

## Has Statistics a Future? If So in What Form?<sup>1</sup>

C. Radhakrishna Rao

*Emeritus Eberly Professor of Statistics, Pennsylvania State University, University Park, PA 16802, USA*

### SUMMARY

The mathematical foundations of statistics as a separate discipline were laid by Fisher, Neyman and Wald during the second quarter of the last century. Subsequent research in statistics and the courses taught in the universities are mostly based on the guidelines set by these pioneers. Statistics is used in some form or other in all areas of human endeavor from scientific research to optimum use of resources for social welfare, prediction and decision-making. However, there are controversies in statistics, especially in the choice of a model for data, use of prior probabilities and subject-matter judgments by experts. The same data analyzed by different consulting statisticians may lead to different conclusions.

What is the future of statistics in the present millennium dominated by information technology encompassing the whole of communications, interaction with intelligent systems, massive data bases, and complex information processing networks? The current statistical methodology based on simple probabilistic models developed for the analysis of small data sets appears to be inadequate to meet the needs of customers for quick on line processing of data and making the information available for practical use. Some methods are being put forward in the name of data mining for such purposes. A broad review of the current state of the art in statistics, its merits and demerits, and possible future developments will be presented.

*Key Words* : Bayesian analysis, Cross validation, Data mining, Decision theory, Estimation, Hypothesis testing, Large data sets, Machine learning.

### 1. STATISTICS AS A SEPARATE DISCIPLINE

#### 1.1 A Paradigm for Statistical Theory and Methods

The word statistics was coined by the German Scholar Gottfried Achenwall about the middle of the eighteenth century in the context of *collection, processing and use of data by government*.

During the nineteenth century, statistics acquired a new meaning as extraction of information from data for decision making. The need arose especially in testing hypotheses or making predictions or forecasts based on information in the observations made on natural phenomena or generated through well designed

experiments. It was realized that the information contained in particular data, however well they are ascertained, is subject to some uncertainty and consequently our conclusions based on observed data could be wrong. How then can we acquire new knowledge? We have to evolve a new methodology of data analysis with a view to estimate the amount of uncertainty in extracted information and to formulate rules for making decisions with minimal risk.

The equation

$$\begin{array}{|c|} \hline \text{Uncertain} \\ \text{knowledge} \\ \hline \end{array} + \begin{array}{|c|} \hline \text{Knowledge} \\ \text{of the extent} \\ \text{of uncertainty} \\ \text{in it} \\ \hline \end{array} = \begin{array}{|c|} \hline \text{Useable} \\ \text{knowledge} \\ \hline \end{array}$$

<sup>1</sup> *Keynote Address for the International Conference on Statistics and Informatics in Agricultural Research to mark the Diamond Jubilee Celebration of the foundation of Indian Society of Agricultural Statistics held at NASC Complex, New Delhi on 27 December, 2006.*

is used as a *new paradigm* for statistical theory and methods. Thus, statistics acquired the status of a new discipline of study for

