Chapter-1

Overview of the Project

§ 1.1 Introduction:

The works done under the project entitled "Probabilistic Forecasting of Time Series" can be classified into three broad areas. They are

(1) the works done in the field of the theory of probability,

(2) the works done in the field of probabilistic projection of population in India

and (3) the works done in the field of probabilistic forecasting of temperature and rain fall in India.

A picture of the works done in each of these three areas has been presented below. Background and overview of the works done in this study have been outlined in each of the pictures presented below. Also, the notable findings obtained here have been mentioned in the respective presentations.

§ 1.2 Overview of the Works in the Field of Theory of probability:

The theory of probability, available in many standard literatures (***Ref.*** 1, 2, 9, 13, 52, 88, 90, 93, 128 & 130) has been found to be a major player in playing the role of understanding various phenomena in almost every branch of science (***Ref.*** 97). The history of the development of probability theory can be classified in to five broad stages viz.

(1) *The Prehistory of Probability Theory*, whose beginnings were lost in the dust of antiquity and which had been terminated in the sixteenth century with the works of *Cardano* (***Ref.*** 17 & 94), *Galileo* (***Ref.*** 83), (*Paccioli* (***Ref.*** 81 & 84), *Tartaglia* (***Ref.*** 120) and many others,

(2) *The Origin of Probability Theory as a Science*, continued from the middle of the seventeenth century up to the beginning of the eighteenth century with the works of *Pascal* (***Ref.*** 95 & 98), *Fermat* (***Ref.*** 81 & 82), *Huygens* (***Ref.*** 57) etc.,

(3) *Bernoullian Era of Probability Theory*, begun with the appearance of *James Bernoulli's* treatise " **Arts Conjectandi (The Act of Conjecturing)** " in 1713 (***Ref.*** 5, 6, 35, 53 & 118) and continued up to the early period of the nineteenth century with the works of *d'Alembert* (***Ref.*** 36, 37, 38 & 39),

deMoiver (*Ref.* 40, 41, 42 & 116), *Laplace* (*Ref.* 72, 73, 74 & 116), *Todhunter* (*Ref.* 121) etc.,

(4) *Probability Theory in Russian School*, continued from the middle part of the nineteenth century up to the first quarter of the 20th century mainly with the works of *Chebyshev* (*Ref.* 25, 26, 27, 28, 29, 30 & 80), *Markov* (*Ref.* 4, 86, 87 & 96), *Lyapounov* (*Ref.* 76, 77, 78, 79 & 80) and many others

and (5) *The Modern Period*, begun with the formulations of axioms mainly by *Bernstein* (*Ref.* 7 & 8) and by *Kolmogorov* (*Ref.* 67, 68 & 69) based on the initial contributions of this period made by *Bernstein* and continuing till now with the works of *von Mises* (*Ref.* 51, 65, 123, 124, 125 & 126), *Borel* (*Ref.* 14, 15 & 16) *Khinchin* (*Ref.* 65 & 66), *Doob* (*Ref.* 43) and many others (*Ref.* 34, 56, 59, 84, 114 & 115).

The three significant approaches, other than the two unscientific ones viz. **Intuitive Approach** (*Ref.* 70, 71, 111 &112) and **Subjective Approach** (*Ref.* 3) through which the theory of probability has been brought to the current stage of its development are

(1) **Classical Approach** introduced by *Bernoulli* (*Ref.* 5, 6, 35, 53 & 118),

(2) **Empirical Approach** developed by *von Mises* (*Ref.* 51, 52, 65, 123, 124, 125 & 126) and by *Fisher* (*Ref.* 46 & 48)

and (3) **Axiomatic Approach** introduced by *Bernstein* (*Ref.* 7 & 8) and by *Kolmogorov* (*Ref.* 67, 68 & 69).

The intuitive approach of probability is based on inductive reasoning and judgment (*Ref.* 45, 70, & 71). The intuitive approach is insufficient for scientific purpose (*Ref.* 45, & 71). A popular approach to statistics, inspiring biter quarrels, is the *Bayesian* approach (*Ref.* 3), which makes probability subjective.

The empirical approach to probability has been developed mainly by *Fisher* and by *von Mises*. The notion of sample space, connected with probability, comes from *von Mises*. The notion makes it possible to build up a strictly mathematical theory of probability based on measure theory (*Ref.* 11, 54, 89 & 104). The axiomatic approach to probability was introduced first by *Bernstein* in 1927 and then by *Kolmogorov* in 1933. The problem of an axiomatic construction of probability theory was first posed and solved by *Bernstein*. *Kolmogorov* considered a different approach, to the axiomatic treatment to probability, which clearly relates the probability theory to the set theory and with the metric theory. Axiomatic approach is too abstract and too general to be useful (*Ref.* 45). Prior to these developments the theory of probability was based, for several centuries, on a fundamental definition known as the classical definition of probability, which had been introduced as a consequence of the principle of insufficient reason by *Bernoulli* in 1713.

The use of common sense and engineering and scientific observations leads to a definition of probability, based on the idea of statistical regularity (*Ref.* 32, 33 & 62), as a stable value of the relative frequency of occurrence of some event (*Ref.* 101). This definition of probability is nothing but its empirical definition (also known as the statistical definition) due to *Fisher*. However, probability can never be calculated exactly by the empirical approach while the axiomatic approach to probability cannot say how the probability of an event is to be calculated (*Ref.* 10). Again, the classical definition of probability is not valid when the possible outcomes of a random experiment are not equally likely (*Ref.* 45). These lead to think of searching for some approach of defining the probability that can help in determining the exact value of the probability of an event associated to a random experiment. In the present study, an attempt has been made to search for an approach of defining probability that can be a basis of searching for method / methods of determining the exact value of the probability of an event. It has been found that the probability of an event can be defined theoretically in terms of the number of occurrence of the event among a number of repeated trials. It also has been found that the basic properties of probability, already established, can be derived from this definition. In this context, the next attempt has been made on searching for a method of determining the value of the probability of an event. It has been found that there exists method by which the value of the probability of an event associated to a random experiment, which is still supposed to be undeterminable, can be determined.

Empirical approach provides the way for determining the value of the probability of an event with the help of actual experimentation while the classical approach provides the way for determining this value a priory without the actual experimentation. This implies that the empirical approach is dependent on the actual experimentation while the classical approach is free from it. The empirical approach has the least contribution to the theoretical development of probability. The basic concept by which the theory of probability has been developed either through the classical approach or through the axiomatic approach is the concept of sample space due to *von Mises* as already mentioned. By the definition of sample space due to *von Mises*, the sample points of the sample space are mutually exclusive and exhaustive. However, they may or may not be equally likely. The classical definition of probability, introduced by *Bernoulli*, holds good if and only if the sample points are equally likely. In the situation where the sample points are not equally likely, one cannot define the probability of an event with the help of sample space. To overcome this trouble an attempt has been made to extend the classical definition of probability introduced by *Bernoulli* to the situation where the possible outcomes are not equally likely. It has been found that probability can also be defined in the sense of classical approach in the situation where / when the possible outcomes of a random experiment are not equally likely. Also, another attempt has been made to search for if there exists any link between the two definitions of probability viz. the classical definition and the empirical definition. In this context, it has been found that the classical definition of probability due to *Bernoulli* is a consequence of its empirical definition due to *Fisher* under the special situation when / where the possible outcomes of a random experiment are equally likely.

§ 1.3 Overview of the Works in the Field of Probabilistic Projection of Population in India:

The theory of probability, which is the backbone of the theory of statistics, has been found to be a major player in playing the role of understanding various phenomena in almost every branch of science. The first application of probability theory was its application in population studies in the middle of the seventeenth century (***Ref.*** 84).

The research on population theory was first done, with the application of probability theory, during the second stage viz. *The Origins of Probability Theory as a Science* in the history of the development of probability theory that had been continued from the middle of the seventeenth century up to the beginning of eighteenth century (***Ref.*** 84). The modern theory of population, which has its base on the works of *Malthus* (***Ref.*** 85), has been brought to the current stage of its development through the contributions of many pioneers like *Lotka* (***Ref.*** 75), *Pearl* (***Ref.*** 99 & 100), *Feldman* and *Roughgarden* (***Ref.*** 44), *Rhodes* (***Ref.*** 110), *Tuckwell* (***Ref.*** 122), *Prajneshu* (***Ref.***

107 & 108), *Chakrabarty* (*Ref.* 20, 22 & 24), *Chakrabarty* and *Baruah* (*Ref.* 18, 19 & 21) and many others which are available in many standard literatures (*Ref.* 63, 64 & 106). The works, on the theory of population, done till the end of the ninth decade of the nineteenth century correspond to point estimations and point projections on various characteristics of a population in a region. The projected figures obtained by the above attempts may not be identical with the actual figures but accidental to be identical with the actual ones. This leads to think of searching for means of obtaining projected intervals on various characteristics of a population in a region within which the corresponding true values lie. *Chakrabarty* and *Baruah* (*Ref.* 23) have innovated a method of determining projected interval on the total population in a region. They have innovated a method of obtaining projected interval for total population of a region within which the corresponding actual figure lies if the change in the concerned population is not influenced by any irregular factor/factors.

The study on population estimation / projection, done so far, is free from probabilistic idea. Projection is, however, probabilistic in reality. Historically, the original purpose of the theory of probability was to describe the exceedingly narrow domain of chance and the main effort was directed to the calculation of certain probabilities (*Ref.* 45). One purpose of the theory of probability is it described and predict of various events associated to random phenomena (*Ref.* 45). In the current study, probabilistic idea has been introduced in the study of estimation / projection of population. The study, done here, is basically based on the stability property of relative frequency. Stability property of relative frequency that is the basis of determining the value of the probability of an event associated to a random experiment in the empirical approach to probability has been applied to determine the probability that an individual in a region if selected belongs to a particular subregion of it. With the help of the value of the probability of such type of event corresponding to each of the states in India the number of persons in each of them has been projected for the years 2011, 2021 & 2031. It has been found that the relative frequency approach of probability can yield the estimates of total populations of India and of its states, which are statistically not rejectable if the populations figures are considered in thousands or in higher units and accordingly the corresponding projected populations determined are not statistically rejectable if the population figures are considered in the same units. It has also been found that the application of the relative frequency approach to probability yields estimated / projected figures on the populations of India and of its states age wise, sex wise and age-sex wise which are not statistically rejectable. Of course, total population of the whole region (India in the instant case) cannot be projected by this approach. This creates a necessity of searching for some method of projecting point value on total population of the whole region.

The theory of population due to *Malthus* states that population increases in geometric progression while subsistence increases in arithmetic progression. Though this theory suffers from some shortcomings, one need to think that it has significant contribution to the researchers in the respective field since it is the root source of further thinking for the researchers. In the current study, it has been thought of that the changes in total population over intervals (of some length) of time can be represented by an arithmetic progression. With the help of this thinking, a method of projecting point value on the total population in a region has been framed of. The method, framed of, has been applied to project point values on the total populations of India till 2031. It has been found that the change in the total population of India over time, as the pattern since 1951 tells, can be represented by an arithmetic progression if the said change is considered over interval of time and the length of the interval is taken as (1/5.6) year.

In the method of determining projected interval on the total population in a region innovated by *Chakrabarty* and *Baruah*, projected interval is determined from the underestimates and overestimates of the same. The underestimates and overestimates, used there, suffer from error. It has been thought of that by eliminating the errors involved there, it might be possible to obtain projected point value on the same. Hence, due to the necessity of some method of projecting point value on total population of the whole region, another attempt has been made to innovate a method of projecting point value on total population in the whole region based on underestimates and overestimates of the same that can be determined by the techniques developed by *Chakrabarty* and *Baruah* and that have been used by them in determining interval projection on total population. It has been found that projected point value on total population can be obtained if a set of underestimates and a set of overestimates are known or can be determined. The method innovated has been applied to project point values on total population of India for the years 2011, 2021 & 2031.

It has been found that the method of interval population projection, innovated by *Chakrabarty* and *Baruah*, has yielded acceptable results in the context of total population of India (*Ref.* 23). In the current study, therefore, projected intervals on the total population of India for the years 2011, 2021 & 2031 have been computed by applying this method using the data from 1951 to 2001

§ 1.4 Overview of the Works in the Field of Probabilistic Forecasting of Temperature and Rainfall in India:

Like the study on population projecting, the study on the forecasting of two other time series viz. temperature and rainfall, done so far, is also free from probabilistic idea. There is no certainty that the forecasted results on these two components of weather will be identical with the actual ones. Thus, it would be more realistic if one thinks of determining forecasted results on these two components in terms of probability. In this study, attempt has been made to do so. Four, among many others, vital characteristics of temperature in a region are

(1) Mean Maximum Temperature,

(2) Highest Maximum Temperature,

(3) Mean Minimum Temperature

and (4) Lowest Minimum Temperature.

Similarly, three, among many others, vital characteristics of rainfall in a region are

(1) Total Rainfall,

(2) Heaviest 24 Hours Rainfall in the Month

and (3) Number of Rainy Days in the Month.

The discovery of normal probability distribution discovered by *Gauss*, available in many standard literatures (*Ref.* 50, 60, 115 & 117), is the most significant discovery in the theory of statistics. It has been thought of that it may be possible to apply the area property of normal distribution in developing the literature on how to know whether there exists any significance assignable cause in a region which forces the temperature of the region to be changed as well as on how to determine forecasted interval value with desired probability (i.e. with desired confidence). Similar is the case for rainfall also. In this study, an attempt has been made to develop literature how to know whether there exists any significance assignable cause in a region which forces the temperature of the region to be changed and how to determine forecasted interval with desired confidence. It has been found that that it is possible to apply the area property of normal distribution to know whether there exists any significance assignable cause in a region which forces the temperature (also the rainfall) of the region to be changed as well as to determine forecasted interval value on various characteristics of temperature (also of rainfall) with desired probability. Then the literature developed has been applied to analyze the data on the characteristics of temperature mentioned above in the context of India to know whether there exists any significance assignable cause in a India which forces the temperature of the nation to be changed and then to determine forecasted interval value with specified probability confidence. Similar study has been done for rainfall in the context of India also. It can be recommended on the basis of the findings that there is no any significant cause that influences upon the changes in temperature, in India, over years i.e. temperature in India has not been changing (since 1969) over years significantly. The changes in temperature, occurred since 1969, over years are due to the chance (random) causes only. It has been possible to determine confidence intervals (99% confidence intervals and 99.73% confidence intervals have been computed) on each of the six characteristics viz.

(1) Mean Maximum Temperature (Monthly),
(2) Highest Maximum Temperature (Monthly),
(3) Mean Minimum Temperature (Monthly),
(4) Lowest Minimum Temperature (Monthly),
(5) Heaviest 24 Hours Rainfall (Monthly)
(6) Total Rainfall (yearly)
& (7) Number of Rainy Days (Monthly).

in the context of India for each of the twelve months in a year.

The analysis of variance, discovered by *Fisher* (***Ref.*** 31, 49, 102, 103 & 113), is a statistical tool that can be used to know whether there exists any significance assignable cause in a region which forces the temperature of the region to be changed. This tool can also be applied to determine forecasted interval value with desired probability. Similar is the case for rainfall also. In this study, this statistical tool has been applied to analyze the temperature in the context of India to know whether there exists any significance assignable cause in a region which forces the temperature of the region to be changed and then to determine forecasted interval value with specified probability. Similar study has been done for rainfall in India also. It has been found that variation in each of

(1) Mean Maximum Temperature (Monthly),

(2) Highest Maximum Temperature (Monthly),

(3) Mean Minimum Temperature (Monthly),

(4) Lowest Minimum Temperature (Monthly),

(5) Heaviest 24 Hours Rainfall (Monthly)

& (6) Number of Rainy Days (Monthly).

has been found to be significant among the months.

In order to obtain a picture on temperature and rainfall in India, some locations are to be selected in such a way that the selection locations together can cover India on the whole. Accordingly, 42 locations (called stations in the language of meteorological science) have been selected. The 42 stations are

AGARTALA, AHMADABAD, ALLAHABAD, AMRITSAR, BAIRAGARH, BANGALORE, BHUBANESHWAR, BHUNTER(A), BHOPAL, BOMBAY/SANTACRUZ, CALCUTTA, CHANDIGARH(A), CHANDIGARH, DEHRA DUN, DHUBRI, DIBRUGARH/MOHANBAR, GUWAHATI/BHORJOR, HISSAR, HYDARABAD(A), IMPHAL/TULIHAL, JAIPUR/SANGANER, JAMMU(A), KOHIMA, LUCKNOW/AMAUSI, MADRAJ/MINAMBAKKAM, NAGPUR/SONEGAON, NEW DELHI/SAFDARJU, PALAM(A), PANJIM, PATIALA, PATNA(A), PONDICHERRY, PORT BLAIR, PUNE, SHILLONG(A), SHILLONG/C.S.O., SILCHAR, SIMLA, SRINAGAR(A), TEZPUR, TRIVANDRUM/THIRUVA, UDAIPUR(DABOK), VARANASI/BABATPUR.

The study on temperature and rainfall in the context of India has been carried out on the basis of the data on the characteristics of temperature and rainfall mentioned above from these 42 stations only.