

The Lecture Contains:

- ☰ Basic Introduction
- ☰ Basic Probability space, sample space concepts and order of a Stochastic Process
- ☰ Examples
- ☰ Definition of Stochastic Process
- ☰ Marginal Distributions
- ☰ Moments
- ☰ Gaussian Process
- ☰ Random Walks
- ☰ Markov Chain
- ☰ Further definition of Markov Chain
- ☰ Transition probability and transition matrix

◀ Previous Next ▶

Basic Introduction

We are all aware that in applied statistics after we collect the empirical data, a theoretical probability distribution is fitted in order to extract more information from the data. If the fit is good (which depends on some tests), then the properties of the set of data can be approximated by that of the theoretical distribution. In a similar way, a real life process may have the characteristics of a **stochastic process** (what we mean by a **stochastic process** will be made clear in due course of time), and our aim is to understand the underlying theoretical stochastic processes which would fit the practical data to the maximum possible extent. Hence a good knowledge of the characteristics and behaviour of stochastic processes is required to understand many real life situations.

In general there are examples where probability models are suitable and very often a better way of representation of the probability model would be to consider a collection or family of random variables (r.v.'s) that are indexed by a parameter such as **time** or **space**. These models are what we define as **stochastic process** or random or chance process.



Module 1: Concepts of Random walks, Markov Chains, Markov Processes

Lecture 1: Introduction to Stochastic Process

Thus a **stochastic process** is a family of random variables (r.v's) $\{X(t), t \in T\}$ indexed by the **parameter** $t \in T$. The values assumed by the stochastic process are called the **states** and the set of all possible values is called **state space**. On the other hand the set of possible values of the indexing parameter is called the **parameter space**, which can be either discrete or continuous. When the indexing parameters are discrete we denote it by n and the stochastic process as $\{X_n, n = 0, 1, 2, \dots\}$, and this process is what we call a **stochastic sequence**. In most physical problems time, t , is the natural index parameter. Other kinds of parameters such as space may also arise, e.g., number of defects on a sheet of metal which can be considered as a stochastic process with the area as the index parameter. But since time is the parameter used in majority of problems, we will use the expression **time parameter** in a generic sense.

Remember, like the **parameter space**, the **state space** may also be either discrete or continuous, hence the stochastic process may be any one of the following four (4) types shown in Table 1.1.

Table 1.1: Different types of Stochastic Processes

SNo.	Parameter Space	State Space	Combination	Examples
1	Discrete	Discrete	{Discrete, Discrete}	Markov Chain
2	Discrete	Continuous	{Discrete, Continuous}	Markov Process
3	Continuous	Discrete	{Continuous, Discrete}	
4	Continuous	Continuous	{Continuous, Continuous}	Brownian Motion

Thus the nomenclature for denoting the stochastic processes # 1 and # 2 (Table 1.1) is usually, $\{X_n, n = 0, 1, 2, \dots\}$ while for # 3 and # 4 (Table 1.1), it is $\{X(t), t \in T\}$, but in general one uses the latter representation, i.e., $\{X(t), t \in T\}$, to represent all the **four** types of stochastic processes, such that, depending on the domain space of t , one can refer whether the process is discrete or continuous. For example we can have $T = \{0, 1, 2, \dots\}$ or $T = \{\dots, -2, -1, 0, +1, +2, \dots\}$ or $T = \{-\infty, +\infty\}$ or $T = [0, +\infty)$, etc. One should remember that the main emphasis for this lecture series/course would be on **Markov chain** and **Markov process**, hence we will study # 1 and # 2 in details.



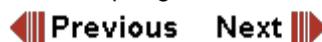
Basic Probability space, sample space concepts and order of a Stochastic Process

We all know that for a given $t \in T$, $X(t)$ is a random variable (r.v) on some probability space, denoted by (Ω, \mathcal{A}, P) , where Ω is the sample space, \mathcal{A} is the σ field of subsets of Ω which generates the events, and $T = [0, +\infty)$ the probability defined on \mathcal{A} . Thus one can view a stochastic process, $X(\omega, t)$ as a family of random variables (r.v's), $\{X(\cdot, t), t \in T\}$ on (Ω, \mathcal{A}, P) . Hence every fixed value of argument $\omega \in \Omega$, i.e., every sample point, $X(\omega, t)$ depends only on t and is simply a function of one real argument, so that for each fixed value of t , $X(\cdot, t)$ is a real valued function defined on Ω . Furthermore for a fixed t , $X(\cdot, t)$ is a random variable (r.v) on (Ω, \mathcal{A}, P) and is a function on Ω . On the other hand for fixed, ω , $X(\omega, \cdot)$ is a function of t , which represents a possible observation of the stochastic process, and this function $X(\omega, \cdot)$ is said to be a realization or a sample function of the process. When several quantities, $X_1(t), X_2(t), \dots, X_n(t)$ are required for the complete description of the state of the system at a fixed parameter point t , a generalization can be made accordingly.

For simplicity we will restrict ourselves to a single quantity $X(t)$ and T to be one dimension. Thus for a given value of time parameter, t , of the stochastic process, $X(t)$, it is a simple random variable (r.v) and its probability distribution can be obtained as for any other random variable (r.v). But when t varies in a space T , the information about the process $X(t)$ is not provided by a simple distribution for a given t , but one needs the *joint distribution* of the basic random variables (r.v's) of the family $\{X(t), t \in T\}$ to get the complete information about the process. Obtaining such a joint distribution is impossible if the membership of the family is infinite in number. It then seems reasonable to assume that the behavior of the process can be obtained by studying it at discrete sets of points and accordingly a joint distribution function defined at these points seems reasonable.

So if (t_1, t_2, \dots, t_n) with $t_1 < t_2 < \dots < t_n$ be such a discrete sets of points within T , then the joint distribution of the process $X(t)$ at these points can be defined as $P\{X(t_1) \leq x_1, X(t_2) \leq x_2, \dots, X(t_n) \leq x_n\}$, and this distribution has the simplest form when the random variables (r.v's) are **independent**.

The study of stochastic process does reduce to the study of simple random variable (r.v). However, in most practical cases, we are faced with the problem of assuming some sort of dependence among the random variables (r.v's). We shall restrict to the simplest type of dependence, called the **first order dependence** or **Markov dependence**, which may be understood with the example given below.



Example 1.1

Consider we throw an unbiased dice n number of times and note the number of times different faces, i.e., $i = 1, 2, 3, 4, 5, 6$ occur, where n can be **100, 1000, 10000**, etc. Now if a theoretical distribution fits this experimental data set, then the expected frequencies/relative frequencies/probability and observed expected frequencies/relative frequencies/probability are compared. In case the fit is good we use the properties of the theoretical distribution to explain the characteristics of the experimental data. Remember in this example we have a static process. Let us illustrate this example in more details, i.e., we have X_n which denotes the random variable (r.v) associated with the n^{th} throw of the dice such that:

$$X_n = \begin{cases} 1 & \text{with probability } \frac{1}{6} \\ 2 & \text{with probability } \frac{1}{6} \\ 3 & \text{with probability } \frac{1}{6} \\ 4 & \text{with probability } \frac{1}{6} \\ 5 & \text{with probability } \frac{1}{6} \\ 6 & \text{with probability } \frac{1}{6} \end{cases} \quad \forall n = 1, 2, 3, \dots$$

When we have a sequence of X_n we have a stochastic process which is denoted by $\{X_n; n = 1, 2, 3, \dots\}$. For the interested reader we would like to mention that this is a Bernoulli process, denoted by $\{X_n; n = 1, 2, 3, \dots\}$. Stated simply, what we have is a collection or family of random variables (r.v's) which is a stochastic process. But in general to be more specific we have the concept of time also, hence it is a dynamic process, such that a general stochastic process is denoted by $X(t)$, $t \in T$, where T is the parameter space.

Example 1.2

Consider a PBX/EPBAX (Electronic Public Branch Automatic Exchange) where we denote $X(t)$, $t \in T$, as the number of incoming calls arriving during a particular time c . Individually $X(t=1), X(t=1.25), \dots$ are the realized values of a random variable (r.v). In case one is interested to know how X changes as time, t , changes then we look into a dynamic process, which can be termed as a stochastic process.

In both these examples one would also be interested to know about the joint distribution of the members of this family of random variables (r.v), and how many random variables (r.v's) are there, that depend on T . One should remember that this parameter space T may be discrete/denumerably infinite or non denumerable infinite. Few examples for discrete parameter space are $T = (0, 1, 2, \dots)$, $T = (0, \pm 1, \pm 2, \dots)$. When T is discrete we usually use n to denote the parameter, e.g., $\{X_n, n = 1, 2, 3, \dots\}$ and call this a **stochastic sequence**. When T is continuous, say for example, $T = (0, \infty)$, $T = (-\infty, +\infty)$, we use X_t to denote the parameter, e.g., $\{X(t), t \in T\}$.

Measure theoretic approach

Before going into more detail about stochastic process we give here a very brief preview of measure theoretical approach which is relevant for probability and statistics. Let X be a random variable (r.v) on (Ω, \mathcal{B}, P) , where, (i) Ω is the sample space, (ii) \mathcal{B} is the sigma field, i.e., very simply stated it is the subsets of Ω which defines a set of events having certain properties which is predefined and common, (iii) P be the probability which is defined on \mathcal{B} , or in other words it is a function mapping from \mathcal{B} to $[0, 1]$.

Example 1.3

Consider the next example of tossing an unbiased coin, where the outcomes are either, a head, H , or tail, T . The associated sample space for this example is $\Omega = (H, T)$, while X the random variable (r.v) is denoted as $X(H) = 1$ and $X(T) = 0$. There is nothing sacrosanct of the fact that $X(H) = 1$ or $X(T) = 0$. Pictorially the mapping, which denotes the probability function, is denoted as shown in Figure 1.1.

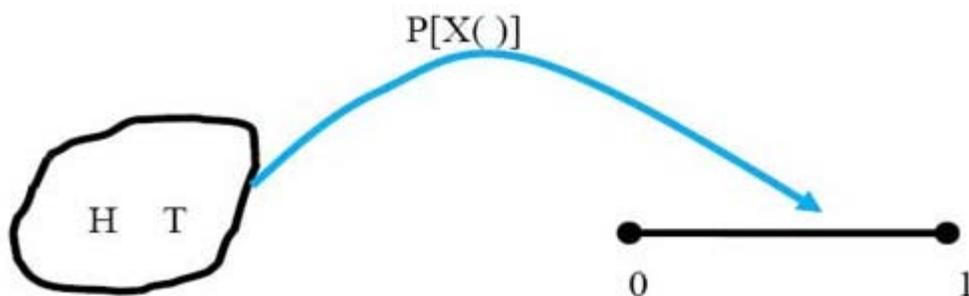


Figure 1.1: Pictorial representation of probability function mapping for the one dimension case

For the case when we toss two unbiased coins, X denotes the number of heads/tails appearing. So if we use the same nomenclature for the random variable (r.v), then $X(H, H) = 2$, $X(H, T) = 1$, $X(T, H) = 1$ and $X(T, T) = 0$ and this simple concept may be illustrated as shown in Figure 1.2.

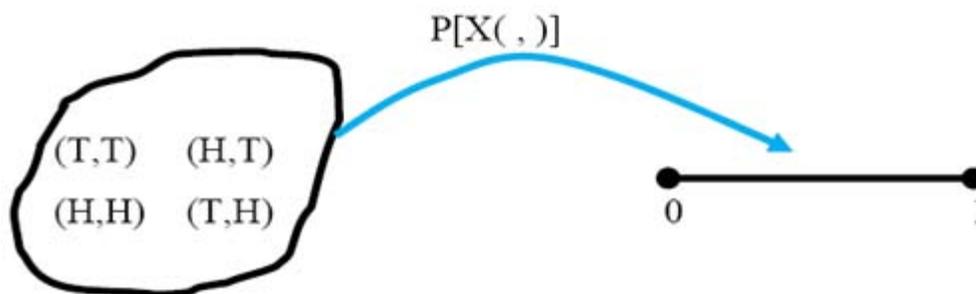


Figure 1.1: Pictorial representation of probability function mapping for the two dimension case

Hence this concept may be extended to the case of n dimensions also, which we omit for the reader to read and clear their concepts from a good book in probability and measure theory.

Module 1: Concepts of Random walks, Markov Chains, Markov Processes

Lecture 1: Introduction to Stochastic Process

We already know that if X is the random variable (r.v), such that $F(x) = P\{X \leq x\}$, $x \in \mathbb{R}$, then $F(x)$ is the distribution function of the random variable (r.v) X . Consider a finite collection of random variables (r.v's) in $\{X(t), t \in T\}$, where $t_1 < t_2 < \dots < t_n \in T$. Then the joint distribution of $\{X(t_1), X(t_2), \dots, X(t_n)\}$ is given by $F(x_1, x_2, \dots, x_n) = P\{X(t_1) \leq x_1, X(t_2) \leq x_2, \dots, X(t_n) \leq x_n\}$. In case $X(t_i)$ s are **independent**, we have

$$F(x_1, x_2, \dots, x_n) = F(x_1) \times F(x_2) \times \dots \times F(x_n) = P\{X(t_1)\} \times P\{X(t_2)\} \times \dots \times P\{X(t_n)\}.$$

In case they are independent and identically distributed, then the following is true $F(x_1, x_2, \dots, x_n) = \prod_{i=1}^n F(x_i) = \prod_{i=1}^n P\{X(t_i)\}$. For example $X \sim B(m, p)$ and $Y \sim B(n, p)$ are independent but not identical, while $X \sim B(m, p)$ and $Y \sim B(m, p)$ are independent and identical, i.e., *i.i.d.*



Definition of Stochastic Process

Defintion 1

As already mentioned a **stochastic process** is a function of two parameters which are from the **sample space** and the **parameter space** respectively, i.e., $X(\omega, t)$, where $\omega \in \Omega$ and $t \in T$, such that the general nomenclature of denoting a stochastic process is $X(\cdot, t), t \in T$. When the realized value is observed or achieved then we say that the realization of the stochastic process has been observed and it is denoted by $X(\omega, \cdot)$.

Defintion 2

A Stochastic process is a family of random (r.v.) and they are usually indexed/identifies by say t , i.e., its representation is $\{X(t): t \in T\}$. Here one should remember that T is some index set in a way such that all the elements of t are elements of T . A realization (or a sample function) of a stochastic process $\{X(t): t \in T\}$ is an assignment to each $t \in T$ of a possible value of $X(t)$. Here T could be finite, countable or uncountable finite. Let us illustrate this with a simple example.

◀ Previous Next ▶

Example 1.4

Consider you toss a coin n number of times. Denote the realization of head as $H = 1$, while that of a tail is denoted as $T = 0$. Then the stochastic process is given as $\{X_1, \dots, X_n\}$ where $T = \{1, \dots, n\}$. Diagrammatically it is shown below

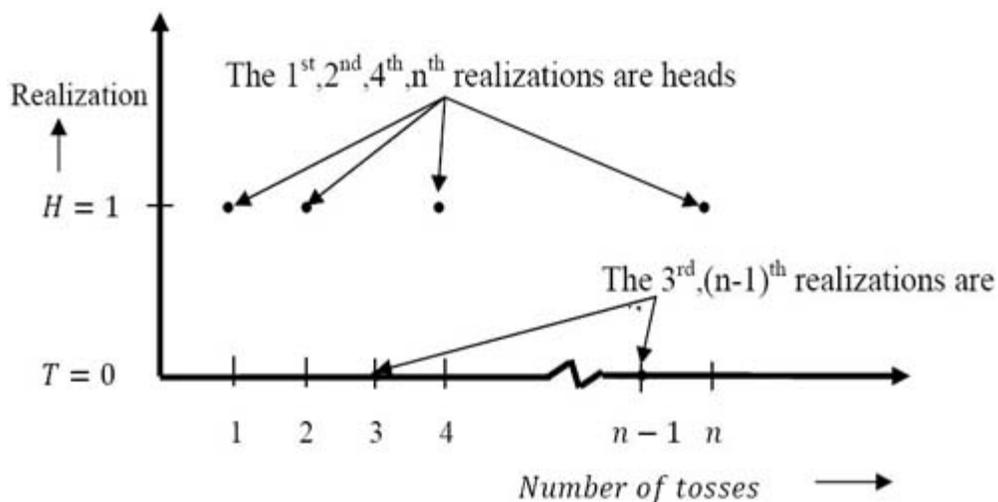


Figure 1.3: Stochastic process depicting the occurrence of head (H) or tail (T) when a coin is flipped number of times

Now if f is the function which maps the relation between the sample space and the real line, $[0, 1]$, then we denote it by $f: \Omega \rightarrow \mathbb{R}$, i.e., $f(\cdot): \Omega \rightarrow \mathbb{R}$, i.e., $f(\omega): \Omega \in \mathbb{R}$. Pictorially it is depicted as shown in Figure 1.4 and Figure 1.5.

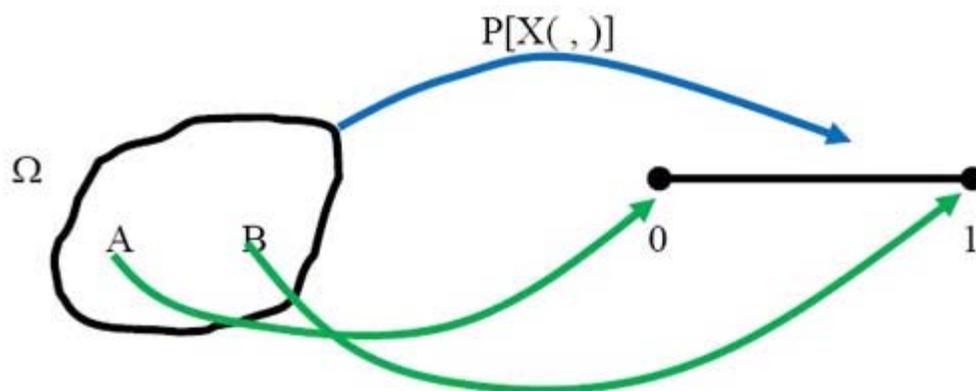


Figure 1.4: Probability mapping for a stochastic process from Ω to $[0, 1]$

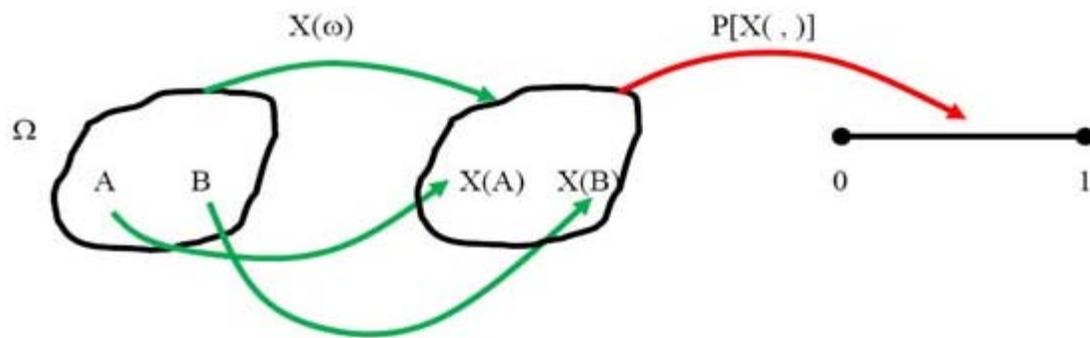


Figure 1.5: Probability mapping for a stochastic process from Ω to $X(\cdot)$ and then to \mathbb{R}

◀ Previous Next ▶

This concept can be generalized for the multivariate case also, i.e., Figure 1.6 and Figure 1.7.

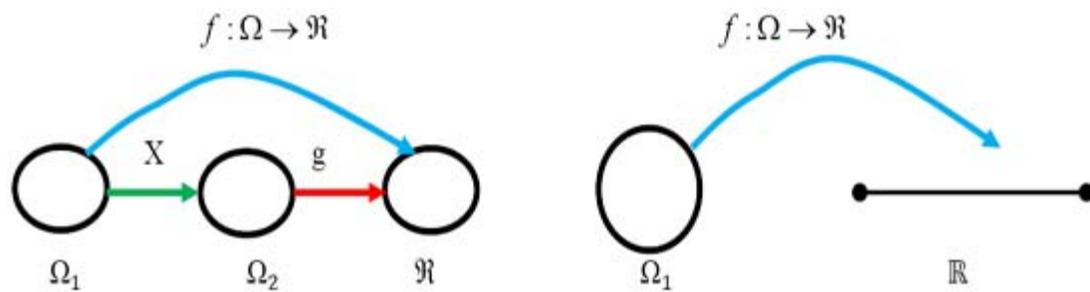


Figure 1.6: Probability mapping from sample space to real line considering $n = 1$

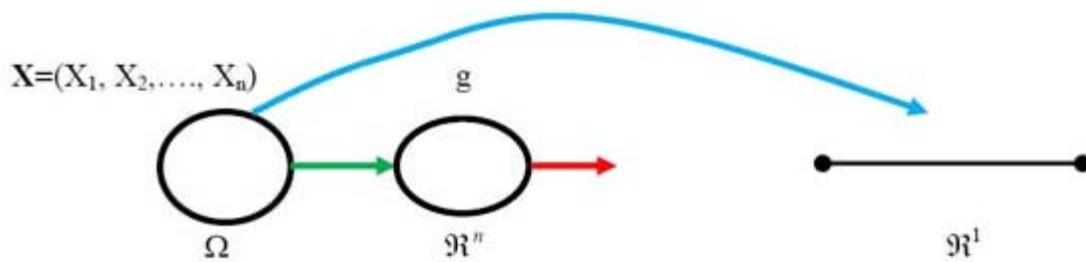


Figure 1.7: Probability mapping from sample space to real line considering $n \geq 2$

For the ease of understanding we have a look at the bivariate case mapping which is shown in Figure 1.8.

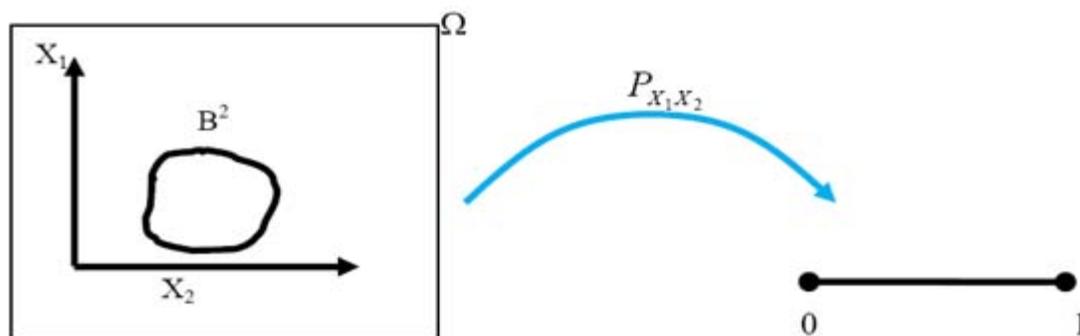


Figure 1.8: Bivariate probability distribution functional mapping

If one considers the tossing of an unbiased coin then $X(\omega)$ is either $X(H)$ or $X(T)$ and as per convention (nothing sacrosanct in the notation as such), we denote $X(H) = 1$ and $X(T) = 0$. Then what would be the value of $P[X(H) = 1]$ or $P[X(T) = 0]$ is what is the finally probability function (generally the distribution function denoted by cumulative distribution function) which is of interest to us for both theoretical as well as practical purposes. So for example (again the same example of tossing the unbiased die) we have $\Omega = \{1, 2, 3, 4, 5, 6\}$, and we can have $X(1) = X(2) = X(3) = 0$ while $X(4) = X(5) = X(6) = 1$. Later on we have the mapping, such that $P[X(1)] = \dots = P[X(6)] = 1/6$. This simple concept can be extended for the higher dimension also and we can have the marginal, conditional as well as the joint distribution functions mapped as required.

Marginal Distributions

Let us again define two r.v., X, Y such that the required probability space is (Ω, \mathcal{A}, P) . Let us also define P_X and P_Y as the probability measure induced by X and Y respectively on the space, defined by $(\mathbb{B}_1, \mathbb{B}_2)$. Now if (X, Y) are r.v.s, such that they are defined on $(\mathbb{R}^2, \mathcal{B}^2 = \mathbb{B}_1 \times \mathbb{B}_2)$, where B_1 and B_2 are arbitrary Borel sets defined for X and Y and $F_{X,Y}(x,y)$ be the joint distribution function defined for the r.v., (X, Y) , where $F_{X,Y}(x,y) = P_{X,Y}(X \leq x, Y \leq y) = P_{X,Y}(B_1, B_2)$ is true, then $P_{X,Y}(B_1, \mathbb{R}^1)$ represents the probability that the variable (X, Y) will take a value belonging to the area marked by (B_1, \mathbb{R}^1) , i.e., the probability $X \in B_1$, irrespective of the value of Y (remember Y can take any value in the entire domain of Y , i.e., $Y \in \mathbb{R}^1$). In a similar way we can define $P_{X,Y}(\mathbb{R}^1, B_2)$. We can write $P_{X,Y}(B_1, \mathbb{R}^1) = P_X(X \in B_1) = P_X(B_1)$ and $P_{X,Y}(\mathbb{R}^1, B_2) = P_Y(Y \in B_2) = P_Y(B_2)$.

Thus the probabilities $P_{X,Y}(B_1, \mathbb{R}^1)$ (Figure 1.9) for varying values of X defines the marginal distribution of X relative to the joint distribution of (X, Y) . In a very simple sense it means we project the mass of the joint distribution on the **sub-space of the variable X** . Similarly we can define the probabilities $P_{X,Y}(\mathbb{R}^1, B_2)$ (Figure 1.10) for varying values of Y such that it defines the marginal distribution of Y relative to the joint distribution of (X, Y) and it implies that we project the mass of the joint distribution on the **sub-space of the variable Y** .

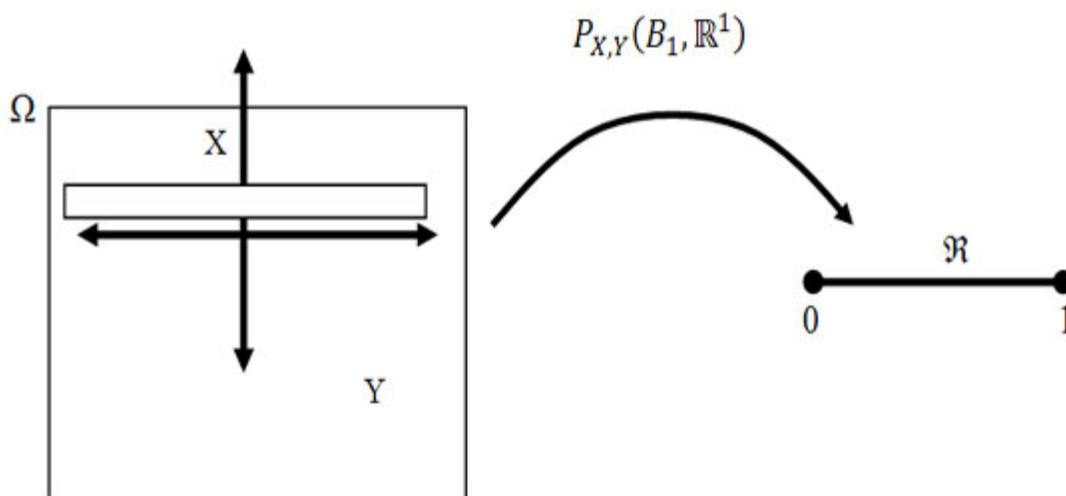


Figure 1.9: Illustration of $P_{X,Y}(B_1, \mathbb{R}^1) = P_X(X \in B_1) = P_X(B_1)$

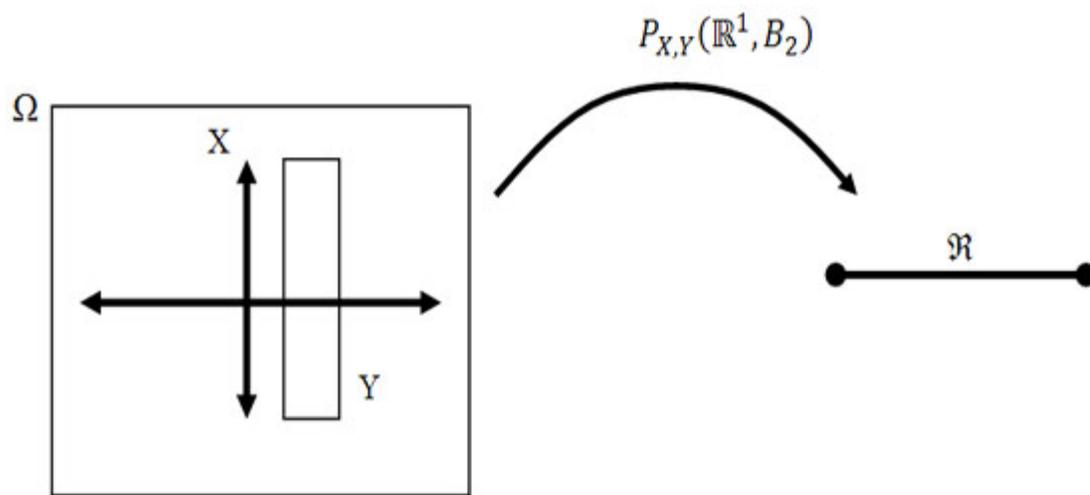


Figure 1.10: Illustration of $P_{X,Y}(\mathbb{R}^1, B_2) = P_Y(Y \in B_2) = P_Y(B_2)$

◀ Previous Next ▶

Module 1: Concepts of Random walks, Markov Chains, Markov Processes

Lecture 1: Introduction to Stochastic Process

One may add that the **marginal distributions** of X and Y relative to their **joint distributions** are identical with the distributions of X and Y taken individually one at a time, i.e.,

$$\begin{aligned} P_{XY}[B_1, \mathfrak{R}^1] &= P_X[X \in B_1] = P_X[B_1] = F_{XY}(x, -\infty \leq Y \leq +\infty) = G_X(x) = F_X(x, \infty) \\ P_{XY}[\mathfrak{R}^1, B_2] &= P_Y[Y \in B_2] = P_Y[B_2] = F_{XY}(-\infty \leq X \leq +\infty, y) = H_Y(y) = F_Y(\infty, y) \end{aligned}$$

and $G_X(x) = F_X(x, \infty)$ and $H_Y(y) = F_Y(\infty, y)$ have the properties of the univariate distributions of X and Y respectively.

In a similar way we can define the marginal distributions in the case when we have n number of r.v., (X_1, X_2, \dots, X_n) , such that each has the required probability space is (Ω, A, P) . Let us also define P_{X_1}, \dots, P_{X_n} as the probability measure induced by X_1, \dots, X_n respectively on the space, defined as (\mathbb{R}, B_i) , where $i = 1, \dots, n$. If (X_1, X_2, \dots, X_n) are r.v., such that they are defined on $(\mathfrak{R}^n, B^n = (B_1 \times \dots \times B_n))$, where $B_1 \times B_2 \times \dots \times B_n$ are arbitrary Borel sets identified for (X_1, X_2, \dots, X_n) then $F_{X_1, \dots, X_n}(x_1, \dots, x_n)$ be the joint distribution function defined for the r.v., (X_1, X_2, \dots, X_n) , where $F_{X_1, \dots, X_n}(x_1, \dots, x_n) = P_{XY}[X_1 \leq x_1, \dots, X_n \leq x_n] = P_{X_1, \dots, X_n}[B_1, \dots, B_n]$. Then $P_{X_1, \dots, X_n}[B_1, \mathfrak{R}^{n-1}]$ represents the probability that the variable (X_1, X_2, \dots, X_n) will take a value belonging to the area marked by $(B_1, \mathfrak{R}^{n-1})$, i.e., the probability $X \in B_1$, irrespective of the value of (X_2, X_3, \dots, X_n) (remember (X_2, X_3, \dots, X_n) can take any value in the entire domain of (X_2, X_3, \dots, X_n) , i.e., $X_i \in \mathfrak{R}^1$, where $i = 2, 3, \dots, n$). In a similar way we can define $P_{X_1, \dots, X_i, \dots, X_n}[\mathfrak{R}^1, \dots, B_i, \dots, \mathfrak{R}^1]$ and one can write $P_{X_1, \dots, X_i, \dots, X_n}[\mathfrak{R}^1, \dots, B_i, \dots, \mathfrak{R}^1] = P_{X_i}[X_i \in B_i]$, for some r.v., X_i .

Thus the probabilities $P_{X_1, \dots, X_i, \dots, X_n}[\mathfrak{R}^1, \dots, B_i, \dots, \mathfrak{R}^1]$ for varying values of X_i defines the **marginal distribution** of X_i relative to the joint distribution of $(X_1, \dots, X_{i-1}, X_{i+1}, \dots, X_n)$. In a very simple sense it means we project the mass of the joint distribution on the **sub-space of the variable X_i** . In a similar sense we can also say that the marginal distributions of X_i on $(X_j, j \neq i)$ relative to their joint distribution is identical with the distribution of X_i taken individually one at a time, i.e.,

$$\begin{aligned} P_{X_1, \dots, X_i, \dots, X_n}[\mathfrak{R}^1, \dots, B_i, \dots, \mathfrak{R}^1] &= P_{X_i}[X_i \in B_i] = P_{X_i}[B_i] \\ &= F_{X_1, \dots, X_i, \dots, X_n}(-\infty \leq X_1 \leq +\infty, \dots, X_i, \dots, -\infty \leq X_n \leq +\infty) \\ &= F_{X_i}(\infty, \dots, X_i, \dots, \infty) \end{aligned}$$

For a stochastic process we will consider the joint distribution, Thus given $\{X(t), t \in T\}$ s.t. $t_1 < t_2 < \dots < t_n$ we are interested to find the joint distribution of $\{X(t_1), X(t_2), \dots, X(t_n)\}$.

Note

We say a **stochastic process** is a **stationary joint distribution** if it is invariant to the shift of time, i.e., if the joint distribution of $\{X(t_1), X(t_2), \dots, X(t_n)\}$ and $\{X(t_1 + h), X(t_2 + h), \dots, X(t_n + h)\}$ are the same $\forall h$. This is usually called **stationary of order n** as we have n ordered time points. If a stochastic process is said to be of order n for every value of $n \geq 1$, then the stochastic process is called **strictly stationary**.

Moments

Few important moments which are of interest to us for any theoretical as well as empirical distributions are: mean, variance/standard deviation, covariance, skewness (3rd order moment), kurtosis (4th order moments) and other higher moments. Now if $E[X(t)]$ is **independent of t** , then we know that $E[X(t)]$ is a **constant**.

For moments we know the following

- 1st order moment is expected value or mean value or average value (we are being very generous in our terminology is trying to define the 1st moment, even though in the strict sense expected value, mean value, average value are slightly different concepts more pertaining to the property of the population and sample). The symbol of the 1st moment is $E(X)$
- 2nd order moment is used to find variance and the variance is given by $E\{(X - E(X))^2\}$. The concept of standard deviation, sample variance, standard error are also some metrics which can be calculated using the 2nd moment.
- 3rd order moments can be used to find skewness, and the general formulae is $E\{(X - E(X))^3\}$.
- 4th order moments can be used to find kurtosis and the general formulae is $E\{(X - E(X))^4\}$.
- r th order moment is denoted by $E(X^r)$.
- r th order central moment is denoted by $E\{(X - E(X))^r\}$. In this case the central moment is **calculated about the mean**, $E(X)$. In many cases we calculate the general moment called the raw moment and we have $E\{X - a\}^r$ which is the moment of order r , where this a can be any point.

Similarly for a stochastic process we can find the expected value, variance as well as the higher moments. Now the variance is given by

$\text{var}\{X(t)\} = E\{[X(t) - E(X(t))]^2\} = E\{[X(t) - E(X(t))] \times [X(t) - E(X(t))]\}$, while covariance is given as $\text{cov}\{X(t), X(s)\} = E\{[X(t) - E(X(t))] \times [X(s) - E(X(s))]\}$.

Note

1. If we take $E[X(t)]$ as **constant**, then we have

- $\text{var}\{X(t)\} = E\{X(t) \times X(t)\} - E\{X(t)\} \times E\{X(t)\}$
- $\text{cov}\{X(t), X(s)\} = E\{X(t) \times X(s)\} - E\{X(t)\} \times E\{X(s)\}$

2. If we take $E[X(t)]$ as zero, then we have

- $\text{var}\{X(t)\} = E\{X(t) \times X(t)\} = c(t)$, where $c(t)$ is a constant
- $\text{cov}\{X(t), X(s)\} = E\{X(t) \times X(s)\} = c(t, s)$, where $c(t, s)$ is a constant

3. If $E\{X(t)\}$ and $V\{X(t)\}$ does not depend on t , then $\text{cov}\{X(t), X(s)\}$ only depends on $|t - s|$ and is called **covariance stationary** or **weekly stationary** or **stationary** and in the wide sense stationary. We will discuss about these later on.

Gaussian Process

A stochastic process is called **Gaussian process** if the joint distribution of $\{X(t_1), X(t_2), \dots, X(t_n)\}$ is n variate normal distribution. Now a n variate normal distribution is specified by its mean vector

$[E\{X(t_1)}, E\{X(t_2)}, \dots, E\{X(t_n)}]$ and the variance-covariance matrix

$$D_{n \times n} = \begin{pmatrix} \text{cov}\{X(t_1), X(t_1)} & \text{cov}\{X(t_1), X(t_2)} & \dots & \text{cov}\{X(t_1), X(t_n)} \\ \text{cov}\{X(t_2), X(t_1)} & \text{cov}\{X(t_2), X(t_2)} & \dots & \text{cov}\{X(t_2), X(t_n)} \\ \vdots & \vdots & \vdots & \vdots \\ \text{cov}\{X(t_n), X(t_1)} & \text{cov}\{X(t_n), X(t_2)} & \dots & \text{cov}\{X(t_n), X(t_n)} \end{pmatrix}$$

Note

If a Gaussian process is **covariance stationary**, then it is **strictly stationary**.



Example 1.5

Let $\{X_n, n \geq 1\}$ be a **stochastic process** with $E(X_n) = 0$ and $V(X_n) = 1, \forall n$ and **uncorrelated** random variables (r.v's), then

$$\text{cov}(X_n, X_m) = \begin{cases} 0 & \text{if } n \neq m \\ 1 & \text{if } n = m \end{cases}$$

Hence

$$\text{cov}(X_n, X_m) = \begin{cases} 0 & \text{if } |n - m| \neq 0 \\ 1 & \text{if } |n - m| = 0 \end{cases}, \text{ which implies that it is covariance stationary}$$

Furthermore if the random variables (r.v's) are *i.i.d.*, then the stochastic process is **strictly stationary**

Note : A stochastic process is not stationary in any sense is called **evolutionary** or **non-stationary**.

One should remember that modelling with continuous stochastic process is easier than working with stochastic processes which are not continuous, i.e., discrete. For example one may consider stock prices movement to be continuous. But in general the prices move in quantum of say $\left(\frac{1}{8}\right)$ or $\left(\frac{1}{6}\right)$ of a Re. 1. So in real sense the stock price movement is not continuous, even though for all practical purposes when trying to model stock prices or study them, we consider stock price movement as continuous. We will study stochastic processes which are called **diffusion processes**, the illustration of which is given below (Figure 1.11).

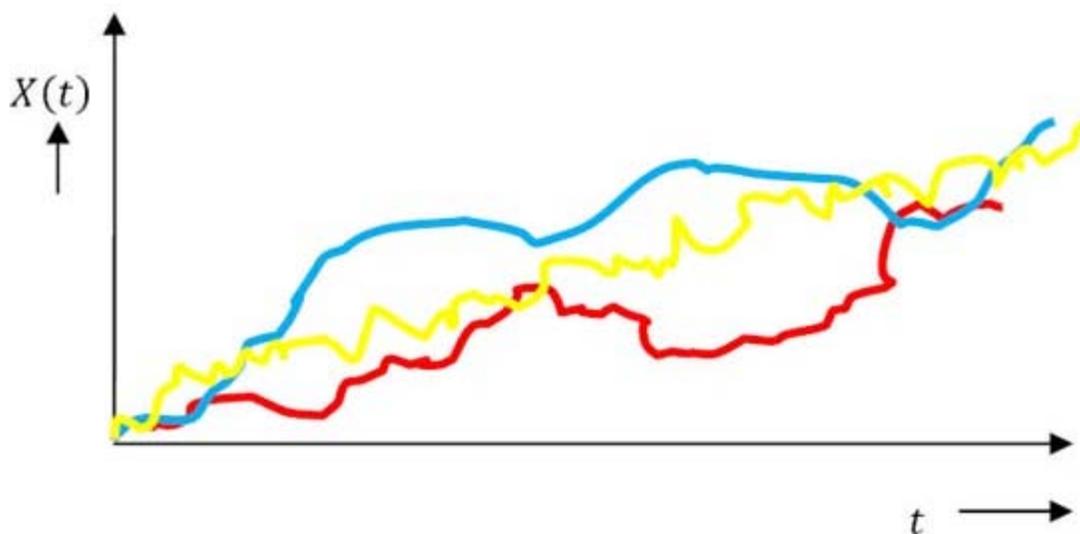


Figure 1.11: Diffusion process which may be used to define stock price movements

Random Walks

Let us start the discussion with a simple illustration and few practical examples.

Illustrative example for explaining the concept of random walk

Assume there is a particle which at time $n = 0$ is at any particular position or state, say i , i.e., $X_{n=0} = i$, and assume the movement of the particular is along the x-axis. This being the first simple example we will consider the movement is along one dimension only. Now suppose at time $n = 1$, we are in the j^{th} position or state, i.e., $X_{n=1} = j$. Assume the movement that the particle has from the i^{th} state to the j^{th} state in the time between $n = 0$ to $n = 1$ happened due to the fact that the particular had a jump or moved a step, which we denote by Z_1 . Furthermore at $n = 2$ assume the particular undergoes another jump Z_2 , such that Z_1 and Z_2 are **identically** and **independently** distributed (*i.i.d*). For ease of explanation consider the diagrams below (Figure 1.12 and Figure 1.13) which is self explanatory. In the first figure (Figure 1.12) we see how the particle moves from the i^{th} state (which is occupies at $n = 0$) to subsequent states denoted by j, k, l, m and so on at instances of time denoted by $n = 1, n = 2, n = 3, n = 4$, etc. The colour schemes makes it clear how the movement of the particle takes place at each instance and these jumps or morevement are one step at a time. While on the other hand Figure 1.13 gives an arbitrary snapshot of the movement of the particle, and obviously there are other combinations of movements of the particle also.

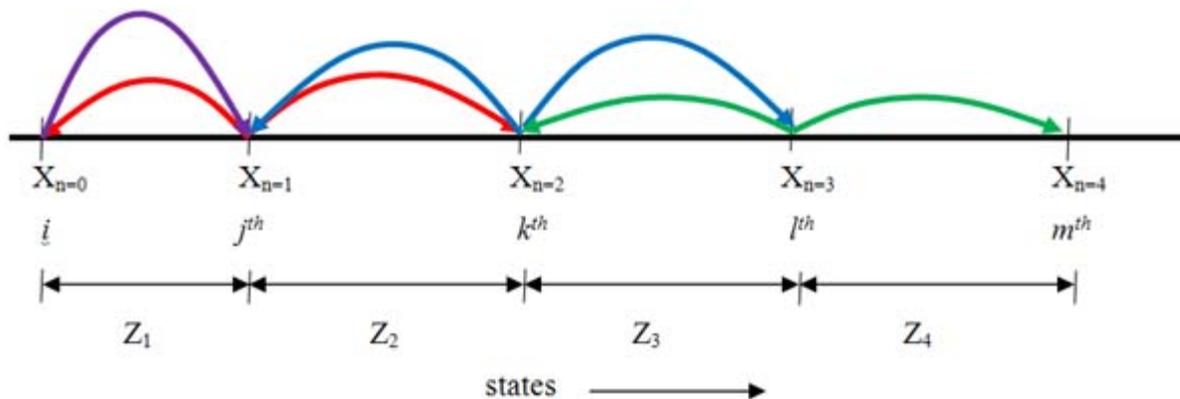


Figure 1.12: Illustration of random walk

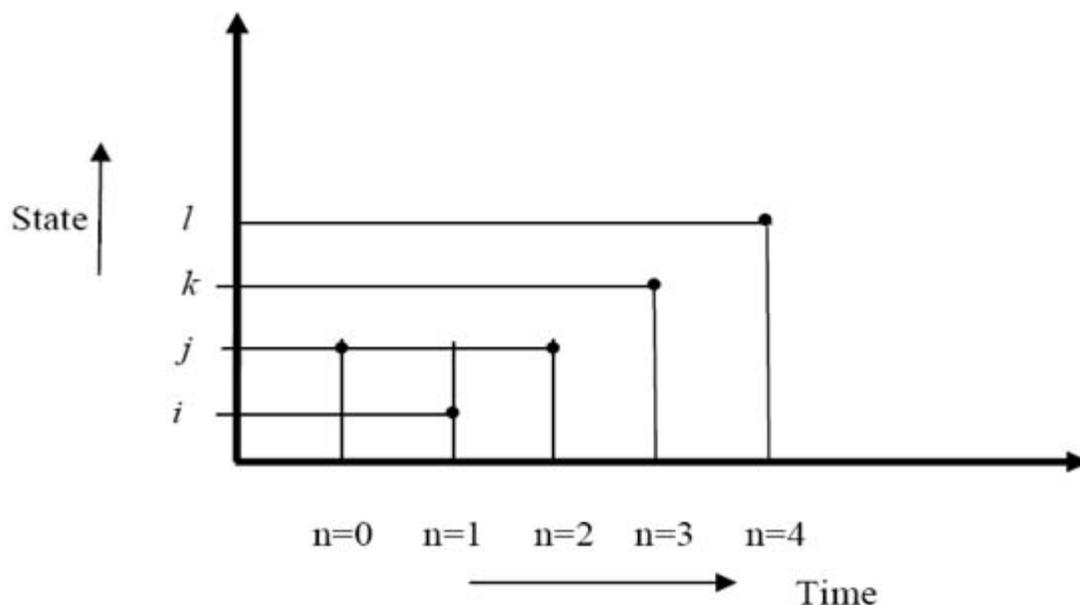


Figure 1.13: Arbitrary snapshot of the particle under going random walk

In case we simplify the above notations then we consider the case where at the starting of the process, i.e., $n = 0$, the particle is at any given point say, i , such that $X_{n=0} = i$. At time $n = 1$ the particle undergoes a jump of quantum Z_1 (which is a random variable with a particular distribution). Furthermore at time $n = 2$ the same particle undergoes another jump Z_2 where Z_2 is independent of Z_1 but has the same distribution as Z_1 . Thus the particle undergoes jumps in a manner that after end of the first time period, $n = 1$ it is at the position $X_0 + Z_1$, after the second time period, $n = 2$ it is at the position $X_0 + Z_1 + Z_2$. Thus for $X_n = X_0 + Z_1 + Z_2 + \dots + Z_n$, $\{Z_n\}$ is a sequence of mutually independent and identically distributed random variables.

More generally $X_n = X_{n-1} + Z_n$, $n = 1, 2, \dots$ and in case $Z_i = -1$, or 0 or $+1$, then what is of interest to

us is $Pr\{Z_i = -1\}$, $Pr\{Z_i = 0\}$ and $Pr\{Z_i = +1\}$, such that $Pr\{Z_i = -1\} + Pr\{Z_i = 0\} + Pr\{Z_i = +1\} = 1$. There may be instances where $Pr\{Z_i = 0\} = 0$, in which case we should have $Pr\{Z_i = -1\} + Pr\{Z_i = +1\} = 1$

Module 1: Concepts of Random walks, Markov Chains, Markov Processes

Lecture 1: Introduction to Stochastic Process

Example 1.6

Consider we put an item into use and obviously it will fail after some time. When it fails it is replaced by another item of the same type. We may safely assume that $n \geq 0$ which denotes the lives of the items (they are all similar remember) are **independent** (a restrictive assumption) and each item fails at the n^{th} time with a probability denoted by $p_n \geq 0$, such that $n = 0, 1, 2, \dots$. Also let us consider that the distribution, $\{p_n\}$ is **aperiodic** such that $\sum_{n=0}^{\infty} (n \times p_n) < \infty$. If X_n is the age of the item at time n (here is where we bring the time concept as relevant for the Stochastic Process), i.e., the number of periods (including the n^{th}) it has been in use, then we may assume $\{X_n, n \geq 0\}$ is a random walk (a good example of **Markov chain**) with the transition probabilities which can be calculated using $\lambda_i(n) = \left(\frac{p_n}{\sum_{m=0}^{\infty} p_m} \right)$.

Example 1.7

Consider an insurance company has just started, say at time period $n = 0$, and the initial fixed capital the company has is $X_0 = 1,00,00,000$. In due course of time the company receives amounts Y_1, Y_2, \dots in form of premiums and other incomes, while at the same time due to insurance claims and other expenses it has to shelf out amounts denoted by W_1, W_2, \dots . One should remember that these inputs and outputs are happening at same instances of time, which are $1, 2, \dots$. Now the amount of money the insurance company has at any point of time n is given by $X_n = X_0 + (Y_1 - W_1) + (Y_2 - W_2) + \dots + (Y_n - W_n)$.

It is obvious that the insurance company goes bankrupt in case $X_n \leq 0$. Thus if we assume that $\{Y_n\}$ and $\{W_n\}$ are two separate sequences then X_n is a random walk starting at X_0 with jumps $(Y_1 - W_1)$, $(Y_2 - W_2)$ and so on. An interesting thing is the fact that for this random walk we have an **absorbing barrier at the origin** such that the random walk is denoted by:

$$X_n = \begin{cases} X_{n-1} + Y_n - W_n & (X_{n-1} > 0, X_{n-1} + Y_n - W_n > 0) \\ 0 & \text{otherwise} \end{cases}$$

◀ Previous Next ▶

Example 1.8

Let us now concentrate on an example from civil engineering. Let X_n represent the amount of water in a dam at the end of n units of time. Suppose during day r , Y_r units of water flows into the dam in form of rainfall, supply from rivers, reservoirs, etc., and also assume that Y_r has a particular statistical distribution. As we all know any amount of water cannot be stored in a dam, hence water is discharged from the dam based on the following rule which is, if $X_{r-1} + Y_r - b > a$ holds true then a amount of water is discharged, where b is the capacity of the dam.

The situation may be represented as follows:

$$X_n = \begin{cases} 0 & X_{n-1} + Y_n < 0 \\ X_{n-1} + Y_n & 0 < X_{n-1} + Y_n < b \\ b & X_{n-1} + Y_n \geq b \end{cases}$$

It is easy to note that the dam continues to remain full (i.e., with capacity b) until the first negative Z_r , i.e., the first subsequent day when the amount released exceeds the inflow. On the other hand it will continue to remain empty until the first positive value of Y_r . If $\{Y_r\}$ is a sequence of mutually independent and identically distributed random variables that may describe X_n as a simple random walk on the interval $[0, b]$ with two **reflecting barriers** at 0 and b .

Example 1.9

The next example is from astronomy. Consider that during one revolution around the earth the satellite under goes a change in its energy level. In each successive revolutions the quantum of energy change is assumed to be identical and independently distributed and is denoted by Z_n . Thus with an initial energy level of $X_0 > 0$ for the satellite, the energy level at time n is given by $X_n = X_0 + Z_1 + \dots + Z_n$. The satellite escapes the earth's gravitational pull if its energy is more than a threshold value of b else it falls to the earth if its energy level is less than a . In that case the situation may be represented by:

$$X_n = \begin{cases} a & X_{n-1} + Z_n < a \\ X_{n-1} + Z_n & a < X_{n-1} + Z_n < b \\ b & X_{n-1} + Z_n > b \end{cases}$$

Where a and b are the **absorbing states** for this random walk.



Example 1.10

Define $X_n = \begin{cases} 1 & \text{if the market goes up on day } n \\ 0 & \text{otherwise} \end{cases}$

Now the question is, is the market persistent? To answer this we first denote the **transition probability matrix** $P = \begin{pmatrix} q & 1-q \\ 1-p & p \end{pmatrix}$, where the 1st row & 1st column element which is q , denotes what is the probability that provided price was at level 0 (this zero is just a notional concept and has nothing to do with its actual value) today, it would continue to be at that price tomorrow also. Similarly the other probabilities signify the other price movements. Now for any particular row (denoted by the suffix i) we have $\sum_{j=0}^{\infty} p_{ij} = 1, \forall i$. Hence the corresponding probabilities of 1st row & 2nd column or for that matter 2nd row & 1st column which are $(1-q)$ and $(1-p)$ can be easily found given q and p values. This means that $p_{0,1} = 1 - p_{0,0} = (1-q)$ and $p_{1,0} = 1 - p_{1,1} = (1-p)$. Even though a repetition we would like to again mention that as $p_{i,j} = P(X_{n+1} = j | X_n = i)$, hence $p_{1,1} = P(X_{n+1} = 1 | X_n = 1)$.

Example 1.11

In line with example 1.9, let us conduct a simple thought experiment, where suppose we are given the state of the prices of a particular stock for 3 consecutive days and assume the price fluctuations are such that they are either 0 or 1. Then the transition probability matrix, P would be denoted as

$P = \begin{pmatrix} p & 1-p & 0 \\ q & p & 1-(p+q) \\ 0 & q & 1-q \end{pmatrix}$. The question which comes to anyone's mind at this stage is how do

we estimate the values of p or q . In case we have the matrix $P = \begin{pmatrix} 6 & 4 \\ 2 & 8 \end{pmatrix}$, which is a sample instance of the number of times of movement of stock prices, then the probability values in the long run may be estimated as $p_{0,0} \cong 0.60$, $p_{0,1} \cong 0.40$, $p_{1,0} \cong 0.20$, $p_{1,1} \cong 0.80$. Another method to calculate the values of $p_{i,j}$ is by using the method of maximum likelihood estimation (MLE) principle, where we look into the likelihood of the data and maximize the value, i.e., $L(p_{0,0}, p_{0,1}, p_{1,0}, p_{1,1}) = \{p_{0,0}^6 \times p_{0,1}^4 \times p_{1,0}^2 \times p_{1,1}^8\}$.

One would immediately recognize that $\log_e \{L(p_{0,0}, p_{0,1}, p_{1,0}, p_{1,1})\}$ is a monotonic transformed function of $L(p_{0,0}, p_{0,1}, p_{1,0}, p_{1,1})$. Moreover based on the following facts that (i) $p_{0,0} + p_{0,1} = 1$, (ii) $p_{1,0} + p_{1,1} = 1$ and (iii) $p_{0,0} + p_{0,1} + p_{1,0} + p_{1,1} = 2$ we equate the equations $\frac{\partial \log_e \{L(p_{0,0}, p_{0,1}, p_{1,0}, p_{1,1})\}}{\partial p_{0,0}}$, $\frac{\partial \log_e \{L(p_{0,0}, p_{0,1}, p_{1,0}, p_{1,1})\}}{\partial p_{0,1}}$, $\frac{\partial \log_e \{L(p_{0,0}, p_{0,1}, p_{1,0}, p_{1,1})\}}{\partial p_{1,0}}$ and $\frac{\partial \log_e \{L(p_{0,0}, p_{0,1}, p_{1,0}, p_{1,1})\}}{\partial p_{1,1}}$ individually to zero and solve them to find $p_{0,0}, p_{0,1}, p_{1,0}$ and $p_{1,1}$, utilizing their respective estimated values given by $\hat{p}_{0,0}, \hat{p}_{0,1}, \hat{p}_{1,0}$ and $\hat{p}_{1,1}$. To find the maximum or minimum value we verify the properties of the Hessian matrix $(H = \begin{vmatrix} g_1(\hat{p}_{0,0}, \hat{p}_{0,1}, \hat{p}_{1,0}, \hat{p}_{1,1}) & 0 \\ 0 & g_2(\hat{p}_{0,0}, \hat{p}_{0,1}, \hat{p}_{1,0}, \hat{p}_{1,1}) \end{vmatrix})$, where g_1 and g_2 are functions of $\hat{p}_{0,0}, \hat{p}_{0,1}, \hat{p}_{1,0}, \hat{p}_{1,1}$. For the benefit of the reader we would like to mention that $-H^{-1}$ which is the variance/covariance of $\hat{p}_{0,0}, \hat{p}_{0,1}, \hat{p}_{1,0}$ and $\hat{p}_{1,1}$ is asymptotically normal. One can also calculate the standard errors of $\hat{p}_{0,0}, \hat{p}_{0,1}, \hat{p}_{1,0}$ and $\hat{p}_{1,1}$ using the MLE approach.

Markov Chain

Consider the following where we have (i) State Space : $S = \{0, 1, 2, \dots\}$, (ii) Index Set : $T = \{0, 1, 2, \dots\}$ and (iii) event is occurring when X_n belongs to state ε . Figure 1.14 and Figure 1.15 give a diagrammatic view of **Markov chain** for an ease of understanding of the reader.

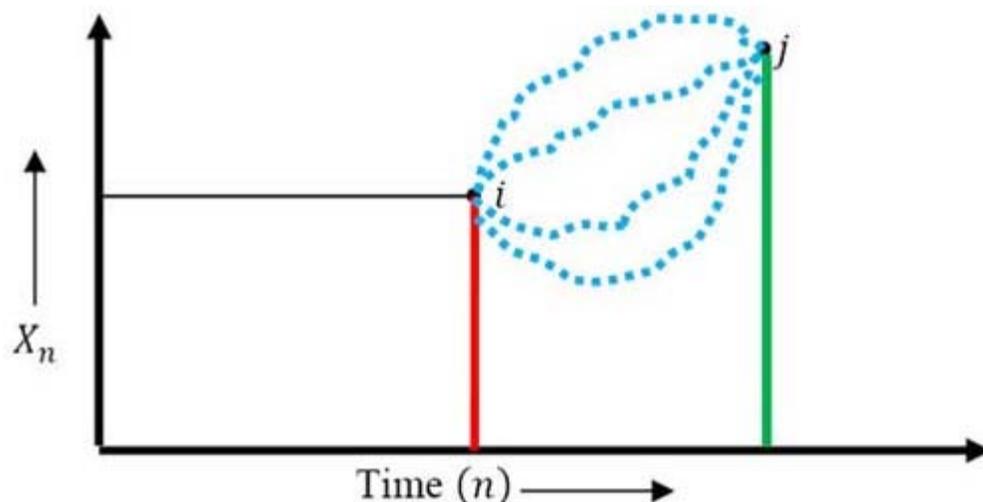


Figure 1.14: Illustration of Markov chain

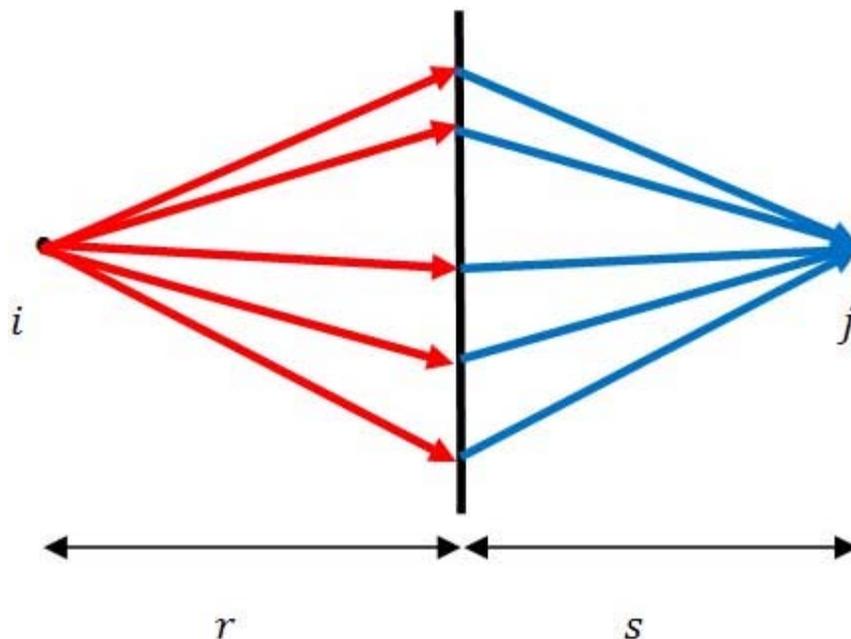


Figure 1.15: Illustration of Markov chain considering initial state i and final state j

In Figure 1.14 it shows that the movements from state i to state j can take place through any direction and they are continuous. But to make our life simple we consider the jumps to be in discrete time and the movements are also considered to be discrete, which is depicted in Figure 1.15.

The probability of X_{n+1} being in state j given that X_n is in state i (the concept of **one step transition probability**) is given by $P_{i,j}^{n,n+1} = P(X_{n+1} = j | X_n = i)$. Using matrix notation it may be

represented by the standard matrix representation as shown $\begin{bmatrix} p_{0,0} & p_{0,1} & \dots \\ p_{1,0} & p_{1,1} & \dots \\ \vdots & \vdots & \ddots \\ \dots & \dots & \dots \end{bmatrix}$. Few importance

properties hold for this transition probability matrix and they are as follows:

1. $p_{i,j} \geq 0, \forall i,j$
2. $\sum_{j=0}^{\infty} p_{ij} = 1, \forall i$

◀ Previous Next ▶

Module 1: Concepts of Random walks, Markov Chains, Markov Processes

Lecture 1: Introduction to Stochastic Process

Claim: Given X_0 (or its probability) and $p_{i,j}$ ($\forall i, j$), then the whole probabilistic structure of the process is determined.

Proof: We need to compute $P(X_0 = i_0, X_1 = i_1, \dots, X_n = i_n), \forall i_0 < i_1 < \dots < i_n$

Thus:

$$P(X_0 = i_0, X_1 = i_1, \dots, X_n = i_n) = P(X_n = i_n | X_{n-1} = i_{n-1}, \dots, X_0 = i_0) \times P(X_{n-1} = i_{n-1}, \dots, X_0 = i_0)$$

i.e., we use the concept of $P(A \in B) = P(A|B) \times P(B)$

Now

$$P(X_n = i_n | X_{n-1} = i_{n-1}, \dots, X_0 = i_0) = p_{i_{n-1}i_n}$$

Similarly we have:

$$P(X_{n-1} = i_{n-1} | X_{n-2} = i_{n-2}, \dots, X_0 = i_0) = p_{i_{n-2}i_{n-1}}$$

Using this we can find the p_{i_1, i_0} . Finally the probability of starting from the initial position will be given by the problem or from the practical situation, based on which we have formulated our problem.

◀ Previous Next ▶

Example 1.12

The concept of **transition probability** values can be used quite nicely in finance domain, say for example in interest calculation models. Consider the example where we have the interest rate at time period t is given by $r_t = r_{t-1} + \varepsilon_t$, where ε_t is the error. This is a simple model which is a pure Markov chain example. Next assume $r_t = a + br_{t-1} + \sigma\varepsilon_t$, where a and b are constants while σ is the volatility. This is a simple AR(1) model. In the second model the volatility can change with respect to time. Furthermore we can assume that this volatility term also follows a Markov chain process. In the simple example we may model the volatility as follows :

Table 1.2: Scenarios of Volatility

	High σ	Low σ
High σ	q	$1 - q$
Low σ	$1 - p$	p

Different researchers have found methods to find the interest rate using different concepts of interest rate calculations. Another method of calculating interest rate can be $r_t = a + br_{t-1} + \sigma\varepsilon_t r_{t-1}^\gamma$, where γ is some index. We can bring more complication in this model by considering a as well as b also vary with respect to time. A closer look at the equation will immediately reveal that when $b = 1$ the interest rate will blow up. An important method named the **unit root test method** has quite a lot of application in interest rate problem, using which we can test for the stationary/non-stationary of the time series, especially interest rate or rate change of stock prices.

Further definition of Markov Chain

A stochastic process with **discrete state space** and **discrete parameter space** is called a **Markov chain**. Let us consider a stochastic process with $t_1 < t_2 < \dots < t_n < t, \forall t, t_i \in T$. Now if we have

$P[X(t) = j | X(t_1) = i_1, \dots, X(t_n) = i_n] = P[X(t) = j | X(t_{n-r}) = i_{n-r}, \dots, X(t_n) = i_n]$ then the stochastic process is called a **Markov process** of order r .

Remember a **homogeneous Markov chain** of order one is also called a **Markov chain**, where we denote it simply as $\{X_n, n = 0, 1, 2, \dots, K\}$

Transition probability and transition matrix

Let us consider the following, i.e., $P[X_{n+1} = j | X_n = i, X_{n-1} = i, \dots] = P[X_{n+1} = j | X_n = i] = p_{ij}$, which is the transition probability of the transition from i th to the j th state in a **single step**. If the transition property is independent of n , then it is called the **homogeneous Markov chain**.

One step transition probability

Now $p_{ij} = P[X_n = j | X_{n-1} = i] = \dots = P[X_2 = j | X_1 = i] = P[X_1 = j | X_0 = i]$, denotes the probability corresponding to the chance that the stochastic process will move from the i th to the j th state considering that the i th to the j th states can be reached between any two consecutive positions, but generally that may not always be the case. So without any further complication consider p_{ij} is fixed irrespective of n , such that we have $P[X_{n+m} = j | X_m = i] = p_{ij} \times \dots \times p_{ij} = p_{ij}^n$, which is the n th step transition probability which denotes the transition of the particle or body as it goes from state i to state j after n steps. Here m denotes the time instance when the particle is at state i and n is the time units after which it reaches state j from state i . Hence the total time elapsed is $(n + m)$ periods.

Now if we bring in the concept of probability mass function, then we can easily add that $P[X_0 = i] = a_i$ for $i \in I$ (where I is the state space). Here $a_i \geq 0 \forall i$, and $\sum_{i \in I} a_i = 1$. The second term which is

$\sum_{i \in I} a_i = 1$ means that the process should definitely start from any one of the positions which is an element of I . Utilizing this concept for this simple case we can easily extend this and have the concept of transition probability, which is given by $P[X_{n+1} = j | X_n = i] = p_{ij}, i, j \in I$. Here one can deduce that

$p_{ij} \geq 0 \forall i, j$ and $\sum_{j \in I} p_{ij} = 1$. If we extend the concept of transition probability we have the transition probability matrix given by $P = (p_{ij})$, such that in matrix formulation it is given as

$$P = \begin{pmatrix} p_{11} & p_{12} & \dots \\ p_{21} & p_{22} & \dots \\ \vdots & \vdots & \dots \end{pmatrix}$$
. One should remember that the row sum is 1, i.e., $\sum_{j \in I} p_{ij} = 1$, which is very

intuitive. Thus if you are at any i th state then you can move to either 1st or 2nd or any other state, including i th state (i.e., remains at the same position) and **no where else**. Another important thing to remember about this matrix P , is the fact that it is not an symmetric matrix, as $p_{ij} \neq p_{ji}$.

Now $P[X_n = i] = \sum_{j \in I} P[X_n = i, X_0 = j]$

i.e., $\sum_{j \in I} P[X_0 = j] \times P[X_n = i | X_0 = j] = \sum_{j \in I} a_j \times p_{ij}^{(n)}$, where $p^* = p_{ij}^{(n)}$ is the n step transition

probability matrix and $p^* = p \times p \times \dots \times p$, n times, where, $p = (p_{ij})$, i.e., $p = \begin{pmatrix} p_{11} & p_{12} & \dots \\ p_{21} & p_{22} & \dots \\ \vdots & \vdots & \dots \end{pmatrix}$

◀ Previous Next ▶