U.P.Rajarshi Tandon Open University, Prayagraj

PGSTAT – 106 /MASTAT – 106 Stochastic Process

Unit – 1 : Introduction

Block: 1 Markov Dependent Trials or Two State Markov Chain

Unit – 2 : Markov Dependent Trials

Unit – 3 : n-step Transition Probabilities

Unit – 4 : Stationary probability distributions and Expected Number of Visits to a State

Block: 2 Markov Chain with more than two states and Random Walk (Gamblers ruin problem

Unit – 5: n-step transition probabilities and Chapman-Kolmogorov Equations

Unit – 6: First Passage and First Return Probabilities

Unit – 7: Classification of States

Unit – 8 : Random Walk and Gambler's Ruin Problem

Block 3: Poisson Process and Simple Branching Process

Unit – 9: Conditions and derivation of Poisson Process

Unit – 10: Interarrival Time Distributions

Unit – 11: Simple Branching Process Introduction, Probability Generating Function and Moments

Unit – 12: Probability of Extinction of Simple Branching Process

Block: 4 Queuing Process and Martingales

Unit – 13: M/M/1 Queuing Process: Introduction and Steady State Analysis

Unit – 14: Waiting time distributions of M/M/1 Queuing Process

Unit – 15: Martingales: Introduction

Unit- 16: Optimal Sampling Theorem

Course Design Committee

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

Prof. Anup 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.

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

Course Preparation Committee

Prof. Anoop Chaturvedi **Writer**
Department of Statistics
University of Allahabad, Prayagraj
Prayagraj, U.P.

Prof. Shalabh **Editor**
Department of Statistics
IIT, Kanpur, U.P.

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

PGSTAT – 107/ MASTAT – 107 **STOCHASTIC PROCESS**
©UPRTOU
First Edition: ***July 2023***
ISBN :

Printed By:

Blocks & Units Introduction

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

The Unit – 1 *Introduction to Stochastic Process*, introduces the concept of stochastic processes and discusses the related definitions and examples.

The Block 1 *Markov Dependent Trials or Two State Markov Chain* Considers the two state Markov chain and discusses various related distributions, limiting distributions and behaviour of Markov trials.

Unit – 2*Markov Dependent Trials* explains the basic concepts of Markovian property, two state-Markov Chains/ Markov dependent trials and definitions of various terms.

In Unit – 3*n-step Transition Probabilities* the n-step transition probabilities of a two state Markov Chain are derived when (i) the initial probability vector is given, (ii) when the initial probability vector is not given.

The Unit – 4*Stationary probability distributions and Expected Number of Visits to a State* derives the limiting probability distribution of a two-state Markov Chain, discusses the stationarity property, and obtains the results related to expected number of visits to a state.

The Block 2 *Markov Chain with more than two states and Random Walk (Gambler's ruin problem)* Considers the Markov chains with more than two states and discusses various results related to it. The block also considers random walk model as a gambler's ruin problem.

In Unit – 5 *n-step transition probabilities and Chapman-Kolmogorov Equations* the n-step transition probabilities and Chapman Kolmogorov equations for a Markov Chain are derived.

The Unit – 6 *First Passage and First Return Probabilities* focusses on the derivation of first passage and first return probabilities of a Markov Chain and presents various related results.

The Unit – 7 *Classification of States* discusses classification of states such as periodic, aperiodic states, the property of ergodicity, recurrent or transient states etc. and various results related to them.

In Unit – 8 *Random Walk and Gambler's Ruin Problem* we discuss gambler's ruin problem and derive the results related to probability of ruin.

The Block 3 *Poisson Process and Simple Branching Process* covers two different topic, the Poisson process and simple branching process.

The Unit – 9 *Conditions and derivation of Poisson Process* defines Poisson process, discusses its various conditions, and provides the derivation of the Poisson process.

In Unit – 10 *Interarrival Time Distributions* the derivations of various results related to interarrival time distributions are given.

The Unit – 11 *Simple Branching Process Introduction, Probability Generating Function and Moments* defines simple Branching process and gives definitions of various terms. The probability generating function of the process and its moments are derived.

In Unit – 12 *Probability of Extinction of Simple Branching Process* the probability of extinction and various results related to the probability of extinction of the simple Branching Process are derived.

The Block 4 *Queuing Process and Martingales* covers two different topics, the Queuing process and Martingales.

In Unit – 13 *M/M/1 Queuing Process: Introduction and Steady State Analysis* the simple M/M/1 queuing process is introduced and the definitions of various terms are given. The steady state analysis of the M/M/1 queuing model is also presented.

The Unit – 14 *Waiting time distributions of M/M/1 Queuing Process* derives the waiting time distribution and different results related to waiting time distribution of the M/M/1 queuing process.

The Unit – 15 *Martingales: Introduction* defines Martingales explains with several examples.

In Unit- 16 *Optimal Sampling Theorem* the derivation of optimal sampling theorem is given and it has been explained with several examples.

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

Unit – 1: Introduction to Stochastic Processes

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:

(i) T countable, S countable
(ii) T countable, S uncountable
(iii) T uncountable, S countable
(iv) T uncountable, S uncountable

Examples:

(i) 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$S = \{1,2, \dots .., 6\}; T = \{1,2,3, \dots \dots .\}$. Both S and T are countable.

(ii) 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 $S = \{1,2,3, \dots .\}$. Hence S is countable while$T = (0, \infty)$ is uncountable.

(iii) 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.

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

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

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

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

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

(ix) 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.

(x) $\{N_t, t \in T\}$, 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.

Unit –2:Markov Dependent Trials

Example 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

$$X_n = \begin{cases} 1 \; if \; outcome \; of \; the \; n \; th \; trail \; is \; S \\ 2 \; if \; outcome \; of \; the \; nth \; trail \; is \; F \end{cases} (1)$$

Then $\{X_n, n = 1,2, \ldots\}$is a stochastic process.

Further

$$P\{X_{n+1} = j_{n+1} | X_1 = j_1, \ldots, X_n = j_n\}$$
$$= P\{X_{n+1} = j_{n+1}\}, \text{(because differnt trails are independent).}$$

$j_r = 1,2; \; r = 1, \ldots., n$. 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 $\{X_n, n = 0,1,2, \ldots\}$, such that $P\,(X_n = 1) = p \; and \; P(X_n = 0) = q\; (= 1 - p), \forall\, n = 0,1,2, \ldots$Further $n = 0,1,2 \ldots$ and for each possible value of $j_0, j_1, \ldots j_n, j_{n+1}$, we have

$$P(X_{n+1} = j_{n+1} | X_0 = j_0, X_1 = j_1, \ldots \ldots \ldots, X_n = j_n,)$$
$$= P(X_{n+1} = j_{n+1} | X_n = j_n) (1)$$

Then $\{X_n, n = 0,1,2, \ldots\}$ 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 ofthe first $(n-1)$trials as "PAST", then the property (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 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

$$p^{(0)} = \left(p_1{}^{(0)}, p_2{}^{(0)}\right)$$

Let

$$p_n(S) = p_1^{(n)} = P(X_n = 1) \text{Probability of } S \text{ at the} n^{th} \text{trial}$$

$$p_n(F) = p_2^{(n)} = P(X_n = 2): \text{ Probability of } F \text{ at the} n^{th} \text{trial}$$

$$= 1 - p_1^{(n)}$$

$p^{(n)} = \left(p_1{}^{(n)}, p_2{}^{(n)}\right)$

If we write

$$p_{11}^{(n)} = P(X_n = 1 | X_0 = 1)$$

$$p_{12}^{(n)} = P(X_n = 2|X_0 = 1)$$

$$= 1 - p_{11}^{(n)}$$

$$p_{22}^{(n)} = P(X_n = 2|X_0 = 2)$$

$$p_{21}^{(n)} = P(X_n = 2|X_0 = 1)$$

$$= 1 - p_{22}^{(n)}$$

Then the matrix

$$P^{(n)} \begin{pmatrix} p_{11}^{(n)} & p_{12}^{(n)} \\ p_{21}^{(n)} & p_{22}^{(n)} \end{pmatrix}$$

is called the n-step transition probability matrix.

Unit – 3: n-step Transition Probabilities

The following theorem derives the n-step transition probabilities of a two-state Markov Chain when the initial probability vector is given.

Theorem1: 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

$$\begin{aligned}
& p_n(S) = P(X_n = 1) \\
& = \mathrm{P}(\mathrm{X_n} = 1, \mathrm{X_{n-1}} = 1) + \mathrm{P}(\mathrm{X_n} = 1, \mathrm{X_{n-1}} = 2) \\
& = P(X_n = 1 | X_{n-1} = 1) P\,(X_{n-1} = 1) + P\,(X_n = 1 | X_{n-1} = 2) P(X_{n-1} = 2) \\
& = p_{11} p_{n-1}(S) + p_{21}\;\; p_{n-1}(F) \\
& = p_{11} p_{n-1}(S) + p_{21}\,[1 - p_{n-1}(S)] \\
& = p_{11} p_{n-1}(S) + (1-p_{22})[1 - p_{n-1}(S)] \\
& = a\, p_{n-1}(S) + b
\end{aligned}$$

where $a = p_{11} + p_{22} - 1, \;\; b = 1 - p_{22}$.

Writing $p_n = p_n(S)$, we get the difference equation

$$p_n = a\, p_{n-1} + b\,, n \ge 1 \qquad (2)$$

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

$$p_n = u_n + \frac{b}{1-a}, \qquad n = 0,1,2 \ldots \tag{3}$$

Hence from (2) and (3), 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

$$\begin{aligned} p_n &= p_n(S) \\ &= u_n + \frac{b}{1-a} \\ &= a^n u_0 + \frac{b}{1-a} \\ &= a^n \left[p_0(S) - \frac{b}{1-a}\right] + \frac{b}{1-a} \\ &= (p_{11} + p_{22} - 1)^n \left\{p_0{}^{(S)} - \frac{1-p_{22}}{2-p_{11}-p_{22}}\right\} + \frac{1-p_{22}}{2-p_{11}-p_{22}} \end{aligned}$$

Interchanging the roles of S and F, we obtain

$$\begin{aligned} p_n(F) &= (p_{11} + p_{22} - 1)^n \left\{p_0{}^{(F)} - \frac{1-p_{11}}{2-p_{11}-p_{22}}\right\} + \frac{1-p_{11}}{2-p_{11}-p_{22}} \\ &= 1 - p_n(S). \end{aligned}$$

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 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

$$\begin{aligned}
p_{11}^{(n)} &= P(X_n = 1|X_0 = 1) \\
&= P(X_n = 1, X_{n-1} = 1|X_0 = 1) + P(X_n = 1, X_{n-1} = 2|X_0 = 1) \\
&= P\{X_n = 1|X_{n-1} = 1\}P\{X_{n-1} = 1|X_0 = 1\} \\
&\qquad + P\{X = 1|X_{n-1} = 2\}P\{X_{n-1} = 2|X_0 = 1\} \\
&= p_{11}p_{11}^{(n-1)} + p_{21}p_{12}^{(n-1)} \\
&= p_{11}p_{11}^{(n-1)} + (1 - p_{21})\left[1 - p_{11}^{(n-1)}\right] \\
&= a\, p_{11}^{(n-1)} + b \qquad (4)
\end{aligned}$$

where $a = p_{11} + p_{22} - 1, \; b = 1 - p_{22}$

For solving this difference equation (4), we write

$$p_{11}^{(n)} = u^{(n)} + \frac{b}{1-a}, n \geq 1$$

so that (4) reduces to

$$u^{(n)} = au^{(n-1)} = a^2 u^{(n-2)} \ldots\ldots a^{n-1} u^{(1)} = a^{n-1}\left[p_{11}^{(1)} - \frac{b}{1-a}\right]$$

Hence

$$p_{11}^{(n)} = a^{n-1}\left[p_{11}^{(1)} - \frac{b}{1-a}\right] + \frac{b}{1-a}$$

$$= (p_{11} + p_{22} - 1)^{n-1}\left[p_{11} - \frac{1-p_{22}}{2-p_{11}-p_{22}}\right] + \frac{1-p_{22}}{2-p_{11}-p_{22}}, \left(p_{11}^{(1)} = p_{11}\right)$$

$$= \frac{(p_{11} + p_{22} - 1)^n (1-p_{11})}{2-p_{11}-p_{22}} + \frac{1-p_{11}}{2-p_{11}-p_{22}} \quad (5)$$

Interchanging the roles of S and F, we obtain

$$p_{22}^{(n)} = \frac{(p_{11} + p_{22} - 1)^n (1-p_{22})}{2-p_{11}-p_{22}} + \frac{1-p_{11}}{2-p_{11}-p_{22}} \quad (6)$$

Further

$$p_{12}^{(n)} = 1 - p_{11}^{(n)}$$

$$= -\frac{(p_{11} + p_{22} - 1)^n (1-p_{22})}{2-p_{11}-p_{22}} + \frac{1-p_{11}}{2-p_{11}-p_{22}} \quad (7)$$

$$p_{21}^{(n)} = 1 - p_{22}^{(n)}$$

$$= -\frac{(p_{11} + p_{22} - 1)^n (1-p_{22})}{2-p_{11}-p_{22}} + \frac{1-p_{22}}{2-p_{11}-p_{22}} \quad (8)$$

Combining (5), (6) (7) and (8) 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■

Unit – 4: Stationary probability distributions and Expected Number of Visits to a State

First, we derive the limiting n-step transition probability distribution as $n \to \infty$.

Theorem 3: If $|p_{11} + p_{22}-1|<1$

$$\lim_{n\to\infty} P^{(n)} = A = \begin{bmatrix}\pi_1 & \pi_2\\ \pi_1 & \pi_2\end{bmatrix} \qquad (9)$$

where

$$\pi_1 = \frac{(1-p_{22})}{(2-p_{11}-p_{22})}, \pi_2 = \frac{(1-p_{11})}{(2-p_{11}-p_{22})}. \qquad (10)$$

Proof: Since $\lceil p_{11} + p_{22} - 1\rceil < 1$, 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 3 we see that

$$\lim_{n\to\infty} p_{11}^{(n)} = \lim_{n\to\infty} p_{21}^{(n)} = \pi_1, \text{and} \lim_{n\to\infty} p_{22}^{(n)} = \lim_{n\to\infty} p_{12}^{(n)} = \pi_2$$

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 probabilitydistribution 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 $p_{11} = p_{22}$

$$(\pi_1 =) \lim_{n\to\infty} p_n(S) = \frac{1}{2} = \lim_{n\to\infty} p_n(F) = \frac{1}{2} (= \pi_2)$$

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:

$$\left.\begin{array}{l} a_1 = a_1 p_{11} + a_2 p_{21} \\ a_2 = a_1 p_{12} + a_2 p_{22} \end{array}\right\} (11)$$

Suppose $P(X_0 = 1) = a_1, P(X_0 = 2) = a_2$, where a_1, a_2 satisfy (11),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

$$P(X_n = 1) = a_1, P(X_n = 2) = a_2 \ \forall n \geq 0.$$

Theorem 4:The limiting probabilitydistribution $\pi = (\pi_1, \pi_2)$ 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 (11) holds for the probability distribution π, so that $\pi = (\pi_1, \pi_2)$ is a stationary probability distribution for the Markov Chain■

Theorem 5: 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

$$\pi_1 = \frac{(1 - p_{22})}{(2 - p_{11} - p_{22})}, \pi_2 = \frac{(1 - p_{11})}{(2 - p_{11} - p_{22})} \begin{bmatrix} \pi_1 p_{11} + \pi_2 p_{21} = \pi_1, \\ \pi_2 p_{12} + \pi_2 p_{22} = \pi_2 \\ \pi_1 + \pi_2 = \pi_1 \end{bmatrix}$$

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■

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 Chainmakes to state j starting initially in state i, in the first n transitions.

Let

$$\mu_{ij}^{(n)} = E\left(N_{ij}^{(n)}\right)$$

Theorem 6: 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 be $\{X_0, X_1, \dots\dots\}$ a two state Markov Chain. Define a random variable

$$y_{ij}^{(m)} = \begin{cases} 1 & if\ X_m = j, X_0 = i \\ 0 & if\ X_m \neq j, X_0 = i \end{cases};\ m = 1,2,\dots$$

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

$$N_{ij}^{(n)} = y_{ij}^{(1)} + y_{ij}^{(2)} + \cdots \dots \dots y_{ij}^{(n)}$$

$$= \sum_{m-1}^{n} y_{ij}^{(m)}$$

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 probabilitythat 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.

Block: 2 Markov Chain with more than two states and Random Walk (Gamblers ruin problem):

In this block we will discuss the (i) Markov processes with more than two states and (ii) gambler's ruin problem as a random walk model.

Unit – 5: n-step transition probabilities and Chapman-Kolmogorov Equations

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 2: 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 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 $\{X_n; n = 0,1,2,3, \ldots\ldots\}$ be a stochastic process with X_n taking discrete values 1,2,3, ...

Definition: The stochastic process $\{X_n; n = 0,1,2,3, \ldots\ldots\}$ is called a Markov chain if for $n = 1,2, \ldots;\ i_0, i_1, i_2, \ldots i_{n-1}, j \in S$,

$$P\{X_n = j | X_{n-1} = i_{n-1}, \ldots, X_0 = i_0\}$$
$$= P\{X_n = j | X_{n-1} = i_{n-1}\}.$$

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.

The probability $p_{ij} = p\{X_n = j | X_{n-1} = i\}, i, j \in S$ is called the (one step) transition probability$i \rightarrow$j at time n. the transition probabilities may or may not be independent of n. if the transition probabilityp_{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 $S = \{1,2,3, \ldots\}$. Then $p_{ij} \geq 0\ \forall\ i, j \in S\ and\ \sum_{j \in S} p_{ij} = 1 \forall i \in S$. The matrix

$$P = \begin{bmatrix} p_{11} & p_{12} & p_{13} & \ldots \\ p_{21} & p_{22} & p_{23} & \ldots \\ \vdots & \vdots & \vdots & \cdots \end{bmatrix}$$

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.

P is a stochastic matrix.

Let

$$p_j^{(n)} = P\ (X_n = j); n = 0,1,2, \dots j \in S = \{1,2, \dots .\}$$

$$p_j^{(0)} = p\ (X_0 = j); j \in S: \text{initial probability distribution}$$

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

$$P^{(n)} = \begin{bmatrix} p_{11}^{(n)} & p_{12}^{(n)} & \cdots \\ p_{21}^{(n)} & p_{22}^{(n)} & \cdots \\ \vdots & \vdots & \cdots \end{bmatrix}$$

is called the n −step TPM of the MARKOV CHAIN.

Higher Transition probabilities:

Chapman – Kolmogorov Equation:

For obtaining the $n -$ step transition probabilities, we have

$$p_{ij}^{(n)} = P\ \{X_n = j | X_0 = i\}$$

$$= \sum_{r \in S} P\ [X_n = j, X_{n-1} = r |\ X_0 = i] (S = \{1,2,3, \dots\})$$

$$= \sum_{r=1}^{\infty} P\ [X_n = j | X_{n-1} = r, X_0 = i]\ P[X_{n-1} = r |\ X_0 = i]$$

$$= \sum_{r=1}^{\infty} P\,[X_n = j | X_{n-1} = r]\; P[X_{n-1} = r |\, X_0 = i]$$

$$= \sum_{r=1}^{\infty} p_{ir}^{(n-1)} p_{rj} \qquad (1)$$

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 (1) in matrix notation as

$$\begin{bmatrix} p_{11}^{(n)} & p_{12}^{(n)} & \cdots \\ p_{21}^{(n)} & p_{22}^{(n)} & \cdots \\ \vdots & \vdots & \cdots \end{bmatrix} = \begin{bmatrix} p_{11}^{(n-1)} & \cdots \\ \vdots & \cdots \\ \vdots & \cdots \end{bmatrix} \begin{bmatrix} p_{11} & p_{12} & p_{13} & \cdots \\ p_{21} & p_{22} & p_{23} & \cdots \\ \vdots & \vdots & \vdots & \cdots \end{bmatrix}$$

or

$$P^{(n)} = P^{(n-1)} P$$

$= P^{(n-2)} P^2$

$\vdots$

$= P^n$

Thus

$$P^{(n)} = P^n (2)$$

Eq. (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

$$p_{ij}^{(m+n)} = \sum_r p_{ir}^{(m)}\, p_{rj}^{(n)}$$

$$= \sum_r p_{ir}^{(n)}\, p_{rj}^{(m)}, \quad (i,j) \in S. \qquad (3)$$

The set of equations (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 thatX_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 p.d. of X_i is

$$P(X_i = r) = \frac{1}{3} \quad r = 1,2,3$$

Further

$$P\left(X_i = r, X_j = s\right) = \frac{1}{9} \quad r, s = 1,2,3$$

$$P(X_1 = r, X_2 = s, X_3 = t) = \begin{cases} \frac{1}{9}, & r,s,t = 1,2,3;\; r = s = t \; or \; r \neq s \neq t \\ 0 \; if \; r = s \neq t \; or \; r = t \neq s \; or \; r \neq t = s \end{cases}$$

Hence

$$P\left(X_i = r, X_j = s\right) = P(X_i = r)P\left(X_j = s\right) = \frac{1}{9}$$

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

$$p_{ij}^{(1)} = p_{ij} = P\ [X_{m+1} = j | X_m = i]$$

$$= P\ [X_{m+1} = j] = \frac{1}{3} (\text{since} X_m, X_{m+1} \text{are pairwise independent})$$

$$p_{ij}^{(2)} = P[X_{m+2} = j | X_m = i] = \frac{1}{3}$$

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 in eqestion.

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.

For obtaining the vector of State occupation probabilities at time n,

$$p^{(n)} = \left(p_1^{(n)}, p_2^{(n)}, \ldots .\right)$$

we have

$$\begin{aligned}
p_j^{(n)} &= P(x_n = j)(n = 0,1, \ldots ., j = 1,2 \ldots) \\
&= \sum_r P(x_n = j, x_{n-1} = r) \\
&= \sum_r P(x_n = j | x_{n-1} = r) P(x_{n-1} = r) \\
&= \sum_r p_{rj}\, p_r^{(n-1)} \\
&= \sum_r p_r^{(n-1)} p_{rj} \quad (4)
\end{aligned}$$

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 (4) as

$$p^{(n)} = p^{(n-1)}P \quad (5)$$

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 $p^{(n)}$.

Unit – 6: First Passage and First Return Probabilities

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.

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 ≥ 2

$$p_{jj}^{(n)} = P\;[X_n = j|X_0 = j]$$

$$= \sum_{r=1}^{n} P[X_1 \neq j, \ldots X_{r-1} \neq j, X_r = j|X_0 = j]P[X_n = j|X_r = j]$$

$$= \sum_{r=1}^{n} f_{jj}^{(r)} p_{jj}^{(n-r)} \left(p_{jj}^{(0)} = P\;[X_0 = j|X_0 = j] = 1\right)$$

$$= f_{jj}^{(n)} p_{jj}^{(0)} + \sum_{r=1}^{n-1} f_{jj}^{(r)} p_{jj}^{(n-r)}$$

$$= f_{jj}^{(n)} + \sum_{r=1}^{n-1} f_{jj}^{(r)} p_{jj}^{(n-r)}$$

Or

$$f_{jj}^{(n)} = p_{jj}^{(n)} - \sum_{r=1}^{n-1} f_{jj}^{(r)} p_{jj}^{(n-r)} ; n = 2,3, \dots \qquad (6)$$

From (6),$f_{jj}^{(2)} f_{jj}^{(3)} \dots$ can be calculated recursively.

Similarly

$$p_{jk}^{(n)} = \sum_{r=1}^{n-1} f_{jk}^{(r)} p_{kk}^{(n-r)} \text{ (verify it)}$$

So that

$$f_{jk}^{(n)} = p_{jk}^{(n)} - \sum_{r=1}^{n-1} f_{jk}^{(r)} p_{kk}^{(n-r)} ; \quad n = 2,3 \dots ..$$

Notice that $n = 1$ $f_{jk}^{(1)} = p_{jk}$

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.

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) $p_k = \frac{1}{k!}\frac{d^k}{ds^k}P(s)|_{s=0}$ $\quad k = 0,1,2,...$

(ii) $E(X) = \frac{d}{ds}P(s)|_{s=1} = P^{'}(1)$

$E[X(X-1)] = P^{''}(1)$

In general, for $r = 1,2,...$

$E[X(X-1) \dots \dots (X-r+1)] = P^{(r)}(1)$

(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

$$g(s) = E(s^{X_i}), \qquad i = 1,2,...$$

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

$$\frac{d^k}{ds^k}s^k = k(k-1)(k-2)\dots 1 = k!$$

For $r < k$,

$$\frac{d^k}{ds^k} s^r = 0$$

For $r > k$,

$$\frac{d^k}{ds^k} s^r = r(r-1)\dots(r-k+1)s^{r-k},$$

Which tends to 0 as $s \to 0$. Hence

$$\frac{d^k}{ds^k} P(s)|_{s=0} = p_k k!$$

$$or\ p_k = \frac{1}{k!}\frac{d^k}{ds^k} P(s)|_{s=0}.$$

(ii) We observe that

$$\frac{d^r}{ds^r} P(s) = \sum_{n=r}^{\infty} p_n\, n(n-1)\dots(n-r+1)s^{n-r}$$

Taking limit $s \to 1$, we obtain

$$\left.\frac{d^r}{ds^r} P(s)\right|_{s=1} = P^{(r)}(1)$$

$$= \sum_{n=r}^{\infty} p_n\, n(n-1)\dots(n-r+1)$$

$$= E[X(X-1)\dots\dots(X-r+1)]$$

(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

$P(s) = E\left(s^{(X+Y)}\right)$

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

$= E(s^X)E(s^Y)$ (since X and Y are independently distributed)

$= P_1(s).P_2(s)$

(iv) We can easily verify that

$$\lim_{s \to 1\cdot-} P(s) = P(1)$$

$$= \sum_{n=0}^{\infty} p_n = 1$$

(v) We have

$$G(s) = E[s^{Y_N}]$$

$$= E[E(s^{YN}|N)]$$

$$= E[E\{s^{X_i} \ldots \ldots s^{X_N}|N\}]$$

$$= E[E(s^{X_1}) \ldots \ldots E(s^{X_N})|N]$$

$$= E[g(s)^N]$$

$$= h[g(s)].$$

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 7: We have

$$P_{jk}(s) = F_{jk}(s) P_{kk}(s); (j \neq k) \qquad (7)$$

$$P_{jj}(s) = \frac{1}{1 - F_{jj}(s)}. \qquad (8)$$

Proof.Let us define

$$\delta_{jk} = \begin{cases} 1 \; if \; j = k \\ 0 \;\; if \; j \neq k \end{cases}$$

We observe that

$$P_{jk}(s) = \sum_{n=0}^{\infty} p_{jk}^{(n)} s^n$$

$$= p_{jk}^{(0)} + \sum_{n=0}^{\infty} p_{jk}^{(n)} s^{(n)} \quad (p_{jk}^{(0)} = 1 \text{ if } j = k \text{ and } 0 \text{ if } j \neq k \text{ or} p_{jk}^{(0)} = \delta_{jk})$$

$$= \delta_{jk} + \sum_{n=0}^{\infty} \left\{ \sum_{m=1}^{n} f_{jk}^{(m)} p_{kk}^{(n-m)} \right\} s^{n-m+m}$$

$$= \delta_{jk} + \sum_{m=1}^{\infty} f_{jk}^{(m)} s^m \sum_{n=m}^{\infty} s^{n-m} p_{kk}^{(n-m)}$$

$$= \delta_{jk} + \sum_{m=1}^{\infty} f_{jk}^{(m)} s^m \sum_{u=0}^{\infty} s^u p_{kk}^{(u)}$$

$$= \delta_{jk} + \sum_{m=1}^{\infty} f_{jk}^{(m)} s^m \quad p_{jj}(s)$$

$$= \delta_{jk} + F_{jk}(s) P_{kk}(s)$$

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

$$P_{jj}(s) = 1 + F_{jj}(s) P_{jj}(s)$$

$$\text{or} P_{jj}(s) = \frac{1}{1 - F_{jj}(s)}.$$

Hence the theorem follows■

Theorem 8: 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 9: 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

$$\sum_{n=0}^{\infty} p_{jk}^{(n)} = P_{jk}(1) = F_{jk}(1)P_{kk}(1)$$

$$= F_{jk}(1)P_{kk}(1)$$

$$= f_{jk} P_{kk}(1) \leq P_{kk}(1)(\text{since} f_{jk} \leq 1)$$

$$= \sum_{n=0}^{\infty} p_{kk}^{(n)} < \infty, \text{since the} k^{th} \text{state is transient.}$$

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 $\sum_{n=0}^{\infty} p_{jk}^{(n)}$.

Unit – 7: Classification of States

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 $i \to j$.

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 10: 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 11: 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 $C_1 \cap C_2 = \emptyset$ 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 $C(i) \subset C(j)$. or $C_1 = C_2$. This proves the theorem■

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 divisior 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.

Definition: A recurrent, non-null and a periodic 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 12: 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 13: 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 αso 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■

Stability of a Markov Chain:

Stationary Distribution: For a Markov Chain with transition probability $\{p_{jk}\,; j, k \in S\}$, a probability distribution $\{u_j\}$ is called *stationary (or invariant)* if

$$u_k = \sum_j u_j p_{jk}\,.\left(u_j \geq 0, \sum_j u_j = 1\right)$$

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

$$u_k = \sum_j u_j\, p_{ik}^{(n)}, n \geq 1.$$

Unit – 8 : Random Walk and Gambler's Ruin Problem

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 $\{0,1,2,\ldots,a\}$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

$$\mu_k = p\,\mu_{k+1} + q\mu_{k-1};\ \ k = 2,3,\ldots,a-2 \qquad (1)$$

$$\mu_1 = q + p\,\mu_2 (2)$$

$$\mu_{a-1} = q\,\mu_{a-2} (\mu_a = 0) (3)$$

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

$$\left.\begin{aligned} 1 &= A + B \\ 0 &= A + B\left(\frac{q}{p}\right)^a \end{aligned}\right\}$$

$$\Longrightarrow B = -\frac{1}{\left(\frac{q}{p}\right)^a - 1}$$

$$A = \frac{\left(\frac{q}{p}\right)^a}{\left(\frac{q}{p}\right)^a - 1}.$$

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:

$$\nu_k = \frac{\left(\frac{q}{p}\right)^k - 1}{\left(\frac{q}{p}\right)^a - 1} \quad (9)$$

We can easily obtain ν_k byreplacing 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.

Unit – 9: Conditions and derivation of Poisson 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 $(0,t]$ 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 $(T, T + t)$ 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 $t_1 < t_2 \dots t_n, N_{t_1}, N_{t_2} - N_{t_1}, \dots, N_{t_n} - N_{t_{n-1}}$ 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

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

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

Where of $P_k(h)$ 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).

Theorem1: Under the conditions (i), (ii), (iii) and (iv), $N(t)$ follows a Poisson distribution with meanλt, *i.e*,$P_n(t)$ is given by.

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

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:

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

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 $(t, t+h]$

$\vdots$

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

Now

$$P(A_1) = P[N(t) = n]\, P\,[N(h) = 0 | N\,(t) = n]$$
$$= P_n(t)P_0(h)\big(\text{from (ii)}\big)$$

$$P(A_2) = P[N(t) = n-1] P\,[N(h) = 1 | N\,(t) = n-1]$$
$$= P_{n-1}(t)P_1(h)$$
$$\vdots$$
$$P(A_{n+1}) = P_0(t)P_n(h)$$

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

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

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

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

$$= P_{>1}(h) = o(h) \big(\text{By condition (iii)}\big)$$

Hence

$$P_n(t+h) = P_n(t) P_0(h) + P_{n-1}(t) P_1(h) + o(h) \qquad (2)$$

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 (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

$$\frac{d}{dt}P_n(t) = P_n^{'}(t) = \lambda[P_{n-1}(t) - P_n(t)]; n \geq 1 \qquad (3)$$

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

$$\frac{P_0(t+h) - P_0(t)}{h} = \lambda P_0(t) + o(h) \qquad (4)$$

As $h \to 0$, (4) reduces to

$$P_n^{'}(t) = -\lambda P_0(t)$$

or

$$\frac{d}{dt}\log P_0(t) = -\lambda \qquad (5)$$

$$\text{or } log P_0(t) = -\lambda t + K \qquad (6)$$

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

$$P_0(t) = e^{-\lambda t} \tag{7}$$

For $n = 1$

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

or

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

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

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

Hence

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

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

$$P_1(t) = \lambda t\, e^{-\lambda t} = \frac{(\lambda t)^1 e^{-\lambda t}}{1!} \tag{8}$$

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

$$P_{k-1}(t) = \frac{(\lambda t)^{k-1} e^{-\lambda t}}{(k-1)!} \tag{9}$$

Then, for $n = k$, the equation (3) becomes

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

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

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

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

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

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

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: 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, ∃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 $n \rightarrow \infty$ so that

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

and we obtain

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

*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

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

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

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

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

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

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

Thus (i), (ii), (iii) imply (iv)■

Unit – 10: Interarrival Time Distributions

Theorem 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

$$P\{X > x\} = P\ \{E_{i+1}\text{did not occur in } (t_i, t_i + x) |\ E_i \text{occured at time} t_i\}$$
$$= P\ \{\text{no event occurs in interval } (t_i, t_i + x) |\ N\ (t_i) = i\}$$
$$= P\ \{N(x) = 0 |\ N\ (t_i) = i\}$$
$$= \ P_0(x) = \ e^{-\lambda t};\ x > 0.$$

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 4: The inter arrival 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 5: 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 exponential distribution with mean $1/\lambda$ and let $Z_1, Z_2, \ldots$ be iid random variables having exponential distribution with mean $1/\lambda$. Then sum $W_n = \sum_{i=1}^{n} Z_i$ is the waiting time upto the n^{th} occurrence, i.e., the time form 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

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

$$P\{N(t) < n\} = P\{W_n = Z_1 + \cdots + Z_n > t\}$$
$$= 1 - P\{W_n \leq t\}.$$

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

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

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

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

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

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

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

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

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

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

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

Proof: We have

$$P[\gamma \leq t] = P\ [\text{The event occurs one time before the time } t]$$
$$= P\ [N(t) = 1]$$
$$= e^{-\lambda t}\ \lambda t$$
$$P[N(T) = 1] = e^{-\lambda T}\ \lambda T$$

and

$$P[N(T) = 1 | \gamma \leq t]$$
$$= P[\text{event does not occur in interval}(t, T)]$$
$$= e^{-\lambda(T-t)}$$

Therefore,

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

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

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

$$= G_\gamma[t|N(T) = 1] \quad \text{(say)}$$

$G_\gamma[t|N(T) = 1]$ 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■

Unit – 11: Simple Branching Process Introduction, Probability Generating Function and Moments

Galton and Watson (1874) developed a mathematical model for the problem of extinction of families.

Let $p_i$be the probability that a man produces i sons let each son has the same probability distribution for sons of his own and so on. What is the probability that the male line is the probability distribution of the number of descendants in the n^{th} generation.?

The simple Branching Process has wide applications in the problems where one is concerned with objects (or individuals) that can generate objects of similar kind; such objects may be biological entities, such as human beings, animals, genes, bacteria and so on, which generate similar objects by biological methods or may be physical particles such as neutrons which yield new neutrons under a nuclear chain reaction. We can say that Branching processes are used to model reproduction.

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 off springs reproduced or the object generated by the objects of the 0^{th} generation are the direct descendant of the ancestors and are said to form the 1^{st} generation; the objects generated by these 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 descendents 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 variable's $\{x_n;\ n= 0,1,2,\ldots.\}$ constitutes a Galton-Watson (G.W.) branching process with off spring distribution {pk; k=0,1,2,……}

Probability Generating Function (pgf) of the Branching Process:

Let

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

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

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

without loss of generality, we assume that $X_0 = 1$, *i.e.,* the process starts with on individuals. Then

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

Theorem 7: We have

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

$$g_n(s) = g\,[g_{n-1}(s)] \qquad (2)$$

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 (1).

Substituting $n = 2,3, \dots$ in (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

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

$$= g_{n-2}\big[g\big(g(s)\big)\big]$$

$$= g_{n-2}\big(g_2(s)\big)$$

$$= g_{n-3}\left(g\big(g_2(s)\big)\right)$$

$$= g_{n-3}\big(g_3(s)\big)$$

$$= \cdots$$

$$= g_{n-k}\big(g_k(s)\big)(k = 0,1,2, \dots, n)$$

For $k = n - 1$

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

This proves result (2) of the theorem■

Moments of X_n*:*

Theorem 8: If we assume that $E(X_1) = \sum_{k=0}^{\infty} k\, p_k \;=\; \mu$ and var (x_1) =σ^2 then,

$$E(X_n) = \mu^n \quad (3)$$

$$Var\ (X_n) = \begin{cases} \dfrac{\mu^{n-1}(\mu^n - 1)}{\mu - 1}\sigma^2 & if\ \mu \neq 1 \\ n\sigma^2 \quad if & \mu = 1 \end{cases} \quad (4)$$

*Proof:*We have

$$g_n(s) = g_{n-1}\big(g(s)\big)(5)$$

Differentiating (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

$$m = g''(1)$$
$$= E[X_1(X_1 - 1)]$$
$$= \ \sigma^2 + \ \mu^2 - \mu.$$

On iterating we obtain

$$g_n''(1) = m\mu^{n-1} + \mu^2[m\mu^{n-2} + \mu^2 g_{n-2}''(1)]$$
$$= m\,(\mu^{n-1} + \mu^n) + \mu^4 g_{n-2}''(1)$$
$$= \cdots$$
$$= m(\mu^{n-1} + \mu^n + \cdots \dots + \mu^{n-2}) + \mu^{2n-2} g_1''(1)$$
$$= m\,\mu^{n-1}(1 + \mu + \cdots \dots + \mu^{n-2}) + \mu^{2n-2}\,m$$
$$= m\,\mu^{n-1}(1 + \mu + \cdots \dots + \mu^{n-2} + +\mu^{n-1})$$
$$= m.n\ \ if\ \mu = 1$$
$$= m\mu^{n-1}\frac{(\mu^{n-1})}{\mu - 1} \qquad if\ \mu \neq 1$$

Hence

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

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

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

and

$$Var\ (x_n) = \sigma^2 n \qquad\qquad if \mu = 1.$$

Hence the theorem follows■

Unit – 12: Probability of Extinction of Simple Branching Process

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 $P\{X_{n+1} = 0 \mid X_n = 0\} = 1$. 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 9(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 the positive root less than unity of the equation

$$g(s) = s \qquad (6)$$

Proof: Let $q_n = P\{X_n = 0\}$. 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

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

q_n: probability that the population starts with one ancestor dies out before the n^{th} generation. Now, if

$$p_0 = P\{X = 0\} = 0, \text{then} X_0 \leq X_1 \leq X_2 \leq \cdots$$

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

$$
\begin{aligned}
&P\{T < n + 1 | X_0 = 1\} \\
&= \; p_0 + p_1 p_0 + p_1^2 p_0 + \cdots + p_1^n \, p_0 \\
&= p_0 \frac{1 - p_1^n}{1 - p_1} \\
&= 1 - p_1^n \rightarrow 1 \text{ as } n \; \rightarrow \; \infty
\end{aligned}
$$

Hence $T < \infty$ almost surely.

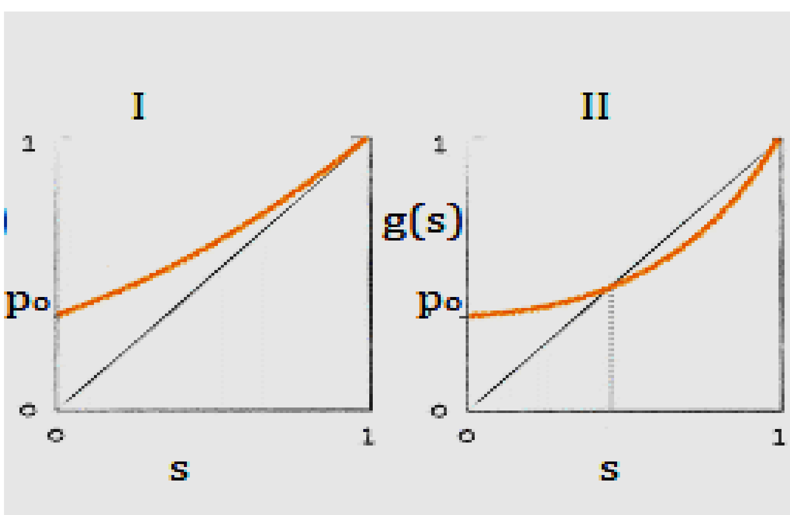
We exclude these trivial cases and assume that $0 < p_0 < p_0 + p_1 < 1$.

Now

$$
\begin{aligned}
&g(s) = \; p_0 + p_1, s + p_2 s^2 + \cdots \ldots; \quad 0 \leq s \leq 1 \\
&g(0) = \; p_0 > 0 \; and \; for \; 0 < s \; \leq 1 \\
&g^{'}(s) > 0 \\
&g^{''}(s) > 0,
\end{aligned}
$$

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

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:



Now

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

substituting $s = 0$,we get

$$g_{n+1}(0) = g\big(g_n(0)\big) \text{or} q_{n+1} = g(q_n) \qquad (7)$$

substituting $n = 0,1,2,\ldots$ respectively in (7), we get

$$q_1 = g\ (0) = p_0 > 0 = q_1 > 0$$

$$q_2 = g\ (q_1)$$

$$> g\ (0) = q_1 (\text{since g(s) is an increasing funciton of s}\}$$

$$\Rightarrow q_1 > q_2$$

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

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

q is the probability of ultimate extinction. From (2) it follows that q satisfies

$$q = g(q) \qquad (8)$$

Thus, the probability of ultimate extinction is a solution of (8).

Let λ be an arbitrary positive root of (8). At leat one such root exists which is λ=1.

Then

$$q_1 = g(0) < g(\lambda) = \lambda \qquad (\lambda \text{ is positive})$$

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

$$q_2 = g(q_1) < g(\lambda) = \lambda \Rightarrow q_2 < \lambda$$

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

Since λ is an arbitrary positive root of (8), it follows that q is the smallest positive root of (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 μ> 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■

Block: 4 Queuing Process and Martingales

Unit – 13: M/M/1 Queuing Process: Introduction and Steady State Analysis

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.

The Simple Queueing Model:

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

M/M/1(∞, FIFO)model.

Input Service Number of Servers

(Markovian) (Poisson process

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 $n=0$

$$\begin{aligned} &p_0(t+\Delta t)\\ &= p_0(t)P\left(\text{no arrival in}(t, t+\Delta t)\right)+ p_1(t)P\left(1\, departure\ in\ (t, t+\Delta t)\right)\\ &\quad+\sum_{k=2}^{\infty} p_k(t)P(k\ departures\ in\ (t, t+\Delta t)\\ &= p_0(t)[1-\lambda\Delta t+o(\Delta t)]+p_1(t)[\mu\Delta t+o(\Delta t)]+o(\Delta t) \end{aligned}$$

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

$$\frac{d}{dt}p_0(t) = \mu p_1(t) - \lambda p_0(t)(1)$$

For $n \geq 1$

$$p_n(t+\Delta t)$$
$$= p_{n-1}(t)P \text{ [one arrival, no departure } in \ \ (t+t+\Delta t)]$$
$$+p_n(t)P \text{ [no arrival, no departure in}(t, t+\Delta t)]$$
$$+p_{n+1}(t)P\text{[no arival, one departure in}(t, t+\Delta t)] + o(\Delta t)$$
$$= p_{n-1}(t)[\lambda\Delta(t) + o(\Delta t)] + p_n(t)[1 - \lambda\Delta(t) - \mu\Delta(t) + o(\Delta t)]$$
$$+p_{n+1}(t)[\mu\Delta(t) + o(\Delta t)] + o(\Delta t)$$

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

$$\frac{d}{dt}p_n(t) = \lambda p_{n-1}(t) - (\lambda+\mu)p_n(t) + \mu p_{n+1}(t)(2)$$

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

$$\left.\begin{matrix} 0 = \mu \, p_1 - \lambda \, p_0 \\ 0 = \ \lambda \, p_{n-1} - (\lambda+\mu)p_n + \mu p_{n+1}; \ (n \geq 1) \end{matrix}\right\}(3)$$

or

$$\left.\begin{aligned} p_1 &= \rho\, p_0 \\ p_{n+1} &= \rho p_n + (p_n - \rho\, p_{n-1});\ \ (n \geq 1) \end{aligned}\right\} (4)$$

where

$$\rho = \frac{\lambda}{\mu} = \frac{\frac{1}{\mu}}{\frac{1}{\lambda}} = \frac{\text{mean service time}}{\text{mean interarrival time}} \ (5)$$

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

$$\begin{aligned} p_0 &= p_0 \\ p_1 &= \rho\, p_0 \\ p_2 &= \rho\, p_1 + (p_1 - \rho\, p_0) \\ &= \rho\, p_1 \\ &= \rho^2 p_0 \\ p_3 &= \rho\, p_2 + (p_2 - \rho\, p_1) \\ &= \rho\, p_2 = \rho^3 p_0 \\ &\vdots \\ p_n &= \rho^n p_0 \end{aligned}$$

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

$$L_s = \sum_{n=0}^{\infty} n\rho^n(1-\rho)$$
$$= \frac{\rho}{1-\rho} = \frac{\lambda}{\mu-\lambda}. \quad (6)$$

The probability that the server is free $= 1 - \rho$.

Unit – 14: Waiting time distributions of M/M/1 Queuing Process

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

$$g\left(\frac{t}{n}\right) = \frac{\mu\, e^{-\mu t}\,(\mu t)^n}{n!}\,(0 < t < \infty)(8)$$

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 theprobability 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*

$$g(w) = \sum_{n=1}^{\infty} h(w|n)p_n$$

$h(w|n)$) 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

$$g(w) = \begin{cases} 0 \quad if\ w < 0 \\ 1 - \rho\ e^{-\mu(1-\rho)w} & if\ w \geq 0. \end{cases}$$

Unit – 15: Martingales: Introduction

ConditionalExpectation:

Let$X_1, X_2, \ldots$ beasequenceofrandomvariablesand$\mathcal{F}_n$ denotestheinformationcont ainedin$X_1, X_2, \ldots, X_n$.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 variable(s) with mean 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

$$E(X_1 + X_2 + \ldots + X_m \mid \mathcal{F}_m)$$
$$= X_1 + X_2 + \cdots + X_m$$
$$= S_m$$

Since $X_{m+1} + \ldots + X_n$is independent of $X_1, X_2, \ldots, X_m$, we get

$$E(X_{m+1} + \ldots + X_n \mid \mathcal{F}_m)$$

$$= E(X_{m+1} + \ldots + X_n) = (n - m)\mu$$

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 ⍰ $= 0$ and $Var(X_i) = E(X_i^2) = \sigma^2$. For $m < n$ we shall have

$$E(S_n^2|\mathcal{F}_m) = E[\{S_m + (S_n - S_m)\}^2|\mathcal{F}_m]$$
$$= E(S_m^2|\mathcal{F}_m) + 2E(S_m(S_n - S_m)|\mathcal{F}_m) + E((S_n - S_m)^2|\mathcal{F}_m)$$

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

$$E(S_m^2|\mathcal{F}_m) = S_m^2$$
$$E((S_n - S_m)^2|\mathcal{F}_m) = E(S_n - S_m)^2$$
$$= Var(S_n - S_m)$$
$$= (n - m)\sigma^2$$

$$E(S_m(S_n - S_m)| \mathcal{F}_m)$$
$$= E\big(S_m(S_n - S_m)\big)$$
$$= S_m E(S_n - S_m)$$
$$= 0$$

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

$$P(X_i = 1) = p,$$
$$P(X_i = 0) = 1 - p$$

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 showthat

$$P(X_i = 1|S_n = k) = \frac{k}{n}$$

Hence

$$E(X_i = 1|S_n) = \frac{S_n}{n}$$

and

$$E(S_m|S_n) = E(X_1|S_n) + \cdots + E(X_m|S_n) = S_n \frac{m}{n}$$

Martingale

*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

1. each M_n is measurable with respect to $X_0, X_1, \ldots, X_n$;
2. 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 □. Let $S_0 = 0$ and for $n > 0, S_n$be the partial sum $S_n = X_1 + \ldots + X_n$, then $M_n = S_n - n$□ is a martingale with respect to $\mathcal{F}_n$(information in $X_1, X_2, \ldots, X_n$). By using Example 1,

$$E(M_{n+1}|\mathcal{F}_n) = E(S_{n+1} - (n+1)\square|\mathcal{F}_n)$$
$$= E(S_{n+1}|\mathcal{F}_n) - (n+1)\square$$
$$= (S_n + \square) - (n+1)\square$$
$$= M_n$$

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

$$P\big(W_{n+1} = 2^n - (2^n - 1)\big|W_n = -(2^n - 1)\big)$$
$$P\big(W_{n+1} = 1\big|W_n = -(2^n - 1)\big)$$
$$= \frac{1}{2}$$
$$P\big(W_{n+1} = -(2^{n+1} - 1)\big|W_n = -(2^n - 1)\big) = \frac{1}{2}$$
$$E[W_{n+1}|\mathcal{F}_n] = \frac{1}{2} \times 1 + \frac{1}{2} \times \big(-(2^{n+1} - 1)\big)$$
$$= -(2^n - 1) = W_n.$$

Therefore$W_n$is a martingale with respect to $\mathcal{F}_n$.

Example 6 Suppose$X_1, X_2, \ldots$, areas 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

$$E(B_{n+1}X_{n+1}| \mathcal{F}_n) = B_{n+1}E(X_{n+1} | \mathcal{F}_n)$$
$$= 0$$

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 f0llows:

> 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,

$$E(M_{n+1}|\mathcal{F}_n) = E((n+3)^{-1}X_{n+1}|X_n)$$
$$= \frac{1}{n+3}\left[X_n + \frac{X_n}{n+2}\right]$$
$$= \frac{X_n}{n+2}$$
$$= M_n$$

Submartingale and Supermartingale

Definition: A process M_n with $E(|Mn| < \infty)$ 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 submartingale is a game in one’s favor and a supermartingale 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 submartingale and a supermartingale.

Unit- 16: 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 $P(T < \infty) = 1$,

$$E(|M_n| < \infty)(6)$$

and

$$\lim_{n\to\infty} E\left(|M_n| I(T > n)\right) = 0 \qquad (7)$$

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

Proof: Let F_n be the information contained in $X_0, X_1, \cdots, X_n$ and $I(T > n)$, 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 $j \le (K-1), M_j I(T = j)$ 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 $T_n = min(T, n)$

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

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

Since T_n is a bounded stopping time, hence $E\left(M_{T_n}\right) = M_0$, the $P(T > n) \to 0$ 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 $\lim_{n\to\infty} E\left(|M_T|I(T > n)\right)$, 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 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 $T = min\{j : X_j = 0$ 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 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

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

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 (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

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

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