UNIT 1 APPLIED PROBABILITY I

Structure

1.0 OBJECTIVES

After going through this unit, you will be able to:

- appreciate the application of probability theory in uncertain scenario; and
- formulate problems with different probabilistic distributions

1.1 INTRODUCTION

You have already seen the basic theory of prgbability in statistics course (i.e., EEC-13) as well as in quantitative methods (MEC-03). The present discussion is intended for a review of those themes. Since we will rely on probability-based approaches most of the time in our discussion of financial and insurance related themes, the topics covered below aim at providing you the basic inputs for understanding the analysis being carried on in those fields. While developing the study material in this unit, we have used the notations, examples and proofs given in Bialas (2005) and Imbens (2000). For your reference, these sources will be useful.

1.2 SET

Basic Definitions

We need to use **set theory** for describing possible outcomes and scenarios in our model of uncertainty.

***Remember** that a set is a collection of elements.*

Notation: Sets are denoted by capital letters ($A, B, \Omega, \ldots$) while their elements are denoted by small letters $(a, b, \omega, \ldots)$. If x is an element of a set A, we write $x \in A$. If x is not an element of A, we write $x \notin A$.

There are two sets of special importance:

i) The **universe** is the set containing all points under consideration and is denoted by Ω.

ii) The **empty** set (the set containing no elements) is denoted by $\emptyset$.

If every point of set A belongs to set B, then we say that **A** is a subset of B (B is a **superset** of A). We write

$A \subseteq B$; $B \supseteq A$

$A=B$ if and only if $A \subseteq B$ and $B \supseteq A$.

***A* is** a **proper** subset of B *if* $A \subseteq B$ *and* $A \neq B$.

Two sets A and B are disjoint (or mutually exclusive) if $A \cap B = \emptyset$

Cardinality* of *Sets

The *cardinality* of a set A (denoted by $|A|$) is the number of elements in A.

For some (but not all) experiments, we have

$|\Omega|$ is finite.

Each of the outcomes of the experiment is equally likely.

In such cases, it is important to be able to enumerate (i.e., count) the number of elements of subsets of Ω.

Sample spnces and events

Using set notation, we can introduce the first components of our probability model.

A **sample space**, Ω, is the set of all possible outcomes of an experiment.

An **event** is any subset of a sample space. For example,

when you roll a die once, Ω for this experiment is given by

$\Omega = \{1,2,3,4,5,6\}$

The subsets of Ω that represent the following events are:

1) The die turns up even: $A = \{2,4,6\}$

2) The die turns up less than 5: $B = \{1,2,3,4\}$

3) The die turns up 6: $C = \{6\}$

Sets of evenis

It is sometimes useful to be able to talk about the set of all possible events $\mathcal{F}$. The set $\mathcal{F}$ is very different than the set of all possible outcomes, Ω. Since an event is a set, the set of all events is really a set of sets. Each element of $\mathcal{F}$ is a subset of Ω.

For example, suppose we toss a two-sided coin twice. Then the sample space is

$\Omega = \{HH, HT, TH, TT\}$.

If you were asked to list all of the possible events for this experiment, you would need to list all possible subsets of Ω, namely,

$A_1 = \{\{H\ H\}\}$	$A_9 = \{\{H\ T\}, \{T\ T\}\}$
$A_2 = \{\{H\ T\}\}$	$A_{10} = \{\{T\ H\}, \{T\ T\}\}$
$A_3 = \{\{T\ H\}\}$	$A_{11} = \{\{H\ H\}, \{H\ T\}, \{T\ H\}$
$A_4 = \{\{T\ T\}\}$	$A_{12} = \{\{H\ H\}, \{H\ T\}, \{T\ T\}\}$
$A_5 = \{\{H\ H\}\}, \{H\ T\}\}$	$A_{13} = \{\{H\ H\}, \{T\ H\}, \{T\ T\}\}$
$A_6 = \{\{H\ H\}, \{T\ H\}\}$	$A_{14} = \{\{H\ T\}, \{T\ H\}, \{T\ T\}\}$
$A_7 = \{\{H\ H\}, \{T\ T\}\}$	$A_{15} = \{\{H\ H\}, \{H\ T\}, \{T\ H\}\} \{T\ T\}\} = \Omega$
$A_8 = \{\{H\ T\}, \{T\ H\}\}$	$A_{16} = \{\} = \varnothing$

Thus we have $\mathcal{F} = \{A_1, A_2, ..., A_{16}\}$

Power Set

For a given set A, the set of all subsets of A is called the **power set** of **A**, and is denoted by 2^A.

If Ω is a sample space, then we often use the set of all possible events $\mathcal{F} = 2$". When Ω contains a finite number of outcomes, this often works well. However, for some experiments, Ω has an infinite number of outcomes (either countable or uncountable). As a result, $\mathcal{F} = 2^{\Omega}$ is much too large. In those cases, we select a smaller collection of sets, $\mathcal{F}$, to represent those events we actually need to consider.

1.3 PROBABILITY

1.3.1 Probability Spaces

In the above, we have seen how to use sample spaces Ω and events $A \subseteq \Omega$ to describe an experiment. In following, we will define what we mean by the *probability of an event*, $P(A)$ and the mathematics necessary to compute $P(A)$ for all events $A \subseteq \Omega$.

For infinite sample spaces, for example when $\Omega = \mathbb{R}$, it is not possible to reasonably compute $P(A)$ for all $A \subseteq \Omega$. In such cases, we restrict our definition of $P(A)$ to **some** smaller subset of events, $\mathfrak{Z}$, that does not necessarily include every possible subset of Ω.

Measuring Sets

Let us start with some important definitions.

1) *The* **sample space,** Ω, *is the set of allpossible outcomes of an experiment.*
2) **An event** *is any subset of* ***the*** *sample space.*
3) **A probability measure,** P(.), is a function, defined on subsets of Ω, that, assigns to each event a real number P(E) such that the following three probability axioms hold:

   i) $P(E) \geq 0$ *for all events E*

   ii) $P(\Omega) = 1$

   iii) $f E \cap F = \varnothing$

If the events E and F are mutually exclusive, then $P(E \cup F)=P(E)+P(F)$

The above three specifications are called the *Kolmogorov axioms* for probability measures. We will discuss more about the probability measures in the next unit. For the purposes of the present unit, you need to remember that the probability of an event A for every $A \subseteq \Omega$ lies in the closed interval of $[0,1]$ i.e.. $0 \leq P(A) \leq 1$. Therefore you can take $P(.)$ as a function which assigns to each element of the set of events $\mathcal{F}$, **a** single number from the interval $[0,1]$. That is, $P:\mathcal{F} \rightarrow [0,1]$.

In calculus, you worked with real-valued functions $f(.)$. That is, $f:\mathbb{R} \rightarrow \mathbb{R}$.

Such functions assign each number in $\mathbb{R}$ a single number from $\mathbb{R}$. For example, $f(x)=x^3$, maps $x=2$ into $f(2)=8$. In probability, the function, $P(.)$, is a function whose argument is a subset of Ω (namely, an event) rather than a number. In mathematics, we call functions that assign real numbers to subsets, *measures.* Following such a logic. we call $P(.)$ as a *probability measure.*

Interpretation of Probability

Logical Probability. Take Ω as a finite set

$$\Omega=\{\omega_1,\omega_2,\ldots,\omega_n\}$$

and each of the outcomes in Ω is equally liliely. That is.

$$P(\{\omega_i\})=\frac{1}{n} \text{ for } i=1,\ldots,n \text{ and for any } A \subseteq \Omega \text{ assign } P(A)=\frac{|A|}{n}$$

For example, consider the outcome of rolling a fair six-sided die once, Then we have $\Omega=\{\omega_1,\omega_2,\ldots,\omega_6\}$ where ω_i = the die shows i number for $i=1,2,\ldots,6$. So we can assign. for example, $P(\{\omega_i\})=\frac{1}{6}$ for $i=1,\ldots,6$

$$P(\{\text{even number}\})=P(\{\omega_2,\omega_4,\omega_6\})=\frac{3}{6} \text{ and, in general,}$$

$$P(A)=\frac{|A|}{6} \text{ for any } A \subseteq \Omega$$

Experimental Probability. Think of an experiment that can be identically reproduced over and over again. Each replicate is called an *experimental trial.* We also assume that the outcome of any **one** of the trials does not affect the probabilities associated with the outcomes of any of the other trials. For example, take motors being produced by an assembly line, with each motor an experimental trial. The outcome of one of our experimental trials is whether that motor is either good or bad. We record the outcome as

T_i = 0 if motor i is bad

T_i = 1 if motor i is good

Thus, Ω = {bad, good)., If A = {good), the logical ,probability $P(A)$ is assigned from the infinite sequence of experimental outcomes $T_1, T_2, \ldots$ such that

$$P(A)=\lim_{n\to\infty}\frac{\sum_{i=1}^{n} T_i}{n}$$

Subjective Probability. In some experiments, it is not possible for replication. For example, consider the experiment of taking MEC-03, with the outcome being the grade that you get in the course. Let

A = {you get an "A" in MEC-03)

To determine the subjective probability of A, you need to construct an equivalent case that is amenable to the format of selecting the probability. For **example,** you may consider a lottery consisting of an urn, and an unlimited quantity of red and white balls. The decision maker is asked to place in the urn any number of red balls and white balls he or she desires, with the goal of making the event of drawing a red ball equivalent to getting an "A" in MEC-03.

Once it is decided how many balls of each type to place in the urn, we simply count the balls in the urn and assign

$$P(A) = \frac{\text{number of red balls in the urn}}{\text{total number of balls,in the urn}}$$

Some Important Results

i) For any finite number of pairwise disjoint sets $A_1, A_2, \ldots, A_n$ we have

$$P(A_1 \cup A_2 \cup \ldots \cup A_n) = P(A_1) + P(A_2) + \ldots + P(A_n)$$

ii) For any two events A and B

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

iii) $P(\emptyset)=0$

If you have a *countable* collection of disjoint sets,

i.e., if $A_1, A_2, \ldots$ is a countable sequence ofdisjoint sets, then

$$P\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} P(A_i)$$

Example. Consider the experiment of counting the number of accidents in a factory in a month. Let E_i be the event that exactly i accidents occur in a month with i=0, 1, 2, ... Suppose that

$$P(E_i) = \frac{e^{-1}}{i!} \text{ for i = 0, 1, 2 \ldots}$$

and we want to find the probability that the factory has at **least** one accident in a month.

That is, we want to compute $P\left(\bigcup_{i=1}^{\infty} E_i\right)$.

Since $\sum_{i=0}^{\infty} \frac{a^i}{i!} = e^{a}$ ***for any*** *a*, ***we have***

$$\sum_{i=0}^{\infty} P(E_i) = \sum_{i=0}^{\infty} \frac{e^{-1}}{i!}$$

$$= e^{-1} \sum_{i=0}^{\infty} \frac{1}{i!}$$

$$= e^{-1} e^{1} = 1$$

1.3.2 Conditional Probability

The probability of an event A given the event B, denoted $P(A|B)$, is the **value** that solves the equation $P(A \cap B) = P(A|B)P(B)$ provided that $P(B) > 0$.

Thus, $$P(A|B) = \frac{P(A \cap B)}{P(B)}$$

Example. Find the probability of two heads given that you have al least one head in the two tosses.

Sample space $E_1 = \{HH, TH, HT\}$ with probability $\frac{3}{4}$ and $E_2 = (HH)$ with probability $\frac{1}{4}$. So, $Pr\left(\frac{E_2}{E_1}\right) = \frac{Pr(E_1 \cap E_2)}{} = \frac{\frac{1}{4}}{\frac{3}{4}} = \frac{1}{3}$

Generalising the above result, we have for any events $A_1, A_2, \ldots, A_n$

$$P(A_1A_2 \ldots A_n) = P(A_1)P(A_2|A_1)P(A_3|A_1A_2)\ldots P(A_n|A_1A_2\ldots A_{n-1})$$

Remember that $P(.)$ is a probability measure if and only if

1) $P(E) \geq 0$ for all events E
2) $P(\Omega) = 1$
3) If $E \cap F = 0$, then $P(E \cup F) = P(E) + P(F)$

1.3.3 Independence

Two events A and B are **independent** if and only if

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

That is, if **two events** A and B are independent and $P(B) > 0$, then $P(A|B) = P(A)$. Essentially *it says*, **knowing** whether B occurs does not change the probability of A

Moreover, if A and B are independent events, then

- A^c *and* B^c *are independent events,*
- A^c *and B are independent events, and*
- A^c and B^c are independent events.

Bayes' Theorem

Take a complete set of alternatives to a collection of events

$B_1, B_2, \ldots\ldots\ldots, B_n$ such that

1) $B_1 \cup B_2 \cup \ldots\ldots \cup B_n = \Omega$, *and*
2) $B_i \cap B_j = \emptyset$ *for any* $i \neq j$.

Law of Total Probability,

Given an event A and a complete set of alternatives $B_1, B_2, \ldots, B_n$,

$$P(A) = \sum_{i=1}^{n} P(A \cap B_i) = \sum_{i=1}^{n} (A|B_i)P(B_i)$$

Bayes' Theorem: *Let* $B_1, B_2, \ldots B_n$ *be a complete set of alternatives, then*

$$P(B_k \mid A) = \frac{P(A \mid B_k)P(B_k)}{\sum_{i=1}^{n} P(A \mid B_i)P(B_i)}$$

Independence of more than two events

We use the definition of independence given above to get results on **more** than two events.

i) A finite collection of events $A_1, A_2, \ldots, A_n$ are **independent** if and only if

$P(A_{k_1} \cap \ldots \cap A_{k_j}) = P(A_{k_1}) \ldots P(A_{k_j})$

for all $2 \leq j \leq n$ and all $1 \leq k_1 < \ldots < k_j \leq n$.

ii) For $n=3$, the events A_1, A_2, A_3 are independent if and only if

- $P(A_1 \cap A_2) = P(A_1)\,P(A_2)$

  $P(A_1 \cap A_3) = P(A_1)\,P(A_3)$
- $P(A_2 \cap A_3) = P(A_2)\,P(A_3)$
- $P(A_1 \cap A_2 \cap A_3) = P(A_1)P(A_2)P(A_3)$

All of the above must hold for A_1, A_2, A_3 to be independent. If only the first three conditions hold, then we say that A_1, A_2, A_3 are **pair wise independent.** We will discuss this concept with greater details in the next unit.

Check Your Progress 1

1) Write the definitions of the terms: universe set, empty set, disjointed set and power set.

....

....

....

....

....

2) If you toss a coin twice, what would be the sample space? List all the possible events for this experiment.

....

....

....

....

....

3) For a given set A, what is set of all sets?

....

....

....

....

4) Show that for any two events **A** and B,

$$P(A \cup \beta) = P(A) + P(\beta) - p(P \cap \beta)$$

....

....

....

....

....

1.4 RANDQM VARIABLES

Basic Concepts

Definition. A random variable (r.v.) is a function X (.), which assigns to **each** element ω of **Ω** *a* real value X (ω). Thus, we write

$$P(\{\omega \in \Omega : X(\omega) = a\}) \equiv P(\{X(\omega) = a\}) \equiv P(X = a)$$

Examples.

1) Flip a coin *n* times. Here $\Omega = \{H,T\}^n$. The random variable $X \in \{0,1,2,...,n\}$ can be defined to be the number of heads

2) Let $\Omega = \mathbb{R}$, define the two r.v.

   a. $X = \omega$,

   b. $Y = \begin{cases} 1 & \text{for } \omega \geq 0 \\ -1 & \text{otherwise} \end{cases}$

3) Packet arrival times in **the interval** $(0,T]$: Here **Ω** is **the** set of all finite ***length*** strings $(t_1, t_2, ..., t_n) \in (0,T]^*$, define the random variable X to be the **length** of the string $n = 0,1,...$

Cumulative Distribution Function

Definition. For **any** random variable *X*, we define the **cumulative distributian function (CDF)**, $F_x(a)$ as

$$F_x(a) \equiv P(X \leq a)$$

Properties of any cumulative distribution function

1) $\lim_{a \to \infty} F_x(a) = 1$

2) $\lim_{a \to -\infty} F_x(a) = 0$

3) $F_x(a)$ is a non-decreasing function of *a*.

4) $F_x(x)$ is a right-continuous function of *a*. In other words, $\lim_{x \downarrow a} F_x(x) = F_x(a)$

5) For any random variable X and real values $a < b$,
   $P(a < X \leq b) = F_X(b) - F_X(a)$

Discrete **Random** Variables

A random variable is said to be discrete if for some countable set $\mathcal{X} \subset \mathbb{R}$, i.e., $\mathcal{X} = \{x_1, x_2, \ldots\}$, $P\{X \in \mathcal{X}\} = 1$.

Example. From above take

n coin flips: Here $\Omega = \{H, T\}^n$; define the random variable $X \in \{0, 1, 2, \ldots, n\}$ and it can be defined to be the number of heads.

A discrete random variable is completely specified by the probability mass **function** (pmf). Thus, *pmf* of variable X is given by

$$p_X(x_i) \equiv P(X = x_i).$$

Properties of any probability mass function

1) $p_X(x) \geq 0$ for every x

2) $\sum_{all\, x_i} p_X(x_i) = 1$

3) $P(E) = \sum_{x_i \in E} p_X(x_i)$

Examples of Discrete Random Variables

- Bernoulli *r.v.*: $X \sim \mathrm{Br}(p)$, for $0 \leq p \leq 1$ has pmf

$$p_X(1) = p, \text{ and } p_X(0) = 1 - p$$

Geometric *r.v.*: $X \sim \mathrm{Geom}(p)$, for $0 \leq p \leq 1$ has pmf

$$p_X(k) = p(1-p)^{k-1}, \text{ for } k = 1, 2, \ldots.$$

This r.v. represents, for example, the number of coin flips until the first heads shows up (assuming independent coin flips)

Binomial *r.v.*: $X \sim \mathrm{B}(n, p)$, for integer $n > 0$, and $0 \leq p \leq 1$ has pmf

$$p_X(k) = p\binom{n}{k} p^k (1-p)^{(n-k)}, \text{ for } k = 0, 1, 2, \ldots, n$$

This r.v. represents, for example, the number of heads in n independent coin flips.

- Poisson *r.v.*: $X \sim \text{Poisson}(\lambda)$, for $\lambda > 0$, has pmf

$$p_X(k) = \frac{\lambda^k}{k!} e^{-\lambda}, \text{ for } k = 0, 1, 2, \ldots$$

This represents the number of random events in interval $(0, 1]$, e.g., arrivals of packets, photons, customers, etc.

We will discuss these distributions elaborately along with some others subsequently.

Contiauous Random Variables

A random variable X is said to have a continuous distribution if there exists a nonnegative function f such that $P(a \leq x \leq b) = \int_a^b f(x)\,dx$ for every a and b.

To specify a random variable it is sufficient to specify its probability density function (pdf). The pdf of a continuous random variable X is given by

$$fx(a) = \frac{d}{da} F_x(a).$$

Properties of any probability density function

1) $fx(x) \geq 0$ for every x

2) $\int_{-\infty}^{\infty} fx(x)dx = 1$

3) $p(E) = \int_E fx(x)dx$ E

Note that if X is *a discrete' random variable, then*

$$P(X < a) = P(X \leq a) \text{ and}$$

if X is a continuous random variable, then

$$F_x(a) = \int_{-\infty}^{a} fx(x)dx.$$

Examples of contittuous distributions

Uniform distribution

The random variable X is said to have a **uniform distribution** if it has the probability density function given by

$$f_X(x) = \begin{cases} \frac{1}{b-a} & \text{for } a \leq x \leq b \\ 0 & \text{otherwise} \end{cases}$$

Exponential distribution

The random variable X is said to have an **exponential distribution** with parameter $\lambda > 0$ if it has the probability density function given by

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

Gaussian Distribution

Gaussian r.v.: $X - \mathcal{N}(\mu, \sigma^2)$ has pdf

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{\frac{(x-\mu)^2}{2\sigma^2}},$$

where μ is the mean and a is the standard deviation

Transformation of Distribution Functions

Take a random variable X with distribution function $F_X(x)$ and pmf or pdf $f_x(x)$, where $f_X(x) > 0$ for $x \in X$. Suppose that we are interested in the distribution of a function of $Y = g(X)$. The distribution of Y is defined by the equation,

$$Pr(Y \in A) = Pr(g(X) \in A) = Pr(\omega \in \Omega \mid g(X(\omega)) \in A).$$

This is not very helpful, as we have not specified the actual experiment and the random variable. The question is how we can go from these functions for X to the corresponding functions for Y. Let us discuss the general case for such transformation:

Result 1. ***Define the mapping*** $g^{-1}(A)$ ***as:***

$$g^{-1}(A) = \{x \mid g(x) \in A\}.$$

Then

$$F_Y(y) = Pr(\omega \in \Omega \mid Y(\omega) \in (-\infty, y]) = Pr(\omega \in \Omega \mid g(X(\omega)) \in (-\infty, y])$$
$$= Pr(\omega \in \Omega \mid X(\omega) \in g^{-1}((-\infty, y])).$$

However, we need more useful results to rely on transformations that satisfy particular conditions:

Result 2. ***Suppose*** $g(x)$ ***is a strictly monotone functions with inverse*** $g^{-1}(\cdot)$. ***Let be*** $\mathcal{Y}$ ***be*** $\{y \mid g(x) = y$ for some $x \in \mathcal{X}\}$.

(i) ***If X is a discrete random variable with pmf*** $f_X(x)$ ***on*** $\mathcal{X}$, ***then Y is a discrete random variable with pmf.***

$$f_Y(y) = f_X(g^{-1}(y)), \text{ on } \mathcal{Y},$$

(ii) ***If X is a continuous random variable with pdf*** $f_X(x)$ ***on*** $\mathcal{X}$, ***then Y is a continuous random variable with pdf***

$$f_Y(y) = f_X(g^{-1}(y)) \cdot \left| \frac{\partial g^{-1}}{\partial y}(y) \right|,$$

on $\mathcal{Y}$.

In case of continuous **variables,** we **proceed** as follows: **assume** $g(\cdot)$ is increasing, so that

$$F_Y(y) = Pr(Y \le y) = Pr(g(X) \le y)$$
$$= Pr(x \le g^{-1}(y)) = F_X(g^{-1}(y)).$$

Then use the chain **rule** to take the derivative with respect to y you get the **result.**

Example 1. Suppose X has a Poisson distribution with parameter λ. That is, the pmf of X is

$$f_X(x) = \begin{cases} \dfrac{e^{-\lambda}\lambda^x}{x!} & x = 0, 1, 2, \ldots \\ 0 & \text{otherwise,} \end{cases}$$

for some positive λ. Usually we use Poisson distribution for counting events such as arrival of insurance claims.

Take the distribution of $Y = g(X) = 2 \cdot X$. Then inverse of this transformation is $X = g^{-1}(Y) = Y/2$ and the pmf of Y is'given as

$$f_X(x) = \begin{cases} \dfrac{e^{-\lambda}\lambda^{(y/2)}}{(y/2)!} & y = 0,\ 2,\ 4, \ldots \\ 0 & \text{otherwise.} \end{cases}$$

Example 2. Let X be an exponential distribution with pdf

$$f_X(x) = \begin{cases} e^{-x} & x > 0 \\ 0 & \text{otherwise.} \end{cases}$$

Then cdf its distribution is $F_X(x) = 1 - e^{-x}$. Units this distribution, when we write $f_X(x) = \lambda \exp(-\lambda x)$, for positive A, we can use it is widely used for modelling mortalities.

Uniform Distribution

Consider the distribution of $Y = g(X) = 1 - e^{-X}$. This is a **monotone** transformation with inverse $X = g^{-1}(Y) = \ln(1-y)$. Then,

$$\frac{\partial g^{-1}}{\partial y}(y) = \frac{1}{1-y}.$$

The range of Y is $\mathcal{Y} = (0,1)$ and the pdf of Y is

$$f_Y(y) \begin{cases} f_X\left(g^{-1}(y)\right) \cdot \left|\dfrac{\partial g^{-1}}{\partial y}(y)\right| = \exp\left(\ln(1-y)\right) \cdot \dfrac{1}{1-y} = 1 & 0 < y < 1 \\ 0 & \text{otherwise.} \end{cases}$$

Note that its pdf is constant over its range.

Example 3. Consider a non-monotone transformation. Let X be a random variable with pdf

$$f_X(x) = \begin{cases} 1/2 & -1 < x < 1 \\ 0 & \text{otherwise,} \end{cases}$$

It has a uniform distribution on the interval $[-1,\ 1]$. Consider the distribution of non-monotonic formal **action** $Y = g(X) = X^2$. Because of the non-monotonicity we have to do this on a more ad **hoc** basis. In this example we **work** directly through the cumulative distribution functions. For its elements we can split things up into intervals where the transformation is monotone. First, we calculate the cdf for X: .

$$F_X(x) = \begin{cases} 0 & x \le -1 \\ (x+1)/2 & -1 < x \le 1 \\ 1 & 1 < x. \end{cases}$$

Then the cdf for $Y = X^2$ is

$$Pr(Y \le y) = Pr(X^2 < y) = Pr(-\sqrt{y} \le X \le \sqrt{y}) = F_X(\sqrt{y}) - F_X(-\sqrt{y}).$$

so, $$F_Y(y) = \begin{cases} 0 & y \le 0 \\ (\sqrt{y}+1)/2 - (-\sqrt{y}+1)/2 = \sqrt{y} & 0 < y \le 1 \\ 1 & 1 < y. \end{cases}$$

Then the pdf is

$$f_Y(y) = \begin{cases} y^{-1/2}/2 & 0 < y < 1 \\ 0 & \text{otherwise.} \end{cases}$$

Check Your Progress 2

1) Write the definitions of CDF and PMF.

....

....

....

....

2) How do you write the PMF of a Bernoulli random variable?

....

....

....

....

3) What is PDF? Write down its properties.

....

....

....

....

4) How do you write the PDF of

(i) exponential distribution

(ii) Gaussian distribution

....

....

....

....

5) What is meaning of transformation of distribution function?

....

....

....

....

5) What is meaning of transformation of distribution function?

....

....

....

....

....

1.5 EXPECTATIONS

The ekpectation of a random variable X is

(i), if X is a discrete random variable *with pmf* $f_X(x)$, *equal to:*

$$\sum x \cdot f_X(x),$$

provided the sum $\sum_x |x| f_X(x)$ exists, .

(ii), if X is a *continuous random* variable *with pdf* $f_X(x)$, *equal to*

$$\int_x x \cdot f_X(x)\,dx.$$

provided *the integral* $\int_x |x| f_X(x)\,dx$ *exists.*

Example 1. Suppose you toss a **single** die. Let X, be the number (X) on top of the die, has a distribution with *pmf*:

$$f_X(x) = \begin{cases} 116 & x = 1,2,3,4,5,6, \\ 0 & \text{otherwise.} \end{cases}$$

Then the expected value is

$$E(X) = \sum_{k=1}^{6} k \cdot (1/6) = 7/2.$$

Example 2. Suppose X has an exponential distribution with parameter λ and *pdf*

$$f_X(x) = \begin{cases} \lambda \exp(-Ax) & x > 0 \\ 0 & \text{otherwise.} \end{cases}$$

As we have seen above the exponential distribution'with $\lambda = 1$ can be used to fit distributions of durations such as waiting times of insurance claim arrivals. Its expected value is

$$= -x\exp(-\lambda x)\Big|_0^\infty \int_0^\infty \exp(-\lambda x)\,dx$$

$$= 0 + \int_0^\infty \exp(-\lambda x)\,dx$$

$$= \frac{-1}{\lambda}\exp(-\lambda x)\Big|_0^\infty$$

$$= \frac{1}{\lambda}$$

If you are interested in finding expectations of transformations of X, then the expected values for transformations have to be defined implicitly:

$$E[r(X)] = E[Y], \text{ where } Y = r(X).$$

That means we do not need to define expectations of transformations. To calculate these expectations we take the help of the following results:

Result 1. The *expectation of* a *function* $r(\cdot)$ *of a random variable* X *is*

(i), if X is a discrete *random variable with pmf* $f_X(x)$,

$$\sum_x r(x)\cdot f_X(x).$$

(ii), if X *is* a *continuous random* variable *with pdf* $f_X(x)$,

$$\int_x r(x)\cdot f_X(x)\,dx.$$

For the discrete case with a monotone transformation:

$$E(Y) = \sum_y y.f_X\left(r^{-1}(y)\right) = \sum_x r(x) f_X(x).$$

For the continuous case with $r(\cdot)$, a monotone (increasing) transformation, is given below:

Let $Y = r(X)$. Then

$$E[r(X)] = E(Y) = \int_y y.f_Y(y)\,dy = \int_y y f_X\left(r^{-1}(y)\right)\cdot\frac{\partial r^{-1}}{\partial y}(y)\,dy.$$

Transform the integrand from y to $z = r^{-1}(y)$ with inverse transformation $y = r(z)$ to get

$$\int_y y f_X\left(r^{-1}(y)\right)\cdot\frac{\partial r^{-1}}{\partial y}(y)\,dy = \int_z r(z) f_X(z)\cdot\frac{\partial r^{-1}}{\partial r(z)}\left(r(z)\right)dr(z)$$

$$= \int r(z) f_X(z)\cdot\frac{\partial r^{-1}}{\partial y}\left(r(z)\right)\frac{\partial r(z)}{\partial z}(z)\,dz$$

$$= \int_z r(z) f_X(z)\,dz.$$

For the last step, we have taken the following: Since $z = r^{-1}(r(z))$, *by* the chain rule for differentiation we differentiate both sides with respect to

$$z,\ 1 = \frac{\partial r^{-1}}{(z)}\left(r(z)\right)\frac{dr(z)}{\partial z}(z).$$

Then $\frac{\partial r^{-1}}{\partial r(z)}\left(r(z)\right) = 1\Big/\frac{\partial r}{\partial z} r(z).$

Note some special expectations. In all such cases X is a random variable with pdf/pmf equal to $f_X(x)$:

1) Mean is another name for the expectation or expected value of X, $\mu = E(X)$.

2) K^{th} *moment* of X is $\mu_k = E(X^k)$ (i.e., $\mu_1 = \mu$).

3) *Variance* of X is $V(X) = \sigma^2 = E\left((X-\mu)^2\right) = \mu_2 - \mu_1^2$.

4) **Moment generating function** (mgf), denoted by $M_X(1)$. It is the expected value of $\exp(t \cdot X)$. We are interested in this function for values of t around zero. It has a number of interesting properties:

   (a) It uniquely defines the distribution of a random variable: if two random variables have the same mgf for all t in an interval around zero, they have the same cdf and pmf/pdf (up to sets of measure zero).

   (b) K^{th} moment is equal to k^{th} derivative of the mgf evaluated at zero:

   $$\mu_k = \frac{\partial^k M_X}{\partial t^k}(0).$$

   In particular, using $M_X^k(t)$ as shorthand for $\frac{\partial^k M_X}{\partial t^k}(t)$, we have $M_X(0) = 1$, $\mu = M_X^1(0), \sigma^2 = M_X^1(0) - M_X(0)^2$.

5) **Cumulative generating function** is the log of the moment generating function, $K,(t) = \ln M_X(t)$. Here $\mu = K_X^1(0)$ and $\sigma^2 = K_X^2(0)$.

6) **Characteristic function** is defined as

   $$\psi_X(t) = E\left[\exp(X\,it)\right],$$

   where $i = \sqrt{-1}$. The advantage working with. the characteristic function rather than the moment generating function is that the former always exists, while the latter does not exist.

7) Expectation of a linear function of a random variable is the linear function of the expectation:

   $$E(a + b \cdot X) = a + b \cdot E(X).$$

8) Variance of a linear function of a random variable is the variance of the random variable multiplied by the square of the slope coefficient:

   $$V(a + b \cdot X) = b^2 \cdot V(X).$$

Example 3. Consider the exponential distribution with *pdf* $\lambda\exp(-\lambda x)$ for $x > 0$ and zero otherwise. We have already calculated its expected value above. Using the moment generating function, we can derive its expectation and avoid the procedure integration by parts.

Thus,

$$M_X(t) = \int_0^\infty \lambda\exp(-\lambda x)\exp(tx)\,dx = \int_0^\infty \lambda\exp(-x(\lambda - t))\,dx$$

$$= \frac{-\lambda}{\lambda - t}\exp(-x(\lambda - t))\Big|_0^\infty = \frac{\lambda}{\lambda - t}.$$

However, it only works for $t < \lambda$. If you are interested only about values of t around zero, you can use this result. The derivative of the *mgf* is

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

so that

$$E\left[X^k\right] = M_X^k(0) = \frac{k!}{\lambda^k}.$$

Its mean is $1/\lambda$, the second moment $2/\lambda^2$, and the variance $1/\lambda^2$.

Example 4. Suppose that an experiment can have one of two outcomes, success or failure. Let **p** be the probability of success. Let Y be the indicator for success, equal to one if the experiment is a success and zero otherwise. This is called as a Bernoulli trial. When we repeat this experiment n times, and assume that the repetitions are independent, we have the following result:

Let Y_i for $i = 1,\ldots,n$ denote the success in the *i*th Bernoulli trial. Define $X = \sum_{i=1}^n Y_i$ as the total number of successes. Then X has a binomial distribution. To find its *pmf*, consider the probability of particular sequence of x successes and $n - x$ failures. For example, first x one's and then $n - x$ zero's. **The** probability of **any** one such a sequence is $p^x(1-p)^{n-x}$. To get **the** probability of x successes and $n - x$ failures, we need to count the number **of** such sequences. This is equal to the number of ways you can select x objects out of a set of n, which is $\binom{n}{x}$. Hence the *pmf* of X is

$$f_X(x) = \binom{n}{x}\cdot p^x\cdot(1-p)^{(n-x)},$$

for $x = 0, 1, 2, \ldots, n$, and zero otherwise.

The mean of the binomial distribution is given by *pmf.*

$$\sum_{x=0}^n \binom{n}{x}\cdot p^x\cdot(1-p)^{(n-x)} = \sum_x f_X(x) = \ldots$$

Therefore,

$$E(X)=\sum_{x=0}^{n}x\cdot\binom{n}{x}\cdot p^{x}\cdot(1-p)^{(n-x)}=\sum_{x=1}^{n}x\cdot\binom{n}{x}\cdot p^{x}\cdot(1-p)^{(n-x)}$$

$$=\sum_{x=1}^{n}x\cdot\frac{n!}{x!(n-x)!}p^{x}\cdot(1-p)^{(n-x)}=\sum_{x=1}^{n}x\cdot\binom{n}{x}\cdot p^{x}\cdot(1-p)^{(n-x)}$$

$$=\sum_{x=1}^{n}n\cdot p\cdot\frac{(n-1)!}{(x-1)!((n-1)-(x-1))!}\cdot p^{(x-1)}\cdot(1-p)^{((n-1)-(x-1))}$$

Let $m=n-1$ and $y=x-1$. Then instead of summing from $x=1$ to $x=n-1$ we take the sum from $y=0$ to $y=m$ and get:

$$E(X)=n\cdot p\cdot\sum_{y=0}^{m}\frac{m!}{y!(m-y)!}\cdot p^{y}\cdot(1-p)^{(m-y)}$$

$$=n\cdot p\cdot\sum_{y=0}^{m}\binom{m}{y}\cdot p^{y}\cdot(1-p)^{(m-y)}=np.$$

Another approach to get the above result is to write X as the sum of independent and identically distribution random variables, $X=\sum_{i=1}^{n}Y_i$, with

$$f_Y(y)=p^{y}\cdot(1-p)^{(1-y)},$$

and mean p. Then the mean of X is given as

$$E(X)=E\left(\sum_{i=1}^{n}Y_i\right)=\sum_{i=1}^{n}E(Y_i)=np.$$

Result 2. Chebyshev's Inequality. ***For any random variable Y with mean μ and variance σ^2, and for any $k>0$***

$$Pr(|Y-\mu|\geq k\cdot\sigma)\leq 1/k^2.$$

Proof. Let X be a nonnegative random variable. Then, for any positive c,

$$Pr(X\geq c)\leq\frac{E(X)}{c}.$$

Taking the continuous case, we get the above result as,

$$E(X)=\int_0^{\infty}x f_X(x)dx$$

$$=\int_c^{\infty}x f_X(x)dx+\int_0^{c}x f_X(x)dx$$

$$\geq\int_c^{\infty}x f_X(x)dx+$$

$$\geq\int_c^{\infty}c f_X(x)dx=c\cdot Pr(X\geq c).$$

When $X = (Y - \mu)^2$ and $c = k^2\sigma^2$, we rewrite the inequality as

$$Pr(X \geq c) \leq E[X]/c$$

$$\Rightarrow \quad Pr\left((Y-\mu)^2 \geq k^2\sigma^2\right) \leq 1/k^2,$$

1.6 SPECIAL DISTRIBUTIONS

1) **Gamma** Distributions. In contrast to an exponential distribution representing the waiting time till the occurrence of first event, Gamma distribution represents the waiting time till the rth event. If X is this waiting time, then

$$F_X(x) = Pr(X \leq x) = 1 - \text{Pr (less than r events in interval } [0, x]).$$

Thus, if we consider the probability that a Poisson random variable with arrival rate λx is less than r, then

$$F_X(x) = 1 - \sum_{k=0}^{r-1} \frac{e^{-\lambda x}(\lambda x)^k}{k!}$$

The probability density function can be obtained by differentiating with respect to x:

$$f_X(x) = \sum_{k=0}^{r-1} \frac{\lambda^{k+1} x^k e^{-\lambda x}}{k!} - \sum_{k=1}^{r-1} \frac{\lambda^k x^{k-1} e^{-\lambda x}}{(k-1)!}.$$

Note that the second summand is only from $k = 1$ to $r - 1$, not from $k = 0$ to $r - 1$. Changing the second summation from $k = 1$ to $k = r - 1$ the summation from $k = 0$ to $r - 2$, we can write this as

$$f_X(x) = \sum_{k=0}^{r-1} \frac{\lambda^{k+1} x^k e^{-\lambda x}}{k!} - \sum_{k=0}^{r-2} \frac{\lambda^{k+1} x^k e^{-\lambda x}}{k!}$$

$$= \frac{\lambda^r}{(r-1)!} x^{r-1} e^{-\lambda x}.$$

So, we get a gamma distribution with parameters r and A.

The gamma distribution is more general. We don't require r to be an integer while formulating it. But when we do this, the interpretation of waiting for the rth event cannot be made.

The general pdf for a Gamma distributed random variable X is obtain when $\beta = \frac{1}{a}$ and $\alpha = r$:

$$f_X(x) = \frac{x^{\alpha-1} e^{-x/\beta}}{\Gamma(\alpha)\beta^\alpha},$$

for $x > 0$. In this formulation the parameters α and β are both positive and the function is defined as

$$\Gamma(a) = \int_0^\infty t^{\alpha-1} e^{-t} dt.$$

It's properties are:

$$\Gamma(\alpha+1)=\alpha\cdot\Gamma(\alpha),\quad \Gamma(1)=\int_0^\infty e^{-1}dt=1,\text{ and }\quad \Gamma(1/2)=\sqrt{\pi},$$

so that

$$\Gamma(n)=(n-1)!,$$

for integer n. Take a random variable Y with an exponential distribution with arrival rate A. Then its moment generating function is

$$M_Y(t)=\mathrm{E}\left[\exp(tY)\right]=\frac{\lambda}{A-t}=\frac{1}{1-\beta t},$$

where $\beta=1/\lambda$ is the mean of the exponential distribution.

Let $Y_1,\ldots,Y_\alpha$ be independent exponential random variables with mean β. Then the moment generating function for the gamma distribution is

$$M_X(t)E\left[\exp\left(\sum_{i=1}^{\alpha}Y_i t\right)\right]=\prod_{i=1}^{\alpha}E\left[\exp(Y_i t)\right]=\prod_{i=1}^{\alpha}\frac{1}{1-\beta t}=\frac{1}{(1-\beta t)^{\alpha}}.$$

The cumulant generating function is

$$K_X(t)=-a\ln(1-\beta t),$$

with

$$K'_X(t)=\frac{\alpha\beta}{1-\beta t},\qquad K''_X(t)=\frac{\alpha\beta^2}{(1-\beta t)^2}.$$

The mean and variance are $\alpha\beta$ and $\alpha\beta^2$ respectively.

2) **Chi-squared Distribution.** Chi-squared distribution is a special case of Gamma distribution if we take $a=k/2$, where **k** is a positive integer, and. $\beta=2$, than we have a Chi-squad distribution with degrees *of freedom* equal to ***k*. Its** *pdf* is

$$f_X(x)=\frac{x^{k/2-1}e^{-x/2}}{\Gamma(k/2)2^{k/2}},$$

for x positive. Its moment generating function is

$$M_X(t)=\frac{1}{(1-2t)^{k/2}},$$

and its mean and variance are k and ***2k*** respectively,

3) **Normal Distribution.** The normal distribution is of fundamental importance as an approximation to a large number of statistics through tht **central limit theorem.** A'random variable **X** has a normal distribution with parameters μ and σ^2, denoted by $\mathcal{N}(\mu,\sigma^2)$, if its *pdf* is

$$f_X(x)=\frac{1}{\sqrt{2\pi^2}}\exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right),$$

for $-\infty < x < \infty$, and $\sigma^2 > 0$. If we consider the moment generating function, then the fact that thepdf and *pmf* add up to one; viz.,

$$M_X(t) = \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(xt - \frac{(x-\mu)^2}{2\sigma^2} \right) dx$$

$$= \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-\mu)^2 - 2\sigma^2 xt}{2\sigma^2} \right) dx$$

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

$$= \exp\left(\mu t + \sigma^2 t^2/2\right) \cdot \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-\mu-\sigma^2 t)^2}{2\sigma^2} \right) dx$$

$$= \exp\left(\mu t + \sigma^2 t^2/2\right).$$

As the cumulant generating function is $K_X(t) = \mu t - 0^2 t'/2$ and hence the mean is μ and the variance σ^2.

Properties of normal distribution: Linear transformations of normal random variables are also normally distributed. Take a random variable X with a $\mathcal{N}(\mu, \sigma^2)$ distribution, and consider the transformation $Y = a + bX$. Then,

$$M_Y(t) = E\left(e^{tY}\right) = E\left(e^{(a+bX)t}\right) = e^{at} \cdot E\left(e^{(bt)X}\right)$$

$$= e^{at} \cdot M_X(\text{bi}) = \exp(a + \mu bt + \sigma^2 b^2 t^2/2)$$

$$= \exp((a + b\mu)t - b^2\sigma^2 t^2/2)$$

$$= \exp\left(\tilde{\mu} t + \tilde{\sigma}^2 t^2/2\right).$$

Hence Y has a normal distribution with mean $\tilde{\mu} = a + b\mu$ and variance $\tilde{\sigma}^2 = b^2\sigma^2$. When you take $Y = (X-\mu)/\sigma$, then the transformation is called $Y - \mathrm{N}(0,1)$, the standard normal distribution with mean zero and unit variance.

See that the normal distribution and the chi-squared distribution are closely connected. If X has a standard normal distribution $\mathcal{N}(0,1)$, then $Y = X^2$ has a Chi-squared distribution with degrees of freedom equal to one. If we work with the moment generating function of *Y*, then

$$M_Y(t) = E\left(e^{tY}\right) = E\left(e^{tX^2}\right) = \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}x^2 + tx^2 \right) dx$$

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

$$= \sqrt{1/(1-2t)} \int_{-\infty}^{\infty} \frac{1}{\sqrt{2/(1+2t)}} \exp\left(-\frac{1}{2/(1-2t)} x^2\right) dx = \frac{1}{(1-2t)^{1/2}}.$$

We get the *mgf* for a Chi-squared distribution with degrees of freedom equal to one.

4) **Cauchy Distribution. A** random variable X has a Cauchy distribution centered around θ if it has ***pdf***

$$f_X(x) = \frac{1}{\pi} \cdot \frac{1}{1+(x-\theta)^2},$$

for $-\infty < x < \infty$. This distribution has no moments. The median and mode are both equal to θ. The ***pdf*** is similar to the ***pdf*** for the normal distribution with a thicker tail. We use it for modelling variables with high kurtosis, that is, which infrequently take on extremely large values, such as stock prices, event like loss in Tsunami. An interesting property of the Cauchy distribution is that if the sequence of independent random variables $X_1, X_2, \ldots, X_n$ all have the Cauchy distribution centered around θ, then the average $\bar{X} = \sum_{i=1}^{n} X_i / n$ also has that same Cauchy distribution centered around θ, with the same quantile. In other words, laws of large numbers will not to apply to the ones like Cauchy distributions.

5) **Beta Distribution.** If **two** independent random variables Y_1 and Y_2 have Gamma distributions with parameters α and 1 and β and 1, then the ratio $X = Y_1/(Y_1+Y_2)$ has a Beta distribution with parameters α and β. Its *pdf* is

$$f_X(x) \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\cdot\Gamma(\beta)} x^{\alpha-1} \cdot (1-x)^{\beta-1},$$

for $0 < x < 1$ and zero elsewhere. The mean and variance are $\alpha/(\alpha+\beta)$ and $\alpha\beta/\left((\alpha+\beta)^2(\alpha + \beta+1)\right)$ respectively. For modelling variables that take on values in the unit interval, we can apply this distribution.

6) Let X have a standard normal distribution (i.e., a normal distribution with mean zero and unit variance). Let W have a Chi-squared distribution with r degrees of freedom, and let X and W be independent. Then $Y = X/\sqrt{(W/r)}$ has a t-distribution with degrees of freedom equal to r. The probability density function of Y is:

$$f_Y(y) = \frac{\Gamma(r+1/2)}{\sqrt{\pi r}\,\Gamma(r/2)} \frac{1}{(1+t^2/r)^{(r+1)/2}},$$

for $-\infty < y < \infty$. The t-distribution has thicker tails than the normal distribution. As the degrees of freedom gets larger, then the t-distribution gets closer to the normal distribution. (For $E[W/r] = 1$, and $V(W/r) = 2/r$, as $r \to \infty$, the mean of the denominator is one, but the variance is zero.)

7) If V and W are independent Chi-squared random variables with degrees of freedom equal to r_1 and r_2 respectively, then $Y = (V/r_1)/(W/r_2)$ has an F-distribution with degrees of freedom equal to r_1 and r_2. The probability density function is

$$f_Y(y) = \frac{\Gamma((r_1+r_2)/2)(r_1/r_2)^{r_1/2}}{\Gamma(r_1/2)\Gamma(r_2/2)} \cdot \frac{y^{r_1/2-1}}{(1+wr_1/r_2)^{(r_1+r_2)/2}},$$

for positive values of y, and zero elsewhere.

Important Exponential of Distributions.

Definition 1. A *family of pdf's (or pmf's) is an exponential family if it can be written as*

$$f_X(x \mid \theta) = h(x) \cdot c(\theta) \exp\left(\sum_{i=1}^{k} w_i(\theta) \cdot t_i(x)\right),$$

for $-\infty < x < \infty$.

The key feature of this representation is that we can separate the distribution into functions of the parameters θ and functions of the variables x. If we parameterise the above exponential distribution as

$$f_X(x \mid \eta) = h(x) \cdot c(\eta) \exp\left(\sum_{i=1}^{k} \eta_i \cdot t_i(x)\right),$$

for $-\infty < x < \infty$, then to η as the natural parameters. Consequently we have cases like:

1) Binomial distribution:

$$f_X(x) = \binom{n}{x} \cdot p^x \cdot (1-p)^{(n-x)},$$

for $x = 0, 1, 2, \ldots, n$, and zero otherwise. This can be written as

$$f_X(x) = 1\{x \in \{0, 1, 2, \ldots, n\}\} \cdot \binom{n}{x} \cdot (1-p)^n \cdot \exp(x \ln p/(1-p)),$$

so, $k = 1$, and

$$h(x) = 1\{x \in \{0, 1, 2, \ldots, n\}\} \cdot \binom{n}{x},$$

$$c(p) = (1-p)^n,$$

$$w_1(p) = \ln(p/(1-p)),$$

$$t_i(x) = x.$$

2) Poisson distribution:

$$f_X(x) = 1\{x \in \{1, 2, \ldots\}\} \cdot (1/x!) \cdot \exp(-\lambda) \exp(x \ln \lambda).$$

3) Exponential distribution:

$$f_X(x) = 1\{x > 0\} \cdot \lambda \cdot \exp(-x\lambda).$$

4) Normal distribution:

$$f_X(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(\mu^2 / (2/\sigma^2)\right) \cdot \exp\left(-x^2/(2\sigma^2) + x\mu/\sigma^2\right).$$

A distribution that does not fit in the above group is the uniform distribution:

$$f_X(x) = 1\{a < x < b\} \cdot \frac{1}{b-a}.$$

In case of indicator function $1\{a < x < b\}$ you cannot factor into a function of the parameters and a function of the variable.

**Joint Distributions Conditional Distributions,
Independence of Random Variables**

Definition 1. Let any N – dimensional *random* variable be *a function* from the sample space to R^N.

Definition 2. Take (X,Y) be a discrete bivariate vector. Then the function $F_{XY}(x,y)$ from R^2 to R, defined by

$$f_{XY}(x,y) = Pr(\omega \in \Omega \mid X(\omega) = x, Y(\omega) y),$$

is the joint probability mass function of (X,Y).

Example 1. Toss a coin three times. The sample space is

$$\Omega\{HHH, HHT, HTH, HTT, THH, THT, TTH, TTT\}.$$

An example of a bivariate random variable is (X,Y), where $X(\omega)$ is the number of heads in the first two tosses, and $Y(\omega)$ is the number of heads in the two tosses. Assuming the coin is a fair one, and the probability of heads is $1/2$ at every toss, with different tosses independent events, the joint *pmf* is

$$f_{XY}(x,y) = \begin{cases} 1/8 & \text{if } (x,y) = (0,0) & (\omega \in \{TTT\}), \\ 1/8 & \text{if } (x,y) = (0,1) & (\omega \in \{TTH\}), \\ 1/8 & \text{if } (x,y) = (1,0) & (\omega \in \{HTT\}), \\ 2/8 & \text{if } (x,y) = (1,1) & (\omega \in \{HTH, THT\}), \\ 1/8 & \text{if } (x,y) = (1,2) & (\omega \in \{THH\}), \\ 1/8 & \text{if } (x,y) = (2,1) & (\omega \in \{HHT\}), \\ 1/8 & \text{if } (x,y) = (2,2) & (\omega \in \{HHH\}), \\ 0 & \textit{otherwise} & \end{cases}$$

Note ihat even was we define two random variables, the definition of the probability mass function for any of the variable does not change. Ignoring the definition of $Y(\omega)$, the *pmf* for x when you do not consider $X(\omega)$ is

$$f_X(x) = \begin{cases} 1/4 & \text{if } x = 0 \quad (\omega \in \{TTT, TTH\}), \\ 2/4 & \text{if } x = 1 \quad (\omega \in \{HTH, HTT, THH, THT\}), \\ 1/4 & \text{if } x = 2 \quad (\omega \in \{HHT, HHH\}), \\ 0 & elsewhere. \end{cases}$$

Similarly we can calculate the *pmf* for Y without x. In case of joint random variables, we typically refer to these variable probability mass functions as marginal probability mass functions. These functions general they can be calculated from the joint *pmf* as

$$f_X(x) = Pr(\omega \in \Omega \mid X(\omega) = x)$$

$$= Pr(\omega \in \Omega \mid X(\omega) = x, -\infty < Y(\omega) < \infty) = \sum_y f_{XY}(x,y),$$

Similarly,

$$f_Y(y) = \sum_x f_{XY}(x,y).$$

Given the joint *pmf*, we can calculate probabilities involving both random variables. For example,

$$Pr(X \le 1, Y \ge 1) = Pr(\omega \in \Omega \mid X(\omega) \le 1, Y(\omega) \ge 1) = \sum_{x\le 1, y\ge 1} f_{XY}(x,y).$$

$$= f(0,1) + f(1,1) + f(1,2) = 1/2.$$

Alternatively we can have joint continuous random variables:

Definition 3. *Let (X,Y) be **a** continuous bivariate vector.* ***Then the*** *function $f_{XY}(x,y)$ from R^2 to R is the joint probability density function of (X,Y) if it satisfies*

$$\int_A f_{XY}(x,y)\,dxdy = Pr(\omega \in \Omega \mid (X(\omega), Y(\omega)) \in A),$$

for ail sets A.

Example 2. Pick a point randomly in the triangle $(0,0),(0,0),(0,1),(1,0)$. Let X be the x-coordinate and y be the y-coordinate. A reasonable choice for the joint pdf appears to be

$$f_{XY}(x,y) = c,$$

for $\{(x,y) \mid 0 < x < 1, 0 < y < 1, x + y < 1\}$ and zero otherwise (that is, constant on the domain).. What is the appropriate value for c? Since we know that the probability of the entire sample space is equal to one, we can write

$$1 = Pr(\omega \in \Omega) = Pr((X,Y) \in \{(x,y) \mid 0 < x < 1, 0 < y < 1, x + y < 1\})$$

$$= \int_{\{(x,y)|0<x<1,0<y<1,x+y<1\}} f_{XY}(x,y)\,dxdy = \int_0^1\int_0^{1-y} c\,dxdy$$

$$= \int_0^1 c\cdot(1-y)\,dy = c/2.$$

Hence $c = 2$.

Another way of writing such integrals differently is:

$$\int_{\{(x,y)|0<x<1,0<y<1,x+y<1\}} g(x,y)dxdy = \int_0^1\int_0^{1-y} g(x,y)\,dxdy = \int_0^1\int_0^{1-x} g(x,y)\,dydx,$$

depending on the order of integration, without changing the result. In such a case the marginal *pdf*'s are defined as

$$f_X(x) = \int_{-\infty}^{\infty} f_{XY}(x,y)\,dy = \int_0^{1-x} 2dy = 2-2x,$$

for $0 < x < 1$ and zero otherwise, Similarly,

$$f_Y(y) = 2-2y, \qquad 0 < y < 1, \text{ and zero otherwise.}$$

In joint random variable case, expectations are calculated in the same way as before, but using double sums or integrals at present:

$$E[r(X,Y)] = \int_x\int_y r(x,y)f_{XY}(x,y)\,dydx,$$

for continuous bivariate random variables, and

$$E[r(X,Y)] = \sum_x\sum_Y r(x,y)\,f_{XY}(x,y),$$

for discrete bivariate random variables.

Example 3. Consider the expected value of XY in the previous example:

$$E(XY) = \int_0^1\int_0^{1-x} 2xydydx = \int_0^1 xy^2\Big|_0^{1-x} dx = \int_0^1 x(1-x)^2\,dx$$

$$= \int_0^1 x^3 - 2x^2 + xdx = \frac{1}{4}x^4 - \frac{2}{3}x^3 + \frac{1}{2}x^2\Big|_0^1 = \frac{1}{12}$$

Recall how the conditional probability of events was defined:

$$Pr(E_1 \mid E_2) = \frac{Pr(E_1 \cap E_2)}{Pr(E_2)},$$

provided $Pr(E_2) > 0$.

Case of bivariate discrete random variables.

Definition 4. *Given two* discrete random variables X and Y *with joint probability mass* function $f_{XY}(x,y)$, *the* conditional probability mass function for X given $Y = y$ is

$$f_{X|Y}(x \mid y) = \frac{f_{XY}(x,y)}{f_Y(y)},$$

provided $f_Y(y) > 0$.

Example 4. Consider the distribution of X given that Y = 1. In that case we are conditioning on $\omega \in \{HTH, HHT, THT, TTH\}$

$$f_{X|Y}(x \mid Y = 1) = \begin{cases} 1/4 & x = 0 & (\omega \in \{TTH\}), \\ 2/4 & x = 1 & (\omega \in \{HTH, THT\}), \\ 1/4 & x = 2 & (\omega \in \{HHT\}), \\ 0 & elsewhere. \end{cases}$$

Note that this is the same as the marginal *pmf* of X and will not be true if we condition on Y = 2. In that case we condition on $\omega \in \{HHH, THH\}$ such that.

$$f_{X|Y}(x \mid Y = 2) = \begin{cases} 1/2 & x = 1 & (\omega \in \{THH\}), \\ 1/2 & x = 2 & (\omega \in \{HHH\}), \\ 0 & elsewhere. \end{cases}$$

Definition 5. (Continuous **r.v.**) Given two continuous random variables X ***and*** Y with joint probability density function $f_{XY}(x, y)$, the conditional probability density function for X given Y = y is

$$f_{X|Y}(x \mid y) = \frac{f_{XY}(x,y)}{f_Y(y)}.$$

To see this formulation we first take the conditional cumulative distribution function where we condition on $Y \in (y, y + \Delta)$, which does have positive probability for nonzero, Δ:

$$Pr(X \le x) \mid Y \in (y, y+\Delta) = \frac{Pr(X < x, y < Y < y + \Delta)}{Pr(y < Y < y + \Delta)}$$

$$\frac{\int_{-\infty}^{x}\int_{y}^{y+\Delta} f_{XY}(u,v)\,dv\,du}{\int_{-\infty}^{\infty}\int_{y}^{y+\Delta} f_{XY}(u,v)\,dv\,du}.$$

Now take the limit as $\Delta \to 0$. In that case both numerator and denominator go to zero, so to get the limit we have to use. L' Hospital's rule and differentiate both numerator and denominator with respect to Δ. In that case we get:

$$F_{X|Y}(x \mid y) = \frac{\int_{-\infty}^{x} f_{XY}(u,y)\,du}{\int_{-\infty}^{\infty} f_{XY}(u,y)\,du} = \frac{\int_{-\infty}^{x} f_{XY}(u,y)\,du}{f_Y(y)}$$

Take the derivative with respect to x to get the conditional probability density function:

$$f_{X|Y}(x \mid y) = \frac{f_{XY}(x,y)}{f_Y(y)}.$$

However, this does not yield the conditional density of X given $Y = y$. Its interpretation is that of the limit of the conditional densities, conditioning on $y \le Y < y + \Delta$, and taking the limit as Δ goes to zero.

Definition 6. Let (X,Y) be a *bivariate* random *vector* with *pmf/pdf* $f_{XY}(x,y)$. Then the *random* variables X and Y are independent if

$$f_{XY}(x,y) = f_X(x) \cdot f_Y(y),$$

where $f_X(x)$ *and* $f_Y(y)$ are the *marginal pdf/pmf's of X and Y*.

Result 1. The two random variables X and Y are independent if and only if the joint *pdf/pmf can* be written as

$$f_{XY}(x,y) = g(x) \cdot h(y),$$

for all $-\infty < x < \infty$ and $-\infty < y < \infty$.

Proof:

If X *and* Y are independent, then we can choose, and $g(x) = f_X(x)$, and $h(y) = f_Y(y)$ and the result is trivial. Now suppose we can factor the joint density of X and Y as $f_{XY}(x,y) = g(x) \cdot h(y)$. Then the marginal density of X is

$$f_X(x) = \int_y f_{XY}(x,y)\,dy = g(x)\int_y h(y)\,dy.$$

Similarly

$$f_Y(y) = \int_x f_{XY}(x,y)\,dx = h(y)\int_x g(x)\,dx.$$

Multiply $f_x(x)$ and $f_y(y)$, we get,

$$f_X(x) \cdot f_Y(y) = g(x)h(y)\int_x g(x)\,dx \int_y h(y)\,dx$$

$$= f_{XY}(x,y) \cdot \int_x \int_y f_{XY}(x,y)\,dxdy = f_{XY}(x,y),$$

Result 2. *Let* X and Y be two independent random variables, Then

(i) $f_{X|Y}(x \mid y) = f_X(x)$, *and*

(ii) $f_{Y|X}(y \mid x) = f_Y(y)$.

Example 5. In the previous example we had

$$f_{XY}(x,y) = 2 \cdot 1\{0 < x < 1\} \cdot \{0 < y < 1\} \cdot 1\{x + y < 1\}, \text{ and zero elsewhere.}$$

It may appear that X and Y are independent by the previous results: Choose $g(x) = 2$ and $h(y) = 1$. The reason this does not work is because of the restrictions on the values of X and Y. To see this write

$$f_{XY}(x,y) = 2 \cdot 1\{0 < x < 1\} \cdot 1\{0 < y < 1\} \cdot 1\{x + y < 1\},$$

for all $-\infty < x < \infty$ and $-\infty < y < \infty$. The indicator *function* $1\{A\}$ is equal to one if the argument is true and zero if it is fails. This indicator function $1\{x+y<1\}$ cannot be separated into a function of x and a function of y and therefore X and Y are not independent.

Let there be two random variables X and Y with means μ_X and μ_Y and variances σ_X^2 and σ_Y^2 respectively. Then the covariance of X and Y is defined as

$$C(X,Y) = E\left[(X-\mu_X)\cdot(Y-\mu_Y)\right],$$

and the correlation coefficient as

$$\rho_{XY} = C(X,Y)/(\sigma_X \sigma_Y).$$

The mean and variance of the sum $X+Y$ are

$$E[X+Y] = \mu_X + \mu_Y,$$

and

$$V(X_Y) = \sigma_X^2 + \sigma_Y^2 + 2\cdot C(X,Y)$$

If the two random variables are independent, the covariance is zero (note that the reverse is not necessarily true), and the variance simplifies to the sum of the variances $\sigma_X^2 + \sigma_Y^2$.

To see how tricky conditional probability density functions can be with continuous random variables, consider the **Borel-Kolmogorov paradox.** The joint density of two random variables X and Y is

$$f_{YX}(y,x) = \exp(-x-y),$$

for $x, y > 0$. The marginal density function of X is

$$f_X(x) = \exp(-x),$$

The conditional density of X given $Y=1$ is

$$f_{X|Y}(x \mid y=1) = \exp(-x),$$

'because X and Y are obviously independent.

Taking the transformation from (X,Y) to (X,Z) where Z is $(Y-1)/X$. The inverse of the transformation is $X = g_1(X,Z) = X$ and $Y = g_2(X,Z) = XZ+1$ with Jacobian (the correction factor)

$$\begin{vmatrix} \frac{\partial g_1}{\partial x}(x,z) & \frac{\partial g_2}{\partial x}(x,z) \\ \frac{\partial g_1}{\partial z}(x,z) & \frac{\partial g_2}{\partial z}(x,z) \end{vmatrix} = \begin{vmatrix} 1 & z \\ 0 & x \end{vmatrix} = |x|.$$

the joint density of Z and X is

$$f_{ZX}(z,x) = x\cdot\exp(-x-xz-1),$$

for $z > -1/x$ and $x > 0$.

Let us calculate the marginal density of z. For $z > 0$, the marginal density is

$$f_Z(z) = \int_0^\infty x\exp(-x(z+1))e^{-1}dx$$

$$= e^{-1}\frac{1}{z+1}\int_0^\infty x(z+1)\exp(-x(z+1))$$

$$= \frac{\;}{e(z+1)^2}.$$

For $z < 0$, the marginal density is

$$f_Z(z) = \int_0^{-1/z} x\exp(-x(z+1))e^{-1}dx = e^{-1}\frac{1}{z+1}\int_0^{-1/z} x(z+1)\exp(-x(z+1))$$

$$= \frac{1}{e(z+1)}\cdot\left[-x\exp(-x(z+1))\Big|_0^{-1/z} + \int_0^{-1/z}\exp(-x(z+1))dx\right]$$

$$= \frac{1}{e(z+1)}\cdot\left[\frac{1}{z}\exp((z+1)/z) - \frac{1}{z+1}\exp(-x(z+1))\Big|_0^{-1/z}\right]$$

$$= \frac{1}{e(z+1)}\cdot\left[\frac{1}{z}\exp((z+1)/z) - \frac{1}{z+1}\exp((z+1)/z) + \frac{1}{z+1}\right].$$

Take the limit as $z \downarrow 0$ and $z \uparrow 0$ to get in both cases $f_Z(0) = e^{-1}$.

The conditional density of X given $Z = 0$ is then

$$f_{X|Z}(x \mid z = 0) = \frac{f_{XZ}(x,z)}{f_Z(z)}\Big|_{z=0} = x\exp(-x),$$

for $x > 0$. The paradox is that $Z = 0$ implies, and is implied by, $Y = 1$, yet the conditional density of X given $Z = 0$ differs from that of X given $Y = 1$.

The above results are due to the way the limit is taken. In one case we take the limit for $1 < Y < 1+\delta$. In the other case we take limit $0 < Z < \delta$. The latter is the same as the limit $1 < Y < 1+\delta x$ which is clearly different from the first limit. For fixed δ the latter conditioning set obviously includes more large values of X.

1.7 CONVERGENCE AND LIMIT LAWS

We have already seen that the mean and variance of two independent random variables X and Y:

$$E[X+Y] = E[X] + E[Y],$$

and

$$V(X+Y) = V(X) + V(Y).$$

More generally, if $X_1,\ldots,X_N$ are N independent random variables, we have mean

$$E\left[\sum_{i=1}^{N} X_i\right] = \sum_{i=1}^{N} E[X_i],$$

and variance

$$V\left(\sum_{i=1}^{N} X_i\right) = \sum_{i=1}^{N} V(X_i).$$

Note that we only use the independence for the variance. Without independence the mean is unchanged, but the variance becomks

$$V\left(\sum_{i=1}^{N} X_i\right) = E\left[\left(\sum_{i=1}^{N} X_i - E\left[\sum_{i=1}^{N} X_i\right]\right)^2\right]$$

$$=$$

$$E\left[\left(\sum_{i=1}^{N} X_i - E[X_i]\right)^2\right] = E\left[\sum_{i=1}^{N}\sum_{j=1}^{N}\left(X_i - E[X_i]\right)\cdot\left(X_j - E[X_j]\right)\right]$$

$$= \sum_{i=1}^{N}\sum_{j=1}^{N} E\left[\left(X_i - E[X_i]\right)\cdot\left(X_j - E[X_j]\right)\right]$$

$$= \sum_{i=1}^{N} V(X_i) + \sum_{i=1}^{N}\sum_{j=1}^{N} C(X_i, X_j).$$

Independence insures that all the covariance terms in the double sum are equal to zero.

Supposing all random variables to be independent and have the same mean μ, variance σ^2. Then the sum has mean

$$E\left[\sum_{i=1}^{N} X_i\right] = N\cdot\mu,$$

and variance

$$V\left[\sum_{i=1}^{N} X_i\right] = N\cdot\sigma^2$$

The mean and variance of the average are

$$E\left[\sum_{i=1}^{N} X_i/N\right] = \mu,$$

and

$$V\left[\sum_{i=1}^{N} X_i/N\right] = N\cdot\sigma^2/N^2 = \sigma^2/N.$$

Examine the behavior of the average of the first N random variables, $\overline{X}_N = \sum_{i=1}^{N} X_i/N$ as N gets large. In that Ease the variance of the sample average gets smaller and smaller. Using Chebyshev's inequality, this implies that the probability that the sample average is more than ε away from μ can be made arbitrarily small by taking N large enough, if we fix ε, then

$$Pr\left(\left|\bar{X}_N - \mu\right| > \varepsilon\right) < \sigma^2/\left(N \cdot \varepsilon^2\right),$$

which can be made arbitrarily small for fixed ε by choosing N large enough. It would seem uncontroversial to say that in this case $\bar{X}_N$ *converges* to p.

In other 'cases this result is not so clear. For example, consider the following sequence of random variables $X_1, X_2, \ldots$ with the *pdf* of X_n equal to

$$f_n(x) = \begin{cases} (n-1)/2 & -1/n, x < 1/n, \\ 1/n & n < x < n+1, \\ 0 & elsewhere \end{cases}$$

The mean of x_n is $1 + 1/2n$. The variance increases with n and approaches infinity as n goes to infinity. However, the probability that X_n is more than ε away from zero is at most $1/n$. Thus, X_n does not converge to its asymptotic mean of 1.

This example demonstrates the need for different concepts of convergence. We consider three such concepts.

Definition 1. *A sequence of random variables X_n converges to μ* in probability *if for all $\varepsilon > 0$,*

$$\lim_{n\to\infty} Pr\left(\left|x_n - \mu\right| > \varepsilon\right) = 0.$$

Definition 2. A *sequence of random variables* X_n *converges to μ in quadratic mean if*

$$\lim_{n\to\infty} E\left[\left(x_n - \mu\right)^2\right] = 0.$$

Definition 3. *A sequence of random variables X_n converges to μ almost surely if for all $\varepsilon > 0$,*

$$P\left(\lim_{n\to\infty}\left|X_n - \mu\right| > \varepsilon\right) = 0.$$

Note that in the first example the convergence is clearly in quadratic mean and probability. The independence implies it is also convergence almost surely. The following relations hold between the different convergence concepts:

(i) Convergence in quadratic mean implies convergence in probability.

(ii) Convergence almost surely implies convergence in probability.

(iii) Convergence in quadratic mean does not imply, and is not implied by, convergence almost surely.

Difference between convergence in quadratic mean and convergence almost surely.

Consider the following sequence of random variables, defined as $X_n(\omega)$, for $\omega \in \Omega = [0,1]$, and with the probability of ω in some interval a,b) equal to $b - a$ for $0 \le a \le b \le 1$:

$X_1(\omega)=1$ for $\omega\in[0,1]$ and zero elsewhere,

$X_2(\omega)=1$ for $\omega\in[0,1/2]$ and zero elsewhere,

$X_3(\omega)=1$ for $\omega\in[1/2,1/2+1/3]$ and zero elsewhere,

$X_4(\omega)=1$ for $\omega\in[1/2+1/3,1]\cup[0,1/12]$ and zero elsewhere,

$X_4(\omega)=1$ for $\omega\in[1/2+1/3,1]\cup[1/12,1/12+1/5]$ and zero elsewhere,

For $X_n(\omega)$ the intervals where $X_n(\omega)$ is equal to one have length $1/n$. That means probability is $1/n$. They shift to the right till they hit 1, and then start over again at 0. Then the mean is $1/n$, and the variance is $1/n-1/n^2$. Both of these go to zero, so that we have convergence to zero in quadratic mean and probability. Consider a particular value of ω in the sequence of values $X_1(\omega), X_2(\omega)$. This sequence does not converge, No matter how large n, the sum $\sum_{i=n}^{\infty}1/i$ diverges, implying that the sequence is always going to return to 1. Hence the probability of an ω such that the limit $X_n(\omega)$ even exists, let alone that it equals zero, is zero, and not one as required by almost sure convergence.

With these convergence concepts we can formulate **laws** of **large numbers.**

Result *1. Let $X_1, X_2, \ldots$ be a sequence of independent and identically distributed random variables with common mean μ and variance σ^2. Let $\bar{X}_N=\sum_{i=1}^{N}X_i/N$ be the average up to the N^{th} random variable. Then*

$$\bar{X}_N \xrightarrow{qm} \mu.$$

The result implies convergence in probability as well, as we already showed using Chebyshev's inequality.

A second result produces an even stronger for almost sure convergence, That is; convergence in quadratic mean does not necessarily hold because the variance does not necessarily exist,.

Result *2. Let $X_1, X_2, \ldots$ be a sequence of independent and identically distributed random variables with common mean μ. Let $\bar{X}_N=\sum\frac{X_i}{N}$ be the average up to the N^{th} random variable. Then*

$$\bar{X}_N \xrightarrow{as} \mu.$$

Definition *4. A sequence of random variables $X_1, X_2, \ldots$ converges in distribution to a random variable Y if at all continuity points of $F_Y(y)$,*

$$\lim_{n\to\infty}F_{X_n}(y)=F_Y(y).$$

The restriction to continuity points is for the following reason:

Suppose

$$f_{X_n}(x)=n, \text{ for } 0<x<1/n,$$

and zero elsewhere. The value of $F_{X_n}(0)$ is zero for all n, but X_n converges in distribution to a random variable Y with $F_Y(0) = 1 \neq \lim_{n\to\infty} F_{X_n}(0)$.

If a random variable converges in'distribution to a (degenerate) random variable with all mass in a single point μ, the random variable converges to μ in probability, and vice versa.

Central Limit Theorems:

Result 3. *Let $X_1, X_2, \ldots$ be a sequence of independent and identically distributed random variables with moment generating function $M_X(t)$ in a neighborhood of zero, and with mean μ and variance σ^2. Then*

$$\frac{1}{\sqrt{N}}\sum_{i=1}^{N}(X_i - \mu)/\sigma = \sqrt{N}\cdot(\bar{X}_N - \mu)/\sigma \xrightarrow{d} \mathcal{N}(0,1).$$

That is, the normalized sum converges to a standard normal distribution.

Slutsky's Theorem:

Result 4. *If X, converges in distribution to X , and X,, converges in probability to a constant c , then*

$$X_n + Y_n \xrightarrow{d} c + X.$$

and

$$X_n/Y_n \xrightarrow{d} c + X.$$

If in addition $c \neq 0$

$$X_n/Y_n \xrightarrow{d} X/c.$$

The last result is known as the delta method:

Result 5. *If a sequence of random variables X_n satisfies*

$$\sqrt{N}(X_n - \mu) \xrightarrow{d} \mathcal{N}(0, \sigma^2),$$

then for any function $g(\cdot)$ continuously differentiable in a neighborhood of μ with derivative $g'(\mu)$,

$$\sqrt{N}(g(X_n) - g(\mu)) \xrightarrow{d} \mathcal{N}(0, g'(\mu)^2 \cdot \sigma^2).$$

Check Your Progress **3**

1) Write the meaning:
   i) cumulative generating function; ii) characteristic function

   ....

   ....

   ....

   ....

   ....

2) Explain the meaning of Chebyshev's inequality.

....

....

....

....

....

3) Which distributions come under the family of exponential distribution?

....

....

....

....

....

4) Write the meaning of independent bivariate random variable.

....

....

....

....

....

5) Write themeaning of random variables X_n conversing to be almost surely.

....

....

....

....

....

1.8 LET US SUM UP

This unit reviews the important concepts of probability which find application in actuarial problems . The basic set operations, union and intersections, are given that help describe possible outcomes in model building where component of uncertainty forms a part. The ideas like universe, empty set and set of events that are required for probability space have been documented. The concepts like conditional probability and independence have also been included and their use in the context of random events, discrete and continuous, have been highlighted. The meaning of uniform, exponential and Gaussian distributions have been used to explain the transformation of distribution functions. Use of Poisson distribution to count events has been pointed out. Expectation of a random variable formulated for discrete and continuous cases has been examined. Use of *pmf* and *pdf* has been considered. Some special distributions viz., Gamma Chi-squared, normal and Cauchy, along with the family of exponential distribution have been included.

The joint and conditional distributions of independent random variables and the convergence and limit laws constitute the last part of discussion in the unit.

1.9 KEY WORDS

Borels Paradox: A conditional probability distribution that runs contrary to intuition: that conditional probability density functions are not invariant under coordinate transformations.

Characteristic Function: Function that takes the value 1 for numbers in the set, and value 0 for numbers not in the set.

Chebyshevs Inequality: A probability distribution where all the values are close to the mean value

Conditional Probability: The probability of some event *A*, given the occurrence of some other event *B*.

Cumulative Distribution Function (cdf): The function that completely describes the probability distribution of a real-valued random variable.

Delta Method: A method for deriving an approximate probability distribution for a function of a statistical estimator from knowledge of the limiting distribution of that estimator.

Discrete Uniform Distribution: Eelements of a finite set that are equally likely.

Exponential Distributions: A class of continuous probability distribution often used to model the time between independent events that happen at a constant average rate.

Indicator Function: A function defined on a set X that indicates membership of an element in a subset *A* of *X*.

Joint probability: The probability of both events together.

Marginal probability: One event, regardless of the other event.

Moment-Generating Function: A function used to generate the moments of the probability distribution..

Poisson Distribution: A discrete probability distribution which is the limiting form of the binomial distribution, provided 1) the number of trials is very large in fact tending to infinity; 2) the probability of success in each trial is very small; tending to zero.

Probability Density Function: A continuous random variable having a probability distribution f(x) which is a continuous non-negative function that gives the probability that the random variable will lie in a specified interval when integrated over the interval, i.e., $P(c \le x \le d) = \int_c^d f(x)$. The function f(x) is called the probability density function (p.d.f.) provided it satisfies following two conditions (i) $f(x) \ge 0$; (ii)the range of the continuous random variable **is** (a, b) $\int_a^b f(x) = 1$.

Probability Mass' Function: Often. the probability of the discrete random variable X assuming the value x is represented by f (x), f(x) = P (X = x) = probability that X assumes value x. The function f(x) is known as the probability mass function (p.m.f.) given that it satisfies following two conditions: 1) $f(x) \geq 0$; 2) $\sum_{i=1}^{n} f(x_i) = 1$.

1.10 SOME USEFUL BOOKS

Bialas, Wayne F. (2005), Lecture Notes in Applied Probability, Department of Industrial Engineering, University at Buffalo (see internet)

Freund J.E. (2001), Mathematical Statistics, Prentice Hall of India

Hoel, P (1962), Introduction to Mathematical Statistics, Wiley John & Sons, New York

Imbens, Guido (2000), Probability and Statistics (Lecture Notes), UCLA (see internet)

Olkin, I., L.J. Gleser, and C. Derrnan (1980), Probability Models and Applications, Macmillan Publishing, New York

1.11 ANSWER OR HINTS TO CHECK YOUR PROGRESS

Check Your Progress 1

1) See Section 1.2 and answer.
2) See Section 1.2 and answer.
3) See Section 1.2 and answer.
4) Read Section 1.3 and answer.

Check Your Progress 2

1) See Section 1.4 and answer.
2) See Section 1.4 and answer.
3) See Section 1.4 and answer.
4) See Section 1.4 and answer.
5) See Section 1.4 and answer.

Check Your Progress 3

1) See Section 1.5 and answer.
2) See Section 1.5 and answer.
3) See Section 1.6 and answer.
4) See Section 1.6 and answer.
5) See Section 1.7 and answer.

1.12 EXERCISES

1) Six fair dice are rolled one time. What is the probability that each face appears?

2) If two far dice are tosses, what is the smallest number of throws for which the probability of getting at least one double six exceeds 0.5.

3) Two dice are thrown and three events are defined as follows: A means "odd face with first die"; B means "odd face with second die"; and C means "odd sum" (one face is odd, one face is even). If each of the *36* sample points has probability 1/36, then

   a) Show that the events are pair wise independent.

   b) Show that the events are not jointly independent.

4) Bow I contains 3 red chips and **7** blue chips. Bow! II contains *6* red chips and 4 blue chips. A bowl is selected at random and then I chip is drawn from this bowl. What is the probability that the chip drawn is red? Given that this chip is red, what is the conditional probability that it came from bowl II?

5) Bowl I contains 6 red chips and 4 blue chips. Fives of these 10 chips are selected at random and without replacement and put in bowl II, which was originally empty. One chip is then drawn at random from bowl II. Find the conditional probability that 2 red chips and 3 blue chips were transferred from bowl I to bowl II given that a blue chip is drawn from bowl II.

6) Suppose that the random variable X has an exponential distribution with *pdf* $f_X(x)=\exp(-x), x>0$ and 0 elsewhere.

   a) Find the pdf for $1/X$

   b) Find the pdf for $\ln(X)$

   c) Find the *pdf* for $Y=1-F_X(X)$.

7) Let $f(x)=1/3$ for $-1<x<2$ and zero elsewhere be the ***pdf*** for a random variable X. Find the *pdf* and distribution function for the random variable $Y=X^2$.

8) Suppose that the random variable X has *cdf*

$$F_X(x)=\begin{cases}0 & \text{if} & x<0\\ 1/2 & \text{if} & x=0,\\ (x+1)/2 & \text{if} & 0<x\le 1,\\ 1 & \text{if} & x>1.\end{cases}$$

   Is X a discrete random variable? Is X a continuous random variable: Calculate the mean and variance of X.

9) Let X be a random variable with moment generating function $M_X(t), -h < t < h$. Prove that

$$P(X \geq a) \leq \exp(-at) \cdot M_X(t), 0 < t < h,$$

and

$$P(X \leq a) \leq \exp(-at) \cdot M_X(t), -h < t < 0.$$

Hint: Use Chebyshev's inequality.

10) Let X be a discrete random variable with $P(X = -1) = 118$, $P(X = 0) = 6/8$ and $P(X = 1) = 118$, and $P(X = c) = 0$ for all other values of c. Calculate the bound on $P(|X - \mu_X| \geq k \cdot \sigma_X)$ for $k = 2$ using Chebyshev's inequality.

UNIT 2 APPLIED PROBABILITY II

Structure

2.0 OBJECTIVES

After going through this unit, you will be able to:

- understand the degree of accuracy offered by an expected value of the mean value:
- compute the lower and upper bounds in which the expected value varies; and
- analyse problems with one-dimension random walk.

2.1 INTRODUCTION

In this unit we initiate preparatory wark for understanding the basic tools used in stochastic process while applying to finance and insurance. Basically themes covered in the theory of probability have been extended to a state where information availability varies over time. Our main concern is to highlight the results obtained in simple random walks, The following study material is adopted from the lecture notes of Prof. Albert Mayer and Dr. Radhika Nagpal (2002) borrowing freely the theorems, definitions and examples without mentioning the source. Our reliance on the above mentioned work is heavy, indeed. Credit goes to the authors for presenting the theme for a non-technical learner. We have included the select themes of the notes keeping in view the overall requirement of the course. Perhaps it will be useful to go back to the lecture notes for better appreciation of the themes dealt with.

2.2 WORKING WITH MEAN DEVIATION

We use' to take expectation of a random variable because it indicates the "average value" the random variable will take.

It means, the average value from repeated experiments is likely to be close to the expected value of one experiment. In addition it is more likely to be closer as the number of experiments increases. This result was first formulated and proved by Bernoulli.

Let us perform an experiment n times. We have $G_1,...,G_n$ independent random variables with the same distribution and the same expectation μ. If *A,* be the average of the results, that is,

$$A_n = \frac{\sum_{i=1}^{n} G_i}{n}$$

then by letting n grow large enough, we will be as certain as **we** want that the average will be close to μ. In other words, given any positive tolerance, ε, *A,,* we will be within the tolerance of μ as n grows. Effectively, we are looking for the limit

$$\lim_{n\to\infty} \Pr\{|A_n - \mu| < \varepsilon\}$$

According to Bernoulli, this limit equals one. This result is called the Weak Law of Large Numbers. **Theorem** 2.1 **(Weak Law of Large Numbers).** ***Let $G_1,...,G_n,...$ be independent variables with the same distribution and the some expectation, μ. For any*** $\varepsilon > 0$,

$$\lim_{n\to\infty} \Pr\left\{\left|\frac{\sum_{i=1}^{n} G_i}{n} - \mu\right| \le \varepsilon\right\} = 1.$$

See, however, that the law, does not say anything about the ***rate*** at which the limit is approached. That is, how big must n be to be within a given tolerance of **the** expected value with a specific desired probability?

. In the following we will develop three basic results concerning the above problem:

i) **Markov's Theorem:** It gives upper bound on the probability that the value of the random variable is mare than a certain multiple of its mean. Markov's result holds if we know nothing more than the value of the mean of a random'variable. As such, it is very general, but also is much weaker than results which take more information about the random variable into account.

ii) **Chebyshev's Theorem:** When we not only know the mean, but also another numerical quantity called the variance of the random variable, we consider this theorem as it combines Markov's Theorem and information about the variance to give more refined bounds.

iii) **Pairwise Independent Sampling Theorern:** It provides the adclitional information about rate of convergence we need to calculate numerical answers to questions such as those above. The Sampling Theorem follows from Chebyshev'sTheorem and properties of the variance of a sum of independent variables. Finally, the Weak Law of Large Numbers will be seen as a corollary of the Pairwise Independent Sampling Theorem.

Marltov's Theorem

When we want to consider the problem of bounding the probability because the value of a random variable is far away from the mean. Markov's theorem gives a very rough estimate, based only on the value of tlie mean.

Theorem 2.2 (Markov's Theorem). *If R* **is** *a nonnegative random variable, then for all* $x > 0$

$$\Pr\{R \geq x\} \leq \frac{E[R]}{x}$$

Markov's Theorem is often expressed in an alternative form; we state it as a corollary.

Corollary 2.1. *If R is a nonnegative random variable, then for all* $c = 0$

$$\Pr\{R \geq c \cdot E[R]\} \leq \frac{1}{c}$$

Examples of Markov's Theorem

Mayer and Nagpal (2002) give the following two examples.

Suppose that N men attending a dinner party deposit their hats in the counter. **At** the end of the night, each is given back his own hat with probability $1/N$. What is the probability that x or more men get the right hat?

First, we derive an upper bound with Marltov's Theorem. Let the random variable, R, be the number of men that get tlie right hat. By linearity of expectation $E[R] = 1$. Using Msrkov's Theorem, the probability that x or more men get the right hat is:

$$\Pr\{R \geq x\} \leq \frac{E[R]}{x} = \frac{1}{x}.$$

Consider the Chinese Appetizer problem where N people are eating Chinese appetizers arranged on a circular, rotating tray. Someone then spins the tray so that each person receives a random appetizer. We need to find the probability that everyone gets the same appetizer as before.

See that are N equally likely orientations for the tray after it stops spinning. Therefore everyone is likely getting the right appetizer in just one of these N orientations. Therefore, probability is $1/N$.

Let us use the Markov theorem to get the probability. Let the random variable, R, be the number of people who get the right appetizer. We know that E[R] = 1. Applying Markov's Theorem, we find:

$$\Pr\{R \geq N\} \leq \frac{E[R]}{N} = \frac{1}{N}.$$

Markov's Theorem also gives the same $1/N$ bound for the probability everyone gets their hat in the hatcheck problem. But in reality, the probability of this event is $1/N!$. So Markov's Theorem in this case gives probability bounds that are way off.

Note that the proof of Markov's Theorem requires that the random variable, R, be nonnegative.

2.2.1 Deviation below the Mean

According Markov's Theorem, a random variable may not greatly exceed the mean. There is another theorem which points out that a random variable is not likely to be much smaller than its mean.

Theorem 2.2.1. *Let L he a real number and let R be a random variable such that $R \leq L$. For all $x < L$, we have:*

$$\Pr\{R \leq x\} \leq \frac{L - E[R]}{L - x}.$$

Proof. The event that $R \leq x$ is the same as the event that $L - R \geq L - x$. Therefore,

$$\Pr\{R \leq x\} = \Pr\{L - R \geq L - x\}$$

$$\leq \frac{E[L - R]}{L - x}. \qquad \text{(by Markov's Theorem)} \qquad (1)$$

Markov's Theorem can be applied since $L - R$ is a nonnegative random variable and $L - x > 0$.

Chebyshev's Theorem

We have seen above the Markov's Theorem for deviations above and below the mean. Suppose we want bounds that apply in both directions, that is, bounds on the probability that $|R - E[R]|$ is large.

It is difficult to use Markov's inequality directly to bound the probability that $|R - E[R]| \geq x$, as we have to compute $E[|R - E[R]|]$.

Since $|R|$ and hence $|R|^k$ are nonnegative variables For any R, Markov's inequality also applies to the event $[|R|^k \geq x^k]$.

Since such an event is equivalent to the event $[|R| \geq x]$, we have

Corollary 2.2.1. *For any random variable R, any positive integer k, **and** any $x > 0$,*

$$\Pr\{|R| \geq x\} \leq \frac{E\left[|R|^k\right]}{x^k}.$$

Take $k = 2$ to get a special case of the above corollary. Then we can apply it to bound the random variable, $R - E[R]$, that measures R's deviation from its mean. Thus,

$$\Pr\{|R - E[R]| \geq x\}$$

$$= \Pr\left\{(R - E[R])^2 \geq x^2\right\} \leq \frac{E\left[\left(R - E[R]^2\right)\right]}{x^k}. \qquad (2)$$

where the inequality (2) follows from Corollary 2.2..1 applied to the random variable, $R - E[R]$. That *is* to say, we can bound the probability that the random variable R deviates from its mean by more than x by an expression decreasing as $1/x^2$ multiplied by the constant

$$E\left[(R - E[R])^2\right].$$

Such a constant is variance of R.

Definition 2.2.1. *The variance. Var [R], of a random variable, R, is:*

$$Var[R] = E\left[(R - E[R])^2\right].$$

Accordingly we will rewrite (2) as Chebyshev theorem.

Theorem 2.2.2. **(Chebyshev)** *Let* **R** *be a random variable, and let* **x** *be a **positive** real number. Then*

$$\Pr\{|R - E[R]| \geq x\} \leq \frac{Var[R]}{x^2}.$$

Example: Gambling Games

When we compare the following two gambling games, the usefulness of variance can be easily judged.

Game *A* : We win \$2 with probability 2/3 and lose *\$1* with probability 1/3.

Game *B* : We win \$1002 with probability 2/3 and lose \$2001 with probability 1/3.

Both the games have the same probability, 2/3, of winning. However. We will see these to be very different, once we consider variance. Let random variables *A* and B be the payoffs for the two games. For example, *A* is 2 with probability 2/3 and –1 with probability 1/3. The expected payoff for each game is,

$$E[A] = 2\cdot\frac{2}{3} + (-1)\cdot\frac{1}{3} = 1,$$

$$E[B] = 1002\cdot\frac{2}{3} + (-2001)\cdot\frac{1}{3} = 1.$$

Thus, the expected payoff is the same for both games. Look at the variance however.

Let

$$A - E[A] = \begin{cases} 1 & \text{with} \\ -2 & \text{with probability*} \end{cases}$$

$$(A - E[A])^2 = \begin{cases} 1 & \text{with probability3} \\ 4 & \text{with probability } \frac{1}{3} \end{cases}$$

$$E\left[(A - E[A])^2\right] = 1 \cdot \frac{2}{3} + 4 \cdot \frac{1}{3}$$

$$Var[A] = 2.$$

Similarly, for $Var[B]$:

$$B - E[B] = \begin{cases} 1001 & \text{with probability?} \\ -2002 & \text{with probability } \frac{1}{3} \end{cases}$$

$$(B - E[B])^2 = \begin{cases} 1{,}002{,}001 & \text{with probability } \frac{2}{3} \\ 4{,}008{,}004 & \text{with probability } \frac{1}{3} \end{cases}$$

$$E\left[(B - E[B])^2\right] = 1{,}002{,}001 \cdot \frac{2}{3} + 4{,}008{,}004 \cdot \frac{1}{3}$$

$$Var[B] = 2{,}004{,}002.$$

We may note that the variance of Game A is 2 whereas the variance of Game B is more than two million. This means, the payoff in Game A remains closer to the expected value of \$1. On the contrary the payoff in Game B can deviate very far from this expected value.

In economics high variance is often associated with high risk. For example, in ten rounds of Game A, we expect to make \$10, but could conceivably lose \$10 instead. On the other hand, in ten rounds of game B, we also expect to make \$10, but could actually lose more than \$20,000.

Theorem 2.2.3 (Pairwise Independent Additivity of Variance).
If $R_1, R_2, \ldots, R_n$ *are pairwise independent random variables, then*

$$Var[R_1 + R_2 + \ldots + R_n] = Var[R_1] + Var[R_2] + \ldots + Var[R_n]$$

$$E\left[\left(\sum_{i=1}^{n} R_i\right)^2\right] = E\left[\sum_{i=1}^{n}\sum_{j=1}^{n} R_i R_j\right]$$

$$= \sum_{i=1}^{n}\sum_{j=1}^{n} E[R_i R_j]$$

$$= \sum_{1 \le i \ne j \le n} E[R_i] E[R_j] + \sum_{i=1}^{n} E[R_i^2] \quad \text{(pairwise independence)} \quad (3)$$

In (3), we use the fact that the expectation of the product of two independent variables is the product of their expectations.

Moreover,

$$E^2\left[\sum_{i=1}^{n} R_i\right] = \left(E\left[\sum_{i=1}^{n} R_i\right]\right)^2$$

$$= \left(\sum_{i=1}^{n} E[R_i]\right)^2$$

$$= \sum_{i=1}^{n}\sum_{j=1}^{n} E[R_i]E[R_j]$$

$$= \sum_{1\le i\ne j\le n} E[R_i]E[R_j] + \sum_{i=1}^{n} E^2[R_i] \tag{4}$$

Therefore we get

$$Var\left[\left(\sum_{i=1}^{n} R_i\right)\right] = E\left[\left(\sum_{i=1}^{n} R_i\right)^2\right] - E^2\left[\sum_{i=1}^{n} R_i\right]$$

$$= \sum_{1\le i\ne j\le n} E[R_i]E[R_j] + \sum_{i=1}^{n} E[R_i^2] -$$

$$\left(\sum_{1\le i\ne j\le n} E[R_i]E[R_j] + \sum_{i=1}^{n} E^2[R_i]\right)$$

$$= \sum_{i=1}^{n} E[R_i^2] - \sum_{i=1}^{n} E^2[R_i]$$

$$= \sum_{i=1}^{n}\left(E[R_i^2] - E^2[R_i]\right)$$

$$= \sum_{i=1}^{n} Var[R_i]$$

2.2.2 Variance of a Binomial Distribution

Take a random variable R. It has a binomial distribution, if

$$\Pr\{R = k\} = \binom{n}{k} p^k (1-p)^{n-k}$$

where n and p are parameters such that' $n \ge 1$ and $0 < p < 1$.

If we take R as the sum of n independent Bernoulli variables, then we can regard R as the number of heads that comes up when we toss n independent coins, (where each head comes up with probability p). Thus, we can write $R = R_1 + R_2 + ... + R_n$ where

$$R_i = \begin{cases} 1 & \text{with probabilityp,} \\ 0 & \text{with probability } 1-p. \end{cases}$$

Then variance of the binomially distributed variable R is given by

$$\begin{aligned} Var[R] &= Var[R_1]+Var[R_2]+...+Var[R_n] \\ &= n\ Var[R_1] \\ &= n\left(E\left[R_1^2\right]-E^2[R_1]\right) \\ &= n\left(E[R_1]-E^2[R_1]\right) \\ &= n\left(p-p^2\right). \end{aligned}$$

'From the above result we see that tlie binomial distribution has variance $p(1-p)n$ and standard deviation $\sqrt{p(1-p)n}$. It we consider a special case, an unbiased binomial distribution with $(p=1/2)$, then the variance is $n/4$ and the standard deviation is $\sqrt{n}/2$.

2.3 APPLICATIONS OF CHEBYSHEV'S THEOREM

Let us reformulate Chebyshev's Theorem in terms of standard deviation.

Corollary 5.1. *Let* R *be a random variable, and let* c *he a positive real number.*

$$\Pr\left\{\left|R-E[R]\right|\geq c\sigma_R\right\}\leq\frac{1}{c^2}$$

That is, the "likely" values of R are clustered in an $O(\sigma_R)$-sized region around $E[R]$. This result. therefore, confirms that the standard deviation measures spread out the distribution of R is around its mean.

A One-Sided Bound

Chebyshev's Theorem gives a "two-sided bound". That is we get the bounds with tlie probability that a random variable deviates *above* or ***below*** the mean by some amount. We develop one-sided bound for it below. For example, let us find the probability or a random variable that deviates above the mean.

2.3.1 Deviation of Repeated Trials

Using Chebyshev's Theorem we will show how the average of many trials approaches the mean in a large number of repeated trials.

Estimation from Repeated Trials

Let's take $\boldsymbol{p}$ as proportion of success of variable, a Bernoulli variable. G, with $G=1$ for success and $G=0$ otherwise. In this case, $G=G^2$, so

$$E\left[G^2\right]=E[G]=\Pr\{G=1\}=p$$

and

$$Var[G]=E\left[G^2\right]-E^2[G]=p-p^2=p(1-p)$$

To estimate p, we take a large number of trials, n, and record the fraction of success. Thus, we are taking Bernoulli variables $G_1, G_2, \ldots, G_n$ which are independent each with the same expectation as G. Computing their sum

$$S_n = \sum_{i=1} G_i \, . \tag{5}$$

and then using the average, S_n/n, as our estimate of p.

Speaking more generally, we may take any set of random variables $G_1, G_2, \ldots, G_n$, with the same mean, μ, and use the average, S_n/n, to estimate μ. A critical property of S_n/n is that it has the same expectation as the G_i' s, namely,

$$E\left[\frac{S_n}{n}\right] = \mu \tag{6}$$

Note that the random variables G_i need not be either Bernoulli or independent for (6) to hold.

To derive its variance, suppose that the G_i' s also have the same deviation, σ. In that case, we get the second critical property of S_n/n, that is,

$$Var\left[\frac{S_n}{n}\right] = \sigma_n^2 \tag{7}$$

Thus, the variance af S_n is the sum of the variances of the G_i' s. For example, by Theorem 2.3.1, the variances can be summed if the G_i' s are pairwise independent. Then we derive,

$$\begin{aligned} Var\left[\frac{S_n}{n}\right] &= \frac{1}{n^2} Var[S_n] \\ &= \frac{1}{n^2} Var\left[\sum_{i=1}^{n} G_i\right] \\ &= \frac{1}{n^2} \sum_{i=1}^{n} Var[G_i] \\ &= \frac{1}{n^2} \cdot n\sigma^2 = \frac{\sigma^2}{n}. \end{aligned}$$

With this result, we apply Chebyshev 's Bound and conclude as follows:

Theorem 2.3.1 [Pairwise Independent Sampling]. Let

$$S_n = \sum_{i=1}^{n} G_i$$

where $G_1, \ldots, G_n$ are pairwise independent variables with the same mean, μ, and deviation, σ. Then

$$\Pr\left\{\left|\frac{S_n}{n} - \mu\right| \geq x\right\} \leq \frac{1}{n}\left(\frac{\sigma}{x}\right)^2. \tag{8}$$

Note that Theorem 2.3.1 provides a precise statement about the average of independent samples of a random variable approaching the mean. We can generalise it to many cases when S_n is the sum of independent variables whose mean and deviation are not necessarily all the same.

2.3.2 Proof of the Weak Law

We can state the conclusion of the Weak Law of Large Numbers, Theorem 2.1, is that the probability average differing from the expectation by more than any given tolerance approaches zero.

Theorem 2.3.2 [Weak Law of Large Numbers]. Let

$$S_n = \sum_{i=1} G_i ,$$

where $G_1, \ldots, G_n, \ldots$ are pairwise independent variables with the same expectation, μ and standard deviation, a . For any $\varepsilon > 0$,

$$\lim_{n\to\infty} \Pr\left\{\left|\frac{S_n}{n} - \mu\right| \geq \varepsilon\right\} = 0$$

Proof. Choose x in Theorem 2.3.1 to be ε. Then, given any $\delta > 0$, choose n large enough to make $(\sigma / x^2)/n < 6$. By Theorem 2.3.1,

$$\Pr\left\{\left|\frac{S_n}{n} - \mu\right| \geq \varepsilon\right\} < \delta$$

So at the limit probability must equal zero.

Notice that this version of the Weak Law is slightly different from the version we first stated in Theorem 2.1. Whereas Theorem 2.3.2 requires that the G_i' s lhave finite variance, Theorem 2.1 requires finite expectation only. Again, the original version of Theorem 2.1 requires mutual independence, while Theorem 2.3.2 requires only pairwise independence.

Note that both the Weak Law and Pairwise Independence Sampling Theorem 2.3.1 do not provide any information about the way the average value of the observations may be expected to *oscillate* in the course of repeated experiments. We will see such a feature in *Strong* Law of *Large Numbers* which deals with the oscillations. Such oscillations are critical in gambling situations, where large oscillations can bankrupt a player, even though the player's average winnings are assured in the long run.

Check Your Progress 1

1) What is the result of Weak Law of Large Number?

....

....

....

....

....

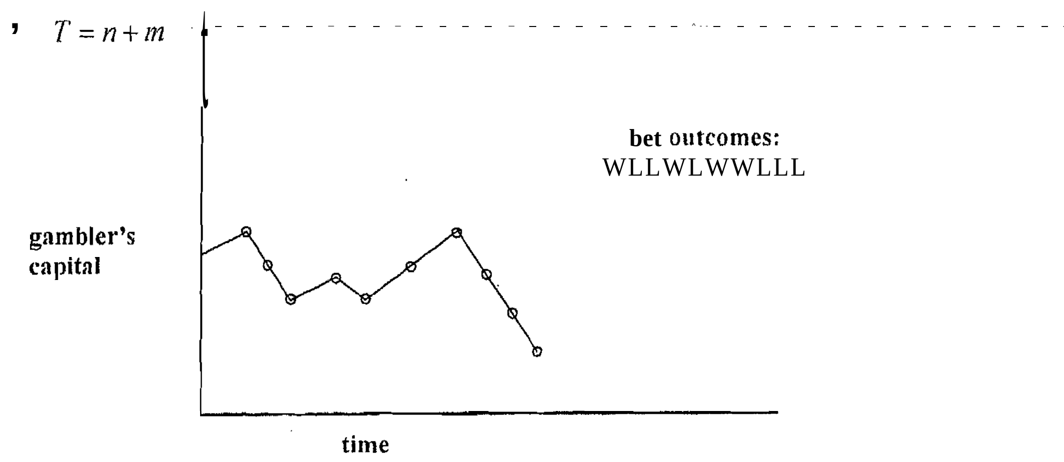
2) Write the meaning of
   (i) Markov's theorem and (ii) Chebyshev's theorem.

3) How do you find two-sided bound given by Chybyshev's theorem?

4) Consider n independent Bernoulli trials with Probability of success p_s. Find the probability of no failures.

2.4 RANDOM WALKS AND GAMBLER'S RUIN

Random walks model processes taking steps in a randomly chosen sequence of directions, For example, in Computer Science, the Google search engine uses random walks through the graph of world-wide web links and determines the relative importance of websites, In case of finance, there are models which consider one-dimensional random walks in the day-to-day fluctuations of market prices. In the following we consider 1-dimensional random walks, i.e., walks along a straight line.



Source: Meyer and Nagpal (2002)

Figure 1: This is a graph of the gambler's capital versus time for one possible sequence of bet outcornes. At each time step, the graph goes up with probability p and down with probability $1-p$. The gambler continues betting until the graph reaches either 0 or $T = n + m$.

Suppose that a gambler starts with n dollars – his initial money (capital). He makes a sequence of \$1 bets. If he wins an individual bet, he gets his money back plus another \$1. If he loses, he loses the \$1. In each bet, he wins with probability $p > 0$ and loses with probability $q = 1 - p > 0$. The gambler plays until either he is bankrupt or increases his capital to a goal amount of. T dollars, If he reaches his goal, then he is called an overall *winner*, and his profit will be $m = T - n$ dollars. If his capital reaches zero dollars before reaching his goal, then we say that he is "ruined".

Figure 1 shows the gambler's position as he proceeds with his \$1 bets. The random walk has boundaries at 0 and T. If the random walk ever reaches either of these boundary values, then it terminates. With these constraints, our objective is to determine the probability, w, that the walk terminates at boundary T. That is, the probability that the gambler is a winner.

It is easy to see that $p = q = 1/2$ in a fair game. In that case random walk is called *unbiased*. The gambler is more likely to win if $p > 1/2$ and less likely to win if $p < 1/2$; these case's give the random walks which are *biased*.

Example 1. Suppose that the gambler is flipping a coin, winning \$1 on Heads and losing \$1 on Tails, Also, the gambler's starting capital is $n = 500$ dollars, arid he wants to make $m = 100$ dollars. That is, lie plays until he goes broke or reaches a goal of $T = n + m = \$600$. What is the probability that he is a winner? We will show that in this case the probability $w = 5/6$. So his chances of winning are really very good, namely, 5 chances out of 6,

Let us suppose that the gambler chooses to play roulette in an American casino, always betting \$1 on red. A roulette wheel has 18 black numbers, 18 red numbers, and 2 green numbers. In this game, the probability of winning a single bet is $p = 18/38 \approx 0.47$. There are two green numbers, which introduce bias to the bets and give the casino an edge. However, the bets are almost fair, and we might expect that the gambler has a reasonable chance of reaching his goal — the 516 probability of winning in the unbiased game surely gets reduced, but perhaps not too drastically.

But our guess is not correct. His odds of winning against the "slightly" unfair roulette wheel are less than 1 in 37,000. No matter how much money the gambler has to start, e.g., \$5000, \$50,000, or $\$5 \cdot 10^{12}$, his odds are still less than 1 in 37,000 of winning a mere 100 dollars. Perhaps it will be better to suggest for not playing the game,

The theory of random walks draws a lot of such counter-intuitive conclusions.

2.4.1 The Probability Space

A game like the one described above follows in Figure 1. It starts at the point $(n, 0)$. A winning path does not touch the x axis: In fact it ends when it first touches the line $y = T$.

In a similar logic, a losing path will not touch the line $y = T$ and will end when it first touches the x axis.

We can take the length k path, which is characterised by the history of wins and losses on individual \$1 bets. So we use a length k string of W's and L' s to model a path, assigning probability $p^r q^{k-r}$ to a string that contains r W's. The outcomes in such a sample space are the string that shows the winning or losing walks.

In an infinite walks in which the gambler plays forever, neither reaches his goal nor goes bankrupt. In such a game the probability of playing forever is zero, so we don 't need to include any such outcomes in our sample space.

2.4.2 The Probability of Winning

The Unbiased Game

Consider the case of a fair coin, So the probability of winning $p = 1/2$. Let us determine the probability, w, that tlie gambler wins. If the random variable G is equal to the gambler's dollar gain, the $G = T - n$ if the gambler wins, and $G = -n$ if lie loses, Thus, the expected value of win is

$$E[G] = w(T - n) - (1 - w)n = wT - n$$

Since the gambler wins or loses, the probability of losing is $1 - w$.

Let G_i be the amount the gambler gains on the ith flip: So, $G_i = 1$ if the gambler wins the flip, $G_i = -1$ if the gambler loses the flip. and $G_i = 0$ if the game has ended before the ith flip. **As** the coin is fair, $E[G_i] = 0$.

Thus, $wT = n = E(G) = \sum_{i=1}^{\infty} E(G_i) = 0$

or, $w = \dfrac{n}{T}$

Theorem 2.4.2. In the unbiased Gambler's Ruin game with probability $p = 1/2$ of winning each individual bet, with initial capital, n, and goal, T,

$$\Pr\{\text{the gambler is a winner}\} = \frac{n}{T} \tag{9}$$

Example 2. Suppose we have $100 and we start flipping a fair coin, betting $1 with the aim of winning $100. Then the probability of reaching the $200 goal is 100 /200 =1 /2 — the same as the probability of going bankrupt. In general, if $T = 2n$, then the probability of doubling your money or losing all your mcney is the same.

Example 3. Suppose we have $500 and we start flipping a fair coin, betting $1 with the aim of winning. $1 00.

So n =500, $T = 600$, and $\Pr\{\text{win}\} = 500/600 = 5/6$.

Example 4. Suppose John starts with $100, and Sita starts with $10. They flip a fair coin, and every time a head appears, John wins $1 from Sita, and vice versa for tails. They play this game until one person goes bankrupt. What is the probability of John winning?

Frame this problem as is done in case of Gambler's Ruin problem with n=100 and T=100+10 =110. The probability of John winning is 1001110=10/11, namely, the ratio of his wealth to the combined wealth. Sita's chances of winning are 1/11.

It is easy to see that although John will win most of the time, the game is still fair. When John wins, he wins $10 only; when he loses, he loses big: $100. The expected win of both players is zero.

Another .result of the analysis is: the larger the gambler's initial stake, the larger the probability that he will win a fixed amount.

Example 5. If the gambler started with one million dollars instead of 500, but aimed to win the same 100 dollars as in the Example **3,** the probability of winning would increase to $1M/(1M + 100) > 0.9999$,

2.4.3 A Recurrence for the Probability of Winning

Let p and T be fixed, and let w_n be the gambler's probability of winning when his initial capital is n dollars. For example, w_0 is the probability that the gambler will win given that he starts off broke; clearly, $w_0 = 0$. Similarly, for $w_T = 1$.

Suppose that, the gambler starts with n dollars, where $0 < n < T$. Consider the outcome of his first bet. The gambler wins the first bet with probability p. In that case, he is left with $n+1$ dollars and becomes a winner with probability w_{n+1}. On the other hand, he loses the first bet with probability $1 - p$. Now he is left with $n - 1$ dollars and becomes a winner with probability w_{n-1}. Thus, he wins with probability $w_n = pw_{n+1} + qw_{n-1}$. Solving for w_{n+1} we have

$$w_{n+1} = \frac{w_n}{p} - w_{n-1}\frac{q}{p}. \tag{10}$$

Homogeneous Linear Recurrence

The definition of a quantity w_{n+1} in terms of a linear combination of values w_k for $k < n+1$ is called a **homogeneous linear recurrence.** These can be solved with the help of a general method. Suppose that the form of the solution is $w_n = c^n$ for some $c > 0$. It can be shown it is value of a constant c

From (10) we have

$$w_{n+1} - \frac{w_n}{p} + w_{n-1}\frac{q}{p} = 0 \qquad (11)$$

For the correct solution, this is equivalent to

$$c^{n+1} - \frac{c^n}{p} + c^{n-1}\frac{q}{p} = 0$$

Factoring out c^{n-1} we get

$$c^2 - \frac{c}{p} + \frac{q}{p} = 0.$$

From the above equation we get two roots, $(1-p)/p$ and 1. If

$$w_n = ((1-p)/p)^n = (q/p)^n,$$

then (11), and hence (10) will be satisfied. Taking $w_n = 1^n$ condition (11) can be satisfied, Since the left-hand side of (11) is zero using either definition, it follows that any definition of the form

$$w_n = A\left(\frac{q}{p}\right)^n + B \cdot 1^n$$

will also satisfy (11). For the values of w_0 and w_T, the boundary conditions we solve for A and B to gel

$$0 = w_0 = A + B,$$

$$1 = w_T A\left(\frac{q}{p}\right)^T + B,$$

Thus, $$A = \frac{1}{(q/p)^T - 1}, \qquad B = -A, \qquad (12)$$

and

$$w_n = \frac{(q/p)^n - 1}{(q/p)^T - 1}. \qquad (13)$$

The solution (13) only applies to biased walks, since we require $p \neq q$ so the denominator is not zero.

Theorem 2.4.3. In the biased Gambler's Ruin game with probability, $p \neq 1/2$, of winning each bet, with initial capital, n, and goal, T,

$$\Pr\{\text{the gambler is winner}) = \frac{(q\ p)^{n} - 1}{(q\ p)^{T} - 1} \qquad (14)$$

To appreciate the idea of the probability that the gambler wins in the biased game, suppose that $p < 112$; that is, the game is biased against the gambler. Then both the numerator and denominator in the quotient in (14) are positive and the quotient is less than one. So adding 1 to both the numerator and denominator increases the quotient, and the bound (14) simplifies to

$$(q/p)^{n} / (q/p)^{T} = (p/q)^{T-n}.$$

Corollary 10.7. In the Gambler's Ruin game biased against the gambler. that is, with probability $p < 1/2$ of winning each bet, with initial capital, n, and goal, T,

$$\Pr\{\text{the gambler is a winner}\} < \left(\frac{p}{q}\right)^{m}, \qquad (15)$$

where $m = T - n$,

the gambler's *intended profit*. Thus, the gambler gains his intended profit, m, before going broke with probability at most $(p/q)^{m}$. In this result, the upper bound does not depend on the gambler's starting capital. To see the consequences of this, we give the following examples:

Example 6. Suppose that the gambler starts with \$500 aiming to profit \$100, this time by making \$1 bets on red in roulette. By (15), the probability, w_n, that he is a winner is less than

$$\left(\frac{18/38}{20/38}\right)^{100} = \left(\frac{9}{10}\right)^{100} < \frac{1}{37,648}.$$

This is in sharp contrast to the unbiased game, where we saw (in Example 3) that his probability of winning was 516.

Example 7. Consider Example 3 and the capital to be \$1,000,000 to start in the unbiased game. The gamble was almost certain to win \$100. But betting against the "slightly" unfair roulette wheel, even starting with \$1,000,000, his chance of winning \$100 remains less than 1 in 37,648. He will almost surely lose all his \$1,000,000. In fact, because the bound (15) depends only on his intended profit, his chance of going up a mere \$100 is less than 1 in 37,648 no *matter **how** much money he starts with.*

The bound (15) is exponential in m. So, for example, doubling his intended profit will square his probability of winning.

Example 8. Suppose that gambler's stake goes up 200 dollars in the above example. What is probability that he wins before going broke in the roulette game?

We have $$(9/10)^{200} = \left((9/10)^{100}\right)^{2} = \left(\frac{1}{37,648}\right)^{2},$$

which is about 1 in '70 billion.

Example 9. Apply the formula (14) and find that the probability of winning \$1 0 before losing \$1 0

Here, $\dfrac{\left(\frac{20/38}{18/38}\right)^{10}-1}{\left(\frac{20/38}{18/38}\right)^{20}-1}=0.2585\ldots$

See that this is somewhat worse than the 1 in 2 chance in the fair game.

What is lesson of fair and unfair game's formulation?

In case of a fair game, it helps if we to have a large capital, whereas in the unfair case, it doesn't.

2.4.4 Intuition

It will be useful to see the intuition behind the gambler's failure to make money when the game is slightly biased against him. Seen in terms of roulette game, the gambler wins a dollar with probability 9/19 and loses a dollar with probability10119.

Therefore, his expected return on each bet is

$$9/19-10/19=-1/19\approx-0.053 \text{ dollars.}$$

That is, on each bet: his capital is expected to drift downward by a little over 5 cents.

If the gambler starts with a trillion dollars, then he will play for a very long time. So at some point there would be a lucky outcome. Perhaps an upward swing will put him \$100 ahead. However, there is a problem in this formulation. His capital drifts downward steadily. If the gambler does not have luck for an upward swing early, then he is doomed. The logic is simple: After his capital drifts downward a few hundred dollars, he needs a huge upward swing to save himself', Perhaps, such a huge swing is extremely unlikely.

To quantify these drifts and swings, consider the position after k rounds, Let the number of wins by the player be a binomial distribution with parameters $p<1/2$ and k. Then his expected win on any single bet is $p-q=2p-1$ dollars. Thus. his expected capital is $n-k(1-2p)$. The actual number of wins must exceed the expected number by $m+k(1-2p)$ to be a winner. Since the binomial distribution has a standard deviation of only $\sqrt{kp(1-p)}$, for the gambler to win, he needs his number of wins to deviate by

$$\frac{m+k(1-2p)}{\sqrt{kp(1-2p)}}=\Theta\left(\sqrt{k}\right)$$

times its standard deviation, which is ~ ~ ~ ~ l idire ltq tail position of binomial distribution.

In contrast, a fair game does not witness drift; swings are the only effect, Dire to this, our earlier result holds good. Therefore, if the gambler starts with a trillion dollars, then almost certainly there will be a lucky swing that puts him \$100 ahead.

2.4.5 Duration of a Walk

The discussion above pointed out that the gambler would be a winner in both fair and unfair games. Let us enquire, how many bets he needs on average to either win or go broke.

Biased Walk Case

Let Q be the number of bets the gambler makes until the game ends. Since the gambler's expected win on any bet is $2p-1$, Wald's Theorem (see Appendix given at the end) should tell us that his game winnings, G, will have expectation $E[Q](2p-1)$. Thus,

$$E[G]=(2p-1)E[Q] \tag{16}$$

In an unbiased game (16) is a trivial result as both $2p-1$ and the expected overall winnings, $E[G]$, are zero. In the unfair case, however, $2p-1\neq 0$.

Moreover,

$$E[G]=w_n(T-n)-(1-w_n)n=w_nT-n$$

So assuming (16), we say:

Theorem 2.4.5. In the biased Gambler's Ruin game with initial capital, n, goal, T, and probability, $p\neq 1/2$ of winning each bet,

$$E[\text{nuber of bets till game ends}]=\frac{\Pr\{\text{gambler is a winner}]T-n}{2p-1} \tag{17}$$

Since $G=\sum_{i=1}^{Q}G_i$ is not a sum of nonnegative variables we do not have a special case of Wald's theorem and if the gambler loses the ith bet, then random variable G_i equals 1. To confirm that (16) applies in general cases, take the random variable G_i+1 is nonnegative, and

$$E[G_i+1\mid Q\geq i]=E[G_i\mid Q\geq i]+1=2p,$$

so by Wald's Theorem

$$E\left[\sum_{i=1}^{Q}(G_i+1)\right]=2pE[Q] \tag{18}$$

But
$$\begin{aligned}E\left[\sum_{i=1}^{Q}(G_i+1)\right] &= E\left[\sum_{i=1}^{Q}G_i+\sum_{i=1}^{Q}1\right]\\ &= E\left[\left(\sum_{i=1}^{Q}G_i\right)+Q\right]\\ &= E\left[\sum_{i=1}^{Q}G_i\right]+E[Q]\\ &= E[G]+E[Q]\end{aligned} \tag{19}$$

Combining (18) and (19) we get the result implied by (16).

Example 10. If the gambler aims to profit \$100 playing roulette with n dollars to start, he can expect to make $((n+100)137,648-n)/(2(18/38)-1) \approx 19n$ bets before the game ends. So he can enjoy playing for a-good while before almost surely going broke.

Duration of an Unbiased Walk

Consider the expected number of bets as a function of the gambler's initial capital. That is, for fixed p and T, let e_n be the expected number of bets until the game ends when the gambler's initial capital is n dollars. Since the game is over in no steps if $n=0$ or T, the corresponding boundary conditions are $e_0 = e_T = 0$.

Alternatively, let the gambler start with n dollars, where $0 < n < T$. By the conditional expectation, the expected number of steps can be broken down. So that,

$$e_n = pE[Q \mid \text{gambler wins first bet}] + qE[Q \mid \text{gambler loses first bet}].$$

When the gambler wins the first bet, his capital is n+1, so he can expect to make another e_{n+1} bets. That is,

$$E[Q \mid \text{gambler wins first bet}] = 1 + e_{n+1}.$$

and similarly,

$$E[Q \mid \text{gambler loses first bet}] = 1 + e_{n-1}.$$

So we have

$$e_n = p(1+e_{n+1}) + q(1+e_{n-1}) = pe_{n+1} + qe_{n-1} + 1,$$

which yields the linear recurrence

$$e_{n+1} = \frac{e_n}{p} - \frac{q}{p}e_{n-1} - \frac{1}{p}.$$

For $p=q=1/2$, this equation simplifies to

$$e_{n+1} = 2e_n - 2e_n - e_{n-1} - 2. \tag{20}$$

There is a general theory for solving linear recurrences like (20) in which the value at $n+1$ is a linear combination of values at some arguments $k < n+1$ plus another simple term — in this case plus the constant — 2. This theory implies that

$$e_n = (T-n)n. \tag{21}$$

Equation (21) can be verified easily from the boundary condition along with (20). Using strong induction on n, we have,

Theorem 2.4.5.1. In the unbiased Gambler's Ruin game with initial capital, n, and goal, T, and probability,

$$E[\text{number of bets till game ends}] = n(T-n) \tag{22}$$

or, $E[\text{number of bets till game ends}] = \text{initial capital intended profit.}$ (23)

To see the result, consider that if the gambler starts with $10 dollars and plays until he is broke or ahead $10, then $10 \cdot 10 = 100$ bets are required on average. If he starts with $500 and plays until he is broke or ahead $100, then the expected number of bets until the game is over is $500 \times 100 = 50,000$.

2.4.6 Right Time to Quit

Let the gambler start with $n > 0$ dollars, without a target T, and let him play until going broke. Then we find that if the game is not favourable, i.e., $p \leq 1/2$, the gambler is sure to go broke. In particular, he is even sure to go broke in a "fair" game with $p = 1/2$.

Lemma 2.4.6.1. If the gambler starts with one or more dollars and plays a fair game until he is broke, then he will go broke with probability 1.

However, if the game is fair, he can "expect" to play for a very long time before going broke; in fact, he can expect to play forever.

Lemma 2.4.6.2. If the gambler starts with one or more dollars and plays a fair game until he goes broke, then his expected number of plays is infinite.

A game without bounds will have a larger expected number of bets compared to the bounded one because, in addition to the possibility that the gambler goes broke, in the bounded game there is also the possibility that the game will end when the gambler reaches his goal, T. That is,

$$u_n > e_n.$$

From (21) we have

$$u_n \geq n(T - n)$$

If $n \geq (T - n)$, and $T \geq n$ then lower bound on u_n can be arbitrarily large, which implies u_n must be infinity.

Take Lemma 2.4.6.1, along with a probability 1. Then the unbounded game ends when the gambler goes broke. Therefore, the expected time for the uhbounded game to end is the *same* as the expected time for the gambler to *go broke* and the expected time to go broke is infinite.

Probability of at least one event

Theorem 2.4.6. Let $A_1, A_2, \ldots A_n$ be independent events, and let T be the number of these events that occur. The probability that none of the events occur is at most $e^{-E[T]}$.

$$T = T_1 + T_2 + \ldots + T_n, \tag{26}$$

where T_i is the indicator variable for the event A_i. We also use the fact that

$$1 + x \leq e^x \tag{27}$$

for all x, which follows from the Taylor expansion

$$e^x = 1 + x + \frac{x^2}{2!} + \frac{x^3}{3!} + \cdots.$$

Two useful special cases of Theorem 2.4.6 are given in the following:

Corollary 15.2. Suppose an event has probqbility $1/m$. Then the probability that the event will occur at least once in m independent trials is at least approximately 1. $1/e \approx 63\%$. There is at least 50% chance the event will occur in $n = m \log 2 \approx 0.69m$ trials.

Check Your Progress 2

1) What is gambler's ruin problem?

....

....

....

....

....

2) What is the probability of a gambler winning in a fair coin toss game?

....

....

....

....

....

....

3) Write the meaning of recurrence for the probability of winning.

....

....

....

....

....

....

2.5 CHERNOFF BOUNDS

Probability of at least k Events

We generalise the above theorem from probability that one event occurs, to probability that k events occur, assuming mutual independence of the events A_i. In other words, we find the probability $\Pr\{T \geq k\}$, given that the events A_i are mutually independent.

Consider two examples, which describe different Scenario.

Example A. Suppose we want to know the probability that at least k heads come up in N tosses of a coin. Here A_i is the event that the coin is heads on the i th toss, T is the total number of heads, and $\Pr\{T \geq k\}$ is the probability that at least k heads come up.

Example B. Suppose that we want the probability of a student answering at least k questions .correctly on an exam with N questions. In this case, A_i is the event that the student answers the i th question correctly, T is the total number of questions answered correctly, and $\Pr\{T \ge k\}$ is the probability that the student answers at least k questions correctly.

The difference between these two examples: In the first, all events A_i have equal probability, i.e., the coin is as likely to come up heads on one flip as on another. So T has a binomial distribution whose tail bounds we have already characterised above.

On the other hand we have a situation where, some examination questions might be more difficult than others. If Question 1 is easier than Question 2, then the probability of event A_1 is greater than the probability of event A_2.

In following we develop a method to handle this moie general situation in which the events A_i may have different probabilities.

Taking Chernoff bound we will show that the number of events that occur is almost never much greater.than the expectation. For example, if we toss N coins, the expected number of heads is $N/2$ heads. The Chernoff Bound implies that for sufficiently large N , the number of heads is almost always not much greater than $N/2$.

Statement of the Bound

Chernoff s Theorem can be stated in terms of Bernoulli variables instead of events. However, we oan take T_i as an indicator for the event A_i.

Theorem 2.5. (Chernoff Bound). Let $T_1, T_2, ..., T_n$ be mutually independent Bernoulli variables, and let

$$T = T_1 + T_2 + ... + T_n .$$

Then for all $c \ge 1$, we have

$$\Pr\{T \ge cE[T]\} \le e^{-(c\ln c - c + 1)E[T]} \tag{28}$$

The formula for the exponent in the bound'is a little awkward. The situation is simpler when $c = e = 2.718...$ In this case,

$$c \ln c - c + 1 = e \ln e - e + 1 = e \cdot 1 - e + 1 = 1,$$

so we have as an immediate corollary of Theorem 2.5:

Corollary 15.2. Let $T_1, T_2, ..., T_n$ be mutually independent Bernoulli variables, and let $T = T_1 + T_2 + ... + T_n$. Then

$$\Pr\{T \ge eE[T]\} \le e^{-E[T]}$$

2.6 STRONG LAW OF LARGE NUMBERS

We described the *Weak* Law of Large Numbers above and took up the question of what was the *strong* law of large numbers we might prove. Roughly speaking, the strong law says that with probability 1, the bound of the weak law will hold for all but a finite number of the S_n simultaneously–there will only be finitely many exceptions to it.

Theorem 2.6. [The Strong Law of Large Numbers]

Let $S_n = \sum_{i=1}^{n} X_i$ where $X_1,...,X_i,...$ are mutually independent, identically distributed random variables with finite expectation, μ. Then

$$\Pr\left\{\lim_{x\to\infty} \frac{S_n}{n} = \mu\right\} = 1.$$

Although Theorem 2.6 can be proven without this assumption, we will assume for simplicity that the random variables X_i have a finite fourth moment. That is, we will suppose that

$$E\left[X_i^4\right] = K < \infty. \tag{29}$$

Check Your Progress 3

1) What do you mean by symmetric random walk?

....

....

....

....

....

....

2) If a gambler is playing with the aim of winning a fixed amount, what are the chances of his success?

....

....

....

....

....

....

3) Suppose an event has probability of $\frac{1}{m}$. Find the probability that the event will occur at least once in independent trials.

....

....

....

....

....

....

4) Take $T_1, T_2, ...T_n$ to be mutually independent Bernoulli variables and let $T = T_1 + T_2 + ... + T_n$. So that

$\Pr\{(T \geq cE[T])\} \leq \exp(-(c \log c - c + 1)E[T])$

....

....

....

....

....

....

5) Write the meaning of Strong Law of Large Numbers.

....

....

....

....

....

....

2.7 LET US SUM UP

In this unit we have discussed the probabilistic idea that the average of a large number of trials is likely to be within an acceptable level of the expectation. For that purpose we have examined the results of Markov's theorem, Chebychev's theorem and the pairwise independent sampling theorem. Building on such findings we extend the coverage to include random walk and discussed the gambler's ruin problem with the different scenarios of initial capital as well as conditions of variance.

2.8 KEY WORDS

Chernoff Bound: A lower bound for the success of majority agreement for n independent, equally likely events. The Chernoff bound is a special case of Chernoffs inequality.

Chernoff's inequality: It states the following:

Let $X_1 X_2, ... X_n$

be independent random variables, such that

$E[X_i] = 0$ and $|X_i| \leq 1$ for all i.

Gambler's Ruin: A 'gamblerts loss of the last of her unit of gambling money and consequent inability to continue gambling. It is is also sometimes used to refer to a final large losing bet placed in the hopes of winning back all the gambler has lost during a gambling session. More generally, the phrase refers to the ever decreasing expected value of a gambler's initial capital as she continues to gamble with her winnings.

Markov Property: A property of a stochastic process in which the past and future are independent.

Random Walk: Taking successive steps, each in a random direction.

Recurrence: A relation in an equation which defines a sequence recursively, i.e., each term of the sequence is defined as a function of the preceding terms.

Stopping Time: With respect to a sequence of random variables X_1, X_2, ... a random variable τ with the property that for each t, the occurrence or non-occurrence of the event $\tau = t$ depends only on the values of $X_1, X_2, ..., X_t$. Thus, at any particular time t, you can look at the sequence so far and tell if it is time to stop.

Strong Law of Large Numbers: As the sample size grows larger, the probability that the sample mean and the population mean will be exactly equal and approaches 1.

Weak Law of Large Numbers: As the sample size grows larger, the difference between the sample mean and the population mean will approach zero.

2.9 SOME USEFUL BOOKS

Meyer, A. and Radhika Nagpal (2002), Mathematics for Computer Science (course Notes 13-14), MIT, Cambridge, MA (see internet)

Ross, Sheldon, (1996), Stochastic Processes, Second Edition, John Wiley & Sons, Inc. New York

2.10 ANSWER OR HINTS TO CHECK YOUR PROGRESS

Check Your Progress 1

1) See Section 2.3 and answer.
2) See Section 2.3 and answer.
3) See Section 2.3 and answer.
4) Hint: Take $\Pr\{F=0\} = p_s^n$ and proceed to find
   $\Pr\{F \le 1\} = p_s^n + n(1-p_s)p_s^{n-1}$. Get
   $E[F] = \sum_{i=1}^{n} i\binom{n}{i}(1-p_s)^i p_s^{n-i} = n(1-p_s)$

Check Your Progress 2

1) See Section 2.4 and answer
2) See Sub-section 2.4.2 and answer
3) See Sub-section 2.4.3 and answer

Check Your Progress 3

1) Read Section 2.4 and answer

2) Hint: Always succeeds in the'long run

3) Hint: Get $1-\frac{1}{e}\approx 63\%$. So there is a 50% chance the event will occur in $n = \log 2m \approx 0.69m$ trials

4) Hint: Find Chernoff Bound

5) See Section *2.6* and answer

2.11 EXERCISES

1) Suppose you roll n dice that are 6-sided, fair and mutually independent. What is the expected value of the largest number that comes up?

   Ans. $6-\frac{1^n+2^n+3^n+4^n+5^n}{6^n}$

2) In casino following game is played: The gambler bets a fixed wager and the owner of casino flips a fair coin until it comes of heads. The gambler receives $1 if the coin turns up heads the first time, $2 if it turns up the first head at second toss. Let the owner tosses the coin k times, and gambler gets $\$2^{k-1}$ to get the first head.

   i) What is the expected amount of money that the gambler will win with a fixed wager of $10?

   Ans. ∞ amount

   ii) What is the probability that the gambler does not lose money in a game when fixed wager is $10,000?

   Ans. $\frac{1}{16,384}$

3) Suppose that all dice are fair and six-sided and that dice rolls are mutually independent. Find the expected value for each of the following random variables.

   i) Sum of the rolls of three dice. (Ans. $\frac{21}{2}$)

   ii) Product of the rolls of three dice (Ans. $\frac{343}{8}$)

4) Two gambling games have the following patterns of bets .

   Game 1: You win $2 with probability 213 and lose $1 with probability 113.

   Game 2: You win $1002 with probability 2/3 and lose $2001 with probability 113.

   i) What is the expected win in each case? (Ans. 1)

   ii) What is the variance in each case (Ans. 2)

5) Suppose you have learned tat the average MEC students total number of marks 200.

a) Use Markov's inequality to find a best possible upper bound for the fraction of MCE students with at least 235 marks. (Ans. $\frac{200}{235}$)

b) Suppose you are told that no student can pass with less than 170 marks. How does this help to improve your previous bound? Show that this is the best possible bound. (Ans. $\frac{30}{65}$)

c) Suppose that you further learn that the standard deviation of the total marks per studdnt is 7. Give a best possible bound on the fractian of students who can graduate with at least 235 marks. (Ans. $\frac{1}{25}$)

6) Let X, Y be independent Binomial random variables with parameters (n,p) and (m,p), respectively. What is $\Pr\{X+Y=k\}$?

(Ans. $\binom{m+n}{k} p^k (1-P)^{m+n-k}$

Wald Theorem

Let $C_1, C_2, \ldots,$ be a sequence of nonnegative random variables, and let Q be a positive interger-valued random variable, all with finite expectations, suppose that

$$E[C_i \mid Q \geq i] = \mu$$

for some $\mu \in \mathbb{R}$ and for all $i \geq 1$. Then

$$E[C_1 + C_2 + \ldots + C_Q] = \mu E[Q].$$

UNIT 3 STOCHASTIC PROCESS

Structure

3.0 OBJECTIVES

After going through this unit, you will be able to:

- understand the features of important processes used for analysing problems which are chasracterised by time and randomness; and
- identify appropriate stochastic process for model building in insurance.

3.1 INTRODUCTION

An important task before the insurance industry is risk management. Events often emerge in course of its functioning which are difficult to predict. Therefore, attempts are made to understand their behaviour using different models. The following discussion introduces some of stochastic processes which have drawn the attention of analysts working in the field of actuarial science. In discussing the themes we have put greater emphasis on the underlying meaning of stochastic processes to help grasp their importance from application perspective. The discussion presented in this unit draws its inputs mainly from two sources. One is a slide presentation entitled, 'Stochastic Models for Actuarial Application' available in the internet without containing the author's name. Some of definitions and examples used in the unit have been taken from the source (and we thank the author for the efforts). The other source we have utilised is Marco (2005). Some of the notations and equations are adopted from it without acknowledging the source. The limited purpose of introducing the concepts has resulted in sidelining many exciting formulations and their derivations from the purview of present unit. For reference an elaborate treatment on the themes Ross (1 996) could be seen.

3.2 STOCHASTIC MODELS

In economic problems. sequences of events take time and occurrence of many events is unpredictable. As a stochastic process means time and randomness, we will discuss the events in the framework stochastic process.

Let a stochastic process $X = \{X(t)\}\ t \in T$ be a collection of random variables. For each t in the index set $T, X(t)$ is a random variable, When we interpret t as time, $X(t)$ represents the state of the process'at time t. If the index set T is a countable set, we have a discrete-time stochastic process, and when non-countable continuous, a continuous-time stochastic process, In most cases, a stochastic variable has both an expected value term (drift term) and a random term (volatility term).

3.2.1 Classification

If the random variable $X(t)$ is

- discrete then the process $\{X(t), t \in T\}$ has a discrete state space; and
- continuous then the process $\{X(t), t \in T\}$ has a continuous state space.

3.2.2 Sample Path

Any realisation of a stochastic process X is called a sample path. Thus, a sample path or a realisation of stochastic process $\{X(t), t \in T\}$ is a particular assignment of possible values for $X(t)$ for all $t \in$ T.

For example, the observed share prices for an asset at the end of each day's trading over a year is a realisation or sample path for the share price.

3.2.3 Increments

Consider two points of time and take

$X(t+u) - X(t), u > 0$.

Then we have two categories of increases:

A stochastic process has independent increments, if the random variables

$X(t_0)$, $X(t_1) - X(t_0)$, $X(t_2) - X(t_1)$, ... $X(t_n) - X(t_{n-1})$

are independent for all shares or, equivalently is independent of $X(s)$ for $s \le r$.

3.2.4 Stationary Increments

A stochastic process has stationary increments, if $X_{(t_2+r)} - X_{(t_1+r)}$ has the same distribution of $X(t_2) - X(t_1)$ for all choices of t_1, t_2 and r > 0.

3.3 GENERAL CONCEPTS

Some general concepts of stochastic processes are presented below:

3.3.1 Lévy Processes

Lévy Processes are stochastic processes with stationary independent increments and continuous in probability. Stationary increment property means that the probability distribution for the changes in the stochastic variable X depends only on the time interval length,

Independent increments means that for all time instant t, the increments are independents. The two most basic types of LCvy processes are **Wiener processes** and **Poisson process.**

3.3.2 Lévy Process

Lévy Processes have the following property: given that its current state is known, the probability of any future event of the process is not altered by additional knowledge concerning its past behavior.

Formally, the probability distribution for X,,, depends only on X, and not additionally on what occurred before time t. That is, it doesn't depend on X_s where $s < t$).

3.3.3 Itô Process

Itô Process is a generalised Wiener process. A Wiener process is also a special case of a **strong diffusion process** that is a particular class of a continuous time Markov process.

A continuous time **Wiener Process** (also called Brownian motion) is a stochastic process with three properties:

- It is a **Markov Process.** This means that all the past information is considered in'the current value, so that a future's value of the process depends only on its current value not on past values. The future values are not affected by the past values history. In finance, this is consistent with the efficient market hypothesis (that is the current prices reflect all relevant information).
- It has **independent increments**. Change in one time interval is independent of any other time interval (non-overlapping).
- The change of value over any finite time interval is normally distributed,

So, it has stationary increments, besides the property of independent increments. Therefore, it is a particular Lévy process (Lévy process with normal distribution for the increments).

The Itô process (or generalised Wiener Process) for the value of a project V is:

$$dV = a(V,t)\,dt + b(V,t)\,dz$$

The generalised term is because the drift $a(V,t)$ and the variance $b(V,T)$ coefficients are functions of the present state and time.

The integral version for this equation is the following **stochastic integral:**

$$V(t) = V(0) + \int_0^t a(V,s)\,ds + \int_0^t b(V,s)\,dz$$

As an illustration of stochastic integral, consider the following mean-reverting process (arithmetic Ornstein-Unlenbeck) process:

$$dx = \eta(m-x)\,dt + \sigma dz$$

where m is an equilibrium level to which the process reverts and η is the reversion speed. The value of the stochastic process at a future date T, that is, $x(T)$, given the starting value

$X(0)$ is: $x(T) = X(0)e^{-\eta T} + (1 - e^{-\eta T})m + \sigma e^{-\eta T}\int_0^{T} e^{\eta t}dz(t)$

3.3.4 Point Process

Point Process is a stochastic process whose realisations are, instead of continuous sample paths, counting measures. Any counting process, which is generated, by an independent identically distributed (iid) sum process (T_n) is called renewal counting process.

The simplest and most fundamental point process is the **Poisson process,** also referred as jump **process** in the financial literature.

The **Poisson process** is a counting process in which interarrival times of successive jumps are independently and, identically distributed (i. i.d) exponential random variables.

The Poisson process is an example of renewal counting process.

Homogeneous Poisson process has the following three properties:

1) It starts at zero. This means that at $t = 0$, there is no jump. From counting of view, $N(0) = 0$ (the number of jumps in the process N is zero at $t = 0$).

2) It has independent stationary increments. That is, a Poisson process is also a special Lévy process.

3) For $t > 0$, the probability of n jumps occurrence until time t is:

$$P[N(t) = n] = (1/n!)(\lambda t)^n e^{-\lambda t}; n = 0, 1, 2...$$

which is a Poisson distribution with parameter λt

The Poisson distribution tends to the normal distribution as the frequency λ tends to infinity. So, a Poisson distribution is asymptotically normal.

A Non-homogeneous Poisson Process is more general than the homogeneous one: the stationary increment assumption is not required (and the independent increment assumption is retained), and the constant arrival rate λ of a Poisson process is replaced by a time-varying intensity function.

3.3.5 Compound Poisson Process

Let Φ_i be a sequence of i.i.d. random variable. These identical probability distributions can be interpreted as the jump-size distributions. Let $N(t)$ be a Poisson process, independent of Φ_i. Then the following process is called a compound Poisson process:

$$X(t) = \sum_{j=1}^{N(t)} \Phi_j$$

The sum of two independent compound Poisson processes is itself a compound Poisson process.

A jump is degenerate when the variable can only jump to a fixed value.

Example 3.1

The case of "sudden death" process is degenerated jump, in which the variable drops to zero forever. It is interesting to note that combination of Poisson process with Brownian motion is related to LCvy process. In the sample paths of LCvy processes the large increments or "jumps" are called **"Lévy flights".**

3.3.6 Poisson-Gaussian Processes

Also named as jump-diffusion processes are a combination of Itô process with a Poisson process, (a mixed, continuous with discrete), can be described by the following equation:

$$dv = a(V,t)\,dt + b(V,t)\,dz + V dq$$

where the terms are as before but the additional dq is a Poisson term defined by:

$dq = 0$ ......with probability $1 - \lambda dt$

$dq = \Phi - 1$ ......with probability Adt

The processes dz and dq are independents.

In the above equation, A is the **arrival** rate of jumps and Φ is the jump size **probability** distribution. Note that the jump size/direction can be random.

In case of jump, the stochastic variable V will change from V to $V\Phi$.

The other equivalent format to write this jump-diffusion is:

$$dV = a(V,t)dt + b(V,t)dz + V(\Phi - 1)dq$$

where the Poisson term dq is here defined by:

dq = 0 ....with probability $1 - \lambda dt$

dq = 1 ....with probability λdt

The integral version for this equation is given by

$$V(t) = V(0) + \int_0^t a(V,s)ds + \int_0^t b(V,s)dz + \sum_{j=1}^{N(t)} c(V,s)dq_j$$

where $N(t)$ is the number of jumps until the instant t drawn from Poisson distribution. In most cases

$$a(V,s) = \text{drift} = \alpha V$$

for the geometric Brownian case; but could be mean-reverting drift, which has other format

$$b(V,s) = \sigma V$$

$$c(V,s) = V(\Phi - 1)$$

For the geometric Brownian with Poisson-jumps, the logarithm version $(v = LnV)$ of the stochastic equation is:

$$d(Ln\,V) = (a - 0.5\sigma^2)dt + \sigma dz + Ln(\Phi)dq$$

The returns generated by a jump-diffusion process are not normally distributed, presenting fatter tails.

When comparing with a normal distribution (from pure Brownian processes), the above distribution presents higher peak and consequently fatter tails.

Even being small tails (low probability density), its size could be large losses (or large gains) for financial or real derivatives (for meaning see Unit 4). In 'that case, jumps can be very important.

In financial applications, the most relevant jump-diffusion process is a combination of mean-reversion (continuous process) with jumps (discrete Poisson process).

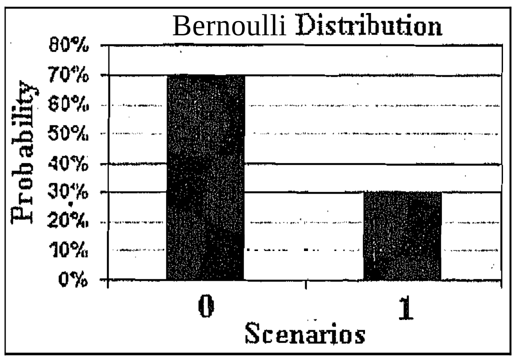
In real options applications, in general, we are interested in low frequency large jumps (rare large jumps), which can be jump-up or jump-down.

Less realistic for commodity prices is the combination of large jumps with high frequency of jumps arrival. We say less realistic because the idea is to associate jumps with "abnormal" rare events (very important event), whereas the continuous process (mean-reversion) is associated with the "normal event".

The problems with jump-diffusion processes are the parameters estimation and the portfolio hedging. Even with some practical problems, the jump-diffusion process is perhaps the best for most uncertainty of commodities prices.

Bernoulli jump process: Poisson process as limit of Bernoulli process

Let us recall that Bernoulli distribution is a univariate discrete distribution with only two discrete scenarios, one occurrence of events with value zero (failure) and another with value 1 (success). The parameter p (probability of the event with value 1) is also the mean of this distribution. The figure below illustrates the Bernoulli distribution for $p = 0.3$ (or 30%)..



Bernoulli distribution is the building block to construct many other distributions in probability, such as binomial, geometric, hypergeometric, negative binomial, and Poisson. Here we will use the Bernoulli distribution to set the occurrence (event $X = 1$) of a jump in a stochastic variable,

Let us first define the concept of a function f being $0(h)$ ("order h"). The function f is called to be $0(h)$ if:

$$\lim_{h \to 0} \frac{f(h)}{h} = 0$$

The link between Bernoulli and Poisson distribution can be developed as follows:

Let $N(t)$ be the number of events (jumps) arriving in a time interval of length t. Denote n as a positive integer, 'and consider the subdivision of the time interval $(0,t)$ into n equal subintervals of length h, so that $h = t/n$. Let X_j be the number of jumps (events) that occurs in subinterval j. By the stationary independent increment assumption, we have

$$N(t) = \sum_{j=1}^{n} X_j$$

This is a counting process summing n i.i.d. random variables with the properties:

$*N(0)=0$

$*Prob[X_j=0]=1-\lambda h+o(h)$

$*Prob[X_j=1]=\lambda h+o(h)$

$*Prob[X_j>1]=o(h)$

If n is a large number, the probability of a jump in this subinterval j is negligible so that each X, is approximately a Bernoulli random variable with parameter $\lambda h=\lambda t/n$.

There is an alternative way to define a Poisson, which is equivalent to the previous definition, as we will see below.

Being a sum of independent and identically distributed Bernoulli random variable, N has a binomial distribution. The binomial distribution is given by:

$$Prob[N=K]\cong\binom{n}{k}\left(\frac{\lambda t}{n}\right)^k\left(1-\frac{\lambda t}{n}\right)^{n-k}$$

where $k=0,1,2\ldots.$

What happens if the time subinterval is very small, that is, if n tends to be infinity? The binomial distribution tends to the Poisson distribution with parameter $\lambda\ t$. Thus,

$$\lim_{n\to\infty}\binom{n}{k}\left(\frac{\lambda t}{n}\right)^k\left(1-\frac{\lambda t}{n}\right)^{n-k}=\frac{\exp(-\lambda t)(\lambda t)^k}{k!}$$

Let us assume that the time interval t is very small. Consequently, N is well approximated by the Bernoulli random variable X with parameter At.

That is, X - $Be(\lambda t)$ given by:

$*Prob[X=0]=1-\lambda t$

$*Prob[x=1]=\lambda t$

The main feature of the Bernoulli jump process is that over a period of time t (i.e., Iday), either no abnormal (jump) information occurs or only one relevant abnormal (jump) information occurs.

Under regular conditions, the parameter estimated are asymptotically unbiased, consistent, and efficient, which can be obtained through maximum likelihood estimation.

In addition, the statistical test of the null hypothesis $\lambda=0$, can be implemented.

The parameter from the Poisson-Gaussian process of mixing jump process with a geometric Brownian motion (GBM) can be estimated with a Bernoulli mixture of **normal densities.** Over $t=1$day, the stock return density **f (X)**s in this case:

$$f(X)=(1-\lambda)\Phi(\sigma,\sigma^2)+\lambda\Phi(\sigma,\sigma^2+\xi^2)$$

where the mean jump size equal =0, jump size variance $=\xi^2$, and daily volatility of the GBM $=\sigma$. The above density expression can be interpreted as the convex combination of two normal densities one of them with the variance augmented by the jump-size variance.

Assuming n daily stock return $X=(X_i, i=1,2,\dots n)$, and denoting by y the vector of parameters to estimate (λ,σ,ξ,a), the logarithm of the likelihood function $L(x;\gamma)$ is

$$\ln L(\underline{x};\underline{\gamma})=\sum_{i=1}^{n}\ln f(X_i;\underline{\gamma})$$

So, the empirical job of maximum likelihood estimation consists in maximising the above function for the four parameters, This is performed by deriving the above expression in relation to each one of the four parameters to estimate and equating these equations to zero. The sufficient conditions for the existence of maximum require that the matrix $H(X;\gamma)$ be positive definite, where $H(X;y)$ is formed by the second derivatives of the ln $L(\underline{x};\underline{\lambda})$ with relation to the parameters to be estimated,

3.3.7 Markov Process

Definition 3.1

$$\text{Prob}\left(X(t_n)\le x_n \middle| \begin{matrix} X(t_1)=x_1, X(t_2)=x_2 \\ X(t_{n-1})=x_{n-1}\end{matrix}\right)$$

$$=\text{Prob}\left(X(t_n)\le x_n \middle| X(t_{n-1})=x_{n-1}\right)$$

for $\quad t_1<t_2<\dots<t_n$

A process with in dependent increments has the Markov property

Proof:

$$P_r\left(X(t_n)\le x_n \middle| X(t_1)=x_1, X(t_2)=x_2,\dots,X(t_{n-1})=x_{n-1}\right)$$

$$=P_r\left(X(t_n)-X(t_{n-1})+x_{n-1}\le x_n \middle| X(t_1)=x_1,\dots,X(t_{n-1})=x_{n-1}\right)$$

$$=P_r\left(X(t_n)-X(t_{n-1})+x_{n-1}\le x_n \middle| X(t_{n-1})=x_{n-1}\right)$$

$$=P_r\left(X(t_n)\le x_n \middle| X(t_{n-1})=x_{n-1}\right)$$

3.4 MARTINGALE

Definition 4.1. **A** discrete-time stochastic process is a martingale if

$E[X(t_n)]<\infty$ for all n and

$$E\left(X(t_n)\middle| X(t_1)=x_1, X(t_2)=x_2,\dots,X(t_{n-1})=x_{n-1}\right)=x_{n-1}$$

for $\quad t_1<t_2<\dots<t_n$

Note that for a martingale

$$E\left[X(t_n)\middle|X(t_1),X(t_2)...,X(t_m)\right]$$

$$= E\left[X(t_m)\middle|X(t_1),X(t_2)...,X(t_m)\right]$$

$$= E\left[X(t_m)\right] \text{ for } m < n$$

Therefore,

$$E\left[\left(X(t_m) = E\left\{E\left[X(t_n) \mid X(t_1),....,X(t_m)\right]\right\}\right.\right.$$

$$= E\left[X(t_n)\right] \text{ for } m < n$$

Also note that

$E[X(t_n)] = E[X(t_0)]$ for all n if $X(t_n)$ is a martingale (see Appendix given at the end).

Example 4.1

- . If you expect to earn no interest or capital gains on an investment, then the share price is a martingale.
- If you gamble at a casino and the games are fair then your gambling wealth will be a martingale.
- If the price of an investment is linked to the CPI then the price of the asset divided by the CPI is a martingale.

3.4.1 Probability Space

We will assume that we are given a probability triple $(\Omega, \mathcal{F}, P)$ consisting of:

1) sample space Ω the set of all possible outcomes, or states of the world.
2) sigma algebra $\mathcal{F}$ the set of all subsets of Ω. It is the set of events that we may be interested in,.
3) probability measure P the real-world probability measure. It assigns probabilities to the set of events (see Appendix given at the end of Unit 1).

3.4.2 Filtration

In stochastic process, we are generally interested in how information is revealed over time. At time t, this information is denoted by F,and it gives the history of the process up to time t. It is true that because the amount of information available increases over time, for any $t < s$, we have

$$\mathcal{F}_t \subseteq \mathcal{F}_s$$

$\mathcal{F}_t$ is the sigma-algebra generated

$$\{X(r), r \in T\} \text{ for all } 0 \le r \le t.$$

The set $(\mathcal{F}_t)$ is called a filtration.

Example 4.2

Consider the share price of Maruti Company over two days. The following table shows the prices and the sample space:

ω_k	**Today**	**Tomorrow**	**Next day**
ω_1	100	110	120
ω_2	100	110	90
ω_3	100	90	100
ω_4	100	90	90

(*Adopted from* internet source)

We have $\Omega = \{\omega_1, \omega_2, \omega_3, \omega_4\}$ which is all the possible price paths over the two days.

Now $\mathcal{F}_t = \{\theta \ \Omega\}$ since you have no information about the actual future price path just about the possible price paths (based on your model).

$\mathcal{F}_t = (\theta \ \Omega, (\omega_1, \omega_2), (\omega_3, \omega_4))$ since you know tomorrow that the price will follow path ω_1 or w_2 if the: price is 110 tomorrow or will follow path w_3 or w_4 if the price is 90.

Finally, $= (\theta \ \Omega, (\omega_1, \omega_2), (\omega_3, \omega_4), (\omega_1), (\omega_2), (\omega_3), (\omega_4))$ since the actual price path is revealed on day 2.

3.4.3 Natural Filtration

In stochastic processes, we are usually interested in expectations conditional on the information revealed at,a specific time as given by the filtration.

For example,

$$E\left[X(t)|\mathcal{F}_s\right]$$

where $s \leq t$ is the conditional expectations of $X(t)$ given the information available at the earlier time s. This information will usually include all the previous values revealed by the process up to time s, $X(r)$ with $r \leq s$.

This is called the natural filtration for $X(t)$

3.4.4 Stopping Time

A stopping time for a stochastic process $\{X(t), t \in T\}$ is a random variable N such that the event $\{N = t\}$ does not depend on $X(s)$ for $s > t$.

If you know that the event has occurred based on information up to and including the time of the event, then it is a stopping time.

On the other hand, if you need to know the future in order to know if an event has occurred, then it is not a stopping time.

Example 4.3

The time for which an interest rate exceeds 5% (or any other fixed interest) is a stopping time. The time for which interest rate reaches a maximum during the year is not a stopping time. We cannot know if an interest rate has reached a maximum at any time point without knowing all the future possible interest rates.

Example 4.4

Consider a stochastic process $\{X(t), t \in T\}$ where

$$X(t) = \begin{Bmatrix} 0, & wp\ p \\ 1, & wp\ 1-p \end{Bmatrix}$$

and define the random variables

$$N_1 = \min\left(n \sum_{t=1}^{n} x(t) = 5\right)$$

$$N_2 = \begin{cases} 2 & \text{if } X(1) = 0 \\ 4 & \text{if } X(1) = 1 \end{cases}$$

$$N_3 = \begin{cases} 5 & \text{if } X(4) = 0 \\ 3 & \text{if } X(4) = 1 \end{cases}$$

Which of these are considered stopping times for the stochastic process?

Stopping times are particularly useful in conjunction with martingales. Since $E[X_1] = E[X_0]$ for all t for is a martingale, it is tempting to conjecture that $E[X_T]$ will also equal $E[X_0]$ when T is a random time.

3.4.5 Optional Stopping Theorem

If $X(t) t = 0,1,2,\ldots$ is a martingale and N is a stopping time and both X and N are bounded, then:

Example 4,4

Assume that a share price is a martingale with current price $1 00. Consider the time N when the price exceeds $100 for the first time. This will be a stopping time and $E[X(N)] = E[X(t_0)] = 100$

3.5 MARKOV CHAIN

Definition 5.1. A Markov Chain (MC) is a Markov process on a discrete index set $T = \{0, 1, 2, \ldots\}$ which we will denote by X_n, $n = 0, 1, 2, \ldots$ The state space can be either finite, $S = (0, 1, 2, \ldots, n)$ or countable, $S = \{0, 1, 2, \ldots)$. If $X_n = i$ then the process is said to be in state i at time n.

The Markov Property

In a Markov chain, given the present state, the past states do not have influence on the future. This is called the Markov property and can be stated as:

$$P_r(X_{n+1} = j | X_0 = x_0, X_1 = x_1, ..., X_n = i) = P_r(X_{n+1} = j | X_n = i)$$

One-step transition probabilities

Definition 5.2. The conditional probability

$$P_{ij}^{(n,n+1)} = P_r(X_{n+1} = j | X_n = i)$$

is the probability that the process will move from state i at time n to state j as time $n + 1$ and is called a one-step transition probability.

These conditional probabilities are the key concepts for describing a Markov chain.

Time-independent one-step Transition Probabilities

In general, the one-step transition probabilities can depend on time n, If the one-step transition probabilities do not depend on time n, .then we drop the n in the notation and write:

$$P_{ij} = P_{ij}^{(n,n+1)} = P_r(X_{n+1} = j | X_n = i)$$

In this case, we said we have a homugenous Markov Chain and we call these transition probabilities stationary.

3.5.1 Homogeneous Markov Chains

Definition 5.3. A homogeneous Markov chain has $P_{ij}^{(n,n+1)} = P_{ij}$ for all n. The one-step transition probabilities do not depend on time n.

Definition 5.4. The transition probability matrix is given by

$$P = [P_{ij}] = \begin{bmatrix} P_{00} & P_{01} & P_{02} & \cdots \\ P_{10} & P_{11} & P_{12} & \cdots \\ P_{20} & P_{21} & P_{22} & \cdots \\ \vdots & \vdots & \vdots & \ddots \end{bmatrix}$$

Example 5.1 (Forecasting the Weather)

Suppose that the chance of rain tomorrow depends on previous weather conditions only through whether or not it is raining today and not on past weather conditions. Suppose also that if it rains today, then it will rain tomorrow with probabilitya ; and if it does not rain today, then it will rain tomorrow with probability β .

If we say that the process is in state 0 when it rains and state 1 when it does not rain,-then the above is a two-state Markov chain whose transition probabilities are given by:

$$P = \begin{Vmatrix} \alpha & 1-\alpha \\ \beta & 1-\beta \end{Vmatrix}$$

Example ***5.2***

One any given day John is either cheerful (C), so-so (S), or glum (G). If he is cheerful today, then he will be C, S, or G tomorrow with respective probabilities 0.5, 0.4, 0.1. If he is feeling so-so today, then he will be C, S, or G tomorrow with probabilities 0.2, 0.3, 0.5. Let Xdenote John's mood on the nth day. Then $\{X_n, n \succ 0\}$ is a three-state Markov chain (state $0 = C$, state1 = S, state 2 = G) with transition probability matrix.

$$P = \begin{Vmatrix} 0.5 & 0.4 & 0.1 \\ 0.3 & 0.4 & 0.3 \\ 0.2 & 0.3 & 0.5 \end{Vmatrix}$$

3.5.2 Transition Probability Matrix

The transition probability matrix is given by

$$P = [P_{ij}] = \begin{bmatrix} P_{00} & P_{01} & P_{02} & \cdots \\ P_{10} & P_{11} & P_{12} & \cdots \\ P_{20} & P_{21} & P_{22} & \cdots \\ \vdots & \vdots & \vdots & \ddots \end{bmatrix}$$

Note that if the state space is finite then the transition probability matrix is a square matrix and satisfies the following:

$$P_{ij} \geq 0, i.j = 0,1,2$$

$$\sum_{j=0}^{\infty} P_{ij} \geq 1, \text{for } i. = 0,1,2...$$

If sums across rows are all equal to then and such a matrix is called a stochastic matrix.

Example 5.3 (Transforming a process into a Markov Chain).

Suppose that whether or not it rains today depends on previous weather conditions through the last two days. Specifically, suppose that if it has rained for the past two days, then it will rain tomorrow with probability 0.7; if it rained today but not yesterday, then it will rain tomorrow with probability 0.5; if it rained yesterday but not today, then it will rain tomorrow with probability 0.4; if it has not rained in the past two days, then it will rain tomorrow with probability 0.2.

Example 5.4 (A Random Walk Model).

A Markov chain whose state space is given by the integers

$$i = 0 \pm 1, \pm 2, \ldots$$

is said to be a random walk if, for some number $0 < p < 1$.

$$P_{i,i+1} = p = 1 - P_{i,i-1} \qquad i = 0 \pm 1 \ldots$$

The preceding Markov chain is called a random walk for we may think of it as being a model for an individual walking on a straight line who at each point of time either takes one step to the right with probabilityp or one step to the left with probability $1 - p$.

n-step Transition Probability

The n-step transition probability P_{ij}^n is the probability that a process in state i will be in state j in steps, that is

$P_{ij}^n = \Pr ob(X_{n+m} = j | X_m = i)$

Example 5.5 (Simple random walk).

Consider a process that starts at the origin and can either increase by 1 with probability p or decrease by 1 with probability q = 1 – p over each time step. The state space is $i = 0 \pm 1, \pm 2, \ldots X_0 = 0; P_{i,i+1}; P_{i,i-1} = q = 1 - p.$

$$P_{ij}^n = \begin{cases} \left(\dfrac{n+j-1}{2}\right) p^{\frac{n+j-1}{2}} q^{\frac{n-j+1}{2}} & \text{if } n+j-i \text{ is even} \\ 0 & \text{otherwise} \end{cases}$$

since there will be $\dfrac{n+j-i}{2}$ positive steps and $\dfrac{n-j+i}{2}$ negative steps in order to go from state i to j in a steps.

The n-step transition probability P_{ij}^n is the probability that a process in state i will be in state j in n steps, that is:

$P_{ij}^n = \Pr ob(X_{n+m} = j | X_m = i)$

3.5.3 Chapman Kolmogorov Equations

The Chapman Kolmogorov equations provide a producer to compute the n-step transition probabilities. These equations are given as follows:

$$P_{ij}^{n+m} = \sum_{j=0}^{\infty} P_{ik}^n P_{kj}^m, \text{for all } n, m \geq 0 \text{ all } i, j.$$

In matrix form, we can write these as:

$$\mathbf{P}^{(n+m)} = \mathbf{P}^{(n)} \bullet \mathbf{P}^{(m)}$$

where $\bullet$ denotes matrix multiplication. Here $\mathbf{P}^{(n)} = (P_{ij}^n)$ is the matrix consisting of the n-step transition probabilities,

Note that beginning with

$$\mathbf{P}^{(2)} = \mathbf{P}^{(1+1)} = \mathbf{P} \bullet \mathbf{P} = \mathbf{P}^2$$

and continuing by induction, we can show that the n-step transition matrix can be obtained by multiplying matrix P by itself n times, that is,

$$\mathbf{P}^{(n)} = \mathbf{P}^{(n-1+1)} = \mathbf{P}^{n-1} \bullet \mathbf{P} = \mathbf{P}^n$$

$$P_{ij}^{(n+m} = P_r(X_{n+m} = j | X_0 = i)$$

$$= \sum_{j=0}^{\infty} \Pr(X_{n+m} = j, X_n = k | X_0 = i)$$

$$= \sum_{j=0}^{\infty} \Pr(X_{n+m} = j | X_n = k, X_0 = i) \Pr(X_n = k | X_0 = i)$$

$$= \sum_{j=0}^{\infty} \Pr(X_{n+m} = j | X_n = k) \Pr(X_n = k | X_0 = i)$$

Thus, we have

$$P_{ij}^{n+m} = \sum_{j=0} P_{ik}^{n} P_{kj}^{m}$$

Example 5.6

Consider Example 5.1 in which the weather is considered as a two-step Markov chain. If $\alpha = 0.7$ and $\beta = 0.4$, then calculate the probability that it will rain four days from today given that it is raining today.

Solution: The one-step transition probability matrix is given by:

$$P = \begin{Vmatrix} 0.7 & 0.3 \\ 0.4 & 0.6 \end{Vmatrix}$$

Hence,

$$\mathbf{P}^{(2)} = \mathbf{P}^2 = \begin{Vmatrix} 0.7 & 0.3 \\ 0.4 & 0.6 \end{Vmatrix} \cdot \begin{Vmatrix} 0.7 & 0.3 \\ 0.4 & 0.6 \end{Vmatrix}$$

$$= \begin{Vmatrix} 0.61 & 0.39 \\ 0.52 & 0.48 \end{Vmatrix}$$

$$\mathbf{P}^{(4)} = (\mathbf{P}^2)^2 = \begin{Vmatrix} 0.61 & 0.39 \\ 0.52 & 0.48 \end{Vmatrix} \cdot \begin{Vmatrix} 0.61 & 0.39 \\ 0.52 & 0.48 \end{Vmatrix}$$

$$= \begin{Vmatrix} 0.5749 & 0.4251 \\ 0.5668 & 0.4332 \end{Vmatrix}$$

and the desired probability P_{00}^{4} equals 0.5749.

Example 5.7

Consider Example 5.3. Given that it rained on Monday and Tuesday, what is the probability that it will rain on Thursday?

Solution: The two-step transition matrix is given by:

$$P^{(2)} = \mathbf{P}^2 = \begin{Vmatrix} 0.7 & 0 & 0.3 & 0 \\ 0.5 & 0 & 0.5 & 0 \\ 0 & 0.5 & 0 & 0.6 \\ 0 & 0.2 & 0 & 0.8 \end{Vmatrix} \cdot \begin{Vmatrix} 0.7 & 0 & 0.3 & 0 \\ 0.5 & 0 & 0.5 & 0 \\ 0 & 0.5 & 0 & 0.6 \\ 0 & 0.2 & 0 & 0.8 \end{Vmatrix}$$

$$= \begin{Vmatrix} 0.49 & 0.12 & 0.21 & 0.18 \\ 0.35 & 0.20 & 0.15 & 0.30 \\ 0.20 & 0.12 & 0.20 & 0.48 \\ 0.10 & 0.16 & 0.10 & 0.64 \end{Vmatrix}$$

Since rain on Thursday is equivalent to the process being in either state 0 or state 1 on Thursday, the desired probability is given by $P_{00}^{2} + P_{01}^{2} = 0.49 + 0.12 = 0.61$.

Forward Equation $P_{ij}^{n+1} = \sum_{k=0}^{\infty} P_{ik}^{n} P_{kj}$ for all $n \geq 0$ all i, j.

Backward Equation $P_{ij}^{n+1} = \sum_{k=0}^{\infty} P_{ik} P_{kj}^{n}$ for all $n \geq 0$ all i, **j.**

Check Your Progress 1

1) What do you mean by a stochastic process?

....

....

....

....

2) Given an example of a sample path.

....

....

....

....

3) Define the terms:
i) stationary increments, ii) Markov process, and iii) Martingale

....

....

....

....

4) How do you present a homogenous Markov Chain?

....

....

....

....

5) Write the meaning of transition probability.

....

....

....

....

3.6 GEOMETRIC BROWNIAN MOTION

Consider random variable that follows a geometric Brownian motion, The stochastic equation for its variation with the time t is given by

$$dV = \alpha V\, dt + \sigma V dz$$

where

$dz =$ Wiener increment $= \varepsilon dt^{1/2}$ (where ε is the standard normal distribution);

α is the drift; and σ is the volatility of V.

In the above equation the first term of the right side gives the expectation (trend) term and the second term the variation (deviation from the tendency or term of uncertainty).

With the above specification, Wiener process leads to changes with jumps in the stochastic variable V. This happens due to the fact that for a small time interval At, the standard deviation of the movement will be much larger than that of the mean. For small A ,) is much larger than At, which determines the behaviour of sample paths of a Wiener process.

A Wiener process does not have time derivative in a conventional sense and

A $= \varepsilon($ A) becomes infinite as *At* approaches zero.

Consider the case of options (see Unit 4 for meaning). There is a dividend like income stream δ for the' holder of the asset. The yield is related to the cash flows generated by the assets.

In that case the equilibrium requires that the total expected return μ be the sum of expected capital gain plus the expected dividend. That is,

$$\mu = \alpha + \delta$$

So that we write, the stochastic as

$$dV = (\mu - \delta) V\, dt + \sigma V\, dz$$

3.6.1 Arithmetic Brownian Model for the Logarithm of the Prices

Following Dixit(1993) if $dV/V = \boldsymbol{a}\, \boldsymbol{dt} + \mathrm{a}dz$, letting, and using $v = In(V)$, **we** find that v follows the arithmetic (or ordinary) Brownian motion:

$$dv = d(InV) = (\mathrm{a} - \tfrac{1}{2}\sigma^2)\, dt + \sigma\, dz$$

So, $\quad dv = \alpha'\, dt + \sigma\, dz$

The logarithm of $V = \mathrm{v}$ follows an Arithmetic Brownian Motion with drift a' and volatility σ.

Although the volatility term is the same of the geometric Brownian for V, as highlighted by Dixit, $d(InV)$ is different from $\mathrm{d}V/V$ due the drift. In reality, by the Jensen's inequality, $\mathrm{d}(InV) < dV/V$ (Itô's effect) .

3.7 MEAN REVERSION MODELS

For commodities and interest rates (and perhaps for exchange rates) mean-reversion model has more economic logic than the geometric Brownian model presented before.

Suppose the price P of a commodity follows a geometric mean-reverting process:

$$dP = \eta P (M - P)\, dt + \sigma P\, dz$$

where M is the long-run equilibrium level (or the long-run mean price which the prices tend to revert); and η is the speed of reversion. The others terms have the same meaning of the geometric Brownian motion (GBM) case, presented before.

The difference between a mean-reverting process and the GBM is the drift term: The drift is positive if the current price level P is lower than the equilibrium level M , and negative if $P > M$.

3.7.1 Arithmetic Ornstein-Uhlenbeck Process.

Sometimes we prefer to work with the arithmetic process for the logarithm of the stochastic variable, mainly for simulations and parameters estimation, due its simplicity.

Let $x = In(P)$. Consider that x follows the' arithmetic Ornstein-Uhlenbeck process and approach an equilibrium level m ,i.e.,

$$dx = \eta(m - x)\,dt + \sigma\,dz$$

The variable x has normal distribution, the expected value of x and the variance are shown below. The expected value equation is:

$$E[x(t)] = m + (x(0) - m)\exp(-\eta t) = x(0)\exp(-\eta t) + m(1 - \exp(-\eta t))$$

Look at the left side of the equation. The expect value is an intermediate point between $x(0)$ and the mean m , weighted by a decay rate. The weights sum to one.

There are two ways to find the'above equation. First, through the ***stochastic integral format for the arithmetic Ornstein-Uhlenbeck process*** and taking expectations (and the It8 integral goes to zero (see Appendix given at the end for the concept)),

$$x(T) = x(0)e^{-\eta T} + (1 - e^{-\eta T})m + \sigma e^{-\eta T}\int_0^T e^{\eta t}dz(t)$$

The second way is to get the same equation for the expected value. We simply integrate directly the deterministic term of the arithmetic Ornstein-Uhlenbeck process $dx/(m - x) = \eta\,dt$.

The variance of the normal distribution of the variable x at the instant t is:

$$VAR[x_t] = [1 - \exp(-2\eta t)]\sigma^2/2\eta$$

See that there is a time decay term for the variance. For long time horizon the variance of this process tends to $\sigma^2/2\eta$. So, unlike the Brownian motion the variance is bounded does not grow to infinity.

3.7.2 Mean Reversion with Jumps Models

For interest rates, and exchange rates, we use the stochastic process presented by a Poisson-Gaussian model of mean-reversion with jumps.

This model has more economic logic and considers large variation of the underlying variable due the arrival of abnormal rare news, The smooth variation is modelled by a mean-reversion process (continuous process), whereas the jumps are modelled with a Poisson process (discrete time process).

The **Poisson process** is **a** counting process in which inter-arrival times of successive jumps are independently and identically distributed (i.i.d.) exponential random variables.

Suppose that the price P follows a geometric mean-reverting with jumps process: $dP = \eta P(M - P)\,dt + \sigma P\,dz + P\,dq$

where M is the long-run equilibrium level (or the long-run mean price which the prices tend to revert); η is the speed of reversion; dq is the Poisson (jump) term; and the other terms have the same meaning as presented before.

The Poisson term dq values zero most of the time, but with frequency λ, it assume a value causing a jump on the underlying variable.

$dq = 0$......... with probability $1-\lambda dt$

$dq = \Phi - 1$... ... ... with probability $\lambda\,dt$

The usual and practical assumption is that the Wiener (dz) and Poisson (dq) processes are not correlated.

The jump size/direction can be random. In the above equation Φ is the jump size probability distribution. Given a jump, the expected jump size is $k = E[\Phi - 1]$.

A model with practical advantages, sometimes used is the compensated Poisson version for geometric mean-reversion with jumps process given as:

$$dP/P = \left[\eta(M - P) - \lambda k\right]dt + \sigma\,dz + dq$$

There is an additional term in the drift, $-\lambda k$, the compensation term.

The **compensated Poisson process** allows that the expected value of dP/P to be independent of the jump parameters.

A Poisson process N_t is a counting process, so increases with the time and cannot be **a** martingale, but a compensated Poisson process $(N^*_t = N_t - \lambda t)$ will be a martingale.

In our Poisson process case, $E(dq) = k\lambda\,dt$, so we get the compensation term above.

3.7.3 Arithmetic Jump-Reversion Model and Probabilistic Moments

This model has some facilities for simulation and for parameters estimation,. due to its simplicity.

Suppose the price is P and let $x = In(P)$. Consider that x follows the arithmetic mean-reversion with jumps:

$$dx = \left[In(P)\right] = \left[\eta\left[m - In(P)\right] - \lambda k\right]dt + \sigma dz + dq$$

where $m = In(M)$ is the long-run mean of the logarithm of the prices,

The expected value of the logarithm of price in this model can be written as:

$$E[In(P)] = In(M) + \{[In(P_0) - In(M)]\exp(-[In(2)t]/H)\}$$

where the half-life $H = [In(2)]/\eta$

Now suppose the following arithmetic jump-reversion, without the compensation term:

$$dx = d[In(P)] = \eta[m - In(P)]dt + \sigma\, dt + \Phi\, dq$$

where Φ is distribution of jump sizes.

The mean jump size is $k = E[\Phi]$.

The Poisson variable dq values 0 most of time and 1 sometimes(= 1 with probability λ dt).

The first probabilistic moment, the mean of this process, is:

$$E(x) = E[Ln(P)] = (m + (\lambda k H/In(2)))$$

$$\times(1 - \exp(-In(2)t/H)) + [In(P_0)\exp(-In(2)t/H)]$$

if the average jump is zero, that is $k = 0$ using this result in the equation above you get the same result of the compensated Poisson case.

In order to find the variance of the process is necessary to solve a PDE (partial differential equation) for the probability density function (Kolmogorov equation),

The variance of the distribution for x, is its second moment (M,) less the square of the first moment, that is, $Var(x) = M_2 - [E(x)]^2$ In other words, the variance is given by:

$$VAR[In(P)] = \frac{\sigma^2 + \lambda.E[\phi^2]}{2.In2/H}\left(1 - \exp\left(-\frac{2.In2}{H}t\right)\right)$$

Note that:

$$E[\phi^2] \neq E[\phi]^2$$

In order to find out its expected value, it is necessary to solve numerically the following integral:

$$E[\phi^2] = \int \phi^2.f(\phi)\,d\phi$$

where ϕ is the jump size and $f(\phi)$ is probability distribution of the jump size.

The higher probabilistic moments from the distributions generated by jump-diffusion processes define characteristics like the typical fat tails of the distribution, Unlike the normal distribution, the distribution of the returns generated by the jump-diffusion process is asymmetric with fatter tails. Asset returns following jump-diffusion, generally presents a positive ***skewness.*** Recall that skewness is a measure of symmetry (skewness is 0 for symmetric distributions like the normal one) and positive skewness means distribution with long positive tails (skewed to the right).

Skewness is related to the third moment (M_3). **A** dimensionless measure is the ratio: $M_3 / \mathbf{a}^3$

The skewness for the jump-reversion process is the following equation:

$$\textit{Skewness} = \frac{\lambda . E\left[\phi^3\right] . H}{\sigma^3 \, 3 In(2)} \left(1 - \exp\left(-\frac{3 In(2)}{H} t\right)\right) + \frac{E\left[LnP\right]^3}{\sigma^3} - $$

$$+ 3E\left[LnP\right] \frac{\left\{\sigma^2 + \lambda . E\left[\phi^2\right]\right\} H}{\sigma^3 \, 2 In(2)} \left(1 - \exp\left(-\frac{2 In(2)}{H} t\right)\right)$$

Kurtosis shows how flat or peaked the distribution is. It is a measure of the distribution shape. Higher peaked is the distribution, higher is the kurtosis,

Kurtosis is related to the fourth moment (M_4).

A dimensionless Kurtosis coefficient is the ratio: M_4 / σ^4.

The normal distribution has a kurtosis coefficient $= 3$.

Check Your Progress 2

1) What are Lévy Processes?

....

....

....

....

....

2) Write the meaning of Point Process.

....

....

....

....

3) What do you mean by the mean Reversion Models?

....

....

....

....

4) How would you formulate the Ornstein-Uhlenbeck Process?

....

....

....

....

....

5) What is a geometric Brownian motion?

....

....

....

....

....

6) Which process is presented by a Poisson-Gaussian model?

....

....

....

....

....

7) What is a Bernoulli jump process?

....

....

....

....

....

8) Deseribe a compound Poisson process.

....

....

....

....

....

3.8 LET US SUM UP

In this unit we have presented the stochastic processes used in insurance modelling. A stochastic process is modelled for time dependent random phenomenon. Thus, it is a collection of random variable for each point of time considering an index set. We can distinguish between discrete time process and continuous time process. The sample path in a process is a realisation of the stochastic process for a particular value at a 'point of time. A process X is stationary if the distribution of X, is the same for all t and has stationary increments if the distribution of $(X_{t+s} - X_t)$ is the same as that of $(X_s - X_0)$ for all t. In a Markov Process, the future occurrence of an event is independent of the past event given the present event. A martingale process states the expected realisation of an event after $(n+1)$ steps to be equal to the realisation after the nth steps, no matter what may have revelation of information overtime which helps taking decision differently. Filtration can be used in that context. If we know that the event has occurred based on information up to and including the time of the event, then it is the stopping time. A Markov Chain tells us about the independence of future anything that has happened in the past, given its present value. Accordingly, we have transition probabilities of different states and space. The Chaporan–Kalmogorov equation provides a procedure to compute n-step transition probabilities. The increments, when seen in terms or underlying probability distributions help identity nature of stochastic process. Accordingly we have Lévy, Poisson, Markov and It6 processes sometimes having jumps or sometimes reverting to mean levels. One of the processes used for determination of stock price is based on a formulation of geometric Brownian motion.

3.9 KEYWORDS

Continuous-time Markov **Process: A** stochastic process $\{ X(t) : t \geq 0 \}$ that satisfies the Markov property and takes values from amongst the elements of a discrete set called the state space.

Compound Poisson **Distribution:** A probability distribution of a "Poisson-distributed number" of independent identically-distributed random variables.

Dominated Convergence **Theorem:** If a sequence $\{ f_n : n = 1, 2, 3, \ldots \}$ of real-valued measurable functions on a measure space S converges almost everywhere, and is "dominated" by some nonnegative function g in L^1. Then

$$\int_S \lim_{n\to\infty} f_n = \lim_{n\to\infty} \int_S f_n .$$

By stating that the sequence is "dominated" by g, it is meant

$$|f_n(x)| \leq g(x)$$

for every n and almost every x (i.e., the measure of the set of exceptional values of x is zero). By g in L^1, we mean

$$\int_S |g| < \infty .$$

So the theorem provides a sufficient condition under which integration and passing to the pointwise limit commutes.

Geometric Brownian motion (GBM) (or, **Exponential Brownian motion): A** continuous-time stochastic process in which the logarithm of the randomly varying quantity follows a Brownian motion, or a Wiener process. It is used to mathematical modelling of some phenomena in financial markets (like option pricing). **A** quantity that follows a GBM may take any value strictly greater than zero, and only the fractional changes of the random variate are significant. This is precisely the nature of a stock **price.**

Generalized Wiener Process: A continuous time random walk with drift and random jumps at every point in time.

Lévy Process: A continuous-time stochastic process that starts at 0 having "stationary independent increments".

Markov Chain: A discrete-time stochastic process with the Markov property. A Markov chain describes at successive times the states of a system. In all these times the system may have changed from the state it was in the moment before to another state, or it may have stayed in the same state. The changes of state are called transitions.

Markov Process: A stochastic process that has the Markov property. The Markov property if the conditional probability distribution of future states of the process, given the present state and all past states, depends only upon the current state and not on any past states, i.e. it is conditionally independent of the past states given the present state.

Martingale Pricing: A pricing approach based on the notions of martingale and risk neutrality, The martingale pricing approach is a cornerstone of modern quantitative finance and can be applied successfully to a variety of derivatives contracts, such as options and futures.

Martingale: A class of betting strategies. The simplest of these strategies was designed for a game in which the gambler wins her stake if a coin comes up heads and loses it if the coin comes up tails. The strategy had the gambler double his bet after every loss, so that the first win would recover all previous losses plus win a profit equal to the original stake. Since a gambler with infinite wealth with probability 1 eventually flips heads, the martingale betting strategy was seen as a sure thing by those who practiced it. However, such practitioners may not possess infinite wealth and the exponential growth of **the** bets would eventually bankrupt those who use the martingale over a long losing streak.

Non-Homogeneaus Poisson Process: is a Poisson process with rate parameter $\lambda(t)$ such that the rate parameter of the process is a function of time.

Ornstein-Uhlenbeck Process: (or, mean-reverting process): **A** stochastic process given by the following stochastic differential equation

$$dr_t = -\theta(r_t - \mu)\,dt + \sigma dW_t,$$

where θ, μ and σ are parameters and W_t denotes the Wiener process.

Partial Differential Equation (PDE): A relation involving an unknown function of several independent variables and its partial derivatives with respect to those variables

Poisson Process: A stochastic process which is defined in terms of the occurrences of events. This counting process, given as a function of time $N(t)$, represents the number of events since time $t = 0$.

Reflection Principle: Finding sets that resemble the class of all sets.

Risk-neutral Measure: a probability measure in which today's fair (i.e, arbitrage-free) price of a derivative is equal to the **discounted** expected value (under the measure) of the future payoff of the derivative.

Stationary Independent Increment: To call the increments of a process independent connotes that increments $X_s - X_t$ and $X_u - X_v$ are independent random variables whenever the two time intervals do not overlap. More generally, any finite number of increments assigned to pairwise non-overlapping time intervals are mutually (not just painvise) independent. To call the increments stationary means that the probability distribution of any increment $X_s - X_t$ depends only on the length $s - t$ of the time interval; increments with equally long time intervals are identically distributed. In the Wiener process, the probability distribution of $X_s - X_t$ is normal.

Stochastic Process: A random function in the most common applications. The domain over which the function is defined is a time interval or a region of space.

Stopping Time: With respect to a sequence of random variables X_1. X_2, ... a random variable τ with the property that for each t, the occurrence or non-occurrence of the event $\tau = t$ depends only on the values of $X_1, X_2, ..., X_t$. Thus, at any particular time t, you can look at the sequence so far and tell if it is time to stop.

Submartingale: A sequence $X_1, X_2, X_3, ...$ of random variables satisfying just

$$E[X_{n+1} \mid X_1, ..., X_n] \geq X_n .$$

Supermartingale: A sequence $X_1, X_2, X_3, ...$ of random variables satisfying

$$E[X_{n+1} \mid X_1, ..., X_n] \leq X_n .$$

(A mnemonic to remember which is which: "Life is a supermartingale; as time advances, expectation decreases.")

Wiener Process: A continuous-time stochastic process. It is often called Brbwnian motion. It is one of the best known Lévy processes with stationary independent increments.

3.10 SOME USEFUL BOOKS

Dixit, A. (1993), *The* Art of Smooth Pasting, Harwood Academic Publishers GmbH, Switzerland

Dixit, A.K. and R.S. Pindyck (1994), Investment under Uncertainty, Princeton University Press, Princeton, N.J.

Marco Antonio Guimarães Dias (2005), Stochastic Processes *with* Focus *in* ***Petroleum*** Applications, Department of Industrial Engineering, PUC-Rio (see internet)

Ross, Sheldon (1996), Stochastic Processes, 2nd ed., John Wiley and Sons, Inc., New York

3 ANSWER OR HINTS TO CHECK YOUR PROGRESS

Check Your Progress 1

I) See Section 3.2.and answer

2) See Sub-section 3.2.2 and answer
3) See Sections 3.2, 3.3 and 3.4 and answer
4) See Section 3.5 and answer
5) See Section 3.5.2 and answer

Check Your Progress 2

1) See Sub-section 3.3.1 and Section 3.5 and answer
2) See Sub-section **3.3.4** and answer
3) See Section 3.7 and answer
4) See Sub-section 3.7.1and answer
5) See Section 3.6 and answer
6) See Sub-section 3.3.6 and answer
7) See Sub-section 3.3.6 and answer
8) See Sub-section 3.3.5 and answer

UNIT 4 APPLICATION TO FINANCE

Structure

4.0 OBJECTIVES

After going through this unit, you will be able to:

- understand the basic concepts use'd in finance; and
- appreciate price determination through stochastic modelling.

4.1 INTRODUCTION

Operation of financial market provides an important example through which we can understand the use of stochastic modelling. If you want to participate in the stock market, the volatile movements of share prices will compel you to associate. their realisation with a stochastic process. Modelling used in finance will then be linked to insurance as source of input for decision making. Keeping this in view the following discussion introduces you to basics of financial derivatives. The coverage of the theme however is far from adequate to understand the ongoing work in the area of quantitative finance. You may like to refer Hull (1992) for some of the themes presented in the unit

4.2 FINANCIAL MARKET

The stock market is a primary capital market. Firms raise finance through stock markets by issuing new securities. It is also a secondary capital market where securities are traded, and signals the return shareholders receive which they provide to firms. Financial instruments can be divided in two basic classes: underlying and derivative. The former can be stocks, bonds or various trade goods, while the latter are financial contracts that promise some payment to the owner, depending on the behaviour of the " underlying" (e.g., a price at a given time T or an average price over certain time period). Derivatives are extremely useful for risk management (apart from obvious investment properties) – we can reduce our financial vulnerability by fixing price for a future transaction now.

A share is a type of financial underlying assets (or underlyings) dealt in financial market. Other underlyings include bonds, commodities and foreign currencies. In financial market not only underlying but also their derivatives are traded. A derivative is a financial instrument, which promises some payment or delivery in the future contingent on an underlying.

4.2.1 Some Important Words and Concepts

- Stock markets. Stock is a printed-paper issued by a company to attract investments. Usually dividends are paid to participants.
- **Bond** markets. "Bond: a printed paper given by a government, city or business company saying that money has been received and will be paid back, usually with interest," which is fixed.
- Currency **markets** or foreign exchange markets.
- Commodity **markets** such as oil, gold, copper, wheat and electricity.
- Futures and options markets. "Futures: goods and stocks bought at prices agreed upon at the time of the purchase but paid for and delivered afterwards."
- buyer = long position; buying = going long
- seller = short position; selling = going short
- **Short** selling involves selling an asset that: is not owned with the intention of buying it later.
- **Arbitrage** = sure profit without investment

Speculator: A trader who enters the futures market in pursuit of profit, accepting the risk.

Hedger: A trader who trades futures to reduce some preexisting risk exposure. They are often producers or major users of a given commodity (e.g., a farmer may hedge by selling his anticipated harvest even before the farmer plants). They often trade through a brokerage firm.

Arbitrageurs: Persons entering into several contracts in different markets to exploit price fluctuations. If a good had two prices, a trader can get an arbitrage

profit: a sure profit with no investment. But prices may differ because of transportation costs, etc.

Spreads: A trading strategy involves taking a position in two or more options of the same type,

portfolio = combination of several assets/securities, etc.

Types of derivatives

- **Options:** An option gives the holder the right (but not the obligation) to buy or sell an asset in the future at a price that is agreed upon today.
- **Futures: A** futures contract is an order that one places in advance to buy or sell an asset or commodity. The price is fixed when the order is placed but the payment is not made until the delivery date.
- **Forwards:** Futures contracts are standardized products bought and sold on organised exchanges. **A** forwards contract is a modified futures contract that is not traded on organized institution. For example, a multinational company which needs to protect itself against a change in the exchange rate usually buys or sells currency forwards through a bank.
- **Swaps:** Swaps are private agreements between two exchange cash flows in the future according to a prearranged formula. Suppose that you want to swap your sterling debt **for** US$ debt. In this case you can arrange for a bank to **pay** you 'each year the sterling's that are needed to service your sterling debt, and in exchange you agree to pay the bank the cost of servicing a US$ loan. Such an arrangements is known as a currency swap.

We continue with the derivative securities introduced above and find some formal definitions to help derive their prices.

4.3 FORWARD CONTRACT

This section is adopted from a discussion found in Internet. We are unable to determine the writer and acknowledge the contribution.

As already seen above, a forward contract is an exchange agreement to buy (or sell) an asset on a specified future date for specified price.

Forwards are not generally traded on exchanges and it costs nothing to enter into a forward contract. **A** future contract is the same as forward except that futures are normally traded on exchanges and the exchange specifies certain standard features of the contract.

Options are **of** different types. The most important of these being the American and European options. An option is so-called because it gives the holder the right, but not the obligation, to under take a transaction.

Definition. A call option gives the holder the right to buy whereas a put option gives the holder the right to sell.

A European call option gives the holder the right, but not the obligation, to **buy** an asset on a certain date for a certain price (the strike price).

An American option is similar except that the holder of an American call has the right to buy the asset for the specified price at any time up until the expiration date. It can be optional to exercise such an option prior to expiration.

Two main uses of options are speculation and hedging. In case of speculation, available funds are invested opportunistically with the aim of making a profit. Hedging, on the other hand, is usually engaged in by companies who have to deal with risky assets such as foreign exchange, gold, oil and so on. For hedgers, the basic purpose of an option is to spread risk.

Suppose that you know that in three months time you need a million gallons of crude oil. However you are worried by the fluctuations in the price of oil. So you decide to buy European call options. See that you know the maximum amount of money needed to buy oil, Denote the price of the underlying asset (oil) at time T when the option expires in three months time by S_T. Let us assume that the strike price is K. If, at time T, $S_T > K$ then you will exercise the option. Such situation is called as the option is in the money. You are buying an asset worth $S_T > K$ dollars for just K dollars. If on the other hand, if $S_T <$ K, then you will buy oil on the open market ignoring the option, which is worthless. Such situation is called as the option is out of the money. The payoff from the option is, thus, $(\max(S_T, K) - K)$ or $S_T - K)_+$

From the above example, it is clear that you have to decide on how much you should be willing to pay for such an option. You can think of it as insurance premium, where you are insuring against the oil price going up.

Let us see how option market operates.

4.3.1 Pricing a Forward

We' have seen above that a forward contract is an agreement to buy (or sell) an asset on a specified future date for a specified price. The pricing problem here is, 'what price for the asset should be specified in the contract?'

Example, Suppose that you enter into a long position buying on a forward contract and agree to buy an asset for price K at time T. Then the payoff at time T is $(S_T - K)$, where S_T is the asset price at time T, since you (must) buy the asset S_T for price K. This pay off could be positive or negative. Since the cost of entering into a forward contract is zero this is also your total gain (or loss) from the contract. What is a fair price of K?

At the time when the contract is written, we don't know S_T. Therefore, we can only guess at it. Typically, we assign a probability distribution to it,

Note that stock prices should stay positive, so a, b are positive. We may think that $\mathbb{E}[S_T]$ would represent a fair price to write into our contract. However, the market price will match the expected price only rarely.

We may assume that market participants can borrow money for the same risk-free rate of interest as they do to lend money. Thus, if you borrow \$1 now, your debt at time T it will be worth be $\$e^r$(say) and similarly if you keep \$1 in the bank then at time T it will be $\$e^r$.

I) Let $K > S_0e^r$. Then the other participant in the deal can adopt the following ctrategy: she borrows $\$S_0$ at time zero and buys with it one unit of stock. At time T, she must repay $\$S_0e^r$ to the bank, but she has the stock to sell to you for $\$K$, leaving her a certain profit of $\$(S_0e^r -$ K).

2) If $K < S_0e^r$, then you can reverse this strategy. Just sell a unit of stock at time zero for $\$S_0$ and put the money in the bank. At time T, your accumulation $\$S_0e^r$ and use $\$K$ to buy a unit of stock. which leaves you with a certain profit of $\$(S_0e^r - K)$.

Unless $K = S_0e^r$, dne of you will be guaranteed to make a profit.

The above argument, selling of stock by you, which may not be yours, is lcnown as short selling. An opportunity to lock into a risk free profit is called an arbitrage.

The starting point in establishing a model in modern finance theory is to specify that there are no arbitrage opportunities. We find that there are people who make their living entirely from exploiting arbitrage opportunities, but such opportunities do not exist for a significant length of tirne before market prices move to eliminate these.

Remember that forwards are a very special sort of derivative. It will no be easy for us to value an option. However. we must look for conditions that don't provide either party with an arbitrage opportunity. Such conditions will be identified in the following.

4.3.2 Discrete Time Models I

The Single Period Binary Model

In the discussion below we try to find the 'fair' price for a European call option. Let us take an example.

Suppose that the current price in Swiss Francs (SFR) of \$100 is $S_0 = 100$. Consider a European call option with strike price $K = 150$ at time T. The fair price to be paid for this option may be obtained as follows:

We know that an option gives the buyer the right to buy \$1 00 for 150SFR at time T. Hence suppose that the true exchange rate at time T is not known. However it can be modelled by a random variable. The simplest model is the single period binary model where S_T takes one of two values with specified probabilities. Let's if

$$S_T = \begin{cases} 180 & with \quad probability \quad P \\ 90 & with \quad probability \quad 1-p \end{cases}$$

Then the payoff of the option will be (180-150) SFR=30/SFR. With probability p and 0 with probability 1-p. Therefore has expectation 30p SFR.

To find the fair price to be paid for the option, let us assume that interest rates are zero and that currency is bought and sold at the same exchange rate. Remember that in the presence of a fair price, there will be no scope of a risk free profit for either the buyer or the seller of the option. In the example above, wc try to arrive at a specific price by supposing that $p = 0.5$. That is S_T taking the value 90SFR, the fair price to pay for the option.

You claim that if the price is 15SFR, then you can make a risk-free profit by the following strategy: you buy the option and you borrow \$33.33 and convert it straight into 50 SFR. (You will have to pay off your loan in dollars at time T.)

Your position at time zero is as follows: you have one option (to buy \$100 for 150SFR at time T), you have 35SFR (50 from the conversion of your dollar loan, less 15 paid for the option) and you have a debt of \$33.33.

At time T, one of two things has happened:

1) If $S_T = 180$, then exercise the option and buy \$100 for 150SFR. Make use \$33.33 to pay off your dollar debt and that leaves a balance of \$66.67. Convert back into SFR' at the current exchange rate. This nets $2/3 \times 1800 = 120SFR$. In total then, if $S_T = 180$, at time T you have 35SFR-150SFR (used to exercise the option)+120SFR, which totals 5SFR clear profit.

2) If $S_T = 90$, then you have to throw away the option (which is worthless) and convert your 35SFR into dollars, netting $\$0.9 \times 35 = \$38.89$. You pay off your debt leaving a profit of \$5.56.

Whatever the true exchange rate at time T, you make a profit.

Determination **of Right Price**

Let's think of things from the point of view of the seller. If you are the seller of the option, then you know that at time T, you will need $\$(S_T - 150)_+$ in order to meet the claim against **me.** The idea is to calculate how much money you need at time zero (to be held in a combination of dollars and Swiss Francs) to guarantee this.

Suppose you hold x_1 and x_2 at time zero. You need this holding to be worth st least $(S_T - 150)_+$ SFR at time T.

1) If $S_T = 180$, then you will need at least 30SFR. That is, we must have

$$x_1 + \frac{180}{100} x_2 \geq 30.$$

2) On the other hand, if $S_T = 90$ then the payoff of the option is zero and you just need not to be out of pocket. That is, you want

$$x_1 + \frac{90}{100} x_2 \geq 0.$$

Solve the simultaneous equation to get the right price. On such a, the seller can exactly meet the claim if $x_1 = -30$ and $x_2 = 300/9$. To purchase \$300/9 at time zero is equal to 50-30=20SFR. Thus, the seller requires exactly 20SFR at time zero to construct a portfolio that will be worth exactly the payoff of the option at time T.

The argument above can be reversed to show that for any lower price, there is a strategy for which the buyer makes a risk-free profit.

The fair price is 20SFR.

Notice that we did not use the probability, p, of the price S, going up to 180SFR at any point in the calculation. We just needed the fact that we could replicate at **any point** in the calculation. We just needed the fact that we could replicate the

claim by this simple portfolio. The seller can hedge the contingent claim $((S_t - 150)+$-usingthe portfolio consisting of $x_1 SFR$ and $\$x_2$

A Characterisation of 'no arbitrage'

The option pricing problem solved above through binary framework in a single period may not help in more complex settings like multi-period situations. Hence we need to take the help of probability theory. We will do that after providing a concise mathematical condition to characterize markets that have no arbitrage opportunities. Let there be a market, which consist of N tradable assets. Their prices at time zero are given by the column vector

$$S_0 = \left(S_0^1, S_0^2, \ldots\ldots, S_0^N\right)^T.$$

Let the market be represented by a finite number of possible states (1,2,..n) in which it might be at time one. The security values at time one are given by an $N \times n$ matrix $D = \left(D_{ij}\right)$, where the coefficient D_{ij} is the value of the i^{th} security at time one if the market is in state j.

In this formulation, you can think of a portfolio as a vector

$$\theta = \left(\theta_1, \theta_2, \ldots, \theta n\right)^T \in \mathbb{R}^N$$

and its market value at time zero is the scalar product

$$S_0.\theta = S_0^1\theta_1 + S_0^2\theta_2 + \ldots + S_0^N \theta N.$$

The payoff of the portfolio at time one is a vector in $\mathbb{R}^n$ whose i^{th} entry is the value of the portfolio if the market is in state i. Thus, the payoff is

$$\begin{pmatrix} D_{11}\theta_1 + D_{21}\theta_2 + \ldots\ldots D_{N1}\theta_N \\ D_{12}\theta_1 + D_{22}\theta_2 + \ldots\ldots D_{N2}\theta_N \\ D_{1n}\theta_1 + D_{2n}\theta_2 + \ldots\ldots D_{Nn}\theta_N \end{pmatrix} = D^T\theta$$

Notation

Foravector $x \in \mathbb{R}^n$ we write $x \geq 0$ to mean $x \in \mathbb{R}_+^n$ and $x > 0$ to mean $x \geq 0, x \neq 0$. Notice that $x > 0$ does not require x to be strictly positive in all its coordinates. We write $x \gg 0$ for vectors, which are strictly positive in all coordinates, i.e., for vectors $x \in \mathbb{R}_{++}^n$. An arbitrage is then a portfolio $\theta \in \mathbb{R}^N$ with

$$\begin{aligned} &S_0.\theta \leq 0, D^T\theta > 0 \\ &or \\ &S_0.\theta \leq 0, D^T\theta > 0 \end{aligned}$$

Definition. A state price vector is a vector $\Psi \in \mathbb{R}_{++}^n$ such that $S_0 = D\Psi$. **That is,**

$$\begin{pmatrix} S_0^1 \\ S_0^2 \\ \vdots \\ S_0^N \end{pmatrix} = \Psi_1 \begin{pmatrix} D_{11} \\ D_{21} \\ \vdots \\ D_{N1} \end{pmatrix} + \Psi_2 \begin{pmatrix} D_{12} \\ D_{22} \\ \vdots \\ D_{N2} \end{pmatrix} + \ldots\ldots + \Psi_n \begin{pmatrix} D_{1n} \\ D_{2n} \\ \vdots \\ D_{Nn} \end{pmatrix}$$

The vector multiplied by Ψ_i is the security price vector if the market is in state i. To interpret Ψ_i we can think it to be the marginal cost of obtaining an additional unit of wealth at the end of the time period if the system is in state i. There is no arbitrage if and only if there is a state price vector.

The Risk Neutral Probability Measure

Recall that all the entries of Yare strictly positive. Therefore, we write

$\Psi_0 = \sum_{i=1}^{n} \Psi_i$ such that

$$\Psi = \left(\frac{\Psi_1}{\Psi_0}, \frac{\Psi_2}{\Psi_0}, \ldots\ldots, \frac{\Psi_n}{\Psi_0}\right)^T \qquad (1)$$

is taken as a vector of probabilities for being in different states. First we will see that Ψ_0 is the discount on riskless borrowing. Suppose that the market allows positive riskless borrowing, and for some portfolio, 8,

$$D^T\bar{\theta} = \begin{pmatrix} 1 \\ 1 \\ \vdots \\ \vdots \\ 1 \end{pmatrix},$$

i.e., the value of the portfolio at time T is one, irrespective of the state of the market. Since Ψ is a state price vector, the cost of such a portfolio at time zero is

$$S_0.\bar{\theta} = (D\Psi).\bar{\theta} = \Psi.(D^T\bar{\theta}) = \sum_{i=1}^{n} \Psi_i = \Psi_0$$

where Ψ_0=discount on riskless borrowing, From the vector (I), the expected value of the i^{th} security at time T is

$$E\left[S_T^i\right] = \sum_{j=1}^{n} D_{ij}\frac{\Psi_j}{\Psi_0} = \frac{1}{\Psi_0}\sum_{j=1}^{n} D_{ij}\Psi_j = \frac{1}{\Psi_0}.S_0^i$$

where the last equality is obtained by taking $S_0 = D\Psi$ That is,

$$S_0^i = \Psi_0 E\left[S_T^i\right], i = 1,\ldots.,n \qquad (2)$$

So, under the probabilities distribution (1), the price of a security is its discounted expected payoff. In the context of contingent claim pricing, we can say that a claim is attainable if it can be hedged.

If there were no arbitrage, the unique time zero price of an attainable claim Cat time T is $\Psi_0\mathbb{E}[C]$ where the expectation is with respect to any probability measure for which $S_0^i = \Psi_0\mathbb{E}\left[S_T^i\right]$ for all iand Ψ_0 is the discount on riskless borrowing rate.

Note that the same value is obtained if the expectation is calculated for any vector of probability such that $S_0^i = \Psi_0\mathbb{E}\left[S_T^i\right]$. In the absence of arbitrage, there is only one riskless borrowing rate.

Moreover, if you find a probability vector for which the present value of each security now is its discounted expected value at time T, then it is possible to find the time zero value of any attainable contingent claim by calculating the expectation. Thus, if there is a portfolio Q, for which we get $Q.S_t = C$, than

$$Q.S_0 = Q.(\psi E[S_T]) = \psi_0 \sum_{i=1}^{n} Q_i E\left[S_T^i\right] = \psi_0 E[Q.S_T]$$

Example. Let us return to the problem given in the previous example.

Recall the conditions in that example: the current price in Swiss Francs (SFR) of \$100 is $S_0 = 150$. We suppose that at time T, the price will have moved to either 90SFR or 180 SFR. Therefore, the problem was to price a European call option with strike price $K = 150$ at time T i.e., holder of such an option has the right, but not the obligation, to buy \$100 for 150SFR at time T. For simplicity, interest rates were obligation to be zero.

The fair price for such an option was worked out to be is 20SFR, with the new approach.

Under our assumption of zero interest rates, we have $\Psi_0 = 1$. That means, we are seeking a probability vector such that

$$\mathbb{E}[S_1] = S_0.$$

Let p be the probability that

$$S_t = 180SFR.$$

Then since S_t can only take values 180 and 90, it must be that

$$\mathbb{P}[S_t = 90] = 1 - p.$$

So, for the price to behave like a fair game,

$$180p + 90(1-p) = 150, \text{ or, } p=2/3.$$

As the claim at time T is $C = (S_t - 150)_+$, the fair price (in SFR) is

$$\mathbb{E}[(S_T - 150)_+] = \frac{2}{3}(180-150)_+ + \frac{1}{3}(90-150)_+ = 20,$$

as before such that given the probability p, it is easy to value other options. For example, a European call with strike price 120SFR instead of 150SFR would be valued at

$$\mathbb{E}[(S_T - 120)_+] = \frac{2}{3}(180-120)_+ + \frac{1}{3}(90-120)_+ = 40$$

For your information, it may be useful to note that the probability measure that assigns probability 2/3 to $S_T = 180SFR$ and 1/3 to $S_T = 180SFR$ is called **equivalent** martingale probabilities for the (discounted) price process $\{S_0, \Psi_0 S_T\}$. The probabilities are also called the risk' neutral probabilities.

4.3.3 The Multi-Period Binary Model

Let the market consists of just two securities: a bond (representing riskless borrowing) and a stock S. Unlimited amounts of either can be bought and sold without transaction costs. Moreover we are allowed to readjust our portfolio instantaneously.

As in the single period binary model, we shall suppose that at each tick of the clock, the stock moves form its current value to one of two possible values (depending on its current value). There are 2^i possible states of the stock price after i ticks of the clock, and we think of them as being arranged in tree as in Figure 1.

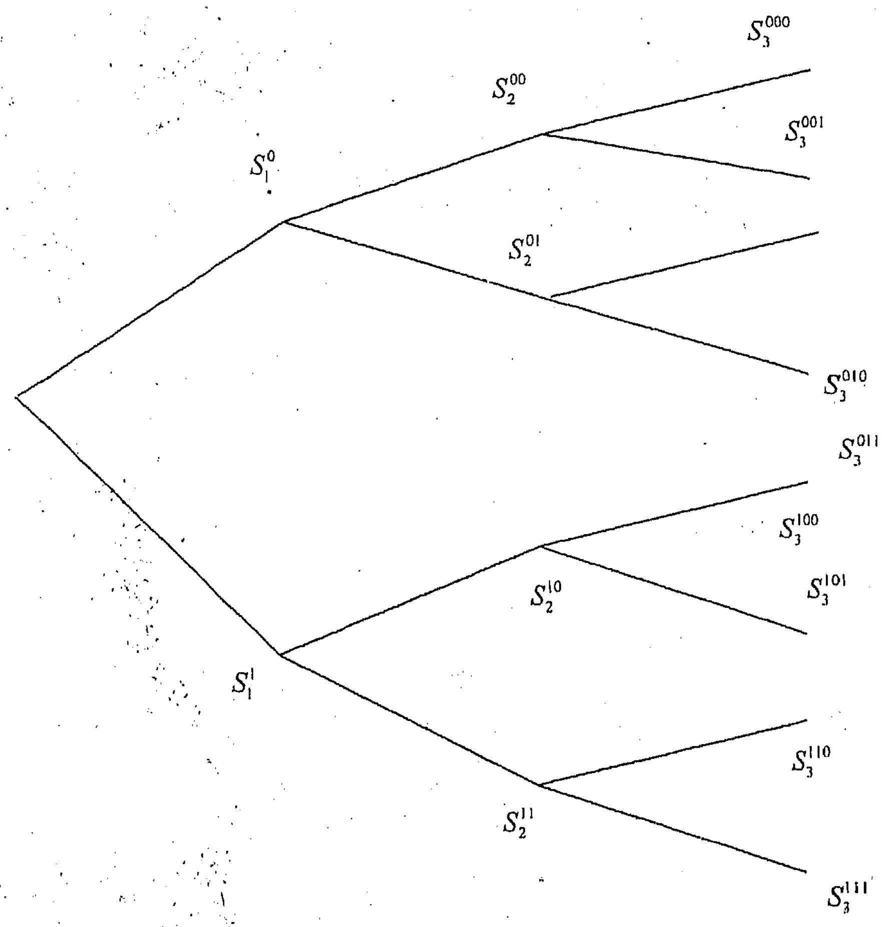


Figure 1

The bond moves as before, so over the i^{th} tick its value is scaled up by $\exp(r_i)$. We assume that all r_i are known.

Suppose again that you are pricing a European option with strike price K at the maturity time, T. We assume now that $T = n$. The payoff of the option is then $(S_n - K)^+$ at time n.

Let us use backward induction on the tree. If we knew the price, S_{n-1} of the stock after $(n-1)$ ticks, then our previous analysis would tell us the value, C_{n-1}, of the claim at time $(n-1)$. Namely,

$$C_{n-1} = \Psi_0^{(n)} \mathbb{E}_{n-1}[C_n]$$

where the expectation is with respect to a probability measure for which

$$S_{n-1} = \Psi_0^{(n)} E_{n-1}[S_n] \text{ and } \Psi_0^{(n)} = e^{-r_n}.$$

So for each state of the market at time $(n-1)$, you know that you need a portfolio worth C_{n-1} if you are to meet the claim against you at time n. You can **now** think of C_{n-1} as a claim at time $(n-1)$. In the same way then, if you know S_{n-2}, in order to meet the claim against me at time $(n-1)$, you need to hold a portfolio worth

$$\Psi_0^{(n-1)} E_{n-2}\left[C_{n-1}\right]$$

where the expectation is with respect to a measure such that

$$S_{n-2} = \Psi_0^{(n-1)} E_{n-2}\left[S_{n-1}\right],$$

and this in turn guarantees that you can exactly meet the claim against you at time n. Proceed following this method. You will able to calculate the cost of a' portfolio. That is, after appropriate readjustment at each tick of the clock, but neither with extra input of wealth not with paying dividends, you will meet the claim against you at time n.

The strategy that readjusts the portfolio in this way is said to be a self-financing strategy. The probability measures used at each stage in the above prescribe exactly one probability for each branch, in our tree of stock prices. For each vertex. of the tree there is a unique path from the vertex, and we specify a probability measure on paths by declaring that the probability of such a path is the product of the probabilities on the branches that comprise it.

Let us assume, for simplicity, that the rate of interest is everywhere zero. It is equivalent to replacing S_i by the discounted security price

$$\widetilde{S}_i = \prod_{j=1}^{i} \Psi_0^{(j)} S_i = e^{-\sum_1^i r_i} S_i.$$

Our risk neutral probabilities then have the property the

$$\mathbb{E}\left[S_k \mid S_{k-1}\right] = S_{k-1} \text{ for each } k = 1, 2, \ldots n.$$

In fact much more is true. To illustrate, consider the two-step model in Figure 2

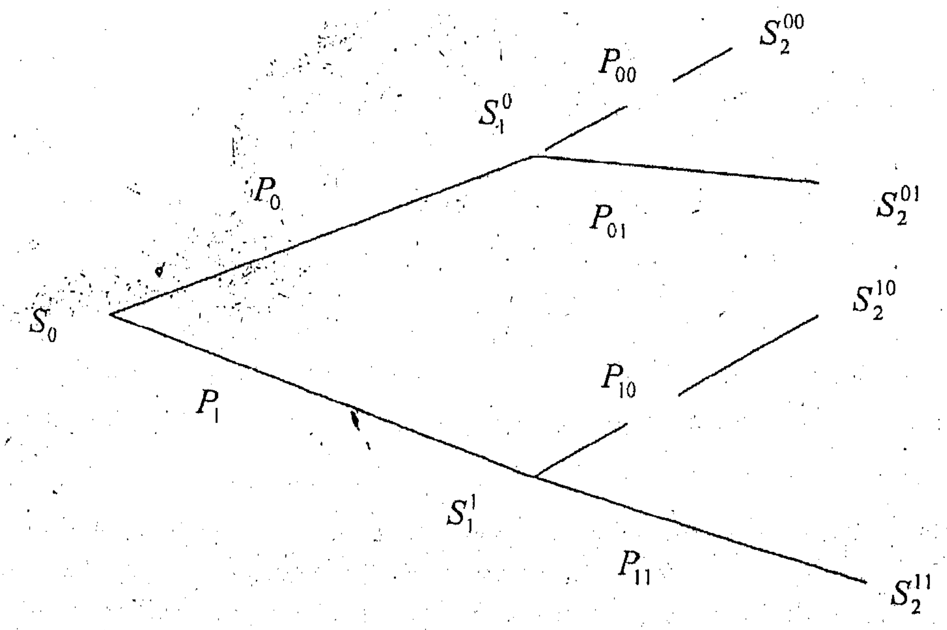


Figure 2

$\mathbf{E}[S_2[$

$$
\begin{aligned}
&= p_0 p_{00} S_2^{00} + p_0 P_{01} S_2^{01} + p_1 p_{10} S_2^{10} + p_1 p_{11} S_2^{11} \\
&= p_0 \left(p_{00} S_2^{00} + p_{01} S_2^{01} \right) + p_1 \left(p_{10} S_2^{10} + p_{11} S_2^{11} \right) \\
&= p_0 \mathbf{E}\left[S_2 | S_1 = S_1^0 \right] + p_1 \mathbf{E}\left[S_2 | S_1 = S_1^1 \right] \\
&= \mathbf{E}\left[\mathbf{E}\left[S_2 | S_1 \right] \right] \\
&= \mathbf{E}[S_1] = S_0.
\end{aligned}
$$

. More generally, if $j > i$,

$$\mathbb{E}\left[S_j | S_i \right] = S_i.$$

Similarly,

$$C_{n-i} = \mathbf{E}\left[C_n | S_{n-i} \right],$$

where the expectation is with respect to the probabilities on paths defined above.

The sequence of (discounted) prices, $\{S_i\}_{i=0}^{n}$, is **a** stochastic **process** and the probability measure that we have defined has the property that $\mathbb{E}\left[S_j | S_i \right] = S_i$. Moreover, in **this** model, the stqck price 'has no memory,' so that the movement of the stock over the next tick of the clock is not influenced by the **way in which** the stock reached its current value and so for $j > i$,

$$\mathbb{E}\left[S_j | S_0, S_1, \ldots S_i \right] = S_i.$$

A sequence of random variables (or stochastic process)

$X_0, X_1, \ldots, X_n$ with $\mathbf{E}\left[|X_r| \right] < \infty$ for each r is martingale if

$$\mathbf{E}\left[X_j | X_0, X_1, \ldots X_{j-1} \right] = X_{j-1} \qquad (3)$$

$$r = 1, 2, \ldots, n.$$

We have seen in Units **2-3** that the idea comes from gambling where. if X_r denotes the capital of the gambler at time r then game is 'fair' only if (3) holds.

The 'information' X_0, X_1, X_{r-1} is often written $\mathcal{F}_{r-1}$. In a continuous setting **we will** be **a little more** careful about the definition, but the idea is the same, $\mathcal{F}_{r-1}$ is the set of events that are 'decidable' by observing the process X, up to time $r-1$.

We now recast the results of the above model in this language.

Suppose that the possible values that the stocks $S_1, \ldots, S_N$ can take on at times 1,2,3.., are known. We denote by Ω the set of all possible 'paths' that the stock price vector can follow in $\mathbb{R}_+^N$. Then **the** absence of arbitrage is equivalent to the existence of a probability measure, $\mathbb{Q}$ and Ω that assigns strictly positive mass to every $\omega \in \Omega$ and

$$S_{r-1} = \Psi_0^{(r)} \mathbb{E}_{\mathbb{Q}}\left[S_r | S_{r-1}\right],$$

where S_r is the vector of stock prices at time r.

If, as above, we consider the discounted stock prices, then

$$\mathbb{E}_{\mathbb{Q}}\left[\tilde{S}_r | \tilde{S}_1, \ldots\ldots, \tilde{S}_{r-1}\right] = \tilde{S}_{r-1}.$$

In other words, the discounted stock price vector is a $\mathbb{Q}$-martingale.

Recalling that two probabilities measure $\mathbb{P}$ and $\mathbb{Q}$ on a space Ω are said to be equivalent if for events $A \subseteq \Omega$

$\mathbb{Q}(A) = 0$ if and only if $\mathbb{P}(A) = 0$ we may observe the following:

Suppose then that we have a market model in which the stock price vector can follow on of a finite number of paths Ω through $\mathbb{R}_+^n$. We may even have our own belief as to how the price will evolve, encoded in a probability measure, $\mathbb{P}$, on Ω. We can say, there is no arbitrage if and only if there is an equivalent martingale measure $\mathbb{Q}$. That is, there is a measure, $\mathbb{Q}$, equivalent to $\mathbb{P}$, such that the discounted price process is a $\mathbb{Q}$-martingale.

In that case, the market price of an attainable claim C to be delivered at time n at time zero is unique and is given by

$$\mathbb{E}_{\mathbb{Q}}\left[\Psi_0 C\right],$$

where

$$\Psi = \prod_1^n \Psi_0^{(i)}$$

is the discount factor over n periods.

The same statement applies in the continuous setting.

The Cox Ross Rubinstein Model

The Cox Ross Rubinstein (CRR) model is a special case of the multi-period binary model in which in each time interval the stock price moves from its current value. S, to one of S_u, S_d, where u and d are fixed constants with $d < e^{r\Delta T} < u$.

It is referred to as the binomial model. Thus, at time k, there are k possible values that the price can take and

$$\mathbb{P}\left[S_K = S_0 u^j d^{k-j}\right] = \binom{k}{j} p^j (1-p)^{k-j}$$

The recombining of the tree makes this highly numerically efficient.

The values of u and d must be calibrated with the market. The usual assumption is that $u = 1/d$ and p can then be determined by risk neutrality as

$$p = \frac{e^{r\Delta t} - d}{u - d}$$

Finally, u is fitted using the variance of the stock price.

Check Your Progress 1

1) What do you mean by financial derivatives?

....

....

....

....

....

2) Write the meaning of the terms: futures and forwards.

....

....

....

....

....

3) List the models used for determining option price.

....

....

....

....

....

4) How do'you formulate a binary model?

....

....

....

....

....

5) What is the meaning of no-arbitrage?'

....

....

....

....

6) Remember the result of multi-period binary model and translate these in the language of a martingale,

....

....

....

....

4.4 OPTION PRICING RESULTS

4.4.1 Properties of Option Pricing

Factors affecting option prices

- the current stock price and the strike price (call options are more valuable if the stock price increases and less valuable if the strike price increases)
- the time to expiration (put and call American options are more valuable as time to expiration increases; European options are not necessarily more valuable)
- the volatility of the stock price, σ, so that $\sigma\sqrt{\Delta t}$ is the standard deviation of the stock price in a short length of time Δt (the owner of a call benefits from price increases but has limited downside risk; the owner of a put benefits from price decreases but has limited risk if price increases; therefore, the value of both calls and puts increases as volatility increases)
- the risk-free interest rate (if the rate increases, the expected growth rate of the stock price tends to increase, however, the present value of any future cash flows decreases, therefore the value of a put option decreases; for call options the first effect tends to increase the price, while the second tends to decrease it, but the first always dominates, so the value increases)
- the dividends expected during the life of an option (reduce the stock price on the ex-dividerid date, decrease the value of a call option, but increase the value of a put option)

Assume that there are no transaction costs; all trading profits (losses) are subject to the same tax rate; borrowing and lending at the risk-free interest rate is possible,

S_0 current stock price

S_t stock price at time t

X strike price of option

T time of expiration of option

r risk-free rate of interest at time T

C value of American call option to buy one share

P value of American put option to sell one share

c value of European call option to buy one share

p value of European put option to sell one share

Upper bounds

$c \le S_0$, $C \le S_0$,

$p \le Xe^{-rT}$, $P \le X$.

Lower bounds

Lower bound for Eurapean calls on non-dividend paying stock

$c \ge S_0 - Xe^{-rT}$.

Example. $S_0 = 20$, $X = 18$, r = 10% p.a., T = 1 year. Then $S_0 - Xe^{-rT} = 3.71$. Assume that the option costs $3. Then an arbitrageur buys the call and shorts the stock. This provides cash 20 − 3 = 17, invested for one year risk-free, it gives $17e^{0.1} = 18.79$. Then the option expires. If the stock price is greater than $18, the arbitrageur exercises the option ,and makes a profit of 18.79 − 18 = 0.79, If the stock price is less than $18, say $17, then the stock is bought in the market and the profit is 18.79 − 17 = 1.79.

Proof of the lower bound.

Consider two portfolios. A (one European call option and cash Xe^{-rT}) and B (one share). At time T, portfolio A is worth max (S_T, X), while portfolio B is worth S_T. Thus A must be worth more than B today, meaning that $c + Xe^{-rT} \geq S_0$.

Lower bound for European puts **on non-dividend paying** stock

$p \geq Xe^{-rT} - S_0$.

For the proof consider two portfolios. A (one put option and one share) and B (cash Xe^{-rT}).

4.4.2 Put Call Parity Case of European Options

Consider two portfolios.

A – one call option plus an amount of cash equal to Xe^{-rT};

B – one put option plus one share.

At time T both portfolios are worth $\max(S_T, X)$, so (because they are European) they must have identical values today.

$c + Xe^{-rT} = p + S_0$.

This is known as **put-call parity.**

Case of American options

Example. A trader owns a call option with strike price $40 **and** the current stock. price $50 with one month to expiry. It is better not to exercise early, because $40 can earn interest for a month, and also the option is a protection against falling price of a stock. It is better to sell the option (or keep the option and short the stock).

Thus

$C \geq S_0 - Xe^{-rT}$.

In contrast to call options, it can be optimal to exercise a put option early. For example, let X = 10 and let the stack. price be zero. Then it is better to exercise immediately to get the maximal profit and invest it. Thus

$P \geq X - S_0$.

There is no put-call parity for American options.

4.4.3 Dividends-paying Stock

Consider two portfolios.

A - one European call option,and cash $D + Xe^{-rT}$;

B - one share.

Then

$$C \geq S_0 - D - Xe^{-rT}$$

Similarly,

$$p \geq D + Xe^{-rT} - S_0.$$

Put-call parity for European options

$$c + D + Xe^{-rT} = p + S_0.$$

For dividends.paying stock, it may be optimal to exercise a call option early.

$$\Delta = \frac{f_u - f_d}{S_0 u - S_0 d}.$$

Let r be the risk-free interest rate. Then the present value of the portfolio should be equal to the cost setting up the portfolio

$$(S_0 u\Delta - f_u)\, e^{-rT} = S_0\, \Delta - f.$$

$$f = S_0\, \Delta - (S_0 u\Delta - f_u)\, e^{-rT}.$$

After substituting the value of A

$$f = e^{-rT}\, [q f_u + (1 - q) f_d],$$

where

$$q = \frac{e^{rI} - d}{u - d}.$$

This option price J'does not refer to the probabilities of the stock moving up or down. The probabilities'are incorporated into the price of the stock. Then q can be interpreted as the probability of up niovement, (I– q) is the probability of a down movement, and $q f_u + (1 - q) f_d$ is the expected payoff from the option.

Then the expected stock price

$$E(S_T) = qS_0 u + (1 - q)\, S_0 d = S_0 e^{rT},$$

meaning that the price grows, on average, risk-free*

The expectation with respect to the risk-free probabilities will be denoted by E_Q, while the expectation in the real world is E_P.

Matching **volatility with *u* and *d*.**

Let μ be the expected return of the stock and σ be its volatility. Assume $S_0 = 1$. Then the probability p of up-movement can be found from

$$pu + (1 - p)d = e^{\mu\Delta t},$$

so that

$$p = \frac{e^{\mu\Delta t} - d}{u - d}.$$

The variance of the stock price after time Δt is σ^2 *At.* On the other hand, this variance is

$$pu^2 + (1 - p)d^2 - [pu + (1 - p)d]^2.$$

From the equality (if higher powers of *At* are ignored),

$u = e^{\sigma\sqrt{\Delta t}}$, $d = e^{-\sigma\sqrt{\Delta t}}$

In a risk-neutral world, the probability of upward movement will be

$$p = \frac{e^{r\Delta t} - d}{u - d}.$$

If q is used instead of p in the formula for the variance, then the variance remains the same. The change of risk references (called also the change of measure) may lead to change of the expected return, but does not change the variance (this is Girsanov's theorem of stochastic calculus).

4.4.4 Delta Hedging

Delta is the ratio of the change in the price of the stock option to the change in the price for the underlying stock, or,

$$\Delta = \frac{\partial f}{\partial S}.$$

Delta hedging requires buying $\Delta \times$ (the value) shares of the stock.

4.4.5 Stochastic Processes on Binomial Trees

Main concepts

- The set of possible stock values is a stochastic process S, it depends on time as S_i (discrete time) or S_t (continuous time).
- The set of probabilities p_j or q_j associated with nodes of the tree is called a measure P or Q. They describe how likely is to jump up or down.
- A filtration $\mathcal{F}_i$ is the history of the stock up until tick-time t = i on the tree.
- A claim X is a function of the nodes at a claim time-horizon T; equivalently, it is an F_T – measurable function. The claim is only defined on the nodes at time T.
- The conditional expectation operator $\mathbf{E}_P(\cdot \mid \mathcal{F}_i)$. For example, $\mathbf{E}_P(X \mid \mathcal{F}_i)$ is the expectation of X along the latter position of paths which have initial segment $\mathcal{F}_i$. This expectation depends on $\mathcal{F}_i$ and so is a random variable itself. The claim X can be converted into a process $\mathbf{E}_P(X \mid \mathcal{F}_i)$ or $\mathbf{E}_Q(X \mid \mathcal{F}_i)$ if the measure P or Q is given.

  A previsible process φ is a process on the same tree whose value at any given node at time tick i depends only on the history up to one time-tick earlier $\mathcal{F}_{i-1}$. Previsible processes play the part of trading strategies where we cannot tell in advance where prices are going to go.
- A process S is a martingale with respect to a measure Q and a filtration F_i if $\mathbf{E}_Q(S_j \mid \mathcal{F}_i) = S_i$ for all $i \le j$.

This means that S has no drift under Q. Then Q is called a martingale measure for S and S is called a Q-martingale.

For any claim X, the process $\mathbf{E}_Q(X \mid \mathcal{F}_i)$ is a martingale.

Binomial Representation Theorem

Theorem. If S is Q-martingale and E is any other **Q-martingale,** then there exists a previsible process φ such that

$$E_i = E_0 + \sum_{k=1} \varphi k \Delta S_k \ ,$$

where $\Delta S_k = S_k - S_{k-1}$.

Proof. Consider a typical node. As there are two values only, the random variables can be transformed to each other by scaling and offset

$$\Delta E_i = \varphi_i \Delta S_i + k;$$

for φ, and k known by F_{i-1}. The conditional expectation given F_{i-1} of ΔS_i and ΔE_i should be zero, as they are martingales. Thus, k = 0.

The bond process

- The bond process B_i represents the value of \$1 at time i. It is a previsible and positive process, $B_0 = 1$.
- The process B_i^{-1} is another previsible process called the discount process.
- $Z_i = B_i^{-1} S_i$ is the discounted stock process.
- $B_T^{-1} X$ is the discounted claim.

Self-financing strategies

Let the discounted stock process $Z_i = B_i^{-1} S_i$ be a **Q-martingale.** Another Q-martingale is $E_i = \mathbf{E}_Q (B_i^{-1} X | F_i)$. Then there exists a previsible process φ such that

$$E_i = E_0 + \sum_{k=1} \varphi k \Delta Z_k \ ,$$

At time i buy the portfolio Π_i with

- ϕ_{i+1} units of stock;
- $\psi_{i+1} = (E_i - \varphi_{i+1} B_i^{-1} S_i)$ units of the cash bond.

At time i, the portfolio is worth $V_i = \varphi_{i+1} S_i + \varphi_{i+1} B_i$. At time zero, portfolio Π_0 is worth

$$\varphi_1 S_0 + \varphi_1 B_0 = \mathbf{E}_Q(B_T^{-1} X).$$

After one time tick Π_0 is worth

$$\varphi_1 S_1 + \varphi_1 B_1 = B_1 (E_0 + \varphi_1 (B_1^{-1} S_1 - B_0^{-1} S_0)) = B_1 (E_0 + \varphi_1 \Delta Z_1) = B_1 E_1$$

by the binomial representation theorem. On the other hand, the portfolio Π_1 scheduled to be acquired at time 1 costs also $B_1 E_1$, so we can cash Π_0 to buy Π_1. This strategy is called **self-financing,** At the end the portfolio will be worth $B_1 B_1^{-1} X = X$, as required. Thus, the claim price is $\mathbf{E}_Q(B_T^{-1} X)$.

For a general step i this argument lead to $V_i = \varphi_i S_i + \psi_i B_i$. Then

$$\begin{cases} V_i = \varphi_i S_i + \psi_i B_i, \\ V_{i-1} = \varphi_i S_{i-1} + \psi_i B_{i-1}. \end{cases}$$

Self-financing Hedging Strategy (φ_i, ψ_i)

- both φ and ψ are previsible
- the change in value V of the portfolio defined by the strategy obeys the difference equation $V_{i+1} - V_i = \Delta V_i = \varphi_{PI}\, \Delta S_i + \psi_{i+1}\, \Delta B_i$
- $\varphi_T S_T + \psi_T B_T$ is identically equal **to** the claim **X**

Option Price Formula (discrete case)

The value at time i of a claim X maturing at date T is

$B_i \, \mathbb{E}_Q (B_i^{-1} X | F_i)$.

A martingale measure Q always exists in the binomial model.

The real measure P (which S follows) is irrelevant.

Continuous Processes

Heuristic Arguments

Let $B_T = e^{rt}$. Assume that over a small time interval *At* the stock moves to either value $Se^{\mu\Delta t + \sigma\sqrt{\Delta t}}$ (if up) or to $Se^{\mu\Delta t - \sigma\sqrt{\Delta t}}$ (if down). Then, if $n = t/\Delta t$,

$$S_t = S_0 \exp\left\{\mu t + \sigma\sqrt{t}\left(\frac{2X_n - n}{\sqrt{n}}\right)\right\},$$

where X_n is the total number of up-jumps.

In the risk-neutral'world, the probability of up-movement is

$$q = \frac{Se^{r\Delta t} - S_{down}}{S_{up} - S_{down}} \approx \frac{1}{2}1 - \sqrt{\Delta t}\left(\frac{\mu + \frac{1}{2}\sigma^2 - r}{\sigma}\right).$$

Then $X_n - \mathbf{Bi}(n, q)$, and $(2X_n - n)/\sqrt{n}$ has mean

$$-\sqrt{t}(\mu + \frac{1}{2}a^2 - r)/\sigma \text{ and the variance } 1,$$

and so converges to $\mathbf{N}(-\sqrt{t}(\mu + \frac{1}{2}\sigma^2 - r)/\sigma, 1)$.

Finally,, log $S_t \sim$ N (log $S_0 + (r - 1 - \frac{1}{2}\sigma^2)t, \frac{1}{2}\sigma^2 t)$ or

$$S_t = S_0 \exp\{\sigma\sqrt{t} \setminus z + (r - \frac{1}{2}\sigma^2)t\},$$

where $z - N(0,1)$ under Q.

Thus $\log(S_t)$ has the normal distribution and so S_t has the log-normal distribution.

$$S_t = S_0 \exp\{\{\sigma\sqrt{t}Z + \mu t\},$$

where μ is called the expected return pn the stock,

Let us discuss the stock price process. It can be assumed that if follows a generalised Weiner process. That means it has constant expected drift rate and constant variance rate. Such an assumption however, fails to satisfy a key feature observed incase of stock prices. That is, the expected percentage return required by investors from a stock is independent of stock price. So we introduce a modification and assume that the expected drift be expressed as a proportion of the stock price is constant. In the real world the stock price is distributed as

Stock Price Process

Assume that the volatility vanishes, i.e., $\sigma = 0$. In this case

$$dS = \mu S dt,$$

where μ is the expected return on the stock. Then

$$\frac{dS}{S} = \mu dt$$

and

$$S_t = S_0 e^{\mu t}$$

Then the volatility of stock prices is modelled using the **Brownian Motion** (also called Wiener process). The standard Brownian motion W_t is a stochastic process, such that

- $W_0 = 0$;
- $W_t - W_s$ is normally distributed with mean zero and the variance $t - s$ for $t \geq s$;
- W_t has independent increments, i.e. $W_{t_n} - W_{t_{n-1}}, \ldots\ldots, W_{t_2}, W_{t_1}$ are jointly independent for any time moments $t_1 \leq t_2 \leq \ldots t_n$.

The Brownian motion has continuous but nowhere differentiable paths,

Let W_t be the standard Brownian motion. Then

$$dS = \mu S dt + \sigma S dW$$

or with time as subscript

$$dS_t = \mu S_t dt + \sigma S_t dW_t \text{ .}$$

4.4.6 Stochastic Differential Equations

In a more general case X_t is a general stochastic process (not necessarily stock price process) such that

$$dX_t = \mu_t dt + \sigma_t \, dW_t$$

The drift μ_t and the volatility σ_t may be random, but they must be adapted to the same filtration, so depend on the events up to the current time (i.e. F_t), but not on the future. The drift and volatility determine uniquely the underlying stock process and can be determined uniquely from X_t.

Stochastic differential equation for X_t

$$dX_t, t = \mu(X_t, t) \, dt + \sigma_t (X_t, t) \, dW_t$$

Example (Due to Hull 1992) A stock pays dividends at 15%p.a. with continuous compounding and has a volatility of 30%p.a. Then

$$dS = 0.15 S dt + 0.30 S dW.$$

Then

$$\frac{\Delta S}{S} = 0.15\Delta t + 0.30Z\sqrt{\Delta t}$$

where Z ~ N(0, 1).

It6 Formula

If $dX_t = \mu_t dt + \sigma_t dW_t$ and **f** is a deterministic twice continuously differentiable function, then $Y_t = f(X_t)$ satisfies

$$dY_t = \left(\mu_t f'(cX_t)\right) + \frac{1}{2}\sigma_t^2 f''(X_t)dt + \left(\sigma_t f'(X_t)dW_t\right)$$

More generally, let $Y_t = F(X_t, t)$ for a twice differentiable function F. Then

$$dY_t = \frac{\partial F}{\partial X_t}dX_t + \frac{\partial F}{\partial t}dt + \frac{1}{2}\frac{\partial^2 F}{\partial X_t^2}\sigma^2 dt$$

$$= \left[\frac{\partial F}{\partial X_t}\mu_t + \frac{\partial F}{\partial t} + \frac{1}{2}\frac{\partial^2 F}{\partial X^2}\sigma_t^2\right]dt + \frac{\partial F}{\partial X_t}\sigma_t \partial W_t$$

This formula transforms a stochastic differential equation for X_t into a stochastic differential equation for Y_t.

Example. X = W, Y = X^2. Then

$$dY_t = d(W^2{}_t) = dt + 2W_t dW_t.$$

This can be used to $\int_0^T W_t dW_t$., since

$$\int_0^T dY_t = \int_0^T dt + \int_0^T 2W_t dW_t$$

whence

$$\int_0^T W_t dW_t = \frac{1}{2}W_T^2 - \frac{1}{2}t$$

Example. Let $dS_t = \mu S dt + \sigma S dW_t$ for the (non-dividend paying) stock price. Find the differential for the forward price $F_t = S_t e^{r(T-t)}$.

Geometric Brownian motion

Consider

$$dS_t = \mu S_t dt + \sigma S_t dW_t.$$

Its solution using Ito's formula is

$S_t = S_0 \exp\{\sigma W_t + (\mu - \frac{1}{2}\sigma^2)t\}$ and

called the **geometric** (or exponential) Brownian motion.

This formula gives the stock price in the real world if the stock follows **the** Geometric Brownian motion with given μ and σ.

Note: W_t has the normal distribution with mean zero and the variance t, so that

$$S_t = S_0 e^{\eta t},$$

where η has the normal distribution with mean $\mu - \frac{1}{2}\sigma^2$ and the variance σ^2/t.

Estimating Volatility

Take a sample of daily returns

$$u_i = \log(S_i/S_{i-1})$$

for i = 1, . . . , n and then the standard deviation of this sample will estimate $\sigma\sqrt{T}$, where T is the length of time interval in years. The choice of n is crucial, more observations do not necessarily mean a better estimate as the volatility may change over time.

When dealing with u_i it is appropriate to give latter observations more weights, for example, use exponentially decreasing weights. It is possible to use various models from Time Series.

Martingales

A process M_t is called a Q-martingale if

- $\mathbf{E}_Q|M_t| < \infty$ for all t;
- $\mathbf{E}_Q(M_t | F_s) = M_s$ for all $s \le t$.

Examples:

1) A constant process is a martingale.
2) Q-Brownian motion is a Q-martingale.
3) $M_t = \exp\{\sigma W_t - \frac{1}{2}\sigma^2 t\}$ is a martingale.
4) For any claim X (with $\mathbf{E}_P|X| < \infty$), the process $N_t = \mathbf{E}_P|X|F_t)$ is a P-martingale.

For pricing options, the central point was to ensure that the process is a martingale with respect to some measure. The price of derivatives then becomes the expectation with respect to this martingale measure.

Self-financing Portfolios

Portfolio is a pair of processes φ_t and ψ_t which describe the number of units of security and of the bond which we hold at time t. The processes can take positive or negative values (short-selling is allowed). The security component φ should be F-previsible. The value of the portfolio is $V_t = \varphi_t S_t + \psi_t B_t$.

A portfolio is self-financing if and only if the change in its value only depends on the change of the asset prices, i.e.

$$dV_t = \phi_t dS_t + \psi_t dB_t$$

Examples: If we assume that $S_t = W_t$ and $B_t = 1$ then

1) $\varphi_t = \psi_t = 1$ is self-financing
2) $\varphi_t = 2W_t$, $\psi_t = -t\, W_T^2$ is self-financing

If X is a claim that depends on events up to time T, then a replicating **strategy** for X is a self-financing portfolio (p;ψ) such that $\int_0^T \sigma_t^2 \psi_t^2 dt < \infty$ and $X = V_T = \varphi_T S_T + \psi_T B_T$. Then the price of X at time t must be

$$V_t = \varphi_t S_t + \psi_t B_t$$

Change of Measure

If γ is a constant, and W_t is a P-Brownian motion, then there exists a measure Q such that

$$\tilde{W}_t = W_t + \gamma_t$$

is a Q-Brownian motion. This change of measure theorem is a particular case of the Cameron-Martin-Girsanov theorem from stochastic calculus.

The change of measure changes only the drift; the volatility remains the same.

Example. $X_t = \sigma W_t + \mu_t$, where W_t is a P-Brownian motion. Then, with $\gamma_t = \mu/\sigma$, there exists a measure Q such that $\tilde{W}_t = W_t + (\mu/\sigma)t$ is a Q-Brownian motion up to time T, i.e., $X_t = \sigma \sim \tilde{W}_t$

4.5 BLACK-SCHOLES MODEL

Black and Scholes derived a differential equation that must be satisfied by the price of any derivative security dependent on a non-dividend paying stock. The fundamental insight gained from their result is, the option is implicitly priced if the stock is traded.

Assumptions:

The assumptions of the Black-Scholes models are: The price of the underlying instrument S_t follows a geometric Brownian motion with constant drift μ and volatility σ in the equation $dS_t = \mu S_t dt + \sigma S_t dw_t$.

Short-selling is permitted, no transaction costs or taxes, all securities are perfectly divisible, there are no dividends, there are no riskless arbitrage opportunities, security trading is continuous, the risk-free interest rate is constant and the same for all maturities.

The PDE

Given the assumptions of the Black-Scholes models, the price V_t of a derivative written on a stock with price process S_t evolves according to the following partial differential equation (PDE):

$$\frac{\partial V}{\partial t} + \frac{1}{2}\sigma^2 S^2 \frac{\partial^2 V}{\partial S^2} + rS\frac{\partial V}{\partial S} - rV = 0$$

It is notable that the PDE does not contain μ, the drift of the stock. An informal derivation of the PDE, explaining this result, is given below.

The Formula

The above discussions lead to the following formula for the price of a call option with exercise price K on a stock currently trading at price S, i.e., the right to buy a share of the stock at price K after T years. The consant interest is r, and the constant stock volatility is σ.

$$C(S,T) = S\Phi(d_1) - Ke^{-rT}\Phi(d_2)$$

where

$$d_1 = \frac{\ln(S/K)+(r+\sigma^2/2)T}{\sigma\sqrt{T}}$$

$$d_2 = d_1 - \sigma\sqrt{T}$$

Here Φ is the standard normal cumulative distribution function.

The price of a put option may be computed from this by put-call parity and simplifies to $P(S,T) = Ke^{-rT}\Phi(-d_2) - S\Phi(-d_1)$.

Assume first that the interest rate is zero, r = 0.

Step 1. Find a measure Q under which S_t is a martingale. Take S_t as a geometric Brownian motion, so that

$$S_t = S_0 \exp\{\sigma W_t + (\mu - ½\, \sigma^2)t\}.$$

To convert it into a martingale we must get rid of μ

$$S_t = S_0 \exp\{\sigma(W_t + \frac{\mu}{\sigma}\, t) - ½\, \sigma^2 t\}$$

$$= S_0 \exp\{\sigma \tilde{W}_t - ½\, \sigma^2 t\},$$

where $\tilde{W}_t$ is the Q-Brownian motion (see change of measure theorem).

Step 2. Put $E_t = \mathbf{E}_Q(X | F_t)$

Step 3. $dE_t = \varphi_t dS_t$

Thus, the replicating strategy is to hold φ_t units of stock at time t and hold

$$\psi_t = E_t = \varphi_t S_t$$

units of the bond at time t.

Assume now that the interest rate is not zero.

Take the discount process B_t^{-1} and form a discounted stock $Z_t = B_t^{-1} S_t$ and a discounted claim $B_T^{-1}X$. Then operate with them as though interest rates were zero, remember that $B^t = e^{rt}$.

Step 1. Make Z_t into a martingale.

Note that

$$Z_t = B_t^{-1} S_t = S_0 \exp\{\sigma W_t + (\mu + r\, ½\, \sigma^2)t\}$$

$$= S_0 \exp\{\sigma W_t + \frac{\mu + r}{\sigma}) + ½\, \sigma^2 t\}$$

$$= S_0 \exp\{\sigma W_t - ½\, \sigma^2 t\}$$

for a Q-Brownian motion W_t, so that Z_t is a Q-martingale. Then

$$S_t = B_t Z_t = e^{rt} Z_t \text{ and so}$$

$$= S_0 \exp\{\sigma W_t + (r - ½\, \sigma 2)t\}$$

determines the stock price process in the risk-neutral world. The corresponding stochastic differential equation is

$$dS_t = rS_t dt + \sigma S_t dW_t$$

Step 2. $E_t = \mathbf{E_Q}(B_t^{-1} X \mid F_t)$

Step 3. By martingale representation theorem, $dE_t = \varphi_t dZ_t$.

Replicating strategy.

- hold φ_t units of the stock at time t
- hold $\Psi_t = E_t = \varphi_t Z_t$ units of ***the*** bond

A strategy $(\phi_t, +\psi_t)$ of holding in a stock S_t and a non-volatile cash bond B_t has value $V_t = \phi_t S_t + \psi_t B_t$ and discounted value $E_t = \varphi_t Z_t + \psi_t$

The strategy is self-financing if either

$dV_t = \varphi_t\, dS_t + \Psi_t\, B_t$ or, equivalently,

$dE_t = \varphi_t dZ_t$.

All claims X, knowable up to some horizon, T have associated replicating . strategies. The arbitrage price of such a claim is given by

$$V_t = B_t\, \mathbf{E_Q}(B_t^{-1} X \mid F_t) = e^{-r(T-t)} = \mathbf{E_Q}(X \mid F_t)$$

where Q is a measure that makes the discounted stock a martingale.

Example: Pricing of a Call Option

Consider a call option with the exercise date T and strike price k. Then the claim is X = $\max(S_T - k, 0)$. Find V_0, the value of the replicating strategy (and thus the option) at time zero as

$$V_0 = e^{-rT}\mathbf{E_Q}(\max(S_T - k, 0)).$$

We should find the marginal distribution of S_T under Q

$$S_T = S_0 \exp\{\sigma W_t + (r - ½\, \sigma^2)T\}$$

Since $W_T \sim N(0, T)$ **under Q,**

$$\sigma W_t + (r - ½\, \sigma^2)T = Z + rT.$$

where $Z \sim \mathbf{N}(½\, \sigma^2 T, \sigma^2 T)$. Then

$$S_T = S_0 e^{Z + rT}$$

so that the claim price is

$$V_0 = e^{-rT}\mathbf{E}((S_0 e^{Z+rT} - k)_+)$$

$$= \frac{1}{\sqrt{2\pi\sigma^2 T}} \int_{\log(k/S_0)-rT}^{\infty} \left(S_0 e^x - ke^{-rT}\right) \exp\left\{-\frac{\left(x + \frac{1}{2}2\sigma^2 T\right)^2}{2\sigma^2 T}\right\} dx.$$

Change of variables $v = -(x + \frac{1}{2}\sigma^2 T)/\sigma\sqrt{T}$ leads to

$$V_0 = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{a} \left(se^{-\sigma\sqrt{T}v - \frac{1}{2}\sigma^2 T}\right) e - \frac{1}{2}\sigma^2 dx.$$

where $a = (\log(s/k) + (r - \frac{1}{2}\sigma^2 T)/\sigma\sqrt{T}$. Then writing

$$-\sigma\sqrt{T}v - \frac{1}{2}\sigma^2 T) - \frac{1}{2}U^2) = \frac{1}{2}(v + \sigma\sqrt{T})^2$$

yields

$$V_0 = \frac{s}{\sqrt{2\pi}} \int_{-\infty}^{a-\sigma\sqrt{T}} e^{-\frac{1}{2}v^2} dv - \frac{ke^{-rT}}{\sqrt{2\pi}} \int_{-\infty}^{a} e^{-\frac{1}{2}v^2} dv.$$

This leads to **Black-Scholes** formula

$$V_0 = S_0\phi\left(\frac{\log\frac{S_0}{k} + \left(r + \frac{1}{2}\sigma\frac{2}{1}\right)}{\sigma\sqrt{T}}\right)^2 - ke^{-rT}\phi\left(\frac{\log\frac{S_0}{k} + \left(r - \frac{1}{2}\sigma^2\right)T}{\sigma\sqrt{T}}\right)$$

We only need that the drift of the stock price is constant, its exact value is immaterial.

The Black-Scholes formula may be written as

$$c = S_0\Phi(d_1) - k\Phi(d_2)$$

Note that $d_2 = d_1 - \sigma\sqrt{T}$.

The corresponding price formula for a European put is

$$p = ke^{-rT}\Phi(-d_2) - S_0\Phi(-d_1).$$

Finding the Replicating Strategy

The value of the claim X at time t is equal to $V(S_t, t)$ as in Black-Scholes formula with S_t (stock price at time t) instead of S_0 and $(T - t)$ (the remaining time) instead of T. Since $dS_t = \sigma S_t dW_t + rS_t dt$, Ito's formula yields

$$dV(S_t, t) = \frac{\partial V}{\partial S_t} dS_t + \left(\frac{\partial V}{\partial t} + \frac{1}{2}\sigma^2 s^2 \frac{\partial^2 V}{\partial s^2}\right)dt.$$

On the other hand, the self-financing condition gives

$$dV_t = \varphi_t dS_t + \Psi_t dB_t.$$

By equating coefficients of dS_t, we get

$$\varphi_t = \frac{\partial V}{\partial S_t}(S_t, t)$$

meaning that the amount of stock at any stage is the derivative of the option price with respect to the stock price,

Compare the values of the replicating strategy $V(S_t, t) = \varphi_t S_t + \Psi_t B_t$ with option price given by the Black-Scholes formula

$$V(S_t, t) =$$

$$\Phi\left(\frac{\log(S_t/k) + (r + \frac{1}{2}\sigma^2)(T-t)}{\sigma\sqrt{(T-t)}}\right) - ke^{-r(T-t)}\Phi\left(\frac{\log(S_t/k) + (r - \frac{1}{2}\sigma^2)(T-t)}{\sigma\sqrt{(T-t)}}\right)$$

to see that the **replicating strategy** is given by

$$\psi_t = \Phi\left(\frac{\log(S_t/k) + (r + \frac{1}{2}\sigma^2)(T-t)}{\sigma\sqrt{(T-t)}}\right)$$

and

$$B_t\psi_t = -ke^{-r(T-t)}\Phi\left(\frac{\log(S_t/k) + (r - \frac{1}{2}\sigma^2)(T-t)}{\sigma\sqrt{(T-t)}}\right)$$

The value of φt is always between zero and one. The amount of borrowing is bounded by the exercise price k.

4.5.1 Black-Scholes Differential Equation

The Black-Schotes differential equation is derived in the following through no-arbitrage or delta-hedging argument.

The Black–Scholes PDE

As per the model assumption above, we assume that the underlying (typically the stock) follows a geometric Brownian motion. That is,

$$dS_t = \mu S_t dt + \sigma S_t dW_t$$

where W_t is Brownian.

Now let V be some sort of option on S-mathematically V is a function of Sand t. V(A, t) is the value of the option at time t if the price of the underlying stock at time t is S. The value of the option at the time that the option matures is known. To determine its value at an earlier time we need to know how the value evolves as we go backward in time.

Another derivation

Consider a portfolio with one long option and short position in the asset

$$\Pi = V(S,t) - \Delta S.$$

Then $d\Pi = dV - \Delta dS$ and using Itô formula for $V_t = V(S_t, t)$

$$dV = \frac{\partial V}{\partial t}dt + \frac{\partial V}{\partial S}dS + \frac{1}{2}\sigma^2 s^2 \frac{\partial^2 V}{\partial s^2}dt.$$

Therefore

$$d\Pi = \frac{\partial V}{\partial t}dt + \frac{\partial V}{\partial S}dS + \frac{1}{2}\sigma^2 s^2 \frac{\partial^2 V}{\partial s^2}dt - \Delta dS.$$

The coefficient before dS is zerb if

$$\Delta = \frac{\partial \Pi}{\partial S}$$

as in Delta hedging. With such A,

$$d\Pi = \left(\frac{\partial V}{\partial t} + \frac{1}{2}\sigma^2 s^2 \frac{\partial^2 V}{\partial s^2}\right)dt,\ dt .$$

This change is riskless and must be equal to the risk-free growth (to avoid arbitrage possibilities)

$$d\ \Pi = r\Pi dt .$$

Thus, $\left(\frac{\partial V}{\partial s} + \frac{1}{2}\sigma^2 s^2 \frac{\partial^2 V}{\partial s^2}\right)dr = r\left(V = \frac{\partial V}{\partial S}\right)dt,$

and we obtain the Black-Scholes equation

$$\frac{1}{2}\sigma^2 s^2 \frac{\partial^2 V}{\partial s^2} + rs\frac{\partial V}{\partial s} - rV + \frac{\partial V}{\partial t} = 0.$$

Dividend Paying Stock

Example (European call option on a dividend paying stock).

Ex-dividend dates in two months and five months. The dividend on each ex-dividend date is expected to be \$0.50. The current share price is \$40, the volatility is 30%p.a., the risk-free rate of interest is 9%p.a. The time of maturity is 6 months. The present rate of the dividends is

$$0.5e^{-2/12\times 0\,09} + 0.5e^{-2/12\times 0\,09} = 0.9741.$$

The option price can be calculated from the Black-Scholes formula with S_0= 39.0259, k = 40, r = 0.09, σ = 0.3, and T = 0.5. The result is \$3.67.

Stocks paying dividends continuously. If the stock is paying dividends **at** rate q (continuously), then the Black-Scholes formula is applicable with S_0 replaced by the dividend-adjusted stock price S_0e^{-qT}. This is also applicable for options on stock indices.

'The dividend payment is assumed to be $\delta S_t d_t$ over a small interval d_t. Then the process S_t is not the value of the asset as a whole. To make it tradable, we can assume that dividends are used to purchase more stock, so at time t the number of stock units will be St. Then the worth of such a portfolio will be

$$S_t = S_0 \exp\{\sigma W_t + (r - \delta - \tfrac{1}{2}\sigma^2)t\}$$

Then the standard arguments are applicable, and, under the martingale measure,

$$S_t = S_0 \exp\{\sigma W_t + (r - \delta - \tfrac{1}{2}\sigma^2)t\}$$

is log-normally distributed. Then the forward price is $F = e^{(r-s)T}S_0$

$$V_0 = e^{-rT}\left\{F\varphi\left(\frac{\log(F/k)\frac{1}{2}A(\mu_1 - \mu_2) + \sigma_2^2 - \rho\sigma_1\sigma_2}{\sigma(\hat{R})}\right)\right\}.$$

and the hedge is $\varphi_t = e^{-\delta(T-t)}\,\phi(d_1)$ units of the stock and have a negative holding of $ke^{-rT}\,\phi(d_2)$ units of the bond.

Note The same result is obtained if one adjusts S_0 for the dividends and uses the standard Black-Scholes formula on non-dividend paying stock with S_0 replaced by S_0e^{-qT}.

4.6 OPTIMAL PORTFOLIOS

In portfolio theory we study investors optimal decision to **diversify** their portfolio, as well as pricing of risky assets. Usually the portfolio theory models the returns of an asset as random variable.

If properly hedged, derivatives provide **riskless** returns. This feature is used by banks who sell their products for a bit more than it is worth. Fund managers buy and sell assets (and also derivatives) with the aim of beating the bank's rate of return. This often involves taking a degree of risk.

4.6.1 Mean-Variance Approach,

The mean-variance approach was developed by H. Markowitz. To see its formulation let us examine the following presentation.

Let $P_i(t)$ be the price of security i at time t.

The returns $R_i = P_i(t)/P_i(0)$ are modelled as random variables with

$$\mu_i = \mathbb{E}(R_i),\ \sigma_{ij} = \mathrm{Cov}(R_i, R_j),$$

The variance-covariance matrix of returns is denoted by Σ. If the first asset is risk-free, then the corresponding variance and covariances are zeros.

Let π_i be the fraction of initial wealth X of the investor invested in security i at time t = 0

$$\pi_i = \frac{\psi_i P_i(0)}{X}$$

where ψ_i is the number of shares of security i held by investor at time t = 0,

$\pi = (\pi_1, \dots, \pi_d)^T$ is called the portfolio vector. Note that

$$\pi_1 + \pi_2 + \dots + \pi_d = 1 .$$

If short-selling is not allowed, portfolios should have all non-negative components. Such portfolios are called **admissible.** The total return is given by

$$(R) = \sum_{i=1}^{d} \pi_i R_i$$

Then

$$E(R) = \mathbb{E}(R) = \sum_{i=1}^{d} \pi_i \mu_i ,$$

$$V(E) = \mathrm{Var}(R) = \sum_{i=1}^{d} \pi_i \sigma_{ij} \pi_i .$$

Two Assets Portfolios

Take $\pi = (\pi_1, \pi_2)^T$ with $\pi_1 + \pi_2 = 1$. For simplicity, denote $\pi_1 = x$ and $4 = 1 - x$. Then

$$E(R) = x\mu_1 + (1 - x)\,\mu_2,$$

$$V(R) = x^2\sigma_1^2 + (1-x)^2\sigma_2^2 + 2x(1-x)\rho\sigma_1\sigma_2,$$

where ρ is the correlation coefficient between the two individual returns. If short sales are forbidden, then $0 \le x \le 1$, otherwise x can be an arbitrary number.

Now $V(R)$ is a quadratic function of $E(R)$, namely,

$$V(R)(\mu_1-\mu_2)^2 =$$
$$(E(R)-\mu_2)^2\,\sigma_1^2 + (\mu_1-E(R))^2\,\sigma_2^2 + 2(E(R)-\mu_2) + (\mu_1-E(R))\,\sigma_1\sigma_2\,.$$

If the short sales are allowed, then the minimum variance is achieved at

$$x^* = \frac{\sigma_2^2 - \rho\sigma_1\sigma_2}{\sigma_1^2+\sigma_2^2-2\rho\sigma_1\sigma_2}.$$

Note that the corresponding **x** may be negative or greater than 1, If short sales are forbidden, then the optimal x can be found as either 0 or 1 or x* given above

(if $x^* \in (0, 1)$).

The objective function may be written as

$$f(x) = -AE\,(R) + V(R),$$

where A is a risk aversion index. If $A = 0$, the portfolio with the lowest variance of return will be selected. As A increases, the investor becomes more willing to accept risk in order to achieve a higher expected return, and if $A = \infty$ the portfolio with the highest expected return will be optimal. In application to two asset

portfolios, this gives

$$f(x) = A(x\mu_1 + (1-x)\mu_2) + 2x\sigma_1^2 + (1-x)^2\sigma_2^2 + 2x(1-x)\rho\sigma_1\sigma_2.$$

Then

$$\frac{df}{dx} = A(\mu_1-\mu_2) + 2x\sigma_1^2 - 2(1-x)\sigma_2^2 + 2(1-2x)\rho\sigma_1\sigma_2.$$

The derivative is zero if

$$x^* = \frac{\frac{1}{2}A(\mu_1-\mu_2)+\sigma_2^2-\rho\sigma_1\sigma_2}{\sigma(R)}$$

If short sales are forbidden, then the optimal value of x is either 0 or 1 or x^* above (if $x^* \in 2\,(0, 1)$).

Risk-free Assets and the Capital Allocation

Assume the portfolio R_p has two assets, a risk-free and another risky asset. The capital allocation line connects all portfolios that can be formed using a risky asset and riskless asset. The second asset R also has several risky assets with return $E(R)$ and variance V(R). Let x be the share of the fund invested in the first (risky) asset. Assume that the risk-free asset provides return R_F. Then the asset of the combined portfolio R_p will be

$$E(R_p) = xE(R) + (1-x)R_F,$$
$$V(R_p) = x^2V(R)\,.$$

Then, if $\sigma(R) = \sqrt{V(R)}$ is the standard deviation of the risky portfolio and

$$\sigma(R) = \sqrt{V(R_r)(R_P)}$$

$$E(R_p) = R_F + \frac{E(R) - R_F}{\sigma(R)}\sigma(R_p).$$

Thus, the set of portfolios comprising **a** combination of the risk-free asset and a risky portfolio is a straight line. The tangent 'point M to the set of admissible portfolios represents the market portfolio or the best combination of risky assets. The optimal position on the capital market line corresponds to the degree of risk preferred. If continued beyond M, it represents a 'borrowing' portfolio.

If the borrowing rate R_B is different from the lending rate R_F, then two tangent points are to be determined.

4.6.2 Capital Asset Pricing Model (CAPM)

The CAPM is based on investor's decision of not to invest in an asset which does not improve the risk-return charactertics (more risk- more return) of her existing portfolio. Thus, CAPM derives the return for an asset in a market given the risk-free rate available to investor and the risk of the market as a whole.

Assume that the return of a security is determined entirely by the market index and random factors

$$R_j = \alpha_j + \beta_j R_M + \varepsilon_j,$$

where R_M **is** the return of the market portfolio and ε_j is a random error term. Then

$$\mathbf{E}(R_j) = r + \beta_j [\mathbf{E}(R_M) - r],$$

where r is the risk-free rate: $\mathbf{E}(R_j)$ is the expected return of asset j; $\mathbf{E}(R_M)$ is the expected return on the market portfolio; β_j is the "beta" of asset j. The beta's can be estimated from the following linear model.

$$r_{j,t} = \alpha_j + \beta_j\, r_{M,t} + \varepsilon_{j,t},$$

where $r_{j,t}$ is the return of asset j in the tth period, $r_{M,t}$ is the return of the market portfolio in the tth period, α_j is the intercept, $\varepsilon_{j,t}$ is the residual error for period t.

The beta of an asset R_j can be also defined as

$$\beta_j = \frac{Cov(R_j, R_M)}{Var(R_M)}.$$

Further,

$$Var(R_j) = \beta_j^2\ Var(R_M) + Var(\epsilon_j)$$

and, for $i \neq j, Cov(R_i, R_j) = \beta_i \beta_j\ Var(R_M)$ **.**

The beta of a portfolio can be found as the weighted sum of beta's of individual assets

$$\beta\Pi = \sum_{i=1}^{d} \Pi_i \beta_i.$$

Cheek Your Progress 2

1) What do you mean by delta hedging?

....

....

....

....

....

2) What is the main finding of Black –Scholes model?

....

....

....

....

....

3) What do you mean by self-financing strategy?

....

....

....

....

....

4) How do you get the replicating strategy in Black-Scholes model?

....

....

....

....

....

5) How do you account for risk in a mean-variance approach?

....

....

....

....

....

6) What is the main finding of capital asset pricing model?

....

....

....

....

4.7 LET US SUM UP

In this unit we have discussed the basic concepts of finance that help understand modelling pricing stochastically. The financial terms frequently used in derivates have been documented before discussing the modelling on option prices in arbitrage free conditions. Binary modelling technique has been introduced with single and multi-period framework. The results have been further seen by translating these in terms of martingale language. The Black-Scholes model which showed that option is implicitly priced if the stock is traded, has been highlighted. The basic formulation of Black-Scholeg model which is built on the assumption of option price following a geometric Brownian motion is deliberated upon and tools of stochastic calculus used have been pointed out. The last part of the unit gives the themes on optimal portfolios and discusses the mean variance approach and capital asset pricing model to show the determination of tlie return of an asset in risky environnient.

4.8 KEYWORDS

Contingent Claims: Assets or securities whose prices depend on the values of other assets or numerical indices,

Derivative: A conditional instrument used bymarket participants to trade or manage an asset. There are essentially two categories of financial instruments that are grouped under the term derivatives: options/futures and swaps.

Financial **Market:** A mechanism which allows people to trade money for securities.

Financial Risk: Any risk associated with money.

Forward: An agreement between two parties to buy or sell an asset at a pre-agreed future point in time.

Futures: A contract to buy or sell a certain underlying instrument at a certain date in the future, at a pre-set price. The future date is called the delivery date or final settlement date. The pre-set price is called the futures price.

Hedge: An investment made to reduce the risk in another investment.

Hedging: A strategy designed to minimise exposure to an unwanted business risk, while still allowing the business to profit from an investment activity.

Long Position: A treader buys an option contract that she has not already written (i.e. sold). Here buyers are referred as the long.

Market **Risk:** Tlie potential to experience financial losses due to fluctuations in the prices at which equities, foreign currencies, 'interest rate linked securities, and commodities can be bought or sold.

Option: A derivative type of tradeable contract in securities markets, whereby one party buys the right to exercise a feature of the contract on or before a future date.

Position: A commitment to buy or sell a given amount of securities or commodities.

Put **Option (or Put):** A financial contract between two parties, the buyer and the writer of the option. The put allows the buyer the right but *not the* obligation to sell a commodity or financial instrument (the underlying instrument) to the writer of the option for a certain time for a certain price (the strike price).

Put–call Parity: A relationship between the price of a call option and a put option - both with the identical strike price and expiry.

Replication: An option trading strategy used to ensure at a certain date the payoff without trading it.

Self-Financing: A strategy when a company is able to release liquidity in order to finance operating charges and development. It is equal to the net profit to which you reintegrate all the entries that are not associated with the movements of cash such as depreciation and provisions, appreciation and depreciation of asset disposals.

Short Position: A trader *selling* an option contract that she does not already own. Here sellers are referred as the short.

Spot Price or **spot rate: A** price of a security quoted for immediate (spot) settlement (payment and delivery).

Stock: Capital raised by a firm through the'issuance and distribution of shares.

Swap: A derivative where two counterparts exchange one stream.of cash flows against another stream. Often used to hedge certain risks, such as interest rate risk or for speculation.

4.9 SOME USEFUL BOOKS

Boshizen, F., A.W. Vaat, H. Zanten 'and K. Banachewtcz (2005), Stochastic ..Process for Finance Risk Management Tools (see internet)

Hull, John (1 992), Options, Futures, and other Derivative Securities, Prentice-Hall International, Inc., New Jersey

Chalasami, P., Somesh Jha (1997), Steven Shreve: Stochastic Calculus Finance Carregie Mellon University (see internet)

4.10 ANSWER OR HINTS TO CHECK YOUR PROGRESS

Check Your Progress 1

1) See Section 4.2 and answer
2) See Sub-section 4.2.1 and answer
3) See Section 4.3 and answer
4) See Section 4.3.2and answer
5) See Sub-section 4.3.2and answer
6) See Sub-section 4.3.3and answer

Check Your Progress 2

1) See Sub-section 4.4.5 and answer
2) See Section 4.5 and answer

3) See Sub-section 4.4.5 and answer
4) See Section 4.5 and answer
5) See Sub-section 4.6.1 and answer
6) See Sub-section 4.6.2 and answer