n)}, p_{22}^{(n)}$ are the transition probabilities at time the n^{th} trial. Then, the n-step transition probability matrix is

$$P^{(n)} = \begin{pmatrix} p_{11}^{(n)} & p_{12}^{(n)} \\ p_{21}^{(n)} & p_{22}^{(n)} \end{pmatrix}.$$

Example 5.3.2.: Suppose that the probability of a dry day (state 1) following a rainy day (State 2) is $\frac{1}{4}$ and that the probability of a rainy day following a dry day is $\frac{1}{2}$. We have a two–state Markov chain such that $p_{21} = \frac{1}{4}$ and $p_{12} = \frac{1}{2}$ and transition probabilities matrix

$$P = \begin{bmatrix} \frac{1}{2} & \frac{1}{2} \\ \frac{1}{4} & \frac{3}{4} \end{bmatrix}$$

We have

$$P^2 = \begin{bmatrix} \frac{3}{8} & \frac{5}{8} \\ \frac{5}{16} & \frac{11}{16} \end{bmatrix}$$

And

$$P^4 = \begin{bmatrix} \frac{43}{128} & \frac{85}{128} \\ \frac{85}{256} & \frac{171}{256} \end{bmatrix}$$

Given that 1 denote a dry day, the probability that May 3 is a dry day is $\frac{3}{8}$, and that May 5 is a dry is $\frac{43}{128}$.

The following theorem derives the n-step transition probabilities of a two-state Markov Chain when the initial probability vector is given.

Theorem 5.3.1.: Given a two state Markov chain with transition probability matrix (TPM)

$$P = \begin{bmatrix} p_{11} & 1-p_{12} \\ 1-p_{21} & p_{22} \end{bmatrix}, 0 \le p_{11}, p_{22} \le 1, |p_{11} + p_{22} - 1| < |$$

and initial provability vector $p^{(0)} = \left(p_1{}^{(0)}, p_2{}^{(0)}\right)$, we have

$$p_n(S) = p_1^{(n)}$$

$$= (p_{11} + p_{22} - 1)^n \left\{p_1^{(0)} - \frac{1-p_{22}}{2-p_{11}-p_{22}}\right\} + \frac{1-p_{22}}{2-p_{11}-p_{22}}$$

and $p_n(F) = 1 - \ p_n(S), i.e., \ p_2^{(n)} = 1 - p_1^{(n)}$.

Proof: For n ≥1, we have

$$p_n(S) = P(X_n = 1)$$

$$= \mathrm{P}(\mathrm{X_n} = 1, \mathrm{X_{n-1}} = 1) + \mathrm{P}(\mathrm{X_n} = 1, \mathrm{X_{n-1}} = 2)$$

$$= P(X_n = 1|X_{n-1} = 1)P\ (X_{n-1} = 1) + P\ (X_n = 1|X_{n-1} = 2)P(X_{n-1} = 2)$$

$$= p_{11}\, p_{n-1}(S) + p_{21}\ \ p_{n-1}(F)$$

$$= p_{11}\, p_{n-1}(S) + p_{21}\, [1 - p_{n-1}(S)]$$

$$= p_{11}\, p_{n-1}(S) + (1-p_{22})[1 - p_{n-1}(S)]$$

$$= a\, p_{n-1}(S) + b$$

where $a = \ p_{11} + p_{22} \ - 1, \ \ b = 1 - p_{22}$.

Writing $p_n = \ p_n(S)$, we get the difference equation

$$p_n = ap_{n-1} + b\ , n \ge 1 \qquad (5.3)$$

For obtaining p_n we solve this difference equation under the restriction |a|<1, (|a|=1, if p_{11} = 1= p_{22} or if p_{11} = 0= p_{22}. If p_{11}=1 we get 1 1 ... or 2 2 ... with probability 1 and if p_{11} = 0= p_{22} we get 12 12… or 2 1 2 1 ...With probability 1.)

Let us define

$$p_n = u_n + \frac{b}{1-a}, \qquad n = 0,1,2 \ldots \tag{5.4}$$

Hence from (5.3) and (5.4), we get

$$u_n + \frac{b}{1-a}$$
$$= a\left(u_{n-1} + \frac{b}{1-b}\right) + b$$
$$= au_{n-1} + \frac{b}{1-a}$$

or

$$u_n = a\ u_{n-1}$$
$$= a^2\ u_{n-2}$$
$$= \cdots$$
$$= a^n\ u_0$$

Hence

$$p_n = p_n(S)$$
$$= u_n + \frac{b}{1-a}$$
$$= a^n\, u_0 + \frac{b}{1-a}$$
$$= a^n\left[p_0(S) - \frac{b}{1-a}\right] + \frac{b}{1-a}$$
$$= (p_{11} + p_{22} - 1)^n \left\{p_0{}^{(S)} - \frac{1-p_{22}}{2-p_{11}-p_{22}}\right\} + \frac{1-p_{22}}{2-p_{11}-p_{22}}$$

Interchanging the roles of S and F, we obtain

$$p_n(F)$$
$$= (p_{11} + p_{22} - 1)^n \left\{ p_0^{(F)} - \frac{1 - p_{11}}{2 - p_{11} - p_{22}} \right\} + \frac{1 - p_{11}}{2 - p_{11} - p_{22}}$$
$$= 1 - p_n(S).$$

Hence the theorem follows.

If the initial probabilities $p_0(S)$and $p_0(F)$are not given then we can compute the transition probabilities $p_{ij}^{(n)} = P\{X_n = j | X_0 = i\}; i, j = 1,2.$

Theorem 5.3.2.: For a two state Markov chain with the transition probability matrix (TPM)

$$P = \begin{bmatrix} p_{11} & 1 - p_{11} \\ 1 - p_{22} & p_{22} \end{bmatrix}, 0 \leq p_{11}, p_{22}, \leq |p_{11} + p_{22} - 1| < 1$$

the n- step TPM is given by

$$P^{(n)} = A + (p_{11} + p_{22} - 1)^n \, B,$$

where,

$$A = \frac{1}{2 - p_{11} - p_{22}} \begin{bmatrix} 1 - p_{22} & 1 - p_{11} \\ 1 - p_{22} & 1 - p_{11} \end{bmatrix}$$
$$B = \frac{1}{2 - p_{11} - p_{22}} \begin{bmatrix} 1 - p_{11} & -(1 - p_{11}) \\ -(1 - p_{22}) & 1 - p_{22} \end{bmatrix}$$

Proof: For n ≥ 2

$$p_{11}^{(n)} = P(X_n = 1 | X_0 = 1)$$
$$= P(X_n = 1, X_{n-1} = 1 | X_0 = 1) + P(X_n = 1, X_{n-1} = 2 | X_0 = 1)$$
$$= P\{X_n = 1 | X_{n-1} = 1\} P\{X_{n-1} = 1 | X_0 = 1\} + P\{X = 1 | X_{n-1} = 2\} P\{X_{n-1} = 2 | X_0 = 1\}$$
$$= p_{11} \, p_{11}^{(n-1)} + p_{21} \, p_{12}^{(n-1)}$$

$$= p_{11}\, p_{11}^{(n-1)} + (1 - p_{21})\left[1 - p_{11}^{(n-1)}\right]$$

$$= a\, p_{11}^{(n-1)} + b \qquad (5.5)$$

where $a = p_{11} + p_{22} - 1,\ \ b = 1 - p_{22}$

For solving this difference equation (5.5), we write

$$p_{11}^{(n)} = u^{(n)} + \frac{b}{1-a}, n \geq 1$$

so that (5.5) reduces to

$$u^{(n)} = a u^{(n-1)}$$
$$= a^2 u^{(n-2)}$$
$$\ldots$$
$$= a^{n-1} u^{(1)}$$
$$= a^{n-1}\left[p_{11}^{(1)} - \frac{b}{1-a}\right]$$

Hence

$$p_{11}^{(n)}$$
$$= a^{n-1}\left[p_{11}^{(1)} - \frac{b}{1-a}\right] + \frac{b}{1-a}$$
$$= (p_{11} + p_{22} - 1)^{n-1}\left[p_{11} - \frac{1-p_{22}}{2-p_{11}-p_{22}}\right] + \frac{1-p_{22}}{2-p_{11}-p_{22}}, \left(p_{11}^{(1)} = p_{11}\right)$$
$$= \frac{(p_{11} + p_{22} - 1)^n (1 - p_{11})}{2 - p_{11} - p_{22}} + \frac{1 - p_{11}}{2 - p_{11} - p_{22}} \qquad (5.6)$$

Interchanging the roles of S and F, we obtain

$$p_{22}^{(n)} = \frac{(p_{11} + p_{22} - 1)^n (1 - p_{22})}{2 - p_{11} - p_{22}} + \frac{1 - p_{11}}{2 - p_{11} - p_{22}} \qquad (5.7)$$

Further

$$p_{12}^{(n)}$$
$$= 1 - p_{11}^{(n)}$$
$$= -\frac{(p_{11} + p_{22} - 1)^n (1 - p_{22})}{2 - p_{11} - p_{22}} + \frac{1 - p_{11}}{2 - p_{11} - p_{22}} \qquad (5.8)$$

$$p_{21}^{(n)}$$
$$= 1 - p_{22}^{(n)}$$
$$= -\frac{(p_{11} + p_{22} - 1)^n (1 - p_{22})}{2 - p_{11} - p_{22}} + \frac{1 - p_{22}}{2 - p_{11} - p_{22}} \qquad (5.9)$$

Combining (5.6), (5.7) (5.8) and (5.9) we get

$$P^{(n)}$$
$$= \begin{bmatrix} p_{11}^{(n)} & p_{12}^{(n)} \\ p_{21}^{(n)} & p_{22}^{(n)} \end{bmatrix}$$
$$= A + (p_{11} + p_{22} - 1)^n \, B,$$

Here A and B are as defined in the theorem. Hence, we follow the theorem.

5.4 Chain Recurrent Event

In many research fields, it is common to observe processes that generate events repeatedly over the follow-up time for a given subject. Such processes are called recurrent event processes and the generated data are referred to as recurrent event data. For example, in bone marrow transplantation, different types of recurrent infections (e.g., bacterial, fungal, and viral infections) can occur after the surgery.

The attributes A defines a recurrent event if:

(i) A occurs at the n^{th} and $(n+m)^{th}$ of the sequence of event, it is necessary and sufficient that A occurs at the last place in each of the two sub sequences i.e., $(E_{i_1}, E_{i_2}, \dots, E_{i_n})$ and $(E_{i_{n+1}}, E_{i_{n+2}}, \dots, E_{i_{n+m}})$.

(ii) If A occurs at the n^{th} place then identically

$$P\left(E_{i_1}, E_{i_2}, \dots, E_{i_{n+m}}\right)$$
$$= P\left(E_{i_1}, E_{i_2}, \dots, E_{i_n}\right) \times P(E_{i_{n+1}}, E_{i_{n+2}}, \dots, E_{i_{n+m}})$$

Suppose that A occurs in the sequence $(E_{i_1}, E_{i_2}, \dots)$ for the first time at the n^{th} place, etc. It is also clear that with each recurrent event A there are associated the two sequences of numbers defined for $n = 1,2, \dots$ as

$$\left.\begin{aligned} v_n &= P[\, A \; occurs \; at \; the \; n^{th} \; trial] \\ f_n &= P[A \; occurs \; for \; the \; first \; time \; at \; the \; n^{th} \; trail] \end{aligned}\right\} \qquad (5.10)$$

Define $f_0 = 0, v_0 = 1$ (5.11)

5.5 Delayed Recurrent Events

Assumed that the probability distributions of the waiting time random variable till the first occurrence of the event and the inter-event time random variables are identical. However, this may not be the situation always and in general waiting time up to the first occurrence of the event has a different distribution than the later inter- recurrence time random variables. Once the event has occurred then the successive waiting time random variables are distributed identically. Such a recurrent event is called a delayed recurrent event where there is a delay in recurrence of the event, which is of same magnitude as the time interval when the event is occurring for the first time.

Consider a recurrent event E and let

$v_n = P[E \text{ occurs at the } n^{th} \; trial]$

Now E occurs for the 1^{st} time at the k^{th} trial with probability $P[X_n = k] = a_k$ and then recurs in subsequent $(n-k)$ trials according to probability distribution $\{p_{n-k}\}$, i.e.,

$$v_n = a_n + a_{n-1}p_1 + a_{n-2}p_2 + \cdots + a_0 p_n \qquad (5.12)$$

i.e., $\{v_n\}$ is convolution of $\{a_n\}$ and $\{p_n\}$.

Let

$$V(s) = \sum_{n=0}^{\infty} v_n s^n \ ;$$

$$A(s) = \sum_{n=0}^{\infty} a_n s^n$$

$$P(s) = \sum_{n=0}^{\infty} P_n s^n$$

be the generating functions of $\{v_n\}$, $\{a_n\}$ and $\{p_n\}$ respectively.

Then

$V(s)$

$= A(s)P(s) = \dfrac{A(s)}{1 - F(s)}$

The following theorem gives the probability $\{v_n\}$ of a delayed recurrent event and specifies the condition when this will be a proper probability distribution.

Theorem 5.5.1.: If $p_n \to b$, then $v_n \to ba$,

where $a = A(1) = \sum_n a_n$

If $\sum_n p_n \to \delta$, then $\sum_n v_n \to \delta a$.

Proof: $\gamma_k = P$(first renewal period is larger than k)

$\gamma_k = a_{k+1} + a_{k+2} + \cdots$

Let k is sufficiently large so that for a pre assigned number $\varepsilon, \gamma < \varepsilon$.

From (5.12), we have

$a_0 p_n + a_1 p_{n-1} + \cdots + a_k p_{n-k}$

$\leq v_n$

$= a_0 p_n + a_1 p_{n-1} + \cdots + a_k p_{n-k} + \{a_{k+1} p_{n-(k+1)} + \cdots + a_{n-1} p_1 + a_n\}$

$\leq a_0 p_n + a_1 p_{n-1} + \cdots + a_k p_{n-k} + \{a_{k+1} + \cdots + a_{n-1} + a_n\}$

$\leq a_0 p_n + a_1 p_{n-1} + \cdots + a_k p_{n-k} + \gamma_k$

$a_0 p_n + a_1 p_{n-1} + \cdots + a_k p_{n-k} \to \{a_0 + a_1 + \cdots + a_k\} b$ since $p_n \to b$

$= (a - \gamma_k) b$

$> ab - \varepsilon b$

$> ab - 2\varepsilon$ as $b < 2$

Now

$a_0 p_n + a_1 p_{n-1} + \cdots + a_k p_{n-k} + \gamma_k$
$\to (a - \gamma_k) b + \gamma_k$

$= ab + \gamma_k (1 - b)$

$< ab + \varepsilon (1 - b)$

$< ab + 2\varepsilon$

Consider

$\Rightarrow ab - 2\varepsilon < v_k < ab + 2\varepsilon$

Then

$\lim_{n \to \infty} v_k \to ab$

Again

$\sum_n v_n = V(1)$
$= A(1) P(1)$

$\rightarrow ab$

as $P(1) = \sum_n p_n \rightarrow b$

5.6 Application of the Theory of Success Runs

Let R be a fixed positive integer and A be the occurrence of a success run of length R in a sequence of Bernoulli trials. In the classical literature a success run of length R meant an uninterrupted sequence of either exactly R, or of at least R, successes. Neither convention leads to a recurrent pattern, for otherwise runs are not recurrent events.

Using (5.10), v is the probability of A at the n^{th} trial and f_n is the probability that the first run of length R occurs at the n^{th} trial.

The probability that the R trials number $n, (n-1), (n-2), \ldots, (n-r+1)$ result in success is p^R. In this case A occurs at one among these R trials; the probability that A occurs at the trial number $(n-k), k = 0,1, \ldots, R-1$, and the k successes, is equal to $v_{n-k}p^k$. Since these R possibilities are mutually exclusive, we get the recurrence relation is

$$v_n + v_{n-1}p + \cdots + v_{n-R+1}p^{R-1} = p^R \qquad n \geq R \qquad (5.13)$$

Then

$v_1 = v_2 = \cdots = v_{R-1} = 0\,, \qquad v_0 = 1$

Multiplying s^n in (5.13), we get

$\{v_n + v_{n-1}p + \cdots + v_{n-R+1}p^{R-1}\}s^n = p^R s^n$

Summing over $n = R, R+1, R+2, \ldots$, we get

$\{v(s) - 1\}\{1 + ps + \cdots + p^{R-1}s^{R-1}\} = p^R(s^R + s^{R+1} + \cdots)$

The two series are geometric and we find that

$$\{v(s) - 1\}\frac{1 - (ps)^R}{1 - ps} = \frac{p^R s^R}{1 - s}$$

or

$$v(s) = \frac{1 - s + qp^R s^{R+1}}{(1-s)(1-p^R s^R)} \qquad (5.14)$$

We know that the generating function $v(s)$ and $F(s)$ are related by

$$v(s) = \frac{1}{1 - F(s)}$$

Then the generating function of the recurrence times is

$$F(s)$$
$$= \frac{p^R s^R (1-ps)}{1 - s + qp^R s^{R+1}}$$
$$= \frac{p^R s^R}{1 - qs(1 + ps + \cdots + p^{R-1}s^{R-1})} \qquad (5.15)$$

We know that $F(1) = 1$. This shows that in a prolonged sequence of trials the number of runs of any length is certain to increase overall bounds.

After differentiation (5.15) and a little algebraic manipulation, the mean and variance of the recurrence times of runs of length R are obtained as

$$\mu = \frac{1 - p^R}{qp^R},$$
$$\sigma^2 = \frac{1}{(qp^R)^2} - \frac{2R+1}{qp^R} - \frac{p}{q^2}$$

5.7 Expected Number of Visits to a Specified State in a Time Period

Let $N_{ij}^{(n)}$ $(i, j = 1,2)$ be a random variable denoting the number of visits the Markov Chain makes to state j starting initially in state i, in the first n transitions.

Let

$$\mu_{ij}^{(n)} = E\left(N_{ij}^{(n)}\right).$$

Theorem 5.7.1.: For a two state Markov Chain. with TPM $P = ((p_{ij})), i, j = 1,2;\ 0 \leq p_{11}, p_{22}, \leq 1, |p_{11} + p_{22} - 1| < 1$, the matrix $\left(\left(\mu_{ij}^{(n)}\right)\right)$, where $\mu_{ij}^{(n)}$ denotes the expected number of visits to state j in the first n transition starting initially from state i, is given by

$$\left(\left(\mu_{ij}^{(n)}\right)\right) = \begin{bmatrix} n\pi_1 + \dfrac{a(1-a^n)\pi_2}{1-a} & n\pi_2 + \dfrac{a(1-a^n)\pi_2}{1-a} \\ n\pi_1 + \dfrac{a(1-a^n)\pi_1}{1-a} & n\pi_2 + \dfrac{a(1-a^n)\pi_1}{1-a} \end{bmatrix}$$

where

$$\pi_1 = \frac{(1-p_{22})}{(2-p_{11}-p_{22})},$$

$$\pi_2 = 1 - \pi_1,$$

$$a = p_{11} + p_{22} - 1.$$

Proof: Let $\{X_0, X_1, \ldots\ldots\}$ be a two state Markov Chain. Define a random variable

$$y_{ij}^{(m)} = \begin{cases} 1 & if\ X_m = j, X_0 = i \\ 0 & if\ X_m \neq j, X_0 = i \end{cases};\ m = 1,2, \ldots$$

For given m

$$P\left[y_{ij}^{(m)} = 0\right] = 1 - p_{ij}^{(m)}$$

$$P\left[y_{ij}^{(n)} = 1\right] = p_{ij}^{(m)}$$

Hence

$$E\left[y_{ij}^{(m)}\right] = p_{ij}^{(m)}$$

Now

$$N_{ij}^{(n)} = y_{ij}^{(1)} + y_{ij}^{(2)} + \cdots \ldots\ldots y_{ij}^{(n)}$$

$$= \sum_{m-1}^{n} y_{ij}^{(m)}$$

Therefore

$$\mu_{ij}^{(n)}$$

$$= E\left[N_{ij}^{(n)}\right]$$

$$= \sum_{m-1}^{n} E\left[y_{ij}^{(m)}\right]$$

$$= \sum_{m-1}^{n} p_{ij}^{(m)}$$

Hence

$$\mu_{ij}^{(n)} = \sum_{m=1}^{n} p_{ij}^{(m)}$$

$$= \sum_{m=1}^{n} \left[\frac{(p_{11} + p_{22} - 1)^m (1 - p_{11})}{2 - p_{11} - p_{22}} + \frac{1 - p_{22}}{2 - p_{11} - p_{22}}\right]$$

$$= \pi_2 \sum_{m=1}^{n} a^m + n\pi_1$$

$$= n\pi_1 + \pi_2 \frac{a(1 - a^n)}{1 - a}$$

which is the $(1,1)^{th}$ element of $\left(\left(\mu_{ij}^{(n)}\right)\right)$. Similarly, we can find other elements of the matrix $\left(\left(\mu_{ij}^{(n)}\right)\right)$. Hence the theorem follows.

Notice that $\lim_{n\to\infty} \frac{\mu_{ij}^{(n)}}{n} = \pi_1$ and $\lim_{n\to\infty} \frac{\mu_{22}^{(n)}}{n} = \pi_2$.

Therefore π_2 may be interpreted as the average fraction of time the process occupies the state $i(i = 1,2)$ in the long run. Hence π_2 has two interpretations:

(i) At a single point of time, as $n \to \infty, \pi_i$ is the probability that the system is in state i.

(ii) Over a long passage of time π_i is the average fraction of time the system is in state i.

5.8 Summary

In this unit, we have discussed the following points:

- We have discussed Two state Markov chain,
- Discuss the chain recurrent events,
- Explain the delayed recurrent events and application to the theory of success runs

5.9 Self-Assessment Exercise

1. Define Two state Markov Chain with example.

2. The transition probability matrix of a Markov Chain is

$$P = \begin{bmatrix} 0.1 & 0.5 & 0.4 \\ 0.6 & 0.2 & 0.2 \\ 0.3 & 0.4 & 0.3 \end{bmatrix}$$

Find $P^{(3)}$.

3. Using above transition probability matrix if the initial probability distribution is

$p^{(0)} = [0.7 \quad 0.2 \quad 0.1]$ then find out $P(X_2 = 3)$ and $P(X_3 = 2, X_2 = 3, X_1 = 3, X_0 = 2)$

4. Let R coins be tossed repeatedly and let A be the recurrent event that for each of the R coins the accumulated number of heads and tails are equal. Is A persistent or transient.

5. Show that $P^{(n)} = A + (p_{11} + p_{22} - 1)^n B$,

where,

$$A = \frac{1}{2 - p_{11} - p_{22}} \begin{bmatrix} 1 - p_{22} & 1 - p_{11} \\ 1 - p_{22} & 1 - p_{11} \end{bmatrix}$$

$$B = \frac{1}{2 - p_{11} - p_{22}} \begin{bmatrix} 1 - p_{11} & -(1 - p_{11}) \\ -(1 - p_{22}) & 1 - p_{22} \end{bmatrix}$$

5.10 References

- Anderson, T. W., and L. A. Goodman. (1957): "Statistical inference about Markov chains", Ann. Math. Stat., Vol. 28, pp. 89-109.
- Bartlett, M. S. (1955): "An Introduction to Stochastic Processes", London: Cambridge University Press.
- Bhanicha-Reid, A. (1960): "T. Elements of the Theory of Markov Processes and Their Applications", New York: McGraw-Hill.
- Biflingsley, P. (1961): "Statistical methods in Markov chains," Ann. Math. Stat., VoL 32, pp. 12-40.
- Cinlar, E. (1975): "Introduction to Stochastic Processes", Prentice Hall.
- Feller, W. (1968): "Introduction to Probability and Applications", New Age India International.
- Feller, W. (1949): "Fluctuation theory of recurrent events," Trans. Amer. Math. Soc, vol. 67, pp. 98-119.
- Hoel, P. G., Port, S. C. and Stone, C. J. (1991): "Introduction to Stochastic Processes", University Book Stall.
- Karlin, S. and Taylor, H. M. (1995): "A First Course in Stochastic Processes", Academic Press.
- Medhi, J. (2002): Stochastic Process, New Age International Publishers.
- Wax, N. (1954): "Noise and Stochastic Processes", New York: Dover.

5.11 Further Reading

- Grimmett, G., & Stirzaker, D. (2001): "Probability and Random Processes", Oxford University Press.
- Lefebvre, M. (2006): "Applied Stochastic Processes", Springer.

- Levin, D. A., Peres, Y., & Wilmer, E. L. (2009): “Markov Chains and Mixing Times”, American Mathematical Society.
- Resnick, S. I. (1992): “Adventures in Stochastic Processes”, Birkhäuser.
- Ross, S. M. (2014): “Stochastic Processes”, Wiley.
- Taylor, H. M., & Karlin, S. (1998): “An Introduction to Stochastic Modelling”, Academic Press.

UNIT–6 TRANSITION PROBABILITIES

Structure

6.1 Introduction

Transition probabilities in a stochastic process are the probabilities of a process transitioning from one state to another in a single time unit. They are often represented in a transition probability matrix, which is doubly stochastic if the rows and columns each sum to one.

6.2 Objectives

After studying this unit, you should be able to:

- Determination the n-step transition probabilities of a Markov process
- State and prove Chapman-Kolmogorov equations
- Define and compute first return and first passage probabilities

6.3 Determination of n-Step Transition Probabilities

So far, we have considered Markov chains with two possible outcomes in each trial. It can be extended to trials with more than two possible outcomes in each trial.

Example: consider a component, such as a valve, which is subject to failure. Let the component be inspected each day and classified as being in one of three states:

State 1: satisfactory

State 2: unsatisfactory

State 3: failed.

Suppose that at time n, the process is at state 1 let the probabilities of being at time $n+1$, in states 1,2,3 be $p_{11}, p_{12}, p_{13};\ p_{11}+p_{12}+p_{13};=1$ and let these probabilities do not depend on n. Next, if the process is in state 2 at time n let the probabilities of being at time $n+1$ in states 1,2,3, be $0, p_{22}, p_{23}$, with $p_{22}+p_{23}=1$. That is once the valve is unsatisfactory, it can never return to the satisfactory state. p_{22}, p_{23} are independent of n and of the history of the process before n. Finally, we suppose that if the process is in state 3 at time n, it is certain to be in state 3 at time $n+1$. Thus, the transition probabilities for transition from time n to time $n+1$ depend on the state given to be occupied at time n and the final state at time $n+1$, but not on what happened before time n. The transition probability matrix is given by

$$P=\begin{bmatrix} p_{11} & p_{12} & p_{13} \\ 0 & p_{22} & p_{23} \\ 0 & 0 & 1 \end{bmatrix}$$

In general, the state space S may consist of k states or even a countably infinite number of states.

Let $\{X_n; n=0,1,2,\dots\}$ be a stochastic process with X_n taking discrete values 1,2, ...

Definition: The stochastic process $\{X_n; n=0,1,2,\dots\}$ is called a Markov chain if for $n=1,2,\dots;\ i_0, i_1, i_2, \dots i_{n-1}, j \in S$,

$$P\{X_n=j|\ X_{n-1}=i_{n-1},\dots,X_0=i_0\}$$
$$=P\{X_n=j|\ X_{n-1}=i_{n-1}\}.$$

If $X_{n-1}=i$ and $X_n=j$, we say that the system has made a transition from state i to the state j at time n.

The probability $p_{ij} = p\{X_n = j | X_{n-1} = i\}, i, j \in S$ is called the (one-step) transition probability $i \to j$ at time n. the transition probabilities may or may not be independent of n. if the transition probability p_{ij} is independent of n, the Markov chain is said to be (time) homogeneous otherwise it is called non-homogeneous. We shall confine to homogeneous Markov chains.

Let the state space $S = \{1,2,3, \dots\}$. Then $p_{ij} \geq 0 \; \forall \; i, j \in S$ and $\sum_{j \in S} p_{ij} = 1 \forall i \in S$. The matrix

$$P = \begin{bmatrix} p_{11} & p_{12} & p_{13} & \dots \\ p_{21} & p_{22} & p_{23} & \dots \\ \vdots & \vdots & \vdots & \dots \end{bmatrix}$$

is called the (one-step) transition probability matrix. The sum of elements in each row of P is unity and each element is non-negative.

Definition: A square matrix satisfying, (i) each element is non-negative (ii) sum of elements in each row in unity, is called a stochastic matrix. If in addition to (i) and (ii), the sum of elements in each column is also unity, then the matrix is called a doubly stochastic matrix.

Here P is a stochastic matrix.

Let

$p_j^{(n)} = P(X_n = j); n = 0,1,2, \dots j \in S = \{1,2, \dots .\}$

$p_j^{(0)} = p(X_0 = j); j \in S:$ initial probability distribution

The conditional probability $P\{X_n = j | X_0 = i\} = p_{ij}^{(n)}$ is called the $n-$step transition probability, $i, j \in S$. The matrix

$$P^{(n)} = \begin{bmatrix} p_{11}^{(n)} & p_{12}^{(n)} & \dots \\ p_{21}^{(n)} & p_{22}^{(n)} & \dots \\ \vdots & \vdots & \dots \end{bmatrix}$$

is called the $n-$step Transition Probability Matrix of the Markov Chain.

6.4 Chapman–Kolmogorov Equation

Let the transition probability $p_{ij}^{(n)}$ gives the probability of n step transition from the state i at a trial to the state j at the next following trial. For obtaining the n −step transition probabilities, we have

$$p_{ij}^{(n)} = P\{X_n = j | X_0 = i\}$$

$$= \sum_{r \in S} P[X_n = j, X_{n-1} = r | X_0 = i] \; ; \qquad (S = \{1,2,3, \dots\})$$

$$= \sum_{r=1}^{\infty} P[X_n = j | X_{n-1} = r, X_0 = i] \, P[X_{n-1} = r | X_0 = i]$$

$$= \sum_{r=1}^{\infty} P[X_n = j | X_{n-1} = r] \, P[X_{n-1} = r | X_0 = i]$$

$$= \sum_{r=1}^{\infty} p_{ir}^{(n-1)} \, p_{rj} \qquad (6.1)$$

Since $p_{rj} \leq 1$, we have

$$\sum_{r=1}^{\infty} p_{ir}^{(n-1)} \, p_{rj} \leq \sum_{r=1}^{\infty} p_{ir}^{(n-1)} = 1 < \infty$$

Therefore $\sum_r p_{rj} p_{ir}^{(n-1)}$ is convergent. We can write (6.1) in matrix notation as

$$\begin{bmatrix} p_{11}^{(n)} & p_{12}^{(n)} & \cdots \\ p_{21}^{(n)} & p_{22}^{(n)} & \cdots \\ \vdots & \vdots & \cdots \end{bmatrix} = \begin{bmatrix} p_{11}^{(n-1)} & p_{12}^{(n-1)} & \cdots \\ p_{21}^{(n-1)} & p_{22}^{(n-1)} & \cdots \\ \vdots & \vdots & \cdots \end{bmatrix} \begin{bmatrix} p_{11} & p_{12} & p_{13} & \cdots \\ p_{21} & p_{22} & p_{23} & \cdots \\ \vdots & \vdots & \vdots & \cdots \end{bmatrix}$$

or

$$P^{(n)} = P^{(n-1)} P$$

$$= P^{(n-2)} P^2$$

$\vdots$

$= P^n$

Thus

$$P^{(n)} = P^n \qquad (6.2)$$

Eq. (6.2) can be used for the computation of $P_{ij}^{(n)}$.

Again

$$P^{(m+n)} = P^{m+n}$$
$$= P^m P^n$$
$$= P^{(m)} P^{(n)}$$
$$= P^{(n)} P^{(m)}$$

or

$$p_{ij}^{(m+n)} = \sum_r p_{ir}^{(m)} \, p_{rj}^{(n)}$$
$$= \sum_r p_{ir}^{(n)} \, p_{rj}^{(m)}, \quad (i,j) \in S. \qquad (6.3)$$

The set of equations (6.3) is known as the Chapman Kolmogorov (C-K) equations. The transition probabilities of a Markov Chain satisfy the Chapman- Kolmogorov equations. However, its converse is not always true, i.e., there exit non- Markovian Chains whose transition probabilities satisfy C-K equations.

Counter Example: Consider the sample space

{(1,2,3), (1,3,2), (2,1,3), (2,3,1), (3,1,2), (3,2,1), (1,1,1), (2,2,2), (3,3,3)}

with a probability mass $\frac{1}{9}$ attached to each triplet. Define the triplet (X_1, X_2, X_3) of random variables such that X_i is the number at the i^{th} place ($i = 1,2,3$). The possible values of X_i are 1,2 and 3. The probability distribution of X_i is

$$P(X_i = r) = \frac{1}{3} \quad r = 1,2,3$$

Further

$$P(X_i = r, X_j = s) = \frac{1}{9} \quad r, s = 1,2,3$$

$$P(X_1 = r, X_2 = s, X_3 = t) = \begin{cases} \frac{1}{9}, & r, s, t = 1,2,3;\ r = s = t \text{ or } r \neq s \neq t \\ 0 \text{ if } r = s \neq t \text{ or } r = t \neq s \text{ or } r \neq t = s \end{cases}$$

Hence

$$P(X_i = r, X_j = s)$$
$$= P(X_i = r)P(X_j = s)$$
$$= \frac{1}{9}$$

but

$$P(X_1 = r, X_2 = s, X_3 = t) \neq P(X_1 = r)P(X_2 = s), P(X_3 = t).$$

Therefore (X_1, X_2, X_3) are pair wise independent but not mutually independent.

Now start with the triplet (X_1, X_2, X_3). Then define another triplet (X_4, X_5, X_6) of random variable's exactly as we have defined (X_1, X_2, X_3) but independent of it. Then define another triplet (X_7, X_8, X_9) in the same manner as above but independent of the first two triplets and so on. Continuing in this manner we obtain a sequence (or family) of random variable's $\{X_1, X_2, X_3, \ldots X_n, \ldots\}$, i.e., a stochastic process. The sequence involves values 1,2 and 3 each with probability $\frac{1}{3}$. We thus have a stochastic process with state space $S = \{1,2,3\}$ and

$p_{ij}^{(1)} = p_{ij}$

$= P\,[X_{m+1} = j | X_m = i]$

$= P\,[X_{m+1} = j]$

$= \frac{1}{3}$ (since X_m, X_{m+1} are pairwise independent)

Similarly,

$p_{ij}^{(2)} = P[X_{m+2} = j | X_m = i] = \frac{1}{3}$

For $n \geq 3$

$p_{ij}^{(n)} = P[X_{m+n} = j | X_m = i]$

$= P[X_{m+n} = j]$

$= \frac{1}{3}$

Thus $\forall\ m, n \geq 1$ and $(i, j) \in S$

$p_{ij}^{(m+n)} = \frac{1}{3}$

and

$$\sum_{r=1}^{3} p_{ir}^{(m)}\, p_{rj}^{(n)} = \sum_{r=1}^{3} \frac{1}{3} \times \frac{1}{3} = \frac{1}{3} = p_{ij}^{(m+n)}$$

So that the C.K. equation holds for the stochastic process.

However, the stochastic process under consideration in non-Markovian. For verifying this, let the first transition takes the system to state 2. Then a transition to state 3 at the next step is possible if and only if the initial state was 1. Thus, the transition following the first step depend not only on the present state but also on the initial state, i.e., the process is non-Markovian.

We can find the n-step transition probability matrix through matrix multiplication. If n is large, it may be more convenient to compute $P^{(n)} = P^n$ using eigen decomposition. The matrix P can be expanded as

$$P = U\Lambda U^{-1}$$

where Λ is the diagonal matrix of eigenvalues and U is the matrix, whose columns are the corresponding eigen-vectors. Then,

$$P^{(n)} = P^n = U\Lambda^n U^{-1}$$

Since Λ is a diagonal matrix, one can easily obtain Λ^n.

For obtaining the vector of State occupation probabilities at time n,

$$p^{(n)} = \left(p_1^{(n)}, p_2^{(n)}, \ldots.\right)$$

we have

$$p_j^{(n)} = P(x_n = j) \quad (n = 0,1, \ldots., j = 1,2 \ldots)$$

$$= \sum_r P(x_n = j, x_{n-1} = r)$$

$$= \sum_r P(x_n = j | x_{n-1} = r) P(x_{n-1} = r)$$

$$= \sum_r p_{rj}\, p_r^{(n-1)}$$

$$= \sum_r p_r^{(n-1)}\, p_{rj} \qquad (6.4)$$

There is no convergent difficulty as

$$\sum_r p_r^{(n-1)}\, p_{rj} \leq \sum_r p_r^{(n-1)} = 1 < \infty$$

In matrix notation we can express (6.4) as

$$p^{(n)} = p^{(n-1)}P \qquad (6.5)$$

On iteration, we obtain

$$p^{(n)} = p^{(n-1)}P$$
$$= p^{(n-2)}P^2$$
$$= \cdots$$
$$= p^{(0)}P^n;\ n = 1,2, \ldots.$$

Hence the initial probability vector $p^{(0)}$ and the TPM P suffice to determine the marginal distribution $p^{(n)}$.

Again, P^n can be obtained using the matrix eigen decomposition.

Example: Let state space of a Markov Chain be $S = (1,2)$, and its transition probability matrix is

$$P = \begin{bmatrix} 7/10 & 3/10 \\ 4/10 & 6/10 \end{bmatrix}$$

Obtain P^2.

Solution: Hence

$$P^{(2)} = P^2$$
$$= \begin{bmatrix} 7/10 & 3/10 \\ 4/10 & 6/10 \end{bmatrix} \times \begin{bmatrix} 7/10 & 3/10 \\ 4/10 & 6/10 \end{bmatrix}$$

$$= \begin{bmatrix} 61/100 & 39/100 \\ 13/25 & 12/25 \end{bmatrix}$$

Example: Consider a communication system which transmits the digits 1 and 2 through several stages. Let, $X_n, n \geq 1$ be the digit leaving the n^{th} stage of system and X_1 be the digit entering the first stage (leaving the 1^{st} stage). At each stage there is a constant probability q that the digit which enters will be transmitted unchanged (i. e. the digit will remain unchanged when it leaves), and probability p otherwise (i. e. the digit changes when it leaves), $p + q = 1$.

Here $\{X_n, n \geq 0\}$ is a homogeneous two–state Markov chain with transition probability matrix

$$P = \begin{bmatrix} q & p \\ p & q \end{bmatrix}$$

It can be shown (by mathematical induction or otherwise) that

$$P^m = \begin{bmatrix} \left\{\frac{1}{2} + \frac{1}{2}(q-p)^m\right\} & \left\{\frac{1}{2} - \frac{1}{2}(q-p)^m\right\} \\ \left\{\frac{1}{2} - \frac{1}{2}(q-p)^m\right\} & \left\{\frac{1}{2} + \frac{1}{2}(q-p)^m\right\} \end{bmatrix}$$

Here

$$p_{11}^{(m)} = p_{22}^{(m)} = \left\{\frac{1}{2} + \frac{1}{2}(q-p)^m\right\}$$

And

$$p_{21}^{(m)} = p_{12}^{(m)} = \left\{\frac{1}{2} - \frac{1}{2}(q-p)^m\right\}.$$

Also as $m \to \infty$,

$$\lim p_{11}^{(m)} = \lim p_{12}^{(m)} = \lim p_{21}^{(m)} = \lim\ p_{22}^{(m)} \to \frac{1}{2}$$

Suppose that the initial distribution is given by

$P\{X_1 = 0\} = a$ and $P\{X_1 = 0\} = b = 1 - a$. Then

$$P\{X_m = 0, X_1 = 0\} = P\{X_m = 0, |X_1 = 1\}P\{\ X_1 = 0\} = a\ p_{11}^{(m)}$$

$$P\{X_m = 0, X_1 = 1\} = P\{X_m = 0, |X_1 = 1\}P\{\ X_1 = 1\} = b\ p_{21}^{(m)}$$

The probability that the digit entering the first stage is 1 given that the digit leaving the m^{th} stage is 1 can be evaluated by applying Bayes' rule. We have

$$P\{X_1 = 0, |X_m = 1\} = \frac{P\{X_m = 0, |X_1 = 0\}P\{X_1 = 0\}}{P\{X_m = 0, |X_1 = 0\}P\{X_1 = 0\} + P\{X_m = 0, |X_1 = 1\}P\{X_1 = 1\}}$$

$$= \frac{a\, p_{11}^{(m)}}{a\, p_{11}^{(m)} + b\, p_{12}^{(m)}}$$

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

$$= \frac{a\{1 + (q-p)^m\}}{a\{1 + (q-p)^m\} + b\{1 - (q-p)^m\}}$$

6.5 First Return and First Passage Probabilities

Suppose that the chain is initially in state j and $f_{jj}^{(n)}$ denotes the probability that next occurrence of state j is at time n, i.e. $f_{jj}^{(1)} = p_{jj}$ and for $n = 2,3 \ldots$

$$f_{jj}^{(n)} = P[X_r \neq j, r = 1,2, \ldots, n-1;\ X_n = j|\ X_0 = j]$$

$f_{jj}^{(n)}$ is called the first return probabilities to state j at time n or recurrence probabilities.

Similarly, we define the first passage probability from state j to state k for time n as $f_{jk}^{(1)} = p_{jk}$ and for $n = 2,3 \ldots$

$$f_{jk}^{(n)} = P\ [X_r \neq k, r = 1,2, \ldots n-1;\ X_n = k|X_0 = j].$$

Now for $n \geq 2$

$$p_{jj}^{(n)} = P\ [\ X_n = j|X_0 = j]$$

$$= \sum_{r=1}^{n} P[X_1 \neq j, \ldots X_{r-1} \neq j, X_r = j|X_0 = j]P[X_n = j|X_r = j]$$

$$= \sum_{r=1}^{n} f_{jj}^{(r)}\, p_{jj}^{(n-r)} \qquad \left(p_{jj}^{(0)} = P\,[X_0 = j | X_0 = j] = 1\right)$$

$$= f_{jj}^{(n)}\, p_{jj}^{(0)} + \sum_{r=1}^{n-1} f_{jj}^{(r)}\, p_{jj}^{(n-r)}$$

$$= f_{jj}^{(n)} + \sum_{r=1}^{n-1} f_{jj}^{(r)}\, p_{jj}^{(n-r)}$$

or

$$f_{jj}^{(n)} = p_{jj}^{(n)} - \sum_{r=1}^{n-1} f_{jj}^{(r)}\, p_{jj}^{(n-r)}\,; n = 2,3,\ldots \qquad (6.6)$$

From (6.6), $f_{jj}^{(2)}\, f_{jj}^{(3)}\ldots$ can be calculated recursively.

Similarly

$$p_{jk}^{(n)} = \sum_{r=1}^{n-1} f_{jk}^{(r)}\, p_{kk}^{(n-r)} \quad (\text{verify it})$$

So that

$$f_{jk}^{(n)} = p_{jk}^{(n)} - \sum_{r=1}^{n-1} f_{jk}^{(r)}\, p_{kk}^{(n-r)}\,; \quad n = 2,3\ldots..$$

Notice that for $n = 1, f_{jk}^{(1)} = p_{jk}$. Given that the chain stats at state j, the sum

$$f_{jj}^{(n)} = \sum_{n=1}^{\infty} f_{jj}^{(n)}$$

is the probability That the process returns to state j at least once.

Definition: Suppose the chain is initially at state j. if the ultimate return to this state is a certain event, the state is called recurrent; in this case the time of first return will be a random variable and called the recurrence time.

Definition: If the ultimate return to a state has probability less than unity the state is called transient (or non-recurrent).

For a recurrent state j, $f_{jj} = 1$ and for a transient state j, $f_{jj} < 1$. $1 - f_{jj}$ gives the probability that the initial state j is never visited again.

In the case of a recurrent state $\{f_{jj}^{(n)};\ n = 1,2, \ldots\}$ is a probability distribution. Thus, for a recurrent state, the expected number of steps required for the first return to state j is given by

$$\mu_{jj} = \sum_{n=1}^{\infty} n\, f_{jj}^{(n)}$$

μ_{jj} is called the mean recurrence time for the state j.

If the mean recurrence time μ_{jj} is finite, the state is called positive recurrent.

If $\mu_{jj} = \infty$, the state is called null recurrent. Similarly

$$f_{jk} = \sum_{n=1}^{\infty} f_{jk}^{(n)}$$

is the probability of ever entering in state k given that the chain starts in state j. we may call f_{jk} the first passage probability from state j to state k. If f_{jk}=1, then

$$\sum_{n=1}^{\infty} n\, f_{jk}^{(n)}$$

is the mean first passage time from state j to state k.

6.5.1 Generating Function

For a sequence of real numbers $\{a_n, n \geq 0\}$, let

$$A(s) = \sum_{j=0}^{\infty} a_j\, s^j$$

converges in some internal $-s_0 < s < s_0$. Then A(s) is called the generating function of the sequence $\{a_n\}$. If $\{a_n\}$ is bounded, i.e., $\sum a_j < \infty$, we have for $|s| < 1$ $A(s) \leq \sum a_j < \infty$.

So that A(s) converges at least for $|s| < 1$.

Let $\{p_n, n \geq 0\}$ be a probability distribution so that $\{p_n, n \geq 0\}$ and $\sum a_j = 1$. Then

$$P(s) = \sum_{n=0}^{\infty} p_n s^n$$

is called the probability generating function (p g f) of the probability distribution $\{p_n\}$. Obviously, for |s|<1

$$|P(s)| = \left|\sum p_n s^n\right|$$

$$\leq \sum p_n |s|^n$$

$$\leq \sum p_n$$

$$= 1 < \infty$$

Therefore P(s) converges absolutely for at least |s|<1.

Let X be a discrete random variable with p.d. $\{p_n\}$, then P(s), the p g f of X, is given by

$P(s) = E[s^X]$.

Now the moment generating function of X is

$\Psi(s) = E[e^{sX}]$

$= E[\{e^s\}^X]$

$= P[e^s]$

Therefore

$$\begin{cases} \Psi(s) = P[e^s] \\ P(s) = \Psi[\log(s)]. \end{cases}$$

Results:

(i) $p_k = \frac{1}{k!}\frac{d^k}{ds^k}P(s)|_{s=0} \qquad k = 0,1,2, \ldots$

(ii) $E(X) = \frac{d}{ds}P(s)|_{s=1} = P'(1)$

$E[X(X-1)] = P''(1)$

In general, for $r = 1,2, \ldots$

$E[X(X-1) \ldots \ldots (X-r+1)] = P^{(r)}(1)$

(iii) If X and Y are independently distributed random variables with p g f's $P_1(s)$ and $P_2(s)$ respectively then the p g f of $X+Y$ is

$$P(s) = P_1(s).P_2(s)$$

(iv) $\lim_{s \to 1-} P(s) = P(1) = 1$

(v) Let $\{X_n\}$ be a sequence of i.i.d. discrete random variables with common p g f

$$g(s) = E(s^{X_i}), \qquad i = 1,2, \ldots$$

Let N be a positive integer valued random variable with p g f

$$h(s) = E(s^N)$$

Define $Y_N = \sum_{i=1}^{N} X_i$. Then the p g f of Y_N is given by

$$G(s) = h[g(s)]$$

Solution:

(i) We have

$$P(s) = \sum_{n=0}^{\infty} p_n \, s^n = p_0 + p_1 s + p_2 s^2 + \cdots + p_k s^k + \cdots$$

Now, differentiating s^k with respect to s, k times we obtain

$$\frac{d^k}{ds^k} s^k$$

$$= k(k-1)(k-2) \ldots 1$$

$$= k!$$

For $r < k$,

$$\frac{d^k}{ds^k} s^r = 0$$

For $r > k$,

$$\frac{d^k}{ds^k} s^r = r(r-1) \ldots (r-k+1) s^{r-k},$$

Which tends to 0 as $s \to 0$. Hence

$$\left.\frac{d^k}{ds^k} P(s)\right|_{s=0} = p_k k!$$

$$\text{or } p_k = \frac{1}{k!} \left.\frac{d^k}{ds^k} P(s)\right|_{s=0}$$

(ii) We observe that

$$\frac{d^r}{ds^r} P(s) = \sum_{n=r}^{\infty} p_n \, n(n-1) \ldots (n-r+1) s^{n-r}$$

Taking limit $s \to 1$, we obtain

$$\left.\frac{d^r}{ds^r} P(s)\right|_{s=1} = P^{(r)}(1)$$

$$= \sum_{n=r}^{\infty} p_n \, n(n-1) \ldots (n-r+1)$$

$$= E[X(X-1) \ldots \ldots (X-r+1)]$$

(iii) Since X and Y are independently distributed random variables with p g f's $P_1(s)$ and $P_2(s)$ respectively, the p g f of X+Y is

$$P(s) = E\left(s^{(X+Y)}\right)$$

$$= E(s^X s^Y)$$

$$= E(s^X)E(s^Y) \text{ (since X and Y are independently distributed)}$$

$$= P_1(s).P_2(s)$$

(iv) We can easily verify that

$$\lim_{s \to 1-} P(s) = P(1)$$

$$= \sum_{n=0}^{\infty} p_n = 1$$

(v) We have

$$G(s) = E[s^{Y_N}]$$

$$= E[E(s^{YN}|N)]$$

$$= E[E\{s^{X_i} \ldots \ldots s^{X_N}|N\}]$$

$$= E[E(s^{X_1}) \ldots . E(s^{X_N})|N]$$

$$= E[g(s)^N]$$

$$= h[g(s)].$$

6.5.2 Generating Functions of $\left\{p_{jk}^{(n)}; n \geq 0\right\}$ and $\left\{f_{jk}^{(n)}; n \geq 1\right\}$

We have

$$p_{jk}^{(n)} = P[x_n = k | x_0 = j]$$

$$p_{jk}^{(n)} = P[x_n = k | x_0 = j, x_1 \neq k, \ldots, x_{n-1} \neq k]$$

For $|s|<1$. the p.g.f. of $\{p_{jk}^{(n)}; n = 0,1, \ldots\}$ is

$$P_{jk}(s) = \sum_{n=0}^{\infty} p_{jk}^{(n)} s^n$$

Similarly, the p.g.f. of $\{f_{jk}^{(n)}; n = 0,1, \ldots\}$ is

$$F_{jk}(s) = \sum_{n=0}^{\infty} f_{jk}^{(n)} s^n.$$

Theorem 6.5.1.: We have

$$P_{jk}(s) = F_{jk}(s)\ P_{kk}(s); (j \neq k) \qquad (6.7)$$

$$P_{jj}(s) = \frac{1}{1 - F_{jj}(s)}. \qquad (6.8)$$

Proof: Let us define

$$\delta_{jk} = \begin{cases} 1 \text{ if } j = k \\ 0 \text{ if } j \neq k \end{cases}$$

We observe that

$$P_{jk}(s) = \sum_{n=0}^{\infty} p_{jk}^{(n)}\ s^n$$

$$= p_{jk}^{(0)} + \sum_{n=0}^{\infty} p_{jk}^{(n)}\ s^{(n)} \quad (p_{jk}^{(0)} = 1 \text{ if } j = k \text{ and } 0 \text{ if } j \neq k \text{ or } p_{jk}^{(0)} = \delta_{jk})$$

$$= \delta_{jk} + \sum_{n=0}^{\infty} \left\{ \sum_{m=1}^{n} f_{jk}^{(m)} p_{kk}^{(n-m)} \right\} s^{n-m+m}$$

$$= \delta_{jk} + \sum_{m=1}^{\infty} f_{jk}^{(m)} s^m \sum_{n=m}^{\infty} s^{n-m} \; p_{kk}^{(n-m)}$$

$$= \delta_{jk} + \sum_{m=1}^{\infty} f_{jk}^{(m)} s^m \sum_{u=0}^{\infty} s^u \; p_{kk}^{(u)}$$

$$= \delta_{jk} + \sum_{m=1}^{\infty} f_{jk}^{(m)} s^m \;\; p_{jj} \;\; (s)$$

$$= \delta_{jk} + F_{jk}(s) \, P_{kk}(s)$$

If $j \neq k, \delta_{jk} = 0$ so that

$$P_{jk}(s) = F_{jk}(s) \, P_{kk}(s)$$

If $j = k, \; \delta_{jk} = 1$ and

$$P_{jj}(s) = 1 + F_{jj}(s) \, P_{jj}(s)$$

or $P_{jj}(s) = \dfrac{1}{1 - F_{jj}(s)}$.

Hence the theorem follows.

Theorem 6.5.2.: The j^{th} state is recurrent, i.e., $f_{jj} = 1$, iff $\sum_{n=0}^{\infty} p_{jj}^{(n)} = \infty$. If j^{th} state is transient, i.e., $f_{jj} < 1$, we have

$$\sum_{n=0}^{\infty} p_{jj}^{(n)} = \frac{1}{1 - f_{jj}}.$$

Proof: For $s = 1$, we have

$$P_{jj}(1) = \sum_{n=0}^{\infty} p_{jj}^{(n)},$$

$$F_{jj}(1) = \sum_{n=1}^{\infty} f_{jj}^{(n)}$$

$$= f_{jj}$$

Since

$$P_{jj}(1) = \frac{1}{1 - F_{jj}(1)},$$

we get

$$\sum_{n=0}^{\infty} p_{jj}^{(n)} = \frac{1}{1 - f_{jj}}$$

Therefore

$$\sum_{n=0}^{\infty} p_{jj}^{(n)} < \infty \Leftrightarrow f_{jj} < 1$$

and

$$\sum_{n=0}^{\infty} p_{jj}^{(n)} = \infty \Leftrightarrow f_{jj} = 1$$

Hence, we get the result.

Theorem 6.5.3.: If the k^{th} state is transient, i.e., $f_{kk} < 1$ then $\sum_{n=0}^{\infty} p_{jk}^{(n)} < \infty. \forall\, j \in S$.

Proof: For $j = k$, the proof is obvious from the previous theorem. If $j \neq k$, we have

$$\sum_{n=0}^{\infty} p_{jk}^{(n)} = P_{jk}(1)$$

$= F_{jk}(1)\, P_{kk}(1)$

$= f_{jk} P_{kk}(1) \leq P_{kk}(1) \quad (\text{since } f_{jk} \leq 1)$

$= \sum_{n=0}^{\infty} p_{kk}^{(n)} < \infty, \text{since the } k^{th} \text{ state is transient.}$

Hence the theorem follows.

Corollary: if k is transient then $\lim_{n\to\infty} p_{jk}^{(n)} = 0$ for every j.

Proof: The proof follows from the convergence of $\sum_{n=0}^{\infty} p_{jk}^{(n)}$.

Example: Determine the probability of ultimate return to the states of the Markov chain having $S = \{1,2,3,4\}$ and transition matrix.

$$P = \begin{bmatrix} 1/3 & 2/3 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 1/2 & 0 & 1/2 & 0 \\ 0 & 0 & 1/2 & 1/2 \end{bmatrix}$$

Solution: The probability of ultimate return of the states is

For state 1

$f_{11}^{(1)} = p_{11} = \frac{1}{3}$

$f_{11}^{(2)} = f_{12} f_{21} = \frac{2}{3} \times 1 = \frac{2}{3}$

$f_{11}^{(3)} = 0$

$\vdots$

$f_{11} = \sum_{n=1}^{\infty} f_{11}^{(n)}$

$$= \frac{1}{3} + \frac{2}{3} + 0 + 0 + \cdots = 1$$

Hence the state 1 is persistent.

For state 2

$$f_{22}^{(1)} = p_{22} = 0$$

$$f_{22}^{(2)} = f_{21}f_{12}$$
$$= 1 \times \frac{2}{3} = \frac{2}{3}$$

$$f_{22}^{(3)} = f_{21}f_{11}f_{12}$$

$$= 1 \times \frac{1}{3} \times \frac{2}{3} = \frac{2}{9}$$

$$f_{22}^{(4)} = f_{21}f_{11}^{(2)}f_{12}$$

$$= 1 \times \left(\frac{1}{3}\right)^2 \times \frac{2}{3} = \left(\frac{1}{3}\right)^2 \frac{2}{3}$$

$$f_{22}^{(5)} = f_{21}f_{11}^{(3)}f_{12}$$

$$= 1 \times \left(\frac{1}{3}\right)^3 \times \frac{2}{3} = \left(\frac{1}{3}\right)^3 \frac{2}{3}$$

$$\vdots$$

$$f_{22}^{(n)} = f_{21}f_{11}^{(n-2)}f_{12}$$

$$= 1 \times \left(\frac{1}{3}\right)^{(n-2)} \times \frac{2}{3} = \left(\frac{1}{3}\right)^{(n-2)} \frac{2}{3}$$

$$f_{22} = \sum_{n=1}^{\infty} f_{22}^{(n)}$$

$$= 0 + \frac{2}{3} + \left(\frac{1}{3}\right)\frac{2}{3} + \left(\frac{1}{3}\right)^2\frac{2}{3} + \cdots + \left(\frac{1}{3}\right)^{(n-2)}\frac{2}{3}$$

$$= \sum_{r=2}^{\infty}\left(\frac{1}{3}\right)^{(r-2)}\frac{2}{3} = 1$$

Hence the state 2 is persistent.

For state 3

$$f_{33}^{(1)} = p_{33} = \frac{1}{2}$$

$$f_{33}^{(2)} = f_{32}f_{23}$$

$$= 0 \times 0 = 0$$

$$f_{33}^{(3)} = f_{32}f_{22}f_{23} = 0$$

$$\vdots$$

$$f_{33} = \sum_{n=1}^{\infty} f_{33}^{(n)}$$

$$= \frac{1}{2} + 0 + 0 + 0 + \cdots = \frac{1}{2} < 1$$

Hence the state 3 is transient.

For state 4

$$f_{44}^{(1)} = p_{44} = \frac{1}{2}$$

$$f_{44}^{(2)} = f_{43}f_{34}$$

$$= \frac{1}{2} \times 0 = 0$$

$$f_{44}^{(3)} = 0$$

$$\vdots$$

$$f_{44} = \sum_{n=1}^{\infty} f_{44}^{(n)}$$

$$= \frac{1}{2} + 0 + 0 + \cdots = \frac{1}{2} < 1$$

Hence the state 4 is transient.

6.6 Summary

In this unit, we have covered the following points:

- We have Defined n-step transition probabilities and calculating n-step transition probabilities.
- We have proved Chapman-Kolmogorov equations.
- We have explained the two valued Process and model for system reliability
- We have described First return and first passage probabilities and finding them.

6.7 Self-Assessment Exercise

1. Show that state j is persistent iff $\sum_{n=0}^{\infty} p_{jj}^{(n)} = \infty$

2. Consider Markov chain with transition probability matrix

$$P = \begin{bmatrix} 0 & 1 & 0 \\ 1/2 & 0 & 1/2 \\ 0 & 1 & 0 \end{bmatrix}$$

Calculate P^3 and P^4.

3. Let the Markov chain consisting of the states $S = \{0,1,2,3\}$ have the transition probability matrix

$$P = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 1/2 & 0 & 1/2 & 0 \\ 0 & 3/4 & 0 & 1/4 \\ 0 & 0 & 1 & 0 \end{bmatrix}$$

Determine which states are transient and which are recurrent.

4. Show that if state k is persistent null, then for every j $\lim_{n\to\infty} p_{jk}^{(n)} \to 0$

And if state k is aperiodic, persistent non null, then

$$\lim_{n\to\infty} p_{jk}^{(n)} \to \frac{F_{jk}}{\mu_{kk}}$$

6.8 References

- Cox, D. R., & Miller, H. D. (1965): "The Theory of Stochastic Processes". Chapman and Hall/CRC.
- Doob, J. L. (1953): "Stochastic Processes". John Wiley & Sons.
- Grimmett, G., & Stirzaker, D. (2001): "Probability and Random Processes". Oxford University Press.
- Karlin, S., & Taylor, H. M. (1981): "A First Course in Stochastic Processes". Academic Press.
- Lawler, G. F. (2006): "Introduction to Stochastic Processes". Chapman and Hall/CRC.
- Pavliotis, G. A. (2014): "Stochastic Processes and Applications: Diffusion Processes, the Fokker-Planck and Langevin Equations". Springer.
- Ross, S. M. (2014): "Introduction to Probability Models", Academic Press.

6.9 Further Reading

- Grimmett, G., & Stirzaker, D. (2001): "Probability and Random Processes", Oxford University Press.

- Karatzas, I., & Shreve, S. E. (1991): “Brownian Motion and Stochastic Calculus”, Springer-Verlag.
- Lefebvre, M. (2006): “Applied Stochastic Processes”, Springer.
- Levin, D. A., Peres, Y., & Wilmer, E. L. (2009): “Markov Chains and Mixing Times”, American Mathematical Society.
- Resnick, S. I. (1992): “Adventures in Stochastic Processes”, Birkhäuser.
- Ross, S. M. (2014): “Stochastic Processes”, Wiley.
- Taylor, H. M., & Karlin, S. (1998): “An Introduction to Stochastic Modelling”, Academic Press.

UNIT–7 CLASSIFICATION OF STATES

Structure

7.1 Introduction

One of the key aspects of studying Markov chains is understanding the nature of the states within the chain. States can be classified based on their properties, such as whether they can be revisited, the time intervals between visits, and the likelihood of staying in or leaving a state. These classifications provide valuable insights into the long-term behaviour of the system, such as whether the system will stabilize in a steady state, keep fluctuating, or eventually reach an absorbing state from which it cannot escape.

Understanding these classifications is crucial for analysing the properties of the system modelled by the Markov chain, predicting future states, and making informed decisions based on probabilistic outcomes. This unit sets the stage for exploring the detailed classifications of states in Markov chains and their implications for the study of stochastic processes.

7.2 Objectives

After studying this unit, you should be able to:

- Classify the state of the Markov chain into communicating classes and closed sets,

- Explain the periodicity and stationary probability distributions,
- Explain the limit theorem,
- Understand the concept of Ergodic chains and Irreducible Ergodic chains

7.3 Classification of States

The states $j, j = 0,1,2, \ldots$ of a Markov chain $\{X_n, n \geq 0\}$ can often be classified according to some fundamental properties of the system. Using such a classification it becomes possible to understand the properties of the Markov chain, predicting future states, and making informed decisions based on probabilistic outcomes.

Definition: A state j is called accessible from the state i iff $\exists$ a positive m such that

$p_{ij}^{(m)} > 0$. We write symbolically $i \to j$. Conversely, if for all m, $p_{ij}^{(m)} = 0$, then j is not accessible from i; i.e., $i \nrightarrow j$.

Definition: Two states i and j are called communicative if j is accessible from i and i is accessible from j. Thus, we say that the states i and j communicate if for some $m, n > 0$, $p_{ij}^{(m)} > 0$, $p_{ji}^{(n)} > 0$. Symbolically we write $i \leftrightarrow j$. Obviously, the communication is symmetric.

Theorem 7.3.1: The communication is transitive, i.e., if $i \leftrightarrow j$, $j \leftrightarrow k$, then $i \leftrightarrow k$.

Proof: Let $i \leftrightarrow j$ and $j \leftrightarrow k$. Suppose m and n are two integers such that

$p_{ij}^{(m)} > 0, p_{jk}^{(n)} > 0$, then by Chapman Kolmogorov equations

$$p_{ik}^{(m+n)} = \sum_{l \in s} p_{il}^{(m)} p_{lk}^{(n)}$$

$$\geq p_{ij}^{(m)} p_{jk}^{(n)}$$

$$> 0$$

so that $i \to k$. Similarly, we can show that if $k \to j$, and $j \to i$, then $k \to i$. Hence $i \leftrightarrow k$■

Definition: For a given state j of a Markov Chain, the set of all states k, which communicate with j, denoted by $C(j)$, is called the communication class of state j. Hence $k \in C(j)$ iff $k \leftrightarrow j$.

Theorem 7.3.2: Let C_1 and C_2 be any two communicating classes of a Markov Chain. Then either $C_1 = C_2$ or $C_1 \cap C_2 = \emptyset$.

Proof: If $C_1 \cap C_2 \neq \emptyset$ then $\exists$ a state k of the Markov Chain belonging to both C_1 and C_2. Let $i, j \in S$ such that $C_1 = C(i)$ and $C_2 = C(j)$. Consider any state $g \in C(i)$. Then $g \leftrightarrow i$. Since $g \leftrightarrow i, i \leftrightarrow k$ by transitivity we have $g \leftrightarrow k$. But $k \leftrightarrow j$, so that $g \leftrightarrow j$, i.e., $g \in C(j)$. Hence $C(i) \subset C(j)$. Similarly, we can show that $C(j) \subset C(i)$. Therefore $C(i) = C(j)$, or $C_1 = C_2$. This proves the theorem.

Definition: If C is a set of states such that no state outside C can be reached from any state in C, then C is said to be closed. If C is closed and $j \in C$ while $k \notin C$, then $p_{jk}^{(n)} = 0 \ \forall \ n$.

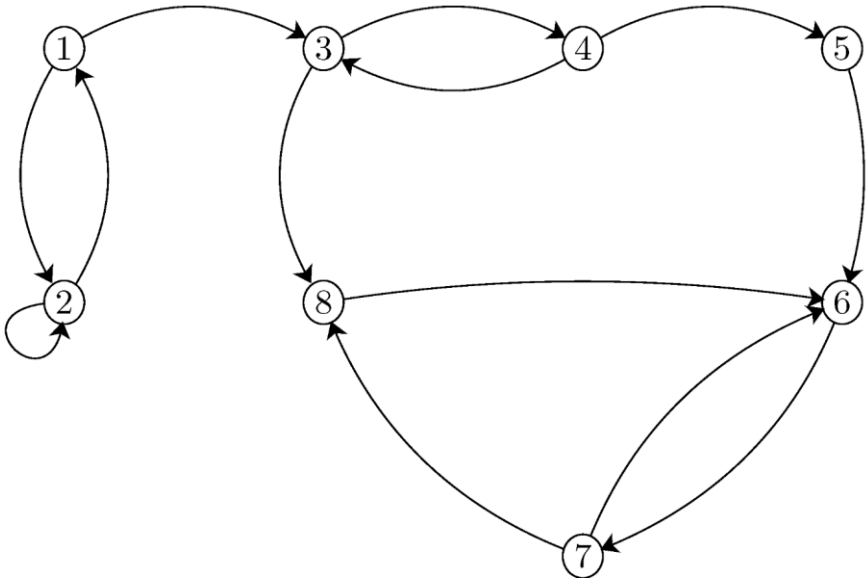
C is closed iff $\sum_{j \in C} p_{ij} = 1$ for every $i \in C$. Then the sub-matrix $P_1 = (p_{ij})$; $i, j, \in C$, is also stochastic and P can be expressed in the canonical form as:

$$P = \begin{bmatrix} P_1 & 0 \\ R_1 & Q \end{bmatrix}$$

A closed set may contain one or more states.

Definition: If a closed set contains only one state j then state j is said to be absorbing. Thus j is absorbing iff $p_{jj} = 1, p_{jk} = 0$, $k \neq j$.

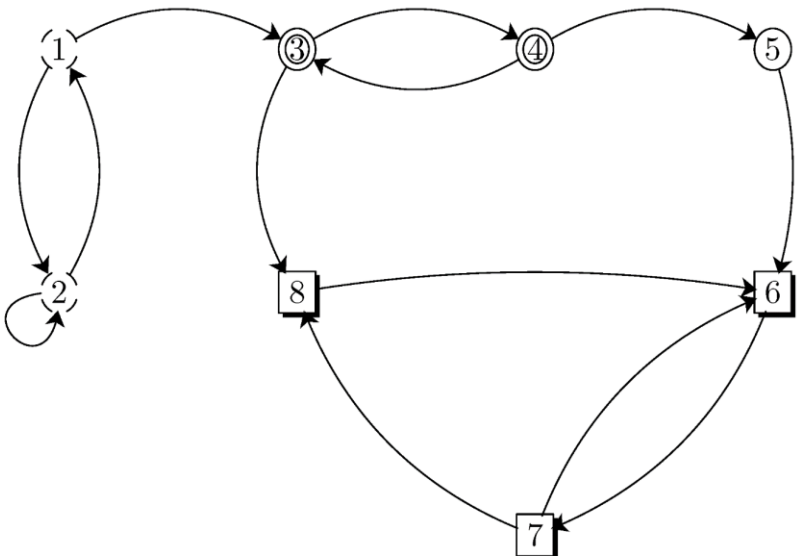
Example 7.3.1.: Consider the Markov Chain shown in the following figure. It is assumed that when there is an arrow from state i to state j, then $p_{ij} > 0$. Find the Equivalence Classes for this Markov Chain.



Solution: The State 1 and 2 communicate with each other, but they do not communicate with any other nodes in the graph. State 3 and 4 communicate with each other, but they do not communicate with any other nodes in the graph. State 5 does not communicate with any other states, so it by itself is a class. Finally, State 6, 7 and 8 construct another class. Thus, here are the classes:

Class 1= {state1, state 2}, Class 2= {state3, state 4}, Class 3={state5},

Class 4= {state6, state 7, state 8}

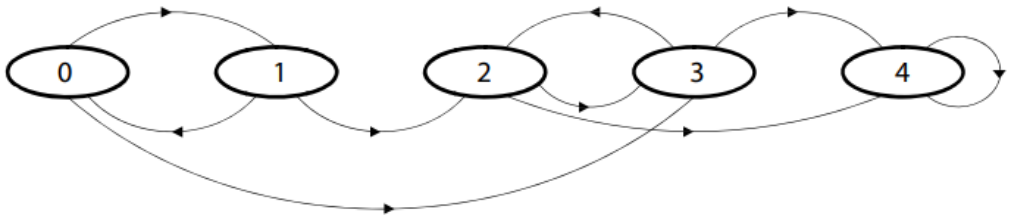


Example 7.3.2.: Consider a Markov Chain with transition Probability Matrix:

$$P = \begin{bmatrix} 0 & 2/7 & 0 & 5/7 & 0 \\ 5/6 & 0 & 1/6 & 0 & 0 \\ 0 & 0 & 0 & 2/5 & 3/5 \\ 0 & 0 & 1/2 & 0 & 1/2 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

Draw the transition graph and find the accessible state and commutative class.

Solution:



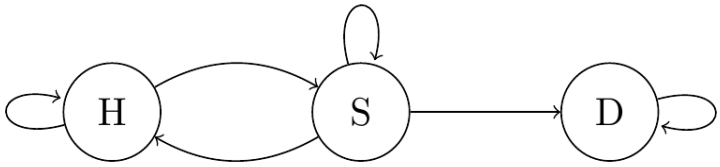
Above figure shows state 0 and state 1 communicate so they belong to the same class. State 2 is accessible from state 1, but not vice versa. So, state 2 does not belong to the class of state 0 and state 1. State 3 and state 2 communicate. Therefore, state 3 does not belong to the class of state 0 and state 1 either. State 2 and State 3 belong to the same class. State 4 is accessible from states 0, 1, 2, and 3, but no state is accessible from State 4. So, State 4 belongs to a class by itself. Thus, this Markov chain consists of three classes: $\{0, 1\}$, $\{2, 3\}$, and $\{4\}$. Note that, for state 4, $p_{44} = 1$. That is, once the process enters 4, it will stay there forever. Such states are called absorbing. In general, state i of a Markov chain is absorbing if $p_{ii} = 1$.

Example 7.3.3.: Consider the following simple model for an epidemic. We have three states: healthy (H), sick (S), and dead (D). This transition matrix is

$$P = \begin{bmatrix} p_{HH} & p_{HS} & 0 \\ p_{SH} & p_{SS} & p_{SD} \\ 0 & 0 & 1 \end{bmatrix}$$

Draw the transition diagram and find state space and communicating classes.

Solution: The transition diagram is



State H and S communicate with each other, while state D only communicates with itself. Hence, the state space $S = \{H, S, D\}$ partitions into two communicating classes: $\{H, S\}$ and $\{D\}$.

7.4 Periodicity

Definition: A state j of a Markov Chain is said to be periodic with period d_j if its return to the state is possible only at $d_j, 2d_j, 3d_j, \ldots$ steps, where d_j is the greatest integer with this property. In other words, if d_j is the greatest common divisor of all integers $n\ (\geq 1)$ for which $p_{jj}^{(n)} > 0$, then j is said to be periodic with period d_j. If $p_{jj}^{(n)} = 0\ \forall\ n$ then we take $d_j = 0$. The state j is said to be aperiodic if no such $d_j (> 1))$ exists. Thus, $d_j = 1$ will correspond to the aperiodic case.

If j is not a recurrent state we do not define its period.

A state j is called ephemeral if $p_{ij} = 0\ \forall\ i \in S$. A chain can only be in an ephemeral state initially and pass out of it in the first transition. An ephemeral state can never be reached from any other state. The column of P corresponding to an ephemeral state is composed entirely of zeros.

Let us exclude the ephemeral states from consideration.

Example 7.4.1.: Consider a Markov chain having $S = \{1,2,3,4\}$ and transition matrix

$$P = \begin{bmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 \\ 1/4 & 1/8 & 1/8 & 1/2 \end{bmatrix}$$

Is this chain (i) irreducible, (ii) aperiodic?

Solution: For this example, start in any state and can still reach any other state, although not necessarily in one step. It means the chain is irreducible. Further

$$d(1) = 0, d(2) = 0, d(3) = 0, d(4) = 1$$

This implies that the state 4 is aperiodic.

Example 7.4.2.: Consider Markov chain having $S = \{1,2,3,4\}$ and transition matrix

$$P = \begin{bmatrix} 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 \\ 1/4 & 0 & 1/4 & 1/2 & 0 \\ 0 & 0 & 0 & 1/4 & 3/4 \\ 0 & 0 & 0 & 1 & 0 \end{bmatrix}$$

Is the chain (i) irreducible, (ii) aperiodic?

Solution: The given Markov chain is reducible.

For state 1

$$d(1) = G.C.D.\left\{m: f_{11}^{(m)} > 0\right\}$$

$$= G.C.D.(2) = 2$$

For state 2

$$d(2) = G.C.D.\left\{m: f_{22}^{(m)} > 0\right\} = G.C.D.(2) = 2$$

For state 3

$$d(3) = G.C.D.\left\{m: f_{33}^{(m)} > 0\right\} = G.C.D.(0) = 0$$

For state 4

$$d(4) = G.C.D.\left\{m: f_{44}^{(m)} > 0\right\} = G.C.D.(1) = 1$$

For state 5

$$d(5) = G.C.D.\left\{m: f_{55}^{(m)} > 0\right\} = G.C.D.(1) = 1$$

Thus, state 1, state 2, and state 3 are periodic states and state 4 and state 5 are aperiodic.

Example 7.4.3.: Markov chain having $S = \{1,2,3,4\}$ and transition matrix. Is the chain irreducible? Which states are aperiodic?

$$P = \begin{bmatrix} 1/3 & 2/3 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 1/2 & 0 & 1/2 & 0 \\ 0 & 0 & 1/2 & 1/2 \end{bmatrix}$$

Solution: The given Markov chain is reducible.

For State 1

$$d(1) = G.C.D.\left\{m: f_{11}^{(m)} > 0\right\}$$
$$= G.C.D.(1) = 1$$

For State 2

$$d(2) = G.C.D.\left\{m: f_{22}^{(m)} > 0\right\} = G.C.D.(1) = 1$$

For State 3

$$d(3) = G.C.D.\left\{m: f_{33}^{(m)} > 0\right\}$$
$$= G.C.D.(0) = 0$$

For State 4

$$d(4) = G.C.D.\left\{m: f_{44}^{(m)} > 0\right\}$$
$$= G.C.D.(0) = 0$$

That means, state 1 and state 2 are aperiodic states and state 3 and state 4 have periodic 0.

7.5 Stationary Probability Distributions

Definition: Suppose a_1 and a_2 are real numbers such that $0 < a_1, a_2 < 1, a_1 + a_2 = 1$. Then, the probability Distribution $a = (a_1, a_2)$ is said to be Stationary with respect to a given two-state Markov Chain with the TPM

$$P = \begin{pmatrix} p_{11} & p_{12} \\ p_{21} & p_{22} \end{pmatrix}$$

if the following condition holds:

$$\left.\begin{aligned} a_1 &= a_1 p_{11} + a_2 p_{21} \\ a_2 &= a_1 p_{12} + a_2 p_{22} \end{aligned}\right\} \qquad (7.1)$$

Suppose $P(X_0 = 1) = a_1, P(X_0 = 2) = a_2$, where a_1, a_2 satisfy (7.1), then

$P(X_1 = 1)$

$= P(X_1 = 1|X_0 = 1)P(X_0 = 1) + P(X_1 = 1|X_0 = 2)P(X_0 = 2)$

$= p_{11}a_1 + p_{21}a_2$

$= a_1$

Similarly

$P(X_1 = 2)$

$= p_{12}a_1 + p_{22}a_2$

$= a_2$

$P(X_2 = 1)$

$= P(X_1 = 1)p_{11} + P(X_1 = 2)p_{21}$

$= a_1 p_{11} + a_2 p_{21}$

$= a_1$

$P(X_2 = 2) = a_2$

In general, $\forall n \geq 0$

$P(X_n = 1) = a_1,$

$P(X_n = 2) = a_2.$

First, we derive the limiting n-step transition probability distribution as $n \to \infty$.

Theorem 7.5.1.: Let $|p_{11} + p_{22} - 1| < 1$ and

$$\lim_{n\to\infty} P^{(n)} = A$$

Then

$$A = \begin{bmatrix} \pi_1 & \pi_2 \\ \pi_1 & \pi_2 \end{bmatrix}$$

where

$$\pi_1 = \frac{(1 - p_{22})}{(2 - p_{11} - p_{22})},$$
$$\pi_2 = \frac{(1 - p_{11})}{(2 - p_{11} - p_{22})}.$$

Proof: Since $|p_{11} + p_{22} - 1| < 1$, we have

$$\lim_{n\to\infty} (1 - p_{11} - p_{22})^n = 0.$$

Hence

$$\lim_{n\to\infty} P^{(n)} = \lim_{n\to\infty} [A + (1 - p_{11} - p_{22})^n B] = A.$$

This proves the required result.

Notice that $\pi_1 + \pi_2 = 1$.

From the above theorem 7.5.1., we see that

$$\lim_{n\to\infty} p_{11}^{(n)} = \lim_{n\to\infty} p_{21}^{(n)} = \pi_1,$$

and

$$\lim_{n\to\infty} p_{22}^{(n)} = \lim_{n\to\infty} p_{12}^{(n)} = \pi_2$$

Therefore, for large n, the probability that system occupies the state i is $\pi_i = (i = 1,2)$ irrespective of whether we started initially in state 1 or state 2. Thus, for large n, there is a state of "Statistical equilibrium" or "Steady State". The steady state probabilities are independent of the initial state of the process. $\tilde{\pi} = (\pi_1, \pi_2)$ Gives the limiting probability distribution of the process when the steady state arrives. The smaller the factor $|p_{11} + p_{22} - 1|$, the faster the approach to the steady state.

Notice that if $p_{11} = p_{22}$

$$(\pi_1 =) \lim_{n\to\infty} p_n(S) = \frac{1}{2} = \lim_{n\to\infty} p_n(F) = \frac{1}{2} (= \pi_2)$$

Theorem 7.5.2: The limiting probability distribution $\pi = (\pi_1, \pi_2)$ of a two state Markov Chain is stationary.

Proof: We have

$$\pi_1 p_{11} + \pi_2 p_{21}$$

$$= \frac{(1 - p_{22})}{(2 - p_{11} - p_{22})} p_{11} + \frac{(1 - p_{11})}{(2 - p_{11} - p_{22})} p_{21}$$

$$= \frac{p_{11}(1 - p_{22}) + (1 - p_{11})(1 - p_{22})}{(2 - p_{11} - p_{22})}$$

$$= 1 - \frac{p_{22}}{2 - p_{11} - p_{22}}$$

$$= \pi_1$$

Further

$$\pi_1 p_{12} + \pi_2 p_{22}$$

$$= \frac{(1 - p_{22})(1 - p_{11})}{(2 - p_{11} - p_{22})} p_{12} + \frac{(1 - p_{11}) p_{22}}{(2 - p_{11} - p_{22})}$$

$$= \frac{1 - p_{11}}{2 - p_{11} - p_{22}}$$

$= \pi_2$

Thus, the stationarity condition (7.1) holds for the probability distribution π, so that $\pi = (\pi_1, \pi_2)$ is a stationary probability distribution for the Markov Chain.

Theorem 7.5.3: The stationary distribution of a two state Markov Chain is unique.

Proof: Suppose $\pi = (\pi_1, \pi_2)$ is stationary with respect to the given two state Markov Chain with

$$\pi_1 = \frac{(1 - p_{22})}{(2 - p_{11} - p_{22})},$$
$$\pi_2 = \frac{(1 - p_{11})}{(2 - p_{11} - p_{22})}$$

Then

$$\pi_1 p_{11} + \pi_2 p_{21} = \pi_1,$$
$$\pi_2 p_{12} + \pi_2 p_{22} = \pi_2$$
$$\pi_1 + \pi_2 = \pi_1$$

Let $\pi' = (\pi_1', \pi_2')$ be any other stationary probability distribution. Then by the definition of stationarity

$$\pi_1' p_{11} + \pi_2' p_{21} = \pi_1',$$
$$\pi_1' p_{12} + \pi_2' p_{22} = \pi_2'$$

Which implies that

$$\pi_1' = \frac{1 - p_{22}}{2 - p_{11} - p_{22}} = \pi_1, \pi_2' = 1 - \pi_1' = \pi_2$$

This proves the theorem.

For a Markov Chain with transition probability $\{p_{jk}; j, k \in S\}$, a probability distribution $\{u_j\}$ is called stationary (or invariant) if

$$u_k = \sum_j u_j \, p_{jk}. \quad \left(u_j \geq 0, \sum_j u_j = 1 \right)$$

Further, we obtain

$$u_k = \sum_j u_j \, p_{jk}$$

$$\sum_j \left\{ \sum_i u_i p_{ij} \right\} p_{jk}$$

$$= \sum_j u_j \left\{ \sum_i p_{ij} \, p_{jk} \right\}$$

$$= \sum_j u_j \; p_{ik}^{(z)}$$

In general, we can easily verify that

$$u_k = \sum_j u_j \, p_{ik}^{(n)}, n \geq 1.$$

Example: A Markov Chain $\{X_t, t = 0,1,2, \cdots\}$ on the state space $S = \{1,2,3,4\}$ has the transition probability matrix

$$P = \begin{bmatrix} \frac{1}{2} & \frac{1}{4} & \frac{1}{4} \\ \frac{1}{3} & 0 & \frac{2}{3} \\ \frac{1}{2} & \frac{1}{2} & 0 \end{bmatrix},$$

Find out the stationary distribution.

Solution:

$$a_1 = \frac{1}{2}a_1 + \frac{1}{3}a_2 + \frac{1}{2}a_3$$

$$\Rightarrow a_1 = \frac{2}{3}a_2 + a_3$$

$$a_2 = \frac{1}{4}a_1 + \frac{1}{2}a_3$$

$$a_3 = \frac{1}{4}a_1 + \frac{2}{3}a_2$$

$$a_2 = \frac{9}{10}a_3$$

$$a_1 = \frac{8}{5}a_3$$

$$a_1 + a_2 + a_3 = 1$$

Putting this value in above equation, we get

$$a_1 = \frac{8}{175}, a_2 = \frac{9}{350}, a_3 = \frac{1}{35}.$$

A Markov Chain $\{X_t, t = 0,1,2,\cdots\}$ on the state space $S = \{1,2,3\}$ has the transition probability matrix

$$P = \begin{bmatrix} 1/2 & 1/2 & 0 \\ 1/3 & 1/3 & 1/3 \\ 0 & 1/3 & 2/3 \end{bmatrix},$$

Find out the stationary distribution.

Solution:

$$a_1 = \frac{1}{2}a_1 + \frac{1}{3}a_2$$

$\Rightarrow a_1 = \frac{2}{3}a_2$

$a_2 = \frac{1}{2}a_1 + \frac{1}{3}a_2 + \frac{1}{3}a_3$

$a_3 = \frac{1}{3}a_2 + \frac{2}{3}a_3$

$\Rightarrow a_3 = a_2$

We know that $a_1 + a_2 + a_3 = 1$

Putting a_1 and a_3 in above equation, we get

$a_2 = \frac{3}{8}$,

Then

$a_1 = \frac{2}{3} \times \frac{3}{8} = \frac{1}{4}$,

$a_2 = \frac{3}{8}$,

$a_3 = \frac{3}{8}$.

7.6 Limit Theorems

Show that if state j is transient, then $\sum_{n=1}^{\infty} p_{ij}^{(n)} < \infty$, for all i. As a consequence, it follows that for j transient $P_{ij}^{(n)} \longrightarrow 0$ as $n \longrightarrow \infty$.

Let μ_{jj} denote the expected number of transition needed to return to state j, i.e.,

$$\mu_{jj} = \begin{cases} \infty & \text{if j is transient} \\ \sum_{n=1}^{\infty} n f_{jj}^{n} & \text{if j is recurrent} \end{cases}$$

7.7 Ergodic Chains and Irreducible Ergodic Chains

Definition: A recurrent, non-null and aperiodic state of a Markov Chain is said to be ergodic. A Markov Chain, all of whose states are ergodic, is called an ergodic chain.

Theorem 7.7.1.: If $i \leftrightarrow j$ then $d_i = d_j$.

Proof: Let $i \leftrightarrow j$. Then $\exists$ integers $m, n > 0$ such that

$$p_{ij}^{(m)} > 0, p_{ji}^{(n)} > 0.$$

Let $p_{ji}^{(n)} > 0$, then by Chapman Kolmogorov equations

$$p_{ji}^{(n+s+m)} = \sum_{l \in S} \sum_{u \in S} p_{jl}^{(n)} p_{iu}^{(s)} p_{uj}^{(m)}$$

$$\geq p_{ji}^{(n)} p_{il}^{(s)} p_{ij}^{(m)} > 0.$$

Again, if $p_{ii}^{(s)} > 0$, we have

$$p_{ii}^{(2s)} = \sum_{u \in S} p_{iu}^{(s)} \, p_{ui}^{(s)}$$

$$\geq \left[p_{ii}^{(s)}\right]^2 > 0.$$

Further $p_{ii}^{(2s)} > 0$ implies that

$$p_{ji}^{(n+2s+m)} > 0.$$

It follows that d_j divides $(n + 2s + m) - (n + s + m) = s$.

This is true $\forall$ s for which $p_{ii}^{(s)} > 0$. Thus, d_j divides d_i. Interchanging the roles of i and j in the above proof, we also conclude that d_i divides d_j. Hence $d_i = d_j$. This leads to the required result.

Theorem 7.7.2.: From a recurrent state a recurrent state can only be obtained.

Proof: Let i be a given recurrent state of the Markov Chain. Let j be any other state which can be obtained from i. Let k be the smallest positive path (length) from i to j such that $p_{ij}^{(k)} = \alpha > 0$. Obviously, the transition from i to j in k steps can not be through i. thus, the probability of a return from j to i must be greater than 0, otherwise the probability of the process not returning to state i must be at least α so that the probability of eventual return to state i is less than $1 - \alpha\ (< 1)$ which contradicts the fact that the i^{th} state is recurrent. Hence $\exists$ a least integer m such that

$p_{jj}^{(m)} = \beta\ (\text{say}) > 0.$

Now for any integer n

$$p_{ii}^{(k+n+m)} \geq p_{ij}^{(k)}\ p_{jj}^{(n)}\ p_{ji}^{(m)}\ \geq\ \alpha\,\beta\,p_{jj}^{(n)}$$

$$p_{jj}^{(m+n+k)} \geq p_{ji}^{(m)}\ p_{ii}^{(n)}\ p_{ij}^{(k)}\ \geq\ \alpha\,\beta\,p_{ii}^{(n)}$$

Thus $\lim_{n\to\infty} p_{ii}^{(n)} = 0$ iff $\lim_{n\to\infty} p_{jj}^{(n)} = 0,$ so that $\sum p_{ii}^{(n)}$ and $\sum p_{jj}^{(n)}$ coverage or diverge together. Since i is recurrent $\sum p_{ii}^{(n)}$ diverges so that $\sum p_{jj}^{(n)}$also diverges. Hence state j is also recurrent. This leads to the required result.

Example 7.7.1.: Consider Markov chain having $S = \{1,2,3,4\}$ and transition probability matrix

$$P = \begin{bmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 \\ 1/4 & 1/8 & 1/8 & 1/2 \end{bmatrix}$$

Show that all states of the above Markov Chain are ergodic.

Solution: The chain is irreducible.

The probability of ultimate return of the states is

For State 1:

$f_{11}^{(1)} = p_{11} = 0$

$f_{11}^{(2)} = f_{12}f_{21} = 0$

$f_{11}^{(3)} = 0$

$\vdots$

$$f_{11} = \sum_{n=1}^{\infty} f_{11}^{(n)} = 0 + 0 + \cdots = 0 < 1$$

Hence the state 1 is transient.

The Mean Recurrence Time is

$$\mu_{11} = 1 \times f_{11}^{(1)} + 2 \times f_{11}^{(2)} + 3 \times f_{11}^{(3)} + \cdots$$

$$= 1 \times 0 + 2 \times 0 + 3 \times 0 + \cdots$$

$$= 0$$

For State 2:

$f_{22}^{(1)} = p_{22} = 0$

$f_{22}^{(2)} = f_{21}f_{12} = 0$

$f_{22}^{(3)} = f_{21}f_{11}f_{12} = 0$

$\vdots$

$$f_{22} = \sum_{n=1}^{\infty} f_{22}^{(n)}$$

$$= 0 + 0 + \cdots = 0 < 1$$

Hence the state 2 is transient.

The Mean Recurrence Time is

$$\mu_{22} = 1 \times f_{22}^{(1)} + 2 \times f_{22}^{(2)} + 3 \times f_{22}^{(3)} + \cdots$$

$$= 1 \times 0 + 2 \times 0 + 3 \times 0 + \cdots = 0$$

For State 3:

$$f_{33}^{(1)} = p_{33} = 0$$

$$f_{33}^{(2)} = f_{32} f_{23} = 0$$

$$f_{33}^{(3)} = f_{32} f_{22} f_{23} = 0$$

$$\vdots$$

$$f_{33} = \sum_{n=1}^{\infty} f_{33}^{(n)}$$

$$= 0 + 0 + 0 + \cdots = 0 < 1$$

Hence the state 3 is transient.

The Mean Recurrence Time is

$$\mu_{33} = 1 \times f_{33}^{(1)} + 2 \times f_{33}^{(2)} + 3 \times f_{33}^{(3)} + \cdots$$

$$= 1 \times 0 + 2 \times 0 + 3 \times 0 + \cdots = 0$$

For State 4:

$$f_{44}^{(1)} = p_{44} = \frac{1}{2}$$

$$f_{44}^{(2)} = \frac{1}{8}$$

$$f_{44}^{(3)} = \frac{1}{8}$$

$$f_{44}^{(4)} = \frac{1}{4}$$

$$\vdots$$

$$f_{44}^{(n)} = 0, \qquad n > 0$$

$$f_{44} = \sum_{n=1}^{\infty} f_{44}^{(n)}$$

$$= \frac{1}{2} + \frac{1}{8} + \frac{1}{8} + \frac{1}{4} + 0 + 0 + \cdots$$

$$= \frac{8}{8} = 1$$

Hence the state 4 is persistent.

The Mean Recurrence Time is

$$\mu_{44} = 1 \times f_{44}^{(1)} + 2 \times f_{44}^{(2)} + 3 \times f_{44}^{(3)} + 4 \times f_{44}^{(4)} + \cdots$$

$$= 1 \times \frac{1}{2} + 2 \times \frac{1}{8} + 3 \times \frac{1}{8} + 4 \times \frac{1}{4} + \cdots$$

$$= \frac{1}{2} + \frac{1}{4} + \frac{3}{8} + 1$$

$$= \frac{17}{8} < \infty$$

State 4 is aperiodic.

Thus, State 4 is ergodic.

Example 7.7.2.: Consider Markov chain having $S = \{1,2,3\}$ and transition matrix

$$P = \begin{bmatrix} 1/2 & 1/2 & 0 \\ 3/4 & 0 & 1/4 \\ 0 & 1 & 0 \end{bmatrix}$$

Show that all states of the above Markov Chain are ergodic?

Solution: The chain is irreducible.

The probability of ultimate return of the states is

For State 1

$$f_{11}^{(1)} = p_{11} = \frac{1}{2}$$

$$f_{11}^{(2)} = f_{12}f_{21}$$

$$= \frac{1}{2} \times \frac{3}{4} = \frac{3}{8}$$

$$f_{11}^{(3)} = f_{12}f_{22}f_{21} = 0$$

$$f_{11}^{(4)} = f_{12}f_{23}f_{32}f_{21}$$

$$= \frac{1}{2} \times \frac{1}{4} \times 1 \times \frac{3}{4} = \frac{3}{32}$$

$$f_{11}^{(5)} = f_{12}f_{23}^{(2)}f_{32}^{(2)}f_{21}$$

$$= \frac{1}{2} \times \frac{1}{4} \times 1 \times \frac{1}{4} \times 1 \times \frac{3}{4} = \frac{3}{128}$$

⋮

$$f_{11} = \sum_{n=1}^{\infty} f_{11}^{(n)}$$

$$= \frac{1}{2} + \frac{3}{2 \times 4} + \frac{3}{2 \times 4^2} + \frac{3}{2 \times 4^3} + \cdots$$

$$= \frac{7}{2 \times 4} + \frac{3}{2 \times 4^2} + \frac{3}{2 \times 4^2} + \cdots = 1$$

Hence the state 1 is persistent.

The Mean Recurrence Time is

$$\mu_{11} = 1 \times f_{11}^{(1)} + 2 \times f_{11}^{(2)} + 3 \times f_{11}^{(3)} + \cdots$$

$$= 1 \times \frac{1}{2} + 2 \times \frac{1}{2} \times \frac{3}{4} + 3 \times 0 + 4 \times \frac{1}{2} \times \frac{1}{4} \times 1 \times \frac{3}{4} + \cdots$$

$$= \frac{11}{6}$$

Thus State 1 is a non-null persistent.

For State 2

$$f_{22}^{(1)} = p_{22} = 0$$

$$f_{22}^{(2)} = f_{21} f_{12} + f_{23} f_{32}$$

$$= \frac{3}{4} \times \frac{1}{2} + \frac{1}{4} \times 1 = \frac{5}{8}$$

$$f_{22}^{(3)} = f_{21} f_{11} f_{12}$$

$$= \frac{3}{4} \times \frac{1}{2} \times \frac{1}{2} = \frac{3}{16}$$

⋮

$$f_{22} = \sum_{n=1}^{\infty} f_{22}^{(n)}$$

$$= 0 + \frac{5}{8} + \frac{3}{16} + \frac{3}{32} + \cdots$$

$$= \frac{1}{4} + \frac{3}{4} = 1$$

Hence the state 2 is persistent.

The Mean Recurrence Time is

$$\mu_{22} = 1 \times 0 + 2 \times f_{22}^{(2)} + 3 \times f_{22}^{(3)} + \cdots$$

$$= 1 \times 0 + 2 \times \frac{5}{8} + 3 \times \frac{3}{16} + 3 \times \frac{3}{16} + \cdots$$

$$= \frac{1}{2} + \frac{9}{4} = \frac{11}{4}$$

Thus State 2 is a non-null persistent.

For State 3

$$f_{33}^{(1)} = p_{33} = 0$$

$$f_{33}^{(2)} = f_{32} f_{23} = 1 \times \frac{3}{4} = \frac{3}{4}$$

$$f_{33}^{(3)} = 1 \times \frac{1}{4} \times \frac{1}{2} = \frac{1}{8}$$

$$\vdots$$

$$f_{33} = \sum_{n=1}^{\infty} f_{33}^{(n)} = 0 + \frac{3}{4} + \frac{1}{8} + \cdots = 1$$

Hence the state 3 is persistent.

The Mean Recurrence Time is

$$\mu_{33} = 1 \times f_{33}^{(1)} + 2 \times f_{33}^{(2)} + 3 \times f_{33}^{(3)} + \cdots$$

$$= 1 \times 0 + 2 \times \frac{3}{4} + 3 \times \frac{1}{8} + \cdots = \frac{15}{8}$$

Thus State 2 is a non-null persistent.

7.9 Summary

We are furnishing, in the following, a summary of the discussions in this unit:

- Introduced the concept of communicating classes and closed sets.
- Defined the irreducible Markov Chain.
- Discuss periodic and aperiodic Markov chains.
- Defined stationary probability distribution.

7.10 Self-Assessment Exercise

1. Consider a Transition Probability Matrix:

$$P = \begin{bmatrix} p & 1-p \\ 1-p & p \end{bmatrix}$$

Where $p \in (0,1)$

Is State 2 accessible from state 1?

2. Consider a Markov Chain with TPM

$$P = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 1/8 & 0 & 7/8 & 0 \end{bmatrix}$$

Draw the transition graph and find the accessible state and commutative class.

3. Using the Markov chain in example 7.3.1. Are Class 1, Class 2 and Class 4 aperiodic?

4.The transition probability matrix of a Markov Chain is

$$P = \begin{bmatrix} 0.1 & 0.5 & 0.4 \\ 0.6 & 0.2 & 0.2 \\ 0.3 & 0.4 & 0.3 \end{bmatrix}$$

Which states are irreducible?

5. Define (i) An Ergodic Markov Chain, (ii) Stationary Markov Chain.

6. Let C_1 and C_2 be two communicative classes of a Markov chain and "s" be a state, which belongs to C_1 but not C_2. Prove that C_1 and C_2 are disjoint.

7. A Markov Chain $\{X_t, t = 0,1,2,\cdots\}$ on the state space $S = \{1,2,3,4\}$ has the transition probability matrix

$$P = \begin{bmatrix} 0 & 1 & 0 & 0 \\ \frac{1}{2} & 0 & \frac{1}{2} & 0 \\ 0 & \frac{1}{2} & 0 & \frac{1}{2} \\ 0 & 0 & 1 & 0 \end{bmatrix},$$

From this transition probability matrix, find out the stationary distribution.

8. A Markov Chain $\{X_t, t = 0,1,2,\cdots\}$ on the state space $S = \{1,2,3,4\}$ has the transition probability matrix

$$P = \begin{bmatrix} 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 \\ 3/4 & 1/4 & 0 & 0 \\ 1/3 & 1/3 & 1/3 & 0 \end{bmatrix},$$

Find out the stationary distribution.

9. Show that the limiting probability distribution $\pi = (\pi_1, \pi_2)$ of a two state Markov chain is stationary.

10. For a two state Markov chain, under suitable assumptions, derive the expression for n-step transition probabilities. Also obtain the limiting probability distribution $\lim_{n\to\infty} p_{ij}$ and show that this limiting probability distribution is stationary.

7.11 References

- Cinlar. E.: "Introduction to Stochastic Processes", Prentice Hall Inc., Englewood Cliffs New Jersey.
- Bartlett, M.S. (1978): "Stochastic processes", 3rd edition, CUP, Cambridge.
- Cox, D. R., & Miller, H. D. (1965): "The Theory of Stochastic Processes". Chapman and Hall/CRC.
- Chung, Kai. Lai.: "Markov Chains with Stationary Transition Probabilities"
- Doob, J. L. (1953): "Stochastic Processes". John Wiley & Sons.
- Grimmett, G., & Stirzaker, D. (2001): "Probability and Random Processes". Oxford University Press.
- Karlin, S., & Taylor, H. M. (1981): "A First Course in Stochastic Processes". Academic Press.
- Lawler, G. F. (2006): "Introduction to Stochastic Processes". Chapman and Hall/CRC.
- Pavliotis, G. A. (2014): "Stochastic Processes and Applications: Diffusion Processes, the Fokker-Planck and Langevin Equations". Springer.
- Ross, S. M. (2014): "Introduction to Probability Models", Academic Press.
- Trivedi, K. S. (2002): "Probability and Statistics in Engineering Systems", Prentice Hall International.

7.12 Further Readings

- Grimmett, G., & Stirzaker, D. (2001): "Probability and Random Processes", Oxford University Press.
- Karatzas, I., & Shreve, S. E. (1991): "Brownian Motion and Stochastic Calculus", Springer-Verlag.
- Lefebvre, M. (2006): "Applied Stochastic Processes", Springer.

- Levin, D. A., Peres, Y., & Wilmer, E. L. (2009): “Markov Chains and Mixing Times”, American Mathematical Society.
- Resnick, S. I. (1992): “Adventures in Stochastic Processes”, Birkhäuser.
- Ross, S. M. (2014): “Stochastic Processes”, Wiley.
- Taylor, H. M., & Karlin, S. (1998): “An Introduction to Stochastic Modelling”, Academic Press.

UNIT–8 CONTINUOUS TIME MARKOV PROCESSES

Structure

8.1 Introduction

A continuous-time Markov process is a type of stochastic process that models the evolution of systems where changes occur at random points in time. Unlike discrete-time Markov chains, which are indexed by discrete steps, continuous-time Markov processes are indexed by continuous time. This means the system can transition between states at any moment, making it more suited to modelling real-world phenomena where events happen unpredictably and continuously.

One of the defining features of a continuous-time Markov process is the Markov property. This property ensures that the future evolution of the process depends only on its current state, not on the history of how it arrived there. Essentially, the process has no "memory" of its past beyond the present moment, simplifying the analysis of such systems.

Continuous-time Markov processes are widely used in various fields, including physics, biology, economics, and queuing theory. They can model a range of phenomena, such as the decay

of radioactive particles, the spread of diseases, or the arrival of customers at a service centre. The behaviour of these processes is typically described using a rate matrix, which specifies the rate at which transitions between states occur, rather than the probabilities of such transitions over a fixed period.

Understanding continuous-time Markov processes is crucial for analysing systems where changes occur continuously over time, allowing for the prediction and control of complex dynamic systems in a variety of applications.

8.2 Objectives

After studying this unit, you should be able to:

- Understand the basic concepts of the continuous time Markov chain and also discuss about the inter-arrival time.
- Define stopping time and explain Wald's equation.
- Describe Doob's optional stopping theorem.
- Discuss forward and backward equations for Kolmogorov differential equation.

8.3 Markov Processes in Continuous-Time

The discrete-time Markov chains $\{X_n, n = 0,1,2,\dots\}$ make transitions at integer times with transition probabilities $P(X_{n+1} = j|X_n = i)$. Thus, the chain can only stay in each state for an integer amount of time before making the next transition.

If we change the integer duration to continuous transition times according to an exponential distribution, then we can obtain a continuous-time Markov chain.

Definition: Let $\{X(t), t \geq 0\}$ be a continuous time stochastic process taking on values in the set of nonnegative integers. Let T_i denotes the amount of time it spends in the state i before making a transition into a different state. Then, it is called a continuous-time Markov chain, if each time it enters state i:

- T_i has an exponential distribution with mean, say, $1/v_i$ i.e., $T_i \sim Exp(v_i)$.
- When the process leaves the state i, it enters state j with some probability p_{ij} satisfying

$$p_{ij} \geq 0, \sum_{j} p_{ij} = 1.$$

Example 8.3.1.: Poisson processes are continuous-time Markov chains having states 0,1,2, ... The process always proceeds from state i to state $i + 1$, i.e., $p_{ii+1} = 1$, where $i \geq 0$. The transition rate vector is $v = (v_i)$ with $v_i = \lambda$. This process is known as a pure birth process since. Whenever a transition occurs, the state of the system is always increased by one.

Definition: The process $\{X(t), t \geq 0\}$ is called a continuous-time Markov chain if $\forall\, s, t \geq 0$ and $\forall$ non negative integers $i, j, X(u), 0 \leq u < s$,

$$P(X(t+s) = j | X(s) = i, X(u) = x(u), 0 \leq u < s)$$

$$= P(X(t+s) = j | X(s) = i)$$

Thus, the probability that the chain will be in the state j after time t depends only on the current state i and is independent of the states the chain has visited in past, and how long the chain has been in the state i.

The continuous-time Markov chain is said to have stationary (or homogeneous) transition probabilities, if $P(X(t+s) = j | X(s) = i)$ is independent of s. Then

$$p_{ij}(t) = P(X(t+s) = j | X(s) = i).$$

8.3.1 Birth and Death Processes

Consider a continuous time stochastic process $\{X(t), t \geq 0\}$ with the state at time t representing the number of people in the system at time t. Thus, the state space of the process is $\{0,1,2,\dots\}$. If there are i people in the system, then we assume that:

(i) The new arrivals enter the system at an exponential rate λ_i, and

(ii) people leave the system at an exponential rate μ_i.

The process is called a birth and death process, with arrival (or birth) rates $(\lambda_i)_{i=0}^{\infty}$ and departure (or death) rates $(\mu_i)_{i=0}^{\infty}$ (with $\mu_0 = 0$).

Theorem 8.3.1.: For any birth and death process, the transition rates are

$v_0 = \lambda_0,\ \ v_i = \lambda_i + \mu_i;\ i > 0$

and the transition probabilities are

$p_{0,1} = 1, \qquad p_{i,i+1} = \frac{\lambda_i}{\lambda_i + \mu_i}, i > 0.$

Proof: Let the process be in the state i at present and the waiting time for the next arrival be $T_i^{(a)} \sim \text{Exp}(\lambda_i)$. Further, the waiting time for the next departure $T_i^{(d)} \sim \text{Exp}(\mu_i)$. Thus, the waiting time for the next transition is

$T_i = \min\left(T_i^{(a)}, T_i^{(d)}\right).$

Then $T_i \sim \text{Exp}(v_i), v_i = \lambda_i + \mu_i$. The transition probabilities are

$$p_{i,i+1} = P\left(T_i^{(a)} < T_i^{(d)}\right)$$

$$= \frac{\lambda_i}{\lambda_i + \mu_i}$$

$$p_{i,i-1} = 1 - p_{i,i+1}$$

$$= \frac{\mu_i}{\lambda_i + \mu_i}, i \geq 1.$$

In particular,

$p_{01} = \frac{\lambda_0}{\lambda_0 + 0} = 1\ (\mu_0 = 0).$

Example: The Poisson process is a birth process with $\lambda_i = \lambda$ and $\mu_i = 0\ \forall i$.

Example: Yule Process

A birth process with a linear birth rate, $\lambda_i = i\lambda, \mu_i = 0$ is called a Yule process.

Example: A model with

$\lambda_i = i\lambda + \theta, \qquad \mu_i = i\mu.$

is called a linear growth process with immigration. Such processes occur in the study of biological reproduction and population growth. Each individual in the population is assumed to give birth at the rate λ. In addition, there is a constant rate of increase θ due to an external source such as immigration. Hence, the total birth rate, where there are i persons in the system, is $i\lambda + \theta$. Deaths are assumed to occur at an exponential rate μ for each member of the population, and hence $i\mu$.

Example: The Queuing System M/M/1: Suppose that the customers arrive at a single server station following a Poisson process having rate λ. The customer goes directly into service if the server is free and, if not, the customer joins the queue. The customer leaves the system after he gets the service and the next customer in the queue, if any, enters service. The service times are assumed to be independent exponential random variables having mean $1/\mu$.

The first M denotes that the interarrival process is Markovian (since it is a Poisson process). The second M denotes that the service distribution is exponential (and, hence, Markovian). The 1 refers to a single server.

If X denote the number in the system (queuing time+service time) at time t, then $\{X(t), t \geq 0\}$ is a birth and death process with $\lambda_i = \lambda, i \geq 0,$ and $\mu_i = \mu; i \geq 1.$

Example (The Queuing System M/M/s): Consider an exponential queuing system in which there are s servers available, each serving at rate μ. An entering customer first waits in line and then goes to the first free server. Assuming customers arrive according to a Poisson process with rate λ, then the process is a birth-death process with parameters arrival rates: $\lambda_i = \lambda,$ for each $i \geq 0,$ and departure rates $\mu_i = \min(i, s).\mu$ for each $i \geq 1.$

Consider now a general birth and death process with birth rates $\{\lambda_i\}$ and death rates $\{\mu_i\}$ with $\mu_0 = 0$.

Let U_i be the time, starting from the state i, it takes for the process to enter the state $i + 1$, for any $i \geq 0$.

Theorem 8.3.2.: We have

$$E(U_0) = \frac{1}{\lambda_0}, \quad E(U_i) = \frac{1}{\lambda_i} + \frac{\mu_i}{\lambda_i} E(U_{i-1}), i \geq 1.$$

Proof: We recursively compute $E(U_i), i \geq 1$ by starting with $i = 0$. Since $U_0 = T_0$ is exponential with the rate λ_0, we have $E(U_0) = \frac{1}{\lambda_0}$.

For $i \geq 1$, we condition on the first transition which takes the process into state $i-1$ or $i+1$:

$$E(U_i) = \frac{1}{\lambda_i + \mu_i} + \frac{\lambda_i}{\lambda_i + \mu_i}.0\frac{\mu_i}{\lambda_i + \mu_i}.[E(U_{i-1}) + E(U_i)]$$

Hence

$$E(U_i) = \frac{1}{\lambda_i} + \frac{\mu_i}{\lambda_i}E(U_{i-1}), \qquad i \geq 1.$$

We state the following recursive relation for the variance of U_i without proof:

$$\text{Var}\,(U_0) = \frac{1}{\lambda_0^2}$$

$$\text{Var}\,(U_i) = \frac{1}{\lambda_i(\lambda_i + \mu_i)} + \frac{\mu_i}{\lambda_i}\text{Var}(U_{i-1}) + \frac{\mu_i}{\lambda_i + \mu_i}[E(U_{i-1}) + E(U_i)]^2, i \geq 1$$

Corollary: If $\lambda_i = \lambda$ and $\mu_i = \mu\ \forall i$, then

$$E(U_i) = \begin{cases} \dfrac{1 - \left(\frac{\mu}{\lambda}\right)^{i+1}}{\lambda - \mu}, & \lambda \neq \mu \\ \dfrac{i+1}{\lambda}, & \lambda = \mu \end{cases}$$

Proof: This is a direct application of the following result:

If $a_{i+1} = c + d.a_i, i = 0,1,2 \ldots$, then

$$a_i = \begin{cases} ci + a_0, & d = 1 \\ c\dfrac{1 - d^i}{1 - d} + d^i a_0, & d \neq 1 \end{cases}$$

The expected time for the process to transition from i to $j > i$ is

$E(U_i) + \cdots + E(U_{j-1})$

and the variance of the overall transition time from i to j is

$Var(U_i) + \cdots + Var(U_{j-1})$.

The transition probability function $p_{ij}(t)$

Let

$p_{ij}(t) = P(X(t+s) = j | X(s) = i)$

$= P(X(t) = j | X(0) = i)$

We consider two different cases for the Markov chain and find formulas for $p_{ij}(t)$:

(i) Pure birth process with distinct birth rates,

(ii) General continuous-time Markov chains.

Pure-Birth Process with Distinct Birth Rates

Now we derive the explicit formula for the transition probability function in the case of a pure birth process ($\mu_i = 0, v_i = \lambda_i$) having distinct birth rates ($\lambda_i \neq \lambda_j$). Before proving the main result, first we determine the distribution of a sum of independent exponential random variable with distinct rates in the following proposition.

Proposition: If $X_i \sim Exp(\lambda_i), i = 1,2$ are independent random variables having distinct rates ($\lambda_1 \neq \lambda_2$), then

$$f_{X_1+\cdots+X_n}(T) = \sum_{k=1}^{n} C_{k,1,n} \lambda_k e^{-\lambda_k t} ,$$

$$C_{k,i,j} = \prod_{i \leq l \leq n, l \neq k} \frac{\lambda_l}{\lambda_l - \lambda_k}$$

Proof: We prove only the special case of n = 2. We have

$f_{X_1+X_2}(t)$

$$= \int_0^t f_{X_1}(s) f_{X_2}(t-s) ds$$

$$= \int_0^t \lambda_1 e^{-\lambda_1 s}. \lambda_2 e^{-\lambda_2 (t-s)} ds$$

$$= \lambda_1 \lambda_2 e^{-\lambda_2 t} \int_0^t e^{-(\lambda_1 - \lambda_2)s} ds$$

$$= \lambda_1 \lambda_2 e^{-\lambda_2 t} \frac{1}{\lambda_1 - \lambda_2} \left(1 - e^{-(\lambda_1 - \lambda_2)t}\right)$$

$$= \frac{\lambda_1}{\lambda_1 - \lambda_2} \lambda_2 e^{-\lambda_2 t} + \frac{\lambda_2}{\lambda_2 - \lambda_1} \lambda_1 e^{-\lambda_1 t}$$

The survival function of the hyper exponential random variable $S = X_1 + \cdots X_n$ is

$$P(S > t) = \sum_{k=1}^{n} C_{k,1,n} e^{-\lambda_k t}.$$

Theorem 8.3.3.: For a pure birth process having distinct rates,

$$p_{ii}(t) = P(T_i > t)$$
$$= e^{-\lambda t}$$

$$p_{ij}(t) = \sum_{k=1}^{j} C_{k,i,j} e^{-\lambda_k t}, \quad i < j$$

where

$$C_{k,i,j} = \prod_{i \le l \le j, l \ne k} \frac{\lambda_l}{\lambda_l - \lambda_k}.$$

Proof: First, we write

$$p_{ij}(t) = P(X(t) < j + 1 | X(0) = i) - P(X(t) < j | X(0) = i).$$

Next, letting T_k represent the duration of the chain in the state k, we have

$$P(X(t) < j+1 | X(0) = i) = P(T_i + \cdots + T_{j+1} > t)$$

and similarly,

$$P(X(t) < j | X(0) = i) = P\left(T_i + \cdots + T_j > t\right).$$

Hence, utilizing the above proposition, the theorem follows.

Note: For $j = i + 1$

$$p_{ij}(t) = \frac{\lambda_{i+1}}{\lambda_{i+1} - \lambda_i} e^{-\lambda_i t} + \frac{\lambda_i}{\lambda_i - \lambda_{i+1}} e^{-\lambda_{i+1} t} - e^{-\lambda_i t}$$

$$= \frac{\lambda_i}{\lambda_{i+1} - \lambda_i} \left(e^{-\lambda_{i+1} t} - e^{-\lambda_i t}\right)$$

For $\lambda_k = k\lambda \ \forall\, k \geq 1$ (Yule process); $\lambda_i \neq \lambda_j$ if $i \neq j$. Suppose X_0=1. Then

$$p_{1j}(t) = e^{-\lambda t}\left(1 - e^{-\lambda t}\right)^{j-1}$$

Hence the conditional distribution of $X(t)$ given $X(0) = 1$ is

$$X(t) | (X(0) = 1) \sim \text{Geometric distribution with } \left(p = e^{-\lambda t}\right).$$

That is, starting with a single individual, the population size at time t has a geometric distribution with mean $e^{-\lambda t}$. If initially there are i individuals, then the population size at time t has a negative binomial distribution with parameters $(i, p = e^{-\lambda t})$.

General Continuous-Time Markov Chains

We prove the Chapman-Kolmogorov equations for continuous Markov Chains.

Chapman-Kolmogorov Equations

Theorem 8.3.4.: For all $s, t \geq 0$

$$p_{ij}(t+s) = \sum_k p_{ik}(t)p_{kj}(s) \qquad (8.1)$$

Proof: We have

$$p_{ij}(t+s) = P(X(t+s) = j | X(0) = i)$$

$$= \sum_k P(X(t+s) = j, X(t) = k | X(0) = i)$$

$$= \sum_k P(X(t+s) = j | X(t) = k, X(0) = i\ )\, P(X(t) = k | X(0) = i)$$

$$= \sum_k P(X(t+s) = j | X(t) = k\ )\, P(X(t) = k | X(0) = i) \quad \text{(Using Markovian property)}$$

$$= \sum_k p_{kj}(s)p_{ik}(t).$$

Hence the result follows.

Note: Chapman–Kolmogorov equation in matrix form is $P(t+s) = P(t)P(s)$.

From the Chapman-Kolmogorov equations we can obtain the following differential equations for all $p_{ij}(t)$.

8.4 Backward and Forward Equations for Homogeneous Case

Kolmogorov's Backward Equations

Theorem 8.4.1.: In any continuous-time Markov chain,

$$p'_{ij}(t) = \sum_{k \neq i} q_{ik}p_{kj}(t) - v_i p_{ij}(t)$$

where $q_{ik} = v_i p_{ik}$ are called the instantaneous transition rates from state i to state k.

Proof: For any $h > 0$

$$\frac{p_{ij}(t+h) - p_{ij}(t)}{h}$$

$$= \frac{1}{h}\left(\sum_{k} p_{ik}(h)p_{kj}(t) - p_{ij}(t)\right)$$

$$= \sum_{k \neq i} \frac{p_{ik}(h)}{h} p_{kj}(t) - \frac{1 - p_{ii}(h)}{h} p_{ij}(t)$$

It remains to show that

$$\lim_{h \to 0} \frac{p_{ik}(h)}{h} = q_{ik}, \quad k \neq i$$

$$\lim_{h \to 0} \frac{p_{ii}(h)}{h} = v_i$$

Since

$$p_{ii}(h) = P(X(h) = i | X(0) = i)$$

$$= P(T_i > h) = e^{-hvi},$$

we have

$$\lim_{h \to 0} \frac{1 - p_{ii}(h)}{h}$$

$$= \lim_{h \to 0} \frac{1 - e^{-hvi}}{h} v_i.$$

Further

$$p_{ik}(h) = P(X(h) = k | X(0) = i)$$

$$= P(T_i < h) p_{ik}$$

$$= \left(1 - e^{-hvi}\right) p_{ik}.$$

Thus

$$\lim_{h\to 0}\frac{p_{ik}(h)}{h}=\lim_{h\to 0}\frac{\left(1-e^{-hvi}\right)p_{ik}}{h}$$

$$=v_i p_{ik}.$$

Note: We have

$$\sum_k q_{ik}=v_i\sum_k p_{ik}=v_i,$$

so that

$$p_{ik}=\frac{q_{ik}}{v_i}=\frac{q_{ik}}{\sum_k q_{ik}}.$$

Note: In a birth and death process, the instantaneous transition rates are just birth and death rates

$$q_{i,i+1}=v_i p_{i,i+1}$$

$$=(\lambda_i+\mu_i)\frac{\lambda_i}{\lambda_i+\mu_i}=\lambda_i$$

$$q_{i,i-1}=v_i p_{i,i-1}$$

$$=(\lambda_i+\mu_i)\frac{\mu_i}{\lambda_i+\mu_i}=\mu_i$$

Example 8.4.1.: A continuous-time Markov chain consisting of two states:

Consider a machine that works for an exponential amount of time having the mean $1/\lambda$ before breaking down. Suppose it takes an exponential amount of time having mean $1/\mu$ to repair the machine. If the machine is in working condition at time 0, what is the probability that it will be working at time t=10?

Solution: The process is a birth and death process with state 0 meaning that the machine is working and state 1 it is being repaired. The parameters are

$\lambda_0 = \lambda, \lambda_1 = 0, \mu_1 = \mu, \mu_0 = 0, \text{and } p_{01} = 1 = p_{10}$

The Chapman-Kolmogorov backward equations are

$$p'_{00}(t) = \lambda\left(p_{10}(t) - p_{00}(t)\right)$$
$$p'_{10}(t) = \mu(p_{00}(t) - p_{10}(t))$$

In matrix form

$$\begin{pmatrix} p_{00}(t) \\ p_{10}(t) \end{pmatrix}' = \begin{pmatrix} -\lambda & \lambda \\ \mu & -\mu \end{pmatrix} \begin{pmatrix} p_{00}(t) \\ p_{10}(t) \end{pmatrix}$$

The initial conditions are

$$p_{00}(0) = 1, p_{10}(0) = 0$$

We decompose the matrix as

$$\begin{pmatrix} -\lambda & \lambda \\ \mu & -\mu \end{pmatrix} = \begin{pmatrix} -1 & \lambda \\ 1 & -\mu \end{pmatrix} \begin{pmatrix} 0 & 0 \\ 0 & -\lambda - \mu \end{pmatrix} \begin{pmatrix} -1 & \lambda \\ 1 & -\mu \end{pmatrix}^{-1}$$

Thus, the solution is given by

$$\begin{pmatrix} p_{00}(t) \\ p_{10}(t) \end{pmatrix}$$

$$= \begin{pmatrix} -1 & \lambda \\ 1 & -\mu \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & e^{-(\lambda+\mu)t} \end{pmatrix} \begin{pmatrix} -1 & \lambda \\ 1 & -\mu \end{pmatrix}^{-1} \begin{pmatrix} 1 \\ 0 \end{pmatrix}$$

$$= \begin{pmatrix} \frac{\mu}{\lambda+\mu} + \frac{\lambda}{\lambda+\mu} e^{-(\lambda+\mu)t} \\ \frac{\mu}{\lambda+\mu} - \frac{\lambda}{\lambda+\mu} e^{-(\lambda+\mu)t} \end{pmatrix}$$

Similarly, we can obtain the following Kolmogorov's forward equations (proof is not given here).

Kolmogorov's Forward Equations

Theorem 8.4.2.: Under suitable regularity conditions,

$$p'_{ij}(t) = \sum_{k \neq i} q_{ik} p_{kj}(t) - v_i p_{ij}(t)$$

Note: For a pure birth process, Kolmogorov's forward equations become

$$p'_{ij}(t) = \lambda_{j-1} p_{ij-1}(t) - \lambda_j p_{ij}(t)$$

which leads to

$$p_{ij}(t) = 0 \quad j < i$$

$$p_{ii}(t) = e^{-\lambda_i t},$$

$$p_{ij}(t) = \lambda_{j-1} e^{-\lambda_j t} \int_0^t e^{\lambda_j s} p_{i,j-1}(s)\, ds, \quad j \geq i+1.$$

Limiting Probabilities

For each state j of a continuous time Markov-chain, let

$$P_j = \lim_{t \to \infty} p_{ij}(t)$$

The limit exists and is independent of the initial state i if all states communicate and chain is positive recurrent.

The $\{P_j\}$ are called the stationary probabilities. Here, $P_j's$ also have the interpretation of being the long run proportion of time that the process is in state j.

We have the following result:

Theorem 8.4.3.: We have

$$v_j P_j = \sum_{k \neq j} q_{kj} p_k \qquad \left(\sum P_j = 1\right)$$

Proof: Let $t \to \infty$ in the forward equations and use $\lim_{t\to\infty} p'_{ij}(t) = 0$.

Example 8.4.2.: For a birth and death process, a necessary and sufficient condition for the limiting probabilities to exist is

$$\sum_{i=1}^{\infty} \frac{\lambda_0\lambda_1 \ldots \lambda_{i-1}}{\mu_1\mu_2 \ldots \mu_i} < \infty$$

In this case, it can be shown that

$$P_0 = \frac{1}{1 + \sum_{i=1}^{\infty} \frac{\lambda_0\lambda_1 \cdots \lambda_{i-1}}{\mu_1\mu_2 \cdots \mu_i}},$$

$$P_i = \frac{\lambda_0\lambda_1 \ldots \lambda_{i-1}}{\mu_1\mu_2 \ldots \mu_i} P_0, \quad n \geq 1$$

For the M/M/1 queue $(\lambda_i = \lambda, \mu_i = \mu)$,

$$P_i = \frac{(\lambda/\mu)'}{1 + \sum_{i=1}^{\infty} (\lambda/\mu)^i} = (\lambda/\mu)^i (1 - \lambda/\mu)$$

provided $\frac{\lambda}{\mu} < 1$.

Solution: For the given birth and death process,

For $j = 0$: $\lambda_0 P_1 = \mu_1 P_1$

For $j \geq 1$: $(\lambda_j \mu_j) P_j = \lambda_{j-1} P_{j-1} + \mu_{j+1} P_{j+1}$

By induction

$$\lambda_j \mu_j = \mu_{j+1} P_{j+1}, \ j \geq 0$$

or equivalently,

$$\frac{P_{j+1}}{P_j} = \frac{\lambda_j}{\mu_{j+1}}, \quad j \geq 0$$

Multiplying such equations from $j = 0$ to $j = i - 1$ gives that

$$P_i = \frac{\lambda_0 \lambda_1 \ldots \lambda_{i-1}}{\mu_1 \mu_2 \ldots \mu_i} P_0, \quad i \geq 1.$$

Using $P_0 + \sum_{i=1}^{\infty} P_i = 1$ we can find

$$P_0 = \frac{1}{1 + \sum_{i=1}^{\infty} \frac{\lambda_0 \lambda_1 \cdots \lambda_{i-1}}{\mu_1 \mu_2 \cdots \mu_i}}.$$

8.5 Interval Arrival Time

Consider a Poisson process $\{X(t), t \geq 0\}$. Let W_n denote the time of n^{th} event, $n = 1,2, \ldots$ The arrival time of the n^{th} event, W_n, is also called the waiting time until the n^{th} event. The process starts in state $X(0) = 0$ and stays in state 0 for a random amount of time W_1 until it jumps to state $X(W_1) = 1$. Then it stays in state 1 for a random amount of time until it jumps to state 2 at time W_2, and so on.

Let $W_0 = 0$ and let $T_n = W_{n+1} - W_n$; $n = 0,1,2, \ldots$. The random variables T_n are known as the interevent times or interarrival times or holding times.

Example 8.5.1.: Let $X_1, X_2, \ldots$ be the interarrival times and $N(t)$ be the Poisson process with intensity $\lambda = 4$.

(i) Calculate the probability that the first arrival occurs after $t = 0.5$, i.e.

$P[X_1 > 0.5]$.

(ii) Find $P[X_1 > 3]$ when no arrivals before $t = 1$.

(iii) Given that the 3^{rd} arrival occurred at time $t = 2$, find the probability that the 4^{th} arrival occurs after $t = 4$.

(iv) Let T be the time of the 1^{st} arrival after $t = 10$. Find the E(T) and Var (T).

(v) Find the conditional expectation and conditional variance of [T|event that the last arrival occurred at $t = 9$].

Solution:

(i) Since $X_1 \sim \text{Exp}(4)$, we have

$P[X_1 > 0.5] = \exp(-4 \times 0.5)$

$= \exp(-2)$

≈ 0.13534

(ii) $P[X_1 > 3 | X_1 > 1] = P[X_1 > 2]$

$= \exp(-2 \times 4)$

$= \exp(-8)$

$= 0.00034$

(iii) Time between the 3^{rd} and the 4^{th} arrival is $X_4 \sim \text{Exp}(4)$. Thus, the conditional probability is

$P[X_4 > 4 | X_1 + X_2 + X_3 = 4] = P[X_4 > 4]$ (Independence of the $X_i's$)

$= \exp(-4 \times 4)$

$= \exp(-16)$

≈ 0.00000

(iv) Watching the process at time $t = 10$, the time of the first arrival from $t = 10$ is $\text{Exp}(4)$ i.e. $T = 10 + X$

Thus,

$$E(T) = E(10) + E(X)$$

$$= 10 + \frac{1}{4}$$

$$= \frac{41}{4} = 10.25$$

$$Var(T) = Var(X) = \frac{1}{4}$$

(v) $E[T|\text{event that the last arrival occurred at } t = 9] = E[T] = \frac{41}{4}$

$$Var\,[T|\text{event that the last arrival occurred at } t = 9] = Var\,[T] = \frac{1}{4}$$

8.6 Stopping Time

Let $X_1, X_2, \ldots$ denote a sequence of independent random variables.

Definition: An integer-valued random variable N is said to be a Stopping Time for the sequence $X_1, X_2, \ldots$, if the event $\{N = n\}$ is independent of $X_{n+1}, X_{n+2}, \ldots$ for all $n = 1,2, \ldots$.

$$P(X_n = 0) = P(X_n + 1) = \frac{1}{2} \;,\; n = 1,2, \ldots$$

Example 8.6.1.: Let $X_n\,, n = 1,2, \ldots$ be independent and such that

Let us take $N = \min\{n: X_1 + X_2 + \cdots + X_n = 10\}$. Then N is a stopping time (Since N is depending upon $X_1, X_2, \ldots, X_n$ and $X_i's$ are independent, therefore N is independent of $X_{n+1}, X_{n+2}, \ldots$).

Example 8.6.2.: Consider the Renewal Process $\{N(t), t \geq 0\}$ induced by $\{X_n, n \geq 1\}$. Also consider the integer-valued random variable $N(t)$ for any t. If $N(t)$ a stopping time for $\{X_n, n \geq 1\}$ look at the event $\{N(t) = n\}$. See that

$$\{N(t) = n\} \Leftrightarrow \{S_n \leq t \leq S_{n+1}\}$$

$\{S_n \leq t \; and \; S_n + S_{n+1} > t\}$ depends on X_{n+1}.

Hence $N(t)$ is not a stopping time for $\{X_n, n \geq 1\}$.

Consider the event $\{N(t) + 1 = n\}$

$\{N(t) = n - 1\} \Leftrightarrow \{S_{n-1} \leq t \ \ and \ \ S_n > t\}$

Since $\{S_{n-1} \leq t \ , \ S_{n-1} + X_n > t\}$ does not depend on X_{n+1}, X_{n+2}, ...

So, $\{N(t) + 1\}$ is a stopping time for $\{X_n, n \geq 1\}$.

8.7 Wald's Equation

Theorem 8.7.1.: Let X_1, X_2, ... be independent and identically distributed random variables with the finite mean $E(X)$. Let $N > 0$ be an integer-valued random variable, independent of $\{X_1, X_2, \ldots\}$, with $E(N) < \infty$. Then

$$E\left(\sum_{n=1}^{N} X_n\right) = E(N)E(X)$$

Proof: We know that,

$$\sum_{n=1}^{N} X_n = \sum_{n=1}^{N} X_n \times 1 + \sum_{n=N+1}^{\infty} X_n \times 0$$

Let

$$I_n = \begin{cases} 1 & if \ n \leq N \\ 0 & if \ n > N \end{cases}$$

Then

$$\sum_{n=1}^{N} X_n = \sum_{n=1}^{N} X_n \times I_n + \sum_{n=N+1}^{\infty} X_n \times I_n$$

$$\Rightarrow \sum_{n=1}^{N} X_n = \sum_{n=1}^{\infty} X_n \times I_n$$

$$\Rightarrow E\left(\sum_{n=1}^{N} X_n\right) = E\left(\sum_{n=1}^{\infty} X_n \times I_n\right)$$

$$= \sum_{n=1}^{\infty} E\,(X_n \times I_n)$$

If $I_n = 1$, then we have not stopped after successively observing $X_1, X_2, X_3, \ldots, X_{n-1}$.

Therefore $I_n$is determined by $X_1, X_2, X_3, \ldots, X_{n-1}$ and it is independent of X_n.

Thus,

$$E\left(\sum_{n=1}^{N} X_n\right) = \sum_{n=1}^{\infty} E\,(X_n)\,E(I_n)$$

$$= E(X) \sum_{n=1}^{\infty} E\,(I_n) \qquad (X_n's \ \ are\ iid)$$

$$= E(X) \sum_{n=1}^{\infty} P[N \geq n]$$

$$\Rightarrow E\left(\sum_{n=1}^{N} X_n\right) = E(X)E(N)$$

8.7.1 Applications of Wald's Equation

1. Suppose $\{X_n\}$ is a discrete distribution that is uniform over the integers $\{1, 2, , \ldots, 10\}; P(X = i) = \frac{1}{10}$, $1 \leq i \leq 10$. Thus $E(X) = 5.5$. Imagine that these are bonuses (in units of Rs.10, 000)

that are given to you by your employer each year. Let $N = \min\{n \geq 1 : X_n = 6\}$, the 1^{st} time that you receive a bonus of size 6. What is the expected total (cumulative) amount of bonus received up to time N?

$$E\left(\sum_{n=1}^{N} X_n\right) = E(X)E(N) = 5.5\,E(N) \qquad (8.2)$$

In (8.2), we have used Wald's equation. Here, N is a stopping time with finite mean.

Now $\{N = 1\} = \{X_1 = 6\}$ and in general $\{N = n\} = \{X_1 \neq 6, X_2 \neq 6, \dots, X_{n-1} \neq 6, X_n = 6\}$ only depends on $\{X_1, X_2, \dots, X_n\}$.

Again

$P[N = 1\} = P[X_1 = 6] = 0.1$ and in general, from the i.i.d. assumption placed on $\{X_n\}$,

$$P[N = n] = P[X_1 \neq 6, X_2 \neq 6, \dots, X_{n-1} \neq 6, X_n = 6]$$

$$= P[X_1 \neq 6]P[X_2 \neq 6] \dots P[X_n = 6]$$

$$= (0.9)^{(n-1)}0.1\,, \quad n \geq 1$$

We conclude that N has a geometric distribution with "success" probability $p = 0.1$, and hence

$$E(N) = \frac{1}{p} = 10$$

Putting this value in (8.2), we get

$$E\left(\sum_{n=1}^{N} X_n\right) = 5.5 \times \frac{1}{10} = 55$$

Notice that before time $N = n$ the random variables $\{X_1, X_2, \dots, X_{n-1}\}$ no longer have the original uniform distribution; they are biased as none of them takes on the value 6.

Then each have the conditional distribution $(X|X \neq 6)$ and thus an expected value different from 5.5. The random variable at time $N = n$ has value 6; $X_n = 6$ and hence is not random at all. All these random variables are biased, in the end, on average, Wald's equation lets the sum be unbiased and independent of N.

We would get the same answer 55 by using the stopping times $N = \min\{n \geq 1 : X_n = k\}$ for any $1 \leq k \leq 10$.

1. Null Recurrence of the Simple Symmetric Random Walk

Let

$R_n = \Delta_1 + \Delta_2 + \cdots + \Delta_n \, , X_0 = 0$

where $\{\Delta_n : n \geq 1\}$ are i.i.d. with $P(\Delta = \pm 1) = 0.5, E(\Delta) = 0$.

We know that this Markov Chain is recurrent (proved via the gambler's ruin problem). i.e., the random time $N_{0,0} = \min\{n \geq 1 : R_n = 0 | R_0 = 0\}$ is proper.

Show that the chain is null recurrent, i.e., $E(N_{0,0}) = \infty$. The chain will, with certainty, return back to state 0, but the expected number of steps required is infinite.

We do so by proving that $E(N_{1,1}) = \infty$, where we define the stopping times $N_{i,j} = \min\{n \geq 1 : R_n = j | R_0 = i\}$. (Since the chain is irreducible, all states are null recurrent together or positive recurrent together; so, if $E(N_{1,1}) = \infty$ then in fact $E(N_{j,j}) = \infty$ for all j.) By symmetry, $N_{j,j}$ has the same distribution (hence mean) for all j.

The first step $\Delta_1 = \pm 1$,

$$E(N_{1,1}) = \frac{\{1 + E(N_{2,1})\}}{2} + \frac{\{1 + E(N_{0,1})\}}{2}$$

$$= 1 + (0.5)E(N_{2,1}) + (0.5)E(N_{0,1})$$

Showing that $E(N_{0,1}) = \infty$, thus proves that $E(N_{1,1}) = \infty$.

By definition, the chain at time $N = N_{0,1}$ has value 1;

$$1 = R_N = \sum_{n=1}^{N} \Delta_n$$

We use Wald's equation. Since $E(N) < \infty$ then we conclude that

$$1 = E(R_N)$$

$$= E\left(\sum_{n=1}^{N} \Delta_n\right)$$

$$= E(N)E(\Delta) = 0$$

yielding the contradiction $1 = 0$. Thus $E(N) = \infty$.

8.8 Doob's Optional-Stopping Theorem (Statement without Proof)

Let $M = \{M_n\}$ be a martingale with respect to $X_0, X_1, \cdots$ and N be a stopping time satisfying any one of the following conditions:

(i) There is a positive integer R such that $N(t) \leq R \ \ for\ all\ t \in \Omega$.

(ii) There is a positive real number K such that $|X_n(t)| < K$ for all n and all $t \in \Omega$, and N is almost surely finite.

(iii) $E(N) < \infty$, and there is a positive real number K such that

$|X_n(t) - X_{n-1}(t)| < K \ \forall\ n\ and\ all\ t \in \Omega$

Then X_N is integrable, and

$E(X_N) = E(X_0)$.

8.9 Summary

In this unit, we have discussed the following points:

- We have discussed Markov processes in Continuous time
- We have state and proved Backward and Forward equations for homogeneous case.
- We have explained Interval arrival time with example.
- We defined Stopping Time and Wald's equation
- We have stated Doob's Optional Stopping theorem

8.10 Self-Assessment Exercise

1. Explain continuous time Markov process with example.

2. A salesman flies between New Delhi, Mumbai, and Kolkata as the following rates:

$$P = \begin{bmatrix} -5 & 3 & 2 \\ 2 & -4 & 2 \\ 6 & 1 & -7 \end{bmatrix}$$

(i) If the salesman takes a trip out of New Delhi, what is the probability of it being a trip to Kolkata?

(i) If the salesman takes a trip out of New Delhi, what is the probability of it being a trip to Kolkata?

(ii) What is the expected time spent in New Delhi?

3. Derive the forward equation for a time-homogeneous process.

4. State and prove the backward equation for a homogeneous case.

5. Define Wald's equations and prove it.

8.11 References

- Anderson, T. W., and L. A. Goodman. (1957): "Statistical inference about Markov chains", Ann. Math. Stat., Vol. 28, pp. 89-109.
- Bartlett, M. S. (1955): “An Introduction to Stochastic Processes”, London: Cambridge University Press.
- Biflingsley, P. (1961): "Statistical methods in Markov chains," Ann. Math. Stat., VoL 32, pp. 12-40.
- Cinlar, E. (1975): “Introduction to Stochastic Processes”, Prentice Hall.
- Dover. Gardiner, C. (2009): “Stochastic methods: A handbook for the natural sciences”. Springer, 4th edition.
- Feller, W. (1968): “Introduction to Probability and Applications”, New Age India International.
- Friedman, A. (2004): “Stochastic differential equations and applications”.
- Hoel, P. G., Port, S. C. and Stone, C. J. (1991): “Introduction to Stochastic Processes”, University Book Stall.
- John, F. (1978): “Partial differential equations”. Spring, 3rd edition.
- Karlin, S. and Taylor, H. M. (1995): “A First Course in Stochastic Processes”, Academic Press.
- Koralov, L. B. and Sinai, Y. G. (2010): “Theory of Probability and Random Processes”. Springer.
- Medhi, J. (2002): “Stochastic Process", New Age International Publishers.
- Oksendal, B. (2005): “Stochastic Differential Equations”. Springer, 6 edition.
- Pavliotis, G. A. (2014): “Stochastic Processes and Applications”. Springer.
- Varadhan, S. R. S. (2007): “Stochastic Processes”, volume 16 of Courant Lecture Notes. American Mathematical Society.

8.12 Further Reading

- Gikhman, I. and Skorokhod, A. V. (2004): “The theory of stochastic processes II. Classics in Mathematics”. Springer-Verlag.
- Grimmett, G., & Stirzaker, D. (2001): “Probability and Random Processes”, Oxford University Press.

- Karlin, S. and Taylor, H. M. (1981): “A second course in stochastic processes”. Academic Press.
- Lefebvre, M. (2006): “Applied Stochastic Processes”, Springer.
- Levin, D. A., Peres, Y., & Wilmer, E. L. (2009): “Markov Chains and Mixing Times”, American Mathematical Society.
- Resnick, S. I. (1992): “Adventures in Stochastic Processes”, Birkhäuser.
- Ross, S. M. (2014): “Stochastic Processes”, Wiley.
- Taylor, H. M., & Karlin, S. (1998): “An Introduction to Stochastic Modelling”, Academic Press.

U.P. Rajarshi Tandon Open University, Prayagraj

MScSTAT – 203(N)/ MASTAT – 203(N) Stochastic Process

Block: 3 Random Walk and Queuing Process

Course Design Committee

Dr. Ashutosh Gupta — **Chairman**
Director, School of Sciences
U. P. Rajarshi Tandon Open University, Prayagraj

Prof. Anoop Chaturvedi — **Member**
Department of Statistics
University of Allahabad, Prayagraj

Prof. S. Lalitha — **Member**
Department of Statistics
University of Allahabad, Prayagraj

Prof. Himanshu Pandey — **Member**
Department of Statistics, D. D. U. Gorakhpur University, Gorakhpur.

Prof. Shruti — **Member-Secretary**
School of Sciences
U.P. Rajarshi Tandon Open University, Prayagraj

Course Preparation Committee

Dr. Pratyasha Tripathi — **Writer**
Department of Statistics
Tilka Manjhi Bhagalpur University, Bhagalpur, Bihar

Prof. Anoop Chaturvedi — **Editor**
Department of Statistics
University of Allahabad, Prayagraj

Prof. Shruti — **Course / SLM Coordinator**
School of Sciences
U. P. Rajarshi Tandon Open University, Prayagraj

First Edition: ***July 2024***
ISBN : 978-93-48270-94-8

MScSTAT – 203N/ MASTAT – 203N
STOCHASTIC PROCESS

Printed and Published by Col. Vinay Kumar, Registrar, Uttar Pradesh Rajarshi Tandon Open University Prayagraj, 2024
Printed by: Cygnus Information Solution Pvt.Ltd., Lodha Supremus Saki Vihar Road Andheri East, Mumbai.

Block & Units Introduction

The ***Block – 3 - Random Walk and Queuing Process*** is the third block of said SLM, which is divided into four units.

The ***Unit – 9 - Random Walk and Gambler's Ruin Problem*** deals with the Random walk, Brownian motion as a random walk, one-dimensional, two-dimensional and three-dimensional random walks, duality in random walk and gambler's ruin problem.

The ***Unit – 10 - Queuing Process*** deals with the Birth and death processes, renewal process, Queuing models- Specification & Effectiveness, Measures, the $E_k/M/1$, $M/E_k/1$; M/M/1; M/M/k & M/G/1 queuing process.

The ***Unit – 11- Distributions*** deals with the Compound distribution, Machine Interference Problem, Waiting Time Distribution for M/M/1 and M/M/k models,

The last unit of this Block is ***Unit – 12 – Martingales*** discussed about the Martingales, Boob – Decomposition, Martingale convergence theorems.

At the end of every block/unit the summary, self-assessment questions are given.

UNIT-9: RANDOM WALK AND GAMBLER'S RUIN PROBLEM

Structure

9.1 Introduction

The concept of the random walk has played a significant role in the development of modern probability theory and has broad applications in various scientific fields. The formal study of random walks began with Karl Pearson's inquiry in 1905, where he posed a seemingly simple question about the distance a person would cover after taking a series of steps in random directions. This laid the groundwork for understanding random processes. Random Walk or Drunkard's Walk is a stochastic process that represents a path consisting of a sequence of random steps on a mathematical space.

Albert Einstein's seminal work in 1906 on Brownian motion further established the importance of random walks by explaining the erratic movement of pollen particles in water, first observed by Robert Brown in 1827. Einstein's model, which treated these movements as continuous-time random walks, provided a crucial link between theoretical mathematics and physical phenomena.

Norbert Wiener's development of the mathematical foundation for Brownian motion in the 1920s, known as the Wiener process, and Paul Lévy's work on extending the theory to include Lévy flights in the mid-20th century, contributed significantly to the field. Lévy flights generalize

random walks to include steps of varying lengths, making the model applicable to more complex systems.

Recent Developments

In recent years, the study and application of random walks have seen substantial advancements:

***Complex Networks*:** Random walk theory has been instrumental in analyzing complex networks, such as social networks, biological networks, and the internet. This has led to insights into network structure and dynamics, influencing algorithms like Google's PageRank.

***Anomalous Diffusion and Lévy Flights*:** Real-world processes often exhibit anomalous diffusion, characterized by non-Gaussian distributions. Models incorporating Lévy flights, which include both short and long steps, have been developed to better represent these processes, with applications in ecology, finance, and optimization algorithms.

***Quantum Random Walks*:** Extending random walks to the quantum realm has introduced new properties and potential applications in quantum computing. Quantum random walks offer possible speed-ups for computational tasks and algorithms, representing a significant frontier in computing research.

***Machine Learning*:** In machine learning, random walk techniques have been integrated into graph-based learning algorithms. Methods like Node2Vec use random walks to learn representations of nodes in large graphs, aiding tasks such as link prediction and node classification.

***Financial Models*:** Random walk theory underpins many financial models, including those based on the Efficient Market Hypothesis (EMH). Recent developments continue to refine these models by incorporating factors like stochastic volatility and jump processes, providing deeper insights into market behavior.

***Biological and Medical Applications**:* Random walks model various biological processes, such as animal movement, disease spread, and cell behavior. Advancements in imaging and tracking

technologies have enabled more detailed studies, leading to improved models, and understanding of these phenomena.

The study of random walks has evolved from a mathematical curiosity to a cornerstone of probability theory with extensive applications. Historical contributions from Pearson, Einstein, Wiener, and Lévy established a robust foundation, while recent advancements continue to expand the theory's applicability. This unit will explore the fundamentals of random walks, their properties, and their applications, providing a comprehensive understanding of this essential concept.

9.2 Objectives

By the end of this unit, you should be able to:

Understand What Random Walks Are:

- Learn what a random walk is and why it is important.
- Know how to describe a random walk using simple mathematics.

Learn About Brownian Motion:

- Find out how Brownian motion relates to random walks.
- Understand the history and key figures like Einstein who studied it.
- See the difference between continuous and step-by-step random walks.

Explore Random Walks in Different Dimensions:

- Discover how random walks work in one, two, and three dimensions.
- Understand how these walks can model different complex systems.

Understand Duality in Random Walks:

- Learn what duality means in the context of random walks.
- Look at examples and see how dual processes can simplify problems.

Study the Gambler's Ruin Problem:

- Understand the gambler's ruin problem and its significance.
- Develop skills to solve this problem.

- Discuss how this problem applies to real-world scenarios like finance.

Solve Problems Using Random Walk Concepts:

- Use your knowledge of random walks to solve practical problems.
- Apply random walk models to areas like physics, biology, finance, and computer science.
- Use computer simulations to study random walks.

Learn About Recent Developments in Random Walks:

- Stay updated on new research and applications of random walks.
- Understand the impact of quantum random walks, complex networks, and unusual diffusion patterns.
- See how these new ideas affect science and technology today.

Test and Reflect on Your Knowledge:

- Check your understanding with self-assessment questions.
- Reflect on how random walk theory applies to different fields.
- Identify areas where you can learn more and improve.

By achieving these objectives, you will have a solid understanding of random walk theory and how it applies to various fields, preparing you for further study or practical use in your work.

9.3 Random Walk

A random walk is a mathematical model that describes a path consisting of a sequence of random steps. This model is used to represent various phenomena in physics, economics, biology, and many other fields. The simplicity and versatility of random walks make them a fundamental concept in probability theory and stochastic processes.

Mathematical Definition

A random walk can be formally defined as follows:

i. One-Dimensional Random Walk:

- Consider a particle starting at position $S_0 = 0$ on a number line.

- At each step n, the particle moves to a new position S_n by taking a step X_n, where X_n is a random variable representing the step size and direction.
- The position after n steps is given by:

$$S_n = S_{n-1} + X_n$$

- Typically, X_n are independent and identically distributed (i.i.d.) random variables. For a simple symmetric random walk, X_n takes values +1 or −1 with equal probabilities 0.5.

One-dimensional random walk can also be defined as:

$n \in \{0,1, \dots, a\}$: $a + 1$ positions on a straight line.

A person starts at k, $(0 < k < a)$.

- The walk starts at stage k at step 0 and moves to either stage $k - 1$ or stage $k + 1$ after one step. The walk continues until either 0 or a is reached.
- p: P(Walker goes forward one place to k + 1)
- $q = 1 - p$: P(Walker goes back one place to $k - 1$)

Position of a walker after having moved n times is known as the state of the walk after n steps.

- A random walk is said to be symmetric if $p = q = 1/2$.
- If walk is bounded, then ends are known as barriers
- If walk ends once a barrier is reached, then barriers are called “absorbing”.

Examples:

1. Brownian Motion ⇒ Path traced by a molecule as it moves through a liquid or gas exhibits characteristics of a random walk, known as Brownian motion. This is a continuous random walk and is essential in the study of diffusion processes.

2. Animal Foraging ⇒ Search pattern of a foraging animal can be modeled as a random walk, helping ecologists understand animal behavior and movement patterns in nature.

3. Stock Prices ⇒ Fluctuating price of a stock is often modeled as a random walk.

4. Gambling $\Rightarrow$ Gain/losses status of a gambler can be described by a random walk. Each bet is a step that either increases or decreases his total money based on the outcome.

Higher-Dimensional Random Walks: A random walk can also be defined in higher dimension.

- In two dimensions, the particle moves on a plane. Each step is represented by a vector

$$\boldsymbol{X}_n = (X_n, Y_n) \text{ where } X_n \; and \; Y_n \text{ are i.i.d. random variables.}$$

- The position after n steps is given by:

$$S_n = S_{n-1} + \boldsymbol{X}_n$$

Where $S_n = \left(S_n^x, S_n^y\right)$ and

$$S_n^x = S_{n-1}^x + X_n$$

$$S_n^y = S_{n-1}^y + Y_n$$

- Similarly, in three dimensions Each step is represented by a vector $\boldsymbol{X}_n = (X_n, Y_n, Z_n)$

- The position after n steps is given by:

$$S_n = S_{n-1} + \boldsymbol{X}_n$$

Where $S_n = \left(S_n^x, S_n^y, S_n^z\right)$ and

$$S_n^x = S_{n-1}^x + X_n$$

$$S_n^y = S_{n-1}^y + Y_n$$

$$S_n^z = S_{n-1}^z + Z_n$$

ii. Unrestricted random walks

Simple random walk on a line ⇒ A step forward $(+1)$ has probability p and a step back (-1) has probability $q(= 1 - p)$. At the i^{th} step a Bernoulli random variable X_i is observed, and the position of the walk at the n^{th} step is the random variable:

$$S_n = S_0 + \sum_{i=1}^{n} X_i = S_{n-1} + X_n$$

S_0: Initial state or origin of the random walk

$$E(X_i) = p - q,\ E(X_i^2) = p + q = 1$$

$$V(X_i) = 1 - (p - q)^2 = (1 - p + q)(1 + p - q) = 4pq$$

Consider *random walks without barriers* or *unrestricted random walks*

State S can take any of the values $\{\dots, -2, -1, 0, 1, 2, \dots\}$

We are interested in (i) the position of the walk after a number of steps, (ii) probability of a return to the origin.

Position of the walk at step n simply depends on the position at the $(n-1)^{th}$ step and has Markov property.

One-step Transition Probabilities:

$$P(S_n = j | S_{n-1} = j - 1) = p$$

$$P(S_n = j | S_{n-1} = j + 1) = q = 1 - p$$

If $S_0 = 0$,

$$E(S_n) = \sum_{i=1}^{n} E(X_i) = n(p - q)$$

$$V(S_n) = 4npq$$

$V(S_n)$ grows with increasing n.

$\max_p V(S_n) = n$ (for $p = 1/2$)

If $p > 1/2$, expect a drift in the positive direction,

If $p < 1/2$, expect a drift in the negative direction.

For symmetric random walk, $p = q = 1/2$,

$$E(S_n) = 0$$

$$V(S_n) = n$$

Since $E(S_n) = n(p-q)$, $V(X_n) = 4npq$, using the central limit theorem, as $n \to \infty$, the asymptotic distribution of

$$Z_n = \frac{S_n - n(p-q)}{4npq} \approx N(0,1)$$

$$\text{or } S_n \approx N(n(p-q), 4npq)$$

Exact Probability Distribution

We assume that $X_0 = 0$ and X_n is position at the n^{th} step.

R_n is the number of right steps (+1),

L_n: Number of left steps (-1)

N: Number of steps

Then

$S_n = R_n - L_n$

$N = R_n + L_n$

$\Rightarrow R_n = \frac{1}{2}(N + S_n),\ L_n = \frac{1}{2}(N - S_n)$

$v_n = P(S_n = x) =$ P(Walk is at position x after n steps)

For reaching $S_n = x$ after $n > |x|$ steps require $r = \frac{1}{2}(n + x)$ (+1) steps and $l = \frac{1}{2}(n - x)$ (-1) steps. If x is an odd (even) integer then n is also odd (even).

Number of ways can $r = \frac{1}{2}(n + x)$ steps be chosen from n is

$$h_{n,x} = \frac{n!}{r!\,(n-r)!} = \binom{n}{r}$$

Then

$$v_{n,x} = \binom{n}{r} p^{\frac{1}{2}(n+x)} q^{\frac{1}{2}(n-x)}$$

Remark: A one-dimensional random walk is recurrent, meaning it will return to the origin infinitely often with probability 1. In two dimensions, the random walk is also recurrent. In three dimensions and higher, the random walk is transient, meaning there is a non-zero probability that the particle will never return to the origin.

Examples and Applications

Brownian Motion: Brownian motion is a continuous-time random walk. It describes the random movement of particles suspended in a fluid and can be modeled as the limit of a discrete random walk where the step size and time interval tend to zero.

Stock Prices: In finance, the random walk hypothesis suggests that stock prices follow a random walk. This implies that price changes are independent of each other and have the same probability distribution, making it difficult to predict future prices based on past trends.

Biological Processes: Random walks model various biological processes, such as the movement of animals searching for food, the spread of diseases, and the diffusion of molecules within cells.

Search Algorithms: Random walks are used in search algorithms and optimization techniques. For example, the PageRank algorithm used by Google to rank web pages is based on a random walk through the web graph.

One-Dimensional Random Walk: Imagine a person standing at the center of a straight path. At each time step, they flip a coin. If it lands heads, they take a step forward; if it lands tails, they take a step backward. Over time, their position forms a zigzag pattern on the path.

Two-Dimensional Random Walk: Picture a drunkard walking in a park. At each step, they choose a random direction and move a fixed distance in that direction. Their path forms a series of random turns and steps, creating a scattered trail over the park.

To better understand random walks, one can simulate them using computational tools. A simple simulation involves generating a sequence of random steps and plotting the resulting path. This can help visualize the behavior and properties of random walks in different dimensions.

By studying random walks, we gain insights into various stochastic processes and their applications across different fields. The concepts and properties of random walks provide a foundational understanding for more complex models and theories in probability and statistics.

9.4 Brownian Motion as a Random Walk

Brownian motion is a continuous-time stochastic process that serves as a mathematical model for the random motion observed in particles suspended in a fluid. This concept can be seen as a natural extension of the discrete random walk to continuous time and space. Brownian motion, also known as a Wiener process, has profound applications in various fields such as physics, finance, and biology.

The phenomenon of Brownian motion was first observed by the botanist Robert Brown in 1827, who noticed the erratic movement of pollen grains in water. It was not until 1905 that Albert Einstein provided a theoretical explanation for this motion, linking it to the random collisions of the pollen particles with molecules in the fluid. This work was pivotal in validating the atomic theory of matter.

Mathematical Definition

Brownian motion can be defined as a limit of discrete random walks, with the step size and time interval tending to zero. Formally, a Brownian motion B(t) is a continuous-time stochastic process that satisfies the following properties:

- $B(0) = 0$. $B(t)$ has independent increments.

- The increments $B(t) - B(s)$ are normally distributed with mean zero and variance $t - s$ for $0 \leq s < t$.
- $B(t)$ has continuous paths.

One-Dimensional Brownian Motion

In one-dimensional Brownian motion, a particle moves along a line where its position at time t, denoted by $B(t)$, evolves as a continuous limit of a discrete random walk. The particle's movement can be described by the following stochastic differential equation:

$$dB(t) = \sigma dW(t)$$

where $W(t)$ is a standard Wiener process and σ is the volatility parameter.

Key Properties include:

Mean and Variance: The expected value $E[B(t)] = 0$, and the variance $Var(B(t)) = t$.

Path Continuity: The paths of $B(t)$ are continuous but nowhere differentiable.

Two-Dimensional Brownian Motion

In two dimensions, Brownian motion can be represented as a vector

$$B(t) = \Big(B_x(t), B_y(t)\Big),$$

where $B_x(t)$ and $B_y(t)$ are independent one-dimensional Brownian motions. The movement of a particle in the plane can be visualized as a random walk with infinitesimally small steps in random directions.

Properties Include:

Isotropy: The motion is isotropic, meaning it has no preferred direction, and the distribution of particle positions is radially symmetric.

Distribution: After time t, the particle's position is normally distributed with mean $(0,0)$ and covariance matrix tI, where I is the identity matrix.

Three-Dimensional Brownian Motion

In three dimensions, Brownian motion extends to a vector

$$B(t) = \Big(B_x(t), B_y(t), B_z(t)\Big),$$

with $B_x(t), B_y(t)$, and $B_z(t)$ being independent one-dimensional Brownian motions. This models the random movement of a particle in space.

Properties include:

Radial Symmetry: The particle's motion is radially symmetric around the origin.

Distribution: The particle's position after time t is normally distributed with mean vector $(0,0,0)$ and covariance matrix tI.

Applications and Examples

Physics: Brownian motion models the random movement of particles in fluids. It is used to describe diffusion processes and to study molecular dynamics.

Finance: The Geometric Brownian Motion model is used to describe the evolution of stock prices. The Black-Scholes option pricing model relies on this concept.

Biology: Brownian motion models the movement of microorganisms and the diffusion of molecules within cells.

Visualization

One-Dimensional: Imagine a particle moving along a straight line, where its position changes continuously but unpredictably over time. This results in a jagged, non-smooth path.

Two-Dimensional: Picture a particle drifting randomly on a flat surface, changing direction at every instant. The path forms a complex, erratic pattern.

Three-Dimensional: Envision a particle moving randomly in all directions in space, creating a tangled, chaotic trajectory.

Simulation

To simulate Brownian motion, one can generate a series of random steps in continuous time. For a one-dimensional Brownian motion, the position B(t) at time t can be approximated by summing small, normally distributed increments over small time intervals. This method can be extended to higher dimensions by generating independent random steps for each coordinate.

Brownian motion extends the concept of a random walk to continuous time and space, providing a powerful tool for modeling random processes. Understanding its properties in one, two, and three dimensions allows us to apply this model to a wide range of phenomena in science and engineering. Through mathematical rigor and simulation, Brownian motion helps us gain insights into the inherent randomness of various systems.

Theorems on Brownian Motion

1. Mean and Variance of Brownian Motion

Theorem: For a one-dimensional Brownian motion B(t):

$$E[B(t)] = 0$$
$$Var(B(t)) = t$$

Proof: Mean: By definition, $B(t)$ has independent increments, and

$$B(t) - B(s) \sim N(0, t-s) \text{ for } 0 \leq s < t.$$

Since $B(t) - B(s)$ is normally distributed with mean 0, we have

$$E[B(t) - B(s)] = 0.$$

Setting $s = 0$, gives $E[B(t) - B(0)] = E[B(t)] = 0$, as $B(0) = 0$. This gives, $E[B(t)] = 0$.

Variance: Since $B(t) - B(s) \sim N(0, t-s)$, the variance of $B(t) - B(s)$ is t−s.

Setting $s = 0$, we get $Var(B(t)) = t - 0 = t$.

2. Brownian Motion Has Continuous Paths

Theorem: Brownian motion $B(t)$ almost surely has continuous paths.

Proof: By definition, a Brownian motion $B(t)$ is a continuous-time stochastic process.

The Wiener process, which defines Brownian motion, has the property that for any sequence of times $t_1, t_2, ...$, the paths $B(t)$ are continuous with probability 1.This property follows from the fact that $B(t)$ is a limit of sums of independent, normally distributed random variables with smaller and smaller time steps, leading to continuous paths in the limit.

3. Self-Similarity of Brownian Motion

Theorem: Brownian motion $B(t)$ is self-similar. For any $c > 0$, the process $\{B(ct)\}_{t\geq 0}$ has the same distribution as $\{\sqrt{c}B(t)\}_{t\geq 0}$.

Proof: Consider $\{B(ct)\}_{t\geq 0}$. By definition, $B(ct) \sim N(0, ct)$. The scaled process $\{\sqrt{c}B(t)\}_{t\geq 0}$ has $\sqrt{c}B(t) \sim N(0, ct)$. Both the processes have the same distribution since their increments are normally distributed with mean 0 and variance ct. Therefore $\{B(ct)\}_{t\geq 0}$ and $\{\sqrt{c}B(t)\}_{t\geq 0}$ are identically distributed, proving the self-similarity.

4. Markov Property of Brownian Motion

Theorem: Brownian motion $B(t)$ has the Markov property. Thus, the future evolution of the process depends only on the current state, not on the past history.

Proof: By definition, Brownian motion has independent increments.

Given the current state $B(t) = x$, the future increment $B(t + s) - B(t) \sim N(0, s)$ is independent of the past values $B(u)$ for $u \leq t$. Therefore, the distribution of $B(t + s)$, given $B(t) = x$, is the same as the distribution of $B(s) + x$. This independence and distribution imply the Markov property.

1. Scaling property of Brownian Motion

Theorem: Brownian motion $B(t)$ satisfies the scaling property, i.e., for any a>0, the process $\left\{aB\left(t/a^2\right)\right\}_{t\geq 0}$ is also a Brownian motion.

Proof: Let $Y(t) = aB\left(t/a^2\right)$. The process $Y(t)$ has increments $Y(t+s) - Y(t) = aB\left((t+s)/a^2\right) - aB\left(t/a^2\right)$.

The increment $B\left((t+s)/a^2\right) - B\left(t/a^2\right) \sim N(0, s/a^2)$. Scaled by a, gives $aN\left(0, \frac{s}{a^2}\right) = N(0,s)$. Thus, $Y(t+s) - Y(t) \sim N(0,s)$ implying that $Y(t)$ has the same distribution as $B(t)$. Therefore, the process $\left\{aB\left(t/a^2\right)\right\}_{t\geq 0}$ is a Brownian motion.

2. Distribution of Maximum of Brownian Motion

Theorem: For a Brownian motion $B(t)$, the maximum value up to *t* is distributed as follows:

$$P(max_{0\leq s\leq t} B(s) \geq a) = 2\ P\ (B(t) \geq a); for\ a > 0$$

Proof: Consider $M(t) = max_{0\leq s\leq t} B(s)$. By the reflection principle, if $B(t)$ hits a at some point and ends below a are equal. Therefore, $P(M(t) \geq a) = 2P(B(t) \geq a)$.

Simulating Brownian motion involves generating a sequence of random steps. For a one-dimensional Brownian motion, we approximate B(t) by summing small, normally distributed increments over small time intervals. This method can be extended to higher dimensions by generating independent random steps for each coordinate.

Brownian motion provides a robust framework for modeling random phenomena in continuous time and space. Understanding its properties and proofs allow for a deeper grasp of its applications in various scientific and engineering fields.

Examples of Brownian Motion as a Random Walk

1. Brownian Motion in Physics: Particle Suspended in a Fluid

Example: Consider a pollen particle suspended in water. The particle undergoes Brownian motion due to collisions with water molecules. This random movement can be modeled as a random walk in continuous time.

Model: Suppose the particle starts at the origin.

Each collision with a water molecule causes the particle to take a small random step in a random direction. Over time, these small steps accumulate to form a continuous path.

Simulation: To simulate this, one can generate a sequence of small, normally distributed random steps. For simplicity, assume steps are taken every small-time interval Δt. Let $X_i \sim N(0, \sigma^2 \Delta t)$ be the step in the x-direction and $Y_i \sim N(0, \sigma^2 \Delta t)$ be the step in the y-direction.

The position of the particle after n step is:

$X_n = \sum_{i=1}^{n} X_i$,

$Y_n = \sum_{i=1}^{n} Y_i$.

2. Brownian Motion in Finance: Stock Prices

Example: The price of a stock is often modeled using a geometric Brownian motion, which incorporates the concept of random walk.

Model: Let $S(t)$ be the price of the stock at time t. The stock price follows the stochastic differential equation:

$$dS(t) = \mu S(t) + \sigma S(t) dB(t)$$

Here μ is the drift term, σ is the volatility and $B(t)$ is a standard Brownian motion.

To simulate this, discretize time into small intervals Δt. The price change over each interval can be approximated as:

$$\Delta S = \mu S \Delta t + \sigma S \Delta B$$

where $\Delta B \sim N(0, \Delta t)$.

3. Brownian Motion in Biology: Movement of Microorganisms

Example: The movement of microorganisms, such as bacteria, in a fluid can be modeled as a random walk.

Model: Assume a microorganism moves in small random steps due to interactions with the fluid. Each step is influenced by random collisions with molecules in the fluid.

Simulation: The position of the microorganism at time t is modeled as:

$X(t) = X(0) + \sum_{i=1}^{n} X_i,$

$Y(t) = Y(0) + \sum_{i=1}^{n} Y_i.$

Where $X_i \; and \; Y_i$ are small, normally distributed steps.

4. Brownian Motion in Chemistry: Diffusion of Molecules

Example: The diffusion of molecules in a liquid or gas can be modeled using Brownian motion.

Model: Consider a molecule diffusing through a liquid.

The molecule undergoes random collisions with other molecules, causing it to move in a random walk pattern.

Simulation: The position of the molecule after n steps is given by:

$$X_n = \sum_{i=1}^{n} X_i, Y_n = \sum_{i=1}^{n} Y_i \; and \; Z_n = \sum_{i=1}^{n} Z_i$$

Where X_i, Y_i and Z_i are normally distributed random variables representing the steps in the x, y and z directions.

9.5 Duality in Random Walk

Duality in the context of random walks is a powerful concept that allows us to relate one stochastic process to another, often simplifying the analysis and understanding of complex systems. The dual process can provide insights into the behavior of the original process, especially when the dual process is easier to analyze or has well-known properties.

Key Concepts and Definitions

Dual Process: A dual process of a random walk is another process that can be used to describe the same system from a different perspective. The relationship between the original process and its dual is characterized by certain symmetries or transformations.

Duality Function: A function $D(x, y)$ that establishes a connection between the states of the original process and the dual process. Typically, if $X(t)$ is the original process and $Y(t)$ is the dual process, the duality function D satisfies $E[D(X(t), y)] = E[D(x, Y(t))]$.

Important Results

1. Duality in Symmetric Random Walks

For a symmetric random walk on the integer lattice Z, the dual process is also a symmetric random walk.

Proof: Let $X(t)$ be a symmetric random walk on Z with transition probabilities

$$P(X(t+1) = x+1 \mid X(t) = x) = P(X(t+1) = x-1 \mid X(t) = x) = 1/2$$

Define the dual process $Y(t)$, as a symmetric random walk on Z with the same transition probabilities.

Duality Function: Consider the duality function $D(x, y) = \delta_{x,y}$ is Kronecker delta function. We need to show $E\left[\delta_{X(t),y}\right] = E[\delta_{x,Y(t)}]$.

For $X(t)$ starting at x and $Y(t)$ starting at y, the probability that $X(t) = y$ and $Y(t) = x$ are the same due to the symmetry of the transition probabilities.

Therefore, $E\left[\delta_{X(t),y}\right] = P(X(t) = y)$ and $E\left[\delta_{x,Y(t)}\right] = P(Y(t) = x)$.

Since both the probabilities are equal for all x and y, the symmetric random walk is self-dual.

2. Duality in Birth-Death Processes

Consider a birth-death process with birth rate λ and death rate μ. The dual process is another birth-death process with swapped rates.

Proof: Original Process: Let X(t) be a birth-death process with birth rate λ and death rate μ.

Dual Process: Define the dual process $Y(t)$ with birth rate μ and death rate λ.

Duality Function: Consider the duality function $D(x, y) = x^y$

For the original process $X(t)$:

$$E[D(X(t+dt), y)|X(t) = x]$$
$$= E[(x+1)^y].\lambda dt + E[(x-1)^y].\mu dt + E[x^y].(1-(\lambda+\mu)dt)$$

For the dual process $Y(t)$:

$$E\big[D\big(x, Y(t+dt)\big)\big|Y(t) = y\big]$$
$$= E[x^{y+1}].\mu dt + E[x^{y-1}].\lambda dt + E[x^y].(1-(\lambda+\mu)dt)$$

Using the binomial theorem and the linearity of expectation, we can show that both expectations are equal, confirming the duality.

Duality in random walks provide a valuable tool for analyzing complex stochastic processes by relating them to simpler or better-understood processes.

Important results, such as the self-duality of symmetric random walks, the duality in birth-death processes, and the duality in the voter model, demonstrate the utility of this concept. Understanding and proving these dualities allows for deeper insights into the behavior and properties of random walks and related stochastic processes.

9.6 Gambler's Ruin Problem

The gambler's ruin problem is a classic problem in probability theory that illustrates the potential outcomes of a gambler who bets repeatedly in a fair game. It examines the likelihood that a gambler, starting with a finite amount of money, will either go broke or reach a target amount. This problem is fundamental in the study of stochastic processes and has applications in finance, economics, and risk management.

The gambler's ruin problem dates back to the 17th century and is often attributed to Blaise Pascal and Pierre de Fermat, who corresponded about probability problems related to gambling. Their work laid the groundwork for modern probability theory. Later, in the 18th century, the problem was further developed by mathematicians such as Abraham de Moivre and Christiaan Huygens, who provided more formal analyses and solutions. The term "gambler's ruin" itself was popularized in the 19th century and has since been a staple example in probability textbooks. The problem is closely related to the concept of random walks, as it can be modeled as a one-dimensional random walk with absorbing barriers.

Consider a gambler I who has an initial capital of k rupees and plays against an opponent, gambler II, whose initial capital is Rs $a-k$. They are playing a game which proceeds by stages. At each step the probability that gambler I wins Re 1 from his opponent is p and the probability that he losses Re 1 to his opponent is $q\ (=1-p)$. The game continuous until the capital of one of the players reduced to zero (*i.e.,* the capital of player I either reduced to zero or increased to "a"). The capital possessed by, say, the player I, performs a random walk on non-negative integers $\{0,1,2,\dots,a\}$ with absorbing barriers at 0 and a. The absorptions being interpreted as the ruin of the one, or the other player. Given the initial capital k, it is of player I, it is either $k-1$ or $k+1$ according as whether player I losses or wins the first game. Let μ_k be the probability that the gambler I, starting with the initial capital k ultimately ruins. Then

$$\mu_k = p\,\mu_{k+1} + q\mu_{k-1};\ \ k = 2,3,\dots,a-2 \qquad (1)$$

$$\mu_1 = q + p\,\mu_2 \qquad (2)$$

$$\mu_{a-1} = q\,\mu_{a-2}\ \ (\mu_a = 0) \qquad (3)$$

We can write equations (1), (2) and (3) jointly as

$$\mu_0 = 1, \mu_a = 0 \text{ (boundary conditions)}$$

$$\mu_k = p\,\mu_{k+1} + q\,\mu_{k-1}; 1 \le k \le a-1 \qquad (4)$$

Now we solve (4) under the boundary conditions.

Case I: Let $p \neq q$ (random walk is asymmetric)

Let $\mu_k = \lambda^k$ be a particular solution of (4). Then auxiliary equations are

$$p\,\lambda^2 - \lambda + q = 0 \qquad (5)$$

or

$$(\lambda - 1)(p\,\lambda - q) = 0 \qquad (6)$$

Equation (6) leads to the roots $\lambda = 1, \lambda = \frac{q}{p}$. Hence, two particular-solutions for μ_k are

$$\mu_k = 1^k = 1, \qquad \mu_k = \left(\frac{q}{p}\right)^k.$$

Then a general solution is

$$\mu_k = A + B\left(\frac{q}{p}\right)^k \qquad (7)$$

Utilizing the boundary conditions $\mu_0 = 1, \mu_a = 0$ in (7), we have

$$\left.\begin{aligned} 1 &= A + B \\ 0 &= A + B\left(\frac{q}{p}\right)^a \end{aligned}\right\}$$

$$\Rightarrow B = -\frac{1}{\left(\frac{q}{p}\right)^a - 1}$$

$$A = \frac{\left(\frac{q}{p}\right)^a}{\left(\frac{q}{p}\right)^a - 1}.$$

Substituting the values of A and B in (7) leads to

$$\mu_k = \frac{\left(\frac{q}{p}\right)^a - \left(\frac{q}{p}\right)^k}{\left(\frac{q}{p}\right)^a - 1}. \qquad (8)$$

Similarly, we can obtain the following expression for the probability of ruin of player II:

$$\nu_k = \frac{\left(\frac{q}{p}\right)^k - 1}{\left(\frac{q}{p}\right)^a - 1} \qquad (9)$$

We can easily obtain ν_k by replacing q by p, p by q and k by $a - k$ in (8).

Since $\mu_k + \nu_k = 1$, the probability of an unending game is 0, *i.e.*,

$$P(unending\ game) = 0$$

Case II: Let $p = q = \frac{1}{2}$, then (5) reduces to

$$\lambda^2 - 2\lambda + 1 = 0, \qquad (10)$$

which has two equal roots $\lambda = 1$. Further when $p = q = 1/2$, if we substitute $\mu_k = k$ in (4), we obtain

$$k = \frac{1}{2}(k+1) + \frac{1}{2}(k-1)$$

Hence $\mu_k = k$ is a second solution of (4). Hence a general solution is

$$\mu_k = C + Dk.$$

Using boundary conditions, we have

$$\left.\begin{array}{c} For\ k = 0, \mu_0 = 1 = C \\ For\ k = a, \mu_a = 0 = C + Da \end{array}\right\}$$

Hence

$$C = 1, D = -\frac{1}{a}$$

This leads to

$$\mu_k = 1 - \frac{k}{a}$$

Similarly, we obtain

$$v_k = \frac{k}{a}$$

Again P (unending game) = 0.

Suppose player II has infinite capital, *i.e.,* $a \to \infty$. An example of player II with infinite capital is Casino. Then, for $p > q$, $\lim_{a\to\infty} \left(\frac{q}{p}\right)^a = 0$ and the probability that player I with initial capital μ_k ultimately ruins, is

$$\mu_k = \left(\frac{q}{p}\right)^k$$

The probability of an unending game is

$$1 - \left(\frac{q}{p}\right)^k.$$

If $p < q$, $\lim_{a\to\infty} \left(\frac{p}{q}\right)^a = 0$ and $\mu_k = 1$.

Further for $p = q$, as $a \to \infty$. $\mu_k \to 1$.

Hence for $p \leq q$, the probability of an unending game is 0 and the probability of ultimate ruin of player I is 1.

Examples and Applications of the Gambler's Ruin Problem

Example 1: Simple Betting Game

A gambler starts with 10 units of currency and bets 1 unit on a fair game (50% chance of winning, 50% chance of losing) until either going broke or reaching 20 units. Determine the probability of the gambler going broke.

Solution: For a fair game, the probability of ruin P_i when starting with i units and aiming to reach N units is given by:

$$P_i = 1 - \frac{i}{N}$$

Here $i = 10$ and $N = 20$

$$P_{10} = 1 - \left(\frac{10}{20}\right) = 0.5$$

So, the probability of the gambler going broke is 0.5 or 50%.

Example 2: Biased Betting Game

A gambler starts with 5 units and bets 1 unit on a game where they have a 40% chance of winning and a 60% chance of losing, until they either go broke or reach 10 units. Determine the probability of the gambler going broke.

Solution: For an unfair game, the probability of ruin P_i is given by

$$P_i = \frac{1 - \left(\frac{q}{p}\right)^i}{1 - \left(\frac{q}{p}\right)^N}$$

Here $p = 0.4, q = 0.6, i = 5\ and\ N = 10$. Substituting all the values and on simplification we get,

$$P_5 \approx 0.116$$

So, the probability of the gambler going broke is approximately 11.6%.

Applications

1. Finance and Investment: In the stock market, investors often face the risk of losing their initial investment. The gambler's ruin problem helps in understanding the likelihood of an investor losing all their capital versus reaching a certain profit target.

An investor starts with \$10,000 and aims to either double their investment or lose it all, making investments that either increase or decrease their capital by \$1,000 with equal probability. The gambler's ruin model can predict the chances of going bankrupt or doubling the investment.

2. Insurance: Insurance companies must manage their reserves to avoid insolvency. The gambler's ruin problem can model the risk of an insurance company depleting its reserves due to excessive claims.

An insurance company starts with a reserve of $1,000,000 and receives claims of $10,000 each. The probability of the company going bankrupt before replenishing its reserves can be calculated using the gambler's ruin framework.

3. Genetics: In population genetics, the frequency of alleles (gene variants) can change over generations due to random sampling. The gambler's ruin problem models the probability that an allele will become fixed (reach 100% frequency) or lost (reach 0% frequency) in a population.

A rare allele starts with a frequency of 0.1 in a population of 100 individuals. The model can predict the probability of the allele becoming fixed or lost over time.

4. Queueing Theory: In queueing systems, such as customer service lines or network packets, the gambler's ruin problem can model the likelihood of a system becoming overwhelmed or stabilizing.

A server processes requests with a probability of success or failure. The probability that the server queue will overflow (ruin) versus clearing all requests can be analyzed using the gambler's ruin model.

5. Gambling and Games of Chance: The original application of the gambler's ruin problem was in gambling, where it models the likelihood of a gambler going broke versus reaching a target profit.

A gambler with $50 bets $5 per game on a fair coin toss. The model can predict the probability of the gambler losing all their money before doubling it.

The gambler's ruin problem is a versatile tool in probability theory, offering insights into the dynamics of stochastic processes where there is a risk of "ruin" or bankruptcy. Through various examples and applications in finance, insurance, genetics, queueing theory, and gambling, we see how this problem helps in understanding and managing risk in different domains. The mathematical formulations provide a structured approach to calculating the probabilities of different outcomes, making it a valuable model for analyzing scenarios involving repeated risks and rewards.

9.7 Summary

This unit provided a comprehensive exploration of random walks and their various applications. We began with an introduction to the concept of random walks, tracing its historical development from early studies by Karl Pearson, Albert Einstein, and Norbert Wiener. These foundational works laid the groundwork for understanding Brownian motion, which describes the random movement of particles suspended in a fluid. We then extended this concept to one-dimensional, two-dimensional, and three-dimensional random walks, emphasizing the differences and applications of each.

In addition to Brownian motion, we explored the notion of duality in random walks, which offers a powerful method for simplifying complex stochastic processes by relating them to more straightforward or better-understood processes. This section included important results such as the duality of symmetric random walks and birth-death processes, highlighting how duality functions can provide significant insights.

The gambler's ruin problem was another key topic covered in this unit. We examined the problem's historical context, mathematical formulation, and solutions for both fair and unfair games. This problem illustrates the risk of a gambler going broke or reaching a target fortune and has important applications in finance, genetics, and risk management.

Throughout the unit, we emphasized the relevance and utility of these concepts in various fields. Brownian motion, for instance, is crucial in physics for modelling diffusion processes, in finance for stock price modelling, and in biology for the movement of microorganisms. The gambler's ruin problem, with its insights into risk and probability, finds applications in financial modelling and genetic drift.

By understanding these fundamental concepts and their applications, students gain valuable insights into stochastic processes and their implications in different scientific and practical domains. This knowledge equips them with the tools to analyse and model random phenomena, providing a solid foundation for further study and research in probability theory and related fields.

9.8 Self-Assessment Questions

1) Define a random walk and provide an example.
2) Explain how Brownian motion can be modeled as a random walk.

3) What are the key differences between one-dimensional, two-dimensional, and three-dimensional random walks?

4) Describe the concept of duality in random walks and provide an example.

5) Solve the gambler's ruin problem for a gambler starting with 5 units and aiming to reach 10 units, with a win probability of 0.4 on each bet.

6) What is the significance of the Markov property in Brownian motion?

7) How does the self-similarity property of Brownian motion simplify its analysis?

8) Describe the applications of the gambler's ruin problem in finance and genetics.

9.9 Reference

- Feller, W. (1968). An Introduction to Probability Theory and Its Applications. Wiley.
- Karatzas, I., & Shreve, S. E. (1991). Brownian Motion and Stochastic Calculus. Springer.
- Ross, S. M. (2014). Introduction to Probability Models. Academic Press.
- Grimmett, G., & Stirzaker, D. (2001). Probability and Random Processes. Oxford University Press.
- Redner, S. (2001). A Guide to First-Passage Processes. Cambridge University Press.

9.10 Further Reading

- Spitzer, F. (2001). Principles of Random Walk. Springer.
- Kloeden, P. E., & Platen, E. (1992). Numerical Solution of Stochastic Differential Equations. Springer.
- Øksendal, B. (2003). Stochastic Differential Equations: An Introduction with Applications. Springer.
- Hammersley, J. M., & Handscomb, D. C. (1964). Monte Carlo Methods. Methuen.
- Pearson, K. (1905). The Problem of the Random Walk. Nature, 72, 294.

UNIT - 10: QUEUING PROCESS

Structure

10.1 Introduction

Queuing theory, the mathematical study of waiting lines or queues, has its origins in the early 20th century. The foundations of this theory were laid by Danish mathematician Agner Krarup Erlang, who, while working for the Copenhagen Telephone Company, developed the first analytical models to describe telephone traffic. Erlang's work, beginning in 1909, focused on understanding the probability of call congestion and determining the number of telephone lines required to handle call volumes without excessive delays. His seminal papers, including "The Theory of Probabilities and Telephone Conversations" (1909) and "Solution of Some Problems in the Theory of Probabilities of Significance in Automatic Telephone Exchanges" (1917), are considered landmark contributions to queuing theory.

The field expanded significantly during and after World War II, driven by the needs of military logistics and operations research. David George Kendall's work in the 1950s further advanced the theory by introducing Kendall's notation, which succinctly describes the characteristics of different queuing systems. This period also saw the development of key models such as the M/M/1 queue, which considers a single server with exponential inter-arrival and service times.

In the 1960s, Leonard Kleinrock's work on the application of queuing theory to computer networks significantly impacted the field, particularly in the development of ARPANET, the precursor to the modern Internet. Kleinrock's contributions helped establishing queuing theory as a fundamental tool in the design and analysis of telecommunications and computer networks.

In recent decades, the advancement of technology and the increasing complexity of systems have driven significant developments in queuing theory. Modern queuing theory incorporates more sophisticated and versatile models to address contemporary challenges in telecommunications, computing, and service systems.

One major area of advancement is the application of queuing theory to network queues and Internet traffic. With the rise of the Internet, queuing theory has become crucial in managing data traffic, ensuring quality of service (QoS), and optimizing network performance. Researchers have developed models to understand and mitigate congestion in packet-switched networks, allowing for more efficient and reliable data transmission.

Healthcare systems have also benefited from advancements in queuing theory. Queuing models are now extensively used to improve patient flow, reduce waiting times, and optimize resource allocation in hospitals and clinics. These models help healthcare providers manage patient admissions, scheduling, and treatment processes more effectively, leading to better patient outcomes and more efficient use of medical resources.

In manufacturing and supply chain management, queuing theory helps design efficient production lines, reduce bottlenecks, and improve inventory management. Queuing models aid in optimizing logistics and distribution networks, ensuring that goods are produced and delivered in a timely and cost-effective manner.

The proliferation of cloud computing services has led to the application of queuing models to manage virtualized resources, optimize load balancing, and ensure service level agreements (SLAs) are met. These models help cloud service providers allocate resources dynamically based on demand, ensuring that users receive consistent and reliable service.

Recent advances in queuing theory also include the use of advanced mathematical and computational techniques such as stochastic processes, simulation, and machine learning. These techniques allow for more accurate and predictive queuing models capable of handling complex and dynamic systems. Methods like Markov chains, Monte Carlo simulations, and reinforcement

learning are increasingly integrated into queuing theory, providing more sophisticated tools for analyzing and optimizing queuing systems.

The advent of big data analytics has enabled the collection and analysis of vast amounts of data on queuing systems. This data-driven approach allows for more precise modelling, real-time monitoring, and adaptive management of queues. By leveraging big data, researchers and practitioners can develop more accurate forecasts and tailor queuing models to specific scenarios, improving system performance and customer satisfaction.

New models that consider multiple servers and different classes of customers have been developed to better reflect real-world scenarios. These models, such as multi-server and multi-class queues, help in understanding and optimizing systems with diverse service requirements and customer profiles. They provide insights into the behavior of queues under various conditions, allowing for more effective resource allocation and service management.

Queuing theory's applications are vast and diverse, encompassing various industries and domains. In telecommunications, it helps to manage call centers, network traffic, and bandwidth allocation. In transportation, queuing models optimize traffic flow, public transport schedules, and airport operations. In retail, they improve checkout processes, customer service, and inventory management. In computing, queuing theory enhances the performance of computer systems, cloud services, and data centers.

The future of queuing theory is promising, with ongoing research focusing on addressing the challenges posed by emerging technologies and complex systems. Potential directions include the integration of artificial intelligence (AI) and machine learning to develop intelligent queuing systems that can learn and adapt to changing conditions in real-time. Sustainable and green queuing models are being explored to consider energy efficiency and sustainability in system design and operation. Additionally, incorporating human factors and behavioral aspects into queuing models aims to better understand and improve user experiences.

A queue is formed when units (or customers, clients) needing some kind of service arrive at a service channel (or counter) which provides such service. Each customer on arrival goes directly into service if the server is free and if not, joins the queue and leaves the system after being served. The basic features characterizing a system are:

(i) The inputs,

(ii) The service mechanism

(iii) The queue discipline and

(iv) The number of service channels.

The input describes the manner in which customers arrive and join the system. The system may have either a limited or an unlimited capacity of holding units. The source from which the customer come may be finite or infinite. The customers may arrive either singly or in group. The interval between two consecutive arrivals is called the *interarrival time.*

The service mechanism describes the way the customers are being served. The customers may be served either singly or in batches. The time required for serving a unit is called the *service time*.

The queue discipline indicates the way customers form a queue and are served. If the customer at the counter leaves the counter after being served and the next customer at the head of the queue enters the service system, the discipline is called the "First come First Service" (FCFS) or "First in First out (FIFO) queue discipline. Some other rules may be adopted, such as last come first served or random ordering before service.

The system may have one channel or s-parallel channels for service. The interarrival and service times may be deterministic or random. Usually, we are concerned with random interarrival and service time.

The following random variables or families of random variables provide important measures of performance of stochastic queueing system:

(i) The number of customers waiting in the queue including the one being served at time t, say $N(t)$.

(ii) The busy period which means the duration of the interval from the moment the service starts with arrival of a customer at any empty counter to the moment the server becomes free for the first time.

(iii) The waiting time W_n for the n^{th} arrival.

(iv) The waiting time $W(t)$ of a customer in the queue which arrived at the instant t.

$\{N(t); t \geq 0\}$ and $\{W(t); t \geq 0\}$ are stochastic processes with continuous time

$\{W_n\ ; n = 0,1,2, \ldots\}$ is a stochastic process with discrete time.

Notation: A queueing system is denoted by a three-part description A/B/C, where the first two symbols denote the interarrival and service time distributions respectively, and the third symbol denotes the number of channels or servers.

Queuing theory has evolved significantly since its inception, driven by the need to address increasingly complex and dynamic systems. Its applications continue to expand, offering valuable insights and solutions across various fields. As technology advances, queuing theory will remain a crucial tool in optimizing system performance and improving efficiency in both traditional and emerging domains.

10.2 Objectives

By the end of this unit, you should be able to:

Understand Fundamental Concepts of Queuing Theory: Grasp the basic principles and terminology of queuing theory, including arrival rates, service rates, and queue disciplines. Comprehend the significance of queuing theory in analyzing and optimizing systems where waiting lines are a critical component.

Differentiate Between Various Types of Queuing Models: Identify and distinguish between different types of queuing models such as M/M/1, M/M/k, M/G/1, Ek/M/1, and M/Ek/1. Understand the specific assumptions and characteristics of each model and how they apply to real-world situations.

Apply Birth and Death Processes in Queuing Theory: Explain the birth and death processes and their relevance to queuing systems. Use birth and death processes to model the arrival and departure of customers in queuing systems, and calculate key performance measures.

Analyze and Solve Problems Involving Renewal Processes: Understand the concept of renewal processes and their application in queuing theory. Solve problems involving renewal intervals and renewal functions to determine system performance and reliability.

Evaluate the Effectiveness of Various Queuing Models: Assess the suitability and effectiveness of different queuing models for specific applications. Compare and contrast the performance of various queuing models under different conditions and constraints.

Calculate Performance Measures for Different Queuing Systems: Determine key performance measures such as utilization (ρ), average number in the system (L), average time in the system (W), average number in the queue (L_q), and average time in the queue (W_q) for different queuing models. Apply these measures to analyse the efficiency and effectiveness of queuing systems and make informed decisions for improvement.

Use Queuing Theory to Optimize System Performance: Apply queuing theory principles to optimize the design and operation of systems involving waiting lines. Develop strategies to reduce wait times, improve service efficiency, and enhance overall system performance.

Integrate Advanced Mathematical Techniques in Queuing Analysis: Utilize advanced mathematical techniques such as stochastic processes, Markov chains, and simulation to model and analyze complex queuing systems. Leverage these techniques to develop more accurate and predictive models for real-world applications.

Incorporate Real-World Applications and Case Studies: Examine real-world applications of queuing theory in various fields such as telecommunications, healthcare, manufacturing, transportation, retail, and computing. Analyze case studies to understand how queuing theory has been applied to solve practical problems and improve system performance.

Develop Critical Thinking and Problem-Solving Skills: Enhance your ability to think critically and solve complex problems related to queuing systems. Apply theoretical knowledge to practical scenarios, making informed decisions based on analytical insights.

Prepare for Advanced Study and Research in Queuing Theory: Build a strong foundation for advanced study and research in queuing theory and related fields. Develop the skills and knowledge necessary to pursue further academic or professional opportunities in operations research, systems engineering, and applied mathematics.

10.3 Birth and Death Processes

The birth and death processes are fundamental concepts in queuing theory, representing a special type of Markov process where transitions can only occur between adjacent states. This model is widely used to describe systems where entities (customers, jobs, etc.) arrive, wait for service, receive service, and then depart.

Definition: The birth and death processes are type of Markov process where changes in state occur through "births" (arrivals) and "deaths" (departures). This process is characterized by its simplicity, where transitions can only occur between adjacent states, typically in the context of queuing systems. These models are fundamental in understanding and analyzing systems where entities arrive, wait for service, receive service, and then leave.

Birth Rate (λ): The birth rate, denoted by λ, is the rate at which new entities (e.g., customers) arrive at the system. It represents the frequency of arrivals per unit time. In a queuing system, the birth rate is crucial for determining how quickly the queue forms.

Death Rate (μ): The death rate, denoted by μ, is the rate at which entities are served and leave the system. It signifies the frequency of departures per unit time. The death rate depends on the service mechanism and the efficiency of the server.

State Transitions: In a birth and death process, the system transitions between states based on the arrival and departure of entities. The state of the system is usually represented by the number of entities in the system (n). The transitions occur as follows:

- Birth Transition (n to n+1): When a new entity arrives, the system moves from state n to state n+1, denoted by a rate λn.
- Death Transition (n to n-1): When an entity is served and leaves, the system moves from state n to state n-1, denoted by a rate μn.

Poisson Process: Birth and death processes often assume arrivals and services follow a Poisson process, meaning the inter-arrival and service times are exponentially distributed. This assumption simplifies the mathematical analysis and is a good approximation for many real-world systems.

Mathematical Representation

The birth and death process can be described using differential equations that model the probability of being in a particular state at a given time. Let $P_n(t)$ be the probability that the system is in state n at time t. The change in $P_n(t)$ over a small-time interval Δt can be expressed as:

$$P_n(t + \Delta t) - P_n(t) = \lambda_{n-1}P_{n-1}(t)\Delta t + \mu_{n+1}P_{n+1}(t)\Delta t - (\lambda_n + \mu_n)P_n(t)\Delta t$$

Dividing through by Δt and taking the limit as Δt approaches zero gives the differential equation:

$$\frac{dP_n(t)}{dt} = \lambda_{n-1}P_{n-1}(t) + \mu_{n+1}P_{n+1}(t) - (\lambda_n + \mu_n)P_n(t)$$

For steady state conditions (where the probabilities do not change over time), we put $\frac{dP_n(t)}{dt} = 0$, leading to:

$$\lambda_{n-1}P_{n-1} + \mu_{n+1}P_{n+1} = (\lambda_n + \mu_n)P_n$$

Stationary Distribution

To find the stationary distribution, we need to solve the balance equations. These equations describe the flow of probability into and out of each state at equilibrium. For a simple M/M/1 queue (where arrivals and departures follow exponential distributions), the stationary probabilities can be derived as follows:

1. **Balance Equations:**

$$\lambda P_0 = \mu P_1$$

$$\lambda P_1 = \mu P_2$$

$$\lambda P_2 = \mu P_3$$

$$\vdots$$

2. **Recurrence Relation:**

Using the balance equations, we can express P_n in terms of P_0:

$$P_n = \left(\frac{\lambda}{\mu}\right)^n P_0$$

3. **Normalization Condition:** The sum of all probabilities must equal 1.

$$\sum_{n=0}^{\infty} P_n = 1$$

$$\text{or } P_0 \sum_{n=0}^{\infty} \left(\frac{\lambda}{\mu}\right)^n = 1$$

$$\text{or } P_0 . \frac{1}{1 - \frac{\lambda}{\mu}} = 1$$

Thus

$$P_0 = 1 - \frac{\lambda}{\mu}$$

4. **Stationary Probabilities:**

   The stationary probability distribution is:

$$P_n = \left(1 - \frac{\lambda}{\mu}\right) . \left(\frac{\lambda}{\mu}\right)^n$$

Performance Measures

Average Number in the System (L): The average number of entities in the system can be found using Little's Law, which states

$$L = \lambda W,$$

where W is the average time, an entity spends in the system.

For an M/M/1 queue:

$$L = \sum_{n=0}^{\infty} nP_n = \frac{\rho}{1-\rho}$$

where $\rho = \lambda/\mu$ is the utilization factor.

Average Time in the system (W):

The average time an entity spends in the system is:

$$W = \frac{1}{\mu - \lambda}$$

Average Number in the Queue (L_q):

The average number of entities in the queue can be found as:

$$\boldsymbol{L_q = L - \rho = \frac{\rho^2}{1-\rho}}$$

Average Time in the Queue (W_q):

The average time an entity spends waiting in the queue is:

$$\boldsymbol{W_q = W - \frac{1}{\mu} = \frac{\lambda}{\mu(\mu - \lambda)}}$$

Applications: Birth and death processes have broad applications in various fields:

Telecommunications:

- Modelling call arrivals and departures in telephone networks.
- Managing data packet flows in computer networks.

Healthcare:

- Analyzing patient arrival and service times in hospitals.
- Optimizing staffing levels in emergency departments.

Manufacturing:

- Managing job arrivals and processing times on production lines.
- Balancing workloads across multiple machines.

Service Industries:

- Reducing wait times in customer service centers.
- Improving efficiency in retail checkout lines.

Transportation:

- Modelling vehicle arrivals and departures at toll booths.
- Optimizing scheduling and dispatching of public transport.

By understanding the fundamental principles and mathematical representations of birth and death process, we can derive key performance measures and apply these insights to optimize real-world systems.

10.4 Renewal Process

The concept of renewal processes was developed to generalize the Poisson process, which was limited to modelling events occurring at constant rates (i.e., exponentially distributed inter-arrival times). Early work in this area was driven by the need to model and analyse systems in which events occur at irregular intervals.

In the 1940s and 1950s, mathematician William Feller made significant contributions to the theory of stochastic processes, including renewal theory. Feller's work helped establish the foundational principles of renewal processes, including the renewal equation and the renewal reward theorem. His contributions were crucial in providing a formal mathematical framework for understanding renewal processes.

In the 1960s, statistician David Cox further advanced the field of renewal theory, particularly in the context of reliability engineering and life testing. Cox's book "Renewal Theory" (1962) is considered a seminal text in the field, offering a comprehensive treatment of the subject.

His work extended the application of renewal theory to practical problems in various domains, making the theory more accessible and useful for practitioners.

Definition:

A renewal process is a type of stochastic process that generalizes the Poisson process by allowing the inter-arrival times between events to follow any probability distribution, not just the exponential distribution. This process models the times at which events, such as arrivals or services, occur. Each event "renews" the process, resetting the system's state. Renewal processes are particularly useful for modelling and analyzing systems where events happen at intervals that are not necessarily memoryless.

In a renewal process, the time between successive events, known as inter-arrival times, are independent and identically distributed random variables. The renewal function, denoted as $M(t)$, represents the expected number of renewals (events) that occur by time t. Mathematically, this is expressed as

$$M(t) = E[N(t)],$$

where $N(t)$ is the number of renewals by time t. The renewal equation, which is integral to renewal theory, relates the renewal function to the distribution of inter-arrival times. For a renewal process with inter-arrival time distribution F, the renewal function $M(t)$ satisfies the integral equation:

$$M(t) = F(t) + \int_0^t M(t-u)f(u)du$$

where $f(u)$ is the probability density function of the inter-arrival times.

Applications of Renewal Processes:

Renewal processes have broad applications across many fields. In reliability engineering, they are used to model the failure and repair of systems. For example, the times between failures of a machine and the subsequent repairs can be modelled using a renewal process. In queuing theory, renewal processes model arrival and service processes in queuing systems where inter-arrival or service times are not exponentially distributed. This is useful for analyzing systems with more complex service requirements and arrival patterns.

In inventory control, renewal processes are used to model the times between replenishments of stock and the demand for products. This helps in optimizing inventory levels and ensuring timely replenishments. In economics and finance, renewal processes model events such as market transactions, arrivals of orders, and occurrences of financial risks. This aids in understanding and managing economic activities and financial uncertainties.

The renewal function $M(t)$ is central to the analysis of renewal processes. It can be computed for specific distributions of inter-arrival times. For example, if the inter-arrival times are exponentially distributed with rate λ, the renewal function is $M(t) = \lambda t$, consistent with the Poisson process.

The renewal process is a versatile and powerful tool in the study of stochastic processes, extending the concepts of the Poisson process to more general inter-arrival time distributions. Its development was motivated by the need to model systems with irregular event timings, leading to broad applications across engineering, economics, and operational research. Key contributions from mathematicians like William Feller and David Cox have solidified its theoretical foundation, providing valuable insights and methods for analyzing and optimizing various real-world systems. The renewal process continues to be an essential component in the toolkit of applied probability and stochastic modelling.

Key Theorems of Renewal Processes

Theorem: Elementary Renewal Theorem

Let $\{X_i\}$ be a sequence of independent identically distributed (*i.i.d.*) non negative random variables with a finite mean $E[X_i] = \mu$. let $N(t)$ denote the number of renewals (or events) that have occurred by time t. The *Elementary Renewal Theorem* states:

$$\lim_{t\to\infty} \frac{N(t)}{t} = \frac{1}{\mu}, \text{ almost surely (a.s.) and in expectation.}$$

Proof: The theorem can be proved using the strong law of large numbers (SLLN).

Let $S_n = X_1 + X_2 + \cdots + X_n$ denote the time of the n^{th} renewal. We need to show that:

$$\lim_{t\to\infty} \frac{N(t)}{t} = \frac{1}{\mu}$$

Define $N(t)$ as the largest integer n such that $S_n \leq t$. Mathematically

$$N(t) = \max\{n: S_n \leq t\}$$

According to the SLLN for *i.i.d.* random variables X_i with finite mean μ;

$\frac{S_n}{n} \to \mu$ a.s. as n tends to infinity.

This implies that

$$\lim_{n\to\infty} \frac{S_n}{n} = \mu \ (a.s.)$$

Taking reciprocal, we get

$$\lim_{n\to\infty} \frac{n}{S_n} = \frac{1}{\mu} \ (a.s.)$$

Since, $S_{N(t)} \leq t \leq S_{N(t)+1}$

we can write it as:

$$\frac{N(t)}{S_{N(t)+1}} \leq \frac{N(t)}{t} \leq \frac{N(t)}{S_{N(t)}}$$

As $t \to \infty, N(t) \to \infty$ because the process continues indefinitely. Thus, by the SLLN:

$$\lim_{t\to\infty} \frac{S_{N(t)}}{N(t)} = \mu \text{ and } \lim_{t\to\infty} \frac{S_{N(t)+1}}{N(t)} = \mu$$

Therefore,

$$\lim_{t\to\infty} \frac{N(t)}{S_{N(t)+1}} = \frac{1}{\mu} \text{ and } \lim_{t\to\infty} \frac{N(t)}{S_{N(t)}} = \frac{1}{\mu}$$

Combining these results, we get

$$\frac{1}{\mu} \leq \lim_{t\to\infty} \frac{N(t)}{t} \leq \frac{1}{\mu}$$

Hence

$$\lim_{t\to\infty} \frac{N(t)}{t} = \frac{1}{\mu}, a.s.$$

Theorem: Renewal Reward Theorem

Let $\{X_i\}$ be the inter renewal times and $\{R_i\}$ be the corresponding rewards, where $\{X_i\}$ and $\{R_i\}$ are *i.i.d.* sequences with $E[X_i] = \mu \; and \; E[R_i] = v$. Let $S_n = X_1 + X_2 + \cdots + X_n$ denote the time of the n^{th} renewal and $W(t) = R_1 + R_2 + \cdots + R_{N(t)}$ denote the total reward received by time t. the Renewal Reward Theorem states that

$$\lim_{t\to\infty} \frac{W(t)}{t} = \frac{v}{\mu}, \text{ a.s. and in expectation.}$$

Proof: Let $N(t)$ be the number of rewards by time t, and $W(t)$ is the total reward by time t. then $W(t)$ is given by:

$$W(t) = \sum_{i=1}^{N(t)} R_i$$

By SLLN, we have

$$\frac{W(t)}{N(t)} \to E[R] = v \; (a.s.)$$

From the Elementary Renewal Theorem, we know that

$$\frac{N(t)}{t} \to \frac{1}{E[X]} = \frac{1}{\mu} \; (a.s.)$$

We express $W(t)$ as

$$\frac{W(t)}{t} = \frac{W(t)}{N(t)} \frac{N(t)}{t}$$

Taking limits as t tends to infinity, we get

$$\log_{t\to\infty}\frac{W(t)}{t}=\log_{t\to\infty}\left(\frac{W(t)}{N(t)}\frac{N(t)}{t}\right)=v.\frac{1}{\mu}=\frac{v}{\mu}$$

Thus, the Renewal Reward Theorem is proven, establishing that the long-run average reward per unit time is the expected reward per renewal divided by the expected inter-renewal time.

These theorems provide fundamental insights into the behavior of renewal processes and are widely used in applications across various fields, including reliability engineering, queuing theory, and inventory control.

10.5 Queuing Models: Specification & Effectiveness

Queuing models are mathematical representations that help analyse and predict the behaviour and performance of queuing systems, which are ubiquitous in various domains such as telecommunications, transportation, healthcare, manufacturing, and service industries. These models are essential for understanding system performance, optimizing resource utilization, reducing waiting times, and improving service efficiency. By using queuing models, organizations can make informed decisions to enhance their operations and customer satisfaction.

Specification of Queuing Models: To specify a queuing model, several key elements must be defined. These elements describe the structure and behaviour of the queuing system. A widely used method for specifying queuing models is Kendall's notation, which is of the form A/B/c/K/N/D. This notation helps succinctly describe the different characteristics of a queuing system.

The arrival process (A) describes the statistical distribution of time between successive arrivals of entities (such as customers, jobs, or packets) into the system. Common arrival distributions include Markovian (denoted as M), which assumes exponential inter-arrival times, deterministic (D) for constant inter-arrival times, and general (G) for arbitrary inter-arrival time distributions. The Erlang-k distribution (Ek) represents k-stage exponential inter-arrival times.

The service process (B) characterizes the statistical distribution of service times. Like the arrival process, it can follow various distributions, including Markovian (M) for exponential

service times, deterministic (D) for constant service times, general (G) for arbitrary service time distributions, and Erlang-k (Ek) for k-stage exponential service times.

The number of servers (c) specifies the number of parallel servers providing service in the system. For example, a single-server system is denoted as c=1, while a system with multiple servers is denoted by the number of servers present, such as c=3.

The system capacity (K) indicates the maximum number of entities that can be in the system, including those in service and those waiting. If not explicitly stated, the system capacity is typically assumed to be infinite.

The population size (N) denotes the size of the population from which entities arrive. Like system capacity, if not explicitly stated, the population size is usually assumed to be infinite.

The queue discipline (D) describes the order in which entities are served. Common queue disciplines include "First In, First Out (FIFO)", "Last In, First Out (LIFO)", "Service In Random Order (SIRO)", and "Priority-based Schemes (PR)".

Common Queuing Models: One of the most basic and widely used queuing models is the M/M/1 model, which represents a single server with exponential inter-arrival and service times. This model is suitable for systems with a single service point, such as a single teller in a bank. The M/M/k model generalizes this to multiple servers, making it applicable to systems with several parallel service points, like multiple checkout counters in a supermarket.

Another important model is the M/G/1, which describes a single server with exponential inter-arrival times and a general service time distribution. This model is useful for systems where service times do not follow an exponential distribution. The Ek/M/1 model represents a single server with Erlang-k inter-arrival times and exponential service times, suitable for systems where arrivals occur in stages. Conversely, the M/Ek/1 model describes a single server with exponential inter-arrival times and Erlang-k service times, appropriate for systems where services are completed in stages.

Effectiveness of Queuing Models: The effectiveness of a queuing model is determined by its ability to accurately represent the real-world system it is modelling and provide useful insights for improving system performance. Several key performance measures are used to evaluate the effectiveness of queuing models. Utilization (ρ) is a crucial measure that indicates the fraction of

time the server is busy. It is calculated as $\rho = \lambda/(c\mu)$ for an M/M/c queue, where λ is the arrival rate and μ is the service rate.

The average number of entities in the system (L) is another important measure, representing the average number of entities both in the queue and being served. For an M/M/1 queue, this is calculated as $L = \lambda/(\mu - \lambda)$. The average time an entity spends in the system (W) is calculated similarly and is given by $W = 1/(\mu - \lambda)$ for an M/M/1 queue.

Other significant performance measures include the average number of entities waiting in the queue (Lq) and the average time an entity spends waiting in the queue (Wq). For an M/M/1 queue, $L_q = \frac{\lambda^2}{\mu(\mu-\lambda)}$ and $W_q = \frac{\lambda}{\mu(\mu-\lambda)}$. The probability of having n entities in the system (P_n) is also critical, calculated as $P_n = (1-\rho)\rho^n$ for an M/M/1 queue.

Applications of Queuing Models: Queuing models have a wide range of applications across different industries. In telecommunications, they are used to manage call centres and network traffic, ensuring efficient use of resources, and minimizing call drops. In healthcare, queuing models optimize patient flow in hospitals, reduce waiting times, and improve resource allocation. In manufacturing, they help design production lines to minimize bottlenecks and enhance throughput.

In transportation, queuing models are used to manage traffic flow, optimize public transport schedules, and reduce congestion. In retail, they enhance customer service by reducing checkout times and optimizing staffing levels.

Optimizing Queuing Systems: To optimize a queuing system, it is essential to balance the trade-offs between service quality and resource utilization. This can involve adjusting service rates by improving efficiency or adding more servers, managing arrival rates through demand management or scheduling, implementing priority schemes to serve critical customers faster, and redesigning the queue layout to enhance flow and reduce wait times.

By accurately specifying and effectively utilizing queuing models, organizations can significantly enhance their operational efficiency, improve customer satisfaction, and achieve better resource management. Understanding and applying these models allows for the optimization of various systems, ensuring they operate smoothly and meet the demands placed on them.

10.6 Measures and Specific Queuing Models

In queuing theory, various measures are used to evaluate the performance of queuing systems. These measures include utilization, average number of entities in the system, average time in the system, average number in the queue, and average time in the queue. Different queuing models are specified based on the distributions of inter-arrival and service times, as well as the number of servers. Here, we discuss the measures and characteristics of the Ek/M/1, M/Ek/1, M/M/1, M/M/k, and M/G/1 queuing processes in detail.

Common Performance Measures:

i. ***Utilization (ρ):*** Utilization represents the fraction of time the server(s) is busy. It is a crucial measure as it indicates the load on the system. High utilization (close to 1) implies the server is busy most of the time, which can lead to long queues and delays. Low utilization indicates underutilization of resources. For a system with c servers, the fraction of time the server is busy can be defined as $\rho = \frac{\lambda}{c\mu}$, where λ is the arrival rate and μ is the service rate.

ii. ***Average number in the system (L):*** This measure gives the expected number of entities present in the system (both in service and waiting) at any given time. It helps in understanding the overall load on the system. The average number of entities in the system, including those in service and those waiting. Formula: $L = \lambda W$, where W is the average time, an entity spends in the system.

iii. ***Average Time in the System (W):*** This measure represents the total time an entity spends in the system, from arrival to departure. It includes both waiting time and service time. It is a critical performance metric for evaluating the efficiency of the system. The average time an entity spends in the system, including both waiting time and service time. Formula: $W = W_q + \frac{1}{\mu}$, where W_q is the average time spent waiting in the queue.

iv. ***Average Number in the Queue (L_q):*** This measure indicates the expected number of entities waiting in the queue. It helps in understanding the congestion in the system. The average number of entities waiting in the queue can be defined as $L_q = \lambda W_q$.

v. ***Average Time in the Queue*** (W_q)***:*** This measure represents the expected waiting time for an entity before it starts receiving service. It is a key indicator of the service quality perceived by the entities. The average time an entity spends waiting in the queue can be given as: $W_q = W - \frac{1}{\mu}$.

By using these measures, one can evaluate and compare different queuing models to choose the most appropriate one for a given system. Proper understanding and application of these models and measures enable organizations to optimize their operations, improve resource utilization, and enhance customer satisfaction.

1. **Ek/M/1 Queue Process:** The Ek/M/1 queuing process is characterized by Erlang-k inter-arrival times and Markovian (exponential) service times with a single server. Erlang-k distribution is a special case of the gamma distribution, representing the sum of k exponential phases, making it suitable for systems where arrivals occur in stages. In this model-
   i. Erlang-k Inter-Arrival Times (Ek): Inter-arrival times follow an Erlang-k distribution, which is the sum of k exponential phases.
   ii. Markovian Service Times (M): Service times follow an exponential distribution.
   iii. Single Server (1): There is one server.

Measures:

i. Utilization $(\boldsymbol{\rho})$: $\rho = \frac{\lambda}{\mu}$.

ii. Average number in the system (L): $L = \frac{\rho+\rho^2}{2(1-\rho)}$.

iii. The Average time in the system (W): $W = \frac{L}{\lambda}$.

iv. Average number in the Queue (L_q): $L_q = L - \rho$.

v. Average Time in the Queue (W_q): $W_q = W - \frac{1}{\mu}$.

Applications: The Ek/M/1 queuing model, with Erlang-k inter-arrival times and exponential service times, is particularly useful in scenarios where arrivals happen in stages, making the arrival process more deterministic.

Examples:

i. Telecommunication Networks: In telecommunication systems, call arrivals can often be modelled using Erlang-k distributions because the call setup process involves several stages. The Ek/M/1 model helps in determining the optimal capacity needed to handle the call traffic without excessive delays.

ii. Manufacturing Systems: In manufacturing, components may arrive in batches after passing through several stages of processing. The Ek/M/1 model can optimize the flow of components through a single machine or workstation, ensuring that the system is neither underutilized nor overly congested.

2. **M/Ek/1 Queuing Process:** The M/Ek/1 queuing process features Markovian (exponential) inter-arrival times and Erlang-k service times with a single server. This model is suitable for systems where services are completed in stages. In this model,

i. Markovian Inter-Arrival Times (M): Inter-arrival times follow an exponential distribution.

ii. Erlang-k Service Times (Ek): Service times follow an Erlang-k distribution, which is the sum of k exponential phases.

iii. Single Server (1): There is one server.

Measures:

i. Utilization ($\boldsymbol{\rho}$): $\rho = \frac{\lambda}{\mu}$.

ii. Average number in the system (L): $L = \frac{\rho+\rho^2(2k-1)}{2(1-\rho)}$.

iii. The Average time in the system (W): $W = \frac{L}{\lambda}$.

iv. Average number in the Queue (L_q): $L_q = L - \rho$.

v. Average Time in the Queue (W_q): $W_q = W - \frac{1}{\mu}$.

Applications: The M/Ek/1 model, with exponential inter-arrival times and Erlang-k service times, is suitable for systems where services are completed in stages, providing a more detailed representation of service processes.

Examples:

i. Healthcare: In hospitals, certain diagnostic tests or treatment procedures may occur in multiple stages (e.g., initial consultation, testing, and follow-up). The M/Ek/1 model helps in scheduling these stages to minimize patient wait times and improve the utilization of healthcare resources.

ii. Banking Services: In banks, loan approval processes often involve several steps, including initial screening, detailed evaluation, and final approval. The M/Ek/1 model can help manage the workflow to ensure timely processing of loan applications.

3. **M/M/1 Queuing Process:** The M/M/1 queuing process is one of the simplest and most widely used models, characterized by Markovian (exponential) inter-arrival and service times with a single server. This model is applicable to systems with a single service point, such as a single teller in a bank. In this model,

i. Markovian Inter-Arrival Times (M): Inter-arrival times follow an exponential distribution.

ii. Markovian Service Times (M): Service times follow an exponential distribution.

iii. Single Server (1): There is one server.

Measures:

i. Utilization ($\boldsymbol{\rho}$): $\rho = \frac{\lambda}{\mu}$.

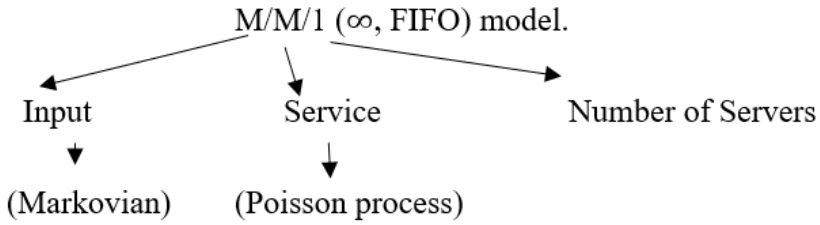
ii. Average number in the system (L): $L = \frac{\lambda}{\mu-\lambda}$.

iii. The Average time in the system (W): $W = \frac{1}{\mu-\lambda}$.

iv. Average number in the Queue (L_q): $L_q = \frac{\lambda^2}{\mu(\mu-\lambda)}$.

v. Average Time in the Queue (W_q): $W_q = \frac{\lambda}{\mu(\mu-\lambda)}$.

Suppose the customers arrive at a single server service system in according with a Poisson process having rate λ with FIFO discipline. Thus, the time between successive arrivals has exponential distribution with mean $1/\lambda$. The successive service times are assumed to be *iid* exponential random variables with mean $1/\mu$. The service does not stop as long as there are customers to be served. The population of customers and the systems capacity are assumed to be infinite. We also assume that the customer does not leave before getting the service and the arrivals and service are independent. This is the simple queueing model denoted as



Steady State Analysis of the M/M/1 (∞,FIFO)

Consider the M/M/1 queueing model with the assumptions stated before:

Let X_t be the number of customers in the queue including the one being served.

Let

$$P(X_t = n) = p_n(t).$$

$\{X_t; t \geq 0\}$is a stochastic process with continuous time parameter and dicrete state space.

In many practical situations one needs to know the limiting distribution as $t \to \infty$, *i.e.*

$$p_n = \lim_{t\to\infty} p_n(t)$$

which is referred to as the *Steady state* probability exactly n customers in the system.

Since the "arrival process" and the "completion process" are both Poisson with rates λ and μ respectively, we have the following:

(i) In the time interval $(t, t + \Delta t)$, the probability of one arrival is $\lambda\Delta t + o(\Delta t)$.

(ii) In the time interval $(t, t + \Delta t)$, the probability of more than one arrival $o(\Delta t)$.

(iii) In the time interval $(t, t + \Delta t)$, the probability of no arrival $1 - \lambda\Delta t + o(\Delta t)$.

(iv) In the time interval $(t, t+\Delta t)$,the probability of one departure is $\mu\Delta(t) + o(\Delta t)$.

(v) In the time interval (t, t+Δt), the probability of more than one departure is $o(\Delta t)$.

(vi) In the time interval $(t, t+\Delta t)$), the probability of no departure is $1 - \mu\Delta t + o(\Delta t)$.

(vii) In the time interval $(t, t+\Delta t)$, the probability of no arrival and no departure is $\big(1 - \lambda\Delta t + o(\Delta t)\big)\big(1 - \mu\Delta t + o(\Delta t)\big) = 1 - \lambda\Delta t - \mu\Delta t + o(\Delta t)$.

(viii) In the time interval $(t, t+\Delta t)$, the probability of one arrival and one departure is $\big(\lambda\Delta t + o(\Delta t)\big)\big(\mu\Delta t + o(\Delta t)\big) = o(t)$.

(ix) In the time interval $(t, t+\Delta t)$, the probability of one arrival and no departure is $\big(\lambda\Delta t + o(\Delta t)\big)\big(1 - \mu\Delta t + o(\Delta t)\big) = \mu\,\Delta\,(t) + o(\Delta t)$.

(x) In the time interval $(t, t+\Delta t)$, the probability of no arrival and one departure is $\big(1 - \lambda\Delta t + o(\Delta t)\big)\big(\mu\Delta(t) + o(\Delta t)\big) = \mu\,\Delta\,(t) + o(\Delta t)$.

(xi) In the time interval $(t, t+\Delta t)$, the probability of r arrival and s departure is $o(\Delta t)$, where at least one of r and s is ≥ 2.

Equation for $p_n(t)$:

For $n = 0$

$$p_0(t+\Delta t)$$

$$= p_0(t)P\left(\text{no arrival in } (t, t+\Delta t)\right) + p_1(t)P\left(1 \textit{ departure in } (t, t+\Delta t)\right)$$

$$+ \sum_{k=2}^{\infty} p_k(t)P(k \textit{ departures in } (t, t+\Delta t)$$

$$= p_0(t)[1 - \lambda\Delta t + o(\Delta t)] + p_1(t)[\mu\Delta t + o(\Delta t)] + o(\Delta t)$$

Then

$$\frac{p_0(t+\Delta t) - p_0(t)}{\Delta t} = \mu p_1(t) - \lambda p_0(t) + \frac{o(\Delta t)}{\Delta t}$$

Let $\Delta(t) \to 0$, then

$$\frac{d}{dt}p_0(t) = \mu p_1(t) - \lambda p_0(t) \qquad (1)$$

For $n \geq 1$

$$\begin{aligned}
&p_n(t+\Delta t)\\
&= p_{n-1}(t)P\ [\text{one arrival, no departure } in\ \ (t+t+\Delta t)]\\
&+p_n(t)P\ [\text{no arrival, no departure in } (t, t+\Delta t)]\\
&+p_{n+1}(t)P[\text{no arival, one departure in } (t, t+\Delta t)] + o(\Delta t)\\
&= p_{n-1}(t)[\lambda\Delta(t) + o(\Delta t)] + p_n(t)[1 - \lambda\Delta(t) - \mu\Delta(t) + o(\Delta t)]\\
&+p_{n+1}(t)[\mu\Delta(t) + o(\Delta t)] + o(\Delta t)
\end{aligned}$$

Hence

$$\frac{p_n(t+\Delta t) - p_n(t)}{\Delta t} = \lambda p_{n-1}(t) - (\lambda+\mu)p_n(t) + \mu p_{n+1}(t) + \frac{o(\Delta t)}{\Delta t}$$

Letting $\Delta t \to 0$, we obtain

$$\frac{d}{dt}p_n(t) = \lambda p_{n-1}(t) - (\lambda+\mu)p_n(t) + \mu p_{n+1}(t) \qquad (2)$$

The system of differential difference equations represented by (1) and (2) govern the stochastic behavior of the M/M/1 queueing process over a passage of time.

Let us assume the existence of a "steady state". Then, as $t \to \infty, p_n(t))$ tends to a limit p_n, independent of t. The equations of steady-state probabilities p_n can be obtained by putting $p_n'(t) = 0$ and $p_n(t) = p_n$ in (1) and (2) we get

$$\left.\begin{array}{c} 0 = \mu\, p_1 - \lambda\, p_0 \\ 0 = \lambda\, p_{n-1} - (\lambda + \mu) p_n + \mu p_{n+1};\ (n \geq 1) \end{array}\right\} \tag{3}$$

or

$$\left.\begin{array}{c} p_1 = \rho\, p_0 \\ p_{n+1} = \rho p_n + (p_n - \rho\, p_{n-1});\ \ (n \geq 1) \end{array}\right\} \tag{4}$$

where

$$\rho = \frac{\lambda}{\mu} = \frac{\frac{1}{\mu}}{\frac{1}{\lambda}} = \frac{\text{mean service time}}{\text{mean interarrival time}} \tag{5}$$

ρ is called the "traffic intensity".

ρ can be interpreted as the expected number of arrivals in the mean service time. $\left(\lambda \times \frac{1}{\mu}\right)$. Notice that λ is expected number of arrivals per unit time and $1/\mu$ is mean service time. Thus $\lambda \times \frac{1}{\mu}$ is expected number of arrivals in the mean service time.

From (4), we obtain

$$\begin{aligned} p_0 &= p_0 \\ p_1 &= \rho\, p_0 \\ p_2 &= \rho\, p_1 + (p_1 - \rho\, p_0) \\ &= \rho\, p_1 \\ &= \rho^2\, p_0 \end{aligned}$$

$$\begin{aligned} p_3 &= \rho\, p_2 + (p_2 - \rho\, p_1) \\ &= \rho\, p_2 = \rho^3\, p_0 \\ &\vdots \\ p_n &= \rho^n\, p_0 \end{aligned}$$

Hence

$1 = \sum_{n=0}^{\infty} p_n = p_0(1-\rho)^{-1}$; assuming $\rho < 1$.

Therefore, if $\rho < 1$,

$$p_0 = 1 - \rho,$$
$$p_n = \rho^n(1-\rho), n \geq 1.$$

Notice that for the existence of a steady state solution ρ must be less than 1. The steady state distribution is geometric. Further, as $t \to \infty$, let L_s be the expected number of units in the system. Then

$$L_s = \sum_{n=0}^{\infty} n\rho^n(1-\rho)$$
$$= \frac{\rho}{1-\rho} = \frac{\lambda}{\mu - \lambda}. \qquad (6)$$

The probability that the server is free $= 1 - \rho$.

Applications: The M/M/1 queuing model, characterized by exponential inter-arrival and service times with a single server, is one of the simplest and most widely used models. It is applicable in scenarios where there is a single point of service and both arrivals and service times follow a memoryless (exponential) distribution.

Examples:

i. **Customer Service:** A single customer service representative handling queries can be modelled using the M/M/1 queuing system. This helps in estimating the average wait time for customers and determining the representative's workload.

ii. **Single Teller Bank:** A bank with a single teller where customers arrive randomly can use the M/M/1 model to predict queue lengths and waiting times, allowing the bank to adjust staffing as needed.

4. **M/M/k Queuing Process:** The M/M/k queuing process extends the M/M/1 model to multiple servers, characterized by Markovian (exponential) inter-arrival and service times

with k servers. This model is suitable for systems with several parallel service points, like multiple checkout counters in a supermarket. In this model,

i. **Markovian Inter-Arrival Times (M):** Inter-arrival times follow an exponential distribution.

ii. **Markovian Service Times (M):** Service times follow an exponential distribution.

iii. **Multiple Servers (k):** There are k servers.

Measures:

i. Utilization ($\boldsymbol{\rho}$): $\rho = \frac{\lambda}{k\mu}$.

ii. Average number in the system (L): Derived using the Erlang B and C formulas, which account for the probabilities of all servers being busy and at least one server being free.

iii. The Average time in the system (W): $W = \frac{L}{\lambda}$.

iv. Average number in the Queue (L_q): Calculate using the probability of having to wait and the average number of entities waiting..

v. Average Time in the Queue (W_q): $W_q = W - \frac{1}{\mu}$.

Applications: The M/M/k model extends the M/M/1 model to multiple servers, making it suitable for systems with several parallel service points. This model is useful in environments where services are provided simultaneously by multiple servers, and both arrival and service processes are memoryless.

Examples:

i. **Call Centers:** A call center with multiple operators receiving calls can be modelled using the M/M/k system. This helps in determining the number of operators required to handle call volumes while minimizing customer wait times.

ii. **Supermarket Checkouts:** A supermarket with several checkout counters can use the M/M/k model to manage queues and optimize the number of open counters, balancing customer service speed and staffing costs.

5. **M/G/1 Queuing Process:** The M/G/1 queuing process is characterized by Markovian (exponential) inter-arrival times and a general service time distribution with a single server. This model is useful for systems where service times do not follow an exponential distribution. In this model,

   i. Markovian Inter-Arrival Times (M): Inter-arrival times follow an exponential distribution.

   ii. General Service Time Distribution (G): Service times follow a general distribution.

   iii. Single Server (1): There is one server.

Measures:

i. Utilization ($\boldsymbol{\rho}$): $\rho = \lambda E[S]$.

ii. Average number in the system (L): Using the Pollaczek-Khinchine formula:

$$L = \lambda E[W] = \lambda\left(\frac{1}{\mu} + \frac{\lambda E[S^2]}{2(1-\lambda E[S])}\right)$$

iii. The Average time in the system (W): $W = E[S] + \frac{\lambda E[S^2]}{2(1-\lambda E[S])}$

iv. Average number in the Queue (L_q): $L_q = \frac{\lambda^2 E[S^2]}{2(1-\lambda E[S])}$

v. Average Time in the Queue (W_q): $W_q = \frac{\lambda E[S^2]}{2(1-\lambda E[S])}$.

Applications: The M/G/1 model, with exponential inter-arrival times and a general service time distribution, is suitable for systems where service times do not follow an exponential distribution. This model provides a more flexible framework for analysing queues with varied service processes.

Examples:

i. **Computer Systems:** In computing, tasks processed by a single server (such as a printer or a database server) often have service times that follow a distribution other than exponential. The M/G/1 model helps in evaluating the performance of these systems, predicting delays, and optimizing resource allocation.

ii. **Repair Services:** A repair shop with a single technician where repair times vary significantly depending on the complexity of the problem can use the M/G/1 model. This helps in understanding the expected wait times for customers and optimizing scheduling.

10.7 Summary

Queuing theory is a crucial mathematical study focusing on the analysis of waiting lines or queues. Originating from the work of Agner Krarup Erlang in the early 20th century, the field has expanded to incorporate a wide range of models and applications. Queuing models like M/M/1, M/M/k, M/G/1, Ek/M/1, and M/Ek/1 are fundamental tools used to evaluate and optimize systems in telecommunications, healthcare, manufacturing, transportation, and service industries. Each model has specific characteristics and applications, making it suitable for different types of queuing systems.

The M/M/1 model, with its single server and exponential inter-arrival and service times, provides a straightforward yet powerful tool for analysing simple systems. The M/M/k model extends this to multiple servers, making it ideal for systems with parallel service points. The M/G/1 model allows for a more general service time distribution, offering flexibility in analysing systems with varied service processes. The Ek/M/1 and M/Ek/1 models cater to scenarios where arrivals or services occur in stages, providing a more detailed representation of such processes.

Performance measures such as utilization, average number in the system, average time in the system, average number in the queue, and average time in the queue are critical in evaluating the effectiveness of queuing models. These measures help in understanding system performance, identifying bottlenecks, and making informed decisions to optimize resource utilization and reduce waiting times.

Queuing theory's applications are vast and diverse. In telecommunications, it helps manage call traffic and optimize network resources. In healthcare, it improves patient flow and resource allocation. In manufacturing, it optimizes production lines and reduces bottlenecks. In retail and banking, it enhances customer service by managing queues and optimizing staffing levels. Advanced mathematical techniques, such as Markov chains and simulation, further enhance the analysis of complex queuing systems, providing deeper insights and more accurate predictions.

Overall, queuing theory remains a vital tool for improving operational efficiency and service quality across various industries. By understanding and applying the appropriate queuing models, organizations can achieve significant improvements in their systems, leading to better customer satisfaction and resource management.

10.8 Self-Assessment Questions

1. Define queuing theory and explain its importance in various fields.
2. What are the key elements specified in Kendall's notation for queuing models? Explain each element.
3. Compare and contrast the M/M/1 and M/M/k queuing models. Under what circumstances would you use each?
4. Explain the significance of the utilization factor (ρ) in queuing models. How is it calculated for an M/M/1 system?
5. Describe the Erlang-k distribution. In what scenarios would an Ek/M/1 queuing model be more appropriate than an M/M/1 model?
6. What are the primary performance measures used in queuing theory? How do they help in analysing the performance of queuing systems?
7. Discuss the differences between M/G/1 and M/M/1 queuing models. Provide examples of systems that would be best modelled by M/G/1.
8. Explain the Renewal Reward Theorem and its significance in the context of queuing theory. Provide a detailed proof of the theorem.
9. How can queuing models be used to optimize the performance of a call center? Illustrate with an example using an M/M/k model.

10. Discuss the role of advanced mathematical techniques, such as Markov chains and simulation, in queuing theory. How do they enhance the analysis of complex queuing systems?

10.9 References

- Gross, D., Shortle, J. F., Thompson, J. M., & Harris, C. M. (2008). Fundamentals of Queueing Theory (4th ed.). John Wiley & Sons.
- Kleinrock, L. (1975). Queueing Systems, Volume 1: Theory. Wiley-Interscience.
- Cooper, R. B. (1981). Introduction to Queueing Theory. North Holland.
- Feller, W. (1971). An Introduction to Probability Theory and Its Applications (Vol. 2). John Wiley & Sons.
- Cox, D. R., & Smith, W. L. (1961). Queues. Methuen.
- Kelly, F. P. (1979). Reversibility and Stochastic Networks. John Wiley & Sons.
- Bolch, G., Greiner, S., de Meer, H., & Trivedi, K. S. (2006). Queueing Networks and Markov Chains: Modeling and Performance Evaluation with Computer Science Applications. John Wiley & Sons.
- Walrand, J. (1988). An Introduction to Queueing Networks. Prentice Hall.

10.10 Further Reading

- Bose, S. K. (2002). An Introduction to Queueing Systems. Springer.
- Harchol-Balter, M. (2013). Performance Modeling and Design of Computer Systems: Queueing Theory in Action. Cambridge University Press.
- Kingman, J. F. C. (1969). On the Mathematics of Queueing. Methuen.
- Takagi, H. (1991). Queueing Analysis: A Foundation of Performance Evaluation, Volume 1: Vacation and Priority Systems. North-Holland.
- Whitt, W. (2002). Stochastic-Process Limits: An Introduction to Stochastic-Process Limits and Their Application to Queues. Springer.
- Medhi, J. (2003). Stochastic Models in Queueing Theory (2nd ed.). Academic Press.
- Asmussen, S. (2003). Applied Probability and Queues (2nd ed.). Springer.

UNIT - 11 DISTRIBUTIONS

Structure

11.1 Introduction

Distributions are fundamental to the field of probability and statistics, playing a crucial role in modelling and analyzing random phenomena. The study of distributions allows us to understand and predict the behavior of systems under uncertainty, which is essential in various domains such as finance, engineering, operations research, and the natural sciences. This unit delves into specific types of distributions that are particularly useful for complex modelling scenarios, including compound distributions and queuing models.

The concept of probability distribution has evolved over centuries, starting with early work on gambling and games of chance. In the 17th century, the formal study of probability began with mathematicians such as Blaise Pascal and Pierre de Fermat, who laid the groundwork for probability theory through their correspondence on gambling problems. The 18th century saw further development with contributions from Jakob Bernoulli and Abraham de Moivre. U Bernoulli's Law of Large Numbers and de Moivre's Normal Distribution were significant milestones in the understanding of probability distributions.

During the 19th century, Carl Friedrich Gauss's work on the normal distribution and Pierre-Simon Laplace's development of the central limit theorem were pivotal. This period also saw the emergence of statistical inference with contributions from Ronald A. Fisher and Karl Pearson, who

developed techniques for estimating population parameters and hypothesis testing. The 20th century marked a significant expansion in the field, with the introduction of new distributions and models to address more complex problems. The Poisson distribution, introduced by Siméon-Denis Poisson, became a cornerstone for modelling rare events. Queuing theory, developed by Agner Krarup Erlang, provided the basis for modern telecommunications and service systems analysis.

In recent decades, the study of distributions has seen significant advancements due to the advent of modern computing and the development of sophisticated mathematical tools. Computational statistics has revolutionized the field, enabling the simulation and analysis of complex distributions that are intractable by analytical methods alone. Techniques such as Monte Carlo simulations and Markov Chain Monte Carlo (MCMC) methods are widely used for approximating distributions and solving complex probabilistic models.

The rise of machine learning has further transformed the landscape, with probabilistic models at the heart of many algorithms. Bayesian networks, Gaussian processes, and hidden Markov models rely heavily on understanding and manipulating probability distributions to make predictions and infer hidden states. The explosion of data in various fields has led to the development of new distributions and models to handle large-scale and high-dimensional data. Techniques for dealing with overdispersion, heavy tails, and multimodality have become increasingly important in the era of big data.

Advances in queuing theory have addressed more complex systems, including networks of queues, priority queues, and queues with time-varying arrival and service rates. These developments are critical for optimizing performance in computer networks, telecommunications, and manufacturing systems. Queuing models are essential for analyzing and designing systems where resources are shared among competing demands, ensuring efficient and effective operations. Understanding distributions is essential for making informed decisions in the presence of uncertainty. They are used to model various phenomena, from the number of customers arriving at a service center to the time between failures of mechanical systems. Applications of probability distributions are vast and varied, including finance, engineering, operations research, and health sciences. In finance, distributions are used for modelling stock prices, assessing risk, and optimizing portfolios. In engineering, they are critical for reliability analysis, quality control, and risk management. Operations research relies on distributions for inventory management, logistics,

and production planning. In the health sciences, distributions are used in epidemiology, survival analysis, and clinical trials.

The study of distributions is a cornerstone of probability and statistics, with a rich historical background and significant modern advancements. This unit will provide a comprehensive overview of compound distributions, machine interference problems, and waiting time distributions for M/M/1 and M/M/k models, equipping you with the tools to model and analyze various stochastic systems effectively.

11.2 Objectives

The objectives of this unit are to equip you with a comprehensive understanding of specific types of distributions that are vital in the analysis and modelling of stochastic systems. By delving into compound distributions, you will learn how to combine multiple distributions to model complex random phenomena, such as the total claim size in insurance or aggregated risks in finance. This knowledge will enable you to handle scenarios where outcomes depend on a random number of underlying events, enhancing your ability to analyze and predict system behavior under uncertainty.

Furthermore, the unit aims to provide you with the skills to tackle the machine interference problem, which involves optimizing the allocation of repairmen in systems with multiple machines that may fail and require repair. Understanding this problem will help you minimize downtime and improve operational efficiency in manufacturing, telecommunications, and other service-oriented industries.

A critical part of this unit is the study of waiting time distributions for M/M/1 and M/M/k queuing models. You will learn to calculate and interpret these distributions, which are essential for analyzing and optimizing queuing systems with single or multiple servers. This knowledge is crucial for managing customer wait times, improving service levels, and enhancing overall system performance in various applications, including customer service centers, healthcare, and computer networks.

By achieving these objectives, you will be well-prepared to apply these concepts and techniques to real-world scenarios, making informed decisions based on probabilistic models and improving the efficiency and reliability of complex systems.

11.3 Compound Distribution

Compound distributions are a powerful tool in probability and statistics, particularly useful in scenarios where the total outcome depends on a random number of events. These distributions arise when one or more parameters of a distribution are themselves random variables following another distribution. This section will explore the definition, mathematical formulation, important results, and examples of compound distributions.

A compound distribution combines two or more distributions to model complex stochastic processes. The key idea is that the parameter of one distribution (e.g., the number of events) is governed by another distribution, which adds an additional layer of randomness. This concept is crucial in various fields such as actuarial science, finance, and risk management, where outcomes are influenced by a multitude of uncertain factors.

Mathematical Formulation

Consider a random variable N that follows a distribution $P(N = n)$, and let X_i be independent and identically distributed ($i.i.d$) random variables with distribution $F(x)$. The compound distribution S is the sum of N such random variables:

$$S = X_1 + X_2 + \cdots + X_N$$

Here, N can be thought of as the number of claims, events, or occurrences, and X_i represents the size, cost, or magnitude of each event.

Important Results

Theorem: Expectation of Compound Distribution

The expected value of the compound distribution S is the product of the expected values of the number of events N and size of each event X.

$$E[S] = E[N].E[X]$$

where $E[N]$ is the expected value of N and $E[X]$ is the expected value of X_i.

Proof: Let $S = \sum_{i=1}^{N} X_i$ where N is a random variable representing the number of events, and X_i are $i.i.d.$ random variables representing the size of each event.

Using the law of total expectation:

$$E[S] = E\left[\sum_{i=1}^{N} X_i\right]$$

$$= E\left[E\left[\sum_{i=1}^{N} X_i \,|N\right]\right]$$

Given $N = n$, the inner expectation becomes:

$$E\left[\sum_{i=1}^{N} X_i \,|N = n\right]$$

$$= \sum_{i=1}^{n} E[X_i]$$

$$= nE[X]$$

Therefore,

$$E[S]$$

$$= E\big[N.E[X]\big]$$

$$= E[N].E[X]$$

Thus, the expected value of the compound distribution S is $E[N].E[X]$.

Theorem: Variance of Compound Distribution

The variance of S is given by

$$Var\,(S) = E[N].Var(X) + (E[X])^2.Var(N)$$

This result highlights the contribution of both the variability in the number of events and the variability in the size of each event to the overall variance.

Proof: Let $S = \sum_{i=1}^{N} X_i$. To find the variance, we use the law of total variance:

$$Var(S) = E[Var(S|N)] + Var(E[S|N])$$

First consider, $Var(S|N = n)$

$$\begin{aligned} & Var(S|N = n) \\ & = Var\left(\sum_{i=1}^{n} X_i\right) \\ & = n.Var(X) \end{aligned}$$

Taking the expectation of this with respect to N

$$\begin{aligned} & Var(S|N = n) \\ & = E[N.Var(X)] \\ & = E[N].Var(X) \end{aligned}$$

Next consider $Var(E[S|N])$

$$\begin{aligned} & Var(E[S|N]) \\ & = Var(N.E[X]) \\ & = (E[X])^2 Var\,(N) \end{aligned}$$

Combining these results,

$$Var\,(S) = E[N].Var(X) + (E[X])^2.Var(N)$$

Theorem: Probability Generating Function

The probability generating function (PGF) of S can be derived using the PGFs of N and X. If $G_N\,(t)$ and $G_X(t)$ are the PGFs of N and X respectively, the PGF of S is:

$$G_S(t) = G_N(G_X(t))$$

This property is particularly useful for deriving the distribution of S in closed form when the PGFs are known.

Proof: Let $S = \sum_{i=1}^{N} X_i$. The PGF of S is defined as

$$G_S(t) = E[t^S]$$

Given the law of total expectation, we have

$$G_S(t) = E[t^S]$$

$$= E\big[E[t^2|N]\big]$$

Given $N = n$, the inner expectation becomes:

$$E[t^S|N = n]$$

$$= E\big[t^{\sum_{i=1}^{n} X_i}\big]$$

$$= (E[t^X])^n$$

$$= (G_X(t))^n$$

Therefore,

$$G_S(t) = E\left[\big(G_X(t)\big)^N\right]$$

$$= G_N(G_X(t)).$$

Examples:

Poisson-Exponential Distribution: Suppose the number of claims *N* follows a Poisson distribution with parameter λ, and the size of each claim X_i follows an exponential distribution with rate μ. The total claim size S then follows a compound Poisson-exponential distribution.

Expectation: $E[S] = \lambda/\mu$

Variance: $Var(S) = \lambda/\mu^2$

This model is commonly used in insurance and risk management to assess the total risk from multiple claims.

Negative Binomial-Gamma Distribution: Consider a scenario where the number of events N follows a negative binomial distribution with parameters r and p, and each event magnitude X_i follows a gamma distribution with shape parameter α and rate parameter β. The resulting distribution S is a compound negative binomial-gamma distribution.

$$\text{Expectation: } E[S] = r.\frac{\alpha}{\beta}.\frac{1-p}{p}$$

$$\text{Variance: } Var\ (S) = r.\frac{\alpha}{\beta^2}.\left(\frac{1-p}{p} + \frac{1-p}{p^2}\right)$$

This model is useful in modelling aggregated loss distributions where both the frequency and severity of losses are random.

Binomial-Normal Distribution: Suppose the number of trials N follows a binomial distribution with parameters n and p, and the outcome of each trial X_i follows a normal distribution with mean μ and variance σ^2. The total outcome S is a compound binomial-normal distribution.

Expectation: $E[S] = n.p.\mu$

Variance: $Var(S) = n.p.(\sigma^2 + \mu^2)(1-p)$

This model can be applied in quality control and reliability testing where the number of successful trials and their outcomes are of interest.

Daily Life Examples of Compound Distributions:

Insurance Claims: Imagine you have an insurance company. Each day, the number of claims you receive is random. For example, on some days, you might receive many claims, while on other days, you might receive only a few. Additionally, the amount of money each claim requires to be paid out is also random. One day, you might have a few small claims, and another day, you might have a few large claims. The total amount of money you pay out in a day is a compound distribution

because it depends on both the number of claims (a random number) and the size of each claim (also random).

Grocery Store Checkout: At a grocery store, the number of customers who come to a checkout counter in an hour is random. Some hours, many customers might come, while in other hours, fewer customers might come. The amount of money each customer spends is also random, as some customers buy a lot, and others buy only a few items. The total sales at the checkout counter in an hour is a compound distribution because it depends on both the number of customers (random) and the amount each customer spends (random).

Counting Daily Steps: Suppose you track the number of steps you take each day. The number of times you go for a walk or move around is random each day. Additionally, the number of steps you take each time you walk is also random. The total number of steps you take in a day is a compound distribution because it depends on both the number of walking sessions (random) and the number of steps in each session (random).

Weather and Rainfall: Imagine you are interested in how much rain falls in your city each month. The number of rainy days in a month is random. Some months have many rainy days, and others have only a few. Additionally, the amount of rain that falls on each rainy day is also random. The total rainfall in a month is a compound distribution because it depends on both the number of rainy days (random) and the amount of rain on each day (random).

Car Repairs: If you own a car repair shop, the number of cars that need repairs each week is random. Some weeks, many cars might come in, while in other weeks, only a few cars might come in. The cost to repair each car is also random because some cars need minor repairs, and others need major repairs. The total revenue from car repairs in a week is a compound distribution because it depends on both the number of cars (random) and the cost of repairing each car (random).

These examples illustrate how compound distributions are present in everyday situations, where outcomes depend on a combination of multiple random factors.

Compound distributions offer a versatile framework for modelling complex stochastic processes where outcomes depend on a random number of underlying events. By combining different distributions, we can capture the intricacies of real-world phenomena, making compound

distributions invaluable in various applications such as insurance, finance, and risk management. Understanding the mathematical properties and key results of compound distributions enables us to analyse and interpret these models effectively, providing deeper insights into the behavior of complex systems under uncertainty.

11.4 Machine Interference Problem

The machine interference problem, also known as the repairman problem, is a significant topic in operations research and industrial engineering, dealing with scenarios where multiple machines operate and occasionally require repair. This problem's study is crucial for optimizing system performance and minimizing downtime in manufacturing, telecommunications, and other service-oriented industries.

The origins of the machine interference problem can be traced back to the early 20th century with the advent of queuing theory, developed by Agner Krarup Erlang. Erlang's work on telephone exchange systems laid the groundwork for understanding how to manage resources in systems with random demands. Over the decades, this initial work was extended to more complex systems involving multiple machines and repairmen, leading to the formulation of the machine interference problem. During the mid-20th century, researchers like D.G. Kendall and J.R. Jackson made significant contributions by developing the mathematical frameworks for analysing such systems. Kendall introduced the notation for queuing models, and Jackson developed network queuing theory, both of which are fundamental to understanding and solving the machine interference problem.

In recent years, the field has seen advancements driven by increased computational power and the development of more sophisticated mathematical techniques. Modern approaches leverage simulation methods and advanced algorithms to handle more complex and realistic scenarios. Techniques such as Monte Carlo simulations, Markov Chain Monte Carlo (MCMC), and stochastic optimization have enhanced our ability to model and solve machine interference problems effectively.

Additionally, advancements in data analytics and machine learning have provided new tools for predictive maintenance and real-time optimization of repair schedules. These technologies enable more precise and efficient management of machine failures and repairs, significantly reducing downtime and improving overall system performance.

Problem Setup: Consider a system with m machines and r repairmen. Each machine operates independently and can fail according to a Poisson process with rate λ. When a machine fails, it requires repair, which takes an exponentially distributed amount of time with rate μ. The objective is to analyse the system's performance, focusing on metrics such as the average number of machines in operation, the average number of machines under repair, and the average waiting time for repair.

Important Results:

1. **Probability of k Machines Operating:** The steady-state probability P_k that k machines are operating can be derived using a birth-death process. Let n be the number of machines operating, where $0 \leq n \leq m$.

The birth rate (repair completion) is $\mu(m-n)$, and the death rate (failure) is $\lambda.n$. The steady state probability P_n is given by:

$$P_n = \frac{\left(\frac{\lambda}{\mu}\right)^n}{\sum_{k=0}^{m}\binom{m}{k}\left(\frac{\lambda}{\mu}\right)^k}.$$

Proof: To derive the steady-state probabilities, we use the balance equations of the birth-death process. For n machines in operation, the rate at which the system moves from n to $n-1$ (due to a machine failure) is λn, and the rate at which it moves from n to $n+1$ (due to a repair completion) is $\mu(m-n)$.

At steady state, the rate of flow into state n equals the rate of flow out of state n:

$$\lambda(n-1)P_{n-1} = \mu(m-n+1)P_n$$

Rearranging and solving for P_n

$$P_n = \frac{\lambda(n-1)P_{n-1}}{\mu(m-n+1)}$$

Using the boundary condition P_0 (when all machines are operating):

$$P_n = P_0 \frac{\left(\frac{\lambda}{\mu}\right)^n}{\prod_{k=1}^{n}(m-k+1)}$$

To find P_0, we use the normalisation condition $\sum_{n=0}^{m} P_n = 1$

$$P_0 . \sum_{n=0}^{m} \frac{\left(\frac{\lambda}{\mu}\right)^n}{\prod_{k=1}^{n}(m-k+1)} = 1$$

$$P_0 = \frac{1}{\sum_{k=0}^{m} \binom{m}{k} \left(\frac{\lambda}{\mu}\right)^k}$$

And

$$P_n = \frac{\left(\frac{\lambda}{\mu}\right)^n}{\sum_{k=0}^{m} \binom{m}{k} \left(\frac{\lambda}{\mu}\right)^k}.$$

2. **Expected number of Machines in Operation:** The expected number of machines in operation $E[N]$ can be calculated using the steady state probabilities:

$$E[N] = \sum_{n=0}^{m} n. P_n$$

Since the distribution is Binomial, we have

$$E[N] = m. \frac{\mu}{\lambda + \mu}$$

3. **Expected number of Machines Under Repairmen:** The expected number of machines under repair $E[R]$ is obtained as

$$E[R] = m - E[N]$$
$$= m. \frac{\lambda}{\lambda + \mu}$$

4. **Utilization of Repairmen:** The utilization ρ of repairmen is given by

$$\rho = \frac{\lambda m}{r\mu}$$

For a system to be stable, $\rho < 1$. Ensuring ρ<1 guarantees system stability, meaning the repairmen can handle the incoming repairs without an infinite backlog.

The machine interference problem highlights the critical need for efficient resource allocation in systems with multiple operational and repairable components. From its early roots in queuing theory to recent advancements driven by computational power and data analytics, understanding and solving this problem remain essential for optimizing performance and minimizing downtime in complex systems. By applying principles from birth-death processes and leveraging modern techniques, we can effectively manage and improve the operational efficiency of such systems.

Examples of the Machine Interference Problem

Home Appliances: In a household with multiple appliances such as a refrigerator, washing machine, dishwasher, and oven, each appliance operates independently but can occasionally break down. When an appliance breaks down, it needs to be repaired by a technician. If the household has a maintenance contract with a company that sends repairmen, the machine interference problem arises. The company needs to allocate its limited number of repairmen to multiple households to ensure that appliances are repaired promptly, minimizing the inconvenience to the residents.

IT Support in a Company: In a large company with hundreds of computers and network devices, IT support staff must handle hardware failures, software issues, and network problems. Each device or software can fail at random times, requiring the attention of IT personnel. The IT department has a limited number of support staff who must prioritize and attend to these issues. Efficiently managing the IT support team to minimize downtime for employees' devices and systems is an example of solving the machine interference problem.

Elevators in a High-Rise Building: In a high-rise building with multiple elevators, each elevator can occasionally break down and require maintenance. The building management has a team of technicians responsible for repairing the elevators. Since the elevators break down randomly and need to be repaired quickly to ensure smooth operation for residents and employees, the management must allocate the repair team efficiently. This situation reflects the machine interference problem, where the goal is to minimize the waiting time for elevator repairs and ensure the availability of elevators.

Agricultural Equipment on a Farm: A large farm uses various machines like tractors, harvesters, and irrigation systems, which can break down randomly. The farm employs a limited number of mechanics to repair these machines. To keep the farming operations running smoothly, the farm must manage the mechanics' workload effectively, ensuring that broken machines are repaired quickly and downtime is minimized. This is another instance of the machine interference problem, where efficient allocation of repair resources is crucial.

Hospital Equipment Maintenance: In a hospital, critical medical equipment such as MRI machines, X-ray machines, and ventilators can fail and require repair. The hospital's maintenance department has a limited number of technicians who must ensure that the equipment is repaired promptly to avoid disruptions in patient care. Managing the repair schedule and prioritizing equipment based on urgency and impact on patient care involves solving the machine interference problem.

Manufacturing Plant Operations: A manufacturing plant operates several machines on the production line. These machines can fail randomly and need repair to keep the production process running smoothly. The plant has a team of maintenance workers who must attend to the broken machines. Efficiently managing the maintenance workers to minimize production downtime and maintain a steady flow of operations is a classic example of the machine interference problem.

These examples illustrate how the machine interference problem is present in various aspects of daily life, where multiple operational units can fail randomly and require repair by a limited number of maintenance personnel. Efficiently managing these resources ensures minimal downtime and smooth operation of the systems involved.

11.5 Waiting Time Distribution for M/M/1 and M/M/k Models

In queuing theory, the M/M/1 and M/M/k models are fundamental for understanding how customers (or jobs) are processed in a system with single or multiple servers. These models help analyze the waiting time distributions, which are crucial for optimizing service efficiency and managing customer satisfaction.

M/M/1 Model: The M/M/1 model represents a single-server queue with Poisson arrivals and exponential service times. Here, "M" stands for "memoryless" (exponential interarrival and service time distributions), and "1" indicates a single server.

Parameters:

λ: Arrival rate (customers per unit time)

μ: Service rate (customers served per unit time)

Queueing time for a customer is the time that lapses between his arrival and the departure on completion of his service.

Theorem 1: For $M/M/1\ (\infty, FIFO)$ queueing model with $\rho < 1$, the steady state probability distribution of the queueing time is exponential with mean

$$\frac{1}{\mu(1-\rho)} = \frac{1}{\mu - \lambda}.$$

Proof: Let T be the queueing the for a customer and $g(t)$ be the *pdf* of T. Let $g(t/m)$ be the conditional pdf of T, given that there are n customers on his arrival. Then, we have

$$g(t) = \sum_{n=0}^{\infty} g\left(\frac{t}{n}\right) p_n \qquad (7)$$

$g\left(\frac{t}{n}\right)$ is the *pdf* of the sum of n, *iid*. exponential random variables with mean $1/\lambda$ plus the remaining service time of the customer being served, which is also exponential (by the memoryless property) with mean $1/\lambda$. Hence

$$g\left(\frac{t}{n}\right) = \frac{\mu\, e^{-\mu t}\, (\mu t)^n}{n!} \quad (0 < t < \infty) \qquad (8)$$

From (7) and (8), we have

$$g(t)$$

$$= \mu\, e^{-\mu t} \sum_{n=0}^{\infty} \frac{(\mu t)^n}{n!} p_n$$

$$= \mu\, e^{-\mu t} \sum_{n=0}^{\infty} \frac{(\mu t)^n}{n!} (1-\rho)\rho^n$$

$$= \mu\,(1-\rho) e^{-\mu t\,(1-e)}, \quad 0 < t < \infty$$

Hence the theorem follows.

Waiting Time in the Queue is the time from the arrival of the customer to the beginning of his service. Let W be the waiting time in the queue. Then $P(W = 0)$ is the probability of no customer on his arrival. Obviously

$$P(W = 0) = 1 - \rho$$

If there is at least one customer on his arrival than he has to wait and the waiting time has the *pdf*

$$g(w) = \sum_{n=1}^{\infty} h(w|n) p_n$$

Here $h(w|n))$ is the conditional *pdf* of the waiting time given that there are n customers on his arrival. Hence

$$g(w) = \sum_{n=1}^{\infty} \frac{\mu\, e^{-\mu w}\, (\mu w)^{n-1}}{(n-1)!}\, (1-\rho)\rho^n$$

$$= \rho(1-\rho)\mu e^{-\mu(1-\rho)w}; \quad 0 < w < \infty$$

Therefore, the waiting time W has the *pdf*

$$g(w) = \begin{cases} 0, & if\ w < 0 \\ 1 - \rho + \int_0^w \rho(1-\rho)\mu e^{-\mu(1-\rho)x} dx\,, & if\ w \geq 0. \end{cases}$$

or

$$g(w) = \begin{cases} 0 & if\ w < 0 \\ 1 - \rho\, e^{-\mu(1-\rho)w} & if\ w \geq 0. \end{cases}$$

Remark: The waiting time distribution for the M/M/k queue is more complex and involves the Erlang distribution. The probability that the waiting time W_q is less than or equal to t is given by the cumulative distribution function (CDF) of the Erlang distribution.

For k servers, the CDF of the waiting time distribution is:

$$P(W_q \leq t) = 1 - \sum_{n=0}^{k-1} \frac{(\mu t)^n e^{-\mu t}}{n!}$$

The M/M/1 and M/M/k models are essential tools in queuing theory, providing insights into the performance and efficiency of queuing systems. The M/M/1 model is simpler and serves as a foundational model, while the M/M/k model addresses more complex scenarios with multiple servers. Understanding the waiting time distributions in these models helps in optimizing resource allocation, minimizing customer wait times, and improving overall system performance.

11.6 Summary

In this unit, we explored the concept of distributions, with a particular focus on compound distributions and their applications in real-world scenarios. Compound distributions are essential in modelling situations where outcomes depend on a random number of events, such as insurance claims or total rainfall in a month. We discussed the mathematical formulation of compound distributions, highlighting key results like the expectation and variance, and provided practical examples to illustrate these concepts.

The machine interference problem, a vital topic in operations research, was examined next. This problem involves managing multiple machines that can fail and require repair, with a limited

number of repairmen available. Historical developments in queuing theory laid the groundwork for understanding this problem, while recent advancements have leveraged computational power and data analytics to optimize resource allocation and minimize downtime. We derived important results such as the expected number of machines in operation and under repair, and the utilization of repairmen, providing proofs to solidify understanding.

The unit also covered the M/M/1 and M/M/k queuing models, and fundamental frameworks for analysing systems with single and multiple servers, respectively. For the M/M/1 model, we discussed key metrics like the average number of customers in the system and queue, as well as the waiting time distribution, which follows an exponential pattern. The M/M/k model extends these concepts to multiple servers, introducing more complexity but also greater applicability to real-world systems. The waiting time distribution in the M/M/k model involves the Erlang distribution, reflecting the increased complexity of multi-server systems.

Throughout the unit, we emphasized the practical applications of these models in various industries, from manufacturing and telecommunications to healthcare and IT support. The self-assessment questions provided aim to reinforce the concepts covered, encouraging further exploration and application of these important probabilistic models.

11.7 Self-Assessment Questions

1. Define a compound distribution and provide a real-world example.
2. Explain the machine interference problem and discuss its practical applications.
3. Derive the expected number of machines in operation in the machine interference problem.
4. What are the key metrics for the M/M/1 queuing model, and how do they relate to each other?
5. Explain the waiting time distribution for the M/M/1 model and derive the probability that the waiting time is less than or equal to a given value.
6. Compare and contrast the M/M/1 and M/M/k queuing models, highlighting their differences in terms of system stability and performance metrics.
7. Describe the utilization factor in the M/M/k model and its significance for system stability.
8. Provide a detailed derivation of the average waiting time in the queue for the M/M/k model.
9. Discuss how advancements in computational methods have impacted the study and application of machine interference problems and queuing theory.

10. Give examples of modern techniques used to manage and optimize queuing systems in real-world applications.

11.8 References

- Gross, D., Shortle, J. F., Thompson, J. M., & Harris, C. M. (2018). Fundamentals of Queueing Theory. Wiley.
- Ross, S. M. (2019). Introduction to Probability Models. Academic Press.
- Tijms, H. (2003). A First Course in Stochastic Models. Wiley.
- Allen, A. O. (1990). Probability, Statistics, and Queueing Theory with Computer Science Applications. Academic Press.
- Medhi, J. (2003). Stochastic Models in Queueing Theory. Academic Press.
- Wolff, R. W. (1989). Stochastic Modelling and the Theory of Queues. Prentice Hall.

11.9 Further Reading

- Allen, A. O. (1990). Probability, Statistics, and Queueing Theory with Computer Science Applications. Academic Press.
- Medhi, J. (2003). Stochastic Models in Queueing Theory. Academic Press.
- Wolff, R. W. (1989). Stochastic Modeling and the Theory of Queues. Prentice Hall.
- Kleinrock, L. (1975). Queueing Systems, Volume 1: Theory. Wiley-Interscience.
- Cooper, R. B. (1981). Introduction to Queueing Theory. North Holland.
- Kingman, J. F. C. (1969). On the Algebra of Queues. Mathematical Proceedings of the Cambridge Philosophical Society, 63(1), 157-170.

UNIT – 12: MARTINGALES AND THEIR APPLICATIONS

Structure

12.1 Introduction

Martingales are a fundamental concept in probability theory and have widespread applications in various fields, including finance, stochastic processes, and statistical inference. Originating from the study of fair games, the martingale theory has evolved into a powerful mathematical framework for analysing and predicting the behaviour of stochastic processes. The concept was first formalized in the mid-20th century by the French mathematician Paul Lévy, and it was further developed by Joseph Doob, whose work laid the groundwork for modern probability theory.

The history of martingales begins with the study of gambling systems in the 18th and 19th centuries. Gamblers were interested in strategies that could maximize their winnings, leading to the development of various betting systems. The term "martingale" itself originated from a particular betting strategy where the gambler doubles the bet after each loss, aiming to recover all previous losses with a single win. However, it was Paul Lévy who, in the early 20th century, formalized the idea of martingales within the framework of probability theory.

Joseph Doob's contributions in the 1940s and 1950s were pivotal. He extended the concept of martingales and developed a comprehensive theory that connected them to other areas of mathematics, such as measure theory and stochastic processes. Doob's work provided the rigorous

mathematical foundation needed to study martingales in a broader context, beyond just gambling and betting.

Martingales are essentially sequences of random variables that represent fair games, where the expected future value, given all past information, is equal to the current value. This memoryless property makes martingales particularly useful in modelling situations where future predictions are based solely on present conditions without any bias from past events. This property also aligns well with the concept of "no arbitrage" in financial markets, where the current price of an asset should reflect all available information, preventing guaranteed profit without risk.

In the mid-20th century, martingale theory found significant applications in the field of finance. The concept of risk-neutral pricing emerged, where the prices of financial derivatives are modelled using martingales. The groundbreaking work of Black, Scholes, and Merton in the 1970s on option pricing leveraged martingale theory, leading to the development of the Black-Scholes model. This model revolutionized financial engineering by providing a method to price options and other derivatives accurately.

Beyond finance, martingales are used in various other domains. In the realm of gambling and betting, martingales model the fairness of games of chance. In the study of stochastic processes, martingales are used to describe phenomena ranging from random walks to Brownian motion. In statistical inference, martingales provide tools for sequential analysis and hypothesis testing.

Recent developments in martingale theory have been driven by advances in computational power and the increasing complexity of applications. Modern financial markets, with their high-frequency trading and complex derivative products, require sophisticated models that can capture intricate dependencies and risks. Computational techniques such as Monte Carlo simulations and algorithmic trading strategies often rely on martingale properties to ensure fair and accurate pricing.

In machine learning and data science, martingales have found applications in online learning algorithms and adaptive data analysis. The ability to model and predict outcomes in real-time, updating predictions as new data arrives, aligns well with the principles of martingales. Techniques such as the martingale method of sequential hypothesis testing are used to control error rates and ensure robust decision-making in dynamic environments.

This unit will delve into the theory of martingales, starting with their definition and basic properties. We will explore significant theorems, such as Doob's Decomposition Theorem, which allows us to break down a sub-martingale into a martingale and a predictable increasing process. This decomposition is crucial for understanding the structure and behaviour of more complex stochastic processes.

We will also examine martingale convergence theorems, which provide conditions under which martingales converge. These theorems are essential for ensuring the stability and predictability of processes modelled by martingales, with implications for long-term predictions and assessments.

Understanding martingales and their properties equips us with powerful tools for analysing and modelling uncertainty. Whether in financial markets, where predicting asset prices is critical, or in statistical methods, where we make inferences based on sequential data, martingales offer a robust framework for dealing with randomness and uncertainty.

The study of martingales bridges theoretical mathematics and practical applications, providing a comprehensive approach to understanding and managing stochastic processes. This unit aims to equip you with the knowledge and skills to apply martingale theory to various real-world problems, enhancing your ability to analyse and make decisions in uncertain environments.

12.2 Objectives

The primary objective of this unit is to provide a comprehensive understanding of martingales, their properties, and their applications in various fields. By delving into the theory of martingales, you will gain a robust foundation in one of the most critical areas of modern probability theory. This unit aims to equip you with the ability to define and identify martingales in different contexts, understand their underlying properties, and apply these concepts to solve complex problems involving stochastic processes.

A significant focus will be on Doob's Decomposition Theorem, which is fundamental in breaking down sub martingales into simpler components. Understanding this theorem will enable you to analyse more complex stochastic processes by decomposing them into martingales and predictable increasing processes. This decomposition is crucial for many applications, particularly in financial modelling and risk management, where it helps in understanding the dynamics of asset prices and their behaviour over time.

Furthermore, the unit will explore martingale convergence theorems, which are essential for determining the long-term behaviour of martingales. By studying these theorems, you will learn the conditions under which martingales converge almost surely or in L^(1)providing a framework for making reliable predictions in uncertain environments. This knowledge is particularly valuable in fields such as finance, where the convergence of martingales is used to model the stability of investment returns and the pricing of derivatives.

Another objective is to illustrate the practical applications of martingales in various domains. In finance, for example, martingales are used to model fair games and the absence of arbitrage opportunities, which are foundational concepts in the pricing of options and other derivatives. By understanding these applications, you will be able to apply martingale theory to real-world financial problems, enhancing your ability to develop effective trading strategies and manage financial risks.

Additionally, the unit aims to provide insights into recent developments and advanced topics in martingale theory. This includes exploring how martingales are used in modern computational techniques, such as Monte Carlo simulations and algorithmic trading, as well as their applications in machine learning and data science. By staying abreast of these advancements, you will be well-prepared to leverage martingale theory in cutting-edge research and practical applications.

Ultimately, this unit seeks to foster a deep understanding of martingales and their significance in probability theory and beyond. By achieving these objectives, you will be equipped with powerful analytical tools for modelling and managing uncertainty, enabling you to make informed decisions and solve complex problems in various fields, from finance and economics to engineering and data science.

12.3 Martingales

Martingales are a fundamental concept in probability theory that describe a specific type of stochastic process. The concept originated in the context of gambling and fair games, but it has since evolved into a powerful mathematical framework with applications in various fields, including finance, economics, and the study of stochastic processes.

Conditional Expectation*:* Let $X_1, X_2, \ldots$ be a sequence of random variables and $\mathcal{F}_n$ denotes the information contained in $X_1, X_2, \ldots, X_n$. If Y is a function of $X_1, X_2, \ldots, X_n$ then

$$E(Y \mid \mathcal{F}_n) = Y; \forall\ Y \qquad (1)$$

$$E(E(Y \mid \mathcal{F}_n) \mid \mathcal{F}_m) = E(Y \mid \mathcal{F}_m) \quad \forall m < n \qquad (2)$$

If Y is independent of $X_1, X_2, \ldots, X_n$, then information about $X_1, X_2, \ldots, X_n$. should not be useful in determining Y

$$E(Y \mid \mathcal{F}_n) = E(Y) \qquad (3)$$

If Y is a random variable and Z is a random variable that is measurable with respect to $X_1, X_2, \ldots, X_n$, then

$$E(YZ \mid \mathcal{F}_n) = ZE(Y) \qquad (4)$$

Example 1: Suppose $X_1, X_2, \ldots,$ are *iid* random variables with mean μ and S_n denote the partial sum

$$S_n = X_1 + X_2 + \cdots + X_n.$$

Then, for $m < n$

$$E(S_n \mid \mathcal{F}_m) = E(X_1 + X_2 + \ldots + X_m \mid \mathcal{F}_m) + E(X_{m+1} + \ldots + X_n \mid \mathcal{F}_m)$$

Since, $X_1 + X_2 + \ldots + X_m$ is measurable with respect to $X_1, X_2, \ldots, X_m$, we obtain

$$\begin{aligned} &E(X_1 + X_2 + \ \ldots + X_m \mid \mathcal{F}_m) \\ &= X_1 + X_2 + \cdots + X_m \\ &= S_m \end{aligned}$$

Since $X_{m+1} + \ldots + X_n$ is independent of $X_1, X_2, \ldots, X_m$, we get

$$\begin{aligned} &E(X_{m+1} + \ \ldots + X_n \mid \mathcal{F}_m) \\ &= E(X_{m+1} + \ \ldots + X_n) \end{aligned}$$

$$= (n-m)\mu$$

Therefore,

$$E(S_n \mid \mathcal{F}_m) = S_m + (n-m)\mu.$$

Example 2: Suppose $X_1, X_2, \ldots,$ and S_n are as defined in Example 1. Suppose $\mu = 0$ and $Var(X_i) = E(X_i^2) = \sigma^2$. For $m < n$ we shall have

$$\begin{aligned}
&E(S_n^2|\mathcal{F}_m) \\
&= E[\{S_m + (S_n - S_m)\}^2|\mathcal{F}_m] \\
&= E(S_m^2|\mathcal{F}_m) + 2E(S_m(S_n - S_m)|\mathcal{F}_m) + E((S_n - S_m)^2|\mathcal{F}_m)
\end{aligned}$$

Since $\mathcal{F}_m$ depends only on $X_1, X_2, \ldots, X_m$ and $S_n - S_m$ is independent of $X_1, X_2, \ldots, X_m$ we have

$$\begin{aligned}
&E(S_m^2|\mathcal{F}_m) = S_m^2 \\
&E((S_n - S_m)^2|\mathcal{F}_m) \\
&= E(S_n - S_m)^2 \\
&= Var(S_n - S_m) \\
&= (n-m)\sigma^2
\end{aligned}$$

$$\begin{aligned}
&E(S_m(S_n - S_m)|\,\mathcal{F}_m) \\
&= E\big(S_m(S_n - S_m)\big) \\
&= S_m E(S_n - S_m) \\
&= 0
\end{aligned}$$

Therefore,

$$E(S_n^2|\mathcal{F}_m) = S_m^2 + (n-m)\sigma^2$$

Example 3: Consider a special case of Example 1 where the random variable X_i has a Bernoulli distribution

$$\begin{aligned}
&P(X_i = 1) = p, \\
&P(X_i = 0) = 1 - p
\end{aligned}$$

Again, assume that $m < n$. For any $i \leq m$, consider $E(X_i|S_n)$. If $S_n = k$, then there are k 1's in first n trial. Given $S_n = k$, we can show that

$$P(X_i = 1|S_n = k) = \frac{k}{n}$$

Hence

$$E(X_i = 1|S_n) = \frac{S_n}{n}$$

and

$$E(S_m|S_n) = E(X_1|S_n) + \cdots + E(X_m|S_n) = S_n \frac{m}{n}$$

Definition: Let $X_0, X_1, \ldots$ be a sequence of random variables and $\mathcal{F}_n$ denote the information contained in $X_1, X_2, \ldots, X_n$. We say that a sequence of random variables $M_0, M_1, M_2, \ldots$ with $E(|M_i|) < \infty$ is a martingale with respect to $\mathcal{F}_n$ if

- Each M_n is measurable with respect to $X_0, X_1, \ldots, X_n$;

and

$$E(M_n|\mathcal{F}_m) = M_m, \qquad \forall\, m < n \tag{5}$$

• The condition $E(|M_i|) < \infty$ is needed to guarantee that the conditional expectations are well defined.

• Sometimes we say that $M_0, M_1, \ldots$ is a martingale without referring to the random variables $X_0, X_1, \ldots$. It will mean that the sequence $\{M_n\}$ is a martingale with respect to itself where $\mathcal{F}_n$ is the information contained in $M_0, M_1, \ldots, M_n$.

Theorem 1: If $E(M_{n+1} \mid \mathcal{F}_n) = M_n \ \forall\, n$ then $M_0, M_1, \ldots$ is a martingale.

Proof: We have

$$E(M_{n+2}|\mathcal{F}_n)$$

$$= E(E(M_{n+2}|\mathcal{F}_{n+1}) \mid \mathcal{F}_n)$$
$$= E(M_{n+1}|\mathcal{F}_n) = M_n$$

and so on. Hence in general,

$$E(M_n \mid \mathcal{F}_m) = M_n, \forall\, m < n$$

Example 4 : Suppose $X_1, X_2, \ldots$, be independent random variables each with mean μ. Let $S_0 = 0$ and for $n > 0, S_n$ be the partial sum $S_n = X_1 + \ldots + X_n$, then $M_n = S_n - n\mu$ is a martingale with respect to $\mathcal{F}_n$ (information in $X_1, X_2, \ldots, X_n$). By using Example 1,

$$E(M_{n+1}|\mathcal{F}_n)$$
$$= E(S_{n+1} - (n+1)\mu|\mathcal{F}_n)$$
$$= E(S_{n+1}|\mathcal{F}_n) - (n+1)\mu$$
$$= (S_n + \mu) - (n+1)\mu$$
$$= M_n$$

Example 5: Suppose $X_1, X_2, \ldots$, are independent random variables with $P(X_i = 1) = P(X_i = -1) = 1/2$. For example, X_i is a result of a game where one tosses a fair coin and wins Rs.1 if the outcome is head and loses Rs.1 otherwise. One way to beat the game is to keep doubling our bet until we eventually win. At this point we stop. Let $W_0 = 0$ and W_n denote the winning (or loses) up to n tosses of the coin using this strategy. Whenever we win, we stop playing. Thus, our winnings stop changing and

$$P(W_{n+1} = 1|W_n = 1) = 1.$$

Suppose tails turned up the first n tosses of the coin. After each toss we have doubled our bet, so we have lost rupees $1 + 2 + \ldots + 2^{n-1} = 2^n - 1$ and $W_n = -(2^n - 1)$. At this time we double our bet again and wager 2^n on the next toss. This gives

$$P\left(W_{n+1} = 2^n - (2^n - 1)\middle|W_n = -(2^n - 1)\right)$$
$$= P\left(W_{n+1} = 1\middle|W_n = -(2^n - 1)\right)$$
$$= \frac{1}{2}$$

$$P\left(W_{n+1} = -(2^{n+1} - 1) \middle| W_n = -(2^n - 1)\right)$$

$$= \frac{1}{2}$$

$$E[W_{n+1}|\mathcal{F}_n]$$

$$= \frac{1}{2} \times 1 + \frac{1}{2} \times \left(-(2^{n+1} - 1)\right)$$

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

$$= W_n.$$

Therefore W_n is a martingale with respect to $\mathcal{F}_n$.

Example 6: Suppose $X_1, X_2, \ldots$, are as in previous example 5 and on the n^{th} toss we make a bet equal to B_n. In determining the amount of bet, we may look at the results of the first $(n - 1)$ tosses but cannot look beyond that. Thus, B_n is a random variable measurable with respect to $\mathcal{F}_{n-1}$. We assume that B_1 is a constant. the winning after n flips, W_n, are given by $W_0 = 0$ and

$$W_n = \sum_{j=1}^{n} B_j X_j$$

For ensuring that the bet at time n always less than some constant C_n assume that $E(|B_n|) < \infty$. Then W_n is a martingale with respect to $\mathcal{F}_n$. Now $E(B_n) < \infty \; \forall n$ implies that $E(|W_n|) < \infty$. Further, W_n is $\mathcal{F}_n$measurable and

$$E(W_{n+1}|\mathcal{F}_n)$$

$$= E\left(\sum_{j=1}^{n+1} B_j X_j | \mathcal{F}_n\right)$$

$$= E\left(\sum_{j=1}^{n} B_j X_j | \mathcal{F}_n\right) + E(B_{n+1} X_{n+1} | \mathcal{F}_n)$$

Using result (1) of conditional expectations

$$E\left(\sum_{j=1}^{n} B_j X_j | \mathcal{F}_n\right) = \sum_{j=1}^{n} B_j X_j = W_n$$

Again, B_{n+1} is $\mathcal{F}_n$ measurable. Hence using (3) and (4), we obtain

$$\begin{aligned} & E(B_{n+1}X_{n+1} | \mathcal{F}_n) \\ &= B_{n+1}E(X_{n+1} | \mathcal{F}_n) \\ &= 0 \end{aligned}$$

Therefore,

$$E(W_{n+1}|\mathcal{F}_n) = W_n.$$

Example 7 (Pyola's Urn): Consider an urn with balls of two colors, red and green. Assume that there is one ball of each color in the urn. We proceed as follows:

At each time step, a ball is chosen at random from the urn. If a red ball is chosen, it is returned and in addition another red ball is added to the urn. Similarly, if a green ball is chosen, it is returned together with another green ball.

Let X_n denote the number of red balls in the urn after n draws. Then $X_0 = 1$ and X_n is a (time homogeneous) Markov chain with transitions

$$P(X_{n+1} = k+1 | X_n = k) = \frac{k}{n+2}$$

$$P(X_{n+1} = k | X_n = k) = \frac{n+2-k}{n+2}$$

Notice that at time n+1 there are n+2 balls in the urn. Let

$$M_n = \frac{X_n}{n+2}$$

Then M_n is the fraction of red balls after n draws. Then M_n is a martingale. We have

$$E(X_{n+1}|X_n)$$

$$= X_n \frac{(n+2-X_n)}{n+2} + X_{n+1} \frac{X_n}{n+2}$$

$$= \frac{1}{n+2}[(n+2)X_n + X_n]$$

$$= X_n + \frac{X_n}{n+2}$$

Since this is a Markov chain, all the relevant information in $\mathcal{F}_n$ for determining X_{n+1} is contained in X_n. Therefore,

$$E(M_{n+1}|\mathcal{F}_n)$$

$$= E((n+3)^{-1}X_{n+1}|X_n)$$

$$= \frac{1}{n+3}\left[X_n + \frac{X_n}{n+2}\right]$$

$$= \frac{X_n}{n+2}$$

$$= M_n$$

Sub-Martingale and Super-Martingale

Definition: A process M_n with $E(|M_n| < \infty)$ is called a submartingale (supermartingale) with respect to $X_0, X_1, \ldots$ if $\forall\ m\ <\ n$,

$$E(M_n|\ \mathcal{F}_n) \geq (\leq)\ M_m.$$

- A sub martingale is a game in one's favor and a super martingale is an unfair game.
- A martingale is a model of fair game.
- M_n is a martingale if and only if it is both a sub martingale and a super martingale.

Optimal Sampling Theorem:

Theorem 1: **(Optional sampling Theorem):** Suppose $M_0, M_1, \cdots$ is a martingale with respect to $X_0, X_1, \cdots$ and T is a stopping time satisfying $P(T < \infty) = 1$,

$$E(|M_n| < \infty) \tag{6}$$

$$\lim_{n\to\infty} E\left(|M_n| I(T > n)\right) = 0 \tag{7}$$

Then, $E(M_T) = E(M_0)$. Here $I(\cdot)$ is an indicator function.

Proof: Let F_n be the information contained in $X_0, X_1, \cdots, X_n$ and $I(T > n)$, the indicator function of event $\{T > n\}$, is measurable with respect to $\mathcal{F}_n$ (Since we need only the information up to time n to determine if we have stopped by time n). M_T is the random variable which equals M_j if $T = j$ we can write

$$M_T = \sum_{j=0}^{K} M_j I(T = j)$$

$$E(M_T|\mathcal{F}_{K-1}) = E(M_K I(T = K)|\mathcal{F}_{K-1}) + \sum_{j=0}^{K} E\left(M_j I(T = j)|\mathcal{F}_{K-1}\right)$$

For $j \leq (K-1), M_j I(T = j)$ is $\mathcal{F}_{K-1}$ measurable; hence

$E\left(M_j I(T = j)|\mathcal{F}_{K-1}\right)$

$= M_j I(T = j)$

Since T is known to be no more than K, then event $\{T = K\}$ is the same as the event $\{T > K-1\}$. The latter event is measurable with respect to $\mathcal{F}_{K-1}$. Hence using eq. (4)

$E(M_K I(T = K)|\mathcal{F}_{K-1})$

$= E(M_K I(T > K-1)|\mathcal{F}_{K-1})$

$= I(T > K-1)E(M_K|\mathcal{F}_{K-1})$

$= I(T > K-1)E(M_{K-1})$

Therefore

$E(M_T|\mathcal{F}_{K-1})$

$$= I(T > K-1)E(M_{K-1}) + \sum_{j=0}^{K-1} E\left(M_j I(T=j)\right)$$

$$= I(T > K-2)E(M_{K-2}) + \sum_{j=0}^{K-2} E\left(M_j I(T=j)\right)$$

$E(M_T|\mathcal{F}_{K-2})$

$= E(E(M_K|\mathcal{F}_{K-1})|\mathcal{F}_{K-2})$

$$= I(T > K-3)E(M_{K-1}) + \sum_{j=0}^{K-3} E\left(M_j I(T=j)\right)$$

We continue this process until we get $E(M_T|\mathcal{F}_0) = M_0$. Now, consider the stopping time $T_n = min(T, n)$. Then

$M_T = M_{T_n} + M_T I(T > n) - M_n I(T > n)$

$E(M_T) = E\big(M_{T_n}\big) + E\big(M_T I(T > n)\big) - E\big(M_n I(T > n)\big)$

Since T_n is a bounded stopping time, we have $E\big(M_{T_n}\big) = M_0$, and $P(T > n) \to 0$ as $n \to \infty$. If $E|M_T| < \infty$ then $E\big(|M_T| I(T > n)\big) \to 0$.

If M_n and T are given so that

$$\lim_{n\to\infty} E\left(|M_T| I(T > n)\right) = 0$$

then, $E(M_T) = E(M_0)$. Hence the theorem follows.

The third term $E\big(M_T I(T > n)\big)$ in $E(M_T)$ is troublesome. There are many examples of interest where the stopping time T is not bounded.

Consider the Example 5 again. $\{T > n\}$ is the event that the first n tosses are tails and has probability 2^{-n}. If this event occurs, the bettor has lost a total $(2^n - 1)$ rupees, *i.e.*, $M_n = 1 - 2^n$. Hence

$$E\big(M_T I(T > n)\big) = 2^{-n}(1 - 2^n)$$

which does not go to 0 as $n \to \infty$.

Example 8: (Gambler's ruin problem revisited): Let X_n be a simple random walk $p = \frac{1}{2}$ on $\{0,1,2, \ldots\}$ with absorbing barriers. Suppose $X_0 = a$ and $M_n \equiv X_n$. Then, X_n is a martingale. Let stopping time $T = min\{j : X_j = 0 \text{ or } N\}$ and since X_n is bounded, we have,

$$\begin{aligned} & E(M_T) \\ & = E(M_0) \\ & = a. \end{aligned}$$

But in this case

$$\begin{aligned} & E(M_T) \\ & = 0P(X_T = 0) + NP(X_T = N) \\ & = NP(XT = N) \end{aligned}$$

Therefore,

$$P(X_T = N) = \frac{a}{N}$$

This gives another derivation of gambler's ruin result for simple random walk.

Example 9: Let X_n be as in Example 8 and $M_n = X_n^2 - n$. Then, M_n is a martingale with respect to X_n. By using Example 2

$$\begin{aligned} & E(M_{n+1}|\mathcal{F}_n) \\ & = E(X_{n+1}^2 - (n+1)|\mathcal{F}_n) \\ & = X_n^2 + 1 - (n+1) \\ & = M_n. \end{aligned}$$

Consider the stopping time $T = min\{j: X_j = 0 \text{ or } N\}$. Since M_n is not a bounded martingale so it is not immediate that (6) and (7) hold. However there exists $C < \infty$ and $\rho < 1$ such that

$$P(T > n) \leq C\rho^n.$$

Since $|M_n| \leq N^2 + n$,

$$E(|M_n|) < \infty$$

and

$$E\big(|M_n| I(T > n)\big)$$

$$\leq C\rho^n(N^2 + n) \to 0$$

Hence, optional sampling theorem holds and $E(M_T) = E(M_0) = a^2$.

$$E(M_T)$$

$$= E(X_T^2) - E(T)$$

$$= N^2 P(X_T = N) - E(T)$$

$$= aN - E(T)$$

Hence, $E(T) = aN - a^2 = a(N - a)$.

12.4 Doob's Decomposition Theorem

Doob's Decomposition Theorem is a fundamental result in the theory of stochastic processes, particularly in the study of sub martingales. It states that any sub martingale can be decomposed uniquely into the sum of a martingale and a predictable, increasing process. This decomposition is crucial for understanding the structure and behaviour of sub martingales, as it allows us to separate the "fair game" component from the trend component.

Theorem (Doob's Decomposition): Let $\{X_n\}$ be a sub martingale with respect to a filtration $\{\mathcal{F}_n\}$. Then there exists a martingale $\{M_n\}$ and a predictable, increasing process $\{A_n\}$ such that: $X_n = M_n + A_n$

where $\{A_n\}$ is adapted to $\{\mathcal{F}_n\}$, non-decreasing, and $A_0 = 0$.

Proof: Let us construct the processes $\{M_n\}$ and $\{A_n\}$ explicitly.

Constructing the Martingale M_n: Define the martingale $\{M_n\}$ as follows:

$$M_n = X_0 + \sum_{k=1}^{n} (X_k - E[X_k|\mathcal{F}_{k-1}])$$

The term $E[X_k|\mathcal{F}_{k-1}]$ is the conditional expectation of X_k given the information up to time $k-1$.

Define the predictable, increasing process $\{A_n\}$ as follows:

$$A_n = \sum_{k=1}^{n} (E[X_k|\mathcal{F}_{k-1}] - X_{k-1})$$

The process A_n is non-decreasing because each term in the summation is non negative (since $\{X_k\}$ is a sub martingale, we have $E[X_k|\mathcal{F}_{k-1}] \geq X_{k-1}$).

We need to verify that $X_n = M_n + A_n$. Substituting the definitions of M_n *and* A_n,

We get,

$$M_n + A_n$$

$$= \left(X_0 + \sum_{k=1}^{n} (X_k - E[X_k|\mathcal{F}_{k-1}]) \right) + \left(\sum_{k=1}^{n} (E[X_k|\mathcal{F}_{k-1}] - X_{k-1}) \right)$$

$$= X_0 + \sum_{k=1}^{n} (X_k - X_{k-1})$$

$$= X_n$$

This leads to

$X_n = M_n + A_n$.

Thus, we have decomposed the sub martingale $\{X_n\}$ into a martingale $\{M_n\}$ and a predictable, increasing process $\{A_n\}$. This decomposition is unique, given the initial conditions.

Doob's Decomposition Theorem is crucial in the study of stochastic processes because it allows for the separation of the "fair game" component from the trend component in a sub martingale. This separation simplifies the analysis and understanding of complex stochastic processes. The theorem is widely used in financial mathematics, particularly in the modelling and pricing of derivative securities, where understanding the martingale component is essential for arbitrage-free pricing. It also finds applications in various fields such as economics, biology, and engineering, wherever stochastic modelling is employed.

Simple Applications Based on Doob's Decomposition Theorem

Doob's Decomposition Theorem allows us to decompose a sub martingale into a martingale and a predictable, increasing process. This decomposition is very useful in various practical scenarios where it is important to separate the "fair game" component from the trend or systematic component. Here are a few simple applications of Doob's Decomposition Theorem:

Gambling Strategies: In a gambling context, suppose a player's total winnings over time can be modelled as a sub martingale (indicating that, on average, the player's winnings tend to increase, perhaps due to a favorable betting system). Doob's Decomposition Theorem allows us to separate the player's winnings into a part that represents the fair game aspect (martingale) and a part that represents the systematic gains (predictable increasing process).

Application: Let X_n be the total winnings after n rounds.

Decompose X_n into $M_n + A_n$, where M_n is the martingale (fair game component) and A_n is the predictable, increasing process (systematic gains). By understanding this decomposition, the player can evaluate how much of their winnings come from genuine luck (martingale) versus a systematic advantage (predictable process).

Stock Market Analysis: In financial markets, suppose the value of a stock or portfolio is modelled as a sub martingale (indicating a general upward trend). Doob's Decomposition Theorem helps in separating the actual investment returns (martingale) from the overall market trend or systematic growth (predictable increasing process).

Application: Let X_n represent the value of a portfolio at time n.

Decompose X_n into $M_n + A_n$, where M_n is the martingale part representing the fair returns and A_n is the predictable part representing market growth or systematic investment strategy.

This decomposition helps investors to understand how much of their portfolio's performance is due to market trends versus active management.

Insurance Risk Management: In an insurance company, the total amount of claims over time can be modelled as a sub martingale if the expected claims tend to increase due to factors like inflation or increasing risk exposure. Doob's Decomposition Theorem can be used to separate the random fluctuation of claims (martingale) from the predictable increase in claims (predictable increasing process).

Application: Let X_n be the total claims by the end of year n.

Decompose X_n into $M_n + A_n$, where M_n is the martingale representing the random component of claims and A_n is the predictable increasing process representing the systematic growth in claims. This helps the insurance company to better predict future claims and manage risk by distinguishing between random variations and systematic trends.

Project Management: In project management, the cumulative cost of a project can sometimes be modelled as a sub martingale if costs tend to increase due to unforeseen events or systematic cost overruns. Doob's Decomposition Theorem can separate the fair component of cost (martingale) from the predictable cost increases (predictable increasing process).

Application: Let X_n be the cumulative cost of the project after n months.

Decompose X_n into $M_n + A_n$, where M_n is the martingale part representing unpredictable cost changes and A_n is the predictable increasing part representing expected cost overruns.

This decomposition allows project managers to differentiate between unexpected costs and systematic budget increases, aiding in better financial planning and control.

Quality Control in Manufacturing: In a manufacturing process, the number of defective items produced over time can be modelled as a sub martingale if defects tend to increase due to machinery wear or other factors. Doob's Decomposition Theorem helps separate the random defect occurrences (martingale) from the predictable increase in defects (predictable increasing process).

Application: Let X_n be the total number of defects detected by the end of day n. Decompose X_n into $M_n + A_n$, where M_n is the martingale part representing random defects and A_n is the predictable increasing process representing systematic increases in defects.

This helps the quality control team to identify and address underlying issues causing systematic increases in defects, improving overall product quality.

These simple applications illustrate how Doob's Decomposition Theorem can be used in practical scenarios to separate random fluctuations from systematic trends, providing valuable insights for decision-making and strategy development.

12.5 Martingale Convergence Theorems

Martingale convergence theorems are fundamental results in probability theory that provide conditions under which martingales converge. These theorems have significant implications for the analysis and long-term behaviour of stochastic processes. We will discuss two primary convergence theorems: the Martingale Convergence Theorem and the L^p Convergence Theorem.

Martingale Convergence Theorem: The Martingale Convergence Theorem states that if a martingale is bounded in L^1, it converges almost surely.

Theorem: Let $\{X_n\}$ be a martingale with respect to a filtration $\{\mathcal{F}_n\}$. If $\{X_n\}$ is bounded in L^1 i.e. $\sup_n E\|X_n\| < \infty$, then there exists a random variable X such that $X_n \to X$ almost surely as $n \to \infty$.

Proof: Since, $\{X_n\}$ is bounded in L^1 it is uniformly integrable. Uniform integrability ensures that the sequence does not lose mass at infinity, which is a necessary condition for convergence.

Construct a non-negative sub martingale $\{Y_n\}\, from\, \{X_n\}$ such that $Y_n = |X_n| + \sum_{k=0}^{n-1} E[|X_{k+1} - X_k||\mathcal{F}_n|]$. The term $\sum_{k=0}^{n-1} E[|X_{k+1} - X_k||\mathcal{F}_n|]$ ensures that $\{Y_n\}$ is increasing.

By the Sub martingale Convergence Theorem, since $\{Y_n\}$ is non-negative and increasing, it converges almost surely to a limit Y.

Using the fact that $\{X_n\}$ is uniformly integrable and $\{Y_n\}$ converges, it follows that $\{X_n\}$ almost surely to a random variable X.

L^2 Martingale Convergence Theorem: The L^2 Martingale Convergence Theorem provides a convergence result under the assumption that the martingale is bounded in L^2.

Theorem: Let $\{X_n\}$ be a martingale with respect to a filtration $\{\mathcal{F}_n\}$. If $\{X_n\}$ is bounded in L^2 , i.e., $\sup_n E[X_n^2] < \infty$, then there exists a random variable X in L^2 such that $X_n \to X$ almost surely and in L^2 as $n \to \infty$.

Proof: Since $\{X_n\}$ is bounded in L^2, it is also bounded in L^1. This implies uniform integrability, which is necessary for almost sure convergence.

Show that $\{X_n\}$ is a Cauchy sequence in L^2 for $m \geq n$.

$$\mathrm{E}[(\mathrm{X_m} - \mathrm{X_n})^2]$$

$$= \mathrm{E}\left[\sum_{\mathrm{k=n}}^{\mathrm{m-1}} (\mathrm{X_{k+1}} - \mathrm{X_k})^2 \,|\, \mathcal{F}_\mathrm{n}\}.\right]$$

$$\leq \sup_\mathrm{n} \mathrm{E}[\mathrm{X_n^2}] < \infty$$

Since convergence in L^2 implies convergence in probability, and uniform integrability ensures almost sure convergence, $\{X_n\}$ converges almost surely to X.

12.6 Summary

In this unit, we explored the fundamental concepts and theorems related to martingales, an essential topic in probability theory. Martingales represent a sequence of random variables modelling fair games, where the future expected value, given the present, is equal to the current value. This property has made martingales a powerful tool in various fields, including finance, where they are used to model stock prices and derivative pricing under the assumption of no arbitrage.

We began by defining martingales, sub martingales, and super martingales, highlighting their key properties and providing simple examples such as gambling games and stock prices. The Doob's Decomposition Theorem was introduced, which allows us to decompose a sub martingale into a martingale and a predictable, increasing process. This theorem is instrumental in understanding the behaviour of stochastic processes and separating systematic trends from random fluctuations.

The unit also covered the Martingale Convergence Theorem and the L^2 Martingale Convergence Theorem. These theorems provide conditions under which martingales converge almost surely or in L^2, respectively. The convergence theorems are crucial for predicting the long-term behaviour of processes modelled by martingales, with significant implications for financial modelling, risk management, and other applications.

Examples were provided to illustrate the practical applications of these theorems, such as in gambling strategies, stock market analysis, insurance risk management, project management, and quality control in manufacturing. These examples demonstrate how martingale theory can be applied to real-world problems, providing a robust framework for analysing and managing uncertainty.

Understanding martingales and their convergence properties equips us with essential tools for modelling stochastic processes, making informed decisions, and solving complex problems in various domains. The insights gained from this unit are foundational for further studies and applications in probability theory and related fields.

12.7 Self-Assessment Questions

1. Define a martingale and provide an example from a real-world application.
2. Explain Doob's Decomposition Theorem and its significance in the theory of martingales.
3. State and prove the Martingale Convergence Theorem.
4. What is the difference between a sub martingale and a super martingale?
5. Provide a simple application of Doob's Decomposition Theorem.
6. Describe the L^2 Martingale Convergence Theorem and its implications.
7. Give an example of a stochastic process that can be modeled as a martingale.
8. Discuss how martingales are used in financial modeling.

9. How does uniform integrability relate to martingale convergence?
10. Explain the importance of martingale theory in the context of risk management.

12.8 References

- Williams, D. (1991). Probability with Martingales. Cambridge University Press.
- Karatzas, I., & Shreve, S. E. (1991). Brownian Motion and Stochastic Calculus. Springer.
- Doob, J. L. (1953). Stochastic Processes. Wiley.
- Shiryaev, A. N. (1996). Probability. Springer.
- Feller, W. (1971). An Introduction to Probability Theory and Its Applications, Vol. 2. Wiley.

12.9 Further Reading

- Revuz, D., & Yor, M. (1999). Continuous Martingales and Brownian Motion. Springer.
- Elliott, R. J., & Kopp, P. E. (1999). Mathematics of Financial Markets. Springer.
- Protter, P. E. (2005). Stochastic Integration and Differential Equations. Springer.
- Ethier, S. N., & Kurtz, T. G. (1986). Markov Processes: Characterization and Convergence. Wiley.
- Steele, J. M. (2001). Stochastic Calculus and Financial Applications. Springer.

U.P. Rajarshi Tandon Open University, Prayagraj

MScSTAT – 203(N)/ MASTAT – 203(N) Stochastic Process

Block: 4 Applied Stochastic Process

Course Design Committee

Course Preparation Committee

First Edition: ***July 2024***
ISBN : 978-93-48270-94-8

MScSTAT – 203N/ MASTAT – 203N
STOCHASTIC PROCESS

Printed and Published by Col. Vinay Kumar, Registrar, Uttar Pradesh Rajarshi Tandon Open University Prayagraj, 2024
Printed by: Cygnus Information Solution Pvt.Ltd., Lodha Supremus Saki Vihar Road Andheri East, Mumbai.

Block & Units Introduction

The ***Block – 4 - Applied Stochastic Process*** is the last fourth block of said SLM, which is divided into three units.

The ***Unit – 13 - Homogeneous Process*** deals with the random variable technique, homogeneous birth and death process, divergent birth process, the general birth and death process, multiplicative process, effect of immigration for homogeneous process.

The ***Unit – 14 - Non-Homogeneous Process*** is discussed about Simple non homogeneous process, Polya process, effect of immigration for non-homogeneous process, Diffusion, Backward Kolmogorov diffusion equation, Fokker-Planck equation.

The unit of this SLM is ***Unit – 15 – Non-Markovian Process*** is discussed about Some multi-dimensional prey and predator, Non-Markovian Process, Embedded Markov Process, Application to population growth, epidemic and counter models.

At the end of every block/unit the summary, self-assessment questions are given.

UNIT -13: HOMOGENEOUS PROCESS

Structure

13.1 Introduction

In the realm of stochastic processes, the concept of homogeneity plays a crucial role in simplifying and understanding complex systems that evolve over time. Unit 13, titled "Homogeneous Process," is dedicated to exploring this pivotal concept and its applications in various fields. This unit will delve into the mathematical and theoretical aspects of homogeneous stochastic processes, providing a solid foundation for understanding how these processes are characterized and analysed.

A homogeneous process, by definition, is a type of stochastic process where the probability distribution governing the process is time-invariant. This means that the statistical properties of the process do not change over time, making it easier to model and predict. Such processes are ubiquitous, finding applications in diverse fields such as physics, biology, finance, and engineering, where they help in modelling everything from particle movements in a fluid to population dynamics and financial market fluctuations.

The unit begins by laying a foundation with an overview of homogeneous processes, followed by specific objectives aimed at building a deep understanding of the topic. We will explore the fundamental equations governing these processes, namely the forward and backward equations, which are essential for modelling and analysing time-evolution in these systems.

A significant portion of the unit is devoted to understanding various types of homogeneous processes, such as birth and death processes, which are models for systems where events occur at random times. We will also delve into the divergent birth process, a complex variant where the birth rate diverges, and the general birth and death process, which provides a more comprehensive model.

The unit will also cover the multiplicative process, another crucial type of homogeneous process, and the impact of external factors like immigration on these systems. This is particularly relevant in real-world scenarios where external influences play a significant role in the dynamics of the system.

As we progress, the unit will employ the random variable technique, a powerful tool in the analysis of stochastic processes. This approach helps in simplifying complex processes into manageable mathematical models, enabling us to derive meaningful conclusions and predictions.

Towards the end, the unit includes self-assessment questions to test your understanding and comprehension of the concepts. We also provide an extensive list of references for further reading and exploration, and a summary to reinforce the key points covered.

By the end of this unit, learners will have gained a comprehensive understanding of homogeneous processes, equipped with the knowledge to apply these concepts to real-world scenarios and further academic study. Whether you are a student, a researcher, or a professional, this unit will provide valuable insights into the fascinating world of stochastic processes.

13.2 Objectives

Objective 1: Understanding the Basic Concepts of Homogeneous Processes

***Defining Homogeneous Processes*:** Clarifying the definition of a homogeneous process in the context of stochastic processes, emphasizing its time-invariance characteristic.

***Understanding Time-Invariance*:** Exploring the concept of time invariance in more depth, explaining how the statistical properties of the process remain constant over time.

***Differentiating Homogeneous from Non-Homogeneous Processes*:** Highlighting the differences between homogeneous and non-homogeneous processes, with examples to illustrate these distinctions.

***Basic Mathematical Formulation*:** Introducing the fundamental mathematical formulations used in modelling homogeneous processes, including probability distributions and expected values.

Objective 2: Learning About Various Types of Homogeneous Processes and Their Characteristics

***Birth and Death Processes*:** Delving into one of the most common types of homogeneous processes. Discussing how these processes model systems where events, such as births and deaths, occur randomly over time.

***Divergent Birth Processes*:** Exploring a more complex variant of birth processes where the birth rate can diverge, analysing its characteristics and implications.

General Birth and Death Processes: Expanding the concept of birth and death processes to more general and diverse scenarios, including variable rates and multiple states.

Multiplicative Processes: Introducing and defining multiplicative processes, common in finance and economics, and discussing their unique characteristics and applications.

***Homogeneous Processes with Immigration*:** Investigating how external factors like immigration affect homogeneous processes, and understanding the resulting changes in the system's dynamics.

***Real-World Applications*:** Providing examples from various fields such as biology, finance, physics, and engineering to illustrate the practical applications and relevance of homogeneous processes.

Objective 3: Supporting Learning Objectives

***Developing Analytical Skills*:** Enhancing the ability to analyse and interpret the behaviour of homogeneous processes through mathematical and statistical methods.

***Problem-Solving and Modelling*:** Fostering skills in developing and solving models of homogeneous processes relevant to real-world scenarios.

***Critical Thinking and Application*:** Encouraging learners to apply theoretical knowledge to practical situations, enhancing critical thinking skills.

These objectives are designed to provide a comprehensive understanding of homogeneous processes, ensuring that learners not only grasp the theoretical aspects but also develop the ability to apply this knowledge practically. By the end of this section, students should be able to recognize, model, and analyse various types of homogeneous processes, and appreciate their significance in both academic studies and real-world applications.

13.3 Forward and Backward Equations for Homogeneous Case

This section of the unit deals with the mathematical backbone of homogeneous stochastic processes, focusing on the derivation and understanding of both forward and backward equations. These equations are pivotal in describing the evolution of probabilities in a stochastic process over time.

Forward Equations in Homogeneous Processes

Concept and Relevance: The forward equations, also known as the Kolmogorov forward equations, are fundamental in the study of homogeneous processes. They describe how the probability distribution of a process evolves forward in time.

Derivation of Forward Equations: Deriving the forward equations, also known as the Kolmogorov forward equations, for a continuous-time stochastic process involves several steps. These equations describe how the probability distribution of the process evolves over time. Let's focus on a continuous-time Markov chain as an example to derive these equations.

Preliminaries

Markov Property: A process X(t) is Markovian if the future state depends only on the current state, not on the past states.

***State Space*:** Assume the process has a countable state space S.

Transition Probabilities: Define $P_{ii}(t)$=P(X(t+s) = j |X(s)=i), the probability that the process moves from state i to state j in time t.

Chapman-Kolmogorov Equation

The Chapman-Kolmogorov equation relates the probabilities over different time intervals:

$$P_{ij}(t+u) = \sum_{k \in S} P_{ik}(t) P_{kj}(t)$$

This equation is a consequence of the Markov property, integrating over all possible intermediate states k.

Derivation of Forward Equations

1. Infinitesimal Transition Probabilities: Consider the probabilities of transitions over a very small-time interval

   Δt: $P_{ik}(\Delta t) = 1 - \lambda_i$

   Δt+o(Δt), where λi is the rate of leaving state i, and o(Δt) denotes higher-order small terms.

$P_{ij}(\Delta t)$=$q_{ij}(\Delta t)$+o(Δt) for i not equal to j,

where q_{ij} is the rate of transitioning from i to j.

2. Applying Chapman-Kolmogorov Equation

$$P_{ij}(\Delta t) = \sum_{k \in S} P_{ik}(\Delta t) P_{kj}(\Delta t)$$

Expand and rearrange terms to get

$$P_{ij}(t+\Delta t) - P_{ij}(t) = \sum_{k \neq i} P_{ik}(t) q_{kj}(\Delta t) - \lambda_i P_{ij}(t) \Delta t + o(\Delta t)$$

3. Taking the limit: To form the differential equation, take the limit as $\Delta t \to 0$

$$\log_{\Delta t \to 0} \frac{P_{ij}(t+\Delta t) - P_{ij}(t)}{\Delta t} = \log_{\Delta t \to 0} \left(\sum_{k \neq i} P_{ik}(t) q_{kj} - \lambda_i P_{ij}(t) + \frac{o(\Delta t)}{\Delta t} \right)$$

4. Kolmogorov Forward Equation: Simplifying, we get the forward equation

$$\frac{dP_{ij}(t)}{dt} = \sum_{k \neq i} P_{ik}(t) q_{kj} - \lambda_i P_{ij}(t)$$

The Kolmogorov forward equation describes how the probability of transitioning from state i to state j changes over time. The term $\sum_{k \neq i} P_{ik}(t) q_{kj}$ represents the rate of entering state j from all other states, and $\lambda_i P_{ij}(t)$ represents the rate of leaving state j.

This derivation assumes a basic understanding of calculus and probability theory, and it provides the foundation for analysing continuous-time Markov chains and other types of stochastic processes.

Examples and Application: Illustrate the application of forward equations with practical examples, such as simple birth and death processes or Markov chains.

Backward Equations: Concept and Mathematical Formulation

Introduction to Backward Equations: While forward equations provide a future view of probabilities, backward equations, or Kolmogorov backward equations, offer a way to look backward in time. They are particularly useful in scenarios where conditioning on the present state is essential.

Mathematical Formulation of Backward Equations:

- Describe the process of conditioning on the current state and its implications for probability calculations.
- Present the mathematical derivation of backward equations, starting with the basic principles and leading to the formal expression.

- Explain the relationship between forward and backward equations and how they complement each other in the analysis of stochastic processes.

***Real-World Examples*:** Use real-world scenarios to demonstrate how backward equations are applied in practice, such as in financial modelling or in computing the probability of certain events in a time-reversed manner.

Comparative Analysis

***Comparing Forward and Backward Equations*:** Discuss the differences in application and interpretation between the two types of equations, providing a comprehensive understanding of when and why each type is used.

***Practical Implications*:** Emphasize the practical implications of understanding both types of equations in modelling and predicting the behaviour of homogeneous stochastic processes.

This section is designed to equip learners with a thorough understanding of both forward and backward equations in homogeneous processes. By understanding these fundamental tools, students can better analyse and model various stochastic processes, an essential skill in many scientific and engineering fields.

EXAMPLES

***Example 1*: Simple Birth Process:** Imagine a population where individuals reproduce at a constant rate, and there are no deaths. Let us denote λ as the constant birth rate.

Forward Equation: The forward equation for this process is derived from the Kolmogorov forward equations. If $P_n(t)$ is the probability of having n individuals at time t, the forward equation is:

$$\frac{dP_n(t)}{dt} = \lambda P_{n-1}(t)) - \lambda P_n(t)$$

This equation reflects that the rate of change in the probability of having n individuals is influenced by the birth rate from the previous state n−1 and the rate leaving the current state n.

Backward Equation: The backward equation, however, is not typically used in this scenario as there are no transitions backward (i.e., no deaths).

***Example 2*: Queueing System with Arrivals and Services:** Consider a queue where customers arrive at a rate of λ and are served at a rate of μ. This is a birth-death process where arrivals are 'births' and services are 'deaths'.

Forward Equation: The forward equation in this scenario is:

$$\frac{dP_n(t)}{dt} = \lambda P_{n-1}(t) - (\lambda+\mu)\, P_n(t) + \mu P_{n+1}(t)$$

This equation represents the balance between the rate of customers arriving and being served. It incorporates the probabilities of moving to state n from n−1 (arrival) and moving away from state n to n+1 (service) or n−1 (next customer served).

Backward Equation: The backward equation for this system would provide insights into how the probability of being in a certain state at a specific time depends on the initial state of the system. However, in most practical queueing models, the forward equation is more commonly used for analysis.

These examples demonstrate the practical application of forward equations in different homogeneous processes. Forward equations are especially useful in scenarios where understanding the evolution of a system over time is crucial, such as in population dynamics or queueing theory. The backward equations, while less commonly used in these specific examples, are invaluable in other contexts where understanding how past states influence current probabilities is necessary.

13.4 Random Variable Technique

This section covers the application of random variable techniques in the study of homogeneous processes. Random variables are fundamental in probabilistic modelling and provide a powerful tool for describing and analysing the behaviour of stochastic systems.

Introduction to Random Variables in Homogeneous Processes

***Definition of Random Variables*:** A random variable is a function that assigns a real number to each outcome in a sample space of a random experiment. In the context of stochastic processes, random variables are used to represent the state of the process at different points in time.

Homogeneous Processes: In a homogeneous process, the statistical properties (like the mean, variance, etc.) of the process do not change over time. This time-invariance property simplifies the use of random variables in analysis and prediction.

Role of Random Variables: Random variables are used to model various aspects of homogeneous processes, such as waiting times, the number of occurrences of an event, or the state of a system at a particular time.

Probability Distributions: Discussing how different probability distributions (like Poisson, exponential, or normal distributions) can be used to model different types of homogeneous processes.

Examples of Random Variable Techniques in Real-World Scenarios

Queueing Theory:

Example: Modelling customer arrival and service times in a queue.

Application: Using exponential random variables to model the time between arrivals (inter-arrival times) and service times in a queueing system, which is often a Poisson process.

Reliability Engineering:

Example: Assessing the reliability of a system over time.

Application: Employing exponential random variables to model the lifetime of components in a system. This can help in predicting the time until failure of a component or the entire system.

***Population Dynamics*:**

Example: Modelling the growth of a biological population.

Application: Using birth and death processes (which are specific types of Markov processes) to model population dynamics. Here, random variables can represent the number of individuals in a population at a given time.

Finance and Risk Management:

Example: Modelling stock prices or interest rates over time.

Application: Implementing geometric Brownian motion, a continuous-time stochastic process, where the logarithm of the price follows a Brownian motion (or Wiener process). This involves using random variables to represent the price of a stock at different times.

Epidemiology:

Example: Spread of infectious diseases.

Application: Utilizing random variables to model the number of infected individuals over time in a population. The spread of disease can often be modelled as a branching process, which is a type of Markov process.

In each of these examples, the random variable technique simplifies complex real-world phenomena into a mathematical framework that can be analysed and used for prediction. This section aims to provide learners with a clear understanding of how random variables are applied in various fields to model and analyse homogeneous processes.

13.5 Homogeneous Birth and Death Process

This section focuses on the homogeneous birth and death process, a fundamental concept in stochastic processes. We shall explore the definition, characteristics, and practical applications of these processes.

Definition and Characteristics of Birth and Death Processes

Birth and Death Processes Defined: In stochastic modelling, a birth and death process is a type of continuous-time Markov chain that models two types of transitions: "births", which increase the state by one, and "deaths", which decrease the state by one.

Homogeneity in Birth and Death Processes: In homogeneous birth and death processes, the rates of births and deaths are constant over time, making the process time-invariant. This means the probability of a birth or death occurring in a small interval of time is proportional to the length of the interval, independent of when the interval occurs.

State Space: Typically, the state space for these processes is the set of non-negative integers, where each state represents the number of entities (such as individuals in a population or customers in a queue).

Analysis of Homogeneous Birth and Death Processes

Transition Rates: The transition rates in a homogeneous birth and death process are denoted as λ_n for births (transition from state n to n+1) and μ_n for deaths (transition from state n to (n−1).

Differential Equations: The Kolmogorov forward and backward equations can be applied to derive the differential equations governing the probabilities of being in each state over time.

Steady-State Analysis: In some cases, these processes reach a steady state where the state probabilities remain constant over time. Calculating the steady-state probabilities involves solving a system of linear equations derived from the transition rates.

In other words, it can write as

State Space: Let us denote the state of the process at time t by X(t), where $X(t) \in \{0,1,2,3,...\}$. Each state represents the number of entities (like individuals in a population).

Transition Rates: Birth Rates (λ_n): The rate at which the process transitions from state n to n+1. It's the rate of adding an entity.

Death Rates (μ_n): The rate at which the process transitions from state n to n−1. It's the rate of losing an entity.

Homogeneity: In a homogeneous process, these rates (λ_n and μ_n) are constant over time, though they may depend on the state n.

Kolmogorov Forward Equations: For the probability P_n (t)= $P(X(t) = n)$ of being in state n at time t, the forward equations are given by:

$$\frac{dP_n\,(\mathrm{t})}{dt} = \lambda_{n-1}P_{n-1} - (\lambda_n + \mu_n)P_n + \mu_{n+1}P_{n+1}$$

This equation says that the rate of change of the probability of being in state n is the sum of:

- The rate of entering state n from n−1 (births).
- The rate of leaving state n to either n+1 (births) or n−1 (deaths).

Initial Conditions: Typically, $P_n\,(0)$ is given for all n, which represents the probability distribution of the states at time t=0.

Steady-State Analysis: If a steady-state exists, the probabilities P_n become time-independent, and the system of equations simplifies to a balance equation:

$$\lambda_{n-1}P_{n-1} = (\lambda_n + \mu_n)P_n - \mu_{n+1}P_{n+1}$$

This set of equations can be solved to find the steady-state probabilities P_n .

Example Application: consider a queue where customers arrive at a rate λ and are served at a rate μ. If there is only one server, then $\lambda_n = \lambda \; and \; \mu_n = \mu \; for \; all \; n.$

the forward equations for this system becomes:

$$\frac{dP_n\,(\mathrm{t})}{dt} = \lambda_{n-1}P_{n-1} - (\lambda_n + \mu_n)P_n + \mu_{n+1}P_{n+1}$$

By solving these equations with appropriate initial conditions, one can determine the probability of having n customers in the queue at any time t.

This mathematical formulation provides a powerful tool for analysing systems where events occur randomly but with constant rates, as in the case of homogeneous birth and death processes.

Examples of Homogeneous Birth and Death Processes

Population Dynamics: Modelling animal populations where individuals are born and die at constant rates. For example, a simple model might assume everyone in the population has the same birth and death rate, irrespective of the population size.

Queueing Systems: In a customer service scenario, new customers arrive (births) at a service centre, and customers are served and leave the queue (deaths). If the arrival and service rates are constant, this can be modelled as a homogeneous birth and death process.

Epidemiology: Modelling the spread of a disease, where an infection is considered a birth and recovery or death as a death. This application becomes more complex if the rates of infection and recovery are influenced by external factors or if they vary over time.

Homogeneous birth and death processes are widely used in various fields to model systems where entities are born and die at constant rates. Here are a couple of examples to illustrate the application of homogeneous birth and death processes:

Example 1: Population Dynamics

Consider a simple ecological model where a species in an isolated environment reproduces and dies at constant rates. Let us denote λ as the constant birth rate per individual and μ as the constant death rate per individual.

Modelling: In this model, the rate of change in the population depends solely on the current population size.

If n represents the number of individuals in the population at a given time, the transition rate from state n to n+1 (birth) is $n\lambda$, and from n to n−1 (death) is $n\mu$.

The system can be described by a set of differential equations based on these rates, which will govern the probability of the population being at a certain size over time.

Example 2: Queueing Theory: Imagine a customer service centre where customers arrive randomly to be served by clerks. Assume customers arrive at a rate λ (a 'birth') and are served at a rate μ (a 'death').

Modelling: Here, the 'state' of the system can be represented by the number of customers in the queue.

The arrival of a new customer increases the number in the queue, while the completion of service decreases it.

The transition rates are λ for arrivals (regardless of the current state, as each arrival is independent) and $n\mu$ for services (assuming each clerk serves at rate μ and there are n clerks, each serving one customer).

This queue can be modelled as a birth and death process, and the steady-state probabilities (if they exist) can tell us useful information like the average number of people in the queue or the average waiting time.

These examples showcase how homogeneous birth and death processes are used to model real-world systems where the events (births and deaths) occur randomly but at constant rates. In the population dynamics model, it helps in understanding species growth under ideal conditions, while in queueing theory, it aids in optimizing customer service operations.

13.6 Divergent Birth Process

The concept of a divergent birth process is an intriguing area within stochastic processes, particularly in the study of birth processes where the birth rate is not constant and can increase without bound under certain conditions.

Explanation of Divergent Birth Processes

Basic Concept: Unlike a traditional birth process where the birth rate remains constant or changes in a controlled manner, in a divergent birth process, the birth rate increases as the number of entities in the system increases, and it can potentially grow to infinity.

Characteristics:

Self-Accelerating: The process is self-accelerating; the more entities present, the higher the birth rate.

Non-Linear Growth: The growth of the number of entities in the system can be non-linear and may exhibit exponential or even faster growth.

Unbounded Growth: There is no upper limit to the growth rate, leading to scenarios where the population size can grow indefinitely in a finite time.

Mathematical Modelling

Rate Equations: Let X(t) denote the number of entities at time t. The birth rate, denoted as λ(X(t)), is a function of X(t) itself.

Differential Equations: The change in the number of entities can be modelled by the differential equation:

$$\frac{dX(t)}{dt}=\lambda(X(t))\cdot X(t)$$

Here, λ(X(t)) increases with X(t), indicating a divergent birth rate.

Growth Models: Various models can be used for λ(X(t)), such as:

Linear Growth: $\lambda(X(t)) = a\cdot X(t)+b$, where $a,b>0$.

Exponential Growth: $\lambda(X(t)) = \exp(a.X(t))$, where $a>0$.

Implications in Various Fields

Biology: In ecological models, divergent birth processes can represent scenarios of uncontrolled population growth under certain environmental conditions, leading to phenomena like algal blooms or invasive species outbreaks.

Epidemiology: In disease modelling, especially for highly contagious diseases, the rate of new infections can accelerate as more individuals become infected, resembling a divergent birth process.

Economics: In financial markets, certain asset bubbles can be modelled as divergent birth processes, where the rate of investment or asset value increase accelerates as more capital enters the market.

Physics: In nuclear chain reactions, the rate of neutron production can increase rapidly, leading to a divergent process under uncontrolled conditions.

Mathematical Implications: Divergent birth processes challenge traditional modelling approaches due to their potential for infinite growth in finite time, requiring careful mathematical treatment and consideration of boundary conditions and constraints.

Divergent birth processes represent a significant departure from traditional homogeneous processes, introducing complexity and challenges in both modelling and interpretation. Understanding these processes is crucial in fields where exponential or faster-than-exponential growth patterns are observed, necessitating careful analysis and control mechanisms.

Divergent birth processes are characterized by a birth rate that increases rapidly, often exponentially, with the number of entities in the system. Let's explore a couple of examples to illustrate this concept:

Example 1: **Epidemiological Model of a Highly Contagious Disease**

Imagine a scenario where a highly contagious disease spreads through a population. Initially, the number of infected individuals is low, but as they interact with others, the rate of new infections increases rapidly.

Modelling: In this model, the "birth" rate, representing new infections, increases with the number of currently infected individuals.

If I(t) represent the number of infected individuals at time t, the rate of new infections might be proportional to $I(t)^2$, assuming each infected individual has a chance of infecting any other individual.

The differential equation representing this divergent birth process could be something like:

$\frac{dI(t)}{dt} = I(t)^2$ where α is a constant representing the infection rate.

Such a model leads to exponential or even faster growth in the number of infected individuals, characteristic of a divergent birth process.

Example 2: **Chain Reactions in Nuclear Physics**

Consider a nuclear chain reaction, such as in a nuclear reactor or an atomic bomb, where each fission event causes the release of additional neutrons that can induce further fissions.

Modelling: In this process, the rate of fission events (the "birth" rate of new fissions) increases with the number of free neutrons available.

If N(t) represents the number of free neutrons at time t, the rate of increase in neutrons can be proportional to the current number of neutrons.

The differential equation might look like, $\frac{dN(t)}{dt} = \beta N(t)$, where β is a constant representing the probability of a neutron inducing another fission.

As the reaction proceeds, it can lead to a rapid, exponential increase in the number of neutrons, characteristic of a divergent process.

These examples demonstrate the concept of divergent birth processes in very different contexts: epidemiology and nuclear physics. In both cases, the key feature is that the rate of the "birth" event (new infections or fissions) increases rapidly with the number of entities already present, leading to exponential or super-exponential growth.

13.7 General Birth and Death Process

The general birth and death process is a more comprehensive model in the study of stochastic processes, encompassing a wider range of scenarios than the simpler homogeneous birth and death processes.

Definition and Characteristics:

Basic Concept: A general birth and death process is a type of continuous-time Markov chain where the transition rates (both birth and death) may vary depending on the current state of the system.

State Space: Like the homogeneous case, the state space is typically the set of non-negative integers, where each integer represents the number of entities in the system.

Transition Rates:

Birth Rates (λ_n): The rate of transition from state n to n+1 (an addition to the population). Unlike the homogeneous case, λ_n can vary with n.

Death Rates (μ_n): The rate of transition from state n to n−1 (a subtraction from the population). Similarly, μ_n can also vary with n.

Mathematical Modelling:

Kolmogorov Forward and Backward Equations: The forward and backward equations still form the basis of analysis, but they now incorporate the state-dependent rates:

- Forward Equation:

$$\frac{dP_n\,(\mathrm{t})}{dt} = \lambda_{n-1}P_{n-1}(t) - (\lambda_n + \mu_n)P_n\,(t) + \mu_{n+1}P_{n+1}\,(t)$$

- Backward Equation:

$$\frac{dP_n\,(\mathrm{t})}{dt} = \lambda_n P_n(t-s) - (\lambda_{n+1} + \mu_{n+1})P_n\,(t-s) + \mu_n P_n\,(t-s)$$

Probability Distribution Over Time: The solution to these equations gives the probability distribution P_n (t), the probability of being in state n at time t.

Implications and Applications

Biology and Ecology: Modelling populations with varying birth and death rates due to factors like food availability, predation, or disease.

Queueing Theory: Handling systems where the arrival and service rates change over time, such as in a call centre where arrival rates may vary by hour.

Healthcare and Medicine: Studying the spread of diseases with varying infection and recovery rates, or modelling patient flow in hospitals.

Finance and Insurance: Analysing risk processes where the rate of claims or defaults can change over time depending on various economic factors.

Engineering: Modelling systems with components that have failure rates depending on their age or usage, such as in reliability engineering.

The general birth and death process provides a more realistic model for many real-world scenarios where the assumption of constant rates (as in homogeneous processes) is not valid. This complexity allows for a richer and more accurate representation of systems, but it also requires more sophisticated mathematical and computational tools for analysis and prediction.

13.8 Multiplicative Process

Definition and Significance in Stochastic Modelling

Basic Definition: A multiplicative process is a type of stochastic process where the change in the process's value is proportional to its current value. Mathematically, it can be expressed as $X_{t+1} = X_t(1 + \varepsilon_t)$, where X_t is the value at time t and ε_t is a random variable representing the proportional change.

Characteristics:

Proportional Growth/Decay: Unlike additive processes where increments are constant or independent of the state, in a multiplicative process, the increments are proportional to the current state.

Log-Normal Distribution: Often, the values in a multiplicative process follow a log-normal distribution, especially when the proportional changes are small and occur frequently.

Path Dependency: The future value of the process depends on its history, making it path-dependent.

Significance: Multiplicative processes are crucial in modelling phenomena where growth or decay is proportional to the current size or value, a common occurrence in many natural and social systems.

Applications and Examples

Economic Growth: In macroeconomics, the growth of an economy or a company's revenue over time can often be modelled as a multiplicative process, where growth in one period builds upon the previous.

Population Dynamics: The growth of biological populations under certain conditions can be modelled as a multiplicative process, particularly when resources are abundant, and growth rate is proportional to the current population.

Finance and Investments: Stock prices and investment portfolios are classic examples. The return in each period is often a percentage (positive or negative) of the current value, leading to a multiplicative dynamic.

Compound interest is another example where the interest added to an account is proportional to the current account balance.

Physics and Chemistry: Certain phenomena in physics and chemistry, like chain reactions in nuclear physics or the growth of crystal structures, can be modelled using multiplicative processes.

Environmental Science: The spread of pollutants or the growth of certain types of algae in an ecosystem can follow a multiplicative pattern, where the rate of increase is dependent on the current concentration.

Mathematical Modelling: Multiplicative processes are often modelled using stochastic differential equations in continuous time. In discrete time, the process can be modelled using a geometric Brownian motion, especially in financial applications.

The mathematical treatment typically involves understanding the log-transformed version of the process, which often simplifies the analysis and helps in deriving meaningful insights and predictions.

Understanding multiplicative processes is crucial in fields where proportional change is a fundamental characteristic of the system's dynamics. These processes provide a more accurate and realistic modelling framework compared to additive processes in such scenarios.

Supporting Results/ Theorems: When discussing multiplicative processes in stochastic modelling, there are several key results and important theorems that provide a foundational understanding and support the analysis of these processes. Here are some of the most significant ones:

1. Law of Large Numbers (LLN) for Multiplicative Processes

Statement: The LLN in the context of multiplicative processes states that, under certain conditions, the average of the logarithms of the process converges to the expected value of the logarithm of the process as the number of observations goes to infinity.

Significance: This theorem is crucial in understanding the long-term behaviour of multiplicative processes, particularly in ensuring that predictions based on these processes are robust over time.

2. Central Limit Theorem (CLT) for Log-Transformed Variables

Statement: For a multiplicative process, when considering the logarithm of the process values, the CLT states that the distribution of the sum (or average) of these log-transformed values approaches a normal distribution as the number of observations increases.

Significance: This result is essential for statistical modelling and hypothesis testing in scenarios where multiplicative processes are involved, such as in finance or population dynamics.

3. Geometric Brownian Motion (GBM)

Statement: GBM is a continuous-time stochastic process in which the logarithm of the variable follows a Brownian motion (or Wiener process). It is defined by the stochastic differential equation

$dX_t = \mu X_t dt + \sigma X_t dW_t$, where W_t is a Wiener process.

Significance: GBM is a fundamental model in financial mathematics, especially for modelling stock prices, as it incorporates both the drift and the volatility of the process in a multiplicative fashion.

4. Gibrat's Law of Proportional Growth

Statement: This law states that the growth rate of a variable is independent of its size, which is a key characteristic of multiplicative processes.

Significance: Gibrat's Law is particularly important in economics and business studies, where it is used to model the growth of companies or economies.

5. Exponential Growth and Decay

Statement: In a purely multiplicative process, the size of the variable grows (or decays) exponentially, which can be represented by equations like $X(t) = X(0)e^{rt}$, where r is the growth rate.

Significance: Understanding exponential growth and decay is crucial in fields like biology (population dynamics), physics (radioactive decay), and finance (compound interest).

6. Martingale Property

Statement: In some special cases, a transformed multiplicative process can exhibit a martingale property, meaning that its expected future value, given all past information, is equal to its current value.

Significance: The martingale property is a powerful tool in probability theory and financial mathematics, often used in the pricing of derivatives and in risk management.

These theorems and results are cornerstones in the analysis of multiplicative processes, providing a mathematical and probabilistic foundation for understanding and predicting the behaviour of such processes in various fields.

13.9 Effect of Immigration for Homogeneous Process

In stochastic modelling, the concept of immigration introduces an additional layer of complexity and realism into the analysis of homogeneous processes. This section explores how the incorporation of immigration affects the dynamics of these processes.

Understanding the Role of Immigration

Definition of Immigration in Stochastic Processes: Immigration refers to the external introduction of new entities (individuals, particles, etc.) into the system at a certain rate. This is distinct from 'births' within the system.

Homogeneous Process with Immigration: In a homogeneous process, the system's internal dynamics (like birth or death rates) are time-independent. When immigration is introduced, it adds an external, time-independent rate of new entities entering the system.

Mathematical Modelling

Modified Rate Equations: If X(t) represents the state of the system at time t, and β is the constant immigration rate, the evolution of X(t) can be modelled by the differential equation: dX(t) = [Internal Dynamics] +β

Here, "Internal Dynamics" could represent birth-death processes, with the addition of the constant term β to account for immigration.

Impact on Probability Distributions: The introduction of a constant immigration rate changes the probability distribution of the states over time. This can lead to a shift in the mean, variance, and other moments of the distribution.

Impact and Implications

Population Dynamics: In biological models, immigration can significantly affect population stability and growth patterns. It can prevent extinction in models where the population might otherwise die out.

Queueing Systems: In queueing theory, modelling customer arrivals with an additional immigration term can represent scenarios where customers from external sources join the queue, affecting the total workload and waiting times.

Epidemiological Models: In disease spread modelling, immigration can represent the introduction of infected individuals from outside regions, which can alter the dynamics of disease transmission and control measures.

Economic Models: In macroeconomic models, immigration can influence labor market dynamics, population growth, and overall economic performance.

Challenges in Analysis: The addition of immigration introduces new challenges in the analysis and prediction of homogeneous processes. It requires the re-evaluation of existing models and the development of new methods to accurately capture the impact of external entries into the system.

Incorporating immigration into the analysis of homogeneous processes is crucial for a more realistic and comprehensive understanding of various phenomena in fields such as ecology,

economics, and social sciences. It highlights the importance of external factors and their significant roles in shaping the dynamics of systems modelled by stochastic processes.

13.10 Summary

This unit delves into the intricate world of Homogeneous Processes within the realm of stochastic modelling, offering a comprehensive exploration of their theoretical underpinnings and practical applications. The unit begins by defining homogeneous processes as time-invariant stochastic models and underscores their significance in various fields like biology, finance, and physics.

A pivotal part of the unit is the detailed examination of the forward and backward equations - essential tools for understanding how probabilities evolve in these processes. The forward equations, known as the Kolmogorov forward equations, track the probability distribution's evolution over time. In contrast, the backward equations offer a retrospective view, essential for understanding past states' influences.

The unit then navigates through the concept of random variables in homogeneous processes, highlighting their role in simplifying complex stochastic models into manageable mathematical forms. This segment underscores the practicality of random variable techniques through real-world examples across various domains.

Birth and death processes, a cornerstone of stochastic modelling, are dissected to showcase how they model systems where events occur at random times. This is expanded upon in the discussion of divergent birth processes, where the birth rate increases exponentially, and general birth and death processes, which accommodate more variable and realistic scenarios.

The concept of multiplicative processes is introduced, demonstrating their importance in modelling phenomena where growth or decay is proportional to the current state. This is particularly relevant in financial modelling, where it captures the essence of market dynamics.

A unique aspect of homogeneous processes covered in this unit is the effect of immigration - the introduction of new entities into the system from an external source. This addition alters the dynamics of these processes, as illustrated through examples in ecology and economics.

The unit concludes with a set of self-assessment questions designed to test the learner's understanding of the concepts covered, ranging from theoretical aspects to practical problem-solving scenarios. This comprehensive approach ensures a deep and nuanced understanding of

homogeneous processes, preparing learners for advanced studies or professional applications in fields that rely on stochastic modelling.

13.11 Self-Assessment Questions

1. What are the key characteristics that differentiate a homogeneous process from a non-homogeneous process in stochastic modelling?
2. Explain the significance of forward and backward equations in the analysis of homogeneous processes. How do they differ in terms of their application?
3. Describe the role of random variables in modelling homogeneous processes. Why are they crucial in stochastic analysis?
4. What distinguishes a birth and death process from other types of stochastic processes? Give an example of a real-world scenario that can be modelled as a birth and death process.
5. Discuss the potential implications of a divergent birth process in a biological system. How does it differ from a standard birth process?
6. How do general birth and death processes extend the concept of homogeneous birth and death processes? What kind of real-world phenomena might require the use of a general birth and death model?
7. Explain how a multiplicative process is used in financial modelling. What key aspect of financial markets does it capture?
8. Describe the impact of immigration on the dynamics of a homogeneous process. Provide an example from either ecology or economics.
9. Given a simple homogeneous birth and death process with birth rate λ and death rate μ, calculate the steady-state probabilities for the first three states (0, 1, and 2).
10. Suppose a stock price follows a multiplicative process where the daily return is normally distributed with a mean of 0.5% and a standard deviation of 1%. What is the probability that the stock price will increase by at least 10% over 20 trading days?
11. A queueing system with a single server has an arrival rate of 2 customers per hour and a service rate of 3 customers per hour. If there is an additional external arrival (immigration) of 1 customer every 2 hours, calculate the average number of customers in the system.

12. Consider a divergent birth process where the birth rate λ_n is proportional to the square of the current population size n. Write down the differential equation governing the population size over time and discuss its implications.

13.12 References

- Ross, S. M. (2014). Introduction to Probability Models (11th ed.). Elsevier.
- Karlin, S., & Taylor, H. M. (1981). A Second Course in Stochastic Processes. Academic Press.
- Cox, D. R., & Miller, H. D. (1965). The Theory of Stochastic Processes. Chapman and Hall/CRC.
- Renshaw, E. (1991). Modelling Biological Populations in Space and Time. Cambridge University Press.
- Grimmett, G., & Stirzaker, D. (2001). Probability and Random Processes (3rd ed.). Oxford University Press.
- Mikosch, T. (2009). Elementary Stochastic Calculus With Finance in View. World Scientific Publishing Co.

13.13 Further Reading

- Gardiner, C. W. (2009). Stochastic Methods: A Handbook for the Natural and Social Sciences (4th ed.). Springer-Verlag.
- Renshaw, E. (2011). Stochastic Population Processes: Analysis, Approximations, Simulations. Oxford University Press.
- Kleinrock, L. (1975). Queueing Systems, Volume I: Theory. John Wiley & Sons.
- Daley, D. J., & Vere-Jones, D. (2003). An Introduction to the Theory of Point Processes: Volume I: Elementary Theory and Methods. Springer-Verlag.
- Oksendal, B. (2003). Stochastic Differential Equations: An Introduction with Applications (6th ed.). Springer-Verlag.
- Haccou, P., Jagers, P., & Vatutin, V. A. (2005). Branching Processes: Variation, Growth, and Extinction of Populations. Cambridge University Press.

UNIT - 14: NON-HOMOGENEOUS PROCESS

Structure

14.1 Introduction

In Unit 14, we delve into the complex and dynamic world of Non-Homogeneous Processes, a critical area of study in stochastic modelling. Unlike their homogeneous counterparts, non-homogeneous processes are characterized by probability distributions and parameters that change over time. This dynamic nature makes them more aligned with real-world scenarios, where conditions are seldom constant and evolve in response to various factors.

The study of non-homogeneous processes opens-up a vast array of applications across different fields such as finance, where market conditions fluctuate over time; biology, where population dynamics change due to evolving environmental factors; and physics, where systems are subject to time-dependent forces. The versatility of these processes lies in their ability to model complex phenomena that cannot be adequately represented by time-invariant models.

At the heart of this unit is the exploration of simple non-homogeneous processes. These are the foundational models from which more complex systems are developed. Understanding these simple models paves the way for grasping more intricate processes like the Polya process, an urn model with significant implications in probability and statistics, and reflecting real-world scenarios of contagion and popularity dynamics.

Another focal point of this unit is the examination of how external factors, particularly immigration, impact non-homogeneous processes. In many real-world systems, the introduction of new entities from outside significantly alters the dynamics, and understanding this influence is crucial for accurate modelling.

The unit also delves into advanced mathematical concepts such as the Backward Kolmogorov Diffusion Equation and the Fokker-Planck Equation. These equations are fundamental in the study of stochastic processes, enabling the modelling and analysis of diffusion phenomena and providing insights into the behaviour of complex systems under varying conditions.

By the end of this unit, learners will have a thorough understanding of non-homogeneous processes, equipped with both the theoretical knowledge and practical skills to apply these concepts in various domains. Whether in academic research, professional practice, or practical applications, the insights gained from this unit are invaluable in navigating and modelling the ever-changing dynamics of real-world systems.

14.2 Objectives

1. Understanding the Fundamental Concept of Non-Homogeneous Processes: Grasp the definition and key characteristics that distinguish non-homogeneous processes from homogeneous ones. Comprehend why non-homogeneous processes are crucial in modelling real-world scenarios where conditions change over time.

2. Exploring Various Types of Non-Homogeneous Processes: Gain familiarity with different types of non-homogeneous processes, including simple non-homogeneous processes and more complex models like the Polya process. Understand the specific features and applications of each type of process.

3. Analysing the Impact of External Factors such as Immigration: Investigate how external factors, particularly immigration, influence the dynamics of non-homogeneous processes. Learn to model the effects of immigration in the context of non-homogeneous stochastic processes and understand their implications in real-world scenarios.

4. Mastering the Backward Kolmogorov Diffusion Equation: Delve into the derivation and significance of the Backward Kolmogorov Diffusion Equation in the context of non-homogeneous processes. Understand how this equation is used to analyse diffusion processes in various fields.

5. Understanding the Fokker-Planck Equation: Study the derivation and applications of the Fokker-Planck Equation. Comprehend how this equation models the probability density function of the velocity of particles in a fluid and its broader applications in stochastic processes.

6. Developing Problem-Solving Skills: Enhance the ability to apply theoretical knowledge to solve practical problems related to non-homogeneous processes. Improve skills in mathematical modelling and analysis of complex systems that evolve over time.

7. Critical Thinking and Real-World Application: Encourage critical thinking about how non-homogeneous processes can be applied to model and solve real-world problems in various domains such as finance, biology, and physics. Foster the ability to identify and analyse situations where non-homogeneous stochastic models are applicable.

These objectives are designed to provide learners with a comprehensive understanding of non-homogeneous processes, their mathematical underpinnings, and practical applications. The unit aims to equip students, researchers, and professionals with the necessary tools to analyse and interpret systems where conditions change over time, preparing them for challenges in diverse fields of study and work.

14.3 Simple Non-Homogeneous Process

A Simple Non-Homogeneous Process is a type of stochastic process where the probability laws governing the process change over time. Unlike homogeneous processes, where these laws are time-invariant, simple non-homogeneous processes adapt to evolving conditions, making them more reflective of many real-world scenarios.

Characteristics of Simple Non-Homogeneous Processes

Time-Dependent Transition Rates: The most defining feature is that the transition rates (or probabilities) are functions of time. This means that the likelihood of transitioning from one state to another is not constant but varies as time progresses.

Lack of Stationarity: Unlike homogeneous processes, simple non-homogeneous processes do not exhibit stationarity. Stationarity implies that statistical properties such as mean, variance, and autocorrelation remain constant over time, which is not the case for non-homogeneous processes.

Flexible Modelling: These processes can adapt to a wide range of scenarios where the system's behaviour changes over time, due to external factors, internal dynamics, or evolving conditions.

Mathematical Formulation: In mathematical terms, a simple non-homogeneous process can often be described by time-dependent differential equations or difference equations. For example, consider a non-homogeneous Poisson process, where the rate parameter $\lambda(t)$ is a function of time. The probability of observing k events in a time interval [0, t] is given by:

$$P(N(t)=k)= e^{-\Lambda(t)} (\Lambda(t))^k / k!$$

where

$\Lambda(t)=\int_0^t \lambda(s)ds$ is the cumulative rate function.

Examples

Non-Homogeneous Poisson Process: A classic example of a simple non-homogeneous process is a Poisson process with a time-dependent rate $\lambda(t)$. For instance, this can model customer arrivals at a store where the arrival rate varies depending on the time of day.

Variable Interest Rate Models in Finance: In financial modelling, the interest rate might change over time, reflecting economic conditions. A non-homogeneous stochastic model can represent such a variable interest rate, with the rate at any time t being a function of time.

Seasonal Variations in Population Dynamics: In ecological models, the birth and death rates of a population might vary seasonally. A simple non-homogeneous process can model these rates as time-dependent, reflecting changes in environmental conditions.

Applications: Simple non-homogeneous processes are widely applicable in areas where it is crucial to model the dynamics of a system that changes over time. This includes fields like telecommunications (for modelling varying network traffic), finance (for modelling time-varying

market risks), epidemiology (for disease spread with time-varying infection rates), and many others.

In summary, simple non-homogeneous processes offer a flexible and realistic framework for modelling stochastic systems that change over time, providing an essential tool for understanding and predicting behaviour in various dynamic environments.

Simple non-homogeneous processes are prevalent in various fields, reflecting systems where probabilities or rates change over time. Here are some examples to illustrate this concept:

Example 1: **Non-Homogeneous Poisson Process for Customer Arrivals**

Imagine a retail store where the rate of customer arrivals varies throughout the day. It's busier during the morning and evening but quieter in the afternoon.

Modelling: This scenario can be modelled using a non-homogeneous Poisson process where the arrival rate $\lambda(t)$ is a function of time.

For instance, $\lambda(t)$ could be higher during peak hours (morning and evening) and lower during off-peak hours (afternoon).

The number of customer arrivals in a time interval can then be predicted using the time-dependent rate function.

Example 2: **Variable Interest Rate in Financial Modelling**

Consider a financial model where the interest rate on investments or loans changes over time, influenced by economic factors.

Modelling: The interest rate can be represented as a non-homogeneous process where $r(t)$, the rate at time t, is not constant but varies based on market conditions.

Such a model allows for more realistic financial planning and risk assessment, reflecting the actual fluctuations in the economic environment.

Example 3: Seasonal Disease Spread in Epidemiology

In epidemiology, the rate of disease transmission can vary with seasons, influenced by factors like weather conditions and human behaviour.

Modelling: A non-homogeneous model can be used to represent the disease transmission rate, which varies over time. For example, a higher transmission rate in winter due to indoor crowding and lower in summer.

This time-dependent model helps in understanding the seasonal dynamics of disease spread and planning public health interventions accordingly.

Example 4: Varying Traffic Flow in Telecommunications

In a telecommunications network, data traffic can vary significantly throughout the day, with peak usage during certain hours.

Modelling: The data traffic can be modelled as a non-homogeneous process, where the rate of data packet arrivals varies with time.

Such a model is crucial for network planning and ensuring adequate bandwidth during peak usage times.

These examples demonstrate the versatility of simple non-homogeneous processes in modelling real-world scenarios where the system's behaviour is not constant but varies over time. They provide a more realistic framework for analysis and prediction in various dynamic environments.

14.4 Polya Process

The Polya process, named after the Hungarian mathematician George Polya, is a classic example of a non-homogeneous stochastic process. It is particularly known for exhibiting the "rich get richer" phenomenon and has applications in areas ranging from epidemiology to social network analysis.

Definition of the Polya Process: The Polya process is a type of urn model. It begins with an urn containing balls of different colours. At each step, a ball is drawn at random, its colour is noted, and then it is returned to the urn along with an additional ball of the same colour. This process results in an increasing probability of drawing balls of a colour that is already frequent in the urn, encapsulating a positive feedback loop or reinforcement mechanism.

Mathematical Formulation: Let us denote: $n_i(t)$: Number of balls of color in the urn at time *N(t):* Total number of balls in the urn at time *t*. The probability of drawing a ball of color *i* at time *t* and then adding one more of the same colour is given by: Drawing colour at time P(Drawing color *i* at time *t*)= $N(t)n_i(t)$

Supporting Results and Proofs

Probability Distribution Over Time: As the process evolves, the probability distribution of each colour's presence in the urn converges to a limiting distribution.

Proof: The update rule of the process implies that the fractions $N(t)n_i(t)$ evolve according to a path-dependent random walk. By the Law of Large Numbers, as t→∞, these fractions converge almost surely to a limit. The specific distribution depends on the initial conditions of the urn.

Rich Get Richer Phenomenon: A colour that initially has more balls is more likely to be drawn and thus to have even more balls added, leading to a self-reinforcing mechanism.

Proof: Consider two colours, A and B, with A(0)> nB(0) initially. The probability of drawing colour A at any step is higher than that of drawing B due to the greater initial number. As more A's are added, this probability disparity grows, reinforcing A's dominance.

Applications

Epidemiology: Modelling the spread of disease, where the "balls" represent infected individuals and "colours" represent different strains of the pathogen.

Social Network Analysis: Explaining how popular individuals or nodes in a network become increasingly popular, akin to the concept of preferential attachment.

Economics: Understanding market dynamics where successful products gain increasing market share due to positive customer feedback loops.

The Polya process is a compelling model for systems where success breeds success, and its mathematical properties offer insights into the behaviour of complex, evolving systems. However, the detailed proofs of its properties require advanced knowledge of probability theory and stochastic processes and can be quite complex, often involving limit theorems and martingale

theory. For in-depth mathematical proofs, reference to specialized texts in stochastic processes is recommended.

The Polya process, as a reinforcement model, is governed by several important theorems that help in understanding its behaviour and long-term dynamics.

The key theorems related to the Polya process:

1. Convergence Theorem

Statement: The proportion of balls of a given colour in the urn converges almost surely to a limit as the number of draws goes to infinity.

Implication: This theorem ensures that the relative frequency of each colour stabilizes over time, even though the exact composition of the urn continues to change with each draw.

2. Martingale Convergence Theorem

Statement: The sequence of proportions of a given colour in the Polya urn forms a martingale sequence, and by the Martingale Convergence Theorem, this sequence converges almost surely.

Implication: This provides a strong mathematical foundation for the convergence behavior observed in the Polya process, grounding it in well-established principles of probability theory.

3. Theorem of Exchangeability

Statement: The sequence of draws in a Polya urn process is exchangeable, meaning that the joint probability distribution is invariant under permutations of the sequence.

Implication: This property of exchangeability implies that the order of draws does not affect the overall probability distribution of the colours, which is a key characteristic of the Polya process.

4. Central Limit Theorem for Polya Urn

Statement: Under certain conditions, the number of balls of a particular colour in the urn, after normalization, converges in distribution to a normal distribution as the number of draws increases.

Implication: This theorem provides insight into the variability and distribution of the number of balls of each colour as the process evolves over a long period.

5. Pólya-Eggenberger Urn Model Theorem

Statement: In a generalized version of the Polya urn model, known as the Pólya-Eggenberger urn model, the probability distribution of the composition of the urn converges to a Dirichlet distribution.

Implication: This generalization allows for more complex interactions and dependencies between different types of balls, providing a broader applicability of the model.

These theorems collectively offer a comprehensive understanding of the behaviour of the Polya process. They demonstrate how initial advantages or disadvantages can get amplified over time, leading to a "rich-get-richer" phenomenon. The mathematical treatment of these theorems typically involves advanced concepts in probability theory, such as martingales and limit theorems, and is central to the study of reinforcement processes and their applications in various fields.

Example 1: **Disease Spread in Epidemiology:** Consider the spread of a contagious disease where each infected individual increases the likelihood of infecting others, akin to adding more balls of the same colour to an urn in the Polya process.

Application: The Polya process can model the spread of infection, especially in situations where each new case increases the probability of further infections. This is often seen in tightly-knit communities or regions with high population density.

Example 2: **Online Social Networks and Viral Content:** On social media platforms, content that gains initial popularity is more likely to be recommended to other users, thereby increasing its chances of being shared further.

Application: The Polya process can model the dynamics of content virality, where initial likes or shares (akin to the initial balls in the urn) increase the content's visibility, leading to a self-reinforcing loop of popularity.

Example 3: **Financial Markets and Stock Trading:** In the stock market, a stock that performs well initially may attract more investors, increasing its demand and potentially its price further.

Application: The stock market dynamics, particularly the behaviour of investors who are influenced by past performance, can be modelled using the Polya process. A stock that gains early investors (more balls of its colour in the urn) is more likely to attract additional investors.

***Example 4*: Preferential Attachment in Network Science:** In network science, particularly in the study of the internet and citation networks, new nodes tend to connect to already well-connected nodes.

Application: The concept of preferential attachment, where new nodes in a network are more likely to connect to nodes with higher degrees, can be modelled by the Polya process. It's akin to adding more connections (balls) to nodes (colours) that are already popular.

***Example 5:* Consumer Behaviour in Marketing:** A product that gains early popularity in the market might receive more attention and thus have higher subsequent sales.

Application: In marketing, the Polya process can model consumer behaviour where early adoption of a product increases its visibility and appeal, leading to a positive feedback loop in sales.

These examples demonstrate the versatility of the Polya process in modelling scenarios where the probability of an event increases with the occurrence of that event. This self-reinforcing property makes it a valuable tool in understanding and predicting dynamics in fields ranging from epidemiology and social media to financial markets and network theory.

14.5 Effect of Immigration for Non-Homogeneous Process

The concept of immigration in the context of non-homogeneous processes refers to the introduction of new entities into a system from an external source. This addition can significantly alter the dynamics of the system, particularly in processes where the inherent properties are time-dependent. In non-homogeneous processes, the rates of change or transition probabilities are not constant over time, and the introduction of immigration adds another layer of complexity to this time-variant behaviour.

Characteristics and Implications of Immigration in Non-Homogeneous Processes

External Additions: Immigration represents an external influx of entities (like individuals, particles, etc.) into the system, which is independent of the internal generation (birth) process of the system.

Altering System Dynamics: The continuous introduction of new entities can significantly affect the system's state and its evolution over time. It can prevent extinction in population models, increase the workload in queueing systems, or accelerate the spread of a disease in epidemiological models.

Time-Variant Nature: In non-homogeneous processes, the impact of immigration may vary with time, adding to the complexity of modelling and analysis. For instance, the rate or effect of immigration could be higher during certain periods due to external factors.

Mathematical Modelling: Incorporating immigration into a non-homogeneous process typically involves adding a time-dependent term to the existing model. For example, in a non-homogeneous birth-death process, the differential equations describing the system would include an additional term representing the immigration rate.

If X(t) represents the state of the system at time t, and β(t) represents the time-dependent immigration rate, the evolution of X(t) could be modelled by: $\frac{dX(t)}{dt}$=[Internal Dynamics]+β(t)

Here, "Internal Dynamics" could include time-variant birth and death rates, and β(t) adds the external influx at each time point.

Real-World Scenarios and Applications

Population Dynamics: In ecological models, immigration can introduce new species or individuals into an ecosystem, affecting the existing population dynamics, especially if the ecosystem is changing over time due to environmental factors.

Epidemiology: The model of disease spread can be significantly impacted by the immigration of infected individuals, especially in a non-homogeneous scenario where the infection rate changes over time due to factors like seasonal variations or changing public health policies.

Queueing Theory: In a service system like a call centre, immigration can represent an additional influx of calls or customers from an external source, such as a marketing campaign, which can vary over time.

Economic Models: In macroeconomic studies, the immigration of individuals can impact labour markets, consumption patterns, and overall economic dynamics, particularly in models where these factors are time-dependent.

In summary, the effect of immigration on non-homogeneous processes introduces additional complexity and realism into models, making them more representative of real-world systems where external influxes play a significant role in the dynamics over time. This aspect is crucial for accurately understanding and predicting the behaviour of various systems in fields like ecology, epidemiology, economics, and operations research.

The effect of immigration on non-homogeneous processes can be observed in various real-world scenarios, where the introduction of new entities into a system significantly alters its dynamics. Here are some examples to illustrate this concept:

***Example 1*: Immigration in Population Dynamics:** In an ecological model, consider a wildlife reserve where animals of a particular species are periodically introduced from outside to boost the population.

Application: This scenario can be modelled as a non-homogeneous process where the natural birth and death rates of the species vary due to seasonal changes or environmental factors.

The periodic introduction of new animals represents immigration, altering the population dynamics, potentially preventing extinction, or altering the competition dynamics within the reserve.

***Example 2*: Disease Spread with Immigration of Infected Individuals:** In an area experiencing an infectious disease outbreak, consider the impact of individuals traveling into the area from regions with higher infection rates.

Application: The disease spread can be modelled as a non-homogeneous process, where the infection rate changes over time due to factors like public health interventions or seasonal variations.

The arrival of infected individuals from outside adds an immigration component to the model, potentially increasing the disease's spread or introducing new strains into the population.

Example 3: **Queueing Systems with External Customer Influx:** Imagine a customer service centre experiencing varying call volumes throughout the day, with additional calls directed from other branches or as a result of promotional campaigns.

Application: The call arrival rate is non-homogeneous, varying with business hours, customer behaviour, and other factors.

The additional calls from other branches or campaigns represent immigration, impacting the queue length, waiting times, and required service capacity.

Example 4: **Labor Market Dynamics with Immigration:** In an economic model, consider the effect of skilled workers immigrating into a country and joining the workforce.

Application: The labour market can be modelled as a non-homogeneous process, where demand and supply for labour change due to economic policies, technological advancements, or market trends.

The influx of skilled workers adds an immigration element, potentially affecting wage levels, employment rates, and economic growth.

Example 5: **Online Platforms with New User Registration:** An online social media platform or a content streaming service experiences variable user engagement over time and sees an influx of new users due to marketing efforts.

Application: User engagement (such as content consumption or social interactions) can be modelled as a non-homogeneous process, influenced by factors like trending topics, seasonal content, or algorithm changes.

The influx of new users due to marketing or promotional activities represents immigration, impacting overall platform engagement, content virality, and network dynamics.

These examples demonstrate how immigration can significantly impact the behaviour of various non-homogeneous processes, making modelling and analysis more complex but also more reflective of real-world situations.

14.6 Diffusion: Backward Kolmogorov Diffusion Equation

The Backward Kolmogorov Diffusion Equation is a fundamental concept in the study of stochastic processes, particularly in the context of diffusion processes. It is part of a family of equations known as the Kolmogorov equations, which also include the Forward Kolmogorov (or Fokker-Planck) Equation. These equations are central to the theory of continuous-time stochastic processes.

Definition and Background:

In the realm of stochastic processes, diffusion refers to a process that models the random movement of particles or entities. This movement is typically described using Brownian motion or a Wiener process, which are foundational models for random, continuous paths.

The Backward Kolmogorov Diffusion Equation focuses on the evolution of the conditional expectation of a function of the stochastic process backward in time, given its future state. This contrasts with the forward equation, which describes the forward-in-time evolution of the probability density function of the process.

Mathematical Formulation: Consider a stochastic process X(t) that follows a diffusion process. Let f(x,t) be a function of this process, where x is a state and t is time. The Backward Kolmogorov Diffusion Equation is given by:

$$\frac{\partial}{\partial t}u(x,t) = \frac{1}{2}\sigma^2(x,t)\frac{\partial^2}{\partial x^2}u(x,t) + b(x,t)\frac{\partial}{\partial x}u(x,t)$$

Where, $u(x,t)$ is the conditional expectation of some function f of the process, given its state at a later time.

$\sigma^2(x,t)$ is the diffusion coefficient, representing the variance of the process's increments.

$b(x,t)$ is the drift coefficient, representing the mean of the process's increments.

Physical Interpretation: The Backward Kolmogorov Equation can be interpreted in the context of physical diffusion processes, such as the diffusion of particles in a fluid. In this setting, $\sigma^2(x,t)$

represents the random scattering of particles, while $b(x,t)$ represents a systematic drift, such as movement due to an external force or flow.

Applications

Finance: In mathematical finance, this equation is used to model the evolution of various financial instruments' prices, particularly in option pricing.

Physics and Chemistry: It is used to model the diffusion of particles, heat, or chemicals in a medium.

Biology: Applied in modelling population dynamics and the spread of biological species.

Environmental Science: Used in modelling the spread of pollutants in the environment.

Importance in Stochastic Processes

The Backward Kolmogorov Diffusion Equation is important because it provides a way to calculate the expected value of a function of the stochastic process at a previous time, given its future state. This is particularly useful in scenarios where predicting the past behavior of a system based on its current or future state is necessary, such as in backward analysis or smoothing problems in statistics and data analysis.

Understanding and applying this equation requires a solid foundation in stochastic calculus and differential equations, as it involves complex concepts like Brownian motion and Ito's calculus.

The Backward Kolmogorov Diffusion Equation is applied in various fields to model and understand systems where diffusion processes are significant.

Here are some examples to illustrate its application:

Example 1: **Option Pricing in Financial Markets:** Consider the pricing of a financial derivative, like a stock option, where the future price of the underlying asset follows a stochastic process.

Application: The Backward Kolmogorov Equation is used in the Black-Scholes model to determine the option's price based on the expected future price of the underlying asset.

The equation models how the value of the option evolves backwards in time, given its payoff at maturity.

Example 2: **Heat Distribution in a Solid:** In physics, consider the problem of determining the temperature distribution within a solid object at a previous time, given the current temperature distribution.

Application: The equation can be used to model the diffusion of heat within the solid.

By knowing the current temperature distribution, the equation helps in estimating how the heat was distributed at an earlier time.

Example 3: **Pollution Spread in an Ecosystem:** In environmental science, understanding how a pollutant, like an oil spill, spread in the past based on its current distribution in water.

Application: The Backward Kolmogorov Equation can model the diffusion of the pollutant in the water.

It helps in tracing back the spread, which is crucial for identifying the source and understanding the spread pattern.

Example 4**: Population Dynamics in Ecology:** In biology, consider a situation where we need to understand the past distribution of a species in an ecosystem based on current population data.

Application: The equation can model the spread (dispersion) of the species in the ecosystem.

By applying the equation, biologists can estimate past population distributions, which is vital for conservation and study of migration patterns.

Example 5**: Data Smoothing in Time Series Analysis:** In statistics, smoothing a time series data to understand past trends based on current and future observations.

Application: The Backward Kolmogorov Equation can be applied in the context of stochastic processes to smooth data, providing estimates of past states of a time series. This is particularly useful in financial time series where past trends are inferred for analysis and forecasting.

These examples demonstrate the versatility of the Backward Kolmogorov Diffusion Equation in various fields, providing a means to retrospectively analyse and understand the behaviour of systems modelled by diffusion processes.

14.7 Fokker-Planck Equation

The Fokker-Planck Equation, also known as the Forward Kolmogorov Equation, is a fundamental equation in the study of stochastic processes, particularly in the context of diffusion and continuous-time processes. It describes the time evolution of the probability density function of the velocity of a particle under the influence of forces and random fluctuations.

Statement of the Fokker-Planck Equation

Consider a stochastic process X(t) that represents the state of a system. The Fokker-Planck Equation describes the evolution of the probability density function p(x,t) of this process over time. Mathematically, it is expressed as:

$$\frac{\partial p(x,t)}{\partial t} = -\frac{\partial}{\partial x}[\alpha(x,t)p(x,t)] + \frac{1}{2}\frac{\partial^2}{\partial x^2}[b^2(x,t)p(x,t)]$$

where: $p(x,t)$is the probability density function of finding the system in state x at time t. $\alpha(x,t)$ is the drift coefficient, representing the deterministic part of the process. $b(x,t)$is the diffusion coefficient, representing the random fluctuations.

Proof of the Fokker-Planck Equation

The proof of the Fokker-Planck Equation is derived from the Chapman-Kolmogorov Equation and involves applying Itô's calculus for stochastic differential equations. The detailed proof is quite technical and beyond the scope of a basic explanation, as it requires a deep understanding of stochastic calculus. However, the key idea is to start with the differential form of a stochastic process (usually expressed in terms of a Wiener process or Brownian motion) and then derive the partial differential equation governing the evolution of its probability density function.

Applications in Various Fields

Physics: In the study of thermodynamics and statistical mechanics, the Fokker-Planck Equation is used to describe the behaviour of particles in a fluid, accounting for random Brownian motion and external forces.

Finance: In financial mathematics, it is used to model the evolution of stock prices and interest rates, where the random component represents market volatility, and the drift represents the overall trend.

Biology: The equation is applied in modelling the spread of genes in a population or the movement of organisms in response to environmental factors.

Chemistry: In chemical kinetics, the Fokker-Planck Equation models the stochastic behaviour of molecules in reaction-diffusion systems.

Meteorology: It is used to model the distribution and movement of particles in the atmosphere, such as pollutants or water droplets.

Social Sciences: In economics and social sciences, the equation can model various diffusion processes, like the spread of information or the adoption of new technologies.

The Fokker-Planck Equation provides a powerful framework for understanding and predicting the behaviour of systems characterized by both deterministic trends and random fluctuations. It is an essential tool in fields where modelling the time evolution of probability distributions is crucial.

The Fokker-Planck Equation, with its ability to describe the evolution of probability distributions in systems influenced by random fluctuations, finds diverse applications across multiple fields.

Here are examples illustrating its use:

Example 1: **Financial Markets - Stock Price Modelling:** Modelling the price of a stock in a financial market, which fluctuates due to both systematic market trends and random market volatilities.

Application: The Fokker-Planck Equation is used to model the evolution of the probability distribution of stock prices over time.

The drift term represents the overall market trend or expected return, while the diffusion term models the random volatility of the market.

Example 2: **Physics - Brownian Motion of Particles:** Describing the behaviour of microscopic particles suspended in a fluid, where their movement is influenced by collisions with fluid molecules.

Application: The equation models the probability distribution of particle positions, accounting for both systematic forces (like gravity) and random Brownian motion.The drift term can represent external forces acting on the particles, and the diffusion term encapsulates the random thermal motion.

Example 3: **Ecology - Spread of a Biological Species:** Understanding the spread of a biological species in an ecosystem, influenced by factors like migration and environmental conditions.

Application: The Fokker-Planck Equation is used to describe the distribution of the species over a geographical area.

The drift term can model directed movement (like migration towards favourable conditions), while the diffusion term represents random dispersal.

Example 4: **Chemistry - Reaction-Diffusion Systems:** In chemical kinetics, modelling how the concentration of reactants changes over time and space in a reaction-diffusion system.

Application: The equation models the concentration profile of reactants and products.

The drift term can represent the reaction kinetics, while the diffusion term models the random molecular motion leading to diffusion.

Example 5**: Neuroscience - Neuron Firing Rates :** Modelling the firing rates of neurons, which are influenced by both the neuron's inherent properties and random synaptic inputs.

Application: The Fokker-Planck Equation is used to model the probability distribution of the neuron's membrane potential.

The drift term represents the deterministic dynamics of the membrane potential, and the diffusion term accounts for the random input fluctuations.

These examples demonstrate the broad applicability of the Fokker-Planck Equation in modelling systems where outcomes are influenced by a combination of deterministic trends and

random fluctuations. It provides a crucial tool for understanding the behaviour and evolution of complex systems across various scientific disciplines.

14.8 Summary

This unit delves into the complex yet fascinating world of Non-Homogeneous Processes, expanding the understanding of stochastic processes beyond the realm of time-invariant systems. This unit begins by defining non-homogeneous processes, emphasizing their key feature: the changing probability distributions and parameters over time, which set them apart from homogeneous processes. These changes make them particularly suitable for modelling real-world scenarios where conditions are dynamic and evolve.

A significant portion of the unit is dedicated to exploring various types of non-homogeneous processes, starting with simple non-homogeneous processes that lay the groundwork for understanding more complex models. The Polya process is examined in detail, a quintessential example of a non-homogeneous process known for its "rich get richer" dynamic. This urn model, where the probability of drawing a particular colour increases with the frequency of that colour in the urn, illustrates the reinforcement mechanisms often seen in social, economic, and biological systems.

The unit then shifts focus to the impact of immigration on non-homogeneous processes. This aspect is crucial in understanding how the introduction of new entities from outside sources can significantly alter the dynamics of a system, such as in population dynamics, disease spread, and queueing theory.

Advanced mathematical concepts form the core of later sections, with detailed discussions on the Backward Kolmogorov Diffusion Equation and the Fokker-Planck Equation. These equations are instrumental in modelling diffusion processes and provide a mathematical framework for understanding how probability distributions evolve over time in systems influenced by random fluctuations and external forces.

The unit concludes with self-assessment questions designed to test the learner's understanding of the concepts, followed by a comprehensive list of references for further study. The summary encapsulates the key points covered, highlighting the importance of non-homogeneous processes in stochastic modelling and their relevance in various fields such as finance, biology, environmental science, and more.

Overall, this unit offers a deep dive into the intricacies of non-homogeneous stochastic processes, equipping learners with the knowledge and tools to analyse complex systems that change over time, and preparing them for advanced studies or professional applications in diverse fields.

14.9 Self-Assessment Questions

Theoretical Questions

1. Define a Non-Homogeneous Process: Explain what distinguishes a non-homogeneous process from a homogeneous one. Provide an example from real life that illustrates this difference.

2. Explain the Polya Process: Describe the Polya process in detail. What makes it an example of a non-homogeneous process, and how does it exemplify the "rich get richer" phenomenon?

3. Immigration in Non-Homogeneous Processes: Discuss the impact of immigration on non-homogeneous processes. How does the introduction of new entities from an external source affect the dynamics of these processes?

4. Backward Kolmogorov Diffusion Equation: Explain the significance of the Backward Kolmogorov Diffusion Equation in the context of non-homogeneous processes. How does it differ from the Fokker-Planck Equation?

5. Applications of the Fokker-Planck Equation: List at least three fields where the Fokker-Planck Equation is applied and explain its role in each field.

Problem-Solving Questions

1. Applying the Polya Process: Assume an urn initially contains 3 red balls and 2 blue balls. If a ball is drawn at random and then returned to the urn along with another ball of the same colour, what is the probability that the third ball drawn is red?

2. Modelling with Immigration: A certain species of fish is introduced into a lake at a rate of 50 fish per year. The birth and death rates of the fish are known to change seasonally. Sketch a basic model that represents the fish population over time.

3. Using the Backward Kolmogorov Equation: Suppose the stock price of a company follows a diffusion process with a drift coefficient μ and a diffusion coefficient σ^2. Write down the Backward Kolmogorov Equation for this process.

4. Fokker-Planck Equation Problem: Given a particle undergoing Brownian motion in a fluid with a drift coefficient αx and a constant diffusion coefficient β, write down the Fokker-Planck Equation for the probability density function of the particle's position.

5. Real-World Scenario Analysis: Consider a scenario where a new technology is being adopted in a market. The rate of adoption is influenced by the number of current users (network effect) and random market factors. How would you model this scenario using a non-homogeneous process?

14.10 References

- Gardiner, C. W. (2009). Handbook of Stochastic Methods: for Physics, Chemistry and the Natural Sciences (4th ed.). Springer-Verlag.
- Risken, H. (1996). The Fokker-Planck Equation: Methods of Solution and Applications (2nd ed.). Springer-Verlag.
- Oksendal, B. (2003). Stochastic Differential Equations: An Introduction with Applications (6th ed.). Springer-Verlag.
- Pavliotis, G. A. (2014). Stochastic Processes and Applications: Diffusion Processes, the Fokker-Planck and Langevin Equations. Springer-Verlag.
- Durrett, R. (2019). Probability: Theory and Examples (5th ed.). Cambridge University Press.
- Mahmudov, N. M. (2019). Stochastic Differential Equations: An Introduction with Applications in Population Dynamics Modelling. Wiley.

14.11 Further Reading

For extended learning and deeper insights into the topics covered in Unit 14: Non-Homogeneous Process, the following suggestions for further reading are provided, along with the names of their publishers in standard format:

- Gardiner, C. W. (2009). Handbook of Stochastic Methods: for Physics, Chemistry and the Natural Sciences (4th ed.). Springer-Verlag.
- Risken, H. (1996). The Fokker-Planck Equation: Methods of Solution and Applications (2nd ed.). Springer-Verlag.
- Oksendal, B. (2003). Stochastic Differential Equations: An Introduction with Applications (6th ed.). Springer-Verlag.
- Pavliotis, G. A. (2014). Stochastic Processes and Applications: Diffusion Processes, the Fokker-Planck and Langevin Equations. Springer-Verlag.
- Durrett, R. (2019). Probability: Theory and Examples (5th ed.). Cambridge University Press.
- Mahmudov, N. M. (2019). Stochastic Differential Equations: An Introduction with Applications in Population Dynamics Modelling. Wiley.

These resources provide a comprehensive understanding of stochastic processes, particularly non-homogeneous processes, and their applications in various scientific and engineering fields. The books range from introductory to advanced levels, catering to a wide spectrum of readers from students to professionals in the field.

UNIT - 15: NON-MARKOVIAN PROCESS

Structure

15.1 Introduction

The study of dynamic systems and processes forms the cornerstone of understanding complex behaviours in various disciplines such as physics, biology, economics, and beyond. Within this broad spectrum, non-Markovian processes represent a fascinating and intricate class of stochastic processes that are pivotal in modelling systems where the future state does not solely depend on the present state but also on the history of the system. This unit, "Non-Markovian Process," delves into the core of these processes, offering a comprehensive exploration of their theoretical foundations, mathematical formulations, and applications across different fields.

Non-Markovian processes contrast sharply with their Markovian counterparts, which are characterized by the memoryless property, meaning the future is independent of the past given the present. Non-Markovian processes, on the other hand, incorporate memory effects, making the analysis and prediction of future states more complex and nuanced. This memory aspect allows for a more accurate modelling of real-world phenomena where the history significantly influences the evolution of the system.

The unit begins by introducing some multi-dimensional prey and predator models, illustrating the application of non-Markovian dynamics in ecological systems. These models provide insight into how historical interactions between species can influence current and future population dynamics, offering a richer understanding than traditional Markovian models.

Following this, we delve into the theoretical underpinnings of non-Markovian processes, discussing their properties, classifications, and the mathematical challenges they pose. This section lays the groundwork for understanding the depth and breadth of non-Markovian processes and prepares the learner for more complex applications and analyses.

Embedded Markov Processes are introduced as a bridge between Markovian and Non-Markovian processes, showcasing how certain non-Markovian processes can be approximated or transformed into a Markovian framework under specific conditions. This concept is crucial for developing practical solutions and analytical methods in systems that are inherently non-Markovian.

The application sections focus on population growth and epidemic models, illustrating the power of non-Markovian processes in capturing the complexities of real-world phenomena. From predicting the spread of diseases to understanding the nuances of population dynamics in response to environmental changes, these applications highlight the importance of considering historical data and memory effects.

This unit aims to equip learners with a solid understanding of Non-Markovian processes, from their theoretical foundations to practical applications. Through detailed explanations, mathematical modelling, and real-world examples, learners will gain insights into the critical role these processes play in advancing our understanding of complex systems. Whether you're a student, researcher, or enthusiast, this exploration of Non-Markovian processes offers valuable knowledge and perspectives that bridge the gap between theory and application in the study of dynamic systems.

15.2 Objectives

The unit on Non-Markovian Processes is designed to deepen the understanding of these complex stochastic processes, highlighting their theoretical underpinnings, mathematical

frameworks, and practical applications. The detailed objectives are structured to guide learners through a comprehensive exploration, ensuring a robust grasp of the concepts, techniques, and significance of non-Markovian processes in various scientific and mathematical contexts. Here are the detailed objectives of this unit:

1. Understand the Fundamental Concepts of Non-Markovian Processes

- To introduce the concept of non-Markovian processes and differentiate them from Markovian processes.
- To elucidate the significance of memory and history in determining the future evolution of a system described by non-Markovian dynamics.
- To understand the conditions under which a process is considered non-Markovian.

2. Explore the Mathematical Formulation of Non-Markovian Processes

- To learn about the mathematical models and formulations used to describe non-Markovian processes.
- To understand the role of probability distributions, correlation functions, and other mathematical tools in analysing non-Markovian processes.
- To grasp the complexities involved in solving and interpreting non-Markovian models.

3. Examine Specific Examples and Applications

- To study multi-dimensional prey and predator models as examples of non-Markovian processes in ecological systems.
- To explore the application of non-Markovian processes in population dynamics, including the effects of historical events and environmental changes.
- To analyse the use of non-Markovian processes in modelling epidemic spread and understanding disease dynamics over time.

4. Understand Embedded Markov Processes within Non-Markovian Frameworks

- To learn how Embedded Markov Processes can be utilized to approximate or simplify the analysis of non-Markovian processes.

- To understand the conditions and methodologies for embedding Markov processes within non-Markovian frameworks.

5. Apply Mathematical Techniques to Non-Markovian Processes

- To acquire skills in applying mathematical and computational methods to solve non-Markovian process models.
- To understand the challenges and strategies in numerical simulation and analytical solutions of non-Markovian processes.

6. Critically Analyse Real-world Systems Using Non-Markovian Models

- To develop the ability to identify and model real-world phenomena where non-Markovian processes provide a more accurate or insightful description than Markovian models.
- To enhance problem-solving skills by applying non-Markovian process models to complex systems in biology, finance, physics, and other areas.

7. Engage with Advanced Topics and Current Research

- To familiarize learners with current research trends and advanced topics in the study of non-Markovian processes.
- To encourage further exploration and research in non-Markovian processes and their applications.

By achieving these objectives, learners will be equipped with a deep understanding of non Markovian processes, enabling them to apply these concepts to analyse and model complex systems with historical dependencies. This unit aims to foster analytical thinking, problem-solving skills, and a comprehensive understanding of how non-Markovian dynamics can be applied across various disciplines to address real-world challenges.

15.3 Some Multi-Dimensional Prey and Predator Models

Multi-dimensional prey and predator models extend the classic Lotka-Volterra equations to more complex ecosystems involving multiple species interactions. These models capture the dynamics of multiple prey and predator species, allowing for a richer understanding of ecological

and evolutionary dynamics. They can incorporate various factors such as competition, mutualism, and the spatial structure of habitats.

Definition and Mathematical Formulation: A simple multi-dimensional prey-predator model can be represented by a system of differential equations. Consider an ecosystem with n prey species (xi) and m predator species (yj), the dynamics can be described by:

$$\frac{dx_i}{dt} = x_i\left(r_i - \sum_{j=1}^{m} a_{ij}y_i - \sum_{k=1}^{n} b_{ik}x_k\right)$$

$$\frac{dy_i}{dt} = y_i\left(-b_j + \sum_{i=1}^{n} c_{ij}x_i\right)$$

where: r_i is the intrinsic growth rate of prey species i, d_j is the death rate of predator species j, a_{ij} represents the predation rate of predator j on prey i, b_{ik} represents the competition coefficient among prey species, c_{ij} represents the efficiency of converting consumed prey i into predator j biomass.

Examples

Example 1: Two Prey-One Predator Model: Consider a model with two prey species (x1 and x2) and one predator species (y). This model can represent a scenario where the predator feeds on both prey species, but the prey species also compete for resources.

The equations could look something like this:

$$\frac{dx_1}{dt} = x_1(r_1 - a_{11}y - b_{12}x_2)$$

$$\frac{dx_2}{dt} = x_2(r_2 - a_{21}y - b_{21}x_1)$$

$$\frac{dy}{dt} = y(-d + c_{11}x_1 + c_{21}x_2)$$

This model can demonstrate how a predator can indirectly facilitate the coexistence of competing prey species by preferentially preying on the more abundant species.

Example 2: One Prey-Two Predator Model: This model represents an ecosystem with one prey and two competing predators. It can be used to study the conditions under which multiple predators can coexist.

$$\frac{dx}{dt} = x(r - a_{11}y_1 - a_{12}y_2)$$

$$\frac{dy_1}{dt} = y_1(-d_1 + c_{11}x)$$

$$\frac{dy_2}{dt} = y_2(-d_2 + c_{12}x)$$

Supporting Results and Theorems

Theorem (Existence of Equilibria): For a given multi-dimensional prey-predator model, there exists a non-trivial equilibrium point if the interaction terms allow for a balance between the species' growth and death rates.

Proof Sketch: The existence of equilibria can be proven by setting the right-hand side of each differential equation to zero and solving the resulting system of algebraic equations. The solution(s) to this system represent the equilibrium points where the population sizes remain constant over time.

Theorem (Stability of Equilibria): An equilibrium point of a multi-dimensional prey-predator model is locally stable if the Jacobian matrix of the system at that point has eigenvalues with negative real parts.

Proof Sketch: the major steps involved in the proof:

- Calculate the Jacobian matrix of the system at the equilibrium point.
- Determine the eigenvalues of the Jacobian matrix.
- The equilibrium is locally stable if all eigenvalues have negative real parts, indicating that small deviations from the equilibrium will decay over time.

Multi-dimensional prey and predator models offer a nuanced view of ecological dynamics, allowing for the exploration of complex interactions among multiple species. Through mathematical analysis and simulation, these models can provide insights into the conditions that promote biodiversity, stability, and resilience in ecosystems. The examples and theorems mentioned serve as a foundation for understanding these complex interactions, providing a gateway to more detailed and specific studies within this fascinating area of mathematical biology.

15.4 Non-Markovian Process

Non-Markovian processes represent a class of stochastic processes that are fundamental in modelling systems where the future state depends not only on the current state but also on the history of the system. This dependency on past events makes non-Markovian processes inherently more complex and nuanced compared to their Markovian counterparts.

Mathematical Properties of Non-Markovian Processes

Memory: Unlike Markov processes, non-Markovian processes possess memory. The probability distribution of future states depends on the history of states, not just the present state.

Time-Dependent Transition Probabilities: The transition probabilities in non-Markovian processes can change over time, reflecting the influence of past states on future probabilities.

Long-Range Correlations: Non-Markovian processes often exhibit long-range correlations between events, where events far apart in time can still be statistically dependent.

Comparison with Markov Processes: Markov processes are characterized by the Markov property, where the future state depends only on the current state, making these processes "memoryless." In contrast, non-Markovian processes do not adhere to this property. The dependency on past events allows non-Markovian processes to model systems where history plays a crucial role in the evolution of the system.

Predictability: In Markov processes, predictability is solely based on the current state. For non-Markovian processes, predictability requires more information about past states, making them inherently more complex to analyse and predict.

Transition Probabilities: In Markov processes, transition probabilities are fixed and do not change over time. In non-Markovian processes, these probabilities can vary, reflecting the influence of the system's history.

Applications: Markov processes find applications in areas where the memoryless property is a reasonable approximation of the system's dynamics. Non-Markovian processes are applied in situations where history or memory effects are significant, such as in complex physical systems, financial markets, and certain biological processes.

Examples and Case Studies

Example 1: Financial Markets: In financial markets, asset prices often exhibit non-Markovian dynamics. The price of an asset may depend not only on its current state but also on its history of prices, reflecting the collective memory and reactions of the market participants to past events.

Case Study: The Flash Crash

The 2010 Flash Crash, where stock prices plummeted and partially recovered within minutes, is an example where non-Markovian dynamics can be considered. The rapid decline and recovery in prices were influenced by the interplay of automated trading algorithms, which were responding to both current and past market conditions.

Example 2: Polymer Physics

The dynamics of polymers and complex fluids often exhibit non-Markovian behaviour. The motion of a polymer chain in a solvent depends not only on its current configuration but also on its past configurations, due to the memory effects introduced by the viscosity of the solvent and the elasticity of the polymer chain.

Case Study: Viscoelastic Behaviour of Polymers

The viscoelastic behaviour of polymers, where they exhibit both viscous and elastic characteristics depending on the timescale of observation, is a manifestation of non-Markovian dynamics. The stress response of a polymer to strain depends on the entire history of its deformation, not just its current state.

Non-Markovian processes provide a rich framework for modelling and understanding systems where history and memory effects play a significant role. The mathematical complexity

of these processes allows for a more nuanced description of real-world phenomena, offering insights that Markovian models cannot capture. Through examples and case studies across various disciplines, the significance and applicability of non-Markovian dynamics are evident, highlighting their importance in both theoretical and applied contexts.

15.5 Embedded Markov Process

Embedded Markov processes provide a powerful framework for analysing and simplifying non-Markovian processes by identifying discrete events or states within a continuous-time process, where the memoryless property holds. This approach allows the conversion of complex non-Markovian dynamics into a more tractable form, leveraging the well-established theory of Markov processes.

Introduction to Embedded Markov Processes

An embedded Markov process focuses on a sequence of discrete events or states in a system that evolves over continuous time. By considering only these discrete points, one effectively "embeds" a Markov chain within a non-Markovian process. This technique is particularly useful in systems where state transitions at specific points in time are determined by probabilities that do not depend on the history prior to those points.

Simplifying Non-Markovian Processes

The key advantage of identifying an embedded Markov process within a non-Markovian system lies in simplification. Non-Markovian processes are characterized by memory effects, where future states depend on the entire history of the system. By contrast, an embedded Markov process focuses on transitions between states at particular instances, ignoring the path taken between these instances. This reduction to memoryless transitions at discrete times allows for the application of Markovian analysis techniques, including the calculation of steady-state probabilities and transition matrices.

How Embedding Works

The embedding process involves identifying specific events or criteria that define the transitions between states. For example, in a queueing system, one might consider the system state only at arrival or departure times of customers, ignoring the details of the waiting time. These

discrete events, where customers arrive or depart, form the embedded Markov chain, with the state transitions at these times following the Markov property.

Examples and Practical Applications

Queueing Theory: A classic application of embedded Markov processes is in queueing theory. In many queueing models, the system can be analysed as an embedded Markov chain at customer arrival or service completion times. The system's state can be described by the number of customers in the queue, and the transitions occur only when a customer arrives or leaves. This approach simplifies the analysis of complex service systems, allowing for the calculation of performance metrics such as average queue length or waiting time.

Inventory Management: In inventory management, an embedded Markov process can model the system state at discrete ordering or review times. For instance, the decision to reorder stock can be made at regular intervals based on the current inventory level, ignoring the detailed demand history between these intervals. This simplifies the analysis of inventory policies and helps in optimizing ordering strategies.

Reliability Engineering: Embedded Markov processes are also used in reliability engineering to analyse systems with repair and failure modes. By considering the system state only at transition events, such as failures and repairs, one can model the reliability and availability of complex systems more simply. This approach is useful for designing maintenance schedules and improving system design.

Embedded Markov processes offer a valuable tool for simplifying the analysis of non-Markovian systems by focusing on discrete transition events where the memoryless property holds. This approach enables the application of Markovian techniques to a broader range of problems, providing insights and solutions that would be difficult to obtain directly from the underlying non-Markovian dynamics. Through practical examples in queueing theory, inventory management, and reliability engineering, the utility and versatility of embedded Markov processes are clearly demonstrated, highlighting their importance in operational research and systems analysis.

15.6 Application to Population Growth

Non-Markovian processes offer a nuanced approach for modelling population growth, capturing the complexity of biological systems where the future state depends not only on the current population but also on its history. This section explores the application of non-Markovian models to population dynamics, illustrating their potential through analysis and real-world case studies.

Exploration of Non-Markovian Processes in Population Growth: In traditional population models, such as the logistic model, the growth rate is a function of the current population size, ignoring the historical context. Non-Markovian processes extend this by incorporating memory effects, where the growth rate can depend on past population sizes, environmental conditions, or other historical factors. This approach is particularly relevant for populations that experience delayed effects of predation, resource depletion, or environmental changes.

Analysis of Complex Population Dynamics: Non-Markovian models allow for the inclusion of time-dependent factors and memory effects in population growth equations. For instance, a non-Markovian model might incorporate the effect of a past disease outbreak on the current growth rate, reflecting the lasting impact on the population's health or genetic diversity. Similarly, the model could account for the delayed response of a population to environmental changes, such as gradual habitat degradation or climate shifts.

The mathematical formulation of such models typically involves integra-differential equations or differential equations with time-delayed terms. These equations account for the influence of past states on the current rate of population change, providing a more accurate representation of growth dynamics in complex ecosystems.

Case Studies

Case Study 1: Forest Recovery After Wildfires

Following a significant wildfire, the recovery of a forest ecosystem can be modelled using non-Markovian dynamics. The growth rate of the new vegetation depends not just on the current conditions but also on the history of the soil, the presence of surviving root systems, and the past availability of nutrients. A non-Markovian model can capture the delayed effects of fire on soil fertility and the gradual return of flora and fauna, providing insights into the long-term recovery processes and the resilience of the ecosystem.

Case Study 2: Fisheries Management

Fisheries management often deals with the challenge of predicting fish stock recovery after overfishing. A non-Markovian approach can model the population dynamics by incorporating the history of fishing pressure and its cumulative effect on the genetic diversity and reproductive capacity of the fish population. This model can help in designing sustainable fishing quotas that consider not only the current stock levels but also the historical exploitation rates and their long-term effects on the population.

Case Study 3: Disease Impact on Wildlife Populations

The impact of a disease outbreak on wildlife populations can have long-lasting effects that are well captured by non-Markovian models. For instance, the spread of a disease like white-nose syndrome in bat populations affects not only the current population size but also the future growth potential by reducing the breeding population and altering social behaviours. A non-Markovian model can incorporate these delayed effects, providing a framework for predicting long-term population trends and assessing the effectiveness of conservation strategies.

Non-Markovian processes provide a powerful framework for modelling population growth in complex biological systems, where the effects of past events play a significant role in shaping future dynamics. By incorporating memory effects and time-dependent factors, these models offer a deeper understanding of population dynamics, enabling more accurate predictions and effective management strategies. The case studies in forest recovery, fisheries management, and disease impact demonstrate the practical application of Non-Markovian models in addressing real-world ecological and conservation challenges.

Use of non-Markovian processes in modelling the spread of diseases (epidemics).

The use of non-Markovian processes in modelling the spread of diseases offers a sophisticated framework to capture the complexities and intricacies of epidemic dynamics. Traditional epidemiological models often assume memoryless processes, where the future state of an individual's health status (susceptible, infected, or recovered) depends only on the current state, not taking into account the history of the disease spread or individual contact patterns. However, many infectious diseases and their spread through populations exhibit dependencies on past interactions and events, making non-Markovian processes an essential tool for more accurate and realistic modelling.

Incorporating Memory in Disease Spread Models: Non-Markovian processes allow for the inclusion of various memory effects in disease modelling, such as:

Duration of Infection: The time an individual remains infectious can vary significantly, depending on factors such as the individual's immune response and the nature of the pathogen. Non-Markovian models can incorporate a distribution of infectious periods rather than assuming a constant rate of recovery.

Contact Patterns: The frequency and nature of contacts between individuals can have memory, influenced by social structures, movement patterns, and changes in behaviour over time. Non-Markovian models can account for these dynamic contact networks, improving the modelling of disease transmission.

Latency Periods: Many diseases have an incubation period during which an infected individual is not yet infectious. The duration of this period can depend on the individual's health history and the disease characteristics, which can be modelled using non-Markovian processes.

Advantages of Non-Markovian Models in Epidemic Modelling:

Realistic Representation of Disease Dynamics: By incorporating the history of individual states and interactions, non-Markovian models offer a more realistic representation of how diseases spread through populations.

Improved Predictive Power: These models can provide more accurate predictions of outbreak dynamics, including the timing and magnitude of epidemic peaks, by accounting for the variability in individual behaviours and disease characteristics.

Flexible Framework: Non-Markovian processes offer a flexible framework that can be tailored to include specific features of disease spread, such as super-spreader events or the impact of public health interventions over time.

Practical Applications and Case Studies

Modelling COVID-19 Pandemic: The COVID-19 pandemic is a prime example where non-Markovian processes have been applied to model the spread of the disease. The varying infectious

periods, asymptomatic transmission, and changes in social behaviour over time (due to lockdowns and social distancing measures) necessitated the use of models that could account for these complexities. Non-Markovian models were used to simulate the effects of intervention strategies, predict hospitalization needs, and understand the potential impact of vaccination rollouts.

HIV Transmission Dynamics: The spread of HIV within populations is significantly influenced by long-term relationships and behaviour change over time. Non-Markovian models have been employed to capture the dynamics of transmission within networks, considering the duration of relationships and the changing rates of partner acquisition and loss. These models have helped in understanding the long-term trends in HIV prevalence and the effects of interventions such as antiretroviral therapy and education campaigns.

The use of non-Markovian processes in modelling the spread of diseases provides a nuanced and detailed approach that captures the complexities of real-world epidemics. By incorporating memory effects and time-dependent behaviours, these models offer enhanced predictive capabilities and insights, aiding public health officials and researchers in the effective management and control of infectious diseases. As epidemic modelling continues to evolve, the role of non-Markovian processes will remain crucial in addressing the challenges of emerging and re-emerging infectious diseases.

Introduction to counter models and their significance in studying epidemics.

Counter models, in the context of epidemiology, are analytical or computational frameworks designed to track the progression of diseases within populations by quantifying specific events or interactions that contribute to the spread of an infection. These models are particularly significant for studying epidemics because they allow researchers to dissect complex disease transmission dynamics into quantifiable and observable units. By focusing on counters, such as the number of contacts, infections, recoveries, or vaccinations, these models offer a structured approach to understanding and predicting the behaviour of infectious diseases.

Significance of Counter Models in Studying Epidemics

Counter models serve several crucial roles in the study of epidemics:

***Quantification of Transmission Events*:** They enable the precise quantification of transmission events, allowing for a detailed analysis of how, when, and where infections occur.

***Identification of Key Parameters*:** Counter models help identify key epidemiological parameters, such as the basic reproduction number (R0), which is essential for understanding the potential spread of an epidemic.

***Assessment of Intervention Strategies*:** By simulating the effects of public health interventions (e.g., vaccination, social distancing, contact tracing) on the counters, these models can assess the potential impact of such strategies on controlling the epidemic.

***Enhanced Predictive Capabilities*:** They provide enhanced predictive capabilities by allowing for the simulation of various scenarios under different parameters and interventions, aiding in decision-making and planning.

Detailed Examination of Disease Dynamics Through Counter Models

Modelling Contact Networks: Counter models can explicitly model contact networks within a population, tracking the number of contacts each individual has with others. This approach is invaluable for understanding the role of super-spreaders (individuals who infect disproportionately more secondary cases) and identifying potential hotspots for targeted interventions.

Transmission Chains: By counting and analysing transmission chains, counter models can reveal the pathways through which a disease spreads through a population. This insight is crucial for identifying vulnerable groups and designing targeted containment strategies to break chains of transmission.

Impact of Public Health Interventions: Counter models quantify the impact of public health interventions by simulating changes in counter metrics, such as reduced contacts due to social distancing or increased immunity due to vaccination campaigns. This quantification helps in evaluating the effectiveness of interventions and optimizing resource allocation.

Temporal Dynamics: These models are adept at capturing the temporal dynamics of an epidemic, including the incubation period, the infectious period, and time-dependent changes in population

behaviour or immunity. Counters tracking new infections over time can help in identifying peaks, estimating the duration of the epidemic, and understanding the effects of seasonality.

Practical Applications

COVID-19 Pandemic: During the COVID-19 pandemic, counter models were used extensively to project the course of the epidemic under various scenarios, including lockdown measures, mask mandates, and vaccination rollouts. These models provided critical insights into the timing of intervention strategies and their potential to flatten the curve.

HIV/AIDS Epidemic: In the context of the HIV/AIDS epidemic, counter models have been used to track the number of new infections, the impact of antiretroviral therapy (ART) on reducing viral load and transmission, and the effectiveness of preventive measures like pre-exposure prophylaxis (PrEP). These models have informed policy and contributed to the global response aimed at ending the HIV epidemic.

Counter models represent a powerful tool in the epidemiologist's toolkit, offering a structured and quantifiable approach to understanding the dynamics of disease spread. By breaking down complex transmission networks into countable events and interactions, these models provide invaluable insights into the mechanisms driving epidemics and the potential impact of public health interventions. Their application in recent and ongoing epidemics underscores their significance in aiding epidemiological understanding, guiding public health policy, and ultimately controlling the spread of infectious diseases.

15.9 Summary

This unit delved into the intricate world of non-Markovian processes, contrasting them significantly with Markovian processes by highlighting the dependency of future states on the history of a system, rather than solely on its present state. This distinction is crucial for modelling complex systems across various disciplines, from physics to finance, where memory effects and historical dependencies play a pivotal role in the system's evolution.

The exploration began with an introduction to simple multi-dimensional prey and predator models, demonstrating how non-Markovian dynamics could capture interactions within ecosystems more accurately than traditional approaches. This was followed by an in-depth

discussion of non-Markovian processes themselves, including their mathematical properties and how they contrast with Markov processes, particularly in terms of predictability and the handling of past information.

Embedded Markov processes were introduced as a method to simplify or approximate non-Markovian systems. By isolating specific events where the Markov property holds, these embedded processes allow for easier analysis and application of Markovian techniques, offering a bridge between the complexity of non-Markovian dynamics and the simplicity of Markovian models.

The application of non-Markovian processes to population growth provided a practical look at how these models can account for delayed effects and historical dependencies, offering insights into complex population dynamics. This was further illustrated through case studies, including the modelling of disease spread, where non-Markovian processes prove particularly valuable. By incorporating factors like latency periods and variable infectious periods, these models offer a nuanced understanding of epidemic dynamics, surpassing simpler models in predictive power and realism.

Counter models in epidemiology highlighted the utility of non-Markovian processes in quantifying and analysing the spread of diseases. By tracking specific events or interactions, such as the number of infections or recoveries, these models provide detailed insights into epidemic spread and the effectiveness of public health interventions.

Self-assessment questions encouraged reflection and deeper engagement with the content, challenging learners to apply concepts to various scenarios and explore the implications of non-Markovian dynamics in real-world contexts.

The unit concluded with a comprehensive set of references, providing learners with resources to deepen their understanding of non-Markovian processes and explore their applications across different fields further.

Overall, this unit offered a thorough examination of non-Markovian processes, emphasizing their importance in accurately modelling systems where history cannot be ignored. Through theoretical discussions, practical applications, and case studies, learners were equipped with a solid foundation in understanding and applying non-Markovian models to complex real-world problems.

15.9 Self-Assessment Questions

Question 1: Definitions

Q1: Explain the difference between Markovian and Non-Markovian processes. Provide an example of a scenario that could be modelled as a non-Markovian process.

Question 2: Understanding Concepts

Q2: Describe how memory effects in non-Markovian processes can influence the modelling of population dynamics. Why is this significant in the context of ecological modelling?

Q3: Given a system where the future state depends on the cumulative history of past states, how would you determine if the system is better modelled by a non-Markovian process rather than a Markovian process? Illustrate your answer with an example.

Q4: What is an embedded Markov process, and how can it simplify the analysis of non-Markovian systems? Provide an example from queueing theory or inventory management.

Q5: Explain how non-Markovian processes can be used to model the spread of diseases, specifically addressing the incorporation of latency periods and variable infectious periods. How does this approach improve upon simpler models?

Q6: Describe the role of counter models in studying the dynamics of epidemics. What kind of information can these models provide that traditional models might not?

Q7: Consider a non-Markovian process where the rate of change of a population P(t) at time t depends on its size at time t and also on its average size over the past year. Write a differential equation that could represent this scenario.

Q8: How might non-Markovian models be applied to financial markets? Discuss the significance of historical price data and trading volume in modelling asset prices.

Q9: Reflect on the limitations of non-Markovian processes in modelling real-world phenomena. What challenges might researchers face when using these models?

Q10: Identify a recent study or application of non-Markovian processes in any field of your interest. Summarize the objectives and findings of the study.

15.10 References

- R. van Handel, "Stochastic Processes in Physics and Chemistry" (3rd Edition), North-Holland, 2007.
- H.C. Tuckwell, "Introduction to Theoretical Neurobiology: Volume 2, Nonlinear and Stochastic Theories", Cambridge University Press, 1988.
- P. Brémaud, "Markov Chains: Gibbs Fields, Monte Carlo Simulation, and Queues", Springer, 1999.
- J. D. Murray, "Mathematical Biology: I. An Introduction" (3rd Edition), Springer, 2002.
- S.N. Ethier and T.G. Kurtz, "Markov Processes: Characterization and Convergence", Wiley, 2005.
- N.G. van Kampen, "Stochastic Processes in Physics and Chemistry" (3rd Edition), North-Holland, 2007.
- L.E. Kavraki, "Non-Markovian Control in Robotics and Automation: A Survey", IEEE Robotics & Automation Magazine, 2018.
- M. Loecher, "Non-Markovian Effects in Financial Markets", Journal of Economic Dynamics and Control, 2009.
- R.M. Anderson and R.M. May, "Infectious Diseases of Humans: Dynamics and Control", Oxford University Press, 1991.
- H. Kantz and T. Schreiber, "Nonlinear Time Series Analysis", Cambridge University Press, 2004.

These references provide a foundational understanding of non-Markovian processes and their applications across different fields. For the most current research, academic journals and conference proceedings should also be consulted.

15.11 Further Readings

- R. P. Feynman and A. R. Hibbs, "Quantum Mechanics and Path Integrals," Dover Publications, 2010.
- H. Touchette, "The Large Deviation Approach to Statistical Mechanics," Princeton University Press, 2009.
- P. Gaspard, "Chaos, Scattering and Statistical Mechanics," Cambridge University Press, 1998.
- M. Loecher, "Nonlinear Time Series Analysis," Cambridge University Press, 2004.
- N. G. Van Kampen, "Stochastic Processes in Physics and Chemistry," Elsevier, 2007.
- J. D. Murray, "Mathematical Biology," Springer, 2002.
- S. Redner, "A Guide to First-Passage Processes," Cambridge University Press, 2001.
- C. Gardiner, "Stochastic Methods: A Handbook for the Natural and Social Sciences," Springer, 2009.
- E. Scalas, "The Application of Continuous-Time Random Walks in Finance and Economics," Physica A, 2006.
- R. M. Anderson and R. M. May, "Infectious Diseases of Humans: Dynamics and Control," Oxford University Press, 1991.