Journ@l Electronique d'Histoire des Probabilités et de la Statistique

Electronic Journ@l for History of Probability and Statistics

Vol 2, n°2; Décembre/December 2006

www.jehps.net

ABOUT STATISTICS AS A DISCIPLINE IN INDIA

B.L.S.PRAKASA RAO[1]

1 INTRODUCTION

In the preface to the book on " Glimpses of India's Statistical heritage", Ghosh et al. (1992) mention that " The origin of qualitative statistical thinking, according to Mahalanobis (1957) , can be traced back to the *Saptabhanginyaya* or the dialectic of seven-fold predication in *Syadvada* system of logic developed by the Jaina philosopher Badrabahu of the fourth century B.C."

Some articles have been recently written about the evolution of statistics in India. However these publications are not possibly widely available for general readers of this journal. Based on these articles as well as other sources, we propose to briefly review various aspects of the subject of Statistics as it evolved in India.

2 STATISTICS IN ANCIENT TIMES IN INDIA

The subject of statistics and probabilistic ideas in India seem to be of ancient origin in India. Ghosh et al. (1999) mention that " It is interesting and illuminating to note that statistical knowledge and probabilistic ideas were attributed to the kings and rulers mentioned in the great Indian epic, the Mahabharata,." The treatise Arthasastra by Kautilya (probably written during 321 -296 B.C.) during the Mauryan period had a detailed description of the system of data collection relating to agricultural , population and economic censuses in villages and towns during the period. Much later, the tradition of collecting data in detail continued during the period of Mughal emperor Akbar around 1590 A.D. The *Ain-i-Akbari* written by Abul Fazal during 1596-1597 A.D. has the best compilation of that period containing a

[1]University of Hyderabad, HYDERABAD 500046, India

wealth of information. This had details of several government departments including the system of legalized measurements, land classification and crop yields and other information. Abul Fazal was " regarded as a statistician " (Jarret (1894)).

In the preface to his book on "Statistics and Truth", C.R. Rao (1989) writes " Statistics as a method of learning from experience and decision making under uncertainty must have been practiced from the beginning of mankind. But the inductive reasoning involved in these processes have never been codified due to the uncertain nature of the conclusions drawn from given data or information. The break-through occurred only in the beginning of the present century with the realization that inductive reasoning can be made precise by specifying the amount of uncertainty involved in the conclusions drawn. This paved the way for working out an optimum course of action involving minimum risk, in any given uncertain situation, by a purely deductive process. Once this mechanism was made available, the flood gates opened and there was no end to the number of applications impatiently awaiting for methods which could really deliver the goods. From the time of Aristotle to the middle of the 19th century, chance was considered by scientists as well as philosophers to be an indication of our ignorance which makes predictions impossible. It is now regarded that chance is inherent in all natural phenomena, and the only way of understanding nature and making predictions (with minimum loss) is to study the laws (or the inner structure) of chance and formulate appropriate rules of action. Chance may appear as an obstructor and an irritant in our daily life but chance can also create. We have now learnt to put chance to work for the benefit of mankind All knowledge is, in the final analysis, history. All sciences are, in the abstract, mathematics". C. R. Rao (1989) adds " All methods of acquiring knowledge are essentially statistics."

3 STATISTICAL SYSTEM DURING BRITISH INDIA

Ghosh et al. (1999) describe the statistical system in British India in a detailed fashion. British political power was first established by the East-India Company (EIC) in Eastern India and later spread to all of undivided India between 1757 and 1947 except the nominally independent states and a few French and Portuguese settlements. EIC felt a need for acquisition of accounts and a detailed knowledge regarding the territories occupied by it. In 1807, a survey of the provinces was done covering an area of 60,000 square miles and 15 million people (Buchanan (1807)). This report contained a detailed information on topography of each district, the condition of the inhabitants along with their religion, customs, the natural

Journ@l électronique d'Histoire des Probabilités et de la Statistique/ Electronic Journal for History of Probability and Statistics . Vol.2, n°2. Décembre/December 2006

produce of the country, fisheries, mines and quarries, the agricultural situation, the state of the landed property and tenures, the progress made by Indians in arts, the state of industry, the operation of commerce and in addition an indication of rare, useful plants and seeds. Martin (1838) published "The History, Antiquities and Statistics of Eastern India" in three volumes of 2400 pages covering 9 districts of Bengal . A government officer by name A. Shakespeare published the first census report in 1848 relating to the area and revenue of each *paragana* (district) in North-West provinces of India. The first systematic attempt to get detailed information on the whole population of India was made between 1867 and 1872. It was not a census for the whole country and it was also incomplete. The operation of a decennial census for the whole country started in 1881 and is still continuing.

At a later stage, details on topographical data, ethnic divisions and creeds, agricultural situation, commerce, working of district administration, sanitary and medical aspects, meteorological data were collected for each district . About 100 printed volumes aggregating to 36,000 pages covering 240 districts comprising 15 British Indian provinces were published and later condensed in the Imperial Gazetter of India which was released in 1881 in nine volumes. It was observed that " no comparable area of the world has anything like this prodigious compilation of statistical data and demographic and historical material as a country which is almost a continent in the immensity and diversity of its character" (Chaudhuri (1964)).

During the British period, the statistical development was geared to towards administration, tax collection, revenue, trade and commerce and related activities.

4 STATISTICAL SYSTEM AFTER INDIA ATTAINED INDEPENDENCE

After the independence in 1947, the country saw an urgent need for a statistical framework suitable for economic and social development. Mahalanobis, a physicist by training, recognized Statistics as the "key technology" in planning for the economic development of an independent India and pioneered the statistical movement in the country. Mahalanobis was appointed as an Honorary Statistical Adviser in 1949 to the Government of India and Central Statistical Unit was set up under his technical guidance which was later organized named as Central Statistical Organization (CSO) in 1951. This organization was set up mainly to coordinate the statistical work done in various ministries and other government agencies and to advise them, to maintain standards with regard to definitions, concepts and procedures,

Journ@l électronique d'Histoire des Probabilités et de la Statistique/ Electronic Journal for History of Probability and Statistics . Vol.2, n°2. Décembre/December 2006

to provide consultancy, to liaison with international statistical organizations, to prepare and publish a Monthly Statistical abstract and an Annual Statistical Abstract and to inform annual statistical information to public. In order to improve the quality and fill up the gaps in statistical information , it was decided to establish a National Sample Survey Organization (NSSO) . This organization had four divisions: Survey design and research, Field operations, Data processing, and Economic Analysis. The national sample survey (NSS) is the largest multi-purpose socio-economic survey.

5 RESEARCH , TEACHING AND TRAINING IN STATISTICS

The first important work in statistics in the modern sense to be undertaken in India was possibly the statistical analysis of examination results in Calcutta University. Mahalanobis (1922) analyzed the data collected on Anglo-Indians in Calcutta and published a paper on statistical analysis of Anglo-Indian stature. He continued to work on anthropological data and developed the concept of D^2-statistic for classifying populations . The D^2-statistic is a powerful technique in multivariate analysis for classification problems and cluster analysis. The second major contribution of Mahalanobis to the field of statistics is in organization of large scale sample surveys. Starting with exploratory surveys (pilot surveys) confined to a few square miles in Bengal in 1937, Mahalanobis was probably the first person to organize a large scale survey covering the whole of Bengal (about 50000 square miles) in 1941 (cf. Mahalanobis (1944)). The three important contributions of Mahalanobis to the subject of survey sampling are pilot surveys, concept of optimum survey design and interpenetrating network of subsamples (IPNS). All the three ideas are forerunners of important contributions to statistics: pilot surveys as a prelude to Wald's "sequential analysis", optimum survey design as a precursor to "operations research" and IPNS technique as curtain raiser to resampling procedures such as "bootstrap" (Ghosh et al. (1999)).

According to Nair (cf. Ghosh et al. (1992), p.108) who received his B.A. (Hons) degree in Mathematics in the year 1932 from the Maharaja's college of Science, Thiruvananthapuram affiliated to the University of Madras, the only colleges which offered an optional (elective) paper in Statistics in its honours curriculum was the Maharaja's college of Science and the Madras Christian college. There was also a Department of Economics and Statistics at the University of Madras around the years 1934-36. The main responsibility of the Lecturer in Statistics in this Department was to teach Economic Statistics for an Evening Diploma

Journ@l électronique d'Histoire des Probabilités et de la Statistique/ Electronic Journal for History of Probability and Statistics . Vol.2, n°2. Décembre/December 2006

Course in Economics which the University has started for the benefit of graduates employed in government offices, colleges and private firms." Huzurbazar (cf. Ghosh et al. (1992), p.27) writes "..the topic of Probability was not included at any level in the Mathematics curriculum of Bombay University. Later I was a freshman at Rajaram College, Kolhapur, which has just made arrangements for the teaching of a " new course" in Mathematics as an alternative option to the usual "old course" in Mathematics for the first two years in college. The interesting thing about this new course was that it introduced for the first time, elements of Probability and Statistics as optional to Geometry, Trigonometry and Mechanics in the old course in Mathematics. This new course in Mathematics was introduced in Bombay University in 1935...". From these observations, it is clear that there was no undergraduate program specifically catered to the discipline of Statistics even though the subject of elements of Probability and Statistics was taught as an elective in the mathematics curriculum at some of the universities.

From the year 1932, the Indian Statistical Institute started the short training courses in Statistics which were attended by officers on deputation from the government and other organizations from all over India. These were the only training courses available until Mahalanobis started the first post graduate course in Statistics at Calcutta University in 1941. There were not many text books on Statistics and the teachers had to learn by reading original papers published in statistical journals. Apart from teaching, faculty members were also involved in research and made fundamental contributions to Multivariate Analysis, Construction of Designs, Sample surveys and Statistical Inference. Mahalanobis thought that a post graduate programme can be strengthened if the students get trained in Statistics from the undergraduate program. He established an undergraduate Department of Statistics at the Presidency College, Calcutta. The post graduate Department of Statistics was started in 1948 by the Bombay University and in 1953 by the University of Pune. Other major universities which have contributed significantly to the development of the subject of Statistics during the early years are the universities of Madras, Mysore, Kerala, Guwahati, Andhra and Lucknow. Post graduate departments of statistics were later started at other places in the country.

Evolution of the subject of Statistics in India is incomplete without an exhaustive discussion on the Indian Statistical Institute founded by Mahalanobis and its role. The Indian Statistical Institute was founded as a society on December 17, 1931 and it introduced the Bachelor of Statistics (B. Stat.) and Master of Statistics (M.Stat.) and Ph. D. degree programs in the year 1960 after the Institute was declared as an "Institute of National im-

portance" by the government of India in view of the extensive contributions of the Institute to research and training in the subject of Statistics. Fisher's view that "teaching, instruction or training in Statistics, at whatever level, is bound to be, on the one side with fact finding projects in the traditional statistical fields of demography and economics, and on the other side with opportunities to gain first hand familiarity with at least some field in natural sciences" was also held by Mahalanobis. The courses for the B. Stat. degree include some areas of biological, physical and geological sciences. Faculty at the Indian Statistical Institute (ISI) had made fundamental contributions to the subject of Statistics. No treatment of such topics as linear models, estimation, maximum likelihood estimation can be complete without an appropriate coverage of work done at ISI. The Cramer-Rao inequality, Rao-Blackwell theorem and Basu's theorem on independence of an ancillary and a complete sufficient statistic are part of any undergraduate or post graduate course in Statistical Inference.

During the forties, Mahalanobis recognized the need for introducing Statistical Quality Control techniques for improving the quality of products produced by Indian industries. A special course was organized by ISI in 1945-46. He invited Walter Shewart who has done pioneering work in U.S.A. in Statistical Quality Control (SQC) to visit ISI. Shewart conducted a one-week conference on "Standardization in Industrial Statistics" in 1848 under the auspices of the ISI and the Indian Standards Institution. These activities in turn led to the starting of the first SQC unit by ISI in Bombay in 1953 and later at other cities in India. The main objective of the staff of these units was to visit industries and act as consultants in SQC activities of the industries to improve the quality of their products.

Mahalanobis saw the need for training statistical officers from the Middle East, South and South East Asia and from countries in Africa. Under the auspices of UNESCO and the Government of India, the International Statistical Education Centre (ISEC) was started by ISI in 1950. It is now jointly operated by the International Statistical Institute and ISI. It has provided training in statistical methods to a large number of trainees from over fifty countries.

Another Institute which contributed in a substantial measure to the development of Statistics in India is the Indian Agricultural Statistics Research Institute (IASRI). At the end of 1943, the government of India set up an enquiry committee to look into the causes of the devastating Bengal famine. The committee observed that one of the main factors responsible for the famine was the defective statistics of crop production available at that time. Sukhatme and his coworkers began research in the methods of collection of statistics of crop yields by developing survey techniques of yield estimation under random sampling. There was a ma-

jor scientific dispute between ISI and Indian Council of Agricultural research (ICAR) on the best method of crop-cutting experiments for the estimation of crop yields. After considerable experimentation with cuts of different shapes and sizes, Mahalanobis recommended the use of circular cuts of radius 4 feet for yield surveys and ISI had been using the same . As against this, ICAR had been using the rectangular cuts of size of length 33 feet and width 16.5 feet for crop yield surveys conducted by them. Studies conducted, to see the relative efficiency of two systems, however did not reveal significant differences between the two methods (Sastry (1977), Adhikari (1990)).

In order to have an outlet for research contributions in Statistics emanating from India and other countries in the subcontinent, Mahalanobis founded and edited the journal *Sankhya, The Indian Journal of Statistics* from 1931. He was the editor from 1931-1972. The first issue of the journal was published in June 1933. In an editorial published in the first issue, Mahalanobis explained why the name *Sankhya* seemed to be appropriate in the Indian context. In the editorial he said that " We believe that the idea underlying this integral concept of Statistics finds adequate expression in the Indian word *Sankhya*. In Sanskrit the usual meaning is 'number', but the original root meaning was 'deterministic knowledge'. In the *Atharva-veda*, a derivative from *Sankhyata* occurs both in the sense of 'well-known' as well as 'numbered'..." It was and it is now one of the internationally recognized journals in the areas of Probability and Statistics for its quality contributions.

6 SIGNIFICANT CONTRIBUTIONS OF SOME INDIAN STATISTICIANS

We have already discussed the significant contributions to the evolution of the subject of Statistics and the important contributions made by Prasanta Chandra Mahalanobis . A more detailed discussion can be found in the biography of Mahalanobis written by Rudra (1996). The contributions to the areas of Probability and Statistics by other Indian statisticians is quite significant. We try to present only some of these in a nut shell.

Raghu Raj Bahadur received his Ph.D. degree from University of North Carolina, U.S.A. in 1950 after spending an year during 1946-47 at Indian Statistical Institute doing course work in theoretical and applied statistics. His research in the fifties and sixties laid the rigorous foundations for the theory of sufficiency and efficiency. The concept of sufficiency was introduced by Fisher. In a series of papers starting from 1954, Bahadur explored and elaborated the role of sufficiency in statistical decision theory. He investigated the conditions

Journ@l électronique d'Histoire des Probabilités et de la Statistique/ Electronic Journal for History of Probability and Statistics . Vol.2, n°2. Décembre/December 2006

under which maximum likelihood estimates are consistent and presented examples of inconsistency of maximum likelihood estimates. His work on the asymptotic theory of sample quantiles , now known as the Bahadur representation of sample quantiles is well known. In a series of papers Bahadur developed new methods for the comparison of competing statistical procedures based on the theory of large deviations. This idea led to the concept of Bahadur efficiency. During the years 1956-61, he was a Professor at the Indian Statistical Institute and gave inspiring lectures on several topics such as inference, sequential analysis, measure-theoretic probability and limit theorems of probability theory. He migrated to U.S.A. in 1961 and was Professor Emeritus at the University of Chicago, U.S.A. .

Debabrata Basu started his research career at Indian Statistical Institute in 1950. He obtained his Ph. D. under the guidance of C.R. Rao. Basu inspired generations of students at Indian Statistical Institute through his masterly ability to present counter examples . He was well known for his essays on the Foundations of Statistical Inference and his theorems on complete sufficient statistics. He migrated to U.S.A. in 1975 and served at Florida State University, U.S.A. till 1990.

Gopinath Kallianpur received the Ph.D. degree from University of North Carolina, U.S.A. in 1951. After spending an year at the Institute for Advanced Study in Princeton, he joined Indian Statistical Institute in 1953. His main area of research was stochastic processes, in particular, stochastic filtering theory. He is presently an Alumni Distinguished Professor at University of North Carolina, U.S.A.

Debabrata Lahiri was a completely self-taught individual in the field of Statistics. The Lahiri method of sampling with probability proportional to aggregate size leading to unbiased ratio estimation and his work on non-sampling errors are worthy of special mention in Survey sampling. Among his various other contributions are his work on integration of surveys and the introduction of circular systematic sampling in the design of sample surveys.

Pesi Rustom Masani received his Ph. D. from Harvard University, U.S.A. in 1946. A major part of his early work deals with investigation of the factorization problem for matrix-valued holomorphic functions which led him to Wiener-Masani theory of multivariate stationary stochastic processes. His work on helixes in Hilbert space and its application to probability theory and generalized harmonic analysis is significant. Masani retired from University of Pittsburgh, U.S.A. as Distinguished Professor.

Keshavan Raghavan Nair earned his Ph. D. degree from University of London in 1947. He introduced the concept of partially balanced incomplete block designs and discussed the methods for their construction and analysis. Among his other well known contributions is

a paper with C.R.Rao dealing with confounding in asymmetric factorial experiments in the study of Design of Experiments. He retired as the Director of Central Statistical Organization, Government of India in 1968.

Calyampudi Radhakrishna Rao received Ph. D. from Cambridge in 1948 for his thesis " Statistical Problems in Biological Classification". Owing to the Cramer-Rao inequality and Rao-Blackwell theorem, he is well known for any serious student of Statistics. His contributions to the Theory of Estimation. Multivariate Analysis, Characterization Problems, Combinatorics and Design of Experiments, Matrix Algebra, and Mathematical Genetics are well known. Particularly significant are his joint work with K.R.Nair on partially balanced incomplete block designs (PBIBD), his work on efficient score test which is also known as Lagrangian multiplier test, his work on orthogonal arrays. As the Head of the Research and Training School of the Indian Statistical Institute, he created an internationally renowned team of young statisticians, economists and mathematicians. He migrated to U.S.A in 1979 and retired recently from the Eberly Professorship at Pennsylvania State University, U.S.A.

Pandurang Vasudeo Sukhatme received his Ph.D from University of London in 1936. Besides being a sample survey specialist, he is also well known in the field of nutrition for the Sukhatme-Morgan hypothesis. This hypothesis states that at low levels of calorie intake, energy is used with greater metabolic efficiency and efficiency is decreased as the intake increases over the homeostatic range.

Sujit Kumar Mitra received his Ph.D. from University of North Carolina in 1956. He has made pioneering contributions to Linear models, Design of experiments, Statistical Inference, and to the Theory of Generalized inverses for Matrices.

A more extensive discussion on the contributions of some of these Statisticians is given in Ghosh et al. (1992) and Mukhopadhyay (1997).

Significant contributions were made to the areas of Limit Theorems in Probability , Measures on abstract spaces such as metric spaces, Hilbert spaces , Information Theory, by Probabilists V.S. Varadarajan, R.Ranga Rao, K.R. Parthasarathy and S.R.S. Varadhan who were all trained at the Indian Statistical Institute. There are others such as R.C. Bose, S. N. Roy who migrated to U.S.A. early in their career and made significant contributions to the branches of Construction of Designs and Multivariate Analysis.

7 CONCLUSION

Following Ghosh et al. (1999), we note that India had a long historical tradition of collection and use of various types of statistics. The system was strengthened during the British period. The development and evolution of the subject of Statistics and its applications is significant during the years 1930 to 1960 and later. The mixture of theory and applications and interaction between institutions, academia and the government was a source of real problems. This attracted the best Indian students to Statistics .

8 ACKNOWLEDGEMENT

This work was prepared at the invitation of Prof. Laurent Mazliak. As was pointed out at the beginning, this paper has benefited greatly by the papers of Ghosh et al. (1999) and Rao (2003). The author thanks Prof. T.J. Rao for providing reprints and preprints of his papers.

9 REFERENCES

Adhikari, B.P. (1990) Social construction of the statistical estimation of the crop yield , Paper presented at the *XII World Congress of Sociology of the International Sociological Association*, Madrid, Spain.

Buchanan, F. (1807) *Survey of Eastern India*, Report submitted to the Court of Directors, London.

Chaudhuri, S. B. (1964) *History of the Gazetteers of India*, Publication Division, New Delhi.

Ghosh, J.K. , Mitra, S. K. , and Parthasarathy, K. R. (1992) *Glimpses of India's Statistical Heritage*, Wiley Eastern, New Delhi.

Ghosh, J. K, Maiti, P., Rao, T.J., and Sinha, B. K. (1999) Evolution of statistics in India, *International Statistical Review*, **67**, 13-34.

Jarret, H.S. (1894) *Translation of Ain-i-Akbari*, Asiatic Society of Bengal, Vol.II, p.vii.

Mahalanobis, P.C. (1922) Anthropological observations on the Anglo-Indians of Calcutta, Part I, Analysis of male stature, *Rec. Ind. Museum*, **23**, 1-96.

Mahalanobis, P.C. (1944) On large-scale sample surveys, *Philos. Trans. Roy. Soc., London, Ser. B*, **231**, 329-451.

Mahalanobis, P.C. (1957) The foundations of statistics, *Sankhya*, **18**, 183-194.

Martin, M. (1838) *History, Antiquities, Topography and Statistics of Eastern India*, W.H. Allen, London.

Mukhopadhyay, N. (1997) A conversation with Sujit Kumar Mitra, *Statistical Science*, **12**, 61-75.

Rao, C. R. (1989) *Statistics and Truth*, Council of Scientific and Industrial Research, New Delhi.

Rao, T. J. (2003) Origin of Indian official statistical system, Mahalanobis role, Bulletin International Statistical Institute.

Rudra, Ashok (1996) *Prasanta Chandra Mahalanobis , A Biography*, Oxford University Press, Delhi.

Sastry, N.S. (1977) Estimation of agricultural production by sample survey for estimating crop acreage and crop yields: The Indian experience, Bulletin International Statistical Institute.