II. THEORETICAL STATISTICS.

(Section of Mathematics and Physics, in co-operation with the Statistical Conference.)

PROF. S. N. BOSE, Dacca, presided.

1. PROF. P. C. MAHALANOBIS, Calcutta, opened the discussion:

The Technique of Random Sample Survey.

The technique of random sample survey was discussed with special reference to surveys on a geographical or regional basis such as crop census, crop-cutting experiments, forest surveys, etc., and his remarks were illustrated with results obtained during two experimental crop census conducted in Bengal in 1937 and 1938. The problem may be conveniently considered under the following heads:—

(1) Partitioning the whole area into suitable zones.

(2) Determination of the 'best' size of the sampling unit.
(3) Fixing the total number and density of sampling units per unit area (acre or square-mile, etc.); and their allocation to different zones.
(4) Procedure for selecting the sampling units at random.

The 'best' size and number may be conveniently defined as that set of values which, for any assigned total expenditure, would supply a final estimate with the lowest sampling error.

On the side of organization the chief factors are:—

(1) Cost of preparation of sampling units and analysis of results.
(2) Cost of collecting the field information.
(3) In the case of a semi-permanent or permanent organization the question of the most efficient distribution of labour.

It is well known that the efficiency of the sampling can be increased by dividing the whole area into a number of homogeneous zones; but the accuracy with which this can be done depends entirely on the previous knowledge available. In actual practice this problem can be solved only on a trial and success basis.

Questions of some theoretical interest are, however, involved in fixing the 'best' size of the sampling unit, the total number, and the number of units to be allotted to each zone. The total number will, of course, ultimately depend on (*a*) the accuracy which it is desired to attain in the final estimate, (*b*) the area in the different zones, and (*c*) the variability from one spot to another of the area within each zone.

Questions of cost are also involved in fixing the 'best' size and the 'best' number of sampling units or grids. A convenient approach to the theoretical solution of the problem is given by constructing two separate functions:—(*a*) the 'variance' function giving the sampling variance for different sizes of the sampling unit; and (*b*) the 'cost' function giving the total cost involved in collecting, say, 100 sampling units of various sizes. The variance function is determined by the manner in which the elements under survey are actually distributed in space. The cost function is made up of three different kinds of elements: (*a*) cost of statistical portion of the work, (*b*) cost of field enumeration, (*c*) overhead and other expenses. The cost of field work per sampling unit is determined by the time taken for the actual field inspection and enumeration (which will depend on the size of the sampling unit) plus the time taken to move from one unit to another.

If the results of a complete enumeration happen to be available, it is possible to study the 'variance function' by model sampling experiments in the Laboratory. The cost function, however, can only be studied by actual experiments on the field.

Once the variance function and the cost functions have been suitably formulated, it is possible to obtain the best size and number of sampling units by straightforward analytic methods which are illustrated with the help of actual numerical figures obtained during two recent experimental crop surveys in Bengal.

In the case of jute crop in Bengal, which was scattered over 70 lakhs of individual plots out of 11 or 12 crores of plots covering an area of 55,000 sq. miles, the cost of a complete enumeration had been estimated at anything between 12 or 15 and 20 or 25 lakhs of rupees. In a random sample survey, on the other hand, it should be possible to obtain a provincial estimate with an accuracy of the order of three or four per cent. at a cost of about two lakhs of rupees.

It was followed by a brisk discussion relating to the theoretical problems in which a large number of persons including Professors S. N. Bose (Dacca), N. R. Sen (Calcutta), K. Krishnan (Calcutta), K. B. Madhava (Mysore), Dr. K. R. Ramanathan (Poona), and others participated.

2. Mr. S. N. Roy, Calcutta :

Discussed one special problem which had arisen in the same connexion. He showed that the exact distribution of a mean of a number of correlated variates (each pair of which had the same intra-class correlation) was given by the usual Gauss-Laplacian formula.

3. Prof. K. B. Madhava, Mysore :

Discussed the broader aspects of the sampling technique in random surveys, and stressed the importance of proper stratification and the use of necessary 'control'.

Other subjects.

1. Messrs. R. C. Bose *and* S. N. Roy, Calcutta.

The distribution of Studentized D^2-statistic appropriate to any pair out of a given set of k equidispersional population distinct in sets of means.

2. Mr. R. C. Bose, Calcutta.

The application of the theory of Galois fields and associated finite projective geometries to the construction of completely orthogonalized hyper-Graeco-Latin Squares.

3. Messrs. R. C. Bose *and* K. Kishen, Calcutta.

The application of the theory of Galois fields and associated finite projective geometries to investigate the generalized interaction in a s^n factorial arrangement, where s is a prime or a power of a prime.