1. Introduction

1.1 General Introduction

"Even when students believe they can learn materials in a class, and when they know the appropriate strategies for learning, they won't do it — until we give them a reason to do it" — Tuckman.

The field of history of education is nearly as broad as the field of history itself, for its subject is the study of how societies have transmitted cultures to young and how ideas, beliefs and faith have been taught to young and old alike.

Education, in general, aims at shaping the personality of children in the desired direction. It is done through a variety of inputs like curriculum, textbooks, other instructional materials, organisation of teaching learning through direct and indirect contact with children. Other meaningful inputs come from peers, parents, siblings and the interaction with the society. But, however, major emphasis is always given on “class room teaching”.

It is rightly said that while it matters what is taught, it matters much more how it is taught and it matters still more how it is evaluated.
The method of examination therefore determines the method of teaching to a great extent, so that, aspects of curriculum which are not subjected to examination are usually neglected in teaching. Ideally speaking, whatever is taught should be tested, but in practice it is reverse, – whatever is tested is taught.

Main purpose of evaluation is envisaged by Education Commission (1964-1966) and re-iterated in National Policy of Education (1986) is to help to determine and raise the standards of attainments in state and in national levels. A country requires specialists in different fields for the smooth functioning of the whole system, and to specialise in a particular field, a person must come out successfully from the institutions meant for that particular skill. Educational institutions of all types are therefore bound to be involved in measuring and reporting the efforts of their students. The symbols and signals used in such reporting are usually considered as marks and organisation and operation of the process for establishing, recording, and transmitting such marks is the “Grading System”.

In general a Grade is defined as a degree or step in a scale as of quality, ability, dignity etc. Measurement experts such as Peter Aerasian (1994) explains that educators use grades primarily (i) for administrative purpose, (ii) to give students feedback about their progress and achievement, (iii) to provide guidance for students about their future course work, (iv) to provide guidance to teachers for instructional planning and (v) to motivate students.
According to him, grades have served a variety of administrative functions, most dealing with decisions about students including their passing out and retention or student’s entrance in an institution.

Another important purpose for grades is to provide feedback about student’s achievement.

Again for guidance purpose grades help counsellors to provide direction for students to recommend courses they should take or not to.

Teachers also use grades for initial decisions about student’s weaknesses in order to group them for discussion.

Again grades motivate students to try harder from both negative and positive perspectives.

But according to Austin and Mc Cann (1992), using grades to provide feedback about students achievement should be the primary function of grades, guidance ranked second, motivation and instructional planning are tied third and administration the last.

Anderson and Block (1977), Bloom (1971) engaged in development of mastery learning have shown that when time to learn is allowed to vary, a student’s prior knowledge is most important, other researchers have shown that when time to learn is held constant, as it is in most learning environments, then a
student's intelligence and academic ability is most important. Other student characteristics that have been found to be important include study habits, age, sex, motivation, learning style, socio-economic development etc.

There are a wide variety of other context variables that influence the teaching learning process. They include, home, peer groups, community, society, culture and infrastructural condition of the institution etc. The variable related to home environment seems to be especially important, because they include items like educational level of parents, family income, amount of technology at home, number of books and magazines at home etc. One of the variables that best predicts student's achievement seems to be the level of mother's education.

1.2. Aim of the thesis:

In this modern age the whole educational system has reached a stage where it has achieved the status of an industry. With more and more involvement of private sectors in the educational field of the country with growing number of private institutions, some questions always touch one's mind as to, "which institution will be the most reliable one?". "Is it possible for an institution to fulfil the dreams of the students in putting their footprints on the path they desire to travel?".
As these private institutions are charging quite exorbitant fees, behind all these queries is the basic human desire to obtain the value for money, and to avoid frustrations and possible damages associated with the unacceptable results from the students.

In past or even in recent times different authors tried to analyse examination results on the basis of the available data, which cannot provide us with any concrete solution to the above mentioned queries. That is, the theoretical aspect of the problem was never discussed which can help one to provide answers to such questions. The proper modelling of the system will always be necessary to study its theoretical aspect.

A model takes into account the essentials of a phenomenon or a system ignoring minor details. That is, a model is a simplification of the real world with most important elements included. A good model enables us to workout, what is likely to happen, if we take up certain actions, so that we can avoid actions which have undesirable consequences.

Since the model is an approximation, the conclusions drawn from the model are also approximations, even though the conclusions are logically deduced from the model. In fact, if the deductions following the model agree with actual observations, we consider the model to be satisfactory and use it for further study of the phenomena. Many times improvement can be obtained by introducing random variables or chance factors in the model, because most of the real life phenomena are governed by chance factors. A model involving chance
factors is called a 'stochastic model' or a 'probability model'. Even though, they are difficult to analyse, they fit the observations well. As government has the models of the country's economy, where they can try out any changes, they are thinking of making, the grading system should also follow a flexible model which can help in solving different problems associated with the system.

In this thesis we have considered the grading pattern of the students in the two most important examinations of student life, viz., the high school leaving certificate examination and the pre-degree examination. As we know, depending on the scores of the student in the pre-degree examination, a student can choose his or her future course of study and as most of these courses take up entrance tests which set-up 50% marks in the pre-degree examination as the minimum requirements to be eligible to sit for these tests, we have divided the whole marking system into three grades viz., grade '0' for the students scoring above 75% of marks, grade '1' for those scoring between 50% and 75% and grade '2' for the students scoring below 50%.

1.3 Plan of the thesis

A brief survey on the development in the same field is given in the next section of the first chapter.

In the second chapter we have started with a case study in order to examine the nature of the dependency between the results of H.S.L.C. and pre-
degree examinations and a Markovian model is formulated and is analysed probabilistically.

The third chapter deals with the extension of the model discussed in the second chapter, which was a discrete time Markovian model, to a continuous time Markovian model under certain conditions to be fulfilled. Also the probability of attaining a particular grade at time \( t \) is obtained under different initial status. Again different characteristics viz., expected number of visits to the state of excellence, probability of sure success etc. were obtained both theoretically and numerically.

In the fourth chapter the examination results are analysed from a stochastic point of view. Also analysis of results are made for tests held at random interval of time.

The fifth chapter deals with Bayesian analysis of the examination results, where the transition probabilities \( p_{ij} \)'s are assumed to be random variables. The M.L.E. of the transition probabilities are obtained and are compared with the Baye’s estimators of \( p_{ij} \)'s under different priors. Also two case studies are carried out in this chapter with the help of real life data.

The sixth i.e. last chapter of this thesis gives a brief discussion and future scope of studies. Relevant graphs and numerical results are presented with different chapters.
Finally, a list of references is displayed at the end of the thesis.

1.4 Brief Survey of the existing literature

In recent years quite a number of studies were made in the field of education to study its' different aspects.

Shyam Uday Singh et.al. (1996) carried out a study on creative ability and its impact on achievement in science to study the significant mean difference on the criterion of creativity and to determine the significant mean difference among different levels of achievement in science and on the criterion of creativity.

Vijaya Lakshmi (1996) on the other hand carried out a comparative study of intellect abilities of tribal and non-tribal students by taking a sample of randomly selected school students both tribal and non-tribal from a district in the state of Bihar and showed that tribal students will differ significantly from non-tribal students in intellectual abilities.

Again Suniti Khare (1996) in her work showed significant correlation between home environment and achievement of students in their school examination.
W.R. Callen *et. al.* (1994) carried out a statistical analysis of students’ performance in New Engineering science core course where they introduced some modified courses along with some traditional courses. To measure students preparation in the subject a pre-test was designed and administered in the first week of the term. Analysis revealed that there was no significant difference between the final scores of the individual courses whether taught by the new method or by the traditional method. They used general linear model procedure by taking final examination mark as the dependent variable and other such as gender, major subject, cumulative grade point average as independent variables. Also there was no significant difference between the pre-test scores and the final scores.

Munoz – Repiso, M. *et. al.* (1991) made a study to know the average grades of the students in the university admission examination and to evaluate the congruency of the grades of the University admission test with the average marks in last two years secondary examinations test. They found good amount of correlation between the marks of the two last years secondary school examination and the grades in the admission test.

On the other hand Gonzalez, B. *et.al.* (1990) studied the higher education access system in six European countries viz., Belgium, Spain, France, Italy, U.K and Germany and found that all the countries except Belgium have a single examination system that is valid for the whole country and which has to be passed in order to start higher education because only a limited number of seats are there in the universities.
Deely and Smith (1998) has discussed a method of comparison of institutional performance using the studies presented by Goldstein and Spiegelhalter (1996). Here in this paper they defined two ranking criterias and their performances are illustrated.


A follow-up of 2600 students in twenty institutions of higher education was presented by Prof. N.J. Entwistle of Lancaster University to identify characteristics of students successful in different courses and in different institutions.

Again Prof. R.F. Kempa of the department of education, Keele University presented a article on the development of noval procedures for curriculum evaluation, student characteristic and their effect on science learning behaviour.

Our purpose here in this thesis is not in any way to dispute their findings nor necessarily to endorse particular model but rather to add other methods of modelling which provide a more refined way of analysis of grading system.
However in this context it may be mentioned that stochastic process 
(Kemeny, J. and Snell, J. (1960))
has been applied to "Easte's Learning Model" and has been exploited to some 
extent. Below we are giving a small and a brief survey of the Easte's model, with 
only some special cases, not the most general theory.

The theory had been developed to explain certain kind of learning 
which can be illustrated by experiments of the following kind.

Suppose, a human subject is given a sequence of heads (H) and tails 
(T) and each time it is asked to guess what the next choice will be. He is to try to 
get as many right as possible. Again there are various ways that the experimenter 
can produce his sequence of H's and T's and the interest lies in how the subject 
will react to different choices.

In the Easte's model it is assumed that there are a finite number of 
elements called the stimulus elements. At any given time each of these elements 
is connected either to a response $A_1$ or to a response $A_2$.

In a single experiment there is certain probability $\theta$ ( $0 < \theta < 1$ ), 
that any particular stimulus element will be sampled by the subject. It is assumed 
that elements sampled and connected to $A_1$ influences the subject in the direction 
of producing an $A_1$ response and so on. The sampling of various elements are 
assumed to be an independent trials process. We also assume that the 
experimenter takes one of the two possible reinforcing actions, $A_1$ or $A_2$. 

Subject learns about the choice of the experimenter only after he has made his own choice. In some experiment the subject is rewarded if he does guess correctly. Two basic assumption are made in this context.

(a) The probability that the subject makes response $A_1$ is equal to the proportion of elements in the set sampled that are connected to $A_1$. If no element is sampled, the responses are assumed to be the same as if all elements are sampled.

(b) If, in a given experiment, the experimenter chooses $E_1$, then all the elements that were sampled on this experiment, and that were connected to $A_2$ have their connections changed to $A_1$. If the experimenter choose $E_2$, then all the elements sampled and connected to $A_1$ have their connections changed to $A_2$.

We assume that the experimenter chooses $E_2$ with probability $a$, if the subject made response $A_1$ on the previous experiment, and chooses $E_1$, with probability $b$, if the subject made response $A_2$ on the last experiment. We can represent choices of experiment for each choice of the subject by the matrix,

$$
\begin{pmatrix}
1-a & a \\
 1-a & a \\
 b & 1-b
\end{pmatrix}
$$
Here a method is developed for studying the response process by introducing a Markov Chain. If there are two stimulus elements, then there will be 3 possible states, say 2,1,0 indicating the number of stimulus elements connected to an $A_1$ response. In this manner, the transition probability matrix was obtained as, say,

$$
P = \begin{pmatrix}
2 & 1 & 0 \\
2(1-\theta^2)a + (1-a) & 2\theta(1-\theta)a & \theta^2a \\
\frac{1}{2}\theta^2(1-a) + \frac{1}{2}\theta(2-\theta)b & (1-\theta^2) + \theta(1-\theta)(1-a) + \theta(1-\theta)(1-b) & \frac{1}{2}\theta^2(1-b) + \frac{1}{2}\theta(2-\theta)a \\
\theta^2b & 2\theta(1-\theta)b & (1-\theta^2)b + (1-b)
\end{pmatrix}
$$

Importance of grading pattern which is the main subject of our study may be actually understood if we go through the available materials on grading patterns as studied by different scholars. Here we present a glimpse of several works on grading patterns presented by different authors.

According to Valen, E. Johnson, (1997), who seriously and excitedly attempted research on grading patterns, G.P.A. or the grade point average is the most widely used summary of the undergraduate student performance in American educational system. But since G.P.A. is obtained through simple average schemes, it rewards students for selecting less challenging courses. Thus as a result there is substantial reduction of students taking up subjects like mathematics, natural sciences and other upper level undergraduate courses.
Again inflating grades is a reasonable strategy for faculty members, especially junior faculty. Because, as a consequence of assigning higher than average grades, course enrolments increase, student complaints are minimised and also salary increases, promotions and tenure decisions are often tied to student course evaluation. So there seems to be little faculty incentives for not inflating grades.

To remove the disincentives for learning caused by differential grade assignment, a number of alternate measures for student performance have been proposed in the educational literature. Here we discuss some of these including a Bayesian model for student assessment proposed by V. Johnson (1997), which incorporates essential ideas for other methods available. This model represents a variation of the Bayesian ordinal data proposed in Albert and Chib (1993), Johnson (1996) and Cowles, Carlin and Connett (1996) and may be regarded as a Bayesian extension of the graded response models. By adopting this alternate measure of student performance, penalties imposed on students for taking challenging courses can be eliminated.

A brief review of these proposals are given below.

**Pairwise Comparative Method**

Goldman and Widawski (1976) proposed a grade adjustment method based on pairwise comparisons of grades obtained by the same students
across the multiple departments. According to them, the differences in students' grades for classes taken in different departments provide information about the relative grading standards between the departments. They averaged all such differences obtained from the transcripts of 475 students of University of California, to obtain the grade adjustment factors for 17 academic fields. From this analysis they concluded that, the departments with high ability students tended to grade more stringently than fields with less able students.

Strenta and Elliot (1987), Elliot and Strenta (1988) in their studies of Dartmouth College of undergraduates, confirmed Goldman and Widawski's results of stable trend in differential department grading standards. They also incorporated both within and between department course comparisons and estimated grade adjustment factors for a large number of departments. And the resulting indices produced adjusted GPA measures that correlates more strongly with both student ability test (SAT) and high school GPA, than did the standard GPA.

**Graded Response Method**

J.W. Young (1989) in his doctoral thesis adapted a model derived from Item Response Theory (IRT; e.g. Lord and Novick, 1968) called the graded response model (GRM; Samejima, 1969) for applications to undergraduate grade data. There it was assumed that,
$k = \text{Total number of grades which can be assigned to students, which are ordered from 1 to } k.$

$Y_{ij} = \text{the grade assigned to the student } i \text{ in the } j^{th} \text{ class.}$

$\beta_i = i^{th} \text{ student's ability}$

$\eta_j = \text{discrimination parameter of the } j^{th} \text{ class grade}$

$\xi_{jk} = \text{upper grade cut offs for grade } k \text{ in class } j.$

With the help of these notations, the basic assumptions of the Graded Response Model is that,

$$P_r \{ Y_g \leq k \} = \pi_{yk} = \frac{\exp[\eta_j(\beta_i - \xi_{jk})]}{1 + \exp[\eta_j(\beta_i - \xi_{jk})]}$$

Young applied this model to a cohort of Stanford undergraduate grades and found that the estimates of student abilities obtained using this model correlates better with external measures of student ability than did raw GPA (Young, 1990, 1993)
Regression Models

Recently Larkey and colleagues investigated some linear techniques for adjusting student GPA's to account for the difficulty of the courses taken (Caulkin, Larkey and Wei, 1996; Larkey and Caulkin, 1992; Larkey, 1991; Young 1992).

Here an additive adjustment is made to each student's GPA based on estimates of the difficulty of the students courses. The difficulty of the courses is estimated from a linear regression of student grades on true student GPA and course difficulty parameters.

Assuming $Y_{ij}$ to be the grade of the $i^{th}$ student in the $j^{th}$ class and $g_i$ and $c_j$ be the $i^{th}$ student's true GPA and the difficulty of the $j^{th}$ class; the additive model to estimate these adjustment factors takes the form

$$Y_{ij} = g_i - c_j + e_{ij}$$

Where, $e_{ij}$ is the error term which is normally distributed with 0 mean.

Caulkin, Larkey and Wie found that predictions obtained using additive adjustment model produced estimates of student performance that correlates more highly with high school GPA and SAT scores than did, estimates obtained using GRM.
A Bayesian Model for Graded Data

The proposed model begins with the assumption that an instructor assigns grades by first ordering perceived student performance from best to worst, possibly with ties for groups of students who performed at the same level approximately.

After ordering students according to their estimated classroom achievement, instructors next group students into grade categories by fixing grade cut offs between the estimated achievement levels of students in the class. The following variables are used to model this mechanism for grade generation:

1. The variable $Y_{ij}$ denotes the grade assigned to the $i^{th}$ student in the $j^{th}$ class. In general, there are $k$ possible grades ordered from 1 to $k$.

2. The variable $X_i$ represents the mean classroom achievement of the $i^{th}$ student, in classes selected by student $i$. This variable is called the achievement index of student $i$.

3. Grade cutoffs for class $j$ are denoted by $\gamma^j_0, \gamma^j_1, \ldots, \gamma^j_k$.

4. Random variation in the performance of student $i$ in class $j$ is denoted by $\varepsilon_{iy}$. This term accounts for the fact that student achievement varies from class to class and that instructor assessment of student achievement is also
subject to error. It is assumed that $\varepsilon_{ij}$ is Gaussian with 0 mean and variance $\sigma^2_{ij}$.

Also it is assumed that,

$$\sigma^2_y = \sigma^2_{n(j)}, \text{ independent of } i.$$ 

With these notations, the model for grade generation may be summarised as follows.

Student $i$ gets a grade of $Y_{ij} = k$ in class $j$ if and only if,

$$\gamma_{k-1}' < X_i + \varepsilon_{ij} < \gamma_k'$$

The prior distributions used in this model are,

$$X_i \sim N(0,1)$$

$$\varepsilon_{ij} \sim N(0, \sigma^2_{n(j)})$$

$$\sigma^2_{n(j)} \sim IG(\alpha, \lambda) \propto (\sigma^2_{n(j)})^{-(\alpha+1)} \exp\left(-\frac{\lambda}{\sigma^2_{n(j)}}\right)$$

Besides these several models have been tried by several authors viz. Gilks, Richardson and Speigelhalter (1996) who tried posterior simulation by Markov Chain Monte Carlo technique. Tanner and Wong (1987) proposed a Bayesian data augmentation scheme which was exploited by several authors later

1.5 Important theorems, definitions and results used in this thesis

Stochastic Process

A stochastic process \( \{x(t); \ t \in T\} \) is a collection of random variables. That is for each \( t \in T \), \( X(t) \) is a random variable. That index \( t \) is often interpreted as a time and as a result, we refer to \( X(t) \) as the state of the process at time \( t \). The set \( T \) is called the index set of the process. When \( T \) is a countable set, the stochastic process is said to be discrete time process. If \( T \) is an interval of the real line, the stochastic process is said to be a continuous time process.

The state space of the stochastic process is defined as the set of all possible values that the random variable \( X(t) \) can assume.

Markov Chain

Consider a system that can be in any one the finite or countably infinite number of states. Let \$ \$ denote this set of states and assume that \$ \$ is a subset of integers. The set \$ \$ is called the state space of the system.
The Markov property states that given the present state, the past states have no influence on the future. A stochastic process having this property is said to be a Markov Chain.

That is a stochastic process \( \{X_n, n \in \mathbb{N}\} \) is called a Markov Chain provided that,

\[
P_r\{X_{n+1} = j \mid X_0 = i_0, X_1 = i_1, \ldots, X_n = i\}
\]

\[
= \Pr\{X_{n+1} = j \mid X_n = i\}
\]

\[
= p_{ij}; i, j \in E,
\]

Where \( E \) is a finite set and \( \mathbb{N} \) is the set of non-negative integers.

The probabilities \( p_{ij} \) are called the transition probabilities, which are basic for the study of the Markov Chain. These transition probabilities \( p_{ij} \) satisfy,

\[
p_{ij} \geq 0, \sum_j p_{ij} = 1 \quad \forall \; i
\]

These probabilities may be written in the matrix form
A state $j$ of a Markov Chain is said to be persistent or recurrent, if the return to the state $j$ is certain and transient if the return to the state $j$ is uncertain.

Again a persistent state is said to be null persistent, if the mean recurrence time is infinite, and is said to be non-null persistent if the mean recurrence time is finite.

**Markov Process**

The stochastic process $X = \{X_t, t \in \mathbb{R}_+\}$ is said to be a Markov process with state space $\mathcal{E}$ provided that for any $t, s \geq 0$ and $j \in \mathcal{E}$,

$$
\Pr\{X_{t+s} = j \mid X_u, u \leq t\} = \Pr\{X_{t+s} = j \mid X_t\} \quad \cdots (1.4.1)
$$

The conditional probability appearing in (1.4.1) may in general, depend on both $t$ and $s$ (in addition to $j$ and the value of $X_t$) when

$$
\Pr\{X_{t+s} = j \mid X_t = i\} = P_s(i, j)
$$
Is independent of $t \geq 0$ for all $i, j \in E$ and $s \geq 0$, the process $X$ is said to be a time homogeneous Markov Process.

**Poisson Process**

A stochastic process $\{X_t, t \geq 0\}$ is said to be a counting process, if $X_t$ represents the total number of "events" that have occurred up to time $t$. One of the most important counting processes is the Poisson process which is defined as follows.

The counting process $\{X_t, t \geq 0\}$ is said to be a Poisson process having rate $\lambda, \lambda > 0$, if

(i) $X_0 = 0$

(ii) The process has independent increments,

(iii) The number of events in any interval of length $t$ is Poisson distributed with mean $t$. That is, for all $s, t \geq 0$

$$P \{X_{t+s} - X_s = n\} = \frac{e^{-\lambda t} (\lambda t)^n}{n!}, \quad n = 0, 1, 2, \ldots \ldots \ldots \ldots$$

**Compound Poisson Process**

A stochastic process $\{X_t, t \geq 0\}$ is said to be a Compound Poisson process if it can be represented as,
\[ X_t = \sum_{i=1}^{N_t} Y_i, \quad t \geq 0 \]

Where, \( \{N_t, \ t \geq 0\} \) is a Poisson process, and \( \{Y_n, \ n \geq 0\} \) is a family of independently and identically distributed random variables which are also independent of \( \{N_t, \ t \geq 0\} \)

Random Walk: Consider a Markov Chain whose state space consists of the integers \( i = 0, \pm 1, \pm 2, \ldots \) and have the transition probabilities given by,

\[ p_{i,i+1} = p = 1 - p_{i,i-1}; \quad i = 0, \pm 1, \pm 2, \ldots \]

Where, \( 0 \leq p \leq 1 \). In other words, on each transition, the process either moves one step to the right with probability \( p \) or one step to the left with probability \( 1 - p \).

Laplace Transformation

Suppose \( F(t) \) is a real valued function defined over the interval \((-\infty, \infty)\), such that, \( F(t) = 0 \ \forall \ t < 0 \). The Laplace transform of \( F(t) \), denoted by \( L\{F(t)\} \) is defined as,

\[ L\{F(t)\} = \int_{0}^{\infty} e^{-st} F(t) \, dt = f(s) \]
Inverse Laplace Transformation

If the Laplace transform of the functions \( F(t) \) is \( f(s) \), i.e., \( L\{F(t)\} = f(s) \), then \( F(t) \) is called the inverse Laplace transform of \( f(s) \). We write it as, \( F(t) = L^{-1}\{f(s)\} \)

Initial Value Theorem

If \( L\{F(t)\} = f(s) \), then,

\[
\lim_{t \to \infty} F(t) = \lim_{s \to 0} s f(s),
\]

Wherever the limit exists.

Discrete Uniform Distribution

A random variable \( X \) is said to have a discrete uniform distribution, over the range \([1, N]\), if its p.m.f is given by,

\[
P\{X = x\} = \begin{cases} 
\frac{1}{N} & \text{for } x = 1, 2, \ldots, N, \\
0, & \text{otherwise}
\end{cases} \quad \ldots \ (1.4.2)
\]

\( N \) is the parameter of this distribution and lies in the set of all positive integers.
Multinomial Distribution

Suppose an experiment results in k possible outcomes \( O_1, O_2, \ldots, O_k \) with probabilities \( p_1, p_2, \ldots, p_k \) respectively, \( \sum_{i=1}^{k} p_i = 1 \). N such experiments are performed. Let \( X_i \) be the number of experiments that result in outcome \( O_i \) (\( i = 1, 2, \ldots, k \)) \( \sum_{i=1}^{k} X_i = N \) then,

\[
\Pr \{ X_1 = x_1, X_2 = x_2, \ldots, X_k = x_k \} = P(x_1, x_2, \ldots, x_k) = \frac{N!}{x_1! \cdots x_k!} \prod_{i=1}^{k} p_i^{x_i} \quad \ldots \quad (1.4.3)
\]

Where, \( x_i = 0, 1, 2, \ldots, N \) and \( \sum_{i=1}^{k} X_i = N \)

Binomial Distribution

When \( k = 2 \) in (1.4.3) we get the p.m.f for binomial distribution.

Trinomial Distribution

When \( k = 3 \) in (1.4.3) we get the p.m.f. for trinomial distribution.
Hypergeometric Distribution

A random variable $X$ is said to have a hypergeometric distribution with parameter $N$, $M$ and $n$, if its p.m.f. is given by,

$$P (X = x) = H (x; N, M, n)$$

$$= \begin{cases} \frac{\binom{M}{x} \binom{N-M}{n-x}}{\binom{N}{n}} & \text{for max (0,n-N+M) } \leq x \leq \min (M,n) \\ 0, & \text{otherwise} \end{cases}$$

Where $N$ is a positive integer, $M$ is a positive integer not exceeding $N$ and $n$ is a positive integer that is at most $N$.

Beta Distribution of first type

A random variable $X$ is said to have a beta distribution of first type with parameters $a$, $b$ ($a>0$, $b>0$) if its p.m.f. is given by

$$f (x, a, b) = \begin{cases} \frac{1}{\beta(a,b)} x^{a-1} (1-x)^{b-1}, & 0 < x < 1 \\ 0, & \text{otherwise} \end{cases}$$
Baye’s Estimators

In case of point estimation problem we assume that our random sample has come from a density \( f(x,\theta) \) and \( \theta \) is fixed. In some real world situation there is often additional information about \( \theta \). For example in our present study we have evidence that the transition probability \( p_{ij} \) (\( i, j = 0, 1, 2 \)) itself acts as a random variable, for which we are able to present a realistic density function. What is important here is, how this additional information about \( \theta \) of the density function \( f(x,\theta) \) may be used to estimate \( \theta \). We may assume that the unknown parameter \( \theta \) is the value of some random variable \( H \)

Prior and Posterior Distribution

The density \( g(\theta) \) of the random variable \( H \) is called the prior distribution of \( \theta \).

Again the conditional density of \( H \) given \( X_1 = x_1, X_2 = x_2, \ldots, X_n = x_n \) denoted by \( f(\theta | H) \) is called the posterior distribution of \( H \) and is denoted by \( f(\theta | x) \), given by,

\[
f(\theta | x) = \frac{f(x_1, x_2, \ldots, x_n | \theta) g(\theta)}{f(x_1, x_2, \ldots, x_n)}
\]

Again, if \( \int g(\theta) d\theta \neq 1 \),
then $g(\theta)$ is called the improper prior distribution of $\theta$.

**Posterior Baye’s estimator**

Let $X_1, X_2, \ldots, X_n$ be a random sample from the density function $f(x \mid \theta)$, where $\theta$ is the value of the random variable $\Theta$ with known density $g(\theta)$. The posterior Baye’s estimator $\tau(\theta)$ with respect to prior $g(\theta)$ is defined to be,

$$E(\tau(\Theta) \mid x_1, x_2, \ldots, x_n)$$

$$= \int \tau(\theta) f(\theta \mid x) \, d\theta$$

$$= \frac{\int \tau(\theta) \prod f(x_i \mid \theta) g(\theta) \, d\theta}{\int \prod f(x_i \mid \theta) g(\theta) \, d\theta}$$