
Lecture 3 Markov Branching Process

Let $Z(t)$ be number of particles at time t . Let

$$\delta_{1k} + a_k h + o(h), \quad k = 0, 1, \dots$$

represent the probability that a single particle will split producing k particles during a small time interval $(t, t+h)$ of length h . δ_{1k} denotes the Kronecker delta function. Assume that $a_1 \leq 0, a_k \geq 0$ for $k = 0, 2, 3, \dots$ and $\sum_{k=0}^{\infty} a_k = 0$. We further postulate that individual particles act independent of each other, always governed by the infinitesimal probabilities. Note that we are also assuming time homogeneity for the transition probabilities since a_k is not a function of the time at which the conversion or splitting occurs. Each particle lives a random length of time following an exponential distribution with mean $1/\lambda = a_0 + a_2 + a_3 + \dots$. On completion of i -th lifetime, it produces a random number D of descendants of like particles. The probability mass function of D is

$$P(D = k) = \frac{a_k}{a_0 + a_2 + a_3 + \dots}, \quad k = 0, 2, 3, \dots$$

The lifetime and progeny distribution of separate individuals are independent and identically distributed. Using the independence assumptions, we get

$$P(Z(t+h) = n+k-1 \mid Z(t) = n) = na_k h + o(h), \quad k = 0, 2, 3, \dots$$

and

$$P(Z(t+h) = n \mid Z(t) = n) = 1 - na_1 h + o(h).$$

Probability Generating Function

Let

$$P_{ij}(t) = P(Z(t+s) = j \mid Z(s) = i), \quad i, j = 0, 1, \dots, t, s \geq 0$$

and let

$$\phi(t; s) = \sum_{j=0}^{\infty} P_{1j}(t) s^j$$

the probability generating function for $P_{1j}(t)$.

THEOREM 1. *The probability generating function for $P_{1j}(t)$, $\phi(t; s)$, satisfies:*

$$\phi(t + v; s) = \phi(t; \phi(v; s)).$$

This is the continuous time analog of Theorem 7, in the case of discrete time branching processes.

Proof. Since individual particles act independently, we have the fundamental relation

$$\sum_j P_{ij}(t) s^j = \left[\sum_j P_{1j}(t) s^j \right]^i = [\phi(t; s)]^i.$$

The formula means that the population $Z(t; i)$ involving in time t from i initial parents is same, in the sense of probability, as the combined sum of i population each having one initial parents. Also, this formula characterizes and distinguishes branching processes from other continuous time Markov chains. By the time homogeneity, the Chapman - Kolmogorov equations take the form

$$P_{ij}(t + v) = \sum_k P_{ik}(t) P_{kj}(v).$$

Now

$$\begin{aligned} [\phi(t + v; s)]^i &= \sum_j P_{ij}(t + v) s^j \\ &= \sum_j \sum_k P_{ik}(t) P_{kj}(v) s^j \\ &= \sum_k P_{ik}(t) \sum_j P_{kj}(v) s^j \\ &= \sum_k P_{ik}(t) [\phi(v; s)]^k \\ &= [\phi(t; \phi(v; s))]^i. \end{aligned}$$

When $i = 1$, we obtain the result. □

THEOREM 2. *Let $u(s) = \sum_k a_k s^k$. Then $\phi(t; s)$ satisfies:*

$$\frac{\partial \phi(t; s)}{\partial t} = \frac{\partial \phi(t; s)}{\partial s} u(s)$$

$$\frac{\partial \phi(t; s)}{\partial t} = u(\phi(t; s))$$

with the initial condition

$$\phi(0; s) = \sum_j P_{1j}(0) s^j = s.$$

Proof. We have

$$\begin{aligned} \phi(h; s) &= \sum_j P_{1j}(h) s^j \\ &= \sum_j (\delta_{1j} + a_j h + o(h)) s^j \\ &= s + h \sum_j a_j s^j + o(h) \\ &= s + hu(s) + o(h). \end{aligned}$$

Now

$$\phi(t + h; s) = \phi(t; \phi(h; s)) = \phi(t; s + hu(s) + o(h)).$$

By Taylor's theorem, we expand the right-hand side with respect to the second variable

$$\phi(t + h; s) = \phi(t; s) + \frac{\partial \phi(t; s)}{\partial s} hu(s) + o(h).$$

Hence

$$\frac{\phi(t + h; s) - \phi(t; s)}{h} = \frac{\partial \phi(t; s)}{\partial s} u(s) + \frac{o(h)}{h}.$$

Taking $h \rightarrow 0^+$, we get

$$\frac{\partial \phi(t; s)}{\partial t} = \frac{\partial \phi(t; s)}{\partial s} u(s).$$

This is a partial differential equation for the function of two variables $\phi(t; s)$ with the initial condition

$$\phi(0; s) = \sum_j P_{1j}(0) s^j = s.$$

Proof of part 2:

Now consider,

$$\phi(v + h; s) = \phi(v; \phi(h; s)).$$

By Taylor's theorem, we get

$$\phi(v+h; s) = \phi(v; s) + hu(\phi(v; s)) + o(h).$$

Hence

$$\frac{\phi(v+h; s) - \phi(v; s)}{h} = u(\phi(v; s)) + \frac{o(h)}{h}.$$

Taking $h \rightarrow 0^+$, then substituting $v = t$, we get

$$\frac{\partial \phi(t; s)}{\partial t} = u(\phi(t; s)).$$

This is an ordinary differential equation with initial condition

$$\phi(0; s) = \sum_j P_{1j}(0)s^j = s.$$

□

Mean of $Z(t)$

Consider

$$\frac{\partial \phi(t; s)}{\partial t} = \frac{\partial \phi(t; s)}{\partial s} u(s).$$

By differentiating with respect to s and interchanging the order of differentiation on the left side, we get

$$\frac{\partial^2 \phi(t; s)}{\partial t \partial s} = u(s) \frac{\partial^2 \phi(t; s)}{\partial s^2} + u'(s) \frac{\partial \phi(t; s)}{\partial s}.$$

If $s = 1$, $u(1) = \sum_k a_k = 0$, and then

$$\frac{\partial m(t)}{\partial t} = u'(1)m(t)$$

where

$$m(t) = E[Z(t)] = \frac{\partial \phi(t; s)}{\partial s} \Big|_{s=1}.$$

But, since $Z(0) = 1$, then $m(0) = 1$ and the solution is

$$m(t) = e^{u'(1)t}.$$

Probability of Extinction

DEFINITION 3. *The extinction probability, q , is defined by*

$$q = \lim_{t \rightarrow \infty} P_{10}(t).$$

Now, we assume that $a_0 > 0$, otherwise extinction is impossible. It is enough to consider the case where the process starts with a single individual at time zero. In fact, with $s = 0$

$$P_{i0}(t) = [P_{10}(t)]^i.$$

Now, $P_{i0}(t)$ is non-decreasing in t ,

$$\begin{aligned} P_{i0}(t+v) &= [\phi(t+v; 0)]^i \\ &= [\phi(t; \phi(v; 0))]^i \\ &\geq [\phi(t; 0)]^i \\ &= P_{i0}(t). \end{aligned}$$

Derivation of Probability of Extinction

Let t_0 be a fixed positive number. Consider the discrete time process

$$Z(0), Z(t_0), Z(2t_0), \dots, Z(nt_0), \dots$$

where $Z(t)$ is the population size at time t .

Assume that, the population size at $t = 0$ is one. Since $Z(t)$ is assumed to be a Markov process, the discrete process $Y_n = Z(nt_0)$ will be a Markov chain which is also a discrete time branching process. By the hypothesis of homogeneity of the probability function of $Z(t)$ and

$$\sum_j P_{ij}(t) s^j = \left[\sum_j P_{1j}(t) s^j \right]^i = [\phi(t; s)]^i.$$

We obtain

$$\begin{aligned}
 \sum_k P(Y_{n+1} = k \mid Y_n = i) s^k &= E [S^{Y_{n+1}} \mid Y_n = i] \\
 &= E [S^{Z((n+1)t_0)} \mid Z(nt_0) = i] \\
 &= E [S^{Z(t_0)} \mid Z(0) = i] \\
 &= [\phi(t_0; s)]^i \\
 &= [E [S^{Z(t_0)} \mid Z(0) = 1]]^i \\
 &= [E [S^{Y_1} \mid Y_0 = 1]]^i.
 \end{aligned}$$

This shows that $\{Y_n, n = 0, 1, \dots\}$ is a branching process.

The probability generating function of the number of offspring of a single individual in this process is $\phi(t_0; s)$. By Theorem 2, we know that the probability of extinction for the Y_n process is the smallest non-negative root of the equation

$$\phi(t_0; s) = s.$$

But

$$\begin{aligned}
 P(Y_n = 0 \text{ for some } n) &= \lim_{n \rightarrow \infty} P(Y_n = 0) \\
 &= \lim_{n \rightarrow \infty} P(Z(nt_0) = 0) \\
 &= \lim_{t \rightarrow \infty} P(Z(t) = 0) \\
 &= q.
 \end{aligned}$$

The extinction probability q of the continuous time branching process $Z(t)$ is the smallest non-negative root of the equation

$$\phi(t_0; s) = s$$

where t_0 is any positive number.

Hence, we expect that we should also be able to calculate q from an equation that does not depend on time.

THEOREM 4. *The probability of extinction q is the smallest non-negative root of the*

equation

$$u(s) = 0$$

Hence, $q = 1$ if and only if $u'(1) \leq 0$.

Proof. Since q satisfies

$$\phi(t_0; s) = s$$

for any $t_0 > 0$, we have

$$\frac{\phi(v+h; s) - \phi(v; s)}{h} = u(\phi(v; s)) + \frac{o(h)}{h}.$$

If $s = q$, then

$$\frac{q - q}{h} = u(q) + \frac{o(h)}{h}.$$

Hence, for any $h > 0$,

$$0 = u(q) + \frac{o(h)}{h}$$

Taking $h \rightarrow 0^+$, we obtain

$$u(q) = 0.$$

Since

$$u''(s) = \sum_{k=1}^{\infty} a_k k(k-1)s^{k-2} \geq 0$$

$u(s)$ is convex in the interval $[0, 1]$. As $u(1) = 0$ and $u(0) = a_0 > 0$, $u(s)$ may have at most one zero in the interval $(0, 1)$. According to whether $u'(1) \leq 0$ or $u'(1) > 0$ holds, we have the cases $q = 1$ or $q < 1$ respectively. Note that $E[Z(t_0)] = E[Y]$ if and only if $u'(1) > 0$. This means that for the discrete time branching process $Z(nt_0)$, $n = 0, 1, 2, \dots$ ($t_0 > 0$ fixed), extinction occurs with probability less than one, and therefore the same is true for the process $Z(t)$. Here, the probability of extinction q is the smallest zero of $u(s)$ in $[0, 1]$. In a similar manner we conclude that, if $u'(1) \leq 0$, q must be equal to one. In either case q is the smallest non-negative root of $u(s) = 0$. $\square$

Limit Theorem

If $u'(1) = 0$ and $u''(1) < \infty$, then

$$P(Z(t) > 0 \mid Z(0) = 1) \sim \frac{2}{tu''(1)}, \quad t \rightarrow \infty$$

and

$$\lim_{t \rightarrow \infty} P\left(\frac{2Z(t)}{tu''(1)} > \lambda \mid Z(t) \neq 0\right) = e^{-\lambda}, \quad \lambda > 0.$$

If $u'(1) > 0$ and $u''(1) < \infty$, then

$$\frac{Z(t)}{e^{tu'(1)}}$$

has a limit distribution as $t \rightarrow \infty$.

Some Other Important Branching Processes

1. Bellman-Harris Processes

Consider a classical branching process in which progeny are born at the moment of the parents death. Let $Z(t)$ be the number of particles alive at time t . The distribution of particle lifetime τ is an arbitrary non-negative random variable, the resulting process is called an “age-dependent” or Bellman - Harris process. Assume that all particles reproduce and die independently of each other. This model generalizes the birth and death process in two respects: first, the lifespan of individual particles need not have the exponential distribution, and second, more than one particle can born. The process $Z(t)$ is not a Markov process and its analysis is usually done by using renewal theory.

2. Bellman-Harris Processes with Disasters

Consider the population model which follows a Bellman - Harris process. At random times, disasters beset the population and each particle alive at the time of the disaster survives with probability p . The survival of any particle is assume to be independent of the survival of any other particle. Measure of interest is limiting behavior when extinction does not occur.

3. Bellman - Harris Processes with Immigration

Consider a Bellman - Harris process $\{Z(t), t \geq 0\}$ where $Z(t)$ denotes the population size at time t . In addition, we allow the sudden appearance into the system of newly born particles called immigrants. Immigrants are assumed

to arrive in groups of various sizes with the probability of n immigrants in a group immigrating at time t given by $p_n(t)$, $n = 1, 2, \dots$. Once these particles arrive, they reproduce and die according to the Bellman-Harris process. Measures of interest are the mean, the limiting distribution and the asymptotic behavior. This branching process is widely used to describe growth and decay of biological populations.

References

1. G. Sankaranarayanan, "Branching Processes and Its Estimation Theory", Wiley Eastern Ltd, New Delhi, 1989.
2. K.B. Athreya and P.E. Ney, "Branching Processes", Dover Publications, New York, 2004.
3. K.B. Athreya, P.R. Parthasarathy and G Sankaranarayanan, "Supercritical Age-Dependent Branching Processes with Immigration", Journal of Applied Probability, Vol. 11, No. 4, pp. 695 - 702, 1974.
4. P.R. Parthasarathy and A. Krishnamoorthy, "Compounding Bellman - Harris allowing Immigration", Mathematical Biosciences, Vol. 37, pp. 105 - 111, 1977.
5. J Medhi, "Stochastic Processes", 3rd edition, New Age International Publishers, 2009.
6. T.E. Harris, "Theory of Branching Processes", Dover Publications, New York, 1989.

Stochastic Processes

NPTEL - II Web Course

N. Selvaraju

Department of Mathematics
Indian Institute of Technology Guwahati
&

S. Dharmaraja

Department of Mathematics
Indian Institute of Technology Delhi

March 9, 2015

Modulewise Contents

1	Probability Essentials	1
1	Introduction	1
1	Probability Spaces	1
2	Conditional Probability and Independence	4
2	Random Variables and Probability Distributions	7
1	Random Variables and Distribution Functions	7
2	Probability Mass and Density Functions	10
3	Functions of Random Variables	13
3	Random Vectors and Expectations	15
1	Random Vectors	15
2	Expectations	18
3	Generating Functions	22
4	Sequences of Random Variables	24
1	Sequences of Random Variables	24
2	Limit Theorems	26
2	Introduction to Stochastic Processes	1
1	What are Stochastic Processes?	1
2	Classes of Stochastic Processes	7
1	Independent Process	7

MODULEWISE CONTENTS

iv

2	Bernoulli Process	7
3	Processes with Independent Increments	7
4	Processes with Stationary Increments	8
5	Markov Processes	8
6	Poisson Process	8
7	Simple Random Walks	9
8	Second-Order Processes	9
9	Stationary Processes	10
10	Gaussian Processes	11
11	Martingales	11
12	Brownian Motion	11
3	Discrete-Time Markov Chains	1
1	Markov Chains (MCs)	1
4	Continuous-Time Markov Chains	1
1	Definition, Infinitesimal Generator Matrix and Kolmogorov Equations	1
1	Introduction	1
2	Definition	1
3	Chapman-Kolmogorov equations	4
2	Limiting and Stationary Distributions	8
1	Limiting distribution	9
2	Stationary distribution	10
3	Some notes	11
3	Poisson Processes - I	13
1	Definition	15
2	Properties of Poisson processes	16
4	Poisson Processes - II	19

1	Examples	19
2	Non-homogeneous Poisson process	20
3	Compound Poisson process	21
5	Birth Death Process	23
1	Definition	23
2	Finite BDP	24
3	Pure birth process	25
4	Pure death process	25
6	$M/M/1$ Queueing Model	27
1	Queueing systems	27
2	$M/M/1$ Queueing model	29
3	Little's law	31
7	Simple Markovian Queueing Models	35
1	$M/M/c$ Queueing model	35
2	$M/M/1/N$ Queueing model	37
3	$M/M/c/K$ Queueing model	38
4	$M/M/c/c$ loss system	39
5	$M/M/\infty$ self service system	39
5	Martingales	1
1	Introduction, Filtrations and Conditional Expectations	1
1	Filtrations	1
2	X_t is $\mathcal{F}_t$ -measurable	3
3	Adaptability	3
4	Conditional Expectations	4
2	Conditional Expectations	6
1	Rules on Conditional Expectations	6

MODULEWISE CONTENTS

vi

2	Conditional Expectation Given σ -field	8
3	Generated σ -fields	10
1	σ -field generated by a random vector $\mathbf{Y}$	10
2	Properties of Conditional Expectations given σ -field	12
4	Martingales, Sub-martingales and Super-martingales	15
1	Martingales	15
2	Doob's Martingale Process	17
3	Sub-martingales and Super-martingales	18
5	Stopping Times and Inequalities	19
1	Stopping Times	19
2	Jenson's Inequalities	21
6	Convergence Theorems	23
1	Markov Time	23
2	Optional Sampling Theorem	23
3	Martingale Convergence Theorem	25
4	Martingale in Continuous Time	25
5	References	28
6	Brownian Motion	1
1	Brownian Motion	1
2	Properties of Brownian Motions	6
3	Processes Derived from Brownian Motion	12
4	Stochastic Differential Equations	16
5	Ito Integrals	23
6	Ito's Formula	28
7	SDEs and their Applications in Finance	34
7	Renewal Processes	1

MODULEWISE CONTENTS

vii

1	Renewal Function and Renewal Equation	1
2	Generalized Renewal Processes and Renewal Limit Theorems	8
3	Markov Renewal and Markov Regenerative Processes	16
4	Non Markovian Queues	23
5	Non Markovian Queues (contd.)	29
6	Non Markovian Queues (contd.)	34
8	Branching Processes	1
1	Galton - Watson Process	1
2	Properties of GW Process	7
3	Markov Branching Process	11
9	Stationary Processes	1
1	Stationary Processes	1
1	Introduction	1
2	Important Definitions	1
3	Strict Sense and Wide Sense Stationary Process	2
4	Examples of Stationary Process	3
2	Autoregressive Processes	7
1	Moving Average Process	7
2	Autoregressive Process	8

Module 1: Probability Essentials

Lecture 1 Introduction

In our everyday life, we encounter many situations that we feel should see improvements. The system that we observe may be the planes hovering around airports waiting to land or to take off, or may be the price of a particular stock in a stock exchange, or may be speed of the internet in your home, or may be production of some systems in a manufacturing company. In order to improve the behaviour of such systems, one needs to study them first by abstracting their essential features. These are typical examples of systems which vary in time in a random manner and interest in studying them has never vanished and will never be. The mathematical models that abstract the essential features of such systems are known as stochastic processes or random processes.

In this module, we will briefly review the probability essentials that form backbone to the study of stochastic processes. Those who are familiar with the material covered in this module may either skip this module entirely or gloss over the module quickly to understand the concepts and notations used throughout the course.

Keywords: Sigma-fields, probability spaces, conditional probability, independence.

1 Probability Spaces

Probability theory is concerned with modelling systems that behave in an unpredictable manner. In probabilistic modelling, the first basic concept is that of a ‘random experiment’, like tossing a coin or throwing a dice. For example, when we toss a coin or throw a dice, we do not know what is going to be the outcome. They are thus known as random experiments which essentially means that we can enumerate all possible outcomes of the experiment but not sure which one of them actually will happen. We abstract all possible outcomes of a random experiment by a set Ω and call it the *sample space* of the experiment. The elements of Ω are called sample points or elementary events.

We are often interested in the study of a group of sample points, rather than a single sample point only. These are called as events of the experiment and the events

satisfy some consistency requirements among themselves. In its full generality, they form a collection known as a σ -algebra (or σ -field) over Ω which is defined as below.

DEFINITION 1. *A nonempty collection $\mathcal{F}$ of subsets of Ω is called a σ -algebra (over Ω) if $\mathcal{F}$ is closed under the operations of countable unions and complementations, i.e.,*

1. $\cup_{n=1}^{\infty} A_n \in \mathcal{F}$ whenever $A_n \in \mathcal{F}, n \geq 1$, and
2. $A^c \in \mathcal{F}$ whenever $A \in \mathcal{F}$.

The pair $(\Omega, \mathcal{F})$ is called a measurable space.

EXAMPLE 2. *The largest σ -field of subsets of a fixed set Ω is the collection of all subsets of Ω , i.e., the power set P^Ω . The smallest σ -field consists of the two sets $\emptyset$ and Ω . These σ -fields are sometimes called as trivial σ -fields.*

EXAMPLE 3. *Let Ω be an infinite set, and let $\mathcal{F}$ be the collection of all finite subsets of Ω . Then $\mathcal{F}$ does not contain Ω and is not closed under complementation, and so is not an algebra (or a σ -algebra) on Ω .*

In the definition given above, if we replace the ‘countable union’ part by a ‘finite union’, then the collection is called as a field. It can be seen easily that each σ -field on Ω is a field on Ω since, for example, the union of the finite sequence $A_1, A_2, \dots, A_n$ is the same as the union of the infinite sequence $A_1, A_2, \dots, A_n, A_n, A_n, \dots$. Similarly, every finite field (a field with a finite number of elements) is a σ -field too. But, in general, a field may not be a σ -field as shown below.

EXAMPLE 4. *Let A be a nonempty proper subset of Ω , and let $\mathcal{F} = \{\emptyset, \Omega, A, A^c\}$. Then $\mathcal{F}$ is the smallest σ -field containing A . For if $\mathcal{G}$ is a σ -field and $A \in \mathcal{G}$, then by definition of a σ -field, $\Omega, \emptyset$, and A^c belong to $\mathcal{G}$, hence $\mathcal{F} \subseteq \mathcal{G}$. But $\mathcal{F}$ is a σ -field, for if we form complements or unions of sets in $\mathcal{F}$, we invariably obtain sets in $\mathcal{F}$. Thus $\mathcal{F}$ is a σ -field that is included in any σ -field containing A , and the result follows.*

EXERCISE 5. *Let A, B be nonempty proper subsets of Ω . Determine the smallest σ -field containing A and B . How many elements are there in it? Extending the idea, describe explicitly the smallest σ -field containing a finite number of subsets $A_1, A_2, \dots, A_n$ of Ω .*

Generalizing the idea of the previous example, one can talk about the *smallest*

σ -field containing a class of sets. If $\mathcal{C}$ is a class of subsets of a set Ω , the smallest σ -field containing the sets of $\mathcal{C}$ will be written as $\sigma(\mathcal{C})$, and will be called the *minimal σ -field over $\mathcal{C}$* or the *σ -field generated by $\mathcal{C}$* . And, $\mathcal{C}$ will be called a generator for the σ -field $\sigma(\mathcal{C})$.

EXAMPLE 6. (*Borel σ -field*) In probability theory, the σ -field of interest is what is known as the Borel σ -field (especially in the case when $\Omega = \mathbb{R}$). Denoted as $\mathcal{B}$ (or $\mathcal{B}(\mathbb{R})$), the Borel σ -field over $\mathbb{R}$ is the σ -field generated by the class of all intervals of the form (a, b) , $a, b \in \mathbb{R}$. A Borel set is an element of the Borel σ -field.

Note that the class of open intervals is only one of the many generators of the Borel σ -field. And, though practically all the subsets of $\mathbb{R}$ that we encounter are Borel sets, there exists non-Borel sets. But we will not worry about that.

The third idea concerns the measurement of these events, i.e, the probability, which we define below in an axiomatic way.

DEFINITION 7. A function $P : \mathcal{F} \rightarrow \mathbb{R}$ is called a probability measure (or simply probability) if

1. $0 \leq P(A) \leq 1$, $\forall A \in \mathcal{F}$,
2. $P(\Omega) = 1$, and
3. $A_1, A_2, \dots \in \mathcal{F}$ are pairwise disjoint, then $P(\cup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} P(A_i)$.

The triplet $(\Omega, \mathcal{F}, P)$ is called a probability space.

If you are familiar with measure theory, you can realize immediately that the probability measure is a measure of total mass one. You should keep in mind the fact that there is always an underlying probability space in every probabilistic modelling. Sometimes it can be described easily and sometimes it may not be.

EXAMPLE 8. In the case of throwing a die twice, $\Omega = \{(i, j) : i, j = 1, 2, \dots, 6\}$, where i is the outcome of the first throw and j the outcome of the second throw. If we take $\mathcal{F} = P^\Omega$ then we get a particular probability space.

If we take $\mathcal{F} = \{\{(1, 1)\}, \{(1, 1)\}^c, \Omega, \emptyset\}$, then we get a different probability space.

2 Conditional Probability and Independence

Consider families with two children. What is the probability that a family has another boy, given that it has one boy? The answer however is not $\frac{1}{2}$. Here, $\Omega = \{BB, BG, GG, GB\}$ and let $H = \{BB, BG, GB\}$ be the event that the family has at least one boy. If H is considered as the new sample space and all outcomes are considered to be equally likely then $P(A | H) = \frac{1}{3}$, where A is the event that the family has two boys, and in that case $P(A | H) = \frac{P(A \cap H)}{P(H)} = \frac{1}{3}$. In general, we have the following definition.

DEFINITION 9. Given a probability space $(\Omega, \mathcal{F}, P)$, let $B \in \mathcal{F}$ be some fixed event such that $P(B) > 0$, then the conditional probability of $A \in \mathcal{F}$ given B (denoted by $P(A | B)$) is equal to $\frac{P(A \cap B)}{P(B)}$.

THEOREM 10. Given $(\Omega, \mathcal{F}, P)$, consider a fixed $B \in \mathcal{F}$ such that $P(B) > 0$. Then the function $P(\cdot | B)$ behaves like an ordinary probability function, that is it satisfies all the three axioms.

Proof. 1. $P(A | B) \geq 0$ for any $A \in \mathcal{F}$

2. $P(\Omega | B) = 1$

3. If A_i for $i = 1, 2, \dots$ is a sequence of mutually exclusive (disjoint) events then

$$P(\cup_{i=1}^{\infty} A_i | B) = \sum_{i=1}^{\infty} P(A_i | B).$$

This is because

$$P(\cup_{i=1}^{\infty} A_i | B) = \frac{P(\cup_{i=1}^{\infty} (A_i \cap B))}{P(B)} = \sum_{i=1}^{\infty} \frac{P(A_i \cap B)}{P(B)} = \sum_{i=1}^{\infty} P(A_i | B).$$

Also the original probability function P on $\mathcal{F}$ actually is of the same form where $B = \Omega$.

□

We have the following theorem which is known as the theorem of total probability.

THEOREM 11. Let $A_i \in \mathcal{F}$ for $i = 1, 2, \dots$ be a sequence of mutually exclusive and

exhaustive events such that $P(A_i) > 0$ for all i . If $B \in \mathcal{F}$ be any other event then

$$P(B) = \sum_{i=1}^{\infty} P(B | A_i)P(A_i).$$

The proof of the above theorem is simple on noting that $P(B) = P(B \cap \Omega) = P(B \cap (\cup_{i=1}^{\infty} A_i)) = P(\cup_{i=1}^{\infty} (A_i \cap B)) = \sum_{i=1}^{\infty} P(A_i \cap B)$, since the A_i s are mutually exclusive. So, $P(B) = \sum_{i=1}^{\infty} P(B | A_i)P(A_i)$.

A very useful result is the following theorem.

THEOREM 12. (*Bayes theorem*) Let $A_i \in \mathcal{F}$ for $i = 1, 2, \dots$ be a sequence of mutually exclusive and exhaustive events such that $P(A_i) > 0$ for all i . If $B \in \mathcal{F}$ be any other event then

$$P(A_i | B) = \frac{P(B | A_i)P(A_i)}{\sum_{i=1}^{\infty} P(B | A_i)P(A_i)}.$$

The theorem follows from the fact that $P(A_i | B) = \frac{P(A_i \cap B)}{P(B)}$, and then using total probability result.

EXAMPLE 13. In a bolt factory machines $M1, M2$ and $M3$ manufacture respectively 25%, 35% and 40% of the total. Of their output 5, 4, 2 percent are defective bolts. A bolt is drawn from a day's production and found to be defective. What is the probability that it was manufactured by machine $M3$?

Solution: Let A_1, A_2 and A_3 represent the events that a bolt selected at random is manufactured by the machines $M1, M2$ and $M3$ respectively, and let B denote the event of its being defective. It is given that $P(A_1) = 0.25, P(A_2) = 0.35, P(A_3) = 0.40, P(B|A_1) = 0.05, P(B|A_2) = 0.04, P(B|A_3) = 0.02$. Therefore, the probability that a defective bolt selected at random is manufactured by machine $M3$ is computed using Bayes' rule as

$$P(A_3|B) = \frac{P(A_3)P(B|A_3)}{\sum_{i=1}^3 P(A_i)P(B|A_i)} = \frac{(0.40)(0.02)}{(0.25)(0.05) + (0.35)(0.04) + (0.40)(0.02)} = \frac{16}{69}.$$

DEFINITION 14. Given a probability space $(\Omega, \mathcal{F}, P)$, two events A and B are said to be statistically independent or stochastically independent or (simply) independent if

$$P(A \cap B) = P(A)P(B).$$

It can be easily shown that the above is the same as saying that if $P(B) > 0$ then A and B are independent if $P(A | B) = P(A)$ and if $P(B^c) > 0$ then A and B are independent if $P(A | B^c) = P(A)$.

A is said to be independent of itself if $P(A)$ is either 0 or 1.

Three events A_1, A_2, A_3 are said to be independent if they are pairwise independent and also if $P(A_1 \cap A_2 \cap A_3) = P(A_1)P(A_2)P(A_3)$.

The following is an example which illustrates that pairwise independence need not imply independence of the events if the number of events is more than 2.

EXAMPLE 15. Let $\Omega = \{\omega_1, \omega_2, \omega_3, \omega_4\}$ be a sample and $\mathcal{F}$ be the power set of Ω . Let all the outcomes in the sample space be equally probable, let $A_1 = \{\omega_1, \omega_2\}$, $A_2 = \{\omega_2, \omega_3\}$, $A_3 = \{\omega_3, \omega_4\}$, then $P(A_1 \cap A_2) = P(A_1)P(A_2) = \frac{1}{2}$, $P(A_2 \cap A_3) = P(A_2)P(A_3) = \frac{1}{2}$, $P(A_1 \cap A_3) = P(A_1)P(A_3) = \frac{1}{2}$ and $P(A_1 \cap A_2 \cap A_3) = P(\phi) = 0$. But $P(A_1)P(A_2)P(A_3) = \frac{1}{8}$ and hence these three events are not statistically independent although they are pairwise independent.

Another question which needs to be addressed is that if $P(A_1 \cap A_2 \cap A_3) = P(A_1)P(A_2)P(A_3)$, then does it imply that the events are pairwise independent? Again, the answer is no. To see how, repeat the above experiment with Ω having 8 elements such that A_1, A_2, A_3 have four elements each and $A_1 \cap A_2 \cap A_3$ has one element.

Similarly, if we have n events say $A_1, A_2, \dots, A_n$ then the A_i 's are said to be independent if they satisfy the following equalities.

1. $P(A_i \cap A_j) = P(A_i)P(A_j)$ for all $i < j$, there are n_{C_2} such pairs.
2. $P(A_i \cap A_j \cap A_k) = P(A_i)P(A_j)P(A_k)$ for all $i < j < k$, there are n_{C_3} such pairs.
3. $\vdots$
4. $P(A_1 \cap A_2 \cap \dots \cap A_n) = P(A_1)P(A_2) \dots P(A_n)$

So in total one has to check $2^n - n - 1$ such identities.

Lecture 2 Random Variables and Probability Distributions

Keywords: Random variables, probability distributions, distribution functions, probability mass and density functions, functions of random variables, random vectors.

1 Random Variables and Distribution Functions

Often, one is interested to know answers to some questions related to some random experiment. If the interest is only the concerned questions, then we may be able to condense the probability space into something smaller so that it can be studied more easily. The abstraction of the concerned question is what we call a random variable, whose precise definition is given below.

DEFINITION 1. A function $X : \Omega \rightarrow \mathbb{R}$ is called a random variable if $X^{-1}(B) = \{\omega : X(\omega) \in B\} \in \mathcal{F}$ for every $B \in \mathcal{B}(\mathbb{R})$, the Borel σ -field over $\mathbb{R}$.

Equivalently, one can also define a random variable (r.v.) X by $\{\omega : X(\omega) \leq x\} \in \mathcal{F}$ for all real x . In the case, when X takes only a countable number of values $x_1, x_2, \dots$, then this is equivalent to $\{\omega : X(\omega) = x_i\} \in \mathcal{F}$ for all i . In effect, what this means is that X is a measurable mapping from the measurable space $(\Omega, \mathcal{F})$ to the measurable space $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$.

Let us consider the example of tossing a coin twice. Then $\Omega = \{HH, HT, TT, TH\}$, let $\mathcal{F} = \{\{HH, HT\}, \{TH, TT\}, \Omega, \phi\}$ and P is some probability measure defined on $\mathcal{F}$. If we define a function X on Ω such that $X(HH) = 2, X(TH) = 1, X(HT) = 1, X(TT) = 0$, so basically the function X gives the number of heads which occurred. Then the question is whether X is a random variable or not. Here,

$$X^{-1}(-\infty, x] = \begin{cases} \phi & \text{for } x < 0 \\ \{TT\} & \text{for } 0 \leq x < 1 \\ \{TT, HT, TH\} & \text{for } 1 \leq x < 2 \\ \Omega & \text{for } x \geq 2 \end{cases}$$

Suppose if we consider $B = (-\infty, \frac{1}{2}]$ which is a Borel set, then $X^{-1}(B) = \{TT\}$ is not an event. Hence this function X is not a random variable with respect to this probability space. However if we take $\mathcal{F}$ to be the power set of Ω , then this

function will be a random variable, for that matter any function from Ω to $\mathbb{R}$ will be a random variable.

Remark:

If X is a random variable then $X^{-1}(-\infty, x) = \{\omega : X(\omega) < x\}$, $X^{-1}(\{x\}) = \{\omega : X(\omega) = x\}$, $X^{-1}(a, x] = \{\omega : a < X(\omega) \leq x\}$, $X^{-1}[a, x] = \{\omega : a \leq X(\omega) \leq x\}$ are all events for any x and a real numbers.

Result: If X is a random variable on the probability space $(\Omega, \mathcal{A}, P)$ then $aX + b$ is also a random variable for any two real numbers a and b .

To see this, consider $\{\omega : aX(\omega) + b \leq x\} = \{\omega : X(\omega) \leq \frac{x-b}{a}\}$, if $a > 0$. But $\{\omega : X(\omega) \leq \frac{x-b}{a}\} \in \mathcal{A}$ since X is a random variable on the probability space $(\Omega, \mathcal{A}, P)$. $\{\omega : aX(\omega) + b \leq x\} = \{\omega : X(\omega) \geq \frac{x-b}{a}\}$, if $a < 0$. $= X^{-1}[\frac{x-b}{a}, \infty) = X^{-1}(-\infty, \frac{x-b}{a})^c = X^{-1}(-\infty, \frac{x-b}{a})^c \in \mathcal{A}$. If $a = 0$ then $\{\omega : aX(\omega) + b \leq x\} = \Omega$ if $x \geq b$ and is $\emptyset$ if $x < b$.

In fact, the above fact which was considered for a linear function can be generalized to a much broader class of functions. The above is true if the function, for example, is continuous, or piecewise continuous.

The random variable X on a probability space $(\Omega, \mathcal{F}, P)$ induces a probability space $(\mathbb{R}, \mathcal{B}, P_X)$ such that $P_X(B) = P(X^{-1}(B)) = P(\{\omega : X(\omega) \in B\})$ for any Borel set B . This P_X is called the probability distribution of the random variable X . Since this is a probability measure so it should satisfy the three axioms.

1. $P_X(B) \geq 0$, is obvious.
2. $P_X(\mathbb{R}) = P(X^{-1}(\mathbb{R})) = P(\omega : X(\omega) \in \mathbb{R}) = P(\Omega) = 1$
3. Let B_i for $i = 1, 2, \dots$ be a sequence of mutually exclusive Borel sets, then $P_X(\cup_i B_i) = \sum_i P_X(B_i)$. This is because

$$P_X(\cap_i B_i) = P(X^{-1}(\cup_i B_i)) = P(\cup_i (X^{-1}(B_i))) = \sum_i P(X^{-1}(B_i)) = \sum_i P_X(B_i),$$

since P is a probability measure and if $B_i \cap B_j = \emptyset$, then $X^{-1}(B_i) \cap X^{-1}(B_j) = \emptyset$.

A very useful function connected with P_X is the distribution function of X defined as follows.

DEFINITION 2. A real-valued function F defined on $\mathbb{R}$ which satisfies the conditions

1. non-decreasing,
2. right continuous,
3. $F(-\infty) = \lim_{x \rightarrow -\infty} F(x) = 0$ and $F(\infty) = \lim_{x \rightarrow \infty} F(x) = 1$,
4. $0 \leq F(x) \leq 1$

is called a distribution function.

The function $F(x) = P(\omega : X(\omega) \leq x) = P(X \leq x)$, $x \in \mathbb{R}$ is called the (cumulative) distribution function of the random variable X .

Indeed, it can be verified that the function $F(\cdot)$ defined above satisfies the properties of a distribution function.

EXAMPLE 3. Let $\Omega = \{HH, HT, TH, TT\}$ and $\mathcal{A} = P^\Omega$ Let us define a function X on Ω as $X(HH) = 0, X(HT) = X(TH) = 1$, and $X(TT) = 2$. Then $F(x) = P[\{\omega : X(\omega) \leq x\}] = P(\phi) = 0$ if $x < 0$

$$F(x) = P[\{\omega : X(\omega) \leq x\}] = P[\{HH\}] = \frac{1}{4} \text{ if } 0 \leq x < 1$$

$$F(x) = P[\{\omega : X(\omega) \leq x\}] = P[\{HH, HT, TH\}] = \frac{3}{4} \text{ if } 1 \leq x < 2$$

$$F(x) = P[\{\omega : X(\omega) \leq x\}] = P[\Omega] = 1 \text{ if } 2 \leq x.$$

We can draw the graph of this function defined on $\mathbb{R}$ given by

$$F(x) = \begin{cases} 0 & \text{for } x < 0 \\ \frac{1}{4} & \text{for } 0 \leq x < 1 \\ \frac{3}{4} & \text{for } 1 \leq x < 2 \\ 1 & \text{for } x \geq 2 \end{cases}$$

which is discontinuous at only a finite number of points. In fact it is a property of the distribution function, since it is a monotonically nondecreasing bounded function that it can be discontinuous at only a countable number of points.

Result: If F is any distribution function then there exists a probability space and a random variable defined on that space, say X , such that F will be the cu-

mulative distribution function of the random variable X . That is, for every $x \in \mathbb{R}$, $F(x) = P(\{X \leq x\})$.

2 Probability Mass and Density Functions

Two important class of random variables that we usually encounter are given below.

A discrete random variable is one where $P\{\omega : X(\omega) = x\} = 0$ for all but countably many x 's, say, $x_1, x_2, \dots$, and $\sum_i P\{\omega : X(\omega) = x_i\} = 1$. In defining sets one usually omits the ω ; thus $(X = x)$ means the same as $\{\omega : X(\omega) = x\}$. For such a random variable, the distribution function will be a piecewise constant function with at most a countable number of values. The jumps would correspond to the probability masses at the jump points. The masses $P(X = x)$ then defines the probability distribution of X and hence they are referred as 'probability mass functions'.

DEFINITION 4. *Given a probability space, $(\Omega, \mathcal{F}, P)$, a real valued function defined on $\mathbb{R}$ by $p(x) = P(\{X = x\})$ is called the probability mass function (PMF) of the random variable X , and x is called a possible value of X if $P(\{X = x\}) > 0$.*

So, for any $B \in \mathcal{B}$, $P_X(B) = P(X^{-1}(B)) = P(\bigcup_{x_n \in B} X^{-1}\{x_n\}) = \sum_{x_n \in B} P(X^{-1}\{x_n\}) = \sum_{x_n \in B} P(\{X = x_n\})$. So $F(x) = \sum_{x_n \leq x} P(\{X = x_n\})$.

The properties of p are:

1. $0 \leq p(x) \leq 1$.
2. $\{x : p(x) \neq 0\}$ is a finite or a countably infinite set.
3. $\sum_i p(x_i) = \sum_i P(\{X = x_i\}) = P(\bigcup_i \{\omega : X(\omega) = x_i\}) = P(\Omega) = 1$.

Any function $p(\cdot)$ that satisfies the above is a probability mass function.

Some examples of discrete random variables are given below.

1. Let us consider the experiment of tossing a coin n times with p probability of head (success) and $1 - p$ is the probability of tail (failure) in a single toss and the tosses are independent. Trials of this type are called Bernoulli trials. If

we denote the number of heads in n tosses by X , then X is a discrete random variable with PMF

$$p(x) = P(X = x) = \begin{cases} nC_k(1-p)^{n-k}p^k, & \text{for } x = 0, 1, 2, \dots, n \\ 0 & \text{otherwise} \end{cases}$$

This is known as binomial distribution. A Bernoulli distribution is a special case of binomial with $n = 1$.

2. A random variable X is said to be discrete uniform on $\{x_1, x_2, \dots, x_n\}$ if the PMF of X is

$$p(x) = \begin{cases} \frac{1}{n}, & \text{for } x = x_1, x_2, \dots, x_n \\ 0 & \text{otherwise} \end{cases}$$

3. A coin is tossed repeatedly till the occurrence of a head (success), with probability of head (success) given by p in each toss. Let X be the function giving the number of tails preceding the first head. So here X can take values $0, 1, 2, \dots$, and

$$p(x) = \begin{cases} p(1-p)^x, & \text{for } x = 0, 1, 2, \dots \\ 0 & \text{otherwise} \end{cases}$$

is the PMF of X . Here X is said to follow a geometric distribution with parameter p .

There is a variant of geometric distribution which counts the number of tosses required to get the first success.

4. The PMF of a Poisson random variable with parameter λ is defined as

$$p(x) = \begin{cases} \frac{\lambda^x e^{-\lambda}}{x!}, & \text{for } x = 0, 1, 2, \dots \\ 0 & \text{otherwise} \end{cases}$$

A discrete random variable is used to describe a random experiment with finite or countably infinite number of possible outcomes. We simply introduce X as a random variable which takes values $x_1, x_2, \dots$ and take $X = x_i$ if and only if the experiment gives the i th outcome. So conducting an experiment with finite or countably infinite number of possible values is conceived as observing the values of a discrete random variable X .

The other important class of random variables are called continuous random variables.

DEFINITION 5. A random variable X is said to be of the continuous type if its distribution function is continuous at every $x \in \mathbf{R}$. A random variable X is said to be of the absolute continuous type if there exists a nonnegative function on $\mathbb{R}$ such that for every real x we have $P(\{X \leq x\}) = F(x) = \int_{-\infty}^x f(t)dt$. Here, the distribution function F is said to have a probability density function (PDF) f .

Hence $P(\{a < X \leq b\}) = F(b) - F(a) = \int_a^b f(t)dt$. And, for any Borel set A , $P(X \in A) = \int_A f(t)dt$.

Note: All continuous distribution functions do not have a density function. We will be considering only random variables of absolute continuous type. From now on, we will refer to these random variables simply as continuous random variables (omitting the work ‘absolute’, for convenience).

Properties of the density function f are:

1. $f(x) \geq 0$ for all $x \in R$.
2. $\int_{-\infty}^{\infty} f(x)dx = 1$.

Wherever f is continuous or in other words F is differentiable, the function f is obtained by $f(x) = \frac{d}{dx}F(x)$.

Some examples of (absolutely) continuous distributions are given below.

1. The PDF of a uniform distribution on (a, b) is defined by

$$f(x) = \begin{cases} \frac{1}{b-a} & \text{for } a < x < b \\ 0 & \text{elsewhere} \end{cases}$$

Note that the function f is discontinuous at the points a and b and $\int_{-\infty}^{\infty} f(t)dt = 1$. So this is indeed a density function and this is said to be a uniform density function with parameters a and b .

2. The standard normal probability density function is given by

$$f(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{(x)^2}{2}} \quad \text{for } -\infty < x < \infty$$

The density function satisfies the condition that $f(x) = f(-x)$, also the distribution function of the random variables X and $-X$ are same. Also $F(-x) = P(X \leq -x) = P(-X \geq x) = P(X \geq x) = 1 - P(X < x) = 1 - P(X \leq x) = 1 - F(x)$.

3. The exponential density function with parameter $\lambda > 0$ is given by

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

Here the random variable X is said to have exponential distribution with parameter λ . Exponential distribution has a very special and unique property which is known as memoryless property and is given below. For any $s, t \in \mathbb{R}$ such that $s \geq t$, $P(\{X > s | X > t\}) = P(\{X > s - t\})$. This property of the exponential distribution has a lot of applications in, for example, queueing theory.

3 Functions of Random Variables

Let $g(x)$ be a real valued function defined on $\mathbb{R}$. Define $Y = g(X)$ such that $Y(\omega) = g(X(\omega))$, where X is a random variable defined on the probability space $(\Omega, \mathcal{A}, P)$. Here $X : \Omega \rightarrow \mathbb{R}$ and $g : \mathbb{R} \rightarrow \mathbb{R}$ so $Y = g(X)$ is a function from $\Omega \rightarrow \mathbb{R}$.

Let us consider the following example: $\Omega = \{-1, 0, 1\}$, $\mathcal{A} = P^\Omega$ and let all the outcomes be equally likely. Let $X(\omega) = \omega$ and let $Y = g(X) = X^2$ so for every $\omega \in \Omega$, $Y(\omega) = \omega^2$, and the possible values of Y are 0, 1. In order to show that Y is a random variable one has to verify that

$$\{Y = y\} \in \mathcal{A} \text{ for } y = 0, 1$$

$$\{Y = 0\} = \{\omega : Y(\omega) = 0\} = \{0\} \in \mathcal{A}.$$

$$\{Y = 1\} = \{\omega : Y(\omega) = 1\} = \{1, -1\} \in \mathcal{A}.$$

Hence Y is a random variable and $P(\{Y = 0\}) = \frac{1}{3}$, $P(\{Y = 1\}) = \frac{2}{3}$. So in general if X is a discrete random variable and $g : \mathbb{R} \rightarrow \mathbb{R}$ is a real valued function then $Y = g(X)$ will be a random variable if $\{Y = y\} \in \mathcal{A}$, for every y belonging to the range of the function Y .

Since $\{Y = y\} = \{g(X) = y\} = \{X \in g^{-1}\{y\}\}$ so if $g^{-1}\{y\}$ is a Borel set for every y in the range of Y , then Y will be a random variable (here $g^{-1}\{y\}$ is the inverse image of y under g). In that case, $P(\{Y = y\}) = P(\{X \in g^{-1}\{y\}\}) = \sum_{i: g(x_i)=y} P(\{X = x_i\})$ and hence $P(\{Y = y\}) = \sum_{i: x_i \in g^{-1}\{y\}} P(\{X = x_i\})$.

In general if X is a random variable on $(\Omega, \mathcal{A}, P)$ and g is a single valued function from $R \rightarrow R$ then $g(X)$ is a random variable if and only if $\{\omega : g(X(\omega)) \in B\} \in \mathcal{A}$ for every Borel set $B \in \mathcal{B}$. That is $\{\omega : X(\omega) \in g^{-1}(B)\} \in \mathcal{A}$ for every Borel set $B \in \mathcal{B}$. So if the function g is such that $g^{-1}(B)$ is a Borel set for every Borel set B , then since X is a random variable Y will also be a random variable on the same probability space.

Let X be an absolutely continuous random variable with density function f . If g be such that $Y = g(X)$ is also a random variable then how to find the density function of Y ? We have the following result.

THEOREM 6. *Let g be a differentiable function either strictly increasing or strictly decreasing on an interval I with an inverse function g^{-1} , and let $g(I)$ denote the range of g . Let X be a continuous random variable with density function f such that $f(x) = 0$ for x outside I . Then $Y = g(X)$ has a density function h given by*

$$h(y) = f(g^{-1}(y)) \left| \frac{d}{dy} g^{-1}(y) \right| \quad \text{if } y \in g(I)$$

One can extend the above result to the case when g is neither increasing nor decreasing, but a smooth enough function.

EXAMPLE 7. *Let X be a random variable having normal distribution with parameters 0 and σ^2 , and let $Y = X^2$. We want to find the density function of the random variable Y .*

The density function as in the uniform distribution case will be given by

$$h(y) = \left| \frac{1}{2\sqrt{y}} \right| f(\sqrt{y}) + \left| -\frac{1}{2\sqrt{y}} \right| f(-\sqrt{y}), \text{ for } y > 0.$$

Here $I = (-\infty, \infty)$ and $g(I) = [0, \infty)$. So $h(y) = \frac{1}{\sigma\sqrt{2y}} e^{-\frac{y}{2\sigma^2}}$, for $y > 0$.

This is said to be the gamma density function with parameters $\alpha = \frac{1}{2\sigma^2}$ and $p = \frac{1}{2}$.

Lecture 3 Random Vectors and Expectations

Keywords: Random vectors, conditional distributions, mathematical expectations, transforms and generating functions.

1 Random Vectors

DEFINITION 1. The function $X = (X_1, X_2)$ defined on $(\Omega, \mathcal{F}, P)$ into $\mathbb{R}^2$ by $X(\omega) = (X_1(\omega), X_2(\omega))$, $\omega \in \Omega$ is called a two-dimensional random variable or a random vector if the inverse image of every two-dimensional interval, $I = \{(a_1, a_2) : -\infty < a_1 \leq x_1, -\infty < a_2 \leq x_2\}$ belongs to $\mathcal{F}$. That is, for every $x_1, x_2 \in \mathbb{R}$,

$$X^{-1}(I) = X^{-1}((-\infty, x_1] \times (-\infty, x_2]) = \{\omega : -\infty < X_1(\omega) \leq x_1, -\infty < X_2(\omega) \leq x_2\} \in \mathcal{F}.$$

Result: If X_1 and X_2 are two random variables defined on the probability space $(\Omega, \mathcal{F}, P)$ then (X_1, X_2) is a two-dimensional random variable on $(\Omega, \mathcal{F}, P)$.

This is because $X^{-1}(I) = X^{-1}(A_1 \times A_2) = X^{-1}[(A_1 \times \mathbb{R}) \cap (\mathbb{R} \times A_2)] = X^{-1}(A_1 \times \mathbb{R}) \cap X^{-1}(\mathbb{R} \times A_2) = \{\omega : X_1(\omega) \leq a_1\} \cap \{\omega : X_2(\omega) \leq a_2\} \in \mathcal{F}$, so $\{\omega : -\infty < X(\omega) \leq a_1, -\infty < X(\omega) \leq a_2\} \in \mathcal{F}$ for all $a_1, a_2 \in \mathbb{R}$. The function $F : \mathbb{R}^2 \rightarrow [0, 1]$, defined by $F(x_1, x_2) = P[X_1 \leq x_1, X_2 \leq x_2]$ for all $(x_1, x_2) \in \mathbb{R}$ is known as the distribution of the random variable (X_1, X_2) .

EXAMPLE 2. Consider the experiment of throwing a balanced die twice. Let Y be the function on Ω such that $Y(\omega) = Y((i, j)) = \max\{i, j\}$ and let X be such that $X(\omega) = |i - j|$, then since both X and Y are random variables on the probability space $(\Omega, \mathcal{F}, P)$, where $\Omega = \{(i, j) : 1 \leq i \leq 6, 1 \leq j \leq 6\}$ and $\mathcal{A} = P^\Omega$ and $P(\{(i, j)\}) = \frac{1}{36}$ for all $i = 1, 2, \dots, 6$ and $j = 1, 2, \dots, 6$ the vector (X, Y) is a two dimensional random variable on the same probability space. The joint distribution is given by entries on the table given below.

$X \backslash Y$	1	2	3	4	5	6
1	$\frac{1}{36}$	0	0	0	0	0
2	$\frac{1}{36}$	$\frac{2}{36}$	0	0	0	0
3	$\frac{1}{36}$	$\frac{2}{36}$	$\frac{2}{36}$	0	0	0
4	$\frac{1}{36}$	$\frac{2}{36}$	$\frac{2}{36}$	$\frac{2}{36}$	0	0
5	$\frac{1}{36}$	$\frac{2}{36}$	$\frac{2}{36}$	$\frac{2}{36}$	$\frac{2}{36}$	0
6	$\frac{1}{36}$	$\frac{2}{36}$	$\frac{2}{36}$	$\frac{2}{36}$	$\frac{2}{36}$	$\frac{2}{36}$

Properties: A distribution function (in two-dimensions) satisfies the following properties.

1. It is nondecreasing and right continuous with respect to each of the co-ordinates. That is if $x_1 \leq x_2$, and $y_1 \leq y_2$ then $F(x_1, y) \leq F(x_2, y)$ for all $y \in \mathbb{R}$ Also $F(x, y_1) \leq F(x, y_2)$ for all $x \in \mathbb{R}$ and $F(x_1, y_1) \leq F(x_2, y_2)$.
2. $F(x, -\infty) = \lim_{y \rightarrow -\infty} F(x, y) = 0$.
3. $F(-\infty, y) = \lim_{x \rightarrow -\infty} F(x, y) = 0$.
4. and $F(x, +\infty) = \lim_{y \rightarrow \infty} F(x, y) = F_1(x)$.
5. $F(+\infty, y) = \lim_{x \rightarrow \infty} F(x, y) = F_2(y)$.
6. $F(+\infty, +\infty) = \lim_{x \rightarrow \infty y \rightarrow \infty} F(x, y) = 1$.
7. Continuity from the right with respect to each co-ordinate. That is $F(x+, y) = F(x, y+)$ and $F(x, y) = F(x, y)$. Also $\lim_{\epsilon_1 \rightarrow 0 \epsilon_2 \rightarrow 0} F(x + \epsilon_1, y + \epsilon_2) = F(x, y)$.
8. For every (x_1, y_1) and (x_2, y_2) with $x_1 < x_2$ and $y_1 < y_2$ the inequality $F(x_1, y_1) + F(x_2, y_2) - F(x_1, y_2) - F(x_2, y_1) \geq 0$ holds. The L.H.S is nothing but $P(x_1 \leq X \leq x_2, y_1 \leq Y \leq y_2)$, and that should definitely be nonnegative.

EXAMPLE 3. An example of a function from $\mathbb{R}^2 \rightarrow [0, 1]$ which is not a distribution function is the following.

$$F(x, y) = \begin{cases} 0 & \text{for } x < 0 \text{ or } x + y < 1, \text{ or } y < 0 \\ 1 & \text{elsewhere} \end{cases}$$

$F(1, 1) + F(\frac{1}{3}, \frac{1}{3}) - F(1, \frac{1}{3}) - F(\frac{1}{3}, 1) = 1 + 0 - 1 - 1 = -1 < 0$, although it can be verified that it satisfies all the other properties of a distribution function.

Conditional distributions

Let (X_1, X_2) be a random vector of discrete type on $(\Omega, \mathcal{F}, P)$ with joint probability mass function f . Suppose the marginal p.m.f's be given by f_1 and f_2 . Recall that for any two events A and B on the probability space $(\Omega, \mathcal{F}, P)$, $P(A|B) = \frac{P(A \cap B)}{P(B)}$, provided of course that $P(B) > 0$. Let $A = \{\omega : X_1(\omega) = x_1\}$ and $B = \{\omega : X_2(\omega) = x_2\}$, then suppose if x_2 is a possible value of X_2 then $P(A|B)$ is well defined and is given by

$$P(A|B) = P[X_1 = x_1 | X_2 = x_2] = \frac{P(\{\omega : X_1(\omega) = x_1, X_2(\omega) = x_2\})}{P(\{\omega : X_2(\omega) = x_2\})} = \frac{f(x_1, x_2)}{f_2(x_2)}.$$

For a fixed x_2 , the function $P[X_1 = x_{i_1} | X_2 = x_2] \geq 0$ and also

$$\sum_{x_{i_1}} P[X_1 = x_{i_1} | X_2 = x_2] = \sum_{x_{i_1}} \frac{P[X_1 = x_{i_1} \cap X_2 = x_2]}{P[X_2 = x_2]} = \frac{P[X_2 = x_2]}{P[X_2 = x_2]} = 1.$$

Thus $P[X_1 = x_1 | X_2 = x_2]$ defines a discrete probability density function.

DEFINITION 4. Let (X, Y) be a random vector of the discrete type. If $P[Y = y] > 0$, then the function $P[X = x_i | Y = y] = \frac{P[X=x_i \cap Y=y]}{P[Y=y]}$ is called the conditional p.m.f of X given $Y = y$. Similarly if $P[X = x] > 0$ then $P[Y = y_j | X = x] = \frac{P[Y=y_j \cap X=x]}{P[X=x]}$ is called the conditional p.m.f of the random variable Y given $X = x$.

DEFINITION 5. Two random variables are said to be independent if their joint density function satisfies the condition that

$$f(x_1, x_2) = f_1(x_1)f_2(x_2) \text{ for all } (x_1, x_2) \in \mathbb{R}^2.$$

That is, the joint density function is the product of the marginal density functions for every $(x, y) \in \mathbb{R}^2$.

If X and Y are discrete random variables with joint density function f , then for any $A, B \subseteq \mathbb{R}$, $P(X \in A, Y \in B) = \sum_{x \in A} \sum_{y \in B} f(x, y)$. If X and Y are independent then this reduces to $P(X \in A, Y \in B) = \sum_{x \in A} f_X(x) \sum_{y \in B} f_Y(y)$. Here f_X and f_Y are the marginal p.d.f of the random variables X and Y .

Alternatively X and Y are said to be independent if for any two sets $A, B \subseteq \mathbb{R}$ $P(X \in A, Y \in B) = P(X \in A)P(Y \in B) \iff F(x, y) = P[X \leq x]P[Y \leq y] = F_X(x)F_Y(y)$ for all $(x, y) \in \mathbb{R}^2$. That is the joint distribution function is the product

of the marginal distribution functions.

If X and Y are independent random variables and f and g are functions satisfying the condition that $f^{-1}(A), g^{-1}(A) \in \mathcal{F}$ for any Borel set A then $f(X)$ and $g(Y)$ are also independent random variables. The fact that $f(X)$ and $g(Y)$ are random variables has already been proved, so we just need to show that they are independent random variables.

Now $P[f(X) \leq x, g(Y) \leq y] = P[X \in f^{-1}(-\infty, x], Y \in g^{-1}(-\infty, y]]$, so this equals $P[X \in f^{-1}(-\infty, x)]P[Y \in g^{-1}(-\infty, y]] = P[f(X) \leq x]P[g(Y) \leq y]$. Hence the random variables $f(X)$ and $g(Y)$ are independent.

Also discrete random variables X and Y are also said to be independent if $P[X = x|Y = y] = P[X = x]$ for every $x \in R$ and every y satisfying $P[Y = y] > 0$. So in the previous example one can check that the random variables X and Y are not independent.

EXAMPLE 6. *A die is thrown n times. Let X be the number of times 1 occurs and Y be the number of times 2 occurs, then the joint p.m.f of X and Y is given by*

$$P[X = x, Y = y] = \frac{n!}{x!y!} \left(\frac{1}{6}\right)^x \left(\frac{1}{6}\right)^y \left(\frac{4}{6}\right)^{n-x-y}.$$

Are the random variables X and Y independent?

Since $P[X = x] = n_{C_x} \left(\frac{1}{6}\right)^x \left(\frac{5}{6}\right)^{n-x}$, and $P[Y = y] = n_{C_y} \left(\frac{1}{6}\right)^y \left(\frac{5}{6}\right)^{n-y}$, we see that, in general, $P[X = x]P[Y = y] \neq P[X = x, Y = y]$.

The ideas presented above for a two-dimensional random variable holds good for any finite-dimensional random variables in a analogous way.

2 Expectations

A gambler plays the following game. A balanced coin is tossed. If head turns up then he gets Rs 5 but if tail occurs then he has to pay Rs 2. He plays this game repeatedly say N times. Let N_H be the number of times that head occurs and N_T be the number of times that tail occurs then his average gain is going to be $5\frac{N_H}{N} + (-2)\frac{N_T}{N}$. As $N \rightarrow \infty$ this expression becomes $5P(\{H\}) + (-2)P(\{T\})$ (provided we take

probability to be the long-run relative frequency).

Let X be the random variable giving the win of the gambler in a single play, then $E(X) = 5P(X = 5) + (-2)P(X = -2)$ gives the long-run gain of the gambler, or the weighted average gain of the gambler if he plays the game a large number of times.

Consider a discrete random variable X with p.m.f f . The expectation of the random variable X is said to exist if $E(|X|) = \sum_i |x_i|f(x_i) < \infty$. In that case the expectation of the random variable X denoted by $E(X)$ is given by $E(X) = \sum_i x_i f(x_i) < \infty$, since the series is absolutely convergent. So this means that the series $\sum_i x_i f(x_i)$ should be absolutely convergent, otherwise what could have happened is that by rearranging the terms we could have obtained different values of $\sum_i x_i f(x_i)$. Example of a random variable for which expectation does not exist but for which $\sum_i x_i f(x_i) < \infty$ is given next.

EXAMPLE 7. Let X be the random variable having the probability mass function given by $P(X = (-1)^{j+1} \frac{3^j}{j}) = \frac{2}{3^j}$ for $j = 1, 2, \dots$ then $\sum_j \frac{2}{3^j} = \frac{2}{3} \frac{1}{1-\frac{1}{3}} = 1$, so this is indeed a discrete density function or a probability mass function. But $E(|X|) = \sum_j \frac{3^j}{j} \frac{2}{3^j} = \sum_j \frac{2}{j}$ does not converge so $E(X)$ does not exist, although $\sum_j (-1)^{j+1} \frac{3^j}{j} \frac{2}{3^j} = \sum_j (-1)^{j+1} \frac{2}{j}$ converges.

So for any function g such that $Y = g(X)$ is a random variable, we have $E(g(X))$ will exist if $E(|g(X)|) = \sum_i |g(x_i)|P(X = x_i) < \infty$, and in that case $E(g(X)) = \sum_i g(x_i)P(X = x_i)$. This is because $E(g(X)) = \sum_i y_i P(Y = y_i)$. If g is one-one then $E(g(X)) = \sum_i y_i P(X = g^{-1}\{y_i\}) = \sum_i g(x_i)P(X = g^{-1}\{y_i\})$ here $y_i = g(x_i) = \sum_i g(x_i)P(X = x_i)$. Suppose g is not one-one then $E(g(X)) = \sum_i y_i P(g(X) = y_i) = \sum_i y_i P(X \in g^{-1}\{y_i\}) = \sum_i y_i \sum_{j: g(x_j)=y_i} P(X = x_j) = \sum_i \sum_{j: g(x_j)=y_i} g(x_j)P(X = x_j) = \sum_i g(x_i)P(X = x_i)$ (since $\sum_i \sum_{j: g(x_j)=y_i} g(x_j)P(X = x_j)$ is just a rearrangement of the series $\sum_i g(x_i)P(X = x_i)$, and the latter is absolutely convergent).

Expectation of some well known distributions

Consider the random variable $X \sim B(n, p)$, then since X has only a finite number

of possible values so $E(X)$ exists, and

$$\begin{aligned}
 E(X) &= \sum_x xP(X = x) = \sum_{x=0}^n n_{C_x} (1-p)^{n-x} p^x x \\
 &= np \sum_{x=1}^n \frac{(n-1)!}{(x-1)!(n-x)!} (1-p)^{n-x} p^{x-1} = np \sum_{x-1=0}^{n-1} (n-1)_{C_{x-1}} (1-p)^{n-x} p^{x-1} \\
 &= \sum_{y=0}^{n-1} n_{C_y} (1-p)^{n-1-y} p^y = np \text{ where } y = x - 1.
 \end{aligned}$$

Consider the random variable $X \sim P(\lambda)$ that is X follows Poisson distribution with parameter λ . Then $E(|X|) = E(X) = \sum_{x=0}^{\infty} x \frac{\lambda^x e^{-\lambda}}{x!} = \lambda \sum_{x=1}^{\infty} \frac{\lambda^{x-1} e^{-\lambda}}{(x-1)!} = \lambda \sum_{y=0}^{\infty} x \frac{\lambda^y e^{-\lambda}}{y!} = \lambda$.

Properties of Expectations

1. If X is a random variable taking constant value c then $E(x) = c$.
2. Let X be such that $E(X)$ exists and is finite, then $E(aX + b)$ also exists and is finite where a and b are real constants. This is because

$$\begin{aligned}
 E(|aX + b|) &= \sum_i |ax_i + b|P(X = x_i) \\
 &\leq \sum_i |ax_i|P(X = x_i) + \sum_i |b|P(X = x_i) \\
 &= |a| \sum_i |x_i|P(X = x_i) + |b| \sum_i P(X = x_i) \\
 &= |a| \sum_i |x_i|P(X = x_i) + |b|.
 \end{aligned}$$

Since $E(X)$ exists so $\sum_i |x_i|P(X = x_i) < \infty$, so $E(|aX+b|) = |a| \sum_i |x_i|P(X = x_i) + |b| < \infty$, in that case $E(aX + b) = a \sum_i x_i P(X = x_i) + b \sum_i P(X = x_i) = aE(X) + b$.

3. Note that $E(X - \mu) = 0$.

So in general the n th moment X is said to exist if $E(|X^n|) < \infty$ and if it exists is given by $E(X^n) = \sum_i (x_i)^n P(X = x_i)$.

The variance which measures the square of the deviation of the random variable from the mean if it exists is given by

$$\begin{aligned} E(X - \mu)^2 &= \sum_i (x_i - \mu)^2 P(X = x_i) = \sum_i (x_i^2 - 2\mu x_i + \mu^2) P(X = x_i) \\ &= \sum_i x_i^2 P(X = x_i) - 2\mu \sum_i x_i P(X = x_i) + \mu^2 \sum_i P(X = x_i) \\ &= E(X^2) - \mu E(X) + \mu^2 = E(X^2) - \mu^2. \end{aligned}$$

Variance of a random variable is usually denoted by σ^2 .

Variance of some well known distributions

For Binomial case,

$$\begin{aligned} E(X^2) &= \sum_{x=0}^n x^2 n_{C_x} (1-p)^{n-x} p^x \\ &= \sum_{x=1}^n x(x-1) n_{C_x} (1-p)^{n-x} p^x + \sum_{x=0}^n x n_{C_x} (1-p)^{n-x} p^x \\ &= n(n-1)p^2 \sum_{x=2}^n x(x-1) \frac{(n-2)!}{(x-2)!(n-x)!} (1-p)^{n-x} p^{x-2} + np \\ &= \sum_{y=0}^{n-2} (n-2)_{C_y} (1-p)^{n-2-y} p^y + np = n(n-1)p^2 + np \end{aligned}$$

$$\text{So } \text{Var}(X) = E(X^2) - (E(X))^2 = n(n-1)p^2 + np - n^2p^2 = np(1-p).$$

If $X \sim P(\lambda)$ then

$$\begin{aligned} E(X^2) &= \sum_{x=0}^{\infty} x^2 \frac{\lambda^x e^{-\lambda}}{x!} \\ &= \lambda^2 \sum_{x=2}^{\infty} x(x-1) \frac{\lambda^{x-2} e^{-\lambda}}{x!} + \sum_{x=0}^{\infty} x \frac{\lambda^x e^{-\lambda}}{x!} \\ &= \lambda^2 \sum_{y=0}^{\infty} x \frac{\lambda^y e^{-\lambda}}{y!} + \lambda = \lambda^2 + \lambda. \end{aligned}$$

So $\sigma^2 = \lambda^2 + \lambda - (E(X))^2 = \lambda^2 + \lambda - \lambda^2 = \lambda$. So for a random variable having Poisson distribution both mean and variance are equal.

If X is a random variable taking the constant value c then $\text{Var}(X) = 0$. This is because $\text{Var}(X) = E(X - \mu)^2 = \sum_i (x_i - \mu)^2 P(X = x_i) = 0P(X = c) = 0$. Similarly $\text{Var}(X) = 0 \Rightarrow \sum_i (x_i - \mu)^2 P(X = x_i) = 0 \Rightarrow (x_i - \mu)^2 P(X = x_i) = 0$ for all

i , $\Rightarrow (x_i - \mu)^2 = 0$ whenever $P(X = x_i) > 0$, that is $(x_i - \mu) = 0$ for all i .

Some results

1. If $E(X)$ exists, then $|E(X)| \leq E(|X|)$, as
 $|E(X)| = |\sum_i x_i P(X = x_i)| \leq \sum_i |x_i| P(X = x_i) = E|X|$
2. If $a \leq X \leq b$, then $a \leq E(X) \leq b$, as
 $E(X) = \sum_i x_i P(X = x_i) \leq \sum_i b P(X = x_i) = b$, similarly the other one.
3. If $P(X \geq 0) = 1$ and $E(X)$ exists and $E(X) \geq 0$.
4. For any constant $a \in \mathbb{R}$, $E(X - \mu)^2 \geq E(X - a)^2$, as $E(X - a)^2 = E(X - \mu + \mu - a)^2 = E[(X - \mu)^2 + (\mu - a)^2 + 2(X - \mu)(\mu - a)] = E(X - \mu)^2 + (\mu - a)^2$.
5. $Var(aX + b) = a^2 Var(X)$, as by letting $Y = aX + b$, then $Var(Y) = E(aX + b - aE(X) - b)^2 = E(aX - a\mu)^2 = a^2 E(X - \mu)^2$.

3 Generating Functions

We have seen how to define $E(X)$ or $E(X^n)$. In an analogous way, we can also define expectations of some function of a random variable or even a random vector, as long as the function is smooth enough to satisfy certain technical conditions. For a function g , we define $E(g(X))$, provided it exists, by

$$E(g(X)) = \begin{cases} \sum_i g(x_i) p(x_i) & \text{if } X \text{ is discrete} \\ \int_{\mathbb{R}} g(x) f(x) dx & \text{if } X \text{ is continuous.} \end{cases}$$

Some of such prominent ones are given below.

1. (Moment Generating Function (MGF)): The MGF of a random variable X , if exists, is given by the expectation of $g(x) = e^{ux}$ for $u \in \mathbb{R}$. Whenever it exists, it uniquely determines the distribution. It is much more convenient to work with MGF rather than the distributions themselves in many situations. The moments can also be obtained from the MGF.

2. (Probability Generating Function (PGF)): The PGF of a discrete random variable X is given by the expectation of $g(s) = s^X$ for $|s| \leq 1$. Again, it is sometimes convenient to work with PGF rather than the distributions themselves. The probabilities and factorial moments can be obtained from the PGF.
3. (Laplace Transform of a random variable): The LT of a nonnegative continuous random variable X means the Laplace transform corresponding to the PDF of the random variable and is given by the expectation of $g(t) = e^{-tX}$ for $t \in \mathbb{R}$. Like PGF, this is sometimes convenient to work with. The moments can be obtained from the LT and this is also connected with MGF.
4. (Characteristic Function): The characteristic function of a random variable X exists always and is given by the expectation of $g(u) = e^{iux}$ for $u \in \mathbb{R}$. Like MGF, it uniquely determines the distribution. Whenever MGF does not exist, this is very useful to work with rather than the distributions themselves in many situations. The moments can also be obtained from the characteristic function.

Lecture 4 Sequences of Random Variables

Keywords: Modes of convergence of sequences of random variables, laws of large numbers, central limit theorem.

1 Sequences of Random Variables

The purpose of this section is to give names and definitions to various forms of convergence associated with random variables. This will give us tools to understand two important results, namely, the laws of large numbers and the central limit theorem.

We discuss briefly the four types of convergence in what follows, convergence in distribution, in probability, in mean square, and almost surely (or with probability 1). One may like to visit a probability theory text for a thorough study of these concepts. We will assume that the sequence of random variables and the limiting random variable are all defined on the same probability space.

The weakest form is the convergence in distribution which is defined as follows.

DEFINITION 1. *A sequence of random variables $X_1, X_2, \dots$ converges in distribution (or in law) to a random variable X if $\lim_{n \rightarrow \infty} F_{X_n}(x) = F_X(x)$ at each x for which $F_X(x)$ is continuous. This is written as either $F_n \rightarrow F$ in distribution or $X_n \rightarrow X$ in law.*

Convergence in distribution does not say that the random variables themselves converge in any reasonable sense, but only that their distribution functions converge. For example, let $Y_1, Y_2, \dots$ be i.i.d. r.v.s with the distribution function F_Y . For each $n \geq 1$, if we let $Z_n = Y_n + \frac{1}{n}$, then it is easy to see that $\{Z_n, n \geq 1\}$ converges in distribution to Y . However (assuming Y has variance σ_Y^2 and is independent of each Z_n), we see that $Z_n - Y$ has variance $2\sigma_Y^2$. Thus Z_n does not get close to Y as $n \rightarrow \infty$ in any reasonable sense, and $Z_n - Z_m$ does not get small as n and m both get large. As an even more trivial example, the sequence $\{Y_n, n \geq 1\}$ converges in distribution to Y .

DEFINITION 2. *A sequence of random variables $X_1, X_2, \dots$ converges in probability to a rv X if $\lim_{n \rightarrow \infty} P\{|X_n - X| > \epsilon\} = 0$ for every $\epsilon > 0$.*

Convergence in probability is a special case of convergence in distribution. Convergence in probability says quite a bit more than convergence in distribution. As an important example of this, consider the difference $Y_n - Y_m$ for n and m both large. If $\{Y_n, n \geq 1\}$ converges in probability to 0, then Y_n and Y_m are both close to 0 with high probability for large n and m , and thus close to each other. More precisely, $\lim_{m \rightarrow \infty, n \rightarrow \infty} P\{|Y_n - Y_m| > \epsilon\} = 0$ for every $\epsilon > 0$. If the sequence $\{Y_n, n \geq 1\}$ merely converges in distribution to some arbitrary distribution, then, as we saw, $Y_n - Y_m$ can be large with high probability, even when n and m are large.

Yet another kind of convergence is convergence in mean square, which is a special case of convergence in r th mean.

DEFINITION 3. *A sequence of random variables $X_1, X_2, \dots$ converges in r th mean to a rv X if $\lim_{n \rightarrow \infty} E|X_n - X|^r = 0$.*

If $r = 1$, it is called convergence in mean, and, if $r = 2$, it is called convergence in mean square. Convergence in probability does not imply convergence in r th mean, in general.

Convergence with probability one is often referred to as convergence a.s. (almost surely) and convergence a.e. (almost everywhere).

DEFINITION 4. *A sequence of random variables $X_1, X_2, \dots$ converges almost surely (or with probability one) to a rv X if $P\{\omega \in \Omega : \lim_{n \rightarrow \infty} X_n(\omega) = X(\omega)\} = 1$.*

In trying to understand the above definition, note that each sample point ω of the underlying sample space Ω maps to a sample value $X_n(\omega)$ of each rv X_n , and thus maps to a sample path $\{X_n(\omega), n \geq 1\}$. For any given ω , such a sample path is simply a sequence of real numbers. That sequence of real numbers might converge to $X(\omega)$ (which is a real number for the given ω), it might converge to something else, or it might not converge at all. Thus a set of ω exists for which the corresponding sample path $\{X_n(\omega), n \geq 1\}$ converges to $X(\omega)$, and a second set for which the sample path converges to something else or does not converge at all. Convergence a.s. of the sequence of rvs is thus defined to occur when the first set of sample paths above is an event that has probability one.

It can be shown that convergence a.s. is not implied by any of the other types of convergence we have discussed, and that a.s. convergence does imply convergence in probability and in distribution but not in mean square.

2 Limit Theorems

In this section, we will again briefly discuss two important limit theorems concerning sequences of random variables, the laws of large numbers and the central limit theorems.

The laws of large numbers are a collection of results in probability theory that describe the behavior of the arithmetic average of n rvs for large n . For any n rvs, $X_1, \dots, X_n$, the arithmetic average is the rv $(1/n) \sum_{i=1}^n X_i$. Since in any outcome of the experiment, the sample value of this rv is the arithmetic average of the sample values of $X_1, \dots, X_n$, this random variable is usually called the sample average. If $X_1, \dots, X_n$, are viewed as successive variables in time, this sample average is called the time average. Under fairly general assumptions, the standard deviation of the sample average goes to 0 with increasing n , and, in various ways depending on the assumptions, the sample average approaches the mean.

These results are central to the study of stochastic processes because they allow us to relate time averages (i.e., the average over time of individual sample paths) to ensemble averages (i.e., the mean of the value of the process at a given time). We first discuss what is called the weak law of large numbers for i.i.d. rvs. We also briefly discuss another of these results, the strong law of large numbers. The strong law requires considerable patience to understand. We also discuss the central limit theorem. It enhances our understanding of the weak law, and also is an important result in its own right.

Note that the idea of a sequence of rv's (i.e., a sequence of functions) approaching a constant is clearly much more involved than a sequence of numbers approaching a constant. The laws of large numbers bring out this central idea in a more fundamental, and usually more useful, way. We start the development by considering $X_1, X_2, \dots, X_n$ which are i.i.d. rvs with a finite mean μ and finite variance σ^2 . Let $S_n = X_1 + \dots + X_n$, and consider the sample average S_n/n . Recall that $\text{Var}(S_n) = n\sigma^2$. By applying the Chebyshev inequality to the sample average,

$$P \left\{ \left| \frac{S_n}{n} - \mu \right| > \epsilon \right\} \leq \frac{\sigma^2}{n\epsilon^2}.$$

This is an upper bound on the probability that S_n/n differs by more than ϵ from its

mean μ . Since the upper bound goes to 0 with increasing n , we have the following theorem.

THEOREM 5. (*Weak Law of Large Numbers*) Let $X_1, X_2, \dots$ be a sequence of i.i.d. rvs with a finite mean μ and finite variance σ^2 and let $S_n = X_1 + \dots + X_n$. Then S_n/n converges in probability to μ .

It is to be noted that the above result can be proved in the case of infinite variance too. If the convergence is with probability one, then it is said that the strong law of large numbers hold for the sequence.

The weak law of large numbers says that with high probability, S_n/n is close to μ for large n , but it establishes this via an upper bound on the tail probabilities rather than an estimate of what $F_{S_n/n}$ looks like. If we look at the shape of $F_{S_n/n}$ for various values of n , it can be observed that the function $F_{S_n/n}$ becomes increasingly compressed around μ as n increases. If we normalize the random variable S_n/n to zero mean and unit variance, we get a normalized rv $Z_n = (S_n/n - \mu)\sqrt{n}/\sigma$. Further inspection will show that the normalized distribution functions seem to be approaching a limiting distribution. The critically important central limit theorem states that there is indeed such a limit, and it is the normalized Gaussian distribution function.

THEOREM 6. (*Central limit theorem (CLT)*). Let $X_1, X_2, \dots$ be i.i.d. rv's with finite mean μ and finite variance σ^2 . Then for every real number z ,

$$\lim_{n \rightarrow \infty} P \left\{ \frac{S_n - n\mu}{\sigma\sqrt{n}} \leq z \right\} = \Phi(z),$$

where $\Phi(z)$ is the normal distribution function, i.e., the Gaussian distribution with mean 0 and variance 1,

$$\Phi(z) = \int_{-\infty}^z \frac{1}{\sqrt{2\pi}} e^{-\frac{y^2}{2}} dy.$$

Module 2: Introduction to Stochastic Processes

Lecture 1 What are Stochastic Processes?

The field of stochastic processes is essentially a branch of probability theory, treating probabilistic models that evolve in time. It is best viewed as a branch of mathematics, starting with the axioms of probability and containing a rich and fascinating set of results following from those axioms. Although the results are applicable to many areas, they are best understood in terms of their mathematical structure and interrelationships. Now, to apply the axiomatic probability results to a real-world system, a probability model for the given system needs to be developed. Mathematically precise results can then be derived within the model and translated back to the real world. If the model fits the system sufficiently well, real problems can be solved by analysis within the model. However, the mathematical models are almost always simplified approximations of reality, and therefore, precise results within the model become approximations in the real world. Thus, it is important and desirable to have as many models as possible in hand for selection, and as many results as possible concerning the models for ready inference. Of course, this is only one of many uses of the results within the models. Many a time results can be transported across models based on certain properties. Moreover, it is always interesting mathematically to unearth certain beautiful results within the models.

Sometimes an application area will have customary choices of models. For example, there is a well developed taxonomy of queueing models. A sound knowledge of the application area, combined with a sound knowledge of the behavior of these queueing models, often lets one choose a suitable model for a given system within the application area. In other cases, one can start with a particularly simple model and use the behavior of that model to gain insight about the application, and use this to iteratively guide the selection of more general models. The area of stochastic processes is no different to this back-and-forth process. Each improved version of a model gives new insights about the real world system and each additional complexity in the real world system drives the need for an improved model. What is more important, and herein lies the real power, in choosing a probability model for a real-world system is that a prospective choice depends heavily on prior understanding,

at both an intuitive and mathematical level, of results from the range of mathematical models that might be involved. Therefore, what we aim in this course is to guide the student in both the mathematical and intuitive understanding necessary in developing and using stochastic process models in studying application areas.

Application-oriented students often ask why it is important to understand axioms, theorems, and proofs in mathematical models when the applied results are approximate anyway. One answer is that theorems are transferable between applications, and thus enable insights from one application area to be transferred to another. Given the need for precision in the theory, however, why is an axiomatic approach needed?

If we try to avoid axiomatics, we can simply state and explain the major results. Engineering and science students learn to use calculus, linear algebra and undergraduate probability effectively without axioms or rigor. Why doesn't this work for more advanced probability and stochastic processes?

Probability theory has more than its share of apparent paradoxes, and these show up in very elementary arguments. Undergraduates are content with this, since they can postpone these questions to later study. Graduate students, however, usually want a foundation that provides understanding without paradoxes, and the axioms achieve this.

This course neither uses nor develops measure theory. Measure theory is undoubtedly important in understanding probability at a deep level, but most of the topics useful in many applications can be understood without measure theory. The required background in probability is provided in the previous module as review. Some exposure to linear algebra and analysis (especially concrete topics like vectors, matrices, and limits) is helpful. The most important prerequisite is the mathematical maturity and patience to couple precise reasoning with intuition. We will start from the very basic constructs of the stochastic processes in simple settings.

Let $(\Omega, \mathcal{F}, P)$ be a given probability space. A stochastic process is a collection of random variables $\{X(t), t \in T\}$, where T is some subset of $\mathbb{R}$, indexed by the parameter t , all defined on the same probability space $(\Omega, \mathcal{F}, P)$. The set S of possible values that the random variables $X(t)$ takes for each t of a stochastic process $\{X(t), t \in T\}$ is known as its *state space*, and the set of possible values for the

parameter t (i.e., the set T) is known as *parameter space*. Often, t is referred to as *time*. The adjective discrete/continuous and time/state is generally used depending on T or S is countable or not. For example, a discrete-time continuous-state stochastic process means a stochastic process with a countable T and uncountable S .

In case of discrete time, the parameter generally used is n , and the family is represented by $\{X_n, n = 0, 1, 2, \dots\}$. In case of continuous time both the symbols $\{X_t, t \in T\}$ and $\{X(t), t \in T\}$ (where T is a finite or infinite interval) are used. We will use $\{X(t), t \in T\}$ to denote any general stochastic process.

EXAMPLE 1. Consider the experiment of throwing a fair die repeatedly. If X_n is the outcome of the n^{th} throw, $n \geq 1$, then $\{X_n, n \geq 1\}$ is a family of random variables such that for a distinct value of n , one gets a distinct random variable X_n .

If Y_n denote the number of 6's in the first n throws, then $\{Y_n, n \geq 1\}$ is a stochastic process.

EXAMPLE 2. Consider the number of telephone calls received at a switchboard. Suppose X_t is the random variable which represents the number of incoming calls in an interval $(0, t)$ of duration t units. Then, the family $\{X_t, t \in T\}$ constitutes a stochastic process (where $T = [0, \infty)$).

Broadly, stochastic processes can be classified based on the nature of their parameter and state spaces as

- Discrete-parameter discrete-state stochastic processes
- Discrete-parameter continuous-state stochastic processes
- Continuous-parameter discrete-state stochastic processes
- Continuous-parameter continuous-state stochastic processes

Often, the discrete-state stochastic processes are referred to as stochastic chains or simply chains.

A stochastic process can be one-dimensional, two-dimensional or in general n -dimensional.

Examples:

- Consider a two-dimensional random vector $X(t) = (X_1(t), X_2(t))$, where $X_1(t)$ is the maximum temperature and $X_2(t)$ is the minimum temperature at time t . Then $\{X(t), t \geq 0\}$ is a two-dimensional stochastic process.
- Similarly, consider an n -dimensional random vector $X(t) = (X_1(t), X_2(t), \dots, X_n(t))$, then $\{X(t), t \geq 0\}$ is an n -dimensional stochastic process.

In the example given above, one can also create a complex-valued stochastic processes by considering $X(t) = \{X_1(t) + i X_2(t)\}$, where i is $\sqrt{-1}$.

EXAMPLE 3. (*Examples of discrete-time discrete-state stochastic process*)

- Consider the stochastic process $\{X_n, n = 1, 2, \dots\}$, where X_n = Number of customers in the barber shop after the departure of the n th customer from the shop.
Here, $T = \{1, 2, 3, \dots\}$ and $S = \{0, 1, 2, \dots\}$.
- Consider the stochastic process $\{X_n, n = 1, 2, \dots\}$, where X_n be the number of packets waiting in the buffer at the n th time unit in a communication router.
Here also, $T = \{1, 2, 3, \dots\}$ and $S = \{0, 1, 2, \dots\}$.

EXAMPLE 4. (*Examples of continuous-time discrete-state stochastic process*)

- Consider the stochastic process $\{X(t), t \geq 0\}$, where $X(t)$ = Number of customers in the barber shop at any time t .
Here, $T = [0, \infty)$ and $S = \{0, 1, 2, \dots\}$.
- Consider the stochastic process $\{X(t), t \geq 0\}$, where $X(t)$ = Number of customers eating food in a restaurant at any time t .
Here, $T = [0, \infty)$ and $S = \{0, 1, 2, \dots\}$.

EXAMPLE 5. (*Examples of discrete-time continuous-state stochastic process*)

- Consider the stochastic process $\{X_n, n = 1, 2, \dots\}$, where X_n = Water level in a dam observed at the n time unit (say, n th day).
Here, $T = \{1, 2, 3, \dots\}$ and $S = [0, \infty)$.
- Consider the stochastic process $\{X_n, n = 1, 2, \dots\}$, where X_n = Amount of one US dollar in Rupees at n th time unit.
Here also, $T = \{1, 2, 3, \dots\}$ and $S = [0, \infty)$.

EXAMPLE 6. (*Examples of continuous-time continuous-state stochastic process*)

- Consider the stochastic process $\{X(t), t \geq 0\}$, where $X(t)$ = Temperature of a particular city at any time t .
Here, $T = [0, \infty)$ and $S = \{x \mid -50 < x < 60\}$.
- Consider the stochastic process $\{X(t), t \geq 0\}$, where $X(t)$ = Water level in a dam observed at any time t .
Here, $T = [0, \infty)$ and $S = [0, \infty)$.

There is an alternative way to look at a stochastic process. Since a random variable itself is a function from Ω to $\mathbb{R}$, one can view a stochastic process as a function of two arguments $X(\omega, t)$, $\omega \in \Omega$, $t \in T$, and we have the following two approaches of defining a stochastic process:

1. As a family $\{X(., t), t \in T\}$ of random variables (one for each t).
2. As a collection $\{X(\omega, .), \omega \in \Omega\}$ of functions from T to $\mathbb{R}$ (one for each ω).

What we have seen above is the first approach. The second approach is referred to as the sample path approach to a stochastic process, since the function corresponding to an ω is referred to as a sample path (or realization or trajectory) of the stochastic process. Sometimes, the paths of random processes are also an important object of analysis, as in the case of Brownian motions. Since, in this view, a stochastic process is a collection of functions, one for each ω , one can take the set of all continuous functions to represent a process that changes continuously. If a more general setup, it is assumed that the sample paths of the process is a right-continuous function with left limits (sometimes referred to as regular right-continuous function or a *càdlàg* function).

Now, when do we say that we have complete knowledge of a stochastic process? The answer to this question is almost analogous to a similar question on random variables. Let $X = \{X(t), t \in T\}$ be a stochastic process. Then we would like to compute $P\{X(t_1) \in B\}$ for some fixed t_1 and for some Borel set B (say, for example, for an interval $B = [1, b]$). We would also be interested to know how the process at t_1 affects the process at another time t_2 . Thus, we need to talk about the joint distribution of $X(t_1)$ and $X(t_2)$. More generally, we would like to have all

finite-dimensional distributions of the process X , that is, probabilities of the sets of the form $\{X(t_1) \in B_1, \dots, X(t_n) \in B_n\}$, for any choice of $0 \leq t_1 \leq t_2, \dots \leq t_n$. One has to be careful in defining these probabilities in a consistent way (like we do in a simple two-dimensional case).

It follows immediately from the construction given above that the process is determined by its finite-dimensional distributions. It is important to note that we talking here only about the distributional properties of the process. Finite-dimensional distributions do not determine the continuity property of the sample paths. Let us look at the following example.

EXAMPLE 7. *Let $X(t) = 0$ for $0 \leq t \leq 1$ define a stochastic process X and Z be a uniformly distributed random variable on $[0, 1]$. Let $Y(t) = 0$ for $t \neq Z$ and $Y(t) = 1$ for $t = Z$ define another stochastic process Y . Then, for any fixed t , $P\{Y(t) = 0\} = 1$. It means that all one-dimensional distributions of X and Y are the same and so do all the finite-dimensional distributions. However, all the sample paths of the process X are continuous whereas every path of Y has a jump at the point Z .*

The two processes in the above example are called versions or modifications of one another as $P\{X(t) = Y(t)\} = 1$ for all t . One has continuous sample paths and the other does not. If we agree to pick any version of the process we want, then we can pick the continuous version when it exists. In general we choose the smoothest possible version of the process. Under some conditions on the joint distributions of the process which we do not intend to state here, we can ensure that there exists continuous or regular right-continuous versions of stochastic processes. Throughout the course, we will assume that we are dealing with only such processes.

Lecture 2 Classes of Stochastic Processes

We have seen one way of classifying stochastic processes based on the parameter space and state space. A more meaningful and useful classification is based the nature of random variables in the random process. In some cases, the members of the family $\{X(t), t \in T\}$ are mutually independent, but more often they are not. The random variables in stochastic processes that we come across generally exhibit some sort of dependence relationship among themselves. Based on these dependence relationship various types of processes can be defined. The various classical types of SPs are: Processes with independent increments, Markov processes, renewal processes, stationary processes, etc. We will now give an overview of some such processes. Let $(\Omega, \mathcal{F}, P)$ be the given probability space on which the random variables of a stochastic process are defined.

1 Independent Process

The simplest of all is a process of the form $\{X(t), t \in T\}$, where the random variables are all independent. In some sense, this is simply a sequence of independent random variables and their study is much simpler as one needs to know only the one-dimensional properties of the random variables.

2 Bernoulli Process

Consider a stochastic process $\{X_i, i = 1, 2, \dots\}$, where $X_i \sim$ Bernoulli distribution with parameter p . Also, X_i 's are i.i.d random variables, i.e., $X_i \sim B(1, p)$. Now define $S_n = \sum_{i=1}^n X_i$. Then $S_n \sim B(n, p)$, i.e. number of successes in n trials. Then, $\{S_n, n = 1, 2, \dots\}$ is called a Binomial process or Bernoulli process.

3 Processes with Independent Increments

If, for all $t_1, t_2, \dots, t_n, t_1 < t_2 < \dots < t_n$, the random variables

$$X(t_2) - X(t_1), X(t_3) - X(t_2), \dots, X(t_n) - X(t_{n-1})$$

are independent, then $\{X(t), t \in T\}$ is said to a process with independent increments. This means that the increments in non-overlapping intervals are independent.

EXAMPLE 1. *Poisson process (which we will study in detail later)*

4 Processes with Stationary Increments

If, for all s, t with $s \leq t$, the random variable $X(t) - X(s)$ has the same distribution as $X(t-s) - X_0$, then $\{X(t), t \in T\}$ is said to a process with independent increments. This means that the probability distribution of any increment $X(t)X(s)$ depends only on the length ts of the time interval.

EXAMPLE 2. *Poisson process (which we will study in detail later)*

5 Markov Processes

If, for $t_1 < t_2 < \dots < t_n < t$, for all n and for all $x_1, \dots, x_n$,

$$P\{a \leq X(t) \leq b | X(t_1) = x_1, \dots, X(t_n) = x_n\} = P\{a \leq X(t) \leq b | X(t_n) = x_n\}$$

then the process $\{X(t), t \in T\}$ is a Markov process.

A discrete-parameter discrete-state Markov process is known simply as a Markov chain.

From the above, a Markov Process is a stochastic process $\{X(t), t \in T\}$ such that, given the value of $X(s)$, the values of $X(t)$, $t > s$, do not depend on the values of $X(u)$, $u < s$, then the process is said to be a Markov process.

EXAMPLE 3. *Consider the experiment of throwing a fair die repeatedly. If Y_n denote the number of 6's in the first n throws, then $\{Y_n, n \geq 1\}$ is a Markov chain.*

6 Poisson Process

Consider process of arrival of customers at a railway reservation counter. Let $N(t)$ be the number of arrivals that has occurred during the time interval $[0, t]$. Then $\{N(t), t \geq 0\}$ is a continuous time discrete state stochastic process. Assume that

- The probability of an arrival in a small time interval $(t, t + \Delta t)$ is $\lambda \Delta t + O(\Delta t)$.
- Probability of more than one arrival in a small time interval Δt is $O(\Delta t)$.
- Non overlapping intervals are independent.

It can be shown that $P\{N(t) = k\} = e^{-\lambda t} \frac{(\lambda t)^k}{k!}$, for a fixed t , i.e., $N(t) \sim$ Poisson distribution (λt) . The stochastic process $\{N(t), t \geq 0\}$ is called a Poisson Process.

Let T_k be the time of the k th arrival where $k = 1, 2, \dots$. Let X_k be the successive inter arrival times of k th customer. Then $X_1 = T_1, X_2 = T_2 - T_1, \dots, X_k = T_k - T_{k-1}$. Then what is the distribution of $X_k, k = 1, 2, \dots$?

$$\begin{aligned} P(X_1 > t) &= P(N(t) = 0) = e^{-\lambda t} \\ P(X_1 \leq t) &= 1 - e^{-\lambda t} \end{aligned}$$

This implies that $X_1 \sim$ Exponential distribution (λ) . Similarly, $X_i \sim$ Exponential distribution $(\lambda), i = 1, 2, \dots$. Also, $\{X_i, i = 1, 2, \dots\}$ is a discrete time continuous state stochastic process. Further, since, $X_i \sim \text{Exp}(\lambda)$ and are i.i.d. random variables, $\{X_i, i = 1, 2, \dots\}$ is also a Renewal process.

7 Simple Random Walks

Consider integer valued iid random variables $X_i, i = 1, 2, \dots$. As a special case, consider

$$P(X_i = k) = \begin{cases} p, & k = 1 \\ 1 - p, & k = -1 \end{cases}$$

Define $S_n = \sum_{i=1}^n X_i$. Then $\{S_n, n = 1, 2, \dots\}$ is called a simple random walk. When $p = 1/2$, it is called a symmetric simple random walk. One can define a more general random walk by letting the random variables X_i to take on a finite number of values.

8 Second-Order Processes

A stochastic process $\{X(t), t \in T\}$ is called a second-order process if $E[X(t)]^2 < \infty$.

EXAMPLE 4. Let $X_n, n \geq 1$ be uncorrelated random variables with mean 0 and variance 1. Then, $\{X_n, n \geq 1\}$ is a second-order process.

9 Stationary Processes

Let $\{X(t), t \in T\}$ be a second-order process. The mean function of the process is defined by

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

and the covariance function (also called auto-covariance function) is defined by

$$C(s, t) = \text{Cov}(X(s), X(t)) = E[X(s)X(t)] - E[X(s)] E[X(t)].$$

A second-order process is called covariance stationary or weakly stationary or wide sense stationary if its mean function $m(t)$ is independent of t and its covariance function $C(s, t)$ is a function only of the time difference $|t - s|$, for all t, s , that is, $C(s, t) = f(t - s)$.

If, for arbitrary integer n and arbitrary $t_1, t_2, \dots, t_n$, the joint distribution of the vector random variables

$$(X(t_1), X(t_2), \dots, X(t_n)) \text{ and } (X(t_1 + h), X(t_2 + h), \dots, X(t_n + h))$$

are the same for all $h > 0$, then the SP $\{X(t), t \in T\}$ is said to be strictly stationary.

A process which is not stationary (in any sense) is said to be evolutionary.

EXAMPLE 5. Let $X_n, n \geq 1$ be uncorrelated random variables with mean 0 and variance 1. Then

$$C(n, m) = \text{Cov}(X_n, X_m) = E[X_m X_n] = \begin{cases} 0, & m \neq n; \\ 1, & m = n. \end{cases}$$

and so $\{X_n, n \geq 1\}$ is covariance stationary.

If X_n are also identically distributed, then $\{X_n, n \geq 1\}$ is strictly stationary.

EXAMPLE 6. The Poisson process is evolutionary.

10 Gaussian Processes

If the distribution of $(X(t_1), X(t_2), \dots, X(t_n))$ for all $t_1, t_2, \dots, t_n$ is multivariate normal, then $\{X(t), t \in T\}$ is said to be a Gaussian process.

Result: If a Gaussian process $\{X(t)\}$ is covariance stationary, then it is strictly stationary.

11 Martingales

A discrete parameter stochastic process $\{X_n, n \geq 0\}$ is called a (discrete-time) martingale (or a martingale process) if,

- (i) $E[|X_n|] < \infty$, and
- (ii) $E[X_{n+1}|X_n, X_{n-1}, \dots, X_0] = X_n$.

EXAMPLE 7. Let $\{Z_i, i = 1, 2, \dots\}$ be a sequence of i.i.d. random variables with mean 0 and let $X_n = \sum_{i=1}^n Z_i$. Then $\{X_n, n \geq 1\}$ is a martingale.

12 Brownian Motion

Let $(\Omega, \mathcal{F}, P)$ be the probability space. For each $\omega \in \Omega$, suppose there is a continuous function $W(t)$ of $t \geq 0$ that satisfies $W(0) = 0$ and that depends on ω . Then $\{W(t), t \geq 0\}$ is a (standard) Brownian motion (also called as a Wiener process) if for all $0 = t_0 < t_1 < \dots < t_m$ the increments

$$W(t_1) - W(t_0), W(t_2) - W(t_1), \dots, W(t_m) - W(t_{m-1})$$

are independent and each of these increments is normally distributed with

$$E[W(t_{i+1}) - W(t_i)] = 0, \quad \text{Var}[W(t_{i+1}) - W(t_i)] = t_{i+1} - t_i.$$

This stochastic process is an important subclass of processes that is also a Markov process, a martingale, a process with independent increments. This has applications in many areas and remains a basic element in financial modelling.

References

1. S Karlin and H M Taylor, “A First Course in Stochastic Processes”, 2nd edition, Academic Press, New York, 1975.
2. J Medhi, “Stochastic Processes”, 3rd edition, New Age International Publishers, 2009.
3. I U Narayan Bhatt, “Elements of Applied Stochastic Processes”, John Wiley & Sons, 2nd edition, 1984.

Module 3: Discrete-Time Markov Chains

Lecture 1 Markov Chains (MCs)

The material for this module is to be inserted here.

Module 4: Continuous-Time Markov Chains

Lecture 1 Definition, Infinitesimal Generator Matrix and Kolmogorov Equations

1 Introduction

Continuous-time Markov Chain is a very useful class of stochastic processes. It is abbreviated as CTMC. Many real world process which evolve randomly over time, can be modelled as CTMC. The examples include, the arrival pattern of customers to a coffee shop, the arrival and departure of planes to a airport. After knowing DTMC, a first guess about the CTMC is that it is simply a continuous time Markov chain in which transitions can happen at any time. This guess is indeed correct. As we proceed in the article, we get to know what is CTMC.

A Markov process can be classified into various categories depending on the state space and parameter space. A Markov process with discrete state space is called a Markov chain. A Markov chain with discrete parameter space is called a Discrete time Markov chain and with continuous parameter space is called a Continuous time Markov chain. Brownian motion is an example of a Markov process with continuous parameter space, which shall be discussed in module 7.

2 Definition

A discrete state, continuous parameter stochastic process $\{X(t); t \geq 0\}$ is called a CTMC if for $0 \leq t_0 < t_1 < \dots < t_n < t$ and $x, x_0, x_1, \dots, x_n \in S$ (state space), its conditional distribution satisfies

$$P\{X(t) \leq x \mid X(t_0) = x_0, X(t_1) = x_1, \dots, X(t_n) = x_n\} = P\{X(t) \leq x \mid X(t_n) = x_n\}.$$

When the Markov chain is time-homogeneous, i.e. $P[X(t+T) = j \mid X(T) = i]$ for any $T \geq 0$, does not depend on T , and this probability is denoted by $P_{ij}(t)$ and it is known as the stationary transition probability.

We consider following notations for further discussion.

Suppose state space, $S = \{0, 1, 2, 3, \dots\}$.

Then the random variable $X(t)$ has the probability mass function $\{\pi_j(t)\}$. That is

$$\pi(t) = P(X(t) = 0, X(t) = 1, X(t) = 2, \dots) = [\pi_0(t), \pi_1(t), \pi_2(t), \dots].$$

We have $\pi_j(t) \geq 0$, for $j \in S$ and $\sum_{j \in S} \pi_j(t) = 1$.

Initial state probability vector is given as follows.

$$\pi(0) = P(X(0) = 0, X(0) = 1, X(0) = 2, \dots) = [\pi_0(0), \pi_1(0), \pi_2(0), \dots]$$

Given $\pi_i(0)$ and $P_{ij}(t)$, we get for $j \in S$,

$$\begin{aligned} \pi_j(t) &= \text{Prob}(X(t) = j) \\ &= \sum_{i \in S} P(X(t) = j \mid X(0) = i) P(X(0) = i) \\ &= \sum_{i \in S} \pi_i(0) P_{ij}(t) \end{aligned}$$

Waiting time distribution

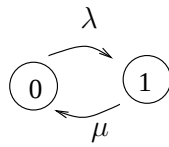
Consider a time-homogeneous CTMC $\{X(t); t \geq 0\}$. At time $t = 0$ initial state probability vector is known.

Let τ be a random variable, known as Waiting time, denoting time taken for a change of state from state i ,

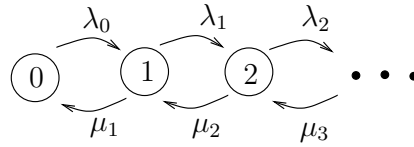
$$\begin{aligned} P[\tau > s + t \mid X(0) = i] &= P[\tau > s + t \mid X(s) = i] P[\tau > s \mid X(0) = i] \\ F_\tau^c(s + t) &= F_\tau^c(t) F_\tau^c(s) \\ \implies F_\tau^c(t) &= \exp(-\lambda t), \quad t > 0, \lambda > 0 \end{aligned}$$

Thus the waiting time τ has exponential distribution with parameter λ . The distribution is same for all i .

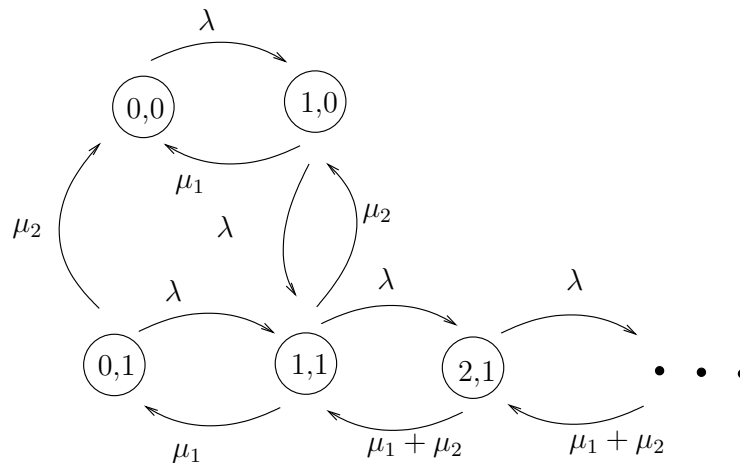
In Example 1, each of the two states 0 and 1 are accessible from the other one and hence are communicating states. Since this Markov chain has only two states which are communicating, it is an irreducible CTMC. Also the time spent in states



Example 1



Example 2



Example 3

0 and 1 is exponentially distributed with parameters λ and μ respectively.

In Example 2 the state space is countably infinite. The time spent in state 0 is exponentially distributed with parameter λ_0 . Whereas in state 1 the system can spend exponential time before moving in state 2 with parameter λ_1 and can spend exponential amount of time before moving to state 0 with parameter μ_1 , the two being independent of each other.

Therefore, the time spent in state 1 is the minimum of the two, which is again exponential with parameter $\lambda + \mu$.

Example 3 shows a two dimensional CTMC.

3 Chapman-Kolmogorov equations

The Chapman-Kolmogorov equations are the important equations for the time homogeneous CTMC. These give relation between transition probabilities of higher order in terms of lower ones.

$$\begin{aligned}
 P_{ij}(t+T) &= P[X(t+T) = j \mid X(0) = i] \\
 &= \sum_{k \in S} P[X(t+T) = j, X(t) = k \mid X(0) = i] \\
 &= \sum_{k \in S} P[X(t+T) = j \mid X(t) = k, X(0) = i] \times P[X(t) = k \mid X(0) = i] \\
 \implies P_{ij}(t+T) &= \sum_{k \in S} P_{kj}(T)P_{ik}(t) \quad \forall i, j \in S, t \geq 0, T \geq 0.
 \end{aligned}$$

Infinitesimal Generator Matrix

Define

$$\begin{aligned}
 q_{ij} &= \frac{d}{dt} P_{ij}(t) \big|_{t=0}, \quad i \neq j \\
 q_{ii} &= \frac{d}{dt} P_{ii}(t) \big|_{t=0}.
 \end{aligned}$$

That is,

$$\begin{aligned}
 q_{ij} &= \lim_{\Delta t \rightarrow 0} \frac{P_{ij}(\Delta t) - P_{ij}(0)}{\Delta t} \\
 q_{ii} &= \lim_{\Delta t \rightarrow 0} \frac{P_{ii}(\Delta t) - P_{ii}(0)}{\Delta t}.
 \end{aligned}$$

Using $P_{ij}(0) = 0$, for $i \neq j$ and $P_{ii} = 1$, we get,

$$P_{ij}(\Delta t) = q_{ij}\Delta t + o(\Delta t), \quad i \neq j$$

$$P_{ii}(\Delta t) = 1 + q_{ii}\Delta t + o(\Delta t).$$

Since $\sum_i P_{ij}(\Delta t)=1$, we get

$$\sum_j q_{ij} = 0 \quad \text{and} \quad q_{ij} \geq 0, \quad \forall i \neq j.$$

$$\text{Hence, } q_{ii} = - \sum_{j \neq i} q_{ij}.$$

Thus, using this we can construct a matrix $Q = [q_{ij}]$ such that,

$$(1) \quad q_{ij} \geq 0, \quad i \neq j$$

$$(2) \quad q_{ii} = - \sum_{j \neq i} q_{ij}.$$

This matrix Q is known as the infinitesimal generator matrix.

Kolmogorov equations

There are two sets of differential equations, namely Kolmogorov forward and Kolmogorov backward equations. The main difference between these two sets is the way these are solved in time. The forward equations is solved by moving forward in time, whereas backward equations are solved by moving backward in time.

Consider

$$P_{ij}(t+T) = \sum_{k \in S} P_{ik}(t) P_{kj}(T).$$

Differentiate w.r.t. T ,

$$P'_{ij}(t+T) = \sum_{k \in S} P_{ik}(t) \frac{d}{dT} P_{kj}(T).$$

Letting $T \rightarrow 0$,

$$\begin{aligned} P'_{ij}(t) &= \sum_{k \in S} P_{ik}(t) q_{kj} \\ P'(t) &= P(t)Q. \end{aligned}$$

The above equation is known as the forward Kolmogorov equation.

Similarly we can get,

$$\begin{aligned} P'_{ij}(t) &= \sum_{k \in S} q_{ik} P_{kj}(t) \\ P'(t) &= QP(t) \end{aligned}$$

which is known as the backward Kolmogorov equation.

Thus

$$\begin{aligned} P'(t) &= P(t)Q \\ P'(t) &= QP(t) \end{aligned}$$

are the forward and backward Kolmogorov equations, respectively.

Also consider

$$\pi_i(t+T) = P(X(t+T) = i) = \sum_{k \in S} \pi_k(t) P_{kj}(T).$$

Differentiate w.r.t. T ,

$$\pi_i(t+T) = \sum_{k \in S} \pi_k(t) \frac{d}{dT} P_{kj}(T).$$

Letting $T \rightarrow 0$,

$$\begin{aligned}\pi_i(t+T) &= \sum_{k \in S} \pi_k(t) q_{kj} \\ \pi'(t) &= \pi(t)Q.\end{aligned}$$

The above equation is known as the forward Kolmogorov equation.

Similarly we can get,

$$\begin{aligned}\pi'_i(t) &= \sum_{k \in S} q_{ik} \pi_k(t) \\ \pi'(t) &= Q\pi(t)\end{aligned}$$

which is known as the backward Kolmogorov equation.

Thus

$$\begin{aligned}\pi'(t) &= \pi(t)Q \\ \pi'(t) &= Q\pi(t)\end{aligned}$$

are the forward and backward Kolmogorov equations, respectively.

Lecture 2 Limiting and Stationary Distributions

Before giving the definition of limiting and stationary distributions, we will see some examples, wherein we find the transition probabilities $P(X(t) = j \mid X(0) = i)$, denoted as $\rho_{ij}(t)$. We will also see what happens to these probabilities as $t \rightarrow \infty$. These examples will pave the direction to give definition of limiting and stationary distributions.

Example 1: Consider a two state time homogeneous Markov chain $\{X(t); t \geq 0\}$. The state space $S = \{0, 1\}$, where 0 is the down or non operational state and 1 is the up or operational state.

Time spent in state 1 or the time spent by system in moving from state 1 to 0 is exponentially distributed with parameter λ and time spent in state 0 or time spent in moving from state 0 to 1 is exponentially distributed with parameter μ .

The Q matrix can therefore be written as,

$$\begin{bmatrix} -\mu & \mu \\ \lambda & -\lambda \end{bmatrix}$$

Now,

$$P_{01}(\Delta t) = \mu\Delta t + o(\Delta t)$$

$$P_{10}(\Delta t) = \lambda\Delta t + o(\Delta t).$$

The forward Kolmogorov equations are given by, for $i = 0, 1$

$$P'_{i0}(t) = -\mu P_{i0}(t) + \lambda P_{i1}(t)$$

$$P'_{i1}(t) = \mu P_{i0}(t) - \lambda P_{i1}(t).$$

Assume that $P_{11}(0)=1$; $P_{10}(0) = 0$.

For $i = 1$,

$$P_{10}(t) + P_{11}(t) = 1$$

$$P'_{11}(t) = -(\lambda + \mu)P_{11}(t) + \mu$$

$$P_{11}(t) = \frac{\mu}{\lambda + \mu} + ke^{-(\lambda + \mu)t}.$$

Hence we get transition probabilities as

$$P_{11}(t) = \frac{\mu}{\lambda + \mu} + \frac{\lambda}{\lambda + \mu}e^{-(\lambda + \mu)t}$$

$$P_{10}(t) = \frac{\lambda}{\lambda + \mu} - \frac{\lambda}{\lambda + \mu}e^{-(\lambda + \mu)t}.$$

Taking limit $t \rightarrow \infty$ we can get the steady state distribution.

Example 2: Consider a stochastic process $\{X(t); t \geq 0\}$ with rate matrix as,

$$Q = \begin{pmatrix} -2 & 1 & 1 \\ 1 & -1 & 0 \\ 2 & 1 & -3 \end{pmatrix}.$$

Eigenvalues of Q are 0, -2 and -4.

Hence $P_{11}(t) = k_1 + k_2e^{-2t} + k_3e^{-4t}$.

Using $P_{11}(0) = 1$; $P'_{11}(0) = q_{11} = -2$; $P''_{11}(0) = q_{11}^2 = 7$ we get,

$$P_{11}(t) = \frac{3}{8} + \frac{1}{4}e^{-2t} + \frac{3}{8}e^{-4t}.$$

1 Limiting distribution

For an irreducible, positive recurrent CTMC, limits $\lim_{t \rightarrow \infty} P_{ij}(t)$ exist. This is called limiting distribution of corresponding CTMC. If it is independent of initial state i , then we can denote $\lim_{t \rightarrow \infty} P_{ij}(t)$ by π_j , independent of i .

We can form the vector $\pi = (\pi_0, \pi_1, \dots)$ and since limiting distribution is a probability distribution, we have ([4])

$$\pi_j \geq 0 \text{ and } \sum_j \pi_j = 1.$$

2 Stationary distribution

The vector π is called a stationary distribution of a CTMC if $\pi = (\pi_0, \pi_1, \dots)$ satisfies

1. $\pi_j \geq 0 \ \forall \ j$
2. $\sum_j \pi_j = 1$
3. $\pi Q = 0$.

Main result for stationary distribution

For an irreducible, positive recurrent CTMC, the stationary distribution exists and it is unique. The vector $\pi = (\pi_0, \pi_1, \dots)$ is uniquely determined by,

$$\pi Q = 0, \quad \sum_j \pi_j = 1.$$

Example 3: Consider a two state CTMC with Q matrix as

$$\begin{pmatrix} -\mu & \mu \\ \lambda & -\lambda \end{pmatrix}.$$

It can be clearly seen that the above CTMC is irreducible and positive recurrent and hence by the previous result the stationary distribution exists and is unique. Stationary distribution is given by

$$\begin{aligned} \pi Q &= 0, \quad \sum_j \pi_j = 1 \\ (\pi_0, \pi_1) \begin{pmatrix} -\mu & \mu \\ \lambda & -\lambda \end{pmatrix} &= (0, 0) \end{aligned}$$

$$\begin{aligned}
&\Rightarrow -\mu\pi_0 + \lambda\pi_1 = 0 \\
&\Rightarrow \lambda\pi_1 = \mu\pi_0 \\
&\Rightarrow \pi_1 = \frac{\mu}{\lambda}\pi_0 \\
\text{Also } \pi_0 + \pi_1 = 1 &\Rightarrow \left(1 + \frac{\mu}{\lambda}\right)\pi_0 = 1 \\
&\Rightarrow \pi_0 = \frac{\lambda}{\lambda + \mu}, \quad \pi_1 = \frac{\mu}{\lambda + \mu}.
\end{aligned}$$

Example 4: Consider the CTMC with Q matrix as

$$Q = \begin{pmatrix} -2 & 1 & 1 \\ 1 & -1 & 0 \\ 2 & 1 & -3 \end{pmatrix}.$$

It can be seen that every state can communicate with every other state and hence the given CTMC is irreducible and positive recurrent.

To find the steady state solution consider $\pi Q = 0$ and $\sum_{j=1}^3 \pi_j = 1$.

$$\begin{aligned}
&\Rightarrow -2\pi_1 + \pi_2 + \pi_3 = 0 \\
&\Rightarrow \pi_1 - \pi_2 = 0 \\
&\Rightarrow 2\pi_1 + \pi_2 - 3\pi_3 = 0.
\end{aligned}$$

Using the above equations we get $\pi_1 = \frac{3}{8}$, $\pi_2 = \frac{4}{8}$, $\pi_3 = \frac{1}{8}$.

3 Some notes

1. Steady-state or equilibrium distribution [1]

For $X(t)$, the number of customers in the system at time t , its probability distribution is denoted by

$$p_n(t) = P(X(t) = n \mid X(0) = 0)$$

$p_n(t)$ is time dependent. For complete description of the CTMC, we need to consider time-dependent or transient solutions. But it is often difficult to obtain such solutions. Further in many practical situations, one needs to know the behavior in steady state, i.e., when the system reaches an equilibrium state, after being in operation for a pretty long time.

It is easier and convenient to determine

$$p_n = \lim_{t \rightarrow \infty} p_n(t)$$

provided the limit exists. When this limit exists, it is said that the system has reached steady state or equilibrium; and corresponding probability distribution $\{p_n\}$, is called steady-state or equilibrium distribution.

2. Time Reversible CTMC

For an irreducible CTMC, if there exist a probability distribution π that satisfy the time-reversibility equations

$$\pi_i q_{ij} = \pi_j q_{ji} \quad \forall i, j \in S$$

then the CTMC has positive recurrent states, time reversible and the solution π is unique stationary distribution.

Lecture 3 Poisson Processes - I

We can model a real life phenomenon, wherein we are interested in counting the number arrival of some random event, by Poisson process. Although some assumptions are there for modelling real life phenomenon by Poisson process, but these can be easily incorporated.

Postulates for Poisson Process

Let $\{N(t); t \geq 0\}$ be continuous time and non-negative integral valued stochastic process.

Let $p_n(t) = P(N(t) = n)$.

Further assume that

1. $N(0) = 0$. That is

$$p_0(0) = 1, \quad p_n(0) = 0 \quad \text{for } n \neq 0 \quad (1)$$

2. Arrivals in non-overlapping intervals are independent.
3. $p_n(t)$ depends only on the length t on interval and is independent of where this interval is situated.
4. In an interval of infinitesimal length h , the probability of exactly one occurrence is $\lambda h + o(h)$ and that of more than one occurrence is of $o(h)$.

That is if the interval between t and $t + h$ is of very short duration h , then

$$p_1(h) = \lambda h + o(h)$$

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

$$\Rightarrow p_0(h) = 1 - \lambda h + o(h).$$

Theorem : Under the above postulates (2), (3) and (4), $N(t)$ follows Poisson distribution with mean λt . [1]

Proof : Consider $p_n(t + h)$ for $n \geq 0$.

n events by epoch $t + h$ can happen in the following mutually exclusive ways

$$A_1, A_2, A_3, \dots, A_n$$

For $n \geq 1$

A_1 : n events by epoch t and no event between t and $t + h$. That is

$$\begin{aligned} P(A_1) &= P(N(t) = n)P(N(h) = 0 \mid N(t) = n) \\ &= p_n(t)p_0(h) \\ &= p_n(t)(1 - \lambda h) + o(h) \end{aligned}$$

A_2 : $n - 1$ events by epoch t and 1 event between t and $t + h$. That is

$$\begin{aligned} P(A_2) &= P(N(t) = n - 1)P(N(h) = 1 \mid N(t) = n - 1) \\ &= p_{n-1}(t)p_1(h) \\ &= p_{n-1}(t)(\lambda h) + o(h) \end{aligned}$$

For $n \geq 2$

A_3 : $n - 2$ events by epoch t and 2 events between t and $t + h$. That is

$$\begin{aligned} P(A_3) &= p_{n-2}(t)p_2(h) \\ &\leq p_2(h) \end{aligned}$$

and so on for $P(A_4), P(A_5), \dots$

We thus have
$$\sum_{k=2}^n P(A_{k+1}) \leq \sum_{k=2}^n p_k(h) = o(h)$$

and so

$$\begin{aligned} p_n(t + h) &= p_n(t)(1 - \lambda h) + p_{n-1}(t)(\lambda h) + o(h), \quad n \geq 1 \\ \frac{p_n(t + h) - p_n(t)}{h} &= -\lambda p_n(t) + \lambda p_{n-1}(t) + \frac{o(h)}{h} \\ \lim_{h \rightarrow 0} \frac{p_n(t + h) - p_n(t)}{h} &= -\lambda p_n(t) + \lambda p_{n-1}(t) + \lim_{h \rightarrow 0} \frac{o(h)}{h} \end{aligned}$$

$$p'_n(t) = -\lambda(p_n(t) - p_{n-1}(t)), \quad n \geq 1. \quad (2)$$

For $n = 0$

$$\begin{aligned} P_0(t+h) &= p_0(t)p_0(h) \\ &= p_0(t)(1 - \lambda h) + o(h) \\ \frac{p_0(t+h) - p_0(t)}{h} &= -\lambda p_0(t) + \frac{o(h)}{h} \\ \lim_{h \rightarrow 0} \frac{p_0(t+h) - p_0(t)}{h} &= -\lambda p_0(t) + \lim_{h \rightarrow 0} \frac{o(h)}{h} \end{aligned}$$

$$p'_0(t) = -\lambda(p_0(t)). \quad (3)$$

From (1), (2) and (3), we get

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

Thus $N(t)$ follows Poisson distribution with parameter λt .

1 Definition

A stochastic process $\{N(t); t \geq 0\}$ is said to be a Poisson process, with intensity or rate $\lambda > 0$, if the following conditions are satisfied:

- It starts from zero, i.e. $N(0) = 0$.
- It has stationary and independent increments.
- For every $t > 0$, $N(t)$ has a Poisson distribution $P(\lambda t)$.

2 Properties of Poisson processes

1. Time taken for first arrival is $Exp(\lambda)$. [1]

Proof : Let T denote the time of first arrival.

$$\begin{aligned} P(T > t) &= P[N(t) = 0] \\ &= \frac{e^{-\lambda t} (\lambda t)^0}{0!} \\ &= e^{-\lambda t} \end{aligned}$$

$\therefore T$ is $Exp(\lambda)$.

2. Superposition [1]

Suppose there are two independent Poisson arrival processes, with the parameters λ_1 and λ_2 . Then the superposition of the two processes will again be a Poisson process with parameter $(\lambda_1 + \lambda_2)$.

Similarly, if there are n independent Poisson streams with parameters, $\lambda_1, \lambda_2, \dots, \lambda_n$, then the combined stream or the superimposed stream is a Poisson stream with parameter $(\lambda_1, \lambda_2, \dots, \lambda_n)$. This finds use in many telecommunication applications.

Proof : Let $N_1(t)$ and $N_2(t)$ be two independent Poisson processes, with the parameters λ_1 and λ_2 , respectively. Let

$$N(t) = N_1(t) + N_2(t).$$

Its quite obvious that $N(t) \in \{0, 1, 2, \dots\}$.

$$\begin{aligned} P(N(t) = n) &= \sum_{r=0}^n P(N_1(t) = r) P(N_2(t) = n - r) \\ &= \sum_{r=0}^n \frac{e^{-\lambda_1 t} (\lambda_1 t)^r}{r!} \frac{e^{-\lambda_2 t} (\lambda_2 t)^{n-r}}{(n-r)!} \\ &= \frac{e^{-(\lambda_1 + \lambda_2)t} \{(\lambda_1 + \lambda_2)t\}^n}{n!}, \quad n = 0, 1, 2, \dots \end{aligned}$$

Thus $N(t)$ follows Poisson distribution with parameter $(\lambda_1 + \lambda_2)t$.

3. Decomposition [1]

A random selection from a Poisson process yields a Poisson process. Suppose that $N(t)$, the number of occurrences of an event E in an interval of length t is a Poisson process with parameter λ . Suppose also that each occurrence of E has a constant probability p of being recorded, and that the recording of an occurrence is independent of that of other occurrences and also of $N(t)$.

If $M(t)$ is the number of occurrences recorded in an interval of length t , then $M(t)$ is also a Poisson process with parameter λp .

Proof : The event $\{M(t) = n\}$ can happen in the following mutually exclusive ways:

A_r : E occurs $(n+r)$ times by epoch t and exactly n out of $(n+r)$ occurrences are recorded, probability of each occurrence recorded being p , ($r = 0, 1, 2, \dots$).

We have

$$\begin{aligned} P(A_r) &= P(E \text{ occurs } (n+r) \text{ times by epoch } t) \times \\ &\quad P(n \text{ occurrences are recorded given that the number of occurrences is } n+r) \\ &= \frac{e^{-\lambda t} (\lambda t)^{n+r}}{(n+r)!} \times^{n+r} C_n p^n q_r \end{aligned}$$

Hence

$$\begin{aligned} P(M(t) = n) &= \sum_{r=0}^{\infty} P(A_r) \\ &= \frac{e^{-\lambda p t} (\lambda p t)^n}{(n)!}. \end{aligned}$$

3. Poisson process and Binomial distribution [1]

If $\{N(t)\}$ is a Poisson process and $s < t$, then $(N(s) \mid N(t) = n)$ is Binomial($n, \frac{s}{t}$).

Proof : For $k = 0, 1, 2, \dots, n$

$$\begin{aligned}
 P(N(s) = k \mid N(t) = n) &= \frac{P(N(s) = k, N(t) = n)}{P(N(t) = n)} \\
 &= \frac{P(N(s) = k, N(t) - N(s) = n - k)}{P(N(t) = n)} \\
 &= \frac{P(N(s) = k)P(N(t) - N(s) = n - k)}{P(N(t) = n)} \\
 &= \frac{\frac{e^{-\lambda s}(\lambda s)^k}{k!} \times \frac{e^{-\lambda(t-s)}(\lambda(t-s))^{n-k}}{(n-k)!}}{\frac{e^{-\lambda t}(\lambda t)^n}{n!}} \\
 &= {}^nC_k \left(\frac{s}{t}\right)^k \left(1 - \frac{s}{t}\right)^{n-k}
 \end{aligned}$$

4. Poisson process and Uniform distribution [7]

Given that $N(t) = n$, the n arrival times $S_1, \dots, S_n$ have the same distribution as the order statistics corresponding to n independent random variables uniformly distributed on the interval $(0, t)$.

Proof : To obtain the conditional density of $S_1, \dots, S_n$ given that $N(t) = n$, note that for $0 < S_1 < \dots < S_n < t$ the event that $S_1 = s_1, S_2 = s_2, \dots, S_n = s_n, N(t) = n$ is equivalent to the event that the first $n + 1$ interarrival times satisfy $T_1 = s_1, T_2 = s_2 - s_1, \dots, T_n = s_n - s_{n-1}, T_{n+1} > t - s_n$.

Now, by using the property that $T_1, T_2, \dots, T_n$ are iid exponential random variables having mean $\frac{1}{\lambda}$.

$$\begin{aligned}
 f(s_1, s_2, \dots, s_n \mid n) &= \frac{f(s_1, s_2, \dots, s_n, n)}{P\{N(t) = n\}} \\
 &= \frac{\lambda e^{-\lambda s_1} \times \lambda e^{-\lambda(s_2 - s_1)} \times \dots \times \lambda e^{-\lambda(s_n - s_{n-1})} \times e^{-\lambda(t - s_n)}}{e^{-\lambda t} \frac{(\lambda t)^n}{n!}} \\
 &= \frac{n!}{t^n}, \quad 0 < s_1 < s_2 < \dots < s_n < t
 \end{aligned}$$

which proves the result.

Lecture 4 Poisson Processes - II

1 Examples

Example 1: Consider the car insurance claims reported to insurer. Assume that, the average rate of occurrence of claims is 10 per day. Also, assume that this rate is constant throughout the year and at different times of the day. Further assume that in a sufficiently short time interval there can be atmost one claim. What is the probability that there are less than 2 claims reported on a given day? What is the probability that time until the next reported claim is less than 2 hours?

Solution : The given situation can be modelled as Poisson process. The arrival of car insurance claims to the insurer is a Poisson process, $\{N(t); t \geq 0\}$, with rate λ as 10 per day.

Required probability that there are less than 2 claims reported on a given day

$$\begin{aligned} P(N(t+1) - N(t) < 2) &= P(N(1) < 2) \\ &= P(N(1) = 0) + P(N(1) = 1) \\ &= \frac{e^{-\lambda \times 1} (\lambda \times 1)^0}{0!} + \frac{e^{-\lambda \times 1} (\lambda \times 1)^1}{1!} \\ &= e^{-\lambda} (1 + \lambda). \end{aligned}$$

Required probability that time until the next reported claim is less than 2 hours,

given that there are 2 claims till now (2 hours = $\frac{2}{24}$ days)

$$\begin{aligned}
 P(N(t + \frac{2}{24}) > 2 \mid N(t) = 2) &= P(N(t + \frac{2}{24}) \geq 3 \mid N(t) = 2) \\
 &= \frac{P(N(t + \frac{2}{24}) \geq 3, N(t) = 2)}{P(N(t) = 2)} \\
 &= \frac{P(N(t + \frac{2}{24}) - N(t) \geq 1) \times P(N(t) = 2)}{P(N(t) = 2)} \\
 &= P(N(t + \frac{2}{24}) - N(t) \geq 1) \\
 &= P(N(\frac{2}{24}) \geq 1) \\
 &= 1 - P(N(\frac{2}{24}) = 0) \\
 &= 1 - \frac{e^{-\lambda \times \frac{2}{24}} (\lambda \times \frac{2}{24})^0}{0!} \\
 &= e^{-\lambda \times \frac{2}{24}}.
 \end{aligned}$$

Example 2: Consider the situation of waiting for a bus on the bus stand. Assume that the bus arrivals(in minutes) follow a Poisson process with parameter 5. That is, the mean inter arrival time between the buses is 5 minutes. Suppose you come to the bus stand at some time. What is the average waiting time to get the bus?

Solution : Suppose one arrives the bus stand at time t.

Let W be the waiting time. Let $X_1, X_2, X_3, \dots$ be the inter-arrival times.

Since exponential distribution is memoryless, hence the waiting time will also follow exponential distribution with parameter λ .

Hence average waiting time is

$$E[W] = \frac{1}{\lambda} = 5.$$

2 Non-homogeneous Poisson process

Definition : A counting process $\{N(t); t \geq 0\}$, with $N(0) = 0$ is called a non-homogeneous Poisson process if it has independent but not necessarily stationary increments, and for $s < t$, $N(t) - N(s)$ has a Poisson distribution with mean $m(t) - m(s)$, where $m(t)$ is a continuous increasing function on $[0, \infty)$ with $m(0) = 0$.

Here a Poisson distribution of mean 0 is defined

Let $N(t)$ denote the number of customers arriving in $(0, t)$. The arrival process $\{N(t); t \geq 0\}$ has a Poisson distribution

$$P(N(t) = i) = \frac{(\lambda(t))^i e^{-\lambda(t)}}{i!}, \quad i = 0, 1, \dots$$

where,

$$\lambda(t) = \int_0^t \lambda(x) dx.$$

Example 3: For a nonhomogenous Poisson process the intensity function is given by

$$\lambda(t) = \begin{cases} 5 & \text{if } t \in (1, 2], (3, 4], \dots \\ 3 & \text{if } t \in (0, 1], (2, 3], \dots \end{cases}$$

Find the probability that the number number of observed occurrences in the time period $(1.25, 3]$ is more than two.

Solution : $N(3) - N(1.25)$ has a Poisson distribution with mean

$$m(3) - m(1.25) = \int_{1.25}^3 \lambda(t) dt = \int_1^2 .255 dt + \int_2^3 3 dt = 6.75.$$

Hence, $P\{N(3) - N(1.25) > 2\} = 1 - e^{-6.75} (1 + 6.75 + \frac{(6.75)^2}{2}) = 0.9642515816$.

3 Compound Poisson process

Definition : Consider a Poisson process $\{N(t); t \geq 0\}$. Let X'_i 's be iid random variables which are independent of $N(t)$, also. Consider $X(t) = X_1 + X_2 + \dots + X_{N(t)}$. Then $\{X(t); t \geq 0\}$ is a compound Poisson process. If $P(X_i = 1) = 1, \forall i$, then $\{X(t); t \geq 0\}$ is a Poisson process.

Also $E[X(t)] = \lambda t E(X_i), \text{Var}[X(t)] = \lambda t E[X_i^2]$

Example 4: Insurance claims occur according to a Poisson process of rate $\lambda = 10$ per month and the amounts of claims are iid of mean 500 and standard deviation 200. Find the expected total claims in a three month period. Based on the normal approximation, estimate the probability that the total claim will not exceed 20,000.[4]

Solution : Let $N(t)$ be the Poisson process that counts the number of claims and let Y_i be the iid claim amounts. Then the total amount $X(t)$ of claims by time t is a compound Poisson process.

The expected total claim = $E[X(3)] = 10 \times 3 \times 500 = 15,000$.

$Var[X(3)] = 10 \times 3 \times (200^2 + 500^2) = 8,700,000$.

Using the normal distribution $N(15,000, 8,700,000)$ with Z being a standard normal random variable, we get

$$P[X(3) \leq 20,000] = P\left(Z \leq \frac{20,000 - 15,000}{\sqrt{8,700,000}}\right) = .9554.$$

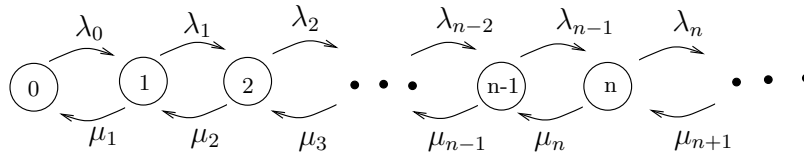
Lecture 5 Birth Death Process

1 Definition

A CTMC $\{X(t); t \geq 0\}$ with the state space $\{0, 1, 2, \dots\}$ is a birth death process if there exist constants λ_i (≥ 0), $i = 0, 1, 2, \dots$ and μ_i (≥ 0), $i = 1, 2, \dots$ with

$$\begin{aligned} q_{i,i+1} &= \lambda_i \\ q_{i,i-1} &= \mu_i \\ q_i &= -(\lambda_i + \mu_i) \\ q_{i,j} &= 0 \text{ for } |i - j| > 1. \end{aligned}$$

State Transition Diagram : The state transition diagramme of a birth death process is shown below. The corresponding infinitely generator matrix is as follows:



BDP

$$Q = [q_{ij}] = \begin{bmatrix} -\lambda_0 & \lambda_0 & 0 & 0 & \dots \\ \mu_1 & -(\lambda_1 + \mu_1) & \mu_1 & 0 & \dots \\ 0 & \mu_2 & -(\lambda_2 + \mu_2) & \mu_2 & \dots \\ \vdots & \vdots & \vdots & \vdots & \vdots \end{bmatrix}$$

Steady state and time-dependent analysis :

$$\begin{aligned} \pi'_0(t) &= -\lambda_0\pi_0(t) + \mu_1\pi_1(t) \\ \pi'_1(t) &= \lambda_0\pi_0(t) - (\lambda_1 + \mu_1)\pi_1(t) + \mu_2\pi_2(t) \\ \pi'_n(t) &= \lambda_{n-1}\pi_{n-1}(t) - (\lambda_n + \mu_n)\pi_n(t) + \mu_{n+1}\pi_{n+1}(t), \quad n = 2, 3, \dots \end{aligned}$$

For steady state probabilities, we take limit $t \rightarrow \infty$,

$$\pi_i = \lim_{t \rightarrow \infty} P(X(t) = i), \quad i = 0, 1, 2, \dots$$

With the assumption of irreducible and positive recurrent, we get

$$\begin{aligned} 0 &= -\lambda_0 \pi_0 + \mu_1 \pi_1 \\ 0 &= \lambda_{n-1} \pi_{n-1} - (\lambda_n + \mu_n) \pi_n + \mu_{n+1} \pi_{n+1}, \quad n = 1, 2, \dots \end{aligned}$$

Solving above system of equations, we get

$$\pi_n = \frac{\lambda_{n-1} \lambda_{n-2} \dots \lambda_0}{\mu_n \mu_{n-1} \dots \mu_1} \pi_0, \quad n = 1, 2, \dots$$

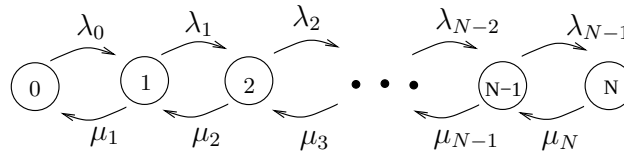
where by using $\sum_{n=0}^{\infty} \pi_n = 1$, we get π_0 as follows

$$\pi_0 = \frac{1}{1 + \frac{\lambda_0}{\mu_1} + \frac{\lambda_0 \lambda_1}{\mu_1 \mu_2} + \dots}$$

provided the denominator in the above expression is finite.

2 Finite BDP

Finite BDP is the BDP with finite state space $\{0, 1, 2, \dots, N\}$.



Finite BDP

We have steady state probabilities as follows:

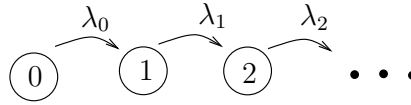
$$\pi_n = \frac{\lambda_{n-1} \lambda_{n-2} \dots \lambda_0}{\mu_n \mu_{n-1} \dots \mu_1} \pi_0, \quad n = 1, 2, \dots, N$$

where

$$\pi_0 = \frac{1}{1 + \frac{\lambda_0}{\mu_1} + \frac{\lambda_0 \lambda_1}{\mu_1 \mu_2} + \dots + \frac{\lambda_{N-1} \lambda_{N-2} \dots \lambda_0}{\mu_N \mu_{N-1} \dots \mu_1}}.$$

3 Pure birth process

A BDP is said to be Pure birth process if the death rates, $\mu_i = 0$ for $i = 1, 2, \dots$



Pure birth process

With this definition of pure birth process we get, Poisson process is pure birth process with same birth rates.

Consider special case of pure birth process with birth rates, $\lambda_i = \lambda$ for $i = 0, 1, \dots$

We have

$$\begin{aligned}\pi'_0(t) &= -\lambda\pi_0(t) \\ \pi'_1(t) &= \lambda\pi_0(t) - \lambda\pi_1(t) \\ \pi'_n(t) &= \lambda\pi_{n-1}(t) - \lambda\pi_n(t), \quad n = 2, 3, \dots\end{aligned}$$

with condition $\pi_0(0) = 1$ and $\pi_i(0) = 0$, $i \neq 0$.

Solving above differential equations, we get

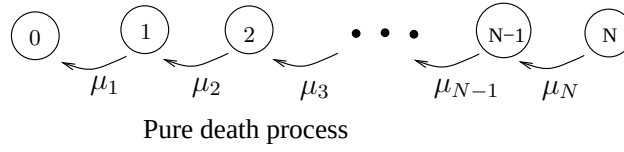
$$\pi_n(t) = \frac{e^{-\lambda t} (\lambda t)^n}{n!}, \quad n = 0, 1, 2, \dots$$

4 Pure death process

A BDP is said to be Pure death process if the birth rates, $\lambda_i = 0$ for $i = 1, 2, \dots, N$.

Here state space is $\{N, N-1, \dots, 0\}$. Initial condition is

$$\pi_N(0) = 1 \quad \text{and} \quad \pi_i(0) = 0, \quad i \neq N$$



Consider special case of pure death process with death rates, $\mu_i = \mu$ for $i = 0, 1, \dots, N$. We have

$$\begin{aligned}\pi'_0(t) &= -\mu\pi_1(t) \\ \pi'_1(t) &= -\mu\pi_1(t) + \mu_2\pi_2(t) \\ \pi'_n(t) &= -\mu_n\pi_n(t) + \mu_{n+1}\pi_{n+1}(t), \quad n = 2, 3, \dots, N\end{aligned}$$

with condition $\pi_0(0) = 1$ and $\pi_i(0) = 0$, $i \neq 0$.

Solving above differential equations, we get

$$\pi_n(t) = {}^N C_n (e^{-\mu t})^n (1 - e^{-\mu t})^{N-n}, \quad n = 0, 1, 2, \dots, N.$$

That is time dependent distribution is Binomial($N, e^{-\mu t}$).

Lecture 6 $M/M/1$ Queueing Model

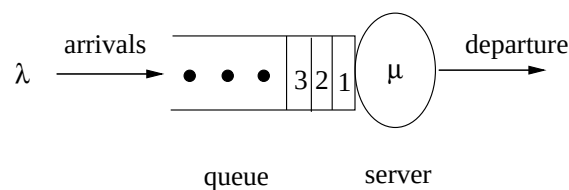
A queueing model is useful for study of queues or waiting lines. It gives nice mathematical formulation of some terms, average no. of customers in a queue, waiting time in a queue, which are useful to study some empirical properties of queue.

1 Queueing systems

Consider a situation where customers (need not be human beings) arrive, wait for their turn to receive service and after being serviced and they leave. The underlying system in such a situation is called, Queueing systems.

Some examples are as follows.

- Supermarket checkouts
- Railway reservation counters
- Computer service center
- Calls allocation in communication system



A queueing system is characterized by the following characteristics.

- Arrival pattern of customers.
- Service pattern of customers.
- Number of service channels.

- System capacity.
- Queue discipline.

For example, consider a room where interview for certain job is being held. Rule is that, on interview day, first 100 candidates who come, will be interviewed. There are 4 interviewers, sitting together, who take interview of one candidate together. The time taken by a candidate to come for the interview, as well as time taken by interviewers to finish an interview, both are exponentially distributed with some fixed parameters. So for this example we have following.

- Arrival pattern of customers - Inter-arrival time is random with exponential distribution.
- Service pattern of customers - Service time is random with exponential distribution.
- Number of service channels - 1 (although there are 4 interviewers)
- System capacity - 100
- Queue discipline - First come first service

Kendall Notation (A/B/X/Y/Z) : Kendall notation is the standard notation used to represent different types of queueing model.

- A: Distribution of inter-arrival times
- B: Distribution of service times
- X: Number of servers
- Y: Maximum number of customers in system
- Z: Queue Discipline

A,B are chosen from the set:

- M=Exponential

- D=Deterministic
- E_k =Erlang type k(k=1,2,...)
- H_k =Hyper Exponential type k
- G=General

Markovian Queues: M/M/1, M/M/c, M/M/c/k

Non-Markovian Queues: M/G/1, G/M/1, GI/G/c, M/D/1

2 M/M/1 Queueing model

M/M/1 queueing model is characterized by following characteristic.

- Arrival process: Poisson process with rate λ
- Service times: Exponential with rate μ
- Service times and inter-arrival times are independent.
- Single server.
- Infinite capacity in a system.
- $N(t)$: Number of customers in a system at time t(state)

CTMC formulation

- (i) Transitions due to arrival or departure of customers.
- (ii) Only nearest neighbors transitions are allowed.
- (iii) State of the process at time t: $N(t) = i, (i \geq 0)$.

(iv) $\{N(t)/t \geq 0\}$ is a continuous-time Markov chain with

$$\begin{aligned} q_{i,i+1} &= \lambda \\ q_{i,i-1} &= \mu \\ q_i &= -(\lambda + \mu) \\ q_{i,j} &= 0 \text{ for } |i - j| > 1. \end{aligned}$$

Stationary Distribution : Since we know how to calculate the distribution of a birth death process, we can find the stationary distribution of M/M/1 queueing model.

Let $\pi = (\pi_0, \pi_1, \dots)$ denote the stationary distribution vector. Then,

$$\pi_i \geq 0, \quad \sum_i \pi_i = 1.$$

Now by $\pi Q = 0$, we get

$$\begin{aligned} 0 &= -\lambda\pi_0 + \mu\pi_1 \\ 0 &= \lambda\pi_{i-1} - (\lambda + \mu)\pi_i + \mu\pi_{i+1}, \quad i = 1, 2, \dots \end{aligned}$$

Solving above set of equations, we get

$$\pi_1 = \frac{\lambda}{\mu}\pi_0, \quad \pi_{i-1} = \frac{\lambda}{\mu}\pi_i = \dots = \frac{\lambda^{i-1}}{\mu^{i-1}}\pi_0, \quad i = 1, 2, \dots$$

Taking $\rho = \frac{\lambda}{\mu}$, we get

$$\pi_0 = 1 - \rho, \quad \pi_i = (1 - \rho)\rho^i \quad ; i = 1, 2, \dots$$

where $\rho < 1$.

ρ : Offered load(traffic intensity)

$1 - \rho$: Server utilization

Average Number in the System

$$\begin{aligned}
E(N) &= \text{Average number of customers in the system in steady state.} \\
&= \sum_n n \rho^n (1 - \rho) \\
&= \rho(1 - \rho) \sum_{n=0}^{\infty} n \rho^{n-1} \\
&= \rho(1 - \rho) \frac{1}{1 - \rho} \\
&\Rightarrow E(N) = \frac{\rho}{1 - \rho}.
\end{aligned}$$

$$\text{Similarly, } Var(N) = \frac{\rho}{(1 - \rho)^2}.$$

Average Number in the Queue

$$\begin{aligned}
E(Q) &= \text{Average number of customers in the queue in steady state.} \\
&= \sum_{n=1}^{\infty} n \pi_{n+1} \\
&= \sum_{n=1}^{\infty} n \rho^{n+1} (1 - \rho) \\
&= \rho^2 (1 - \rho) \frac{1}{(1 - \rho)^2} \\
&\Rightarrow E(Q) = \frac{\rho^2}{1 - \rho}.
\end{aligned}$$

3 Little's law

The long term average number of customers in a stable system($E(N)$) is equal to the long term average effective arrival rate(λ), multiplied by the average time a customer spends in the system($E(R)$); expressed as

$$E(N) = \lambda E(R).$$

Measures for M/M/1 Queue using Little's Law Let R and W be time spend by a customer in the system and waiting time of a customer, respectively. By Little's law,

$$\begin{aligned} E(R) &= \frac{E(N)}{\lambda} \\ &= \frac{\rho}{\lambda(1-\rho)} \\ &= \frac{1}{\mu - \lambda} \end{aligned}$$

Distribution of Waiting Time

$$W = \begin{cases} 0 & \text{if } n = 0 \\ \tilde{S}_1 + S_2 + \dots + S_n & \text{if } n = 1, 2, 3, \dots \end{cases}$$

Then, for $t = 0$, $P(W \leq t) = 1 - \rho$.

Now, $W \mid \{N = n\} \sim \text{Gamma}(n, \mu)$

For $t > 0$

$$P(W \leq t) = \sum_{n=1}^{\infty} \int_0^t \frac{\mu^n x^{n-1} e^{-\mu x}}{(n-1)!} dx (1-\rho) \rho^n.$$

Hence we have,

$$P(W \leq t) = \begin{cases} 0 & , t < 0 \\ 1 - \rho & , t = 0 \\ \sum_{n=1}^{\infty} \int_0^t \frac{\mu^n x^{n-1} e^{-\mu x}}{(n-1)!} dx (1-\rho) \rho^n & , 0 < t < \infty. \end{cases}$$

Further, after simplification it can be seen that for $t > 0$,

$$P(W \leq t) = 1 - \rho e^{-\mu(1-\rho)t}.$$

Hence,

$$P(W \leq t) = \begin{cases} 0 & , t < 0 \\ 1 - \rho & , t = 0 \\ 1 - \rho e^{-\mu(1-\rho)t} & , 0 < t < \infty \end{cases}$$

$$\Rightarrow P[W = 0] = 1 - \rho$$

and

$$f_W(t) = \rho(\mu - \lambda)e^{-(\mu-\lambda)t}, \quad t > 0.$$

Distribution of Response Time

$$R = S + \tilde{S}_1 + S_2 + \dots + S_n$$

Then, for $t \leq 0$, $P(R \leq t) = 0$.

Now, $R \mid \{N = n\} \sim \text{Gamma}(n + 1, \mu)$.

For $t > 0$

$$\begin{aligned} P(R \leq t) &= \sum_{n=0}^{\infty} \int_0^t \frac{\mu^{n+1} x^n e^{-\mu x}}{n!} dx (1 - \rho) \rho^n \\ &= 1 - e^{-\mu(1-\rho)t}. \end{aligned}$$

Hence we have,

$$P(R \leq t) = \begin{cases} 0 & , t \leq 0 \\ 1 - e^{-\mu(1-\rho)t} & , t > 0. \end{cases}$$

Thus we can conclude that, $R \sim \text{Exp}(\mu(1 - \rho))$

$$\therefore E(R) = \frac{1}{\mu(1 - \rho)} = \frac{1}{\mu - \lambda}.$$

Remark : Burke's Theorem

The output of a Poisson input queue with a single channel having exponential service time and in steady state must be Poisson with the same rate as the input.

- Valid for $M/M/1$, $M/m/c$, $M/M/\infty$ queues.
- The number of customers in the queue is independent of the departure process prior to t .

Lecture 7 Simple Markovian Queueing Models

1 M/M/c Queueing model

- Arrival follows Poisson process with rate λ
- Service times follow exponential distribution with parameter μ
- c servers
- Arriving customer finds n customers in system
 - $n < c$; it is routed to any idle server.
 - $n \geq c$; it joins the waiting queue-all servers are busy
- Birth-death process with state-dependent death rates

$$\mu_n = \begin{cases} n\mu & , 1 \leq n \leq c \\ c\mu & , n \geq c \end{cases}$$

Steady-state or equilibrium solution when $\frac{\lambda}{c\mu} < 1$

$$p_n = \begin{cases} \frac{\lambda^n}{n!\mu^n} p_0 & , 1 \leq n \leq c \\ \frac{\lambda^n}{c^{n-c}c!\mu^n} p_0 & , n > c \end{cases}$$

Using normalizing constant

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

$$\Rightarrow p_0 = \left[\sum_{n=0}^{c-1} \frac{1}{n!} \left(\frac{\lambda}{\mu}\right)^n + \frac{1}{c!} \left(\frac{\lambda}{\mu}\right)^c \left(\frac{c\mu}{c\mu - \lambda}\right) \right]^{-1}.$$

The probability that an arriving customer has to wait on arrival

$$= \sum_{n=c}^{\infty} p_n = \frac{p_c}{1-\rho} \quad (\text{say, } C(c, \frac{\lambda}{\mu})).$$

This is known as Erlang's C formula.

$C(c, \frac{\lambda}{\mu})$: fraction of time all the c servers are busy.

Let N_q denote the number of customers in the queue

$$P[N_q = j, W > 0] = P[N = c + j] = \rho^j P_c = \rho^j (1 - \rho) C(c, \frac{\lambda}{\mu}).$$

Thus,

$$\begin{aligned} P[N_q = j \mid W > 0] &= \frac{P[N_q = j, W > 0]}{P[W > 0]} \\ &= (1 - \rho) \rho^j, j = 0, 1, \dots \end{aligned}$$

Expected number of busy servers

$$E(B) = \sum_{n=0}^{c-1} n p_n + \sum_{n=c}^{\infty} c p_n = c\rho.$$

Expected number of idle servers

$$E(I) = E(c - B) = c - c\rho = c(1 - \rho).$$

We have $E(N) = E(B) + E(Q)$.

$$\begin{aligned}
 E(Q) &= \sum_{n=c}^{\infty} (n-c)p_n \\
 &= \sum_{n=c}^{\infty} (n-c) \frac{\left(\frac{\lambda}{\mu}\right)^n}{c!(c^{n-c})} p_0 \\
 &= \frac{\rho}{1-\rho} C\left(c, \frac{\lambda}{\mu}\right)
 \end{aligned}$$

$$\Rightarrow E(N) = c\rho + \frac{\rho}{1-\rho} C\left(c, \frac{\lambda}{\mu}\right).$$

Using Little's law $E(N) = \lambda E(R)$, we get

$$E(R) = \frac{E(N)}{\lambda} = \frac{1}{\mu} + \frac{P_c}{c\mu(1-\rho)^2}.$$

Using $\lambda E(W) = E(Q)$, we get

$$E(W) = \frac{E(Q)}{\lambda} = \frac{1}{\mu} + \frac{P_c}{c\mu(1-\rho)^2}.$$

2 M/M/1/N Queuing model

M/M/1/N queueing model is an M/M/1 queueing model with capacity of the system N. Customers arriving after the queue is full are rejected.

Balance equations for M/M/1/N queueing model in steady state are given by :

$$\lambda\pi_{i-1} = \mu\pi_i \quad , i = 1, 2, \dots, N.$$

Using $\sum_{i=0}^N \pi_i = 1$, we get

$$\pi_0 = \frac{1 - \rho}{1 - \rho^{N+1}}$$

$$\pi_i = \frac{1 - \rho}{1 - \rho^{N+1}} \rho^i, \quad i = 0, 1, 2, \dots, N.$$

Effective Arrival Rate: $\lambda_{eff} = \lambda(1 - \pi_N)$

Throughput: $\mu(1 - \pi_0)$

Blocking Probability: π_N

With these, we have

$$E(R) = \frac{E(N)}{\lambda_{eff}}.$$

3 $M/M/c/K$ Queuing model

- Arrival follows Poisson process with rate λ
- Service times follow exponential distribution with parameter μ .
- c Servers with capacity K .
- Arriving customer find n customers already in system, where if
 - $n < c$: it is routed to an idle server
 - $n \geq c$: it joins the waiting queue
- Customers forced to leave the system if already K present in the system.
- Birth Death process with state dependent rates

$$\mu_n = \begin{cases} n\mu & \text{if } 1 \leq n < c \\ c\mu & \text{if } c \leq n < K \end{cases}$$

- Stead-state or equilibrium solution

$$\pi_n = \begin{cases} \frac{1}{n!} \left(\frac{\lambda}{\mu}\right)^n \pi_0 & \text{if } 0 \leq n < c \\ \frac{1}{c^{n-c} c!} \left(\frac{\lambda}{\mu}\right)^n \pi_0 & \text{if } c \leq n < K \end{cases}$$

We have the following.

1. $E(N) = \sum_{n=1}^k n\pi_n$
2. $E(Q) = \sum_{n=1}^k (n - c)\pi_n$
3. $E(R) = \frac{E(N)}{\lambda_{eff}}, \text{ where } \lambda_{eff} = \lambda(1 - \pi_k)$
4. $E(W) = \frac{E(Q)}{\lambda_{eff}}$

4 $M/M/c/c$ loss system

- c servers, no waiting room.
- An arriving customer that finds all servers busy is blocked.
- Stationary distribution

$$p_n = \frac{(\frac{\lambda}{\mu})^n}{n!} [\sum_{k=0}^c c \frac{(\frac{\lambda}{\mu})^k}{k!}]^{-1}, \quad n = 0, 1, \dots, c$$

5 $M/M/\infty$ self service system

- Number of servers are infinite.
- On arrival, all the customers are taken into service and there is no queue.
- Birth death process with state dependent death rates:

$$\mu_n = n\mu, \quad n = 1, 2, \dots$$

Bibliography

- [1] J. Medhi, Stochastic Process, 3rd edn., New Age International (P) Ltd., New Delhi, (2011)
- [2] K.S. Trivedi, Probability and Statistics with Reliability, Queuing and Computer Science Applications, 2nd edn., Wiley Interscience, New York, (2002)
- [3] L. B. Castaneda, V. Arunachalam, S. Dharmaraja, Introduction to Probability and Stochastic Processes with Applications, 1st edn., John Wiley and Sons, New Jersey, (2012)
- [4] M. Liao, Applied Stochastic Processes, 1st edn., CRC Press, Taylor and Francis group, Boca Raton, (2014)
- [5] P. G. Hoel, S. C. Port, C. J. Stone, Introduction to Probability Theory Hardcover, 1 edn., Cengage Learning, Boston,(1972)
- [6] S. Karlin, H.M. Taylor, A First Course in Stochastic Processes, 2nd edn., Academic press, San Diego, (1975)
- [7] S. M. Ross, Introduction to Probability Models, 8th edn., Academic press, San Diego, (2004)

Module 5: Martingales

Lecture 1 Introduction, Filtrations and Conditional Expectations

In this module, we will introduce and study another important class of stochastic processes, namely, the martingale processes. Martingale referred to a class of betting strategies popular in 18th century. The concept of martingale in probability theory was introduced by **Paul Levy** and much of the original development of the theory was done by **Joseph Doob**. Part of the motivation for that work was to show the impossibility of successful betting strategies.

Our treatment of these processes is restricted to certain basic concepts and related important results. Martingales arise naturally in many branches of the theory of stochastic processes. In particular, they are very helpful tools in the study of the Brownian motion. In this module, the index set T is an arbitrary interval of nonnegative integers or nonnegative real numbers.

As such, martingales models a fair game or something that we consider to be a fair. For example, martingales are ubiquitous in financial mathematics and are very much part of obtaining ‘fair’ prices of financial instruments. For example, security prices and ones wealth will turn out to be examples of martingales. As opposed to martingales, the related concepts of sub-martingales and super-martingales model unfavourable games.

Keywords: Filtrations, $\mathcal{F}$ -measurable, Adaptability, Conditional expectation, Rules on conditional expectation

1 Filtrations

We have a sample space Ω of some random experiment. The information available at time t is represented by the σ -field or σ -algebra of events $\mathcal{F}_t$. We assume that for each t , $\mathcal{F}_t \subseteq \mathcal{F}_{t+1}$; since we are supposed to gain information going from t to $t + 1$, every known event in $\mathcal{F}_t$ is also known at time $t + 1$. Such an expanding collection or family of σ -algebras $\mathcal{F}_t \subseteq \mathcal{F}_{t+1}$ is known as a *filtration*.

One way to construct them is as follows. Recall that a stochastic process is a

family of random variables, $X_t(\omega)$, with $X_t \in \mathcal{F}_t$ (X_t measurable with respect to $\mathcal{F}_t$). Sometimes it happens that the random variables X_t contain all the information in the $\mathcal{F}_t$ in the sense that $\mathcal{F}_t$ is generated by $X_1, \dots, X_t$. This is the *minimal σ -algebra* in which the X_t form a stochastic process. In other cases, $\mathcal{F}_t$ contains more information. Economists use these possibilities when they distinguish between the “weak efficient market hypothesis” (the $\mathcal{F}_t$ are minimal), and the “strong hypothesis” ($\mathcal{F}_t$ contains all the public information in the world, literally). In the case of minimal $\mathcal{F}_t$, it may be possible to identify the outcome, ω , with the path $X = X_1, \dots, X_T$. This is less common when the $\mathcal{F}_t$ are not minimal because the extra information may have to do with processes other than X_t . For the definition of stochastic process, the probabilities are not important, just the algebras of sets and “random variables” X_t .

An example of a filtration would be repeatedly tossing a coin and letting $\mathcal{F}_k$ be the sets that can be determined by the first k tosses. For example, when we toss a two times, one after the other, the sample space Ω has four elements. The filtration will consist of the σ -fields $\mathcal{F}_0 = \{\emptyset, \Omega\}$, $\mathcal{F}_1 = \{\emptyset, \Omega, \{HH, HT\}, \{TH, TT\}\}$ and $\mathcal{F}_2 = \{\emptyset, \Omega, \{HH\}, \{HT\}, \{TH\}, \{TT\}, \text{unions of these four atoms}\}$.

In discrete time, the filtration is an increasing sequence $\mathcal{F}_0 \subset \mathcal{F}_1 \subset \dots$ of σ -fields, one per time instant. The σ -field $\mathcal{F}_n$ may be thought of as the events of which the occurrence is determined at or before time n , the “known events” at time n . The natural filtration of a stochastic process $\{X_n, n = 0, 1, 2, \dots\}$ is defined by

$$\mathcal{F}_n = \{(X_0, X_1, \dots, X_n) \in B; B \subset \mathbb{R}^{n+1}\}.$$

It can also be written as

$$\mathcal{F}_n = \sigma(X_0, X_1, \dots, X_n).$$

EXAMPLE 1. *Toss a coin infinitely many times.*

$$\Omega = \{(HH\dots), (HT\dots), (TT\dots), (TH\dots), \dots\}$$

$$\mathcal{F}_0 = \{\emptyset, \Omega\}$$

$$\Omega_1 = \{A_H, A_T\}$$

$$\mathcal{F}_1 = \{\emptyset, A_H, A_T, \Omega_1\} - \text{contains the information learned by observing the first toss.}$$

$$\Omega_2 = \{A_{HH}, A_{TH}, A_{HT}, A_{TT}\}$$

$$\mathcal{F}_2 = \{\emptyset, A_{HH}, A_{TH}, A_{HT}, A_{TT}, \{A_{HH}, A_{TH}\}, \dots, \Omega_2\} - \text{contains the information}$$

learned by observing the first two consecutive tosses

$$\mathcal{F}_1 \subset \mathcal{F}_2 \subset \mathcal{F}_3 \subset \dots$$

The collection of σ -fields $\{\mathcal{F}_n, n = 0, 1, \dots\}$ is called a filtration.

DEFINITION 2. A random variable X is integrable if $E|X| < \infty$. Given a filtration of σ -fields $\{\mathcal{F}_t\}$, a family of random variables $\{X_t\}$ is adapted if X_t is $\mathcal{F}_t$ measurable for each t .

Recall that a function $f : \mathbb{R} \rightarrow \mathbb{R}$ is Borel measurable on $\mathbb{R}$ if $\{x \in \mathbb{R} : f(x) \in B\} \in \mathcal{B}(\mathbb{R})$ for every $B \in \mathcal{B}(\mathbb{R})$, where $\mathcal{B}(\mathbb{R})$ is the Borel σ -field over $\mathbb{R}$.

2 X_t is $\mathcal{F}_t$ -measurable

DEFINITION 3. Let $\mathcal{F}_t$ be a σ -field of subsets of Ω . A random variable X_t is $\mathcal{F}_t$ -measurable if every set in $\sigma(X_t)$ is also in $\mathcal{F}_t$.

Some Remarks on X_t is $\mathcal{F}_t$ -measurable

1. A random variable is $\mathcal{F}_t$ -measurable if and only if the information in $\mathcal{F}_t$ is sufficient to determine the value of X_t .
2. When a random variable X_t is not $\mathcal{F}_t$ -measurable, the information contained in σ -field $\mathcal{F}_t$ can not determine the values of the random variable X_t .
3. If X_t is $\mathcal{F}_t$ -measurable and g is Borel measurable function, then $g(X_t)$ is also $\mathcal{F}_t$ -measurable.
4. When $\mathcal{F}_t$ is σ -field of all subsets of Ω (the power set of Ω) and Ω is finite or countably infinite, random variable defined on Ω is always $\mathcal{F}_t$ -measurable.

3 Adaptability

DEFINITION 4. The stochastic process $\{X_t, t \geq 0\}$ is said to be adapted to the filtration $\{\mathcal{F}_t, t \geq 0\}$ if

$$\sigma(X_t) \subset \mathcal{F}_t, \text{ for all } t \geq 0$$

i.e., X_t is $\mathcal{F}_t$ -measurable for each $t \geq 0$. In such case, all events concerning the sample paths of an adapted process until time t are contained in $\mathcal{F}_t$.

In discrete case, we say that a stochastic process $\{X_n, n = 0, 1, \dots\}$ is adapted to the filtration $\{\mathcal{F}_n, n = 0, 1, \dots\}$ if $\sigma(X_n) \subset \mathcal{F}_n$ for every n .

For instance, suppose S_n is the price of a stock at the end of n th day then the price process $\{S_n, n = 0, 1, \dots\}$ is adapted to natural filtration $\{\mathcal{F}_n, n = 0, 1, \dots\}$ where $\mathcal{F}_n$ is the history up to the end of n th day.

4 Conditional Expectations

Conditional expectations are in some sense partial averages of random variables. They play a crucial role in the general treatment of martingales. Recall that

$$E(X) = \int_{\Omega} X(\omega) dP(\omega) = \begin{cases} \int_{-\infty}^{\infty} x f_X(x) dx \\ \sum_i x_i P[X = x_i] \end{cases}$$

provided $E(|X|) < \infty$.

DEFINITION 5. Let X and Y be two random variables defined on $(\Omega, \mathcal{F}, P)$. The conditional expectation of X given $Y = y$ can be expressed in the form

$$E(X | Y = y) = \begin{cases} \int_{-\infty}^{\infty} x f_{X/Y}(x/y) dx, & \text{provided } f_Y(y) > 0 \\ \sum_i x_i P[X = x_i / Y = y], & \text{provided } P[Y = y] > 0. \end{cases}$$

EXAMPLE 6. Let $\Omega = \{a, b, c, d\}$, $\mathcal{F} = \mathcal{P}(\Omega)$ and $P(\{w\}) = \frac{1}{4}$, $w \in \Omega$

Define

$$X(w) = \begin{cases} 0, & w = a, d \\ -1, & w = b \\ 1, & w = c \end{cases}$$

$$Y(w) = \begin{cases} -1, & w = a, b \\ 1, & w = c, d \end{cases}.$$

Now,

$$E(X | Y = -1) = \sum_i x_i P[X = x_i / Y = -1]$$

$$= (-1) \times P[X = -1/Y = -1] + (1) \times P[X = 1/Y = -1]$$

$$\begin{aligned} E(X \mid Y = -1) &= -\frac{P[X = -1, Y = -1]}{P[Y = -1]} + \frac{P[X = 1, Y = -1]}{P[Y = -1]} \\ &= -\frac{P(\{a\})}{P(\{a, b\})} + \frac{P(\{c\})}{P(\{a, b\})} \\ &= -\frac{1/4}{1/2} + \frac{0}{1/2} \\ &= -\frac{1}{2}. \end{aligned}$$

Similarly, we get

$$E(X \mid Y = 1) = \frac{1}{2}.$$

Thus,

$$E(X \mid Y = y)(w) = \begin{cases} -\frac{1}{2}, & w = a, b \\ \frac{1}{2}, & w = c, d \end{cases}.$$

From the above example, the observations are as follows:

1. $E(X \mid Y = y)$ is a function of y .
2. $E(X \mid Y = y)$ is a random variable.

Lecture 2 Conditional Expectations

In the last lecture, we discuss the filtrations and conditional expectation of X given Y . In this lecture, we present the rules related to it and the conditional expectation given σ -field.

Keywords: Conditional expectation given σ -field, Properties of conditional expectations

1 Rules on Conditional Expectations

Now, we list the following rules which are useful while computing the conditional expectations.

Rule 1 (Positivity)

$$E(X | Y) \geq 0 \text{ if } X \geq 0 \text{ with probability 1.}$$

Proof. Assuming X and Y are continuous random variables with joint probability density function f . Since $X \geq 0$, $P(X < 0) = 0$ and $f_X(x) = 0$ for all $x < 0$. Hence, $f_{X/y}(x/y) = 0$ for all $x < 0, y \in \mathbb{R}$.

Let $y \in \mathbb{R}$,

$$\begin{aligned} E(X | Y) &= \int_{-\infty}^{\infty} x f_{X/y}(x/y) dx \\ &= \int_0^{\infty} x f_{X/y}(x/y) dx \geq 0. \end{aligned}$$

□

Rule 2 (Linearity)

$$E(\alpha X + \beta Z | Y) = \alpha E(X | Y) + \beta E(Z | Y)$$

where α and β are constants.

Rule 3 (Expectation Law)

$$E(E(X | Y)) = E(X).$$

It says that the expectation of a random variable X can be computed in two steps, first using the information on another random variable Y , and next taking the expectation of the result.

Proof. Assuming X and Y are continuous random variables with joint pdf f ,

$$\begin{aligned} E(E(X | Y)) &= \int_{-\infty}^{\infty} E(X | y) f_Y(y) dy \\ &= \int_{-\infty}^{\infty} \left[\int_{-\infty}^{\infty} x f_{X/y}(x/y) dx \right] f_Y(y) dy \\ &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x f_{(X,Y)}(x, y) dx dy. \end{aligned}$$

Using $\int_{-\infty}^{\infty} f_{(X,Y)}(x, y) dy = f_X(x)$, we get

$$E(E(X | Y)) = \int_{-\infty}^{\infty} x f_X(x) dx.$$

□

Rule 4 (Independence Law) If X and Y are independent random variables, then

$$E(X | Y) = E(X).$$

Proof. Assuming X and Y are continuous random variables,

$$\begin{aligned} E(X | Y) &= \int_{-\infty}^{\infty} x f_{X/y}(x/y) dx \\ &= \int_{-\infty}^{\infty} x f_X(x) dx. \end{aligned}$$

□

Rule 5 (Stability) For any Borel measurable function g ,

$$E(Xg(Y) | Y) = g(Y)E(X | Y).$$

Proof. Assuming X and Y are continuous random variables,

$$\begin{aligned} E(Xg(Y) | Y) &= \int_{-\infty}^{\infty} xg(y)f_{X/y}(x/y)dx \\ &= g(y) \int_{-\infty}^{\infty} xf_{X/y}(x/y)dx \\ &= g(y)E(X | Y). \end{aligned}$$

□

2 Conditional Expectation Given σ -field

The general definition of conditional expectation is as follows.

DEFINITION 1. Let $(\Omega, \mathcal{F}, P)$ be a probability space, let $\mathcal{G}$ be a sub- σ -algebra of $\mathcal{F}$, and let X be a random variable that is integrable. The conditional expectation of X given $\mathcal{G}$, denoted $E(X|\mathcal{G})$, is any random variable that satisfies

- (i) (Measurability) $E(X|\mathcal{G})$ is $\mathcal{G}$ -measurable, and
- (ii) (Partial averaging)

$$\int_A E(X|\mathcal{G})(\omega)dP(\omega) = \int_A X(\omega)dP(\omega) \quad \text{for all } A \in \mathcal{G}.$$

The first questions that immediately arises are: Does there exist such a random variable? If so, is it unique? The existence of the conditional expectations can be shown by using the Radon-Nikodym theorem. The answer to the second question is a qualified yes in the sense that it is unique with probability one. This means that any two conditional expectations differ only on a set of probability zero.

REMARK 2. It may be mentioned here that whenever we write that two random variables are equal, it is to be understood in the almost surely sense.

1. Note that the conditional expectation $E(X | \mathcal{F})$ is a random variable, while the expectation $E(X)$ is a real number.

2. When $\mathcal{F}$ is the σ -field generated by Y , then we also write $E(X \mid \mathcal{F})$ for the random variable $E(X \mid Y)$. Thus $E(X \mid \mathcal{F})$ is the expected value of X given the information $\mathcal{F}$.

EXAMPLE 3. Let $\Omega = \{a, b, c, d\}$, $\mathcal{S} = \mathcal{P}(\Omega)$ and $P(\{w\}) = \frac{1}{4}$, $w \in \Omega$. Define

$$X(w) = \begin{cases} 0, & w = a, d \\ -1, & w = b \\ 1, & w = c \end{cases}$$

Let $\mathcal{F} = \{\emptyset, \Omega\}$. Given $\mathcal{F}$ is the trivial σ -field. The only random variables which are measurable with respect to the trivial σ -field are constants. Hence,

$$E(X \mid \mathcal{F}) = E(X) = k$$

where k is a constant.

EXAMPLE 4. Let $\Omega = \{a, b, c\}$, $\mathcal{S} = \mathcal{P}(\Omega)$ and $P(\{w\}) = \frac{1}{3}$, $w \in \Omega$. Define

$$X(w) = \begin{cases} 0, & w = a, b \\ 2, & w = c \end{cases}.$$

Let $\mathcal{F} = \{\emptyset, \{a\}, \{b, c\}, \Omega\}$. Define

$$Y(w) = \begin{cases} 0, & w = a \\ 1, & w = b, c \end{cases}.$$

Then, we claim that $E(X \mid \mathcal{F}) = E(X \mid Y)$. In fact, Y is $\mathcal{F}$ -measurable, since

$$Y^{-1}((-\infty, x]) = \begin{cases} \emptyset & -\infty < x < 0 \\ a & 0 \leq x < 1 \\ \Omega & 1 \leq x < \infty \end{cases}.$$

Lecture 3 Generated σ -fields

In this lecture, we discuss about σ -fields generated by random vectors.

Keywords: σ -field generated by a random vector $\mathbf{Y}$

1 σ -field generated by a random vector $\mathbf{Y}$

σ -field generated by a collection of subsets of Ω

DEFINITION 1. Let U be a collection of subsets of Ω . Then the smallest σ -field containing U is called the σ -field generated by the collection U of subsets of Ω . This is denoted by $\sigma(U)$.

Consider an example. Let $\Omega = \{a, b, c\}$. Let $U = \{a\}$. Then $\sigma(U) = \{\emptyset, \{a\}, \{b, c\}, \Omega\}$ is a σ -field generated by the collection of sets $\{a\}$, or, generated by the set $\{a\}$.

σ -field generated by a random variable X_t

DEFINITION 2. The σ -field generated by a random variable X_t is defined as the σ -field generated by the family of subsets of the form $\{X_t \in B\}$, where B ranges over the Borel sets of $\mathbb{R}$.

Consider an example. Let $\Omega = \{a, b, c, d\}$. Define-

$$X(w) = \begin{cases} 0.5, & w = a, b \\ 1.5, & w = c, d \end{cases}.$$

Let $A_1 = \{w \in \Omega : X(w) = 0.5\} = \{a, b\}$ and $A_2 = \{w \in \Omega : X(w) = 1.5\} = \{c, d\}$ be the subsets of Ω . Let $U = \{A_1, A_2\}$. Hence,

$$\sigma(X) = \sigma(U) = \{\emptyset, \{a, b\}, \{c, d\}, \Omega\}.$$

σ -field generated by a random vector $\mathbf{Y}$

DEFINITION 3. Let $\mathbf{Y} = (Y_1, Y_2, \dots, Y_n)$ be an n -dimensional random vector, i.e., $Y_1, Y_2, \dots, Y_n$ are random variables. The σ -field $\sigma(\mathbf{Y})$ is the smallest σ -field containing all sets of the form $\{w \mid \mathbf{Y}(w) \in C, w \in \Omega\}$ for all Borel sets $C \subset \mathbb{R}^n$. We call $\sigma(\mathbf{Y})$ the σ -field generated by the random vector $\mathbf{Y}$.

Whenever the conditioning σ -field $\mathcal{G}$ is countably generated, the computations of the conditional expectations and proving the properties are relatively much easier than in the most general case which require ideas from measure theory. It can be shown that when $\mathcal{G}$ is countably generated by the sets $B_1, B_2, \dots$ the definition reduces to

$$E(X|\mathcal{G}) = \sum_i \frac{E(XI_{B_i})}{P(B_i)} I_{B_i}.$$

A special case of this is $E(I_A|\mathcal{G}) = P(A|\mathcal{G})$. Many of the properties that we state below can be easily understood or proved using this special case of the definition, but they hold in general too.

If Y is a discrete random variable, that is, it takes only countably many values $y_1, y_2, \dots$, we let $B_i = \{Y = y_i\}$. These will be disjoint sets whose union is Ω . If $\sigma(Y)$ is the collection of all unions of the B_i , then $\sigma(Y)$ is a σ -field, and is called the σ -field generated by Y . It is easy to see that this is the smallest σ -field with respect to which Y is measurable. We write $E[X|Y]$ for $E[X|\sigma(Y)]$.

EXAMPLE 4. Let $\Omega = \{a, b, c, d\}$, $\mathcal{F} = 2^\Omega$. $P\{\omega\} = \frac{1}{12} \left(2I_{\{a\}} + 4I_{\{b\}} + 3I_{\{c,d\}} \right)$, for $\omega \in \Omega$, and the probability of every other set in $\mathcal{F}$ is the sum of the probabilities of the elements in the set. Define $X(\omega) = I_{\{a,b\}} - I_{\{c,d\}}$ and $Y(\omega) = I_{\{a,c\}} - I_{\{b,d\}}$. We will determine $E(Y|X)$ and verify that the measurability property and the partial averaging property are satisfied.

First, we compute $\mathcal{G} = \sigma(X)$ and is given by

$$\sigma(X) = \{\emptyset, \Omega, \{a, b\}, \{c, d\}\}.$$

Note here that $\sigma(X)$ is generated by 2 sets. Now,

$$\begin{aligned} E(Y|X) &= \frac{E(YI_{\{a,b\}})}{P(\{a,b\})} I_{\{a,b\}} + \frac{E(YI_{\{c,d\}})}{P(\{c,d\})} I_{\{c,d\}} \\ &= \frac{\sum_{\omega \in \{a,b\}} Y(\omega) P\{\omega\}}{P(\{a,b\})} I_{\{a,b\}} + \frac{\sum_{\omega \in \{c,d\}} Y(\omega) P\{\omega\}}{P(\{c,d\})} I_{\{c,d\}} \\ &= -\frac{4}{12} I_{\{a,b\}}. \end{aligned}$$

It is immediate from the above that this is measurable with respect to $\sigma(X)$.

To verify the partial averaging property for every $A \in \sigma(X)$, it is enough to

verify it on the atoms of $\sigma(X)$, i.e., on $\{a, b\}$ and on $\{c, d\}$. We see that

$$\int_{\{a,b\}} E(Y|X)(\omega) dP(\omega) = \sum_{\omega \in \{a,b\}} E(Y|X)(\omega) P(\omega) = -\frac{4}{12} \frac{2}{12} + \left(-\frac{4}{12}\right) \frac{4}{12} = -\frac{2}{12}$$

which is equal to

$$\int_{\{a,b\}} Y(\omega) dP(\omega) = \sum_{\omega \in \{a,b\}} Y(\omega) P(\omega) = 1 \cdot \frac{2}{12} + (-1) \cdot \frac{4}{12} = -\frac{2}{12}.$$

Similarly, we can see that the other required equalities hold:

$$\int_{\{c,d\}} E(Y|X)(\omega) dP(\omega) = 0 = \int_{\{c,d\}} Y(\omega) dP(\omega).$$

An important result that helps in the computation of conditional expectations is what is known as independence lemma, which we state below without proof.

THEOREM 5. (*Independence lemma*) Let $(\Omega, \mathcal{F}, P)$ be a probability space, and let $\mathcal{G}$ be a sub- σ -field of $\mathcal{F}$. Suppose that the random variables $X_1, \dots, X_K$ are $\mathcal{G}$ -measurable and the random variables $Y_1, \dots, Y_L$ are independent of $\mathcal{G}$.

Let $f(x_1, \dots, x_K, y_1, \dots, y_L)$ be a function of dummy variables $x_1, \dots, x_K$ and $y_1, \dots, y_L$, and define

$$g(x_1, \dots, x_K) = Ef(x_1, \dots, x_K, Y_1, \dots, Y_L).$$

Then

$$E[f(X_1, \dots, X_K, Y_1, \dots, Y_L) | \mathcal{G}] = g(X_1, \dots, X_K).$$

2 Properties of Conditional Expectations given σ -field

We now turn to some properties of conditional expectation. Some of the following propositions may seem a bit technical. In fact, they are! However, these properties are crucial to what follows and there is no choice but to master them.

1. If $X \geq Y$, then $E[X|\mathcal{G}] \geq E[Y|\mathcal{G}]$.
2. (Linearity) $E[aX + bY|\mathcal{G}] = aE[X|\mathcal{G}] + bE[Y|\mathcal{G}]$, for real a, b .

3. (Measurability) If X is $\mathcal{G}$ -measurable, then $E[X|\mathcal{G}] = X$.
4. $E[E[X|\mathcal{G}]] = EX$.
5. (Independence) If X is independent of $\mathcal{G}$, then $E[X|\mathcal{G}] = EX$.
6. (Taking out what is known) If Z is $\mathcal{G}$ -measurable, then $E[XZ|\mathcal{G}] = ZE[X|\mathcal{G}]$.
7. (Tower or Iterated Conditioning) If $\mathcal{H} \subseteq \mathcal{G} \subseteq \mathcal{F}$, then $E[E[X|\mathcal{H}]|\mathcal{G}] = E[X|\mathcal{G}] = E[E[X|\mathcal{G}]|\mathcal{H}]$.
8. (Jensen's inequality) If g is convex, then $g(E[X|\mathcal{G}]) \leq E[g(X)|\mathcal{G}]$.

Let us assume, for the moment, that the above properties are true. Then, the conditional expectation may be viewed as the best predictor of a random variable given a σ -field. If X is a r.v., a predictor Z is just another random variable, and the goodness of the prediction will be measured by $E[(X - Z)^2]$, which is known as the mean square error.

THEOREM 6. *If X is a random variable, the best predictor among the collection of $\mathcal{G}$ -measurable random variables is $Y = E[X|\mathcal{G}]$.*

Proof. Let Z be any $\mathcal{G}$ -measurable random variable. We compute, using proposition listed above,

$$\begin{aligned}
 E[(X - Z)^2|\mathcal{G}] &= E[X^2|\mathcal{G}] - 2E[XZ|\mathcal{G}] + E[Z^2|\mathcal{G}] \\
 &= E[X^2|\mathcal{G}] - 2ZE[X|\mathcal{G}] + Z^2 \\
 &= E[X^2|\mathcal{G}] - 2ZY + Z^2 \\
 &= E[X^2|\mathcal{G}] - Y^2 + (Y - Z)^2 \\
 &= E[X^2|\mathcal{G}] - 2YE[X|\mathcal{G}] + Y^2 + (Y - Z)^2 \\
 &= E[X^2|\mathcal{G}] - 2E[XY|\mathcal{G}] + E[Y^2|\mathcal{G}] + (Y - Z)^2 \\
 &= E[(X - Y)^2|\mathcal{G}] + (Y - Z)^2.
 \end{aligned}$$

We also used the fact that Y is $\mathcal{G}$ -measurable. Taking expectations and using again the properties $E[(X - Z)^2] = E[(X - Y)^2] + E[(Y - Z)^2]$. The right hand side is bigger than or equal to $E[(X - Y)^2]$ because $(Y - Z)^2 \geq 0$. So the error in predicting X by Z is larger than the error in predicting X by Y , and will be equal if and only if $Z = Y$. So Y is the best predictor. $\square$

Based on the argument that the conditional expectation is the best predictor for a r.v. X , we can now give interpretations of the properties listed above. If X is larger than Y , then the predicted value of X should be larger than the predicted value of Y . The linearity property says that the predicted value of $X + X$ should be the sum of the predicted values. Measurability says that if we know $\mathcal{G}$ and X is $\mathcal{G}$ -measurable, then we know X and our best prediction of X is X itself. Also, the average of the predicted value of X should be the average value of X . Independence property says that if knowing $\mathcal{G}$ gives us no additional information on X , then the best prediction for the value of X is just EX . Taking out what is known property says that as far as conditional expectations with respect to a σ -field $\mathcal{G}$ go, $\mathcal{G}$ -measurable random variables act like constants: they can be taken inside or outside the conditional expectation at will. The tower property implies that, if we are predicting a prediction of X given limited information, this is the same as a single prediction given the least amount of information.

Lecture 4 Martingales, Sub-martingales and Super-martingales

With the notions of filtrations and conditional expectations in our hand, we turn our attention to martingales. Martingales and stopping times are important technical tools used in the study of stochastic processes such as Markov chains and diffusions.

Keywords: Martingales, Doob's Martingale Process, sub-martingales, super-martingales

1 Martingales

DEFINITION 1. A sequence $\{X_n, n = 1, 2, \dots\}$ of random variables has martingale property with respect to a filtration $\{\mathcal{F}_n, n = 1, 2, \dots\}$ if

1. $E(X_n), n = 1, 2, \dots$ exist
2. X_n is $\mathcal{F}_n$ -measurable, $n = 1, 2, \dots$
3. $E(X_{n+1} | \mathcal{F}_n) = X_n, n = 1, 2, \dots$

and we say that $\{X_n, n = 1, 2, \dots\}$ is a martingale.

Some Remarks

1. The martingale property is equivalent to

$$E(X_n - X_m | \mathcal{F}_m) = 0 \text{ for every } m < n$$

expressing that the increment $X_n - X_m$ given the “past” $\mathcal{F}_m$ has expected value 0. A martingale is a stochastic process that, on the average, given the past, does not grow or decrease.

2. A martingale has a constant mean. i.e., $E(X_n) = E(X_0)$ for all n .
3. A martingale is thus a stochastic process being on average stationary.
4. While martingale concept involves expectations, the Markov process concept involves distributions.

5. A Markov process need not necessarily be a martingale.
6. It has applications in branching process and finance.

EXAMPLE 2. Let $X_1, X_2, \dots$ be a sequence of independent random variables with $E(X_n) = 0$ for all n . Define $S_0 = 0$ and $S_n = X_1 + X_2 + \dots + X_n$ for $n = 1, 2, \dots$. Let $\mathcal{F}_n = \sigma(X_1, X_2, \dots, X_n)$ be the σ -algebra generated by the first n X_i 's.

$$E(S_{n+1} \mid \mathcal{F}_n) = E(S_n + X_{n+1} \mid \mathcal{F}_n)$$

Since S_n is $\mathcal{F}_n$ -measurable and X_{n+1} is independent of $\mathcal{F}_n$,

$$E(S_{n+1} \mid \mathcal{F}_n) = S_n + E(X_{n+1}) = S_n$$

Then $\{S_n, n = 0, 1, \dots\}$ has martingale property with respect to the filtration $\{\mathcal{F}_n, n = 0, 1, \dots\}$.

EXAMPLE 3. Consider the binomial tree model. Let $\{S_n, n = 0, 1, \dots\}$ be a stochastic process and $\{\mathcal{F}_n, n = 0, 1, \dots\}$ be the natural filtration.

$$P(S_{n+1} = uS_n \mid \mathcal{F}_n) = 1 - P(S_{n+1} = dS_n \mid \mathcal{F}_n) = p.$$

Consider the discounted process $\{S_0, e^{-r}S_1, e^{-2r}S_2, \dots\}$. We have

$$E(e^{-(n+1)r}S_{n+1} \mid \mathcal{F}_n) = ue^{-(n+1)r}S_np + de^{-(n+1)r}S_n(1-p)$$

where r is the riskless interest rate.

The discounted process is a martingale only if the right hand side is equal to $e^{-nr}S_n$. This is the case only if $p = \frac{e^r - d}{u - d}$. Since $0 < p < 1$, if $d \leq e^r \leq u$, then the discounted process is a martingale.

EXAMPLE 4. Let $\{\mathcal{F}_n, n = 1, 2, \dots\}$ be a filtration. Let X be any random variable with $E(|X|) < \infty$. Define $X_n = E(X \mid \mathcal{F}_n), n = 1, 2, \dots$

$$\begin{aligned} E(|X_n|) &= E(|E(X \mid \mathcal{F}_n)|) \\ &\leq E(E(|X| \mid \mathcal{F}_n)) \\ &= E(|X|) < \infty. \end{aligned}$$

By the definition of conditional expectation, X_n is $\mathcal{F}_n$ -measurable, for all n . $E(X_{n+1} | \mathcal{F}_n) = E(E[X_{n+1} | \mathcal{F}_{n+1}] | \mathcal{F}_n) = E(X_{n+1} | \mathcal{F}_n) = X_n$

Hence, $\{X_n, n = 1, 2, \dots\}$ is a martingale with respect to the filtration $\{\mathcal{F}_n, n = 1, 2, \dots\}$.

EXAMPLE 5. A player plays against an infinitely rich adversary. He stands to gain Re. 1 with probability p and lose Re. 1 (or gain Re. -1) with probability $q = 1 - p$. X_n - the player's cumulative gain in the first n games. What will be his fortune, on the average, on the next game given that his current fortune?

The game is fair if and only if $\{X_n, n = 1, 2, \dots\}$ is a martingale when $p = q = 1/2$. It says that player's expected fortune after one more game played with the knowledge of the entire past and present is exactly equal to his current fortune.

2 Doob's Martingale Process

Let $\{Y_n, n = 0, 1, \dots\}$ be an arbitrary sequence of random variables and suppose X is a random variable with $E(|X|) < \infty$. Let $\mathcal{F}_n = \sigma(Y_0, Y_1, \dots, Y_n)$. Define

$$X_n = E(X | \mathcal{F}_n), n = 0, 1, \dots$$

We have

$$E(|X_n|) = E(|E(X | \mathcal{F}_n)|) \leq E(E(|X| | \mathcal{F}_n)) = E(|X|) < \infty.$$

Hence

$$E(X_{n+1} | \mathcal{F}_n) = E(E(X | \mathcal{F}_{n+1}) | \mathcal{F}_n) = E(X | \mathcal{F}_n) = X_n.$$

In essence, as n increases X_n approximates X and the approximation becomes more refined because more information has been gathered and included in the conditioning.

Thus $\{X_n, n = 0, 1, \dots\}$ forms a martingale with respect to $\{Y_n, n = 0, 1, \dots\}$, called Doob's process.

3 Sub-martingales and Super-martingales

Some Results

1. A stochastic process $\{X_n, n = 1, 2, \dots\}$ is a sub-martingale with respect to the filtration $\{\mathcal{F}_n, n = 1, 2, \dots\}$ if and only if $\{-X_n, n = 1, 2, \dots\}$ is a super-martingale with respect to the filtration $\{\mathcal{F}_n, n = 1, 2, \dots\}$.
2. A stochastic process $\{X_n, n = 1, 2, \dots\}$ is a martingale with respect to the filtration $\{\mathcal{F}_n, n = 1, 2, \dots\}$ if and only if $\{X_n, n = 1, 2, \dots\}$ is both a sub-martingale and a super-martingale with respect to the filtration $\{\mathcal{F}_n, n = 1, 2, \dots\}$.
3. If $\{X_n, n = 1, 2, \dots\}$ is a martingale, then $E(X_n) = E(X_0)$ for all n . If $m < n$ and if $\{X_n, n = 1, 2, \dots\}$ is a sub-martingale, then $E(X_m) \leq E(X_n)$; if $\{X_n, n = 1, 2, \dots\}$ is a super-martingale, then $E(X_m) \geq E(X_n)$.

EXAMPLE 6. Let $X_1, X_2, \dots$ be a sequence of independent r.v.s with mean 0. (Saying a r.v. X_i has mean 0 is the same as saying $EX_i = 0$; this presupposes that $E|X_1|$ is finite.) Set $\mathcal{F}_n = \sigma(X_1, \dots, X_n)$, the σ -field generated by $X_1, \dots, X_n$. Let $M_n = \sum_{i=1}^n X_i$. Adaptability is easy to see. Since $E|M_n| \leq \sum_{i=1}^n E|X_i|$, the integrability also holds. We now check

$$E[M_{n+1}|\mathcal{F}_n] = X_1 + \dots + X_n + E[X_{n+1}|\mathcal{F}_n] = M_n + EX_{n+1} = M_n,$$

where we used the independence. Hence, $\{M_n\}$ is a martingale.

Our next example will be used many times, so we state it as a result and this called as Doob-Levy martingale.

THEOREM 7. (Doob-Levy Martingale) Let $\{\mathcal{F}_t\}$ be a filtration and let X be a random variable with $E|X| < \infty$, and let $M_t = E(X|\mathcal{F}_t)$. Then $\{M_t\}$ is a martingale.

Proof. Since $E|M_t| \leq E[E(|X| |\mathcal{F}_t)] = E|X| < \infty$, the integrability is clear, and the adaptability is clear from the definition of M_t itself. We now have, for $s \leq t$,

$$E[M_t|\mathcal{F}_s] = E[E(X|\mathcal{F}_t)|\mathcal{F}_s] = E[X|\mathcal{F}_s] = M_s.$$

Therefore, $\{M_t\}$ is a martingale. □

Lecture 5 Stopping Times and Inequalities

Stopping times is an idea that works well together with martingales. A stopping time is a ‘random’ time so that we know at any time t whether to stop or not.

Keywords: Stopping times, Jensen’s Inequalities, Doob decomposition, Markov Time, Optional Sampling Theorem, Martingale Convergence Theorem, Martingale in Continuous Time

1 Stopping Times

DEFINITION 1. A mapping τ from Ω into the nonnegative real number is a stopping time if $\{\tau \leq t\} \in \mathcal{F}_t$ for each t . One sometimes allows τ to also take on the value ∞ .

We can think of a stopping time as the first time something happens. Here is an intuitive description of a stopping time. If I tell you to buy a stock if its price today is equal to the maximum over the last one month, you clearly know whether to buy the stock today or not, as you have with the history of the stock for the last six months. On the other hand, if I tell you to buy a stock if its price today is equal to the maximum over the current month, then either you must be at the end of the month so that you can decide or you must have some ‘insider’ information to decide the behaviour of the stock price for the rest of the month. The first set of instructions forms a stopping time, while the second set does not.

EXAMPLE 2. Consider the case when two coins are tossed, one after the other and assume the usual filtration of the coin tosses. Now, let $\tau(\omega) = 1$, if $\omega \in \{HH, HT\}$, and $\tau(\omega) = 2$, if $\omega \in \{TH, TT\}$. Since τ is discrete, to show that τ is a stopping time, we need to show that $\{\tau = k\} \in \mathcal{F}_k$ for $k = 0, 1, 2$. Now, $\{\tau = 0\} = \emptyset \in \mathcal{F}_0$, $\{\tau = 1\} = \{HH, HT\} \in \mathcal{F}_1$, $\{\tau = 2\} = \{TH, TT\} \in \mathcal{F}_2$. This proves that τ is a stopping time (with respect to this filtration). That is, the information available at time n is sufficient to determine the events connected with τ .

If we consider another r.v., say, $\rho(\omega) = 1$, if $\omega \in \{HH, TT\}$, and $\rho(\omega) = 2$, if $\omega \in \{HT, TH\}$, this is not a stopping time since $\{\rho = 1\} = \{HH, TT\} \ni \mathcal{F}_1$. As we observe, the information available at time 1 is not sufficient to determine the

event $\{\rho = 1\}$ since it requires the information of the second toss also.

From now on, we concentrate our attention more on discrete-time martingales (and filtrations) and consider the general martingales at the end. As a consequence, it only makes sense to consider stopping times that takes values in the set of non-negative integers and ∞ .

As opposed to the concept of adapted, we also have the following idea.

DEFINITION 3. We call a discrete-time process $\{M_n\}$ predictable or pre-visible with respect to the filtration $\{\mathcal{F}_n\}$ if M_n is $\mathcal{F}_{n-1}$ -measurable for every n .

We now introduce the concept of a discrete integral, a useful tool in martingale theory as the subsequent theorem explains.

DEFINITION 4. Let $M = \{M_n\}$ and $X = \{X_n\}$ be two discrete-time processes. We define the process $\{X \cdot M\}$ by $(X \cdot M)_0 = 0$ and

$$(X \cdot M)_n = \sum_{i=1}^n X_i(M_i - M_{i-1}).$$

We call $X \cdot M$ the discrete integral of X with respect to M . If M is a (sub-, super-)martingale, it is often called the martingale transform of M by X .

LEMMA 5. If M is a sub-martingale (resp. super-martingale) and X is a bounded, nonnegative predictable process, then $X \cdot M$ is a sub-martingale (resp. super-martingale) as well. If M is a martingale and X is a bounded, predictable process, then $X \cdot M$ is a martingale.

Proof. Put $Y = X \cdot M$. Clearly, the process Y is adapted. Since X is bounded, say $|X_n| \leq K$ for all n , we have $E|Y_n| \leq 2K \sum_{k \leq n} E|M_k| < \infty$. Now suppose first that M is a sub-martingale and X is nonnegative. Then a.s.

$$E(Y_n | \mathcal{F}_{n-1}) = E(Y_{n-1} + X_n(M_n - M_{n-1}) | \mathcal{F}_{n-1}) = Y_{n-1} + X_n E(M_n - M_{n-1} | \mathcal{F}_{n-1}) \geq Y_{n-1},$$

hence Y is a sub-martingale. If M is a martingale the last inequality is an equality, irrespective of the sign of X_n , which implies that Y is a martingale in that case. $\square$

Using this lemma it is easy to see that a stopped (sub-, super-)martingale is again a (sub-, super-)martingale.

THEOREM 6. *Let M be a (sub-, super-) martingale and τ a stopping time. Then the stopped process $M^\tau = \{M_{n \wedge \tau}\}$ is a (sub-, super-)martingale as well.*

Proof. Define the process X by $X_n = I_{\{\tau \geq n\}}$ and verify that $M^\tau = M_0 + X \cdot M$. Since τ is a stopping time, we have $\{\tau \geq n\} = \{\tau \leq n-1\}^c \in \mathcal{F}_{n-1}$, which shows that the process X is predictable. It is also a bounded process, so the statement follows from the preceding lemma. $\square$

2 Jenson's Inequalities

THEOREM 7. *Let X and Y be random variables on $(\Omega, \mathcal{S}, P)$ with $E(|X|) < \infty$. Let φ be a convex function on $\mathbb{R}$. Then*

$$E[\varphi(X)] \geq \varphi(E[X])$$

$$E[\varphi(X) \mid Y] \geq \varphi(E[X] \mid Y)$$

provided all the expectations exist.

THEOREM 8. *Let X and Y be random variables on $(\Omega, \mathcal{S}, P)$ and let φ be a convex function on $\mathbb{R}$. Suppose both X and φ are integrable and $\mathcal{F}$ is a sub σ -field of $\mathcal{S}$. Then*

$$E[\varphi(X) \mid \mathcal{F}] \geq \varphi(E[X] \mid \mathcal{F})$$

provided all the expectations exist.

THEOREM 9. *Suppose $\{X_n, n = 1, 2, \dots\}$ is a martingale with respect to the filtration $\{\mathcal{F}_n, n = 1, 2, \dots\}$ and φ is a convex function on $\mathbb{R}$. Then, if $\varphi(X_n)$ is integrable for all n , $\{\varphi(X_n), n = 1, 2, \dots\}$ is a sub-martingale with respect to the filtration $\{\mathcal{F}_n, n = 1, 2, \dots\}$.*

Proof. By Jenson's inequality for conditional expectation,

$$\varphi(X_m) = \varphi(E(X_n \mid \mathcal{F}_m)) \leq E(\varphi(X_n) \mid \mathcal{F}_m).$$

$\square$

THEOREM 10. *Suppose $\{X_n, n = 1, 2, \dots\}$ is a submartingale with respect to the filtration $\{\mathcal{F}_n, n = 1, 2, \dots\}$ and ϕ is an increasing convex function on $\mathbb{R}$. Then, if*

$\phi(X_n)$ is integrable for all n , $\{\phi(X_n), n = 1, 2, \dots\}$ is a submartingale with respect to the filtration $\{\mathcal{F}_n, n = 1, 2, \dots\}$.

Proof.

$$X_m \leq E(X_n \mid \mathcal{F}_m)$$

and since ϕ is increasing function,

$$\phi(X_m) \leq \phi(E(X_n \mid \mathcal{F}_m)) \leq E(\phi(X_n) \mid \mathcal{F}_m).$$

□

Now, we present Doob decomposition as follows.

THEOREM 11. *Let $\{X_n, n = 0, 1, \dots\}$ be a submartingale with respect to the filtration $\{\mathcal{F}_n, n = 0, 1, \dots\}$. Then there exists a martingale $M = \{M_n, n = 0, 1, \dots\}$ and a stochastic process $A = \{A_n, n = 0, 1, \dots\}$ such that:*

1. $X_n = M_n + A_n, n = 0, 1, \dots$
2. A_n is an increasing stochastic process, i.e., $A_n \leq A_{n+1}, n = 0, 1, \dots$ a.e.
3. A_n is $\mathcal{F}_{n-1}$ -measurable, for all n
4. $\{M_n, n = 0, 1, \dots\}$ is a martingale with respect to the filtration $\{\mathcal{F}_n, n = 0, 1, \dots\}$

Lecture 6 Convergence Theorems

In this lecture, we present the convergence theorems related to martingales.

Keywords: Martingales, Markov times, convergence theorems

1 Markov Time

DEFINITION 1. A random variable T is called a Markov time with respect to $\{Y_n, n = 0, 1, \dots\}$ if T takes values in $\{0, 1, \dots, \infty\}$ and if, for every $n = \{0, 1, \dots\}$, the event $\{T = n\}$ is determined by $(Y_0, Y_1, \dots, Y_n)$.

We often omit mention of $\{Y_n\}$ and say only “ T is a Markov time”.

EXAMPLE 2. The first time the process $\{Y_0, Y_1, \dots\}$ reaches a subset A of the state space is a Markov time. i.e., $T(A) = \min\{n : Y_n \in A\}$.

Properties of Markov Time

1. If S and T are Markov times, then $S + T$ is also a Markov time. $I_{(S+T=n)} = \sum_{k=0}^n I_{(S=k)} I_{(T=n-k)}$.
2. The smaller of two Markov times, S, T , denoted $S \wedge T = \min\{S, T\}$, is also a Markov time.
Thus, if T is a Markov time, then so is $T \wedge n = \min\{n, T\}$, for any fixed $n = \{0, 1, \dots\}$.
3. If S and T are Markov times, then so is the larger $S \vee T = \max\{S, T\}$.

2 Optional Sampling Theorem

THEOREM 3. Let $\{X_n, n = 1, 2, \dots\}$ be a martingale and T a Markov time. If

1. $\text{Prob } \{T < \infty\} = 1$,
2. $E(|X_T|) < \infty$,
3. $\lim_{n \rightarrow \infty} E(X_n I_{(T > n)}) = 0$,

then $E(X_T) = E(X_0)$.

Here,

$$E(X_0) = \lim_{n \rightarrow \infty} E(X_{T \wedge n}) = E(\lim_{n \rightarrow \infty} X_{T \wedge n}) = E(X_T).$$

Also, referred as Optional Stopping Theorem.

As application, we discuss the following example. Consider the random walk $\{S_n, n = 0, 1, \dots\}$ where $S_0 = 0$ and $S_n = Y_1 + Y_2 + \dots + Y_n, n = 1, 2, \dots$. Here, $Y_0 = 0$ and for $i = 1, 2, \dots$, Y_i are independent identically distributed random variables with

$$\text{Prob}(Y_i = 1) = 1 - \text{Prob}(Y_i = -1) = p.$$

When $p = q (= 1 - p) = 1/2$, $\{S_n, n = 0, 1, \dots\}$ is a martingale. If $T = \min\{n : S_n = 1\}$, then $\text{Prob}\{T < \infty\} = 1$. Since $E(T) = \infty$, optional sampling theorem does not hold, hence $E(S_T) \neq E(S_0) = 0$.

When $p > q$, $\{S_n - n(p - q), n = 0, 1, \dots\}$ is a martingale. If

$$T = \min\{n : S_n = -a \text{ or } S_n = b\} \quad (a, b \text{ positive integers})$$

then $\text{Prob}\{T < \infty\} = 1$ and $E(T) = ab$.

Now, optional sampling theorem holds, hence $E(S_T) = (p - q)E(T)$ for any Markov time T satisfying $E(T) < \infty$.

Now, we present maximal inequality theorem as follows.

THEOREM 4. Let $\{X_n, n = 0, 1, \dots\}$ be a sub-martingale for which $X_n \geq 0$ for all n . Then for any positive λ ,

$$\lambda \text{Prob}\left\{\max_{0 \leq k \leq n} X_k > \lambda\right\} \leq E(X_n).$$

Using

$$T = \begin{cases} \min\{k \geq 0; X_k > \lambda\}, & \text{if } X_k > \lambda, \text{ for some } k = 0, 1, \dots, n \\ n, & \text{if } X_k \leq \lambda, \text{ for } k = 0, 1, \dots, n \end{cases}$$

and optional stopping theorem, we get

$$E(X_n) \geq E(T) = E\left(X_T I\left\{\max_{0 \leq k \leq n} X_k > \lambda\right\}\right) = \lambda \text{Prob}\left\{\max_{0 \leq k \leq n} X_k > \lambda\right\}.$$

3 Martingale Convergence Theorem

THEOREM 5. *Let $\{X_n, n = 1, 2, \dots\}$ be a sub-martingale satisfying $\sup_{n \geq 0} E(|X_n|) < \infty$. Then there exists a random variable X_∞ to which $\{X_n\}$ converges with probability one,*

$$\text{Prob}\left\{\lim_{n \rightarrow \infty} X_n = X_\infty\right\} = 1. \quad (1)$$

If $\{X_n\}$ is a martingale and is uniformly integrable, that is,

$$\lim_{c \rightarrow \infty} \sup_{n \geq 0} E(|X_n| I\{|X_n| > c\}) = 0$$

then, in addition to (1), $\{X_n\}$ converges in the mean, that is,

$$\lim_{n \rightarrow \infty} E(|X_n - X_\infty|) = 0,$$

and

$$E(X_\infty) = E(X_n), \text{ for all } n.$$

Consider Doob's martingale process. Let $\{Y_n, n = 0, 1, \dots\}$ be an arbitrary sequence of random variables and suppose X is a random variable with $E(|X|) < \infty$. Define

$$X_n = E(X | Y_0, Y_1, \dots, Y_n), n = 0, 1, \dots$$

Since $\{X_n, n = 0, 1, \dots\}$ is a martingale and is uniformly integrable, we may conclude that $\lim_{n \rightarrow \infty} X_n = X_\infty$ for some random variable X_∞ . Moreover, $E(X_\infty) = E(X_0)$ and $E(|X_n - X_\infty|) \rightarrow 0$ as $n \rightarrow \infty$.

4 Martingale in Continuous Time

DEFINITION 6. *Let $(\Omega, \mathcal{F}, P)$ be a probability space. Let T be a fixed positive number. Let $\{\mathcal{F}_t, 0 \leq t \leq T\}$ be a filtration of sub σ -fields of $\mathcal{F}$.*

If

1. $E(X_t)$ exist
2. X_t is $\mathcal{F}_t$ -measurable
3. $E(X_t | \mathcal{F}_s) = X_s, 0 \leq s < t \leq T$. (1)

we say $\{X_t, t \geq 0\}$ has martingale property.

EXAMPLE 7. Let $\{N(t), t \geq 0\}$ be a Poisson process with intensity λ and $\{\mathcal{F}(t), t \geq 0\}$ its natural filtration.

$$\begin{aligned}
 E(N(t) | \mathcal{F}(s)) &= E(N(t) - N(s) + N(s) | \mathcal{F}(s)) \\
 &= E(N(t) - N(s) | \mathcal{F}(s)) + E(N(s) | \mathcal{F}(s)) \\
 &= \lambda(t - s) + N(s)
 \end{aligned}$$

Hence,

$$E((N(t) - \lambda t) | \mathcal{F}(s)) = (N(s) - \lambda s)$$

Therefore, $\{N(t) - \lambda t, t \geq 0\}$ is a martingale.

DEFINITION 8. A stochastic process $\{M_t\}$ is called a martingale process (with respect to a filtration $\{\mathcal{F}_t\}$ and with respect to a probability measure P) if: (a) M_t is integrable for all t , (b) M_t is $\mathcal{F}_t$ -measurable for all t , and (c) $E[M_t | \mathcal{F}_s] = M_s$, for all s, t such that $s \leq t$.

And, $\{M_t\}$ is called a sub-martingale process if $E[M_t | \mathcal{F}_s] \geq M_s$, for all s, t such that $s \leq t$, and is a super-martingale if $E[M_t | \mathcal{F}_s] \leq M_s$, for all s, t such that $s \leq t$.

DEFINITION 9. Sub-martingale and super-martingale are defined by replacing the equality in equation (1) with $\geq$ and $\leq$, respectively, i.e., for any $0 \leq s < t \leq T$,

$$E(X_t | \mathcal{F}_s) \geq X_s \quad (\text{sub - martingale})$$

$$E(X_t | \mathcal{F}_s) \leq X_s \quad (\text{super - martingale}).$$

EXAMPLE 10. Let $\{N(t), t \geq 0\}$ be a Poisson process with intensity λ and $\{\mathcal{F}(t), t \geq 0\}$ its natural filtration.

$$\begin{aligned}
E(N(t) \mid \mathcal{F}(s)) &= E(N(t) - N(s) + N(s) \mid \mathcal{F}(s)) \\
&= E(N(t) - N(s) \mid \mathcal{F}(s)) + E(N(s) \mid \mathcal{F}(s)) \\
&= \lambda(t - s) + N(s).
\end{aligned}$$

Hence,

$$E(N(t) \mid \mathcal{F}(s)) \geq N(s)$$

Therefore, $\{N(t), t \geq 0\}$ is a sub-martingale.

EXAMPLE 11. Show that $\{W(t)^2, t \geq 0\}$ is a sub-martingale.

$$E(W(t)^2 \mid \mathcal{F}(s)) = E((W(t) - W(s) + W(s))^2 \mid \mathcal{F}(s)).$$

Using

$$\begin{aligned}
(W(t) - W(s) + W(s))^2 &= (W(t) - W(s))^2 \\
&\quad + 2W(s)(W(t) - W(s)) + W(s)^2
\end{aligned}$$

$$E(W(t)^2 \mid \mathcal{F}(s)) = (t - s) + 0 + W(s)^2.$$

Hence, $E(W(t)^2 \mid \mathcal{F}(s)) \geq W(s)^2$.

Therefore, $\{W(t)^2, t \geq 0\}$ is a sub-martingale.

Sub-martingales tends to increase and super-martingales tend to decrease. The nomenclature may seem like it goes the wrong way; Doob defined these terms by analogy with the notions of subharmonic and superharmonic functions in analysis. Note that the definition of martingale depends on the collection of σ -fields. When it is needed for clarity, one can say that $\{(M_t, \mathcal{F}_t)\}$ is a martingale. To define conditional expectation, one needs a probability, so a martingale depends on the probability as well. When we need to, we will say that $\{M_t\}$ is a martingale with respect to the probability P . This is an issue when there is more than one probability around.

5 References

1. S Karlin and H M Taylor, “A First Course in Stochastic Processes”, 2nd edition, Academic Press, New York, 1975.
2. J Medhi, “Stochastic Processes”, 3rd edition, New Age International Publishers, 2009.
3. Suresh Chandra, S. Dharmaraja, Aparna Mehra, R. Khemchandani, “Financial Mathematics: An Introduction”, Narosa Publication House, 2012.

Module 6: Brownian Motion

Lecture 1 Brownian Motion

One of the well-known and long studied stochastic model is known as Brownian motion and is named after the English botanist Robert Brown. In 1827, Brown described the unusual motion exhibited by a small particle that is totally immersed in a liquid or gas. It is important in the modelling and analysis of the price movements of stocks and commodities in financial mathematics. A formal mathematical description of Brownian motion and its properties was first given by the great mathematician Norbert Wiener beginning in 1918.

In this module, we will first study about the Brownian motions and their properties. In the latter part of the module, we will also deal with the Brownian motion based stochastic calculus. Let us start with the following simple setup.

DEFINITION 1. Consider a trial whose outcomes are success with probability p or failure with probability $1 - p$. Repeat the trial infinitely many times (e.g., toss a fair coin infinitely). The successive outcomes are denoted as $w = (w_1, w_2, w_3, \dots)$ e.g., $w = (w_1, w_2, w_3, \dots) = (H, T, T, \dots)$ or $(T, H, T, \dots)$. Define

$$X_j = \begin{cases} 1 & \text{if } w_j = H \\ -1 & \text{if } w_j = T \end{cases}.$$

The pmf of X_j is given by

$$P(X_j = 1) = p; \quad P(X_j = -1) = 1 - p; \quad p + q = 1.$$

Set $S_0 = 0$ and define

$$S_k = \sum_{j=1}^k X_j, k = 1, 2, \dots$$

where X_i 's are i.i.d. random variables. Then, $\{S_k, k = 0, 1, \dots\}$ is known as a random walk.

Sample path of the random walk is shown in the figure for a outcome w where $w_1 = T, w_2 = T, w_3 = T, \dots$ with $p = 0.45$.

Properties of Random Walk

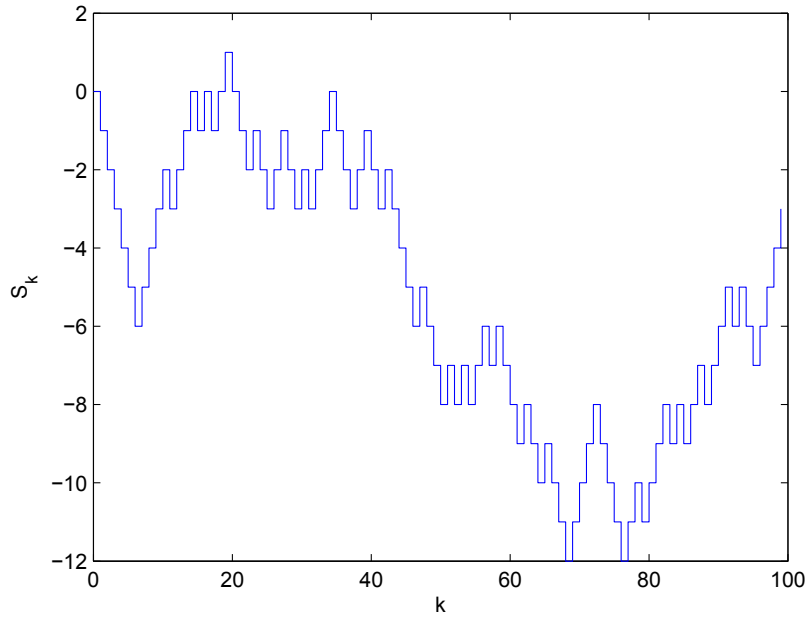


Figure 6.1: A sample path of the random walk

1. Choose non-negative integers $0 = k_0 < k_1 < \dots < k_n$. Then

$$S_{k_{i+1}} - S_{k_i} = \sum_{j=k_i+1}^{k_{i+1}} X_j$$

Since X_j are i.i.d. random variables, $S_{k_1} - S_{k_0}, S_{k_2} - S_{k_1}, \dots, S_{k_n} - S_{k_{n-1}}$ are mutually independent variables. Hence, $\{S_n, n = 0, 1, \dots\}$ has the independent increment property.

2. Similarly, for $1 \leq i < j \leq k < l$

$$S_j - S_i \equiv S_l - S_k.$$

Hence, $\{S_n, n = 0, 1, \dots\}$ has stationary increment property.

3. Also, suppose $S_{k_{i+1}} - S_{k_i} = X_1 + X_2 + \dots + X_n$, one can observe that,

$$E(S_{k_{i+1}} - S_{k_i}) = n(p - q) \text{ and } Var(S_{k_{i+1}} - S_{k_i}) = 4npq.$$

Consider a particle performs a random walk such that in a small interval of time of duration Δt , the displacement of the particle to the right or to the left is also of small magnitude Δx . Let $S(t)$ be denote the total displacement of the particle in time t being x . Let X_j be denote the length of the j th step taken by the particle in a small interval of time Δt with pmf

$$P(X_j = \Delta x) = p; \quad P(X_j = -\Delta x) = 1 - p; \quad p + q = 1$$

$0 < p < 1$, where p is independent of x and t . Partition the interval of length t into n equal subintervals of length Δt . Then $n\Delta t = t$ and the total displacement $S(t)$ is the sum of n i.i.d. random variables X_j , i.e.,

$$S(t) = \sum_{j=1}^{n(t)} X_j, \quad n \equiv n(t) = \frac{t}{\Delta t}.$$

Using $E(X_j) = (p - q)\Delta x$ and $var(X_j) = 4pq(\Delta x)^2$, we get

$$E(S(t)) = nE(X_j) = \frac{t}{\Delta t}(p - q)\Delta x,$$

$$Var(S(t)) = nvar(X_j) = \frac{t}{\Delta t}4pq(\Delta x)^2.$$

$$E(S(t)) = nE(X_j) = \frac{t}{\Delta t}(p - q)\Delta x,$$

$$Var(S(t)) = nvar(X_j) = \frac{t}{\Delta t}4pq(\Delta x)^2.$$

To get a meaningful result; as $\Delta x \rightarrow 0$, $\Delta t \rightarrow 0$, we must have

$$\frac{(\Delta x)^2}{\Delta t} \rightarrow \text{a limit}, \quad (p - q) \rightarrow \text{a multiple of } (\Delta x).$$

As $\Delta x \rightarrow 0$, $\Delta t \rightarrow 0$, per unit time $E(S(t)) \rightarrow \mu$ and $Var(S(t)) \rightarrow \sigma^2$. Hence,

we get

$$\begin{aligned}\Delta x &= \sigma(\Delta t)^{1/2} \\ p &= \frac{1}{2}(1 + \mu(\Delta t)^{1/2}/\sigma) \\ q &= \frac{1}{2}(1 - \mu(\Delta t)^{1/2}/\sigma).\end{aligned}$$

Now, we present the limiting distribution of symmetric random walk is Brownian motion using central limit theorem.

Limiting Case of Random Walk

Let $X_1, X_2, \dots$ be a sequence of independent identically distributed random variables with finite mean μ and finite non zero variance σ^2 and let $S_n = X_1 + X_2 + \dots + X_n$. Then $\frac{S_n - n\mu}{\sigma\sqrt{n}}$ converges in distribution to an $\mathcal{N}(0, 1)$ random variable as $n \rightarrow \infty$. In our setting $\mu = E(X_i) = 0$ and $\sigma^2 = Var(X_i) = 1$. Thus, for large $n(t) (= n)$, $S(t) = \sum_{j=1}^{n(t)} X_j$ converges in distribution to $\mathcal{N}(\mu t, \sigma^2 t)$. Since t represents the length of the interval of time during which the displacement, we conclude for $0 < s < t$, $\{S(t) - S(s)\}$ is $\mathcal{N}(\mu(t-s), \sigma^2(t-s))$. Further, the increments $\{S(s) - S(0)\}$ and $\{S(t) - S(s)\}$ are mutually independent.

DEFINITION 2. A stochastic process $\{W(t), t \geq 0\}$ is said to be a Wiener Process (or Brownian Motion) if

1. For $t > 0$, the sample paths of $W(t)$ are almost surely continuous functions.
2. For $0 \leq t_0 < t_1 < \dots < t_n$ and for all n , increments $W(t_i) - W(t_{i-1})$, $i = 1, 2, \dots, n$ are independent random variables and stationary.
3. For $0 \leq s < t < \infty$, every increment $W(t) - W(s)$ has normal distribution with mean $\mu(t-s)$ and variance $\sigma^2(t-s)$.

A Wiener process $\{W(t), t \geq 0\}$ with $W(0) = 0, \mu = 0, \sigma = 1$ is called a standard Wiener process.

Sample path of Brownian motion is shown in the figure as follows.

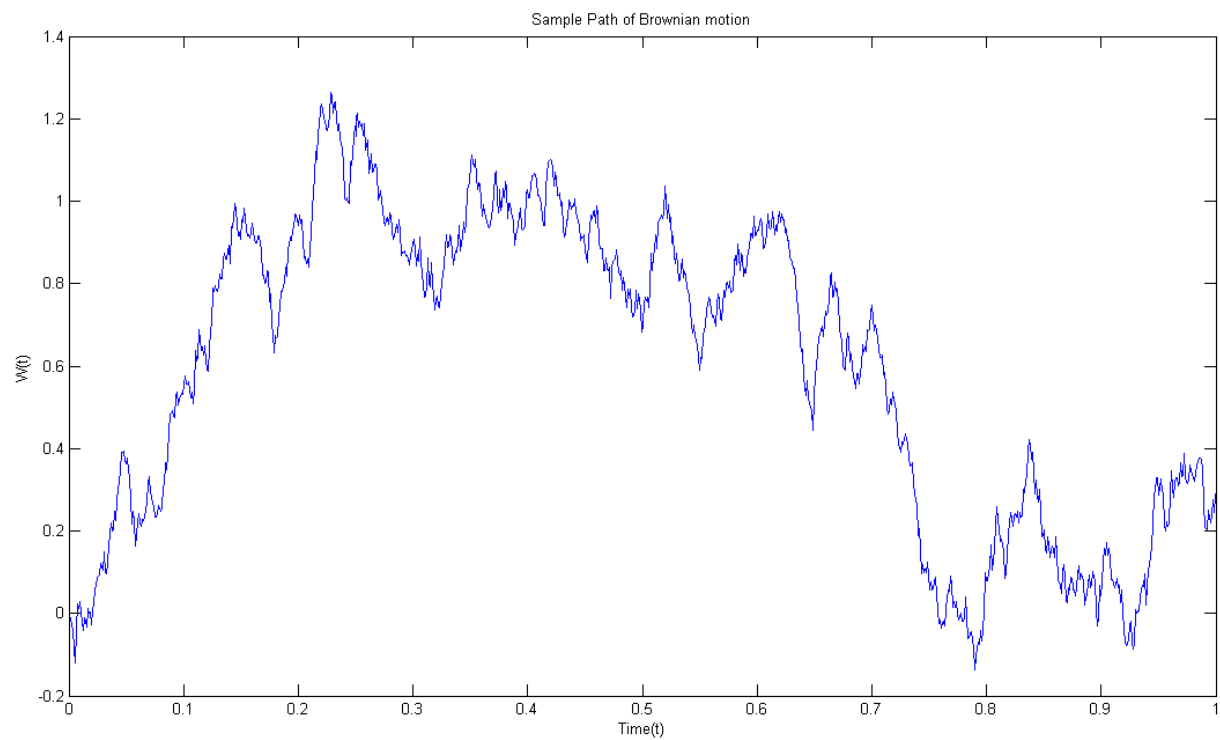


Figure 6.2: A sample path of a Brownian motion

Lecture 2 Properties of Brownian Motions

In this lecture, we present the properties of Brownian motions (BMs).

Nowhere Differential Property

It is not possible to define a tangent line at any point in the sample path. For the proof of the result that with probability one Brownian motion paths are nowhere differentiable, one may refer to advanced texts (see, for example, Breiman, 1968). Here we show a simple statement: For any t almost all trajectories of Brownian motion are not differentiable at t . To see why this is true, consider $\frac{W(t_0+\Delta t)-W(t_0)}{\Delta t} = \frac{\sqrt{\Delta t}Z}{\Delta t} = \frac{Z}{\sqrt{\Delta t}}$, for some standard Normal random variable Z . Thus the ratio diverges to ∞ in distribution, since $P(|\frac{Z}{\sqrt{\Delta t}}| > K) \rightarrow 1$ for any K , as $\Delta t \rightarrow 0$, and this precludes existence of the derivative at t .

Strict-sense Stationary Increments Property

The covariance function $C(s, t)$, for $s \leq t$,

$$\begin{aligned}
 C(s, t) &= E[(W(t) - E(W(t)))(W(s) - E(W(s)))] \\
 &= E[W(t)W(s)] \\
 &= E[(W(t) - W(s) + W(s))W(s)] \\
 &= E[(W(s))^2] \\
 &= s.
 \end{aligned}$$

Hence, $C(s, t) = \min(s, t)$. Therefore, the Wiener process is not wide-sense stationary. But, the Wiener process is strict-sense stationary increments.

Self-similar Property

DEFINITION 1. A stochastic process is said to be H -self-similar for some $H > 0$, if each finite dimensional random vector satisfy the condition, for every $T > 0$, any choice of $t_i \geq 0, i = 1, 2, \dots, n$ and $n \geq 1$

$$(T^H X(t_1), T^H X(t_2), \dots, T^H X(t_n)) = (X(Tt_1), X(Tt_2), \dots, X(Tt_n)).$$

Here, Wiener process is 0.5-self-similar.

Markov property

From the definition, $W(t+s) - W(s)$ is independent of the past, or alternatively, if we know $W(s) = x_0$, then no further knowledge of the values of $W(\tau)$ for $\tau < s$ has any effect on the knowledge of the probability law governing $W(t+s) - W(s)$. Given $W(t)$, the future $W(t+h)$ for any $h > 0$ only depends on the future increment $W(t+h) - W(t)$ and this future is independent of the past. Thus, if $t_0 < t_1 < \dots < t_n < t$,

$$\begin{aligned} P[W(t) \leq x \mid W(t_0) = x_0, W(t_1) = x_1, \dots, W(t_n) = x_n] \\ = P[W(t) \leq x \mid W(t_n) = x_n]. \end{aligned}$$

Hence, the Markov property is satisfied. Thus, $\{W(t), t \geq 0\}$ is a Markov process.

Gaussian Process

A stochastic process $\{X(t), t \geq 0\}$ is called Gaussian process if the distribution of each finite dimensional random vector is multivariate Gaussian (normal) distributed. Then, the joint pdf of $(X(t_1), X(t_2), \dots, X(t_n))$ is given by:

$$f_X(x) = \frac{1}{(2\pi)^{n/2}(\det(\Sigma))^{1/2}} \exp \left[-\frac{1}{2}(X - \mu)\Sigma^{-1}(X - \mu)' \right]$$

where

$$\begin{aligned} \mu &= E(X) = (E(X(t_1)), E(X(t_2)), \dots, E(X(t_n))) \\ \Sigma &= (\text{cov}(X(t_i), X(t_j)); i, j = 1, 2, \dots, n) \\ \text{cov}(X(t_i), X(t_j)) &= E[(X(t_i) - E(X(t_i)))(X(t_j) - E(X(t_j)))]. \end{aligned}$$

Note that, a Gaussian process is independent if and only if uncorrelated, i.e., independent if and only if the covariance matrix Σ is diagonal. Since $\{W(t), t \geq 0\}$ is a Markov process as well as a Gaussian process,

$$\begin{aligned} P[W(t) \leq x \mid W(t_n) = x_n] &= P[W(t) - W(t_n) \leq x - x_n] \\ &= \int_{-\infty}^{x-x_n} \frac{1}{\sqrt{2\pi(t-t_n)}} \exp \left[-\frac{s^2}{2(t-t_n)} \right] ds. \end{aligned}$$

Kolmogorov Equation

We know that Brownian motion is a Markov process with continuous time and continuous state space. Let the transition probability density p be given by

$$p(x_0, s; x, t)dx = P\{x \leq W(t) < x + dx \mid W(s) = x_0\}.$$

We make the following assumptions. For any $\delta > 0$,

$$P\{|W(t) - W(s)| > \delta \mid W(s) = x\} = o(t - s), s < t.$$

In other words, small changes occur during small intervals of time.

$$\begin{aligned} & \lim_{\Delta t \rightarrow 0} \frac{E\{W(t + \Delta t) - W(t) \mid W(t) = x\}}{\Delta t} \\ &= \lim_{\Delta t \rightarrow 0} \int_{|y-x| \leq \delta} (y - x)p(x, t; y, t + \Delta t)dy \\ &= a(t, x). \end{aligned}$$

In other words, the limit of the infinitesimal mean of the conditional expectation of the increment of $W(t)$ exists and is equal to $a(t, x)$, which is known as the drift coefficient.

$$\begin{aligned} & \lim_{\Delta t \rightarrow 0} \frac{E\{[W(t + \Delta t) - W(t)]^2 \mid W(t) = x\}}{\Delta t} \\ &= \lim_{\Delta t \rightarrow 0} \int_{|y-x| \leq \delta} (y - x)^2 p(x, t; y, t + \Delta t)dy \\ &= b(t, x). \end{aligned}$$

In other words, the limit of the infinitesimal mean of the variance of the increment of $W(t)$ exists and is equal to $b(t, x)$, which is known as the diffusion coefficient. A Markov process $\{W(t), t \geq 0\}$ satisfying the above conditions is known as a diffusion process and the partial equation satisfied by its transition pdf is known as diffusion equation. Let $\{W(t), t \geq 0\}$ be a Markov process satisfying the above three conditions. If the transition pdf $p(x_0, t_0; x, t)$ possesses continuous partial derivatives

$$\frac{\partial p}{\partial t}, \frac{\partial a(t, x)p}{\partial x}, \frac{\partial^2 b(t, x)p}{\partial x^2}$$

then $p(x_0, t_0; x, t)$ satisfies the forward Kolmogorov equation

$$\frac{\partial p}{\partial t} = -\frac{\partial a(t, x)p}{\partial x} + \frac{1}{2} \frac{\partial^2 b(t, x)p}{\partial x^2}.$$

This equation is also known as the Fokker-Planck equation.

Connection with Heat Equation

Let $\{W(t), t \geq 0\}$ be a standard Brownian motion. Let the transition probability density p be given by

$$p(x, t) = \frac{1}{\sqrt{2\pi t}} \exp\left[-\frac{x^2}{2t}\right].$$

The solution of diffusion equation

$$\frac{\partial p}{\partial t} = \frac{1}{2} \frac{\partial^2 p}{\partial x^2}$$

is the transition probability density function p .

Joint Distribution of Wiener Process

Consider the joint distribution $(W(t_1), W(t_2))$. We know that, $W(t_1)$ and $(W(t_2) - W(t_1))$ are independent. Also, $W(t_1)$ is $\mathcal{N}(0, t_1)$ and $W(t_2) - W(t_1)$ is $\mathcal{N}(0, t_2 - t_1)$. Then the joint pdf of $(W(t_1), W(t_2))$ is

$$f(x_1, x_2) = p(x_1, t_1)p(x_2 - x_1, t_2 - t_1)$$

where

$$p(x, t) = \frac{1}{\sqrt{2\pi t}} \exp\left[-\frac{x^2}{2t}\right], t > 0; -\infty < x < \infty.$$

Thus,

$$f(x_1, x_2) = \frac{1}{2\pi\sqrt{t_1(t_2 - t_1)}} \exp\left[-\frac{1}{2} \left\{ \frac{x_1^2}{t_1} + \frac{(x_2 - x_1)^2}{(t_2 - t_1)} \right\}\right].$$

$(W(t_1), W(t_2), \dots, W(t_n))$ is jointly normal distributed with cdf for $0 < t_1 < t_2 < \dots < t_n$ is given by

$$P(W(t_1) \leq a_1, W(t_2) \leq a_2, \dots, W(t_n) \leq a_n) =$$

$$\frac{1}{\sqrt{(2\pi)^n t_1(t_2 - t_1) \cdots (t_n - t_{n-1})}} \int_{-\infty}^{a_1} \cdots \int_{-\infty}^{a_n} \exp \left[-\frac{1}{2} \left(\frac{x_1^2}{t_1} + \frac{(x_2 - x_1)^2}{t_2 - t_1} + \cdots + \frac{(x_n - x_{n-1})^2}{t_n - t_{n-1}} \right) \right] dx_1 dx_2 \cdots dx_n.$$

The joint pdf of $(W(t_1), W(t_2), \dots, W(t_n))$ is given by:

$$f_X(x) = \frac{1}{(2\pi)^{n/2} (\det(\Sigma))^{1/2}} \exp \left[-\frac{1}{2} (X - \mu) \Sigma^{-1} (X - \mu)' \right]$$

where

$$\begin{aligned} \mu &= E(X) = (E(W(t_1)), E(W(t_2)), \dots, E(W(t_n))) \\ &= (0, 0, \dots, 0). \end{aligned}$$

$$\begin{aligned} \Sigma &= (\text{cov}(W(t_i), W(t_j)); i, j = 1, 2, \dots, n) \\ &= \begin{pmatrix} t_1 & t_1 & \dots & t_1 \\ t_1 & t_2 & \dots & t_2 \\ \vdots & \vdots & \ddots & \vdots \\ t_1 & t_2 & \dots & t_n \end{pmatrix}. \end{aligned}$$

Martingale Property

Let $\{\mathcal{F}(t), t \geq 0\}$ be the natural filtration. We know that,

1. $E(W(t))$ exist
2. $W(t)$ is $\mathcal{F}(t)$ -measurable .

Also,

$$\begin{aligned} E(W(t) \mid \mathcal{F}(s)) &= E(W(t) - W(s) + W(s) \mid \mathcal{F}(s)) \\ &= E(W(t) - W(s) \mid \mathcal{F}(s)) + E(W(s) \mid \mathcal{F}(s)) \\ &= 0 + W(s). \end{aligned}$$

Therefore, $\{W(t), t \geq 0\}$ is a martingale.

Multidimensional Brownian Motion

DEFINITION 2. *A n -dimensional Brownian motion is a stochastic process*

$$W(t) = (W_1(t), W_2(t), \dots, W_n(t))$$

with the following properties:

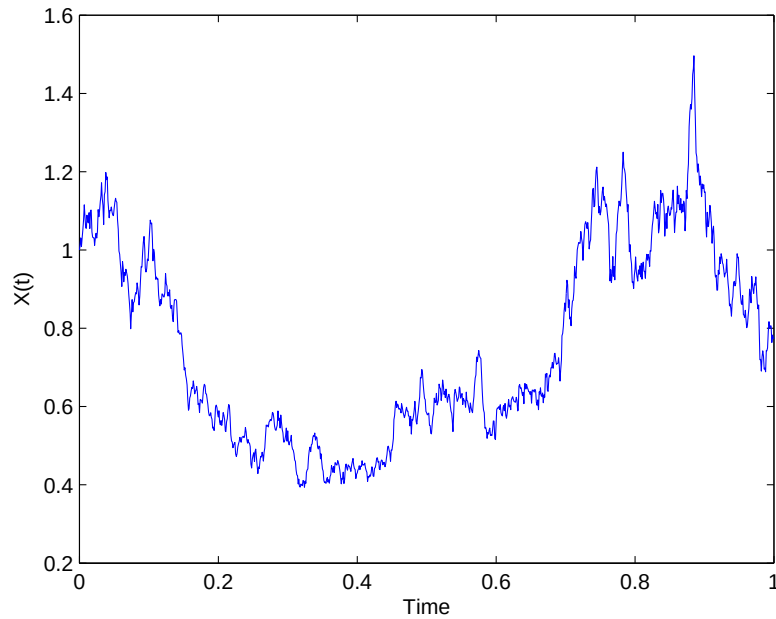
1. *Each $W_i(t)$ is a one-dimensional Brownian motion.*
2. *If $i \neq j$, then the stochastic processes $W_i(t)$ and $W_j(t)$ are independent.*

Lecture 3 Processes Derived from Brownian Motion

Geometric Brownian Motion

DEFINITION 1. A stochastic process $\{X(t), t \geq 0\}$ is said to be a geometric Brownian motion if $X(t) = X(0)e^{W(t)}$ where $W(t)$ is a Brownian motion.

Sample path is shown as follows.



Markov Property

$$\begin{aligned}
 X(t+h) &= X(0)e^{W(t+h)} \\
 &= X(0)e^{W(t)+W(t+h)-W(t)} \\
 &= X(t)e^{W(t+h)-W(t)}.
 \end{aligned}$$

Since BM has independent increments, given $X(t)$, the future $X(t+h)$ only depends on the future increment of the BM. This future is independent of the past. Hence, the Markov property is satisfied. Thus, $\{X(t), t \geq 0\}$ is a Markov process.

Mean and Variance of geometric BM

The moment generating function of a normal distribution random variable X with mean μ and variance σ^2 is given by

$$M_X(s) = E(e^{sX}) = \exp(\mu s + \frac{1}{2}\sigma^2 s^2), \quad -\infty < s < \infty.$$

For Brownian motion with drift, since $W(t) \sim \mathcal{N}(\mu t, \sigma^2 t)$, we have

$$M_{W(t)}(s) = E(e^{sW(t)}) = \exp(\mu t s + \frac{1}{2}\sigma^2 t s^2), \quad -\infty < s < \infty.$$

By using the values $s = 1, 2$, we get

$$E(X(t)) = X(0)M_{W(t)}(1)$$

and

$$E(X^2(t)) = X^2(0)M_{W(t)}(2).$$

After simplifications, we get

$$E(X(t)) = X(0)e^{(\mu + \frac{\sigma^2}{2})t}$$

$$\text{Var}(X(t)) = X^2(0)e^{2\mu + \sigma^2 t} (e^{\sigma^2 t} - 1).$$

Letting $\bar{r} = \mu + \frac{\sigma^2}{2}$, we get

$$E(X(t)) = X(0)e^{\bar{r}t}$$

and more generally

$$E(X(t)/X(s)) = e^{\bar{r}(t-s)}.$$

Because the geometric Brownian motion is nonnegative, it provides for a more realistic model of stock prices. Suppose that the stock price $S(t)$ at time t is given by

$$S(t) = S(0)e^{H(t)}, t \geq 0$$

where $S(0)$ is the initial price and $H(t) = \sigma W(t) + \mu t$ is a Brownian motion with drift. In this case, $H(t)$ represents a continuously compound rate of return of the stock price over the period of time $[0, t]$. Here, $H(t)$ refer to the logarithmic growth of the

stock price, satisfies $H(t) = \log \left(\frac{S(t)}{S(0)} \right)$. Therefore, $H(t)$ has a normal distribution with mean μt and variance $\sigma^2 t$. Letting $\bar{r} = \mu + \frac{1}{2}\sigma^2$, we get

$$E(S(t)) = S(0)e^{\bar{r}t}$$

$$Var(S(t)) = \left(S(0)e^{(\mu + \frac{\sigma^2}{2})t} \right)^2 (e^{\sigma^2 t} - 1).$$

Observe that, the expected stock price depends not only on the drift μ of $H(t)$ but also on the volatility σ . Further, it shows that, the expected price grows like a fixed-income security with continuously compounded interest rate $\bar{r}$. In real scenario, r is much lower than $\bar{r}$, the real fixed-income interest rate, that is why one invests in stocks. But the stock has variability due to the randomness of the underlying Brownian motion and hence there is risk involved here.

EXAMPLE 2. Suppose that stock price $\{S(t), t \geq 0\}$ follows geometric Brownian motion with drift $\mu = 8$ per year and variance $\sigma^2 = 8.5\%$ per annum. Assume that, the current price of the stock is $S(0) = 60$. find

1. $E[S(3)]$
2. $P[S(3) > 100]$

EXAMPLE 3. The price of a stock follows a geometric Brownian motion with parameters $\mu = 0.12$ and $\sigma = 0.24$. If the present price of the stock is Rs. 40, what is the probability that a call option having four months to exercise time and with a strike price $K = \text{Rs.}42$, will be exercised?

Solution A call option on a stock with exercise time T and strike K is the right, but not the obligation, to buy a certain number of shares at time T for the fixed price K . If $S(t)$ denotes the market price of this number of shares then the value of the option at time T is $S(T) - K$ if $S(T) > K$, since the shares can be bought for K and immediately sold in the market for $S(T)$. On the other hand the option is worthless, and will not be exercised, if $S(T) \leq K$. The required probability is $\text{Prob}\{S(1/3) > 42\}$.

Some Remarks

1. $W(t)$ is normal distribution while $X(t)$ is lognormal distribution.

2. Product of independent lognormal distributions is also lognormal distribution.
3. It removes the negativity problem. Also, it justifies from basic economic principles as a reasonable model for stock prices in an ideal non-arbitrage world.

Levy Process

DEFINITION 4. *A stochastic process $\{X(t), t \geq 0\}$ is said to be a Levy process if it satisfies the following properties*

1. $X(0) = 0$
2. *for $t > 0$, $X(t)$ is almost surely continuous*
3. *for $0 \leq t_0 < t_1 < t_2 < \dots < t_n$ and for all n , increments $X(t_i) - X(t_{i-1})$, $i = 1, 2, \dots, n$ are independent and stationary*
4. *for $a > 0$, $P(|X(t) - X(s)| > a) \rightarrow 0$ when $t \rightarrow s$.*

One can observe that, Poisson process and Brownian motion are the examples of Levy processes.

Lecture 4 Stochastic Differential Equations

When in the 19th century the German mathematician Weierstrass constructed a real-valued function which is continuous, but nowhere differentiable, this was considered as nothing else but a mathematical curiosity. High frequency data show that prices of exchange rates, interest rates, and liquid assets are practically continuous. But they are of unbounded variation in every given time interval. In particular, they are nowhere differentiable. The classical calculus is no longer applicable for real-valued functions occurring in mathematical finance. Therefore classical calculus requires an extension to functions of unbounded variation.

Stochastic calculus is a necessary extension of real analysis to cope with functions of unbounded variation. Some important concepts of the stochastic calculus, stochastic differential equations, Ito's integral and Ito's formula will be discussed in this module.

Variations of Real-Valued Function

DEFINITION 1. Consider the real-valued right-continuous functions g on the time interval $[a, b]$ where $0 \leq a < b < \infty$. The value of g at time t is denoted by $g(t)$. Let Π be the set of all finite subdivisions π of the interval $[a, b]$ with $0 = t_0 < t_1 < \dots < t_n = b$. Define $\|\pi\| = \max_i (t_{i+1} - t_i)$. The variation (or 1st variation) of $g(t)$ over the interval $[a, b]$ is defined as

$$V_g([a, b]) = \sup_{\pi \in \Pi} \sum_{i=0}^{n-1} |g(t_{i+1}) - g(t_i)|.$$

DEFINITION 2. The function $g(t)$ is of finite variation if, for every t ,

$$V_g(t) = V_g([0, t]) < +\infty.$$

For all t , if $V_g(t) < K$, a constant independent of t , then $g(t)$ is of bounded variation.

DEFINITION 3. (p -variation of Real-valued Function) p -variation of a real-valued

function $g(t)$ in the interval $[0, t]$ is defined as

$$V_g^p(t) = \sup_{\pi \in \Pi} \sum_{i=0}^{n-1} |g(t_{i+1}) - g(t_i)|^p.$$

Now, we present the important result on p -variation without proof (refer [3]):

THEOREM 4. *Let the p - variation of the function $g(t)$ in the interval $[0, T]$, denoted by $V_g^p(T)$, defined as: If $V_g^p(T) < \infty$*

then $V_g^r(T) = \infty$ $p > r$

and

$V_g^q(T) = 0$ $p < q$.

EXAMPLE 5. *Consider $g(t) = t^2$. We get*

$$g'(t) = 2t, \int_0^1 g'(t) dt = 1$$

$$\int_0^1 |g'(t)|^2 dt = 4 \int_0^1 t^2 dt = 4 \frac{t^3}{3} \Big|_0^1 = \frac{4}{3}.$$

But

$$\lim_{\|\pi\| \rightarrow 0} \|\pi\| \int_0^1 |g'(t)|^2 dt = 0$$

$$V_g(1) = 1; V_g^2(1) = 0$$

By applying Theorem 4, we have

$$V_g^p(1) = \begin{cases} 1, & p = 1 \\ 0, & p > 1 \end{cases}.$$

We have already shown that sample paths of $W(t)$ are nowhere differentiable. The first variation does not make sense because of the above reason. Hence, 1st variation does not exist for Brownian motion.

Quadratic Variation of Brownian Motion

The quadratic variation for Brownian motion over the interval $[0, T]$, denoted by $[W(t), W(t)](T)$, is given by:

$$[W(t), W(t)](T) = V_{W(t)}^2(T) = \lim_{\|\pi\| \rightarrow 0} Q_\pi$$

where

$$Q_\pi = \sum_{i=0}^{n-1} (W(t_{i+1}) - W(t_i))^2.$$

Clearly, Q_π is a function of the sample points $w \in \Omega$. Hence, the quadratic variation calculated for a Brownian motion for each partition itself a random variable. Note that, the limit is taken over all partitions of $[0, T]$, with $\|\pi\| \rightarrow 0$ as $n \rightarrow \infty$.

Since, for each partition π , Q_π is a random variable, how to find the limiting distribution of Q_π for large n ? The question would be what is the proper mode of convergence in these random variables. We shall use convergence in mean square (convergence in L^2). The definition of convergence in L^2 is as follows:

DEFINITION 6. Let $\{X_n, n \geq 1\}$ and X be random variables defined on a common probability space $(\Omega, \mathcal{F}, P)$. We say that X_n converges in L^2 to X if

$$\lim_{n \rightarrow \infty} E(|X_n - X|^2) = 0.$$

In the case of Brownian motion, we will show that

$$\lim_{\|\pi\| \rightarrow 0} E(|Q_\pi - T|^2) = 0. \quad (1)$$

When the above result holds good, we say that the quadratic variation accumulated by the Brownian motion over the interval $[0, T]$ is T almost surely and is denoted as $[W, W](T) = T$.

Let us prove the above result in the following theorem.

THEOREM 7. 1. $E(Q_\pi) = T$

2. $Var(Q_\pi) \leq 2 \|\pi\| T$

3. $E(Q_\pi - T)^2 = Var(Q_\pi)$.

Proof.

$$E(Q_\pi) = \sum_{i=0}^{n-1} E[W(t_{i+1}) - W(t_i)]^2.$$

Since, for fixed i , $W(t_{i+1}) - W(t_i)$ is normal distribution with mean 0 and variance $(t_{i+1} - t_i)$

$$= \sum_{i=0}^{n-1} (t_{i+1} - t_i) = T$$

$$\begin{aligned} \text{Var}(Q_\pi) &= \sum_{i=0}^{n-1} \text{Var}[W(t_{i+1}) - W(t_i)]^2 \\ \text{Var}[W(t_{i+1}) - W(t_i)]^2 &= E[W(t_{i+1}) - W(t_i)]^4 \\ &\quad - 2E[W(t_{i+1}) - W(t_i)]^2(t_{i+1} - t_i) \\ &\quad + (t_{i+1} - t_i)^2. \end{aligned}$$

Using the fourth order moment of normal distribution with mean 0 and variance $(t_{i+1} - t_i)$ is $3(t_{i+1} - t_i)^2$, we get

$$\begin{aligned} \text{Var}[W(t_{i+1}) - W(t_i)]^2 &= 3(t_{i+1} - t_i)^2 - 2(t_{i+1} - t_i)^2 + (t_{i+1} - t_i)^2 \\ &= 2(t_{i+1} - t_i)^2 \\ &\leq 2\|\Pi\|(t_{i+1} - t_i). \end{aligned}$$

Therefore $\text{Var}(Q_\pi) \leq 2 \|\Pi\| T$

$$\begin{aligned} E(|Q_\pi - T|^2) &= E((Q_\pi - T)^2) \\ &= E(Q_\pi - E(Q))^2 \\ &= \text{Var}(Q_\pi). \end{aligned}$$

Then, we get

$$[W, W](T) = \lim_{\|\pi\| \rightarrow 0} Q_\pi = T$$

since

$$\lim_{\|\Pi\| \rightarrow 0} E((Q_\pi - T)^2) = 0.$$

Hence $[W, W](T) = T$. This means it accumulates unit quadratic variation per unit time. In other words, the above phenomenon can be represented in differential form

as $dW(t)dW(t) = dt$. By applying Theorem 4, we have

$$V_W^p(T) = \begin{cases} \infty, & p = 1 \\ T, & p = 2 \\ 0, & p > 2 \end{cases}.$$

This concludes that, the Brownian motion $\{W(t), t \geq 0\}$ is of unbounded variation and finite quadratic variation for every t . $\square$

Introduce uncertainties by introducing as additive white noise term, i.e.,

$$dX(t) = b(t, X(t))dt + dW(t)$$

where $b : [0, T] \times \mathbb{R} \rightarrow \mathbb{R}$ is a continuous function. Time derivative of a Brownian motion is called a *white noise process*. It is used as a model where uncertainties arise from various sources.

Note that $\{X(t), t \geq 0\}$ is a stochastic process. The integral form is

$$X(t) = X(0) + \int_0^t b(s, X(s))ds + W(t)$$

a stochastic integral equation.

In general, if $b : [0, T] \times \mathbb{R} \rightarrow \mathbb{R}, \sigma : [0, T] \times \mathbb{R} \rightarrow \mathbb{R}$ are two suitable functions, then an integral equation of the form

$$X(t) = X(0) + \int_0^t b(s, X(s))ds + \int_0^t \sigma(s, X(s))dW(s) \quad (2)$$

is called a stochastic integral equation.

It is defined by integration of a stochastic process with respect to Brownian motion. Equation (2) can be written as

$$dX(t) = b(t, X(t))dt + \sigma(t, X(t))dW(t), \quad (0 \leq t \leq T) \quad (3)$$

where $b : [0, T] \times \mathbb{R} \rightarrow \mathbb{R}$ and $\sigma : [0, T] \times \mathbb{R} \rightarrow \mathbb{R}$ are two given functions.

The equation (2) is referred to as a stochastic differential equation. The interpretation of (2) tells us that the change $dX(t) = X(t + \Delta t) - X(t)$ is caused by a change dt of time, with factor $b(t, X(t))$ in combination with a change $dW(t) = W(t + \Delta t) - W(t)$ of Brownian motion with factor $\sigma(t, X(t))$.

Strong Solution

Let $(\Omega, \mathcal{F}, P)$ be a probability space, and $\{W(t), t \geq 0\}$ be a Brownian motion defined on it. An adapted process $X(t)$ satisfying (2) is said to be a strong solution (uniquely) if $\{X(t)\}$ and $\{Y(t)\}$ are two solutions on the same probability space satisfying (2) then

$$P\{X(t) = Y(t) \forall t\} = 1.$$

A strong solution is an explicit function f such that $X(t) = f(t, W(s), s < t)$.

Weak Solution

Both strong and weak solutions require the existence of the process $\{X(t), t \geq 0\}$ that solves the integral equation version of the SDE. The difference between the two lies in the underlying probability space $(\Omega, \mathcal{F}, P)$. A weak solution consists of a probability space and the process that satisfies the integral equation, while a strong solution is a process that satisfies the equation and is defined on a given probability space. When no known explicit solution exists for a given SDE, then we can approximate it by a numerical solution, replacing differentials by differences. Hence, approximate solution method is similar to numerical integration.

Existence and Uniqueness of Solution

Now, we discuss the existence of strong solution. Suppose $b : [0, T] \times \mathbb{R} \rightarrow \mathbb{R}, \sigma : [0, T] \times \mathbb{R} \rightarrow \mathbb{R}$ are continuous and satisfying Lipschitz condition

$$|b(t, x) - b(t, y)| + |\sigma(t, x) - \sigma(t, y)| < k |x - y|$$

and $X(0)$ and $W(t)$ are independent random variables. Then the equation (2) has a unique solution.

Ito-Picard Iteration

Note that, the existence and uniqueness follow very closely the standard Picard's method for constructing solutions of ODE. Using the Ito-Picard iteration and $X(0) =$

x_0 , we get for $n = 1, 2, \dots$

$$X_{n+1}(t) = x_0 + \int_0^t b(s, X_n(s))ds + \int_0^t \sigma(s, X_n(s))dW(s).$$

Remark that, the iterations are well defined. By the convergence of iteration scheme, we finally obtain

$$X(t) = \lim_{n \rightarrow \infty} X_n(t).$$

EXAMPLE 8. Consider the stochastic differential equation

$$dX(t) = X(t)dW(t), \text{ with } X(0) = 1.$$

Here, $b(t, x) = 0$ and $\sigma(t, x) = x$. Hence, Lipschitz condition is satisfied. Hence, the strong solution exists. Obtaining the strong solution will be explained in further lectures.

EXAMPLE 9. Let $S(t)$ be the stock price at time t . Consider the SDE

$$dS(t) = \mu S(t)dt + \sigma S(t)dW(t), \quad S(0) \text{ is known}$$

where μ is the constant growth rate of the stock and σ is the volatility. Here, $b(t, x) = \mu x$ and $\sigma(t, x) = \sigma x$. Since μ and σ are constants, Lipschitz condition is satisfied. Hence, the strong solution exists. Obtaining the strong solution will be explained in further lectures.

Lecture 5 Ito Integrals

DEFINITION 1. Let $\{X(t), 0 \leq t \leq T\}$ be a stochastic process. Let $\{X(t), 0 \leq t \leq T\}$ be adapted to the natural filtration $\{\mathcal{F}(t), t \geq 0\}$ of Wiener process $\{W(t), 0 \leq t \leq T\}$, i.e., $X(t)$ be $\mathcal{F}(t)$ -measurable. Define

$$I(t) = \int_0^t X(u) dW(u), \quad 0 \leq t \leq T \quad (1)$$

a stochastic integral with respect to a Wiener process. The above integral is called Ito integral.

Ito Process

DEFINITION 2. Let $\{W(t), t \geq 0\}$ be a Brownian motion and let $\{\mathcal{F}(t), t \geq 0\}$ be an associated natural filtration. An Ito process is a stochastic process $\{X(t), t \geq 0\}$ of the form

$$X(t) = X(0) + \int_0^t \Delta(u) dW(u) + \int_0^t \Theta(u) du$$

where $X(0)$ is a non-random and $\Delta(u)$ and $\Theta(u)$ are adapted processes and $\Delta(u)$ is mean square integrable.

It is now easy to write a stochastic differential equation form of an Ito process which is

$$dX(t) = \Delta(t) dW(t) + \Theta(t) dt.$$

All stochastic processes with no jumps are actually Ito processes. Ito integral and Brownian motion are examples of Ito processes.

If $X(t)$ and $W(t)$ are not stochastic processes, but rather deterministic functions. Assuming $f(s)$ is a differential function and $g(s)$ is a smooth function,

$$\int_0^t g(s) df(s) = \int_0^t g(s) f'(s) ds$$

which is a Reimann integral. But what happens if $f(s)$ is not differentiable.

We can still define the integral. When $f(s)$ is bounded variation, we can prove that the integral is well-defined. Further, if the function $g(s)$ is extremely fluctuating at different points in time, the limit may still diverge.

Since $f(s)$ has bounded variation, one can prove that the limit exists as long as $g(s)$ is not varying too much and is given by

$$\int_0^t g(s)df(s) = \lim_{n \rightarrow \infty} \sum_{i=0}^{n-1} g(s_i) (f(s_{i+1}) - f(s_i)).$$

Remember that, in the integral (1), $X(t)$ and $W(t)$ are stochastic processes as well as $W(t)$ is unbounded variation and is nowhere differentiable. Hence, it is different from Riemann integral. Now, we rewrite the Ito integral (1) in the above form as:

$$I(t) = \lim_{n \rightarrow \infty} \sum_{i=0}^{n-1} X(s_i) (W(s_{i+1}) - W(s_i)).$$

Ito-Integrable

A stochastic process $\{X(t), t \geq 0\}$ is called Ito-integrable on $[0, t]$ if it satisfies

1. $X(t)$ is adapted, i.e., $X(s)$ is $\mathcal{F}(s)$ - measurable, $0 \leq s \leq t$.
2. $\int_0^t E(X^2(s))ds < \infty$.

EXAMPLE 3. Consider the Ito Integral for simple integrand function. Consider the illustration with price may be negative. Let $W(t)$ be the price per share of an asset at time (position) t , $X(t)$ be the number of shares taken in the asset at a time t and t_i be the trading dates of an asset. Assume $X(t)$ be the simple process. It means $X(t)$ is a constant in each $[t_i, t_{i+1}]$. The gain/loss from trading at each time t can be viewed as $I(t)$.

Since $\{X(t), 0 \leq t \leq T\}$ is a simple process, it can be written as

$$X(t, w) = \phi_0(w)1_{\{0\}}(t) + \lim_{p \rightarrow \infty} \sum_{i=1}^p \phi_i(w)1_{(t_{i-1}, t_i]}(t) \quad \forall \quad w \in \Omega$$

where ϕ_i are bounded random variables such that ϕ_0 is $\mathcal{F}(0)$ -measurable, ϕ_i is $\mathcal{F}(t_{i-1})$ -measurable and $0 < t_0 < t_1 < \dots < t_p = T, p \in \mathcal{N}$. For a simple process $\{X(t), t \in [0, T]\}$, the stochastic integral $I(t)$ is defined by

$$I(t) = \int_0^t X(s)dW(s) = \lim_{p \rightarrow \infty} \sum_{i=1}^p \phi_i(W(t_i) - W(t_{i-1})).$$

In particular $X(t) = 1$, $t \in [0, T]$, hence $\phi_i = 1$, $1 \leq i \leq p$, $p \in \mathcal{N}$, we have, for $t \in [0, T]$

$$I(t) = \int_0^t dW(s) = \lim_{p \rightarrow \infty} \sum_{i=1}^p (W(t_i) - W(t_{i-1})) = W(t).$$

EXAMPLE 4. Consider Ito integral of a deterministic integrand. Let $\{W(t), t \geq 0\}$ be a Brownian motion and let $\Delta(t)$ be a non-random function of time. Define $I(t) = \int_0^t \Delta(s) dW(s)$.

$$E(I(t)) = I(0) = 0 \quad (\text{martingale})$$

$$E(I^2(t)) = \int_0^t \Delta^2(s) ds$$

$$\text{Var}(I(t)) = \int_0^t \Delta^2(s) ds$$

From moment generating function, we get

$$E(e^{uI(t)}) = e^{\frac{1}{2}u^2 \int_0^t \Delta^2(s) ds}.$$

Hence, for each $t \geq 0$ the random variable $I(t)$ is normally distributed with mean zero and variance $\int_0^t \Delta^2(s) ds$

$$\text{i.e., } I(t) \sim \mathcal{N}\left(0, \int_0^t \Delta^2(s) ds\right).$$

EXAMPLE 5. Evaluate

$$\int_0^t W(1) dW(t), \quad 0 \leq t \leq 1.$$

Note that, $W(1)$ is not adapted to the filtration $\sigma\{W(s), 0 < s \leq t\}, 0 \leq t \leq 1$. Hence, Ito integral does not exist. This example shows that, assumption of the integrand is adapted to the filtration $\{\mathcal{F}(t), t \geq 0\}$ is need to have existence of the Ito integral.

EXAMPLE 6. Evaluate the Ito integral

$$\int_0^T W(t) dW(t)$$

By using the definition, we get $0 < t_0 < t_1 < \dots < t_n = T, n \in \mathcal{N}$.

$$\int_0^T W(t) dW(t) = \lim_{n \rightarrow \infty} \sum_{i=0}^{n-1} W(t_i) (W(t_{i+1}) - W(t_i)).$$

Note that, for each i , $W(t_i)$ and $W(t_{i+1}) - W(t_i)$ are independent variables and are having normal distributions. Let Π be the set of all finite subdivisions of π of the interval $[0, T]$ with $0 < t_0 < t_1 < \dots < t_n = T$.

$$\begin{aligned} Q_\Pi &= \sum_{j=0}^{n-1} (W(t_{i+1}) - W(t_i))^2 \\ &= \sum_{j=0}^{n-1} ((W(t_{i+1}))^2 - (W(t_i))^2 - 2W(t_i)(W(t_{i+1}) - W(t_i))) \\ &= (W(T))^2 - (W(0))^2 - 2 \sum_{i=0}^{n-1} W(t_i)(W(t_{i+1}) - W(t_i)) \end{aligned}$$

$$\int_0^T W(t) dW(t) = \frac{(W(T))^2 - T}{2}.$$

Hence,

$$\int_0^T W(s) dW(s) = \frac{(W(T))^2 - T}{2}.$$

We see that, unlike the Reimann integral, we have an extra term T . This arises from the quadratic variation of the Brownian motion which is finite. The above integral is defined not only for the upper limit of integration T but also for every upper limit of integration between 0 and t . Construction of the Ito integral is similar to one of the Stieltjes integration. But instead of integrality with respect to a deterministic function (in Stieltjes integral) ito's integral with respect to a random function, more precisely path of a Wiener process.

Properties of Ito Integral

The following results hold by the Ito integral defined in equation (1).

1. The integral $I(t)$ is a martingale with respect to $\{\mathcal{F}(t), t \geq 0\}$
2. $E(I(t)) = E(\int_0^t X(s) dW(s)) = 0$.

3. $E[(I(t))^2] = E[\int_0^t X^2(s)ds] = \int_0^t E(X^2(s))ds$ for all t . This is called Ito isometry.
4. $Var(I(t)) = \int_0^t E(X^2(s))ds$
5. Quadratic variation is given by

$$[I, I](t) = \int_0^t X^2(s)ds.$$

The Ito integral is a random variable $I(t)$, for all $w \in \Omega$

$$\begin{aligned} I(t)(w) &= \int_0^t X(s, w)dW(s, w) \\ &= \lim_{n \rightarrow \infty} \sum_{i=0}^{n-1} X(s_i, w)(W(s_{i+1}, w) - W(s_i, w)). \end{aligned}$$

Actually the convergence takes place only for a subsequence. Its integral depends on the sample path. Note that

$$\begin{aligned} dW(t)dW(t) &= dt \\ dt dt &= 0 \\ dW(t)dt &= 0. \end{aligned}$$

$$dX(t) = b(t, X(t))dt + \sigma(t, X(t))dW(t) \tag{2}$$

The equation (2) is referred to as a stochastic differential equation. The interpretation of (2) tells us that the change $dX(t) = X(t + \Delta t) - X(t)$ is caused by a change dt of time, with factor $b(t, X(t))$ in combination with a change $dW(t) = W(t + \Delta t) - W(t)$ of Brownian motion with factor $\sigma(t, X(t))$.

Lecture 6 Ito's Formula

In this lecture, we present the important Ito's formula.

Derivation of Ito Formula

Assume that $X(t) = f(W(t))$. Assume that f is twice differentiable. Using Taylors theorem, we get

$$\begin{aligned} f(W(t + \Delta t)) &= f(W(t)) + (W(t + \Delta t) - W(t))f'(W(t)) \\ &\quad + \frac{1}{2!}(W(t + \Delta t) - W(t))^2 f''(W(t)) + \dots \end{aligned}$$

We know that

$$dW(t)dW(t) = dt; dW(t)dt = 0; dt dt = 0.$$

Since $f(W(t + \Delta t)) - f(W(t))$ is $dX(t)$, we have

$$dX(t) = f'(W(t))dW(t) + \frac{1}{2}f''(W(t))dt.$$

In Integral form,

$$X(t) = X(0) + \int_0^t f'(W(s))dW(s) + \frac{1}{2} \int_0^t f''(W(s))ds$$

where $\int_0^t f'(W(s))dW(s)$ is a Ito integral and $\int_0^t f''(W(s))ds$ is a Riemann integral.

In 1951, Ito stated and proved what is now known as *Ito's formula* for stochastic process $\{X(t), t \in [0, T]\}$ as

$$df(X(t)) = f'(X(t))dX(t) + \frac{1}{2}f''(X(t))d[X, X](t).$$

EXAMPLE 1. Evaluate $\int_0^T W(t)dW(t)$. Consider

$$\begin{aligned} df(W(t)) &= f'(W(t))dW(t) + \frac{1}{2}f''(W(t))dt \\ f(W(T)) &= f(W(0)) + \int_0^T f'(W(t))dW(t) \\ &\quad + \frac{1}{2} \int_0^T f''(W(t))dt. \end{aligned}$$

From the above equations, take $f(x) = \frac{x^2}{2}$,

$$\int_0^T f'(W(t))dW(t) = \int_0^T W(t)dW(t).$$

$$f'(x) = x; f''(x) = 1.$$

$$\begin{aligned} f(W(T)) - f(W(0)) &= \frac{(W(T))^2}{2} - 0 \\ \int_0^T dt &= T \\ \int_0^T W(t)dW(t) &= \frac{(W(T))^2 - T}{2}. \end{aligned}$$

Hence,

$$\int_0^T W(t)dW(t) = \frac{(W(T))^2 - T}{2}.$$

In variable limit of integration, we have

$$\int_0^t W(s)dW(s) = \frac{1}{2}(W(t))^2 - t, \quad t \geq 0$$

additional term $-\frac{t}{2}$ is because of local increment of the Wiener process over an increment of length Δt is of the size of its standard deviation $\sqrt{\Delta t}$. For smooth continuously differentiable function $f(x)$, the second term in RHS is 0.

EXAMPLE 2. Show that $(W(t))^3$ is an Ito process and find the stochastic differential of $(W(t))^3$. Let $f(x) = x^3$. Apply Ito formula, we get

$$(W(T))^3 = \frac{1}{2} \int_0^T 6W(t)dt + \int_0^T 3(W(t))^2dW(t).$$

One can prove that $3W(t)$ and $3(W(t))^2$ are adapted processes and $3(W(t))^2$ is mean square integrable. Hence, $(W(t))^3$ is an Ito process. In differential form

$$d(W(t))^3 = 3W(t)dt + 3(W(t))^2dW(t).$$

Ito Formula Version 1

The simple form of Ito formula is

$$X(t) = X(0) + \int_0^t f'(W(u))dW(u) + \frac{1}{2} \int_0^t f''(W(u))du$$

where $\int_0^t f'(W(u))dW(u)$ is a Ito integral
and $\int_0^t f''(W(u))du$ is a Riemann integral.

Ito Formula Version 2

Let $f(t, x)$ be continuous partial derivatives of at least second order. We start with $X(t) = f(t, W(t))$. Therefore expression of $f(t, x)$ be

$$\begin{aligned} f(t + \Delta t, x + \Delta x) &= f(t, x) + \frac{\partial f(t, x)}{\partial t}dt + \frac{\partial f(t, x)}{\partial x}dx \\ &+ \frac{1}{2} \left[\frac{\partial^2 f(t, x)}{\partial t^2}(dt)^2 + 2 \frac{\partial^2 f(t, x)}{\partial t \partial x}dtdx \right. \\ &\left. + \frac{\partial^2 f(t, x)}{\partial x^2}(dx)^2 \right] + \dots \end{aligned}$$

$$\begin{aligned} f(t + \Delta t, W(t + \Delta t)) &= f(t, W(t)) + \frac{\partial f(t, W(t))}{\partial W(t)}dW(t) \\ &+ \frac{\partial f(t, W(t))}{\partial t}dt + \frac{1}{2}[\dots]. \end{aligned}$$

Interpret $dW(t)dW(t)$ as dt , $dW(t)dt = 0 = (dt)^2$, we get

$$\begin{aligned} f(t + \Delta t, W(t + \Delta t)) &= f(t, W(t)) + \frac{\partial f(t, W(t))}{\partial W(t)}dW(t) \\ &+ \left[\frac{\partial f(t, W(t))}{\partial t} + \frac{1}{2} \frac{\partial^2 f(t, W(t))}{\partial (W(t))^2} \right] dt. \end{aligned}$$

Ito-Doebelin Formula

$$\begin{aligned}
f(t + \Delta t, W(t + \Delta t)) &= f(t, W(t)) + \frac{\partial f(t, W(t))}{\partial W(t)} dW(t) \\
&\quad + \left[\frac{\partial f(t, W(t))}{\partial t} + \frac{1}{2} \frac{\partial^2 f(t, W(t))}{\partial (W(t))^2} \right] dt.
\end{aligned}$$

In general, Ito process $\{X(t), t \geq 0\}$ where $X(t)$ is given by

$$X(t) = X(0) + \int_0^t A_{(1)}(s) ds + \int_0^t A_{(2)}(s) dW(s)$$

where $A_{(1)}(t)$ and $A_{(2)}(t)$ are stochastic processes which are adapted to filtration $\{\mathcal{F}(t), t \geq 0\}$ and $A_{(2)}(t)$ is mean square integrable.

$$\begin{aligned}
df(t, X(t)) &= \frac{\partial f(t, X(t))}{\partial X(t)} dX(t) + \frac{\partial f(t, X(t))}{\partial t} dt \\
&\quad + \frac{1}{2} \frac{\partial^2 f(t, X(t))}{\partial (X(t))^2} dX(t) X(t).
\end{aligned}$$

Ito Formula Version 2

Assume that $X(t) = f(t, W(t))$. Assume that $f(t, x)$ has continuous partial derivatives at least second order.

$$\begin{aligned}
f(t + \Delta t, W(t + \Delta t)) &= f(t, W(t)) + \frac{\partial f(t, W(t))}{\partial W(t)} dW(t) \\
&\quad + \left[\frac{\partial f(t, W(t))}{\partial t} + \frac{1}{2} \frac{\partial^2 f(t, W(t))}{\partial (W(t))^2} \right] dt.
\end{aligned}$$

EXAMPLE 3. Consider the SDE of the form

$$dX(t) = X(t) dW(t), \text{ with } X(0) = 1.$$

Use the Ito-Doebelin formula

$$\begin{aligned}
df(t, X(t)) &= (f'_t(t, X(t)) + f'_x(t, X(t))) dt \\
&\quad + \frac{1}{2} f''_{xx}(t, X(t)) dt + f'_x(t, x(t)) dW(t)
\end{aligned}$$

with

$$\begin{aligned} f(t, x) &= e^x e^{\frac{-t}{2}}; f'_t(t, x) = \frac{-1}{2} e^x e^{\frac{-t}{2}} \\ f'_x(t, x) &= e^x e^{\frac{-t}{2}}; f''_{xx}(t, x) = e^x e^{\frac{-t}{2}}. \end{aligned}$$

$$\begin{aligned} dX(t) &= df(t, W(t)) \\ &= (f'_t(t, W(t)) + \frac{1}{2} f''_{xx}(t, W(t)))dt + f'_x(t, W(t))dW(t) \\ &= \left(\frac{-1}{2} X(t) + \frac{1}{2} X(t) \right) dt + X(t)dW(t) \\ &= X(t)dW(t). \end{aligned}$$

Hence, the solution is

$$X(t) = e^{W(t)} e^{-\frac{t}{2}}$$

is an Ito process. Also, it is a martingale.

EXAMPLE 4. Let

$$dR(t) = (\alpha - \beta R(t))dt + \sigma \sqrt{R(t)}dW(t)$$

where α, β and σ are positive constants. Closed form solution is not possible.

REMARK 5. It is used to compute stochastic integral for some special cases. Extension into jump process $\{X(t), t \geq 0\}$ where

$$X(t) = X(0) + I(t) + R(t) + J(t)$$

where

$$I(t) = \int_0^t \Gamma(s)dW(s); \quad R(t) = \int_0^t \Theta(s)ds$$

($\Gamma(s)$ and $\Theta(s)$ are adapted processes)

$J(t)$ right-continous pure jump term

$$\begin{aligned} f(X(t)) &= f(X(0)) + \int_0^t f'(X(s))dX^c(s) \\ &\quad + \frac{1}{2} \int_0^t f''(X(s))dX^c(s)dX^c(s) \\ &\quad + \sum_{0 < s \leq t} [f(X(s)) - f(X(s-))] \end{aligned}$$

Extension into multidimensional stochastic processes also.

Lecture 7 SDEs and their Applications in Finance

Geometric Brownian Motion

Let $S(t)$ be the stock price at time t . Let μ be the constant growth rate of the stock and σ be the volatility. Consider the SDE

$$dS(t) = \mu S(t)dt + \sigma S(t)dW(t), \quad S(0) = S_0$$

where constant growth rate of the stock $\mu (-\infty, \infty)$ and volatility $\sigma > 0$ are constants. Assume that $S(t) = f(t, W(t))$. Using Ito's formula (version 2)

$$\begin{aligned} dS(t) &= f_t dt + f_x dW(t) + \frac{1}{2} f_{xx} dt \\ f_x &= \sigma f \text{ and } f_t + \frac{1}{2} f_{xx} = \mu f \\ f &= e^{\sigma x} k(t), f_{xx} = \sigma^2 f, f_t = k'(t) e^{\sigma x} \\ k'(t) e^{\sigma x} &= \left(\mu - \frac{\sigma^2}{2} \right) f \end{aligned}$$

$$k'(t) e^{\sigma x} = \left(\mu - \frac{\sigma^2}{2} \right) e^{\sigma x} k(t)$$

$$k(t) = S_0 e^{(\mu - \frac{\sigma^2}{2})t}.$$

Hence

$$S(t) = S_0 e^{(\mu - \frac{\sigma^2}{2})t + \sigma W(t)}$$

One can prove that $S(t)$ is lognormal distribution. Also,

1. $E(S(t)) = E(S_0) e^{\mu t}$.
2. $E((S(t))^2) = E(S_0^2) \exp((2\mu + \sigma^2)t)$.

Asymptotic behavior of the solution.

1. If $\mu > \frac{\sigma^2}{2}$ then $S(t) \rightarrow \infty$ as $t \rightarrow \infty$ almost surely.
2. If $\mu < \frac{\sigma^2}{2}$ then $S(t) \rightarrow 0$ as $t \rightarrow \infty$ almost surely.

3. If $\mu = \frac{\sigma^2}{2}$ then $S(t)$ will fluctuate between arbitrary large and arbitrary small value as $t \rightarrow \infty$.

Let the interest rate r and the volatility $\sigma > 0$ be constant. Let $S(t)$ be the stock price at any time t . Let

$$S(t) = S(0)e^{(r - \frac{1}{2}\sigma^2)t + \sigma W(t)}$$

be a geometric Brownian motion with mean rate of return r , where the initial stock price $S(0)$ is positive and $W(t)$ is a Brownian motion. Let K be a positive constant which is strike price at maturity T .

One can show that, for $T > 0$,

$$\begin{aligned} E [e^{-rT}(S(T) - K)^+] &= S(0)\Phi(d_+(T, S(0))) \\ &\quad - Ke^{-rT}\Phi(d_-(T, S(0))) \end{aligned}$$

where $X^+ = \text{Max}(X, 0)$,

$$d_{\pm}(T, S(0)) = \frac{1}{\sigma\sqrt{T}} \left[\log \frac{S(0)}{K} + \left(r \pm \frac{\sigma^2}{2} \right) T \right],$$

and Φ is the cumulative standard normal distribution function

$$\Phi(y) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^y e^{-\frac{1}{2}z^2} dz = \frac{1}{\sqrt{2\pi}} \int_{-y}^{\infty} e^{-\frac{1}{2}z^2} dz.$$

Ornstein-Uhlenbeck Process

Consider the SDE

$$dX(t) = -\mu X(t)dt + \sigma dW(t) \quad X(0) = x_0 \quad (1)$$

where μ and σ are constants. This equation is often referred to as Langevin equation. This equation is useful to model the velocity at time t of a free particle that performs a Brownian motion.

If we set $dt = 1$ then

$$\begin{aligned} X(t+1) - X(t) &= -\mu X(t) + \sigma(W(t+1) - W(t)) \\ X(t+1) &= -\mu' X(t) + \sigma(W(t+1) - W(t)) \end{aligned}$$

where $\mu' = 1 + \mu$. This time series model can be considered as a discrete analogue at the solution of the Langerin equation (1). The stochastic differential equation (1) is interpreted as the following stochastic integral equation

$$X(t) = x_0 + \sigma W(t) - \mu \int_0^t X(s) ds.$$

Using the integration method and integration by parts, we can derive the solution $X(t)$ for each sample path of $W(t)$. The computation is somewhat complicated. Here, we derive the solution using Ito's formula which is easy. Let $f(t, x) = e^{\mu t} x$. Then,

$$\frac{\partial f}{\partial t} = \mu e^{\mu t} x, \frac{\partial f}{\partial x} = e^{\mu t}, \frac{\partial^2 f}{\partial x^2} = 0.$$

Use Ito's formula (Version 2), we get

$$X(t) = e^{-\mu t} X(0) + \sigma \int_0^t e^{-\mu(t-s)} dW(s). \quad (2)$$

This stochastic process $\{X(t), t \geq 0\}$ is called Ornstein-Uhlenbeck process. This is strong solution (immaterial of the path of Brownian motion, the solution (2) is given) to (1). If $X(0)$ is constant or if $X(0) \sim$ Normal distribution. Then $X(t)$ is also normally distributed.

If $X(0) = 0$, then

$$E(X(t)) = 0; \text{Var}(X(t)) = \frac{\sigma^2}{2\mu}(e^{-2\mu t} - 1)$$

and

$$\text{Cov}(X(t), X(s)) = \frac{\sigma^2}{2\mu}(e^{-\mu(t+s)} - e^{-\mu(t-s)}), s < t.$$

Since $\{X(t), t \geq 0\}$ is a zero-mean Gaussian process, the covariance function is characteristic for the Ornstein-Uhlenbeck process.

The $\{X(t), t \geq 0\}$ is a Markov process with Gaussian transition probability densities

$$p(t; x, y) = \frac{1}{\sqrt{2\pi}\sigma(t)} \exp \left\{ \frac{-(y - e^{-\mu t}x)^2}{2\sigma^2(t)} \right\}$$

when $\mu > 0$, which is the stable case,

$$\lim_{t \rightarrow \infty} \sigma^2(t) = \frac{\sigma^2}{2\mu} = \theta$$

and

$$\lim_{t \rightarrow \infty} p(t; x, y) = \frac{1}{\sqrt{2\pi\theta}} \exp \left\{ \frac{-y^2}{2\theta} \right\}.$$

Vasicek Interest Rate Model

One of the oldest and commonly used interest rate model. It is used to model short rate (instantaneous rate) (i.e., rate model is valid for short period). $R(t)$ - short rate stochastic process at time t , $t \geq 0$. The model assumes that $R(t)$ follows the SDE

$$dR(t) = (\alpha - \beta R(t))dt + \sigma dW(t)$$

where $\alpha > 0, \beta > 0, \sigma > 0$.

σ is the order of magnitude of the fluctuation of $R(t)$ around α . σ is the volatility. Assume that Lipschitz condition is satisfied. Start with

$$\begin{aligned} d(e^{\beta t} R(t)) &= e^{\beta t} dR(t) + \beta e^{\beta t} R(t) dt \\ &= e^{\beta t} (\alpha - \beta R(t)) dt + e^{\beta t} \sigma dW(t) + \beta e^{\beta t} R(t) dt \\ e^{\beta t} R(t) &= R(0) + \alpha \int_0^t e^{\beta s} ds + \sigma \int_0^t e^{\beta s} dW(s). \end{aligned}$$

The strong solution is

$$R(t) = R(0)e^{-\beta t} + \frac{\alpha}{\beta}(1 - e^{-\beta t}) + \sigma e^{-\beta t} \int_0^t e^{\beta s} dW(s).$$

If $R(0)$ is constant, then $R(t)$ is Gaussian distributed with

$$\begin{aligned} E(R(t)) &= R(0)e^{-\beta t} + \frac{\alpha}{\beta}(1 - e^{-\beta t}) \\ \text{Var}(R(t)) &= \frac{\sigma^2}{2\beta}(1 - e^{-2\beta t}). \end{aligned}$$

As $t \rightarrow \infty$

$$R(t) \sim \mathcal{N}\left(\frac{\alpha}{\beta}, \frac{\sigma^2}{2\beta}\right)$$

If $R(t) \gg \frac{\alpha}{\beta}$ then the process will bring it down back to $\frac{\alpha}{\beta}$ and if $R(t) \ll \frac{\alpha}{\beta}$ then also the process will bring it up to $\frac{\alpha}{\beta}$. Hence, this model is also known as mean reverting Ornstein-Uhlenbeck process.

- If $R(t) = \frac{\alpha}{\beta}$, then drift is zero.
- If $R(t) < \frac{\alpha}{\beta}$, then drift is greater than zero. Implies the process will try to pull back.
- If $R(t) > \frac{\alpha}{\beta}$, then drift is less than zero. Implies the process will try to pull down.

Hence, interest rate does not go too high or too low. This is applicable from economic view point too. Drawback of this model is that, it assumes negative values with positive probability which is undesirable property.

Cox-Ingersoll-Ross Model

$$dR(t) = (\alpha - \beta R(t))dt + \sigma\sqrt{R(t)}dW(t).$$

In this model, volatility is also dependent on the interest rate unlike in Vasicek interest rate model where it was constant. This model is also mean reverting model. As soon as $R(t)$ tends to zero, the model becomes

$$dR(t) = \alpha dt.$$

This implies that-

$$R(t) = R(0) + \alpha t.$$

The process will make sure that $R(t)$ linearly increases as $R(0) > 0$ and $\alpha t > 0$. The model does not have a closed form solution. Only numerical schemes can be used to solve the CIR model. $E(R(t))$ can be obtained as follows:

$$d(e^{\beta t} R(t)) = \beta e^{\beta t} R(t) + e^{\beta t} R(t) + 0dR(t)dR(t).$$

$$\begin{aligned} e^{\beta t} R(t) &= R(0) + \int_0^t \beta e^{\beta s} R(s) ds \\ &\quad + \int_0^t e^{\beta s} ((\alpha - \beta R(s)) ds + \sigma \sqrt{R(s)} dW(s)) \\ &= R(0) + \int_0^t \alpha e^{\beta s} ds + \int_0^t \sigma \sqrt{R(s)} dW(s) \\ &= R(0) + \frac{\alpha}{\beta} (e^{\beta t} - 1) + \int_0^t \sigma \sqrt{R(s)} dW(s). \end{aligned}$$

We know that $\int_0^t \sigma \sqrt{R(s)} dW(s)$ is an Ito integral and is a martingale. Hence, expectation is zero.

$$E(e^{\beta t} R(t)) = R(0) + \frac{\alpha}{\beta} (e^{\beta t} - 1).$$

$$E(R(t)) = R(0)e^{-\beta t} + \frac{\alpha}{\beta} (1 - e^{-\beta t}).$$

Variance of $R(t)$ is given by-

$$Var(R(t)) = \frac{\sigma^2}{\beta} R(0)(e^{-\beta t} - e^{-2\beta t}) + \frac{\alpha \sigma^2}{2\beta^2} (1 - 2e^{-\beta t} + e^{-2\beta t}).$$

As $t \rightarrow \infty$, we have

$$E(R(t)) = \frac{\alpha}{\beta}; \quad Var(R(t)) = \frac{\alpha \sigma^2}{2\beta^2}.$$

If we choose α and β appropriately, we can bring down variance and hence bringing more stability to the model. The bond pricing equation can be solved explicitly.

References

1. S Karlin and H M Taylor, "A First Course in Stochastic Processes", 2nd edition, Academic Press, New York, 1975.

2. J Medhi, “Stochastic Processes”, 3rd edition, New Age International Publishers, 2009.
3. S. E. Shreve “Stochastic Calculus for Finance”, Vol. I & Vol. II, Springer, 2004.
4. Thomas Mikosch “Elementary Stochastic Calculus with Finance in view”, World Scientific, 2006.
5. Suresh Chandra, S. Dharmaraja, Aparna Mehra, R. Khemchandani, “Financial Mathematics: An Introduction”, Narosa Publication House, 2012.

Module 7: Renewal Processes

Lecture 1 Renewal Function and Renewal Equation

Introduction

Renewal theory is the branch of probability theory that generalizes Poisson processes for arbitrary inter-arrival (holding) times. Let X_1 be the time to the first renewal and let $X_n (n = 2, 3, \dots)$ be the time between $(n - 1)$ -th renewal and n -th renewal. Assume that $X_n (n = 1, 2, \dots)$ are i.i.d. non-negative random variables with distribution function F . Let

$$\mu = E(X_n) = \int_0^\infty x dF(x)$$

which we assume to be positive and finite.

DEFINITION 1. Define the time of the n -th renewal by

$$S_n = \sum_{i=1}^n X_i.$$

Let $N(t)$ be the number of renewals by time t so that

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

Then, the counting process $\{N(t), t \geq 0\}$ will be a renewal process. If for some n , $S_n = t$, then a renewal is said to occur at t ; S_n gives the time of the n th renewal and is called the n th renewal time.

In the above, we are assuming that 0 is a renewal time. Sometimes, more general processes can be considered where 0 is not a renewal time.

REMARK 2. 1. A counting process for which the inter-arrival times are i.i.d. with an arbitrary distribution is said to be a renewal process.

2. A Poisson process can be defined as a counting process for which the inter-arrival times are i.i.d. with an exponential distribution. A renewal process is more general counting process than Poisson process.

3. A continuous-time stochastic process in which the embedded jump chain (the discrete process registering what values the process takes) is a Markov chain,

and where the holding times (time between jumps) are random variables with any distribution, whose distribution function may depend on the two states between which the move is made, we say it is called a semi-Markov process or Markov renewal process.

4. When $P(X = c) = 1$ for some $c > 0$, $\{S_n = nc, n \geq 1\}$ is called a deterministic renewal process.

Distribution

Consider a renewal process $\{N(t), t \geq 0\}$ with inter-arrival time distribution F . Note that, the event $\{N(t) \geq n\}$ is equivalent to the event $\{S_n \leq t\}$. Hence,

$$P\{N(t) \geq n\} = P(S_n \leq t) = F^{(n)}(t)$$

where $F^{(n)}$ is n -fold convolution of F with $F^{(0)} = 1$.

$$\begin{aligned} P(N(t) = n) &= P(N(t) \geq n) - P(N(t) \geq n+1) \\ &= F^{(n)}(t) - F^{(n+1)}(t) \end{aligned}$$

For example, when inter-arrival time is uniform distribution between 0 and 1, the n -fold convolution of F is given by

$$F^{(n)}(t) = \frac{t^n}{n!}, \quad n = 1, 2, \dots; \quad 0 \leq t \leq 1$$

which can be shown from the mathematical induction.

$$\begin{aligned} F^{(n+1)}(t) &= P(S_{n+1} \leq t) \\ &= \int_0^t P(S_{n+1} \leq t \mid S_n = x) \frac{x^{n-1}}{(n-1)!} dx \\ &= \int_0^t P(X_{n+1} \leq t-x) \frac{x^{n-1}}{(n-1)!} dx \\ &= \frac{t^{n+1}}{(n+1)!}, \quad 0 \leq t \leq 1. \end{aligned}$$

EXAMPLE 3. Poisson process is a counting renewal process $\{N(t), t \geq 0\}$ whose inter-arrival time has a exponential distribution with parameter λ , i.e., $X \sim \text{Exp}(\lambda)$.

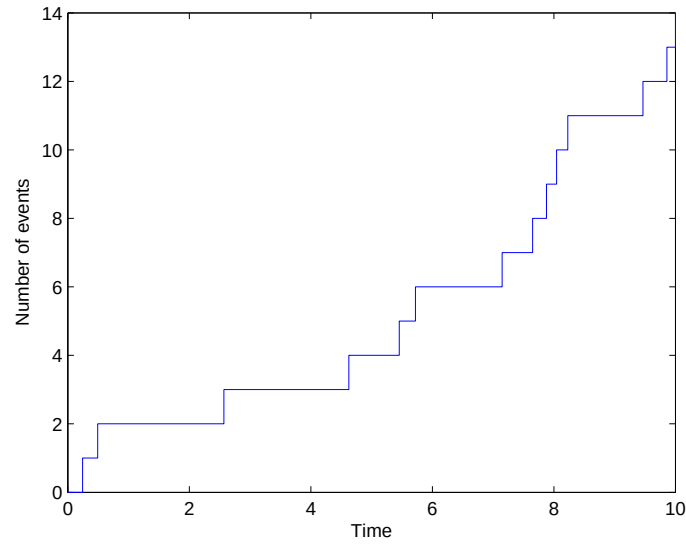


Figure 7.1: Sample path of Poisson process with $\lambda = 2$

Distribution of $N(t)$

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

For arbitrary n ,

$$P(X_n > x) = P(N(x) = 0 \mid N(t_{n-1}) = n - 1) = e^{-\lambda x}, x \geq 0$$

Hence, the distribution of X is $F(x) = 1 - e^{-\lambda x}, x \geq 0$.

Sample path of Poisson process with $\lambda = 2$ is shown in Figure 7.1.

Suppose that in a system, an unit fails, according with a Poisson process with rate $\lambda = 3$ per day. Suppose that there are 6 spare units in an inventory and the next supply is not due in 4 days. The probability that the system will be out of order in the next 4 days is

$$\sum_{n=7}^{\infty} P(N(4) = n) = 1 - \sum_{n=0}^6 P(N(4) = n) = 0.954 .$$

In this module, we discuss various stochastic processes as shown in Figure 7.2.

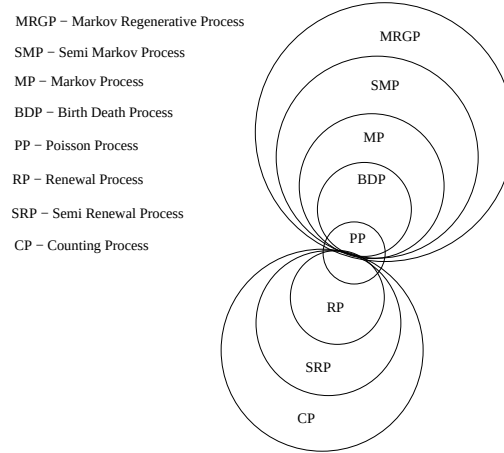


Figure 7.2: Various Stochastic Processes

Renewal Function

Let

$$M(t) = E[N(t)] = \sum_{n=1}^{\infty} P(N(t) \geq n) = \sum_{n=1}^{\infty} F^{(n)}(t)$$

Then $M(t)$ is called a renewal function. The function $m(t) = M'(t)$ is called the renewal density function of the renewal process.

$$\begin{aligned} m(t) &= \lim_{h \rightarrow 0^+} \frac{P(\text{one or more renewals in } (t, t+h))}{h} \\ &= \lim_{h \rightarrow 0^+} \frac{1}{h} \sum_{n=1}^{\infty} P(n \text{ renewals in } (t, t+h)) \\ &= \sum_{n=1}^{\infty} \lim_{h \rightarrow 0^+} \frac{1}{h} [F^{(n)}(t+h) - F^{(n)}(t)] = M'(t). \end{aligned}$$

For example, the renewal process whose inter-arrival time follows i.i.d. with uniform distribution between 0 and 1, the renewal function for $0 \leq t \leq 1$ is given by

$$M(t) = \sum_{n=1}^{\infty} F^{(n)}(t) = \sum_{n=1}^{\infty} \frac{t^n}{n!} = e^t - 1.$$

$$M(t) = \sum_{n=1}^{\infty} F^{(n)}(t)$$

In equivalent form using the Laplace transform,

$$M^*(s) = \sum_{n=1}^{\infty} F^{(n)*}(s) = \sum_{n=1}^{\infty} \frac{f^{(n)*}(s)}{s}$$

where $f^{(n)}(t) = F^{(n)'}(t)$.

$$M^*(s) = \frac{1}{s} \sum_{n=1}^{\infty} [f^*(s)]^n = \frac{f^*(s)}{s[1 - f^*(s)]}.$$

Hence,

$$m^*(s) = \frac{f^*(s)}{1 - f^*(s)} \text{ and } f^*(s) = \frac{m^*(s)}{1 + m^*(s)}.$$

Renewal Equation

It can be shown that the renewal function $M(t)$, $0 \leq t < \infty$, uniquely determine the inter-arrival time distribution F . For example, $M(t) = \lambda t$ corresponds to the exponential distribution with parameter λ . Renewal equations are useful for deriving the quantity of interest associated with a renewal process at a function of time. A renewal equation is expressed by a recursive form through an integral equation. We know that

$$M(t) = E[N(t)].$$

This can be evaluated by conditioning on X_1 , the time of first renewal, i.e.,

$$M(t) = \int_0^{\infty} E[N(t) \mid X_1 = x] dF(x).$$

This integral can be evaluated by dividing into two cases: one is the case where the first renewal occurs after time t and the other is the case where the first renewal occurs before time t . In the former case, since there are no renewals observed by t ,

$$E[N(t) \mid X_1 = x > t] = 0.$$

In the later case, the first renewal occurs before time t and the expected number of renewals between x and t will be $M(t - x)$ from the definition of the renewal function. Hence,

$$E[N(t) \mid X_1 = x < t] = 1 + M(t - x).$$

Hence,

$$M(t) = \int_0^t (1 + M(t-x))dF(x) = F(t) + \int_0^t M(t-x)dF(x).$$

This integral equation is called a renewal equation if $M(\cdot)$ is considered as unknown. The above equation can be written as:

$$M = F + M * f.$$

Now, taking Laplace transform on both sides,

$$M^* = F^* + (M * f)^* = F^* + M^* f^*.$$

Substituting $F^* = \frac{f^*(s)}{s}$ and simplifying, we get

$$f^*(s) = \frac{sM^*(s)}{1 + sM^*(s)}.$$

The renewal equation can be generalized for $Z(t)$, the unknown function associated with a renewal process with distribution function F ,

$$Z(t) = Q(t) + \int_0^t Z(t-x)dF(x)$$

where $Q(t)$ is a known function.

Renewal Times

Let $X_1, X_2, \dots$ be the time between its successive occurrences. Then

$$S_0 = 0; S_{n+1} = S_n + X_{n+1}, n = 1, 2, \dots$$

define the times of occurrence assuming that the time origin is taken to be an instant of such an occurrence. The sequence $\{S_n, n = 0, 1, \dots\}$ is called a renewal process provided that $X_1, X_2, \dots$ be i.i.d. non-negative random variables. Then the S_n , are called renewal times.

Note that, the renewal process $\{S_n, n = 0, 1, \dots\}$ is said to be recurrent if $X_n < \infty$ almost surely for every n ; otherwise is called transient.

Age, Excess and Spread at Time t

Suppose $\{N(t), t \geq 0\}$ is a renewal process.

1. **Age:** The age at t of the renewal process is defined by

$$A(t) = t - S_{N(t)}.$$

2. **Excess:** The excess at t of the renewal process is defined by

$$Y(t) = S_{N(t)+1} - t.$$

3. **Spread:** The spread at t of the renewal process is defined by

$$X_{N(t)+1} = A(t) + Y(t).$$

For example, let $Y(t)$ be the excess at time t . Let $g(t) = E[Y(t)]$. Using the above renewal equation, we can find $g(t)$. Conditioning on X_1 gives

$$g(t) = \int_0^\infty E[Y(t) \mid X_1 = x] dF(x).$$

Now,

$$E[Y(t) \mid X_1 = x > t] = x - t$$

while for $X_1 = x < t$

$$E[Y(t) \mid X_1 = x < t] = g(t - x).$$

Hence,

$$g(t) = \int_t^\infty (x - t) dF(x) + \int_0^t g(t - x) dF(x)$$

is the renewal equation for $g(t)$.

Lecture 2 Generalized Renewal Processes and Renewal Limit Theorems

Reward Renewal Process

Let X_1 be the time to the first renewal and let $\{X_n, n = 2, 3, \dots\}$ be the time between $(n - 1)$ th renewal and n -th renewal. Assume that $\{X_n, n = 1, 2, \dots\}$ are i.i.d. random variables with distribution function F . Let $\mu = E(X_n) = \int_0^\infty x dF(x)$ which will be positive. Let R_n be n th reward at the time of the n th renewal. Usually, R_n may depend on X_n . Let

$$R(t) = \sum_{n=1}^{N(t)} R_n$$

be reward earned by time t . Note that unlike the X_n , each R_n may take negative values as well as positive values.

EXAMPLE 1. Consider an age replacement model. In this model, a component that is used continuously with replacements. Let X be the lifetime of the component, which is random with distribution function F . The component is replaced by a new one upon failure or at a fixed time period T , whichever comes first. This replacement policy is called an age replacement. The cost of a new component c_1 is and the additional cost incurred by a failure is c_2 . Our interest is to find the long-run average cost. Let $N(t)$ be the number of replacements of components by t . Let $R(t)$ be the amount of cost incurred by t , then $\{R(t), t \geq 0\}$ becomes a reward renewal process. The long-run average cost can be expressed as

$$\text{Long-run average cost} = \frac{E(R)}{E(X)} = \frac{E[\text{cost during a cycle}]}{E[\text{cycle length}]}$$

Now,

$$E[\text{cycle length}] = E[\min(X, T)] = \int_0^T (1 - F(x)) dx.$$

Given that, the reward or cost during a cycle, R , as

$$R = \begin{cases} c_1, & X > T \\ c_1 + c_2, & X \leq T \end{cases}.$$

Hence, the expected value is

$$E(R) = c_1 P(X > T) + (c_1 + c_2) P(X \leq T) = c_1 + c_2 F(T).$$

The required long-run average cost is given by

$$\begin{aligned} \frac{E(R)}{E(X)} &= \frac{E[\text{cost during a cycle}]}{E[\text{cycle length}]} \\ &= \frac{c_1 + c_2 F(T)}{\int_0^T (1 - F(x)) dx}. \end{aligned}$$

Markov Reward Model

An Markov Reward Model (MRM) is a labeled continuous time Markov chain (CTMC) augmented with state reward and impulse reward structures. The state reward structure is a function r that assigns to each state $s \in S$ a reward r_s such that if t time-units are spent in state s , a reward of $r_s t$ is acquired. The rewards that are defined in the state reward structure can be interpreted in various ways. They can be regarded as the gain or benefit acquired by staying in some state and they can also be regarded as the cost spent by staying in some state. This type of MRM is called a rate-based MRM. The impulse reward structure, on the other hand, is a function τ that assigns to each transition from s to s' , where $s, s' \in S$ and the spending time in transition of s to s' is positive, a reward $\tau(s, s')$ such that if the transition from s to s' occurs, a reward of $\tau(s, s')$ is acquired. Similar to the state reward structure, the impulse reward structure can be interpreted in various ways. An impulse reward can be considered as the cost of taking a transition or the gain that is acquired by taking the transition.

MRMs are commonly used for the performance, dependability, and performability analysis of computer and communication systems. In general, the reward rate is assigned on the basis of desired measures. Let $Z(t) = r_{X(t)}$ be the instantaneous reward rate of the MRM at time t . Then the expected instantaneous reward rate at time t is given by:

$$E[Z(t)] = \sum_{i \in S} r_i P_i(t).$$

The expected reward rate in steady-state is given by:

$$E[Z] = \sum_{i \in S} r_i \pi_i.$$

The different types of measures namely steady-state measures, transient measures, cumulative measures and performability measures are supported by MRMs. For instance, the steady-state measures are computed from the steady-state behaviour of the Markov chain. Under the assumption that the steady-state distribution of $\{X(t), t \geq 0\}$ equals $\pi = [\pi_0, \pi_1, \dots]$, and can be expressed as

$$M = \sum_{i \in S} r_i \pi_i.$$

Similarly, cumulative measures, denoted by $Y(t)$, express the overall gain that is received from a system over some finite time interval. When transient measures are integrated over a time interval $[0, t]$, $Y(t)$ can be given as

$$Y(t) = \int_0^t r_{X(s)} ds.$$

Alternative Renewal Process

Let $X_1, X_2, \dots$ be i.i.d. random variables which constitutes *on times*. Let $Y_1, Y_2, \dots$ be i.i.d. random variables which constitutes *off times*. Assume that $E(X + Y) < \infty$ and $X + Y$ has distribution F . Suppose that, renewal occurs at the end of every X_i where as no renewals at the end of every Y_i . Assume that, X_i and Y_i are independent random variables. Then $\{X_i + Y_i, i = 1, 2, \dots\}$ is called an alternative renewal process.

EXAMPLE 2. *For example, consider the following situation. A machine works for time X_1 , then breaks down and has to be repaired (which takes time Y_1), then works for a time X_2 , then is down for a time Y_2 , and so on. If we suppose that the machine is as good as new after each repair, then this constitutes an alternative renewal process.*

Delayed Renewal process

It is not always reasonable to insist that the first renewal occurs at time $S_0 = 0$. For instance, in applications where the occurrences are at times when a component

of a system must be replaced, one might well be interested in situations where there is already a working component in place at time 0. For this reason, we define a delayed renewal process to be a sequence $S_0, S_1, S_2, \dots$ where

$$S_n = S_0 + \sum_{i=1}^n X_i \quad \text{and}$$

- (i) the inter-arrival times $X_1, X_2, \dots$ are positive and i.i.d., as in an ordinary renewal process, and
- (ii) the initial delay $S_0 \geq 0$ is independent of the inter-arrival times X_i .

Notice that the distribution of the initial delay random variable S_0 is not required to be the same as that of the inter-arrival time random variables X_i . Hence, a delayed renewal process is a renewal process in which the first arrival time, $X_1 = t_1$, independently, is allowed to have a different distribution $P(X_1 \leq x) = F_1(x); x \geq 0$, than F , the distribution of all the remaining i.i.d. inter-arrival times $\{X_n, n \geq 2\}$. t_1 is then called the delay. When there is no such delay, that is, when $X_1 \sim F$ as usual, the renewal process is said to be a non-delayed version.

As $n \rightarrow \infty$,

$$Z_n = \frac{S_n - n\mu}{\sigma\sqrt{n}} \Rightarrow Z \sim \mathcal{N}(0, 1).$$

As $t \rightarrow \infty$, $N(t)$ becomes normal distributed and both $E(N(t)) \sim \frac{t}{\mu}$ and $Var(N(t)) \sim \frac{\sigma^2 t}{\mu^3}$.

As $t \rightarrow \infty$, using central limit theorem, we have

$$Z(t) = \frac{N(t) - \frac{t}{\mu}}{\sigma\sqrt{\frac{t}{\mu^3}}} \Rightarrow Z \sim \mathcal{N}(0, 1).$$

Long-run Properties

One can study the long-run properties of the renewal process. There are two types of the long-run properties. One is to obtain the long-run (time) average of the quantity of interest and the other is to obtain the pointwise limit. For example, the

long-run average of age is given by

$$\lim_{t \rightarrow \infty} \frac{\int_0^t A(s) ds}{t}$$

while the pointwise limit of the expected age is given by

$$\lim_{t \rightarrow \infty} E[A(t)].$$

Long-run Renewal Rate

One can study the average number of renewals (per unit time) in the long-run. It is called a long-run renewal rate. For a renewal process $\{N(t), t \geq 0\}$ having distribution function F for inter-arrival times, the long-run renewal rate is given by

$$\lim_{t \rightarrow \infty} \frac{N(t)}{t} = \frac{1}{\mu} \text{ with probability } 1$$

where

$$\mu = \int_0^\infty x dF(x).$$

Since $S_{N(t)}$ is the last renewal time prior to t and $S_{N(t)+1}$ is the first renewal time after t .

We know that

$$S_{N(t)} \leq t \leq S_{N(t)+1} \text{ or } \frac{S_{N(t)}}{N(t)} \leq \frac{t}{N(t)} \leq \frac{S_{N(t)+1}}{N(t)}.$$

But,

$$\begin{aligned} \lim_{t \rightarrow \infty} \frac{S_{N(t)}}{N(t)} &= \lim_{t \rightarrow \infty} \frac{X_1 + X_2 + \dots + X_{N(t)}}{N(t)} \\ &= E[X] = \mu \text{ with probability } 1 \end{aligned}$$

and

$$\frac{S_{N(t)+1}}{N(t)} = \lim_{t \rightarrow \infty} \frac{S_{N(t)+1}}{N(t)+1} \frac{N(t)+1}{N(t)} = \mu \text{ with probability } 1.$$

Hence,

$$\lim_{t \rightarrow \infty} \frac{N(t)}{t} = \frac{1}{\mu} \text{ with probability } 1.$$

Renewal Limit Theorems

From the earlier lectures, we note that the arrival time process and the renewal or counting process are related. That is, the arrival time process is the partial sum process for a sequence of independent, identically distributed random variables of inter-arrival times. Hence, it is reasonable that the fundamental limit theorems (such as the law of large numbers and the central limit theorem) for partial sum processes should have analogs for the renewal process. The purpose of this lecture is to explore the limiting behavior of renewal process. The main results that we will study, known as renewal limit theorems, are important for other stochastic processes.

Elementary Renewal Theorem

It states that the basic limit in the law of large numbers in mean as well as with probability 1. For a renewal process $\{N(t), t \geq 0\}$ having mean inter-arrival time $0 < \mu < \infty$,

$$\lim_{t \rightarrow \infty} \frac{E[N(t)]}{t} = \lim_{t \rightarrow \infty} \frac{M(t)}{t} = \frac{1}{\mu}.$$

That is, the limiting mean average rate of arrivals is $\frac{1}{\mu}$.

Note that the limit value of a sequence of random variables is not same as the limit value of a sequence of mean of random variables. Consider the reward renewal processes. Let

$$R(t) = \sum_{n=1}^{N(t)} R_n$$

be reward earned by time t . If $E(R) < \infty$, $E(X) < \infty$, then

$$\frac{E(R(t))}{t} \rightarrow \frac{E(R)}{E(X)} \text{ as } t \rightarrow \infty.$$

Asymptotic property:

$$\frac{R(t)}{t} \rightarrow \frac{E(R)}{E(X)} \text{ as } t \rightarrow \infty, \text{ with probability 1.}$$

Lattice

A nonnegative random variable X is said to be lattice if there exists $d \geq 0$ such that

$$\sum_{n=0}^{\infty} P(X = nd) = 1.$$

The largest d having this property is said to be the period of X . If X is lattice and X has a distribution function of F , then we say that F is lattice. For instance, if X follows Poisson distribution with mean λ , then X is lattice with period 1. If $P(X = \sqrt{2}) = P(X = \sqrt{3}) = 0.5$, then X is not lattice.

Key Renewal Theorem

Let $\{N(t), t \geq 0\}$ be a renewal process with continuous inter-arrival time non lattice distribution function F with mean $\mu < \infty$. If $h : [0, \infty) \rightarrow [0, \infty)$ is integrable and non-increasing, as $t \rightarrow \infty$

$$\int_0^t h(t-x) dM(x) \rightarrow \frac{1}{\mu} \int_0^{\infty} h(x) dx$$

where $M(t)$ is the renewal function.

EXAMPLE 3. Let $Y(t)$ be the excess at t . We know that

$$E[Y(t)] = h(t) + \int_0^t h(t-x) dM(x)$$

where

$$h(t) = \int_t^{\infty} (x-t) dF(x)$$

$$\lim_{t \rightarrow \infty} E[Y(t)] = \lim_{t \rightarrow \infty} \int_0^t h(t-x) dM(x).$$

Assume that, the second moment of an inter-arrival time is finite, then $h(t)$ is integrable and non-increasing. Assume that, F is not lattice.

Now, we can apply Key Renewal Theorem to obtain the limiting mean excess. After simplifications, we get

$$\lim_{t \rightarrow \infty} E[Y(t)] = \frac{E[X^2]}{2E[X]}.$$

Blackwell's Theorem

Let $M(t) = E[N(t)]$ be the renewal function. It states that, asymptotically, the expected number of renewals in an interval is proportional to the length of the interval; the proportionally constant is $\frac{1}{\mu}$. If F is not lattice, then

$$\lim_{t \rightarrow \infty} [M(t+h) - M(t)] = \frac{h}{\mu}.$$

If F is lattice with period d , then

$$\lim_{t \rightarrow \infty} E[\text{number of renewals at } nd] = \frac{d}{\mu}$$

since the number of renewals at time nd will be 1 or 0.

Lecture 3 Markov Renewal and Markov Regenerative Processes

Consider a system $\{Y(t), t \geq 0\}$ with state space $\Omega = \{1, 2, \dots\}$. Suppose that the system is initially in the state X_0 at time t_0 . It stays there for a non-negative random amount of time (which may follow a general distribution), and then, the system jumps to state X_1 (which could be the same as X_0) at the next transition time instant t_1 . It stays there for a non-negative random amount of time and then jumps to state X_2 at next transition time instant t_2 , and the process continues like this. Thus, X_n is the n th state visited by the system at the time instant of the n th transition t_n .

Semi-Markov Process

If $Y(t)$ denotes the state of the system at time t , then $Y(t_n) = X_n$ for $n = \{0, 1, \dots\}$. If the Markov property is satisfied at all the transition time instants t_n , $n = \{0, 1, \dots\}$, then the stochastic process $\{Y(t), t \geq 0\}$ described above is called an semi-Markov process (SMP). That is, the evolution of the system after the time instant $t = t_n$ depends only on the past history of the system till time t_n .

In other words, if $\{Y(t), t \geq 0\}$ is an SMP, then the process $\{Y(t+t_n), t \geq 0\}$ is independent of $\{Y(t), 0 \leq t \leq t_n\}$ given the complete past history $\{Y(t), 0 \leq t \leq t_n\}$ and $X_n = i$, $i \in \Omega$, and is identical to the process $\{Y(t), t \geq 0\}$ given $X_0 = i$ because of time homogeneity.

Thus, in an SMP, the Markov property is satisfied only at each of the transition epochs $\{t_n\}$, and not at all times. The fact that Markov property holds at each of the transition epochs t_n , $n = \{0, 1, \dots\}$ of the stochastic process $\{Y(t), t \geq 0\}$ entails that the Markov property also holds for $\{X_n, n = 0, 1, \dots\}$. Hence, $\{X_n, n = 0, 1, \dots\}$ turns out to be a time-homogeneous embedded Markov chain with the state space Ω . Then, we say $\{(X_n, t_n), n = 1, 2, \dots\}$ constitutes Markov renewal process or semi-Markov process with state space Ω .

The process $\{Y(t), t \geq 0\}$ is not a Markov process, although it inherits some important properties of Markov processes. The associated process $\{X_n, n = 1, 2, \dots\}$ is a Markov process. Hence, the name to $\{(X_n, t_n), n = 1, 2, \dots\}$ as a semi-Markov process.

In conclusion, a continuous-time stochastic process in which the embedded jump chain (the discrete process registering what values the process takes) is a Markov

chain, and where the holding times (time between jumps) are random variables with any distribution, whose distribution function may depend on the two states between which the move is made, we say it is called a semi-Markov process or Markov renewal process. Semi-Markov processes are non-Poissonian with a renewal property. This means that the probability of a jump from state i to state j at a certain time depends only on the states i, j and the time t since the last jump occurred. A semi-Markov process where all the holding times are exponentially distributed is called a continuous time Markov chain (CTMC).

Steady-state Measures

The analysis of an SMP is performed in two stages. In the first stage, the SMP stays in a state i , $i \in \Omega$ for a random amount of time. Let us consider that the time spent in state i follows a general distribution with distribution function $H_i(t)$ (sojourn time distribution). In the second stage, the SMP moves from state i to state j with probability $p_{i,j}$

$$(\text{i.e., } p_{i,j} = P\{X_{n+1} = j | X_n = i\}; \quad i, j \in \Omega).$$

The SMP can now be completely described by the vector of sojourn time distributions $H(t)$ and the transition probability matrix $P = [p_{i,j}]$.

To compute the steady-state probability vector $\pi = [\pi_1, \pi_2, \dots]$ of the SMP, we follow the following steps: first calculate the mean sojourn time $h_i = \int_0^\infty (1 - H_i(t)) dt$ for each state i ; next, find the steady-state probability vector $v = [v_1, v_2, \dots]$ for the EMC of the SMP by solving the following system of equations:

$$v = vP \text{ and } \sum_{i \in \Omega} v_i = 1.$$

Finally, compute the steady-state probabilities for the SMP as given below:

$$\pi_i = \frac{v_i h_i}{\sum_{j \in \Omega} v_j h_j}, \quad i \in \Omega. \quad (1)$$

EXAMPLE 1. Consider a stochastic process $\{Y(t), t \geq 0\}$ with state space $\Omega = \{1, 2, 3, 4\}$. $\{(X_n, t_n), n = 1, 2, \dots\}$ as a semi-Markov process where $X_n = Y(t_n), n =$

1, 2, \dots Assume that the time spent in states 1 and 2 follow exponential distributions with distribution function $H_1(t)$ and $H_2(t)$ (sojourn time distribution) respectively and are given by

$$H_1(t) = \begin{cases} 0, & t < 0 \\ 1 - e^{-2t}, & t \geq 0 \end{cases} ; \quad H_2(t) = \begin{cases} 0, & t < 0 \\ 1 - e^{-3t}, & t \geq 0 \end{cases}$$

Assume that the time spent in states 3 and 4 follow general distributions with distribution function $H_3(t)$ and $H_4(t)$ (sojourn time distribution) respectively and are given by

$$H_3(t) = \begin{cases} 0, & t < 1 \\ t - 1, & 1 \leq t < 2 \\ 1, & 2 \leq t < \infty \end{cases} ; \quad H_4(t) = \begin{cases} 0, & t < 2 \\ t - 2, & 2 \leq t < 3 \\ 1, & 3 \leq t < \infty \end{cases} .$$

Here, the SMP moves from state i to state j with probability $p_{i,j}$

$$(i.e., p_{i,j} = P\{X_{n+1} = j | X_n = i\}; \quad i, j \in \Omega).$$

The transition probability matrix $P = [p_{i,j}]$ is given by

$$P = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \end{bmatrix}$$

Solving

$$v = vP \text{ and } \sum_{i \in \Omega} v_i = 1$$

we get

$$v_1 = v_2 = v_3 = v_4 = \frac{1}{4}.$$

Here,

$$h_1 = \frac{1}{2}; h_2 = \frac{1}{3}; h_3 = \frac{3}{2}; h_4 = \frac{5}{2}.$$

Finally, compute the steady-state probabilities for the SMP as given below:

$$\pi_i = \frac{v_i h_i}{\sum_{j \in \Omega} v_j h_j}, i \in \Omega.$$

Markov Regenerative Process

Consider a stochastic process $\{Z(t), t \geq 0\}$ with state space Ω . Suppose that every time a certain phenomenon occurs, the future of the process Z after that time becomes a probabilistic replica of the future after time zero. Such times (usually random) are called regeneration times of Z , and the process Z is then said to be regenerative. Such process is called a regenerative process.

Markov Regenerative Theory

We consider a stochastic process wherein there exist time points where the process satisfies the memoryless property. These time points are referred to as *regeneration points*. In a Markov regenerative process (MRGP) the stochastic evolution between two successive regeneration points depends only on the state at regeneration, not on the evolution before regeneration. Furthermore, due to the time homogeneity of the embedded Markov renewal process, the evolution of the MRGP becomes a probabilistic replica after each regeneration. As a consequence, all memory other than the state must be reset at a regeneration point.

The concepts of MRGP are given in the next two definitions.

A sequence of bivariate random variables $\{(Y_n, S_n), n \geq 0\}$ is called a *Markov renewal sequence* if:

- (i) $S_0 = 0, S_{n+1} \geq S_n; Y_n \in \Omega'$ and
- (ii) for all $n \geq 0$,

$$\begin{aligned} & P\{Y_{n+1} = j, S_{n+1} - S_n \leq t \mid Y_n = i, S_n, Y_{n-1}, S_{n-1}, \dots, Y_0, S_0\} \\ &= P\{Y_{n+1} = j, S_{n+1} - S_n \leq t \mid Y_n = i\} \quad (\text{Markov Property}) \\ &= P\{Y_1 = j, S_1 \leq t \mid Y_0 = i\}. \quad (\text{Time Homogeneity}) \end{aligned} \tag{2}$$

Then, the MRGP is defined as follows:

A stochastic process $\{Z(t), t \geq 0\}$ on Ω is called an MRGP if there exists a Markov renewal sequence $\{(Y_n, S_n), n \geq 0\}$ of random variables such that all conditional finite dimensional distributions of $\{Z(S_n + t), t \geq 0\}$ given $\{Z(u), 0 \leq u \leq S_n, Y_n = i\}$ are the same as those of $\{Z(t), t \geq 0\}$ given $Y_0 = i, i \in \Omega' \subset \Omega$.

Note that the above definition implies that in this case $\{Z(S_n^+), n \geq 0\}$ or $\{Z(S_n^-), n \geq 0\}$ is an *embedded discrete time Markov chain* (DTMC) or just the *embedded Markov chain* (EMC) in $\{Z(t), t \geq 0\}$, and also that S_n is a stopping time (regeneration points) of $\{Z(t), t \geq 0\}$.

Global and Local Kernels

As a special case, the definition implies that

$$\begin{aligned} P\{Z(S_n + t) = j \mid Z(u), 0 \leq u \leq S_n, Y_n = i\} \\ = P\{Z(t) = j \mid Y_0 = i\}. \end{aligned}$$

We denote the conditional probability in equation (2) by $K_{i,j}(t)$, $i, j \in \Omega'$. The matrix $K(t) = [K_{i,j}(t)]$ is called the *global kernel* of the Markov renewal sequence. Define the matrix $E_{i,j}(t)$, $i \in \Omega', j \in \Omega$, as follows:

$$E_{i,j}(t) = P\{Z(t) = j, S_1 > t \mid Y_0 = i\}.$$

This matrix $E(t) = [E_{i,j}(t)]$ describes the behavior of the MRGP between two transition epochs of the EMC, i.e., over the time interval $[0, S_1)$. We call the matrix $E(t)$ the *local kernel*.

Limiting Distribution

We study the limiting behavior of the MRGP by taking the limit as t approaches infinity. We require two new variables to be defined, viz., the mean time $\alpha_{i,j}$ the MRGP spends in state j between two successive regeneration instants, given that it started in state i after the last regeneration:

$$\alpha_{i,j} = E[\text{time in } j \text{ during } (0, S_1) \mid Y_0 = i] = \int_0^\infty E_{i,j}(t) dt, \quad (3)$$

and the steady state probability vector $\nu = (\nu_k)$ of the Embedded Markov Chain (EMC):

$$\nu = \nu P, \quad \sum_{k \in \Omega'} \nu_k = 1, \quad (4)$$

where $P = K(\infty)$ is the one-step transition probability matrix of the EMC.

The following theorem describes the limiting behavior of MRGPs. Let $\{Z(t), t \geq 0\}$ be a MRGP on Ω with Markov renewal sequence $\{(Y_n, S_n), n \geq 0\}$ with kernel $K(\cdot)$. Let $N(t)$ denotes the total number of state changes by time t . i.e., $N(t) = \sup\{n \geq 0 : S_n \leq t\}$. Suppose that

- (i) the sample paths of $\{Z(t), t \geq 0\}$ are right continuous with left limits,
- (ii) the semi-Markov process $\{Y_{N(t)} \in \Omega' \subset \Omega, t \geq 0\}$ is irreducible, aperiodic, and positive recurrent
- (iii) $\nu = (\nu_k)$ is a positive solution to equation (4).

Steady-state Distribution

Then the steady state probability of the MRGP is given by

$$\pi_j = \lim_{t \rightarrow \infty} P\{Z(t) = j\} = \frac{\sum_{k \in \Omega'} \nu_k \alpha_{k,j}}{\sum_{k \in \Omega'} \nu_k \beta_k}$$

$$\text{where} \quad \beta_k = \sum_{l \in \Omega} \alpha_{k,l}.$$

Transient Measures

Now, let us discuss the time dependent behavior of the MRGP. Define $V(t) = [V_{ij}(t)]$ as the conditional state probabilities of the MRGP given as

$$V_{ij}(t) = P\{Z(t) = j \mid Q(0) = Y_0 = i\}; \quad i, j \in \Omega$$

where $\{Y_n, n \geq 0\}$ is the embedded Markov chain of the MRGP $\{Z(t), t \geq 0\}$. Also,

$$\begin{aligned} V(t) &= E(t) + \int_0^t dK(s)V(t-s) \\ &= E(t) + K(t) * V(t) \end{aligned}$$

The solution method is outlined below:

1. Calculate $K(t)$ and $E(t)$.
2. Compute $\tilde{K}(s)$ and $\tilde{E}(s)$ where $\tilde{K}(s)$ and $\tilde{E}(s)$ are the Laplace-Steiltjes transform obtained as

$$\tilde{K}(s) = \int_0^\infty e^{-st} dK(t) \text{ and } \tilde{E}(s) = \int_0^\infty e^{-st} dE(t)$$

3. Solve the following linear system for $\tilde{V}(s)$

$$[I - \tilde{K}(s)]\tilde{V}(s) = \tilde{E}(s)$$

where

$$\tilde{V}(s) = \int_0^\infty e^{-st} dV(t)$$

4. Invert $\tilde{V}(s)$ to obtain $V(t)$
5. Using $p(t)_{1 \times \Omega} = p(0)_{1 \times \Omega'} V(t)_{\Omega' \times \Omega}$, obtain the time dependent state probabilities of the MRGP.

Lecture 4 Non Markovian Queues

M/G/1 Queueing Model

In this queueing model, the inter-arrival times are exponential, but the service times are non-exponential (general) distribution. Assume that, the CDF of service time is $F(t)$ with mean $\frac{1}{\mu}$. Let $N(t)$ be number of customers in the system at time t . In this model, the evolution of system after an arrival of a customer depends not only on the number of the customers $N(t)$ but also on the remaining service time of the customer receiving service, if any.

Whereas the evolution of the system after a service completion depends only on the state of the system. Thus $\{N(t), t \geq 0\}$ is not a semi-Markov process. Suppose that the time origin is taken to be an instant of a departure which left behind exactly j customers. Then, every time a departure occurs leaving behind j customers, the future of $\{N(t)\}$ after such time has exactly the same probability law as the process $\{N(t), t \geq 0\}$ had starting at time 0. One can observe that, $\{N(t), t \geq 0\}$ is a Markov regenerative process (MRGP).

Let n th customer depart at time point t_n . Let X_n be the number of customers in the system just after the departure instant of n th customer leaves. Then $\{(X_n, t_n), n = 0, 1, \dots\}$ is a Markov renewal process, where t_n is the instant when the n^{th} customer departs and $X_n = N(t_n + 0)$. Suppose $Y(t) = X_n, t_n \leq t < t_{n+1}, n = 1, 2, \dots$, then $\{Y(t), t \geq 0\}$ will be a semi-Markov process having embedded discrete time Markov chain (DTMC) $\{X_n, n = 0, 1, \dots\}$.

Steady-state Measures

To obtain steady-state probabilities of $Y(t)$, we need to compute the one-step transition probabilities of the embedded DTMC $\{X_n, n = 0, 1, \dots\}$. Since these points t_n are the regeneration points of the process $\{N(t), t \geq 0\}$, the sequence of points $\{t_n, n = 0, 1, \dots\}$ form a renewal process. Let A_n be a random variable denoting the number of customers that arrive during the service time of the n^{th} customer.

We have

$$X_{n+1} = \begin{cases} A_{n+1}, & X_n = 0 \\ X_n - 1 + A_{n+1}, & X_n \geq 1 \end{cases}.$$

Since service times of all the customers, denoted by B , have the same distribution, the distribution of A_n is same for all n .

Denoting, for all n ,

$$\begin{aligned} a_r &= Pr(A_n = r) \\ &= \int_0^\infty \frac{e^{-\lambda t} (\lambda t)^r}{r!} dF(t), \quad r = 0, 1, \dots \end{aligned}$$

Therefore,

$$\begin{aligned} P_{ij} &= Pr\{X_{n+1} = j / X_n = i\} \\ &= \begin{cases} a_j, & j \geq 0, i = 0 \\ a_{j-i+1}, & i \geq 1, j \geq i - 1 \\ 0, & i \geq 1, j < i - 1 \end{cases}. \end{aligned}$$

Denoting by $P = [P_{ij}]$, we have

$$P = [P_{ij}] = \begin{bmatrix} a_0 & a_1 & a_2 & \cdot & \cdot & \cdot \\ a_0 & a_1 & a_2 & \cdot & \cdot & \cdot \\ 0 & a_0 & a_1 & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \end{bmatrix}.$$

Note that X_n is irreducible Markov chain. When $\rho = \frac{\lambda}{\mu} < 1$, the chain is positive recurrent. Hence the Markov chain is ergodic.

The limiting probabilities

$$v_j = \lim_{n \rightarrow \infty} P_{ij}^{(n)}, \quad j = 0, 1, 2, \dots$$

exist and are independent of the initial state i .

The probability vector $v = [v_0, v_1, \dots]$ is given as the unique solution of $v = vP$ and $\sum_j v_j = 1$. The solution can be obtained by using n generating functions of v_j 's

and a_j 's. Define

$$A(z) = \sum_{j=0}^{\infty} a_j z^j \text{ and } V(z) = \sum_{j=0}^{\infty} v_j z^j.$$

Now,

$$\begin{aligned} A(z) &= \sum_{j=0}^{\infty} a_j z^j \\ &= \sum_{j=0}^{\infty} z^j \left(\int_0^{\infty} \frac{e^{-\lambda t} (\lambda t)^j}{j!} dF(t) \right) \\ &= \int_0^{\infty} e^{-\lambda t} e^{\lambda z t} f(t) dt \\ &= \int_0^{\infty} e^{-t(\lambda - \lambda z)} f(t) dt \\ &= f^*(\lambda - \lambda z). \end{aligned}$$

Here, $f^*(s)$ is the Laplace transform of $f(t)$.

Expected arrival is given by

$$\begin{aligned} E(\text{Arrival}) &= A'(1) = \left. \frac{d}{dz} A(z) \right|_{z=1} \\ &= \left. \frac{d}{dz} f^*(\lambda - \lambda z) \right|_{z=1} \\ &= \left. \frac{df^*}{ds} \cdot \frac{ds}{dz} \right|_{z=1} \\ &= \left. \frac{df^*}{ds} \right|_{s=0} \cdot \left. \frac{d(\lambda - \lambda z)}{dz} \right|_{z=1} \\ &= -\lambda \left. \frac{d}{ds} \left(\int_0^{\infty} e^{-st} f(t) dt \right) \right|_{s=0} \\ &= -\lambda \left(\int_0^{\infty} -t f(t) dt \right) \\ &= \frac{\lambda}{\mu} = \rho. \end{aligned}$$

Now,

$$v_j = v_0 a_j + \sum_{i=0}^{j+1} v_i a_{j-i+1}, \quad j = 0, 1, 2, \dots$$

Multiplying by z_j on both the sides and taking the sum, we get

$$\begin{aligned}
 \sum_{j=0}^{\infty} V_j z^j &= \sum_{j=0}^{\infty} v_0 a_j z^j + \sum_{j=0}^{\infty} \left(\sum_{i=1}^{j+1} v_j a_{j-i+1} \right) z^j \\
 V(z) &= v_0 \sum_{j=0}^{\infty} a_j z^j + \sum_{i=1}^{\infty} \left(\sum_{j=i-1}^{\infty} v_j a_{j-i+1} z^j \right) \\
 &= v_0 A(z) + \frac{1}{z} [V(z) - v_0] A(z) \\
 V(z) - \frac{1}{z} A(z) V(z) &= v_0 A(z) - \frac{1}{z} v_0 A(z) \\
 V(z) &= \frac{v_0 A(z)(z-1)}{z - A(z)}.
 \end{aligned}$$

Pollaczek-Khinchin (P-K) Formula

We know that $V(1) = A(1) = 1$,

$$\begin{aligned}
 V(1) &= \lim_{z \rightarrow 1} v_0 \left(\frac{A'(z)(z-1) + A(z)}{1 - A'(z)} \right) \\
 1 &= \frac{v_0 A(1)}{1 - A'(1)} = \frac{v_0}{1 - A'(1)}
 \end{aligned}$$

provided $A'(1)$ is finite and is less than 1.

Taking $\rho = A'(1)$ we get, $v_0 = 1 - \rho$. Hence,

$$V(z) = \frac{(1 - \rho)(z - 1)f^*(\lambda((1 - z)))}{z - f^*(\lambda(1 - z))}.$$

This is known as **Pollaczek-Khinchin (P-K)** formula.

Limiting Distributions

Note that, the quantities of interest are the state probabilities at a different set of points, namely the arrival epochs. From the point of view of an arriving customer, the number of customers that he finds in the system, not the number he leaves behind, is the quantity of interest. Since the arrival follows Poisson process, the equilibrium distribution of the number of customers found by the arrival and the equilibrium of the number of customers left behind by a departure in this queueing model.

Note that, in the $M/G/1$ queue, the limiting distributions of the number of customers in the system at arrival epochs, at departure epochs, and at arbitrary points in time are the same only since the arrivals occur as a Poisson process.

P-K Mean Value Formula

Average number of customers in the system in steady-state is given by

$$L_s = \frac{dV(z)}{dz} \Big|_{z=1}$$

More generally,

$$L_s = \rho + \frac{\lambda^2(E(B)^2)}{2(1-\rho)}.$$

This is known as the P-K mean value formula.

Here, $E(B^2)$ is the second order moment about the origin for the service time. This result holds true for all scheduling disciplines in which the server is busy if the queue is non-empty. The expected system size L_s can be computed without $V(z)$ also.

$$\begin{aligned} L_s &= E(X_{n+1}) = E(X_n) \\ L_s &= L_s - E(U(X_n)) + E(Arrival) \end{aligned}$$

where

$$U(X_n) = \begin{cases} 1, & X_n > 0 \\ 0, & X_n = 0 \end{cases}$$

$$\begin{aligned} E(U(X_n)) &= E(Arrival) \\ &= \int_0^\infty E(Arrival|s=t)dB(t) \\ &= \int_0^\infty \lambda tb(t)dt \\ &= \lambda E(B) = \frac{\lambda}{\mu} = \rho. \end{aligned}$$

Now,

$$\begin{aligned} X_{n+1} &= X_n - U(X_n) + A_n \\ X_{n+1}^2 &= X_n^2 + U^2(X_n) + A_n^2 \\ &\quad - 2X_n U(X_n) - 2A_n U(X_n) + 2A_n X_n \end{aligned}$$

where $U^2(X_n) = U(X_n)$ and $X_n U(X_n) = X_n$.

Now, taking expectations on both the sides, we get

$$\begin{aligned} E(X_{n+1}^2) &= E(X_n^2) + E(U^2(X_n)) + E(A_n^2) \\ &\quad - 2E(X_n U(X_n)) - 2E(A_n U(X_n)) + 2E(A_n X_n). \end{aligned}$$

P-K Mean Formula

Using $E(X_{n+1}^2) = E(X_n^2)$ and $Var(Arrival) = \rho + \lambda^2 \sigma_B^2$ where σ_B^2 is the variance of the service time distribution, we get

$$L_s = \rho + \frac{\lambda^2 \sigma_B^2 + \rho^2}{2(1 - \rho)}$$

which is the P-K mean formula.

Using this, other measures such as L_q , T_s and T_q can be obtained as follows

$$L_q = L_s - \lambda E(B); \quad E(B) = \frac{1}{\mu}$$

$$T_s = \frac{L_s}{\lambda}; \quad T_q = \frac{L_q}{\lambda}.$$

Lecture 5 Non Markovian Queues (contd.)

Special Case

For the case where the service time is constant, $Var(B) = 0$, then the P-K formula for $M/D/1$ queue reduces to

$$L_s = \rho + \frac{\rho^2}{2(1 - \rho)}$$

where $\rho = \lambda/\mu$ and $1/\mu$ is the constant service time.

Alternatively, the steady state probabilities can be obtained in the following way. When the embedded DTMC is irreducible and ergodic, the limiting probability $v_j = \lim_{n \rightarrow \infty} P_{ij}^{(n)}$, $j = 0, 1, 2, \dots$ are given as unique solutions of

$$v_j = a_j v_0 + \sum_{i=1}^{j+1} a_{j-i+1} v_i, \quad j = 0, 1, 2, \dots \quad (1)$$

Suppose the mean sojourn time in state j , $\frac{1}{\mu_j}$ (finite) and $\{X_n, n = 0, 1, \dots\}$ is irreducible and ergodic, then

$$P_j = \frac{\frac{1}{\mu_j} v_j}{\sum_k \frac{1}{\mu_k} v_k}, \quad j = 0, 1, 2, \dots$$

where v_j are obtained by solving (1).

Special Case

Here also, as a special case, when the service time follows exponential distribution with mean $\frac{1}{\mu}$ can be discussed. Assume that $\lambda < \mu$. The one step transition probability matrix P becomes

$$P = \begin{bmatrix} 1 & 0 & \cdot & \cdot & \dots \\ \frac{\mu}{\lambda+\mu} & 0 & \frac{\lambda}{\lambda+\mu} & \cdot & \dots \\ 0 & \frac{\mu}{\lambda+\mu} & 0 & \frac{\lambda}{\lambda+\mu} & \dots \\ \vdots & \vdots & \vdots & \vdots & \vdots \end{bmatrix}.$$

Then the solution of $V = VP$ is

$$v_j = \left(\frac{\lambda + \mu}{\lambda} \right) \left(\frac{\lambda}{\mu} \right)^j v_0, \quad j = 1, 2, \dots$$

Hence, for the $M/M/1$ queueing system, the limiting probabilities are given by

$$\begin{aligned} \pi_j &= \frac{\left(\frac{\lambda + \mu}{\lambda} \right) \left(\frac{\lambda}{\mu} \right)^j \left(\frac{1}{\mu} \right)}{1 + \sum_{k=1}^{\infty} \left(\frac{1}{\mu} \right) \left(\frac{\lambda + \mu}{\lambda} \right) \left(\frac{\lambda}{\mu} \right)^k}, \quad j = 0, 1, \dots \\ &= \left(1 - \frac{\lambda}{\mu} \right) \left(\frac{\lambda}{\mu} \right)^j, \quad j = 0, 1, \dots \end{aligned}$$

The expected number of customers in the system is given by the Pollazek-Khintchine formula.

The formula states that mean queueing delay is given by:

$$T_q = \frac{\rho(1 + \mu^2 \sigma_B^2)}{2\mu(1 - \rho)}.$$

Hence, the average time spent in the system is:

$$T_s = T_q + \frac{1}{\mu} = \frac{1 + \mu^2 \sigma_B^2}{2(\mu - \lambda)} + \frac{1}{\mu}.$$

By Little's formula, we know: $L_s = \lambda T_s$. Hence,

$$L_s = \rho + \frac{\rho^2 + \lambda^2 \sigma_B^2}{2(1 - \rho)}$$

where $1/\mu$ is the mean and σ_B^2 is the variance of the service time distribution and $\rho = \frac{\lambda}{\mu}$.

Special Case

When service time is exponentially distributed with mean $\frac{1}{\mu}$, then

$$L_s = \rho + \frac{2\rho^2}{2(1 - \rho)} = \frac{\rho}{1 - \rho}.$$

In this derivation, we assume FCFS scheduling to simplify the analysis.

However, the above formulae are valid for any scheduling discipline in which the server is busy if the queue is non-empty, no customer departs the queue before completing service, and the order of service is not dependent on the knowledge about service times. Hence, using Little's formula, the average time spent in the system is given by $\lambda T_s = L_s$.

EXAMPLE 1. *People are entering cricket stadium at New Delhi to watch a cricket match. There is only one ticket line to purchase tickets. Each ticket purchase takes an average of 20 seconds. The average arrival rate is 2 persons per minute. Find the average length of queue and average waiting time in queue assuming $M/G/1$ queueing with service time follows uniform distribution between 15 and 25 seconds. Departure rate: $\mu = 20$ seconds/person or 3 persons/minute. Arrival rate: $\lambda = 2$ persons/minute. $\rho = 2/3$. Given $B \sim U(15, 25)$, $E(B) = 1/3$ minute.*

$M/G/c/c$ System

The $M/G/c/c$ queueing system, also known as the Erlang loss system since its development can be traced primarily to the Danish mathematician, Agner Krarup Erlang (1878-1929). Customers arrive according to a Poisson process with rate λ . Service from one of c parallel servers. The service times are assumed to be independent and identically distributed with a finite mean $1/\mu$. Because there is no waiting space in the system, arriving customers who find all c servers busy are lost. For example, the lost customers might represent vehicles arriving at a parking garage to find no vacant parking spaces.

Steady-state Distribution

The offered load per server, often termed the traffic intensity, is given by

$$\rho = \frac{\lambda}{c\mu} = \frac{a}{c}.$$

Let $N(t)$ be the number of occupied servers at time t . Let $\{N(t), t \geq 0\}$ be a stochastic process with state space $S = \{0, 1, 2, \dots, c\}$. Define

$$\pi_j = \lim_{t \rightarrow \infty} P(N(t) = j), \quad j = 0, 1, 2, \dots$$

The steady-state distribution is given by

$$\pi_j = \frac{a^j/j!}{\sum_{i=0}^c a^i/i!}, \quad j = 0, 1, 2, \dots, c$$

This distribution is truncated Poisson distribution with parameter a .

Erlang B Formula

Because Poisson arrivals see time averages (PASTA), the long-run proportion of arriving customers who see c servers busy is precisely, denotes by Erlang B formula $B(c, a)$, is given by

$$B(c, a) = \frac{a^c/c!}{\sum_{i=0}^c a^i/i!}$$

When a and c are large, it can be difficult to compute due to the presence of factorials and potentially very large powers. Use the following recursive formula

$$B(k, a) = \frac{aB(k-1, a)}{k + aB(k-1, a)}, \quad k = 1, 2, \dots, c$$

where $B(0, a) = 1$.

The Erlang B formula is a fundamental result for telephone traffic engineering problems and can be used to select the appropriate number of trunks (servers) needed to ensure a small proportion of lost calls (customers).

For a fixed, the blocking probability $B(c, a)$ monotonically decreases to zero as c increases, and for c fixed, $B(c, a)$ monotonically increases to unity as a increases. When the service times are i.i.d. exponential random variables with mean $1/\mu$, the system is an $M/M/c/c$ loss system. In this case,

$$B(c, a) = \frac{(\lambda/\mu)^c/c!}{\sum_{i=0}^c (\lambda/\mu)^i/i!}.$$

Erlang C Formula

Related to the Erlang B formula is the Erlang C formula (or Erlang delay formula) for the $M/M/c$ system (or Erlang delay system), which includes an infinite-capacity queue to accommodate arriving customers who find all c servers busy. For this model, P_c is interpreted as the long-run proportion of customers who experience a

delay before their service begins. The model assumes that the customers are willing to wait as long as needed to receive service.

The Erlang C formula is given by

$$C(c, a) = \frac{\frac{a^c}{c!(1-\rho)}}{\sum_{i=0}^{c-1} \frac{a^i}{i!} + \frac{a^c}{c!(1-\rho)}}.$$

Note that above result does not hold for arbitrary service time distributions, and it requires that the traffic intensity ρ does not exceed unity.

Lecture 6 Non Markovian Queues (contd.)

GI/M/1 Queue

Consider the queueing model *GI/M/1*. Customers arrive at time points $0 = t_0, t_1, t_2, \dots$. Let $Z_n = t_{n+1} - t_n, n = 1, 2, \dots$ be i.i.d. random variables with distribution function $A(\cdot)$ with mean $1/\lambda$. Let the service time distribution be exponential with mean $1/\mu$. Let $Q(t)$ be the number of customers in the system at time t .

Define $Q(t_n - 0) = Q_n, n = 1, 2, \dots$. Thus Q_n is the number in the system just before the n th arrival. Then $\{(Q_n, t_n), n = 0, 1, \dots\}$ is a Markov renewal process, where t_n is the instant when the n^{th} customer arrives and $Q_n = Q(t_n - 0)$. Define X_n as the number of potential service completions during the inter-arrival period Z_n . Let $\{b_j, j = 0, 1, \dots\}$ be the distribution of X_n . Then,

$$b_j = P(X_n = j) = \int_0^\infty \frac{e^{-\mu t} (\mu t)^j}{j!} dA(t)$$

The relationship between Q_n and Q_{n+1} is given by

$$Q_{n+1} = \begin{cases} Q_n + 1 - X_{n+1}, & Q_n + 1 - X_{n+1} > 0 \\ 0, & Q_n + 1 - X_{n+1} \leq 0 \end{cases}$$

Q_{n+1} is independent of X_{n+1} and Q_{n+1} does not depend on any random variable with an earlier index parameter than n . $\{Q_n, n = 0, 1, \dots\}$ is a time homogeneous discrete time Markov chain.

The transition probability is given by

$$\begin{aligned} p_{ij} &= P(Q_{n+1} = j \mid Q_n = i) \\ &= \begin{cases} P(X_{n+1} = i - j + 1), & j > 0 \\ P(X_{n+1} \geq i + 1), & j = 0. \end{cases} \end{aligned}$$

In matrix form,

$$P = \begin{bmatrix} \sum_{i=1}^{\infty} b_i & b_0 & 0 & 0 & \dots \\ \sum_{i=2}^{\infty} b_i & b_1 & b_0 & 0 & \dots \\ \sum_{i=3}^{\infty} b_i & b_2 & b_1 & b_0 & \dots \\ \vdots & \vdots & \vdots & \vdots & \ddots \end{bmatrix}$$

For the DTMC to be irreducible, $b_0 > 0$ and $b_0 + b_1 < 1$. We can easily determine that the DTMC is aperiodic. Let

$$\phi(\theta) = \int_0^\infty e^{-\theta t} dA(t), \quad \operatorname{Re}(\theta) > 0.$$

The probability generating function of $\{b_j\}$ is obtained as

$$\beta(z) = \sum_{j=0}^{\infty} b_j z^j = \phi(\mu - \mu z), \quad |z| \leq 1.$$

Hence,

$$E(Z) = \beta'(1) = -\mu\phi'(0) = \frac{\mu}{\lambda}.$$

We define the traffic intensity

$$\rho = (\text{arrival rate})/(\text{service rate}).$$

From above, we get

$$\rho = \frac{\lambda}{\mu}.$$

It can be shown that the DTMC is positive recurrent when $\rho < 1$, null recurrent when $\rho = 1$, and transient when $\rho > 1$. When $\rho < 1$, the DTMC is irreducible, aperiodic and positive recurrent, then the limiting distribution exists. Let $\pi = [\pi_0, \pi_1, \pi_2, \dots]$ be the limiting probabilities defined as $\pi_j = \lim_{n \rightarrow \infty} P_{ij}^{(n)}$. We determine the limiting distribution by solving

$$\pi_j = \sum_{i=0}^{\infty} \pi_i P_{ij}, \quad j = 0, 1, 2, \dots; \quad \sum_{j=0}^{\infty} \pi_j = 1.$$

We get

$$\pi_j = (1 - \zeta)\zeta^j, \quad j = 0, 1, 2, \dots$$

where ζ is the unique root of the equation $z = \phi(\mu - \mu z)$ in the interval $(0, 1)$. Note that the queue length distribution found just before an arriving customer is geometric with parameter ζ . The best computational method for the determination of the limiting distribution seems to be the direct matrix multiplication to get P^n for increasing values of n until the rows can be considered to be reasonably identical. It

is important to note that the imbedded Markov chain analysis gives the properties of the number in the system at arrival epochs.

As pointed out under the discussion of the $M/G/1$ queue, the limiting distributions of the number of customers in the system at arrival epochs, at departure epochs, and at arbitrary points in time are the same only when the arrivals occur as a Poisson process.

Prabhu [1] and Bhat [2] derived the results of limiting distribution at an arbitrary time t .

Let $p_j = \lim_{t \rightarrow \infty} P(Q(t) = j)$ where $Q(t)$ is the number at an arbitrary time t . The limiting distribution $\{p_j, j = 0, 1, 2, \dots\}$ when $\rho < 1$

$$p_0 = 1 - \rho, \quad p_j = \rho(1 - \zeta)\zeta^{j-1}, j \geq 1$$

To determine the distribution of the waiting time of a customer, we need the distribution of the number of customers in the system at the time of that customer's arrival. Replace ρ by ζ in $M/M/1$ results, since the queue length distribution is same. Waiting time distribution is given by

$$F_q(t) = P(T_q \leq t) = 1 - \zeta e^{-\mu(1-\zeta)t}.$$

Mean and variance of waiting time, time spent distribution and its mean and variance can be obtained.

$$E(T_q) = \frac{\zeta}{\mu(1 - \zeta)}; \quad \text{Var}(T_q) = \frac{\zeta(2 - \zeta)}{\mu^2(1 - \zeta)^2}.$$

Since the arrivals are not Poisson process, the mean value approach does not work. Here $E(T) = (E(L_a))1/\mu + 1/\mu$.

$$E(T) = \frac{1}{\mu(1 - \zeta)}; \quad \text{Var}(T) = \frac{1}{\mu^2(1 - \zeta)^2}.$$

$D/M/1$ Queue

For $D/M/1$ queueing model,

$$\phi(\theta) = \int_0^\infty e^{-\theta t} dA(t) = e^{-\theta}.$$

In this model, the limiting distribution is given by

$$\pi_j = (1 - \zeta)\zeta^j, \quad j = 0, 1, 2, \dots$$

where ζ is the root of the equation $z = e^{-\mu(1-z)}$.

GI/M/c Queue

The imbedded Markov chain analysis of the queue *GI/M/1* can be easily extended to the multiserver queue *GI/M/c*. Here, the transition probabilities P_{ij} are different. Consider three cases for the initial value i and the final value j :

$$i + 1 \geq j \geq c; \quad i + 1 \leq c \text{ and } j \leq c \quad \text{and} \quad i + 1 > c \text{ but } j < c.$$

Case 1:

$$i + 1 \geq j \geq c$$

$$P_{ij} = \int_0^\infty e^{-c\mu t} \frac{(c\mu t)^{i+1-j}}{(i+1-j)!} dA(t).$$

This represents $i + 1 - j$ service completions during an inter-arrival period, when all c servers are busy.

Case 2:

$$i + 1 \leq c \text{ and } j \leq c$$

$$P_{ij} = \binom{i+1}{1+1-j} \int_0^\infty e^{-j\mu t} (1 - e^{-\mu t})^{i+1-j} dA(t).$$

This represents $i + 1 - j$ out of $i + 1$ customers complete service.

Case 3:

$$i + 1 > c \text{ but } j < c$$

$$P_{ij} = \int_{t=0}^\infty \int_{\tau=0}^t e^{-c\mu t} \frac{(c\mu t)^{i-c}}{(i-c)!} c\mu \binom{c}{c-j} e^{-j\mu(t-\tau)} (1 - e^{-\mu(t-\tau)})^{c-j} d\tau dA(t).$$

This represents $i + 1 - c$ customers complete service with rate $c\mu$, and then $c - j$ out of c customers complete service.

Refer Gross and Harris [3] for the limiting distribution.

GI/M/1/N Queue

The queueing model $GI/M/1/N$ is similar to $GI/M/1$, but now the system capacity is restricted to N . The relation between Q_n and Q_{n+1} is given by

$$Q_{n+1} = \begin{cases} \min(Q_n + 1 - X_{n+1}, N) & Q_n + 1 - X_{n+1} > 0 \\ 0, & Q_n + 1 - X_{n+1} \leq 0 \end{cases}.$$

The transition probability matrix P is given by,

$$P = \begin{bmatrix} \sum_{i=1}^{\infty} b_i & b_0 & 0 & \dots & 0 & 0 \\ \sum_{i=2}^{\infty} b_i & b_1 & b_0 & \dots & 0 & 0 \\ \vdots & \vdots & \vdots & & & \\ \sum_{i=N}^{\infty} b_i & b_{N-1} & b_{N-2} & \dots & b_1 & b_0 \\ \sum_{i=N}^{\infty} b_i & b_{N-1} & b_{N-2} & \dots & b_1 & b_0 \end{bmatrix}.$$

Note that the last two rows of the matrix P are identical because the Markov chain effectively starts off with N customers from either of the states $N-1$ and N . Since the finite state DTMC is irreducible, aperiodic and positive recurrent, then the limiting distribution exists. Let $\pi = [\pi_0, \pi_1, \pi_2, \dots, \pi_N]$ be the limiting probabilities defined as $\pi_j = \lim_{n \rightarrow \infty} P_{ij}^{(n)}$.

We determine the limiting distribution by solving

$$\pi_j = \sum_{i=0}^N \pi_i P_{ij}, j = 0, 1, 2, \dots; \sum_{j=0}^N \pi_j = 1.$$

GI/G/1 Queue

Consider a single-server queue with arbitrary arrival process and arbitrary service time distribution, with infinite buffer. Let S be a random variable representing the service time and let $E[S] = 1/\mu$, i.e., μ denotes the service rate. Let λ be the mean arrival rate. Assume that $\lambda < \mu$ so that the queue is stable. One can show that, for a stable $GI/G/1$ queue, utilization is given by λ/μ .

We are interested in the distribution of the number of customers in the system. Let p_n be the probability that there are n in the system. The probability that the

system is empty is given by

$$p_0 = 1 - \lambda/\mu.$$

Note that, Little's formula can be applied to obtain average measures such as mean system size, mean queue size and mean delay.

References

1. J Medhi, "Stochastic Processes", 3rd edition, New Age International Publishers, 2009.
2. S Karlin and H M Taylor, "A First Course in Stochastic Processes", 2nd edition, Academic Press, New York, 1975.
3. E Barzen, "Stochastic Processes", Holden-Day, San Francisco, 1962.
4. U N Bhat, "Elements of Applied Stochastic Processes", John Wiley & Sons, New York, 1972.
5. U N Bhat, "A Study of the Queueing Systems $M/G/1$ and $GI/M/1$ ", Lecture Notes in Operations Research and Mathematical Economics, Vol. 2, Springer-Verlag, New York, 1968.
6. D Gross and C M Harris, "Fundamentals of Queueing Theory", 3rd edition, Wiley, New York, 1998.
7. U N Bhat, "An Introduction to Queueing Theory Modeling and Analysis in Applications", Birkhauser, Boston, 2008.
8. Kishor S. Trivedi, "Probability and Statistics with Reliability, Queuing, and Computer Science Applications", John Wiley and Sons, New York, 2001.
9. B R Haverkort and K S Trivedi, "Specification Techniques for Markov Reward Models", Discrete Event Dynamic Systems: Theory and Applications, Vol. 3, pp. 219 - 247, 1993.
10. V. G. Kulkarni, *Modeling and Analysis of Stochastic Systems*, Chapman & Hall, London, 1995.

11. S. Dharmaraja, Kishor S. Trivedi and D. Logothetis: Performance modelling of wireless networks with generally distributed hand-off interarrival times, *Computer Communications Journal*, 26 (2003) pp. 1747-1755.
12. R. Fricks, M. Telek, A. Puliafito and K.S. Trivedi, “Markov renewal theory applied to performability evaluation”, in *State-of-the Art in Performance Modeling and Simulation. Modeling and Simulation of Advanced Computer Systems: Applications and Systems*, K. Bagchi and G. Zobrist (eds.), pp. 193–236; Gordon and Breach Publishers, Newark, NJ, 1998.
13. R. German, *Performance Analysis of Communication Systems : Modeling with Non-Markovian Stochastic Petri Nets*, John Wiley, Chichester, 2000.

Module 8: Branching Processes

Lecture 1 Galton - Watson Process

H.W. Watson in 1874 presented a solution to the “problem of the extinction of families” which is as follows:

“Let $p_0, p_1, p_2, \dots$ be the respective probabilities that a man has $0, 1, 2, \dots$ sons and let each son have the same probability for sons of his own, and so on. What is the probability that the male line is extinct after r generations, and more generally what is the probability for any given number of descendants in the male line in any given generation?”.

Watson gave the solution “every family will die out, even when the population size, on the average, increase from one generation to the next”. Francis Galton prefaced to the solution by Watson, and Galton recognized that a first step in studying the hypothesis would be to determine the probability that an ordinary family will disappear using fertility data for the whole population. The mathematical model of Galton and Watson is known as a Galton - Watson process. In 1938, A. Kolmogorov also treated the above problem and determined the asymptotic form of the probability that the family is still in existence after a large finite number of generations.

In this lecture, we presents the results on extinction and survival probabilities expressed in terms of the mean and the generating function of a random variable whose distribution models the branching process.

DEFINITION 1. *The branching process is a system of particles (individuals, cells, molecules, etc.) which live for a random time and, at some point during lifetime or at the moment of death, produce a random number of progeny.*

DEFINITION 2. *Processes allowing production of new individuals during a parent individuals lifetime are called the general branching processes.*

EXAMPLE 3. *populations of higher organisms, like vertebrates and plants.*

DEFINITION 4. *Processes that assume the production of progeny at the terminal point of the parent entity's lifetime are called the classical branching processes.*

EXAMPLE 5. *e populations of biological cells, genes or biomolecules.*

One of the important notions in the theory of branching processes is that of the type space. The type space is the set, which can be finite, countably infinite, or a continuum, of all possible varieties of particles included in the process. Particles of a given type may produce particles of different types. Restrictions on type transitions, as well as on the type space, lead to different properties of resulting processes.

Galton-Watson Process

Consider a classical branching process in which progeny are born at the moment of the parents death. The distribution of particle lifetime τ has much impact on the behavior and analysis of the process. If $\tau = 1$, then it is called the Galton - Watson process. If τ is exponentially distributed random variable, then the resulting process is called a Markov branching process. If τ is an arbitrary non-negative random variable, the resulting process is called an “age-dependent” or Bellman - Harris process.

Consider particles that can generate additional particles of the same kind. An initial set of particles, which we call the 0-th generation, have offspring that are called the first generation, their offspring are the second generation, and so on. We denote by $Z_0, Z_1, Z_2, \dots$ as the number of particles in the 0-th, 1-st, 2-nd, $\dots$ generations respectively.

Furthermore, we make the following assumptions:

If the size of the n -th generation is known, then the probability law governing later generations does not depend on the sizes of previous generations (i.e., preceding the n -th). It means $\{Z_n, n = 0, 1, \dots\}$ is a discrete time Markov chain.

We make the additional assumption that the transition probabilities for the Markov chain do not vary with time. It means $\{Z_n, n = 0, 1, \dots\}$ is a time homogeneous DTMC. Note that the number of offsprings born to a particle does not depend on how many other particles are present. Also, we assume that $Z_0 = 1$.

If $Z_n = 0$ then $P(Z_{n+1} = 0) = 1$. The probability distribution of Z_1 is given by

$$P(Z_1 = k) = p_k, \quad k = 0, 1, \dots, \quad \sum_{k=0}^{\infty} p_k = 1$$

where p_k is the probability that a particle existing in the n -th generation has k offsprings in the $(n + 1)$ -th generation.

It is assumed that p_k is independent of the generation number n . The conditional distribution of Z_{n+1} given $Z_n = k$ is appropriate to the assumption that different particles reproduce independently; that is, Z_{n+1} is distributed as the sum of k independent random variables, each distributed like Z_1 . Since $\{Z_n, n = 0, 1, \dots\}$ is a time homogeneous DTMC, the one-step transition probabilities are given by-

$$\begin{aligned} p_{ij} &= P(Z_{n+1} = j \mid Z_n = i) \\ &= P(Y_1 + Y_2 + \dots + Y_i = j), \quad i, j = 1, 2, \dots \end{aligned} \quad (1)$$

where $Y_1, Y_2, \dots, Y_i$ are independent random variables, each distributed like Z_1 .

EXAMPLE 6. Consider the Galton - Watson process $\{Z_n, n = 0, 1, 2, \dots\}$ with offspring distribution $\{p_k\}$. Assume that $p_0 = 1/5, p_1 = 3/5$ and $p_2 = 1/5$. Calculate $P(Z_2 = 3 \mid Z_1 = 2)$, $P(Z_3 = 2 \mid Z_2 = 3)$ and $P(Z_2 = 2)$.

Solution: $P(Z_2 = 3 \mid Z_1 = 2)$

$$\begin{aligned} &= P(Y_1 + Y_2 = 3) \\ &= P(Y_1 = 0, Y_2 = 3) + P(Y_1 = 1, Y_2 = 2) \\ &\quad + P(Y_1 = 2, Y_2 = 1) + P(Y_1 = 3, Y_2 = 0) \\ &= P(Y_1 = 0)P(Y_2 = 3) + P(Y_1 = 1)P(Y_2 = 2) \\ &\quad + P(Y_1 = 2)P(Y_2 = 1) + P(Y_1 = 3)P(Y_2 = 0) \\ &= p_0p_3 + p_1p_2 + p_2p_1 + p_3p_0 \\ &= 0 + \frac{3}{25} + \frac{3}{25} + 0 = \frac{6}{25}. \end{aligned}$$

Similarly,

$$\begin{aligned} P(Z_3 = 2 \mid Z_2 = 3) &= \\ P(Y_1 + Y_2 + Y_3 = 2) &= \\ &= p_0p_0p_2 + p_0p_2p_0 + p_2p_0p_0 \\ &\quad + p_1p_1p_0 + p_1p_0p_1 + p_0p_1p_1. \end{aligned}$$

Now,

$$\begin{aligned}
 P(Z_2 = 2) &= P(Z_2 = 2 \mid Z_1 = 0)P(Z_1 = 0) \\
 &\quad + P(Z_2 = 2 \mid Z_1 = 1)P(Z_1 = 1) \\
 &\quad + P(Z_2 = 2 \mid Z_1 = 2)P(Z_1 = 2) \\
 &= 0 + P(Y_1 = 2)P(Z_1 = 1) \\
 &\quad + P(Y_1 + Y_2 = 2)P(Z_1 = 2) \\
 &= p_2p_1 + (p_0p_2 + p_1p_1 + p_2p_0)p_2.
 \end{aligned}$$

Probability Generating Function

Let $\{Y_i, i = 1, 2, \dots\}$ be i.i.d. random variables each having the same distribution like Z_1 . Let $H(s)$ be probability generating function of Y_i .

$$H(s) = \sum_k P(Y_i = k)s^k = \sum_k p_k s^k, \quad |s| \leq 1$$

where s is a complex variable. Let $H_n(s)$ be probability generating function of Z_n .

$$H_n(s) = \sum_k P(Z_n = k)s^k, \quad |s| \leq 1$$

where s is a complex variable.

Since $Z_0 = 1$, we have

$$H_0(s) = s$$

and

$$\begin{aligned}
 H_1(s) &= \sum_k P(Z_1 = k)s^k \\
 &= \sum_k P(Z_1 = k \mid Z_0 = 1)s^k \\
 &= \sum_k p_k s^k \\
 &= H(s).
 \end{aligned}$$

THEOREM 7. $H_n(s) = H_{n-1}(H(s))$ and $H_n(s) = H(H_{n-1}(s))$.

Proof. For $n = 1, 2, \dots$,

$$\begin{aligned} P(Z_n = k) &= \sum_i P(Z_n = k \mid Z_{n-1} = i)P(Z_{n-1} = i) \\ &= \sum_i P(Y_1 + Y_2 + \dots + Y_i = k)P(Z_{n-1} = i). \end{aligned}$$

Now,

$$\begin{aligned} H_n(s) &= \sum_k P(Z_n = k)s^k \\ &= \sum_k \left[\sum_i P(Z_n = k \mid Z_{n-1} = i)P(Z_{n-1} = i) \right] s^k \\ H_n(s) &= \sum_i P(Z_{n-1} = i) \left[\sum_k P(Y_1 + Y_2 + \dots + Y_i = k)s^k \right]. \end{aligned}$$

Since $\{Y_i, i = 1, 2, \dots\}$ are i.i.d. random variables and PGF of Y_i is $H(s)$, then the PGF of $Y_1 + Y_2 + \dots + Y_i$ is $[H(s)]^i$. Hence,

$$H_n(s) = \sum_i P(Z_{n-1} = i)[H(s)]^i = H_{n-1}(H(s)).$$

We know that $H_1(s) = H(s) = H(H_0(s))$ since $H_0(s) = s$. Suppose for $n = j$,

$$H_j(s) = H(H_{j-1}(s))$$

then

$$\begin{aligned} H_{j+1}(s) &= H_j(H(s)) \\ &= H(H_{j-1}(H(s))) \\ &= H(H_j(s)). \end{aligned}$$

By induction, it is proved. □

THEOREM 8. Let $\mu = E(Z_1)$ and $\sigma^2 = \text{Var}(Z_1)$. Then

$$E(Z_n) = \mu^n$$

and

$$\text{Var}(Z_n) = \begin{cases} \frac{\mu^{n-1}(\mu^n - 1)\sigma^2}{\mu - 1}, & \mu \neq 1 \\ n\sigma^2 & \mu = 1 \end{cases}.$$

Proof. By Theorem 7, we have $H_n(s) = H_{n-1}(H(s))$. Then differentiating and putting $s = 1$, $H'_n(1) = H'_{n-1}(H(1))H'(1)$. We know that, $H(1) = 1$, $H'(1) = \mu$ and for $n = 1, 2, \dots$

$$H'_n(1) = H'_{n-1}(1)\mu = \dots = \mu^n.$$

Hence $E(Z_n) = \mu^n$. □

REMARK 9. 1. If $\mu = 1$, then $E(Z_n) \rightarrow 1$ and $\text{Var}(Z_n) \rightarrow \infty$ when $n \rightarrow \infty$.

2. If $\mu < 1$, then $E(Z_n) \rightarrow 0$ and $\text{Var}(Z_n) \rightarrow \frac{\sigma^2}{1-\mu}$ when $n \rightarrow \infty$.

3. If $\mu > 1$, then $E(Z_n) \rightarrow \infty$ and $\text{Var}(Z_n) \rightarrow \infty$ when $n \rightarrow \infty$.

Lecture 2 Properties of GW Process

Criticality

A very important classification is based on the mean progeny count $\mu = E(Z_1)$.

$$E(Z_n) = \mu^n.$$

Therefore, in the expected value sense, the process grows geometrically if $\mu > 1$, stays constant if $\mu = 1$, and decays geometrically if $\mu < 1$. These three cases are called supercritical, critical, and subcritical, respectively:

$$\mu > 1, \text{ supercritical } E[Z_n] \uparrow \infty$$

$$\mu = 1, \text{ critical } E[Z_n] = 1$$

$$\mu < 1, \text{ subcritical } E[Z_n] \downarrow 0.$$

EXAMPLE 1. Consider the Galton - Watson process $\{Z_n, n = 0, 1, 2, \dots\}$ with offspring distribution $\{p_k\}$. Assume that $p_0 = 1/5, p_1 = 3/5$ and $p_2 = 1/5$. Then $\mu = E(Z_1) = 1$. Hence, this process is a critical Galton - Watson process. $E(Z_1^2) = \frac{7}{5}$, then $\sigma^2 = \frac{2}{5}$. Hence, for $n = 1, 2, \dots$

$$E(Z_n) = 1 \text{ and } \text{Var}(Z_n) = \frac{2n}{5}.$$

Probability of Extinction

If $Z_n = 0$ for some value $n \geq 1$, then $Z_m = 0$ for all $m \geq n$ and also

$$P(Z_m = 0 \mid Z_n = 0) = 1, \text{ for all } m \geq n.$$

We say that a branching process is extinguished when $P(Z_n = 0) = 1$, for some value of $n = 1, 2, \dots$

THEOREM 2. If $\mu \leq 1$, then the probability of extinction is one. If $\mu > 1$, then the probability of extinction is the positive root less than unity of the equation $H(s) = s$.

Proof. Let d_n be the probability that extinction occurs at or before the n -th gener-

ation, i.e., $d_n = P(Z_n = 0)$. We know that, $d_0 = 0$. Now,

$$\begin{aligned}
 d_{n+1} &= P(Z_{n+1} = 0) \\
 &= P(Z_{n+1} = 0 \mid Z_n = 0)P(Z_n = 0) \\
 &\quad + P(Z_{n+1} = 0 \mid Z_n \neq 0)P(Z_n \neq 0) \\
 &= 1 \times P(Z_n = 0) \\
 &\quad + P(Z_{n+1} = 0 \mid Z_n \neq 0)P(Z_n \neq 0) \\
 &\geq P(Z_n \neq 0) = d_n.
 \end{aligned}$$

Hence,

$$0 = d_0 \leq d_1 \leq \dots \leq 1$$

$\{d_n, n = 0, 1, \dots\}$ is an increasing and bound above sequence, then there exists $d = \lim_{n \rightarrow \infty} d_n$, $d \in [0, 1]$ and d is the probability of extinction of the process.

$$\begin{aligned}
 d_m &= P(Z_m = 0) \\
 &= \sum_j P(Z_m = 0 \mid Z_1 = j)P(Z_1 = j) \\
 &= \sum_j P(Z_m = 0 \mid Z_1 = j)p_j.
 \end{aligned}$$

We have

$$\begin{aligned}
 P(Z_m = 0 \mid Z_1 = j) &= P(Y_1 + Y_2 + \dots + Y_j = 0) \\
 &= P(Y_1 = 0)P(Y_2 = 0) \dots P(Y_j = 0) \\
 &= [P(Z_{m-1} = 0)]^j \\
 &= d_{m-1}^j.
 \end{aligned}$$

Hence,

$$d_m = \sum_j d_{m-1}^j p_j = H(d_{m-1}).$$

Since $d_m \rightarrow d$ when $m \rightarrow \infty$, then the value d satisfies the equation $s = H(s)$. Note that the solutions to equation $s = H(s)$ represent intersections of the graphs of $y = s$ and $y = H(s)$. $\square$

REMARK 3.

In the case $\mu > 1$, both extinction and survival are possible with non-zero probability. Then,

$$P(Z_1 = i \mid \text{extinct}) = \frac{P(Z_1 = i \text{ and extinct})}{P(\text{extinct})} = \frac{P(Y = i)d^i}{H(d)}.$$

Denote the last expression by $P(\hat{Y} = i)$. Observe that $E(\hat{Y}) < 1$. Hence, the distribution of Galton-Watson process with Y , conditioned on the extinction is the same as the distribution of the Galton-Watson process with $\hat{Y}$.

In the case $1 < \mu < \infty$,

$$\frac{Z_n}{\mu^n} \rightarrow X \quad \text{a.s. } n \rightarrow \infty$$

for some positive random variable X .

EXAMPLE 4. Consider the Galton-Watson process $\{Z_n, n = 0, 1, 2, \dots\}$ with offspring distribution $\{p_k\}$. Assume that $p_0 = 1/5, p_1 = 3/5$ and $p_2 = 1/5$. Then $\mu = E(Z_1) = 1$.

$$H(s) = \sum_k p_k s^k = \frac{1}{5} + \frac{3}{5}s + \frac{1}{5}s^2.$$

For instance, we find

$$P(Z_2 = 2) = \frac{H_2''(0)}{2!}$$

using

$$H_2(s) = H(H(s)) = \frac{1}{5} + \frac{3}{5}H(s) + \frac{1}{5}[H(s)]^2.$$

Since $\mu = 1$, by Theorem 2, then the probability of extinction is one.

EXAMPLE 5. Suppose Y_i 's are Binomial distributions with parameters 3 and p . Now, $E(Y) = 3p$. When $3p > 1$, by Theorem 2, the probability of extinction is the smallest solution of $s = H(s)$. Now,

$$s = H(s) \implies s = (q + ps)^3.$$

That is

$$p^3 s^3 + 3p^2 q s^2 + (3pq^2 - 1)s + q^3 = 0.$$

We know that, $s = 1$ is a solution, we rewrite as

$$(s - 1)(p^3 s^2 + (3p^2 q + p^3)s - q^3) = 0.$$

Suppose $p = 0.5$, then $s^2 + 4s - 1 = 0$ gives $d = 0.236$.

Lecture 3 Markov Branching Process

Let $Z(t)$ be number of particles at time t . Let

$$\delta_{1k} + a_k h + o(h), \quad k = 0, 1, \dots$$

represent the probability that a single particle will split producing k particles during a small time interval $(t, t+h)$ of length h . δ_{1k} denotes the Kronecker delta function. Assume that $a_1 \leq 0, a_k \geq 0$ for $k = 0, 2, 3, \dots$ and $\sum_{k=0}^{\infty} a_k = 0$. We further postulate that individual particles act independent of each other, always governed by the infinitesimal probabilities. Note that we are also assuming time homogeneity for the transition probabilities since a_k is not a function of the time at which the conversion or splitting occurs. Each particle lives a random length of time following an exponential distribution with mean $1/\lambda = a_0 + a_2 + a_3 + \dots$. On completion of i -th lifetime, it produces a random number D of descendants of like particles. The probability mass function of D is

$$P(D = k) = \frac{a_k}{a_0 + a_2 + a_3 + \dots}, \quad k = 0, 2, 3, \dots$$

The lifetime and progeny distribution of separate individuals are independent and identically distributed. Using the independence assumptions, we get

$$P(Z(t+h) = n+k-1 \mid Z(t) = n) = na_k h + o(h), \quad k = 0, 2, 3, \dots$$

and

$$P(Z(t+h) = n \mid Z(t) = n) = 1 - na_1 h + o(h).$$

Probability Generating Function

Let

$$P_{ij}(t) = P(Z(t+s) = j \mid Z(s) = i), \quad i, j = 0, 1, \dots, t, s \geq 0$$

and let

$$\phi(t; s) = \sum_{j=0}^{\infty} P_{1j}(t) s^j$$

the probability generating function for $P_{1j}(t)$.

THEOREM 1. *The probability generating function for $P_{1j}(t)$, $\phi(t; s)$, satisfies:*

$$\phi(t + v; s) = \phi(t; \phi(v; s)).$$

This is the continuous time analog of Theorem 7, in the case of discrete time branching processes.

Proof. Since individual particles act independently, we have the fundamental relation

$$\sum_j P_{ij}(t) s^j = \left[\sum_j P_{1j}(t) s^j \right]^i = [\phi(t; s)]^i.$$

The formula means that the population $Z(t; i)$ involving in time t from i initial parents is same, in the sense of probability, as the combined sum of i population each having one initial parents. Also, this formula characterizes and distinguishes branching processes from other continuous time Markov chains. By the time homogeneity, the Chapman - Kolmogorov equations take the form

$$P_{ij}(t + v) = \sum_k P_{ik}(t) P_{kj}(v).$$

Now

$$\begin{aligned} [\phi(t + v; s)]^i &= \sum_j P_{ij}(t + v) s^j \\ &= \sum_j \sum_k P_{ik}(t) P_{kj}(v) s^j \\ &= \sum_k P_{ik}(t) \sum_j P_{kj}(v) s^j \\ &= \sum_k P_{ik}(t) [\phi(v; s)]^k \\ &= [\phi(t; \phi(v; s))]^i. \end{aligned}$$

When $i = 1$, we obtain the result. □

THEOREM 2. *Let $u(s) = \sum_k a_k s^k$. Then $\phi(t; s)$ satisfies:*

$$\frac{\partial \phi(t; s)}{\partial t} = \frac{\partial \phi(t; s)}{\partial s} u(s)$$

$$\frac{\partial \phi(t; s)}{\partial t} = u(\phi(t; s))$$

with the initial condition

$$\phi(0; s) = \sum_j P_{1j}(0) s^j = s.$$

Proof. We have

$$\begin{aligned} \phi(h; s) &= \sum_j P_{1j}(h) s^j \\ &= \sum_j (\delta_{1j} + a_j h + o(h)) s^j \\ &= s + h \sum_j a_j s^j + o(h) \\ &= s + hu(s) + o(h). \end{aligned}$$

Now

$$\phi(t + h; s) = \phi(t; \phi(h; s)) = \phi(t; s + hu(s) + o(h)).$$

By Taylor's theorem, we expand the right-hand side with respect to the second variable

$$\phi(t + h; s) = \phi(t; s) + \frac{\partial \phi(t; s)}{\partial s} hu(s) + o(h).$$

Hence

$$\frac{\phi(t + h; s) - \phi(t; s)}{h} = \frac{\partial \phi(t; s)}{\partial s} u(s) + \frac{o(h)}{h}.$$

Taking $h \rightarrow 0^+$, we get

$$\frac{\partial \phi(t; s)}{\partial t} = \frac{\partial \phi(t; s)}{\partial s} u(s).$$

This is a partial differential equation for the function of two variables $\phi(t; s)$ with the initial condition

$$\phi(0; s) = \sum_j P_{1j}(0) s^j = s.$$

Proof of part 2:

Now consider,

$$\phi(v + h; s) = \phi(v; \phi(h; s)).$$

By Taylor's theorem, we get

$$\phi(v+h; s) = \phi(v; s) + hu(\phi(v; s)) + o(h).$$

Hence

$$\frac{\phi(v+h; s) - \phi(v; s)}{h} = u(\phi(v; s)) + \frac{o(h)}{h}.$$

Taking $h \rightarrow 0^+$, then substituting $v = t$, we get

$$\frac{\partial \phi(t; s)}{\partial t} = u(\phi(t; s)).$$

This is an ordinary differential equation with initial condition

$$\phi(0; s) = \sum_j P_{1j}(0)s^j = s.$$

□

Mean of $Z(t)$

Consider

$$\frac{\partial \phi(t; s)}{\partial t} = \frac{\partial \phi(t; s)}{\partial s} u(s).$$

By differentiating with respect to s and interchanging the order of differentiation on the left side, we get

$$\frac{\partial^2 \phi(t; s)}{\partial t \partial s} = u(s) \frac{\partial^2 \phi(t; s)}{\partial s^2} + u'(s) \frac{\partial \phi(t; s)}{\partial s}.$$

If $s = 1$, $u(1) = \sum_k a_k = 0$, and then

$$\frac{\partial m(t)}{\partial t} = u'(1)m(t)$$

where

$$m(t) = E[Z(t)] = \frac{\partial \phi(t; s)}{\partial s} \Big|_{s=1}.$$

But, since $Z(0) = 1$, then $m(0) = 1$ and the solution is

$$m(t) = e^{u'(1)t}.$$

Probability of Extinction

DEFINITION 3. *The extinction probability, q , is defined by*

$$q = \lim_{t \rightarrow \infty} P_{10}(t).$$

Now, we assume that $a_0 > 0$, otherwise extinction is impossible. It is enough to consider the case where the process starts with a single individual at time zero. In fact, with $s = 0$

$$P_{i0}(t) = [P_{10}(t)]^i.$$

Now, $P_{i0}(t)$ is non-decreasing in t ,

$$\begin{aligned} P_{i0}(t+v) &= [\phi(t+v; 0)]^i \\ &= [\phi(t; \phi(v; 0))]^i \\ &\geq [\phi(t; 0)]^i \\ &= P_{i0}(t). \end{aligned}$$

Derivation of Probability of Extinction

Let t_0 be a fixed positive number. Consider the discrete time process

$$Z(0), Z(t_0), Z(2t_0), \dots, Z(nt_0), \dots$$

where $Z(t)$ is the population size at time t .

Assume that, the population size at $t = 0$ is one. Since $Z(t)$ is assumed to be a Markov process, the discrete process $Y_n = Z(nt_0)$ will be a Markov chain which is also a discrete time branching process. By the hypothesis of homogeneity of the probability function of $Z(t)$ and

$$\sum_j P_{ij}(t) s^j = \left[\sum_j P_{1j}(t) s^j \right]^i = [\phi(t; s)]^i.$$

We obtain

$$\begin{aligned}
 \sum_k P(Y_{n+1} = k \mid Y_n = i) s^k &= E [S^{Y_{n+1}} \mid Y_n = i] \\
 &= E [S^{Z((n+1)t_0)} \mid Z(nt_0) = i] \\
 &= E [S^{Z(t_0)} \mid Z(0) = i] \\
 &= [\phi(t_0; s)]^i \\
 &= [E [S^{Z(t_0)} \mid Z(0) = 1]]^i \\
 &= [E [S^{Y_1} \mid Y_0 = 1]]^i.
 \end{aligned}$$

This shows that $\{Y_n, n = 0, 1, \dots\}$ is a branching process.

The probability generating function of the number of offspring of a single individual in this process is $\phi(t_0; s)$. By Theorem 2, we know that the probability of extinction for the Y_n process is the smallest non-negative root of the equation

$$\phi(t_0; s) = s.$$

But

$$\begin{aligned}
 P(Y_n = 0 \text{ for some } n) &= \lim_{n \rightarrow \infty} P(Y_n = 0) \\
 &= \lim_{n \rightarrow \infty} P(Z(nt_0) = 0) \\
 &= \lim_{t \rightarrow \infty} P(Z(t) = 0) \\
 &= q.
 \end{aligned}$$

The extinction probability q of the continuous time branching process $Z(t)$ is the smallest non-negative root of the equation

$$\phi(t_0; s) = s$$

where t_0 is any positive number.

Hence, we expect that we should also be able to calculate q from an equation that does not depend on time.

THEOREM 4. *The probability of extinction q is the smallest non-negative root of the*

equation

$$u(s) = 0$$

Hence, $q = 1$ if and only if $u'(1) \leq 0$.

Proof. Since q satisfies

$$\phi(t_0; s) = s$$

for any $t_0 > 0$, we have

$$\frac{\phi(v+h; s) - \phi(v; s)}{h} = u(\phi(v; s)) + \frac{o(h)}{h}.$$

If $s = q$, then

$$\frac{q - q}{h} = u(q) + \frac{o(h)}{h}.$$

Hence, for any $h > 0$,

$$0 = u(q) + \frac{o(h)}{h}$$

Taking $h \rightarrow 0^+$, we obtain

$$u(q) = 0.$$

Since

$$u''(s) = \sum_{k=1}^{\infty} a_k k(k-1)s^{k-2} \geq 0$$

$u(s)$ is convex in the interval $[0, 1]$. As $u(1) = 0$ and $u(0) = a_0 > 0$, $u(s)$ may have at most one zero in the interval $(0, 1)$. According to whether $u'(1) \leq 0$ or $u'(1) > 0$ holds, we have the cases $q = 1$ or $q < 1$ respectively. Note that $E[Z(t_0)] = E[Y]$ if and only if $u'(1) > 0$. This means that for the discrete time branching process $Z(nt_0)$, $n = 0, 1, 2, \dots$ ($t_0 > 0$ fixed), extinction occurs with probability less than one, and therefore the same is true for the process $Z(t)$. Here, the probability of extinction q is the smallest zero of $u(s)$ in $[0, 1]$. In a similar manner we conclude that, if $u'(1) \leq 0$, q must be equal to one. In either case q is the smallest non-negative root of $u(s) = 0$. $\square$

Limit Theorem

If $u'(1) = 0$ and $u''(1) < \infty$, then

$$P(Z(t) > 0 \mid Z(0) = 1) \sim \frac{2}{tu''(t)}, \quad t \rightarrow \infty$$

and

$$\lim_{t \rightarrow \infty} P\left(\frac{2Z(t)}{tu''(1)} > \lambda \mid Z(t) \neq 0\right) = e^{-\lambda}, \quad \lambda > 0.$$

If $u'(1) > 0$ and $u''(1) < \infty$, then

$$\frac{Z(t)}{e^{tu'(1)}}$$

has a limit distribution as $t \rightarrow \infty$.

Some Other Important Branching Processes

1. Bellman-Harris Processes

Consider a classical branching process in which progeny are born at the moment of the parents death. Let $Z(t)$ be the number of particles alive at time t . The distribution of particle lifetime τ is an arbitrary non-negative random variable, the resulting process is called an “age-dependent” or Bellman - Harris process. Assume that all particles reproduce and die independently of each other. This model generalizes the birth and death process in two respects: first, the lifespan of individual particles need not have the exponential distribution, and second, more than one particle can born. The process $Z(t)$ is not a Markov process and its analysis is usually done by using renewal theory.

2. Bellman-Harris Processes with Disasters

Consider the population model which follows a Bellman - Harris process. At random times, disasters beset the population and each particle alive at the time of the disaster survives with probability p . The survival of any particle is assume to be independent of the survival of any other particle. Measure of interest is limiting behavior when extinction does not occur.

3. Bellman - Harris Processes with Immigration

Consider a Bellman - Harris process $\{Z(t), t \geq 0\}$ where $Z(t)$ denotes the population size at time t . In addition, we allow the sudden appearance into the system of newly born particles called immigrants. Immigrants are assumed

to arrive in groups of various sizes with the probability of n immigrants in a group immigrating at time t given by $p_n(t)$, $n = 1, 2, \dots$. Once these particles arrive, they reproduce and die according to the Bellman-Harris process. Measures of interest are the mean, the limiting distribution and the asymptotic behavior. This branching process is widely used to describe growth and decay of biological populations.

References

1. G. Sankaranarayanan, "Branching Processes and Its Estimation Theory", Wiley Eastern Ltd, New Delhi, 1989.
2. K.B. Athreya and P.E. Ney, "Branching Processes", Dover Publications, New York, 2004.
3. K.B. Athreya, P.R. Parthasarathy and G Sankaranarayanan, "Supercritical Age-Dependent Branching Processes with Immigration", Journal of Applied Probability, Vol. 11, No. 4, pp. 695 - 702, 1974.
4. P.R. Parthasarathy and A. Krishnamoorthy, "Compounding Bellman - Harris allowing Immigration", Mathematical Biosciences, Vol. 37, pp. 105 - 111, 1977.
5. J Medhi, "Stochastic Processes", 3rd edition, New Age International Publishers, 2009.
6. T.E. Harris, "Theory of Branching Processes", Dover Publications, New York, 1989.

Module 9: Stationary Processes

Lecture 1 Stationary Processes

1 Introduction

A stationary process is a stochastic process whose joint probability distribution does not change when shifted in time or space. For a stationary process, mean and variance, if they exist, do not change over time or position.

Stationarity is a key concept in time series analysis as it allows powerful techniques for modeling and forecasting to be developed. Time series is a set of data ordered in time usually recorded at regular time intervals of time. In probability theory a time series is a collection of random variable $\{X(t), t \geq 0\}$ indexed by time.

One of the main features of time series is the interdependency of observation over time. This interdependency needs to be accounted in the time series data modeling to improve temporal behavior and forecast of future moment. So stationarity is used as a tool in time series analysis when raw data is often transformed to become stationary.

2 Important Definitions

DEFINITION 1. Mean function It is defined as $m(t) = E(X(t))$, which may or may not be dependent on t .

DEFINITION 2. Second Order Process A stochastic process is called a second order process if it's second order moment is finite for all t i.e., $E(X^2(t)) < \infty$.

DEFINITION 3. Covariance function It is denoted by $C(s,t)$ given by

$$\begin{aligned} C(s,t) &= \text{cov}(X(s), X(t)) \\ &= E(X(s)X(t)) - E(X(s))E(X(t)) \end{aligned}$$

A Stochastic process has to be a second order process for covariance function to exist. The Covariance function satisfies the following properties:

1. $C(s,t) = C(t,s) \forall t, s \in T$

2. Using Schwarz inequality

$$C(s, t) \leq \sqrt{C(s, s)C(t, t)}$$

3. It is non negative definite i.e, for a set of real numbers $a_1, a_2, a_3, \dots, a_n$ and $t_i \in T$

$$\sum_{j=1}^n \sum_{k=1}^n a_j a_k C(t_j, t_k) = E[\sum a_j X(t_j)^2] \geq 0$$

4. Sum and product also covariance functions.

DEFINITION 4. Auto-Correlation function It is denoted by $R(s, t)$, given by:

$$R(s, t) = \frac{E[X(s)X(t)] - E[X(s)]E[X(t)]}{\sqrt{\text{var}(X(s))}\sqrt{\text{var}(X(t))}}$$

Now assuming, $R(s, t)$ depends only on $|t - s|$, $E(X(t)) = \mu$ and $\text{Var}(X(t)) = \sigma^2$ we get,

$$R(\tau) = \frac{(E(X(t)) - \mu)(E(X(t + \tau)) - \mu)}{\sigma^2}$$

DEFINITION 5. Independent Increments If for every $t_1 < t_2 < \dots < t_n$ $X(t_2) - X(t_1), X(t_3) - X(t_2), \dots, X(t_n) - X(t_{n-1})$ are mutually independent random variables $\forall n$.

DEFINITION 6. Ergodic Property It is the time average of a function along a realization or sample exist almost everywhere and is related to the space average. That means whenever the stochastic process is ergodic the time average is same for all almost initial points i.e., the process evolved for a longer time forgets its initial state.

3 Strict Sense and Wide Sense Stationary Process

DEFINITION 7. Strict Sense Stationary If for arbitrary $t_1 < t_2 < \dots < t_n$ the joint distribution of the random vectors $\{X(t_1), X(t_2), \dots, X(t_n)\}$ and $\{X(t_{1+h}), X(t_{2+h}), \dots, X(t_{n+h})\}$ is same $\forall h$ then the stochastic process $\{X(t), t \in T\}$ is said to be strict sense stationary of order n . If the above definition holds for every integer n then the process is called a strict sense stationary Process.

DEFINITION 8. Wide Sense Stationary Process A Stochastic process is said to be wide sense stationary or weakly stationary or covariance stationary if the following properties hold-

- $m(t) = E[X(t)]$ is independent of t .
- $E[X^2(t)] < \infty$.
- $c(s, t)$ is a function of the time difference $|t - s|$ only i.e.,
 $c(s, t) = f(t - s) \forall t, s$.

Also neither a strict sense stationary process imply the wide sense stationary nor wide sense stationary imply the strict sense stationary process.

4 Examples of Stationary Process

EXAMPLE 9. Let X_i be iid $B(1, p)$ random variables. Then for the stochastic process $\{X_i, i = 1, 2, \dots\}$ the following can be seen $m(i) = E[X_i] = p$
 $E[X_i^2] = p$

$$\begin{aligned} c(i, j) &= E[X_i X_j] - E[X_i]E[X_j] \\ &= \begin{cases} 0, & i \neq j \\ p(1 - p), & i = j. \end{cases} \end{aligned}$$

Since all three properties of wide sense stationary process are satisfied and hence the above process is a wide sense stationary process.

Now we consider the joint distribution of the vectors $\{X_{i_1}, X_{i_2}, \dots, X_{i_n}\}$ and $\{X_{i_1+h}, X_{i_2+h}, \dots, X_{i_n+h}\}$.

Since the random variables X_i are iid random variables, therefore the joint distribution of the two vectors will be the product of the distribution of the random variables and hence the joint distribution for the 2 random vectors will be the same. Hence the process $\{X_i, i = 1, 2, \dots\}$ is strict sense stationary process. So in this case the process $\{X_i, i = 1, 2, \dots\}$ is strict sense as well as wide sense stationary process.

EXAMPLE 10. Let $\{X(t), t \in T\}$ be a strict sense stationary process with finite second order moment and let $Y(t) = a + bt + X(t)$, so now we check whether $\{Y(t), t \in$

$T\}$ is a wide sense and strict sense stationary process.

The function $m(t) = E[Y(t)]$ will be a function of t and hence $\{Y(t), t \in T\}$ is not a wide sense stationary process. Similarly it can be verified that the process $\{Y(t), t \in T\}$ is not a strict sense stationary process.

EXAMPLE 11. Let $\{X_n, n = 1, 2, \dots\}$ be uncorrelated random variables with $E[X_n] = k$, where k is a constant and

$$E[X_m X_n] = \begin{cases} \sigma^2 & \text{if } m = n \\ 0 & \text{if } m \neq n \end{cases}$$

It can be verified that all the three conditions of wide sense stationary will be satisfied and hence it's a wide sense stationary process. This process is also called the white noise process.

If for a stochastic process $\{X(t), t \in T\}$ we assume it to have the ergodic property and be a wide sense stationary process then the mean of this process can be estimated from the time average i.e., $\widehat{\mu}_T = \frac{1}{T} \int_{-T}^T X(t) dt$

EXAMPLE 12. Consider the process $\{X(t), t \geq 0\}$, $X(t)$ is defined as:

$$X(t) = A \cos(\theta t) + B \sin(\theta t),$$

where, A and B are known to be uncorrelated random variables with mean and variance as 0 and σ^2 respectively and θ is a constant.

This, $X(t)$ is a random variable for fixed t

Now we check whether the given process is a wide sense stationary process:
For a fixed t , as mean of A and B is 0, $E(X(t)) = 0$.

$$C(s, t) = E(X(t)X(s)) - E(X(t))E(X(s)) = \sigma^2 \cos(\theta(t - s)),$$

which is a function of $(t - s)$.

$$E(X^2(t)) = \sigma^2.$$

From the above three conditions it can be concluded that the given process is wide sense stationary.

EXAMPLE 13. This comes under time series forecasting.

Suppose we are given time series data $X_0, X_1, \dots, X_n$.

Aim is to forecast for X_{n+1} with the given $n+1$ data. By some function f we can predict X_{n+1} such that the mean square error is minimum. It turns out that the function f is:

$$E[X_{n+1}/X_0 = x_0, X_1 = x_1, \dots, X_n = x_n].$$

If we assume that X_i 's are normally distributed then the above conditional expectation will be a linear function of the data of $X_0, X_1, \dots, X_n$.

In general the above conditional expectation can be estimated using the estimator:

$$\widehat{X_{n+1}} = a_0x_0 + a_1x_1 + \dots + a_n * x_n$$

where $a_0, a_1, \dots, a_n$ are parameters which are to be estimated. Now we can assume that the given time series satisfies the wide sense stationary and hence the parameters are invariant with time and therefore they can be estimated from the data itself.

EXAMPLE 14. In this example we consider the Gaussian process, $\{W(t), t \geq 0\}$ which is a stochastic process wherein every random variable is normally distributed, i.e., $W(t) \sim$ normal distributed or we say that the joint distribution $\{W(t_1), W(t_2), \dots, W(t_n)\} \sim$ multivariate normal distribution.

The distribution of the above random vector can be completely determined by the mean vector

$$E[W(t_1), W(t_2), \dots, W(t_n)],$$

and the covariance matrix, $[C_{ij}]$, where $C_{ij} = \text{cov}(X_i, X_j)$, $i = 1, 2, 3, \dots, n$, $j = 1, 2, 3, \dots, n$

Now we determine whether the above process is the covariance or wide sense stationary.

If $\{W(t), t \geq 0\}$ satisfies the following 3 properties, i.e.,

1. $E[W(t)]$ is independent of t .
2. $E[W^2(t)] < \infty$.
3. $\text{cov}(W_{t_i}, W_{t_j})$ is a function of $|t_i - t_j|$.

then the given stochastic process will be wide sense stationary process.

Also then we can conclude that it will be a strict sense stationary process because it depends only on the joint distribution and mean vector and the covariance matrix.

Lecture 2 Autoregressive Processes

1 Moving Average Process

Pure Random process

Pure Random Process or White Noise Process: is a random process $\{X_t, t \geq 0\}$ which has:

$$E[X_t] = m(\text{constant}), \text{ i.e., the mean is independent of } t.$$

$$E[X_t X_{t+k}] = \begin{cases} \sigma^2 & \text{if } k = 0 \\ 0 & \text{if } k \neq 0 \end{cases}$$

It can be verified that the above process is a covariance stationary process.

Moving Average Process

A Moving Average Process $\{X(t), t \geq 0\}$ is a stochastic process satisfying certain properties. It is represented as:

$$X_t = a_0 e_t + a_1 e_{t-1} + \dots + a_h e_{t-h}$$

where, a_i 's are real constants and e_k is a pure random process with mean 0 and variance σ^2 .

This collection $\{X(t), t \geq 0\}$ is known as a moving average process.

When the constant $a_h \neq 0$ then the above process is called a moving average ("MA" in short) of order h .

Now let $C_k = E[X_t X_{t+k}]$ be the correlation coefficient, then

$$C_k = \begin{cases} (a_0 a_k + a_1 a_{k+1} + \dots + a_{h-k} a_h) \sigma^2 & \text{if } k \leq h \\ 0 & \text{if } k > h \end{cases}$$

Now we define the covariance coefficient ρ_k ,

$$\begin{aligned}\rho_k &= \frac{C_k}{C_0} \\ &= \begin{cases} \frac{a_0 a_k + a_1 a_{k+1} + \dots + a_{h-k} a_h}{a_0^2 + a_1^2 + \dots + a_h^2}, & k \leq h \\ 0, & k > h. \end{cases}\end{aligned}$$

It can be verified that the moving average process is a covariance stationary or wide sense stationary process.

First Order Markov Process

One very important Moving Average Process is the first order Markov process which is defined as:

$$X_t + a_t X_{t-1} = e_t, \quad |a_t| < 1$$

where, $\{e_t, t \in T\}$ is a pure random process with mean 0 and variance 1.

Now X_t can be written as

$$X_t = \sum_{k=0}^{\infty} \rho_k e_{t-k}$$

where, $\rho_k = a^{|k|}$ for all k .

Thus the Markov process of first order can be related to a Moving Average Process of infinite order.

2 Autoregressive Process

Autoregressive process(AR)

A stochastic process $\{X_t, t \in T\}$ of the form:

$$X_t + b_1 X_{t-1} + b_2 X_{t-2} + \dots + b_h X_{t-h} = e_t$$

where, e_t is a pure random process with mean 0 and b_i 's are real constants with $b_h \neq 0$, then the corresponding stochastic process is an AR process of order h .

AR process of infinite order: When $X_t = \sum_{r=0}^{\infty} b_r e_{t-r}$, then the process $\{X_t, t \in T\}$ is an AR process of infinite order.

Now we consider a special case of Autoregressive processes i.e. an Autoregressive

process of order 2.

An Autoregressive process of order 2 will be of the form:

$$X_t + b_1X_{t-1} + b_2X_{t-2} = e_t$$

This process is known as the Yule Process.

Autoregressive Moving Average Process

A stochastic process is called an Autoregressive Moving Average Process (ARMA) if it is of the form:

$$\sum_{r=0}^p b_r X_{t-r} = \sum_{s=0}^q a_s e_{t-s}$$

where, $\{e_t, t \in T\}$ is a pure random process with mean 0 and b_i 's are real constants with $b_0 = 1$.

The above process is called an ARMA process of order (p, q) .

If p and q are ∞ then it will be called ARMA process of order infinity.

Power Spectrum

The autoregressive and moving average process are useful in the study of power spectrum. For a wide sense stationary process we define:

Covariance function : $C_k = E[X_t X_{t+k}] - E[X_t]E[X_{t+k}]$

Correlation function:

$$\begin{aligned} \rho_k &= \frac{C_k}{\sqrt{\text{Var}(X_t)}\sqrt{\text{Var}(X_{t+k})}} \\ &= \int_{-\pi}^{\pi} e^{ikw} dF_1(w) \end{aligned}$$

where, $F_1(w)$ is cdf called integrated spectrum and whenever it's absolutely continuous then the derivative can be written in terms of the density function.

Inverse Representation of ρ_k :

$$f_1(w) = \frac{1}{2\pi} \sum_{k=-\infty}^{\infty} \rho_k e^{-ikw}$$

i.e., knowing the correlation coefficient we can determine the density function.

Representation of C_k :

$$C_k = \int_{-\infty}^{\infty} e^{ikw} dF(w)$$

where, $dF(w) = C_0 dF_1(w)$ and whenever $F(w)$ is an absolutely continuous function then the inverse representation of C_k can be given:

$$f(w) = \frac{1}{2\pi} \sum_{k=-\infty}^{\infty} C_k e^{-ikw}.$$

Hence, $f(w)$ is called the spectral density function and $f_1(w)$ is called the normalized spectral density function.

The advantage of studying power spectrum is that when we have a stochastic process $\{X(t), t \in T\}$ having the wide sense stationary property in time domain, the entities C_k, ρ_k can be found but these are not simple to evaluate.

On the other hand, in frequency domain, for the same stochastic process $\{X(t), t \in T\}$ we have the spectral density and the normalized spectral density function and using the inverse relations we can find C_k and ρ_k . This is called the spectrum study.

EXAMPLE 1. Consider the white noise process, $\{X(t), t \in T\}$. Assume $E[X_t] = 0$ and

$$E[X_r X_s] = \begin{cases} \sigma^2 & \text{if } r = s \\ 0 & \text{if } r \neq s \end{cases}$$

Now we try to find few entities in time domain:

$$\begin{aligned} C_k &= 0, \text{ for } k \neq 0 \\ C_0 &= \sigma^2, \rho_k = 0, \text{ for } k \neq 0 \\ \rho_0 &= \frac{C_0}{\sigma^2} = 1 \end{aligned}$$

Now we study the same in frequency domain:

Spectral density: $f(w) = \frac{1}{2\pi} \sum_{k=-\infty}^{\infty} C_k e^{-ikw} = \frac{\sigma^2}{2\pi}$, for $-\pi \leq w \leq \pi$

$$F(w) = \int f(w) dw = \begin{cases} \frac{\sigma^2 w}{2\pi} & , -\pi \leq w \leq \pi \\ 0 & , \text{otherwise} \end{cases}$$

We can also evaluate the derivative:

$$dF(w) = \frac{\sigma^2}{2\pi} dw$$

Now we show that using the inverse relation we can find C_k given $f(w)$.

$$C_k = \int_{-\infty}^{\infty} e^{ikw} dF(w) = \begin{cases} \sigma^2 & \text{if } k = 0 \\ 0 & \text{if } k \neq 0 \end{cases}$$