In this issue

Statistics

The special section in this issue of *Current Science* contains review articles, not too technical, of recent advances in the application of statistics and probability to a variety problems in science, business and industry. Ideas of probability and statistical inference have been in use in scientific research for a fairly long time now and the early applications to demography, actuarial science, econometrics, to agriculture through design and analysis of experiments, to control of quality at manufacturing stage and to other sciences are well documented and readers may be familiar with some of these. This issue concerns relatively recent applications.

J. K. Ghosh and T. Samanta (**page 1135**) begin with the very important primary problem, that of selecting a model, for the various nondeterministic aspects in complex scientific experiments or observational studies. Selection of an appropriate model is a very important problem for the statistical analysis of any observational data. This article surveys the recent advances in this area, but assumes no previous knowledge of the readers in the area of model selection. Six examples and the analysis of some of these examples are presented in this article, in addition to some theory of model selection.

S. Ghosh and P. P. Majumdar (**page 1145**) deal with the advances in the genetic mapping of quantitative traits in human beings. Traits that are highly heritable are generally determined by genetic factors, with environmental factors playing relatively an unimportant role. Genes underlying such traits, known as major genes may vary from one to a few. The problem of determining the location of such major genes on chromosomes is very interesting and very important. In this article the authors explain how statistical methods are used in solving this problem. The performance of these methods has also been examined using computer simulations and it has been demonstrated that they perform efficiently under a wide variety of scenarios.

P. Chaudhuri and S. Das (**page 1161**) discuss how large DNA sequences can be statistically analysed using the distribution of DNA words. It is shown in this article that the DNA word frequencies are simple, yet effective statistical tools to capture information about structural patterns. Moreover they can reveal biologically significant features in a DNA sequence. The analysis presented in this article also demonstrates the use of the distribution of DNA words in large DNA data in detecting the structural signature of a genome and in identifying phylogenetic relationships among different species reflected in the variation of word distributions in their DNA sequences.

P. K. Sen (**page 1167**) presents a survey of current ideas on statistical approaches to assessing levels of toxicity, especially in the environment. Toxicology relates to the study of the intake process of the toxins which abound in nature, environment and in our modern lifestyle, their mode of propagation, biological reactions, molecular level of penetration, genotoxicity and aftermaths. At each of these phases there is a genuine need for statistical appraisal. These statistical perspectives are brought out in this article along with an outline of recent statistical approaches to this assessment task.

R. L. Karandikar (**page 1176**) discusses the current advances in the use of probability and statistics for financial option pricing. This is a timely article as option trading has been allowed in India recently and trading in options is about to begin in Indian stock markets. This article gives an introduction to options, the theory of its pricing and its interplay with probability theory. An example is used to present the ideas which are then formalized in the context of a discrete model. The article concludes with a discussion on continuous time models. The readers would have benefited to a greater extent, had a more elaborate list of references or bibliography been added.

M. Mazumdar (**page 1183**) discusses the application of probability models and statistical methods to analyse the cost of production in generating electric power. Because electric power cannot be stored, and the particular set of units that are used to supply power at any given point of time depend upon the magnitude of demand and the availability of generators, both of which are random, the cost of electric power is a random variable. Computing the basic parameters of its distribution is difficult because of the large state space. He discusses these difficulties and presents several useful approximation methods for estimating the parameters.

The article by J. V. Deshpande and S. G. Purohit (**page 1191**) is semitechnical in nature and deals with the problems of modelling and estimating the lifetime of a unit. The article reviews and summarizes several statistical procedures for modelling and its validation for the lifetime of a unit. The article studies the probability laws that govern the distribution of lifetime and the effects of various covariates on these laws.

The seven articles in this issue thus give a good idea of the widespread use of the concepts of probability and statistics, the current advances and the issues in various areas. It is hoped these articles will motivate the readers to learn more about the use of these concepts.

S. R. Mohan
T. Parthasarathy

GIS and malaria

Geographic Information System (GIS), an effective spatial and temporal analyses tool is designed to capture, store, manipulate, analyse and visualize the diverse sets of georeferenced data that are required to support accurate modelling of the earth's environmental processes. In this regard, remote sensing greatly enhances the ability to generate temporal data and GIS can analyse landscape-level relationships. In spite of productive use of GIS in diverse sectors (e.g. agriculture, natural resources, urban and regional planning, etc.) in India, the health sector has not fully explored its potential utility in research or in the control of disease.

Vector-borne diseases require an intermediate living agent for their transmission. They are a heavy burden on human populations, a major cause of work loss and a serious impediment to economic development and productivity. Their epidemiology is influenced by attributes of their vectors, which in turn are closely linked to environmental conditions.

Malaria is a serious vector-borne disease affecting a greater proportion of the world's population than any other vector-transmitted disease. Over 300 million cases of malaria occur every year, and the number is increasing. Even regions where malaria had been controlled are now suffering again from this significant public health problem. Malaria was a major problem in India contributing 75 million cases, with 0.8 million deaths every year prior to the launch of the National Malaria Control Programme in 1953. After a significant decline in the 60s, malaria reemerged as a major health problem of India in the late 70s and presently it poses a major challenge with 2 to 2.5 million cases annually. According to recent data of National Malaria Eradication Programme (NMEP) about 50% of complicated malaria leads to mortality, if timely treatment is not given.

Over the past decades, the increased demands on the landscape for food and shelter and an increased number of by-products of man's living environment have led to unparalleled changes. Some of these changes have led to an increase in the distribution of several vector-borne diseases, including malaria.

The increase in prevalence of malaria is determined by several factors: mosquito resistance to insecticides, parasite resistance to drugs, changes in land-use patterns and reduction in funding and manpower dedicated to control activities. Most of the determinants are heterogeneously distributed, changing over both space and time. Factors such as topography, temperature, rainfall, land-use, population movement and degree of deforestation have a profound influence on the temporal and spatial distribution of malaria vectors and malaria. The epidemiological mapping of high-risk areas helps national level malaria control progammes to recognize those geographical areas and population where it is possible to identify the main determinants of malaria morbidity and mortality GIS-based distribution identifies areas of occurrence of vectors at micro level which help in species-specific and environmental-friendly control measures.

In this issue (**page 1129**), Aruna Srivastava *et al.* provide an account of the application of GIS used to identify and map landscape elements related to malaria and its vectors. The techniques adopted by the authors are especially useful in mapping vast and inaccessible regions.

T. V. Ramachandra