STATISTICS AS A KEY TECHNOLOGY

By P. C. MAHALANOBIS

I have selected as a subject "Statistics as a key technology"—which is a a quotation from Ronald Fisher—because of its special relevance to India and other underdeveloped countries. The American Statistical Association was started to help in the collection of reliable statistics for use in economic and social affairs. The need of reliable statistics is as great or even greater today for the rapid economic growth of the developing countries. Industrialisation of all the underdeveloped countries is an essential condition for world peace and is as much the concern of the advanced countries as of the underdeveloped. Because of your friendly interest I feel I have a responsibility on the present occasion to explain our difficulties and ask for your support in strengthening the statistical system in India and other developing countries.

Indian statisticians have a high reputation in the world and yet the statistical system in India is weak and is lagging behind in the timely flow of information required for social and economic affairs, Why?

The best way of diagnosing would be to repeat here something I heard this very afternoon while having lunch with a number of distinguished former presidents of your Association. One of them, who visited India recently, had informal discussions in five or six Indian universities with graduate students who were attending courses for the master's degree in statistics. There was, of course, much of theory and mathematical exercises. But when the students were asked what kind of statistical data they were using, they all looked surprised and said they did not use any data at all.

I may recall some of my own experience. On one occasion I was a member of a selection committee which was interviewing candidates for a high post in sample surveys in the Government of India. One candidate had taken a Ph.D. degree from a well-known university in America with a thesis on statistical sampling; and after his return to India was working in a Government statistical office. He gave me a copy of his Ph.D. thesis which was full of mathematical exercises; the results of his investigations were given in an abstract mathematical form and their implications were not immediately clear to me. As the post was for work mostly at an operating

Based on speech as guest speaker at the 125th Anniversary celebration of the American Statistical Association on the 27th November 1964 in Boston, Mass., U.S.A.; to be published in the *American Statistician*, April, 1965.

level, I asked him whether he had any occasion to use his results in practice. He was surprised at this question and said "no". I asked him whether he knew or could think of any situation in which his results could be used in practice; he thought over my question for a while, and then again said "no."

It is not difficult to see what is wrong with official statistics in India. There is gap between theory and practice. There is gap between the means and the end in the absence of any clearly perceived purpose. There is lack of appreciation of the need of cross-examining the data which is the first responsibility of a statistician. Collection and scrutiny of primary information or processing and handling of data are usually considered "dirty work", not fit to be done with one's own hands by statisticians and, therefore, also not fit subjects for training or for acquiring skill and experience. The validity of the data is not questioned. Anything which is supplied or published by a Government office is accepted as reliable. To have any doubt would be a challenge to established authority. The very idea of having cross-checks is frightening as conflicting results arising from independent checks would be "confusing" and must be resisted and is being resisted even today. A meeting of statisticians, not so long ago, recommended that statistics of one type of information must be collected by only one single Government agency so that "confusing" discrepancies could never arise. In this situation the statistical servicing is bound to remain weak in spite of a great deal of knowledge of theoretical statistics.

I shall have to digress a little to explain the contrast between the traditional outlook in underdeveloped countries and the scientific outlook in advanced or advancing countries. Before the emergence of science and technology in the western countries during the last four hundred years or so, there were only two basic types of decisions in any organised society. In one there was freedom of individual choice about food, clothes, recreation, art, music, literature etc., within the limit of physical availabilities or social permissibilities. The second type of decision in every kind of group or organised human activity was governed by the "principle of authority" in which the decision of an authority of higher status must override a decision of an authority of a lower status. This principle had prevailed in every society in the past and must prevail in all communities in all the countries of the world in future. The decision of a first court of law must be accepted and implemented until and unless it is reversed by a higher court of appeal; and the second decision itself would be subject to confirmation or reversal by an appeal to a higher court until the supreme or the last court of appeal is reached. The principle of authority must be accepted for the very existence of society itself.

The same principle of authority must, however, be completely rejected in the domain of science. The discovery of order and regularity in the movement of the planets gave rise in the 16th and 17th centuries to the view that natural phenomena are amenable to rational and unified explanation. The concept also emerged of a world of physical reality in which different parts are connected and events are regula-

ted by laws of nature. To know this world of reality more and more adequately is the guiding motive of scientific enquiry which is the characteristic mark of the modern age. Science can advance only through a patient accumulation of facts and observations and a critical study of their interrelations which have their foundation in nature itself and which cannot be changed or upset by any human authority however high. The findings of the most eminent scientists must be subjected to critical check and must be corroborated or refuted by independent observations and study. The experimental attitude of mind and the urge to establish the objective validity of the data and of the scientific findings through independent checks form the basis of the scientific tradition.

It is of interest to note that the concept of probability also emerged in the 16th and the 17th centuries, at first, in connexion with games of chance, for example, in the attempt to find equitable rules for the division of stakes in unfinished games, with an obviously pecuniary motive. Games of chance were being played from time immemorial but, as far as we know, the question of any order or regularity in the outcome of such games had never been raised. There is no evidence of the concept of probability having emerged in any country before the emergence of science four hundred years ago.

The reason I think is clear. The throw of a dice or of a coin had been looked upon as being decided by the caprice of gods or demons or by 'chance' which was entirely unpredictable. With the emergence of the scientific view of an objective world of physical reality in which events were regulated by laws of nature, the choice was between making an exception of the outcome of games of chance or bringing them within the world of physical reality governed by laws of nature. To resolve this dilemma, it became necessary for the human mind to find some order or regularity in the occurrence of chance events. This effort led to the formulation of the rules of mathematical probability and the normal law of errors as a part of the laws of nature. In this way chaos and random chance were integrated in the world of reality. Mathematical probability emerged only when the need had arisen; it could not have emerged earlier.

In advanced countries with established scientific tradition there is continuing concern with validity of data and validity of conclusions. In underdeveloped countries the principle of authority is still dominant; the question of validity can scarcely arise. Statistics, therefore, necessarily remains a matter of formal or administrative sanctions; anything having the official stamp must be accepted as authoritative. In a village economy the information available within the village, or in neighbouring villages, is enough for all practical purposes. The question of comparability or compilability does not arise. But as industrialisation starts and proceeds with increasing tempo, as it must in the underdeveloped countries, the need of statistics or increasing coverage and of increasing accuracy becomes more and more insistent. India has already reached this stage but the statistical system is unable to meet the need.

The training of Indian statisticians used to be traditionally in economics. During the last twentyfive years statistical courses are being offered at both undergraduate and graduate levels in an increasing number of universities in India. The training is often of the traditional single subject type, oriented to either economics or to mathematics, and usually of an abstract nature without any contact with data and without any clearly perceived purpose. The greatest danger lies in the lack of appreciation of the fact that a master's degree in statistics of this type does not guarantee any knowledge or skill in the professional work of statisticians. The position is roughly somewhat like what it would be to accept a master's degree or a Ph.D. in bio-chemistry as the equivalent of the M.B., or the M.D. degree awarded on the basis of a full course of professional training in medicine. Statistical posts are, however, being filled on the strength of the master's degree in statistics or a Ph.D. degree awarded on the basis of abstract exercises in "statistical mathematics" without any relation to data. The consequences are as much of a tragedy as it would be if a Ph.D. who has worked intensively on the synthesis of a vitamin but had no medical training were to be placed in charge of brain surgery.

It has to be emphasised that statistics must have a purpose. It had its origin in the counting of men or of cattle or in the measurement of land, foodgrains, etc., for purposes of management and administration from time immemorial. The very word statistics shows the connexion with "statecraft". With the emergence of the concept of probability, statistical theory has been characterised by a dual motivation, one of which is utilitarian and concerned with economic gains and the other scientific or logical and concerned with the question of validity of data or conclusions.

In the second half of the 18th century Laplace developed the mathematics of probability with vigour but the question of applications was also very much in his mind. In his great treatise on the analytical theory of probability, published in early 19th century, he had included a "Philosophical Essay on Probability" in two parts; in the first part, of 52 pages in the English translation, he dealt with the abstract and mathematical aspects of the subject, and in the second part, of 142 pages, with applications of probability theory, devoting nearly three quarters of the space to the latter. The concern of statistics with practical affairs was clear in his mind.

After Laplace, developments in the nineteenth century continually showed the dual concern of statistics, on one side with considerations of economy in practical affairs and on the other with questions of validity in scientific inference. The normal distribution, developed in connexion with games of chance, was used for adjustment of observations and led to the theory of errors which became an important part of statistics. The distribution of errors was taken over as a model for the distribution of velocities in the kinetic theory of gases; the statistical approach led to important

advances in physics; for example, in statistical mechanics and its later developments, and the probabilistic view supplied the fundamental principle of uncertainty in theoretical physics to demarcate the limits of physical observations.

Pure mathematics is indispensable for supplying rigorous logical foundations and for exploring the limits and refinements of statistical theory but has never given it essentially new concepts or tools. New statistical concepts and methods have continually emerged only in dealing with observations in the world of reality. To give a classical example, the physical idea of correlation between two biological variates was formulated on the basis of measurements on sweetpeas and on heights of fathers and sons taken by Francis Galton in the eighteen-seventies; Gauss himself had missed this concept although he handled its pure mathematics fifty years earlier in the form of the cross product of two sets of errors of observation.

In the present century new statistical tools have continually extended the scope of applications. At the turn of the century, for example, Karl Pearson's Chi-square supplied a versatile method for testing the agreement between results expected on the basis of any specified hypothesis or theory and those actually observed. Statistical methods also began to be used increasingly in biology and genetics which justified the word 'Biometrika' which Karl Pearson had coined as the name of his new journal.

Forty years ago, in the early nineteen-twenties, a number of remarkable developments started at about the same time under my own eyes. Ronald Fisher in his work on estimation supplied a critical foundation for theoretical statistics which was the starting point of a great deal of mathematical and logical researches which are still continuing. Fisher's work on estimation introduced in statistical theory the concept of 'efficiency' in the form of the quantum of information secured in relation to the sample, a concept which has its basis essentially on economic considerations. On the applied side, Fisher introduced at about the same time the design of experiments, based on the principles of randomisation, replication and local control with analysis of variance for a critical interpretation of the results of agricultural field trials at a much higher level of precision and objective validity than had been possible earlier. This method was rapidly extended to practically every kind of experimental investigations.

Walter Shewhart, also in the early nineteen-twenties, introduced Statistical Quality Control for the improvement of the efficiency of industrial production. On the basis of repeated small samples, the aim was to diagnose whether a dynamic system of production was working under adequate control or was tending to go out of control so that, in the latter case, timely action could be taken to bring the system again under control. SQC supplied a universal tool for the feed-back control of any dynamical system, and is now being used extensively in industry, and also in business management and systems operation of many kinds all over the world. From the very beginning the economic motivation was dominant.

During the last forty years there have been also big advances in the theory and practice of statistical sampling for the collection of socio-economic information on a large geographical scale. The theory of the design of sample surveys is based, as in the case of the design of experiments, on the principles of randomisation and replication, using sometimes stratification and inter-penetrating subsamples, and always trying to minimize the cost of the survey to secure results with a given margin of uncertainty. It is being increasingly appreciated that large scale sample surveys, when conducted in the proper way with a satisfactory survey design, can supply with great speed and at low cost, information of sufficient accuracy for practical purposes and with the possibility of ascertainment of the margin of uncertainty on an objective basis. Sample surveys also supplied for the first time a method for checking, on a scientific basis, the validity of a census or a complete enumeration.

All over the world, particularly during the last twentyfive years, statistical methods are being increasingly used in new fields of scientific work and practical affairs with the dual objectives of economic gain, and of improvement of the validity of data and of conclusions. Statistics is now concerned with the use at the lowest cost of the most efficient methods of observation, measurement, survey and experimentation and of the processing of the data to draw valid inferences. The scope of statistics extends over the whole range of the natural and social sciences, engineering and technology, medicine, management and economic affairs.

I may now come back to the main theme, namely, statistics as a technology. It will be convenient to explain the precise sense in which I am using the word "technology" in the present context. We may look upon science education and scientific research as the effort to know nature more adequately. We may also consider technology and technological research as the effort to use scientific knowledge for the fulfilment of specific purposes either of a practical or a theoretical nature.

A scientist may have a wide or a very narrow range of interest or specialisation. A technologist must have knowledge and experience of a wide range of scientific subjects. University education had been traditionally based on courses in single subjects, such as, physics, chemistry, botany, economics etc. During the latter part of the nineteenth century it became necessary to develop new educational programmes, for example, in medicine and engineering, on the basis of courses in many subjects because it was essential for professional students to have some knowledge and experience in a number of scientific disciplines. It is relevant to note in the present connexion that the Swiss Federal School of Technology of Zurich had to be established in 1850 or 1851 on the basis of multi-discipline courses because the older universities were not yet prepared to offer such programmes. Presumably for the same reason the Massachusetts Institute of Technology also had to be established in 1861, next door to Harvard University, to give multi-discipline courses in engineering.

The time has come to introduce educational programmes appropriate for statistics as a fully developed new technology which calls for the utilization of a wide range of scientific knowledge to help in solving scientific or practical problems. As Fisher had pointed out, "a professional statistician, as a technologist, must talk the the language of both theoreticians and practitioners." The education of a statistician, like that of other technologists, must have a broad base.

In underdeveloped countries, the statistical system is often weak because it is not functional. Knowledge of statistical theory remains isolated from aplications; routine tasks of collection or processing of data are continued often without utilization of the information, or new programmes are started in a superficial imitation of advanced countries without any purpose.

It is essential, in the underdeveloped countries, to make statistics purposive. The only way is to give training in statistics, not as a single abstract subject, oriented either to economics or to mathematics or a little bit to both, but as a fully developed technology of a multi-discipline character. From my own long experience of statistics in India I am convinced of the need of recognising statistics as a key technology in the underdeveloped countries. May be there is need also in the advanced countries, I have no competence to express any strong views.

Educational and research programmes have been started in the Indian Statistical Institute for the above purpose. Degree courses of statistics have been formulated to cover a wide range of subjects somewhat analogous to courses in medicine and engineering. Pure mathematics and theory of probability have an important place. Statistical theory and different branches of applied statistics and economic planning (to suit the special needs of India) form a large part of the teaching programme. In these three groups of subjects, mathematics, statistics and economics, the course includes a good deal of content of knowledge, besides theory and techniques. In addition, facilities are provided for the students to become familiar with a number of scientific subjects. Here the emphasis is not so much on the content of knowledge as on methods, stress is therefore placed on practical courses, statistical analysis and interpretation of data.

Fisher has pointed out "that the science with which the student is to become acquainted must be genuine research in its own right, not what is eloquently called a 'mock up' for the use of students only." It is therefore the policy of the Institute to establish small, high level research units in both the natural and social sciences to offer facilities for training and research in the use of statistical methods in practice. The point has to be emphasised that it is only in the advancing form of genuine research that new types of data would be thrown up offering scope for the extension of known methods or as a challenge for forging new tools.

The above venture is of a somewhat novel nature. It has no prototype in advanced countries. In the absence of recognisable models outside India there

may be difficulty in continuing or developing these new programmes. Support from the advanced countries would be a great help in making these programmes respectable and, therefore, acceptable in the underdeveloped countries.

The American Statistical Association contributed in a most significant way to the advancement of statistics not only in America but also in other countries of the world. I have come here today to ask for your support in getting statistics recognised as a key technology for rapid economic growth in India and other developing countries.

It has been a privilege to have shared in the anniversary celebrations. I take this opportunity of conveying to you, on my behalf and on behalf of my colleagues in India, our best wishes for continuing success of your endeavours in future.