Chapter 5

INFORMATION AND STATISTICS LINKAGE

Statistics and information theory are strongly related. In this chapter, we will first define statistics and its scope. Then we will show the linkage between statistics and information theory with special reference to the model selection technique using Schwarz Information Criterion (SIC). SIC has certain optimal property that will be discussed in the later chapters. A new information criterion CGIC — Chen-Gupta Information Criterion [Chen, Jiahua 1998, 2003] has also been given which again emphasizes the relationship between information and statistics.

1. STATISTICS

Statistics may be defined as a body of methods for making intelligent decisions at the presence of uncertainty. However, this definition does not encompass the different aspects of the subject. A more pragmatic definition is: “Statistics is a branch of knowledge that deals with multiplicity of data, its collection, analysis, and interpretation.” This latter definition purports that statistics, besides being a science, is also an art. Since the interpretation of data analysis is emboldened by the analyst’s understanding of the random experiment. Statistical inferences are uncertain since the logic used is inductive. The main merit of statistical theory is that under certain conditions, it enables us to get a more or less precise idea about the measures of this uncertainty in terms of probability statements.

In statistical methods, inference is drawn from the information contained in a sample (data set) about the population, whereas the sample (data set) is only one part of the population. Statistical data are collected to help answer the questions of practical action, or questions in scientific research. Statisticians develop research tools. These tools are then applied to specific
problems in different fields such as agriculture, anthropology, biology, economics, education, medicine, physics, psychology, etc.

Hypothesis testing is an important aspect of statistical reasoning that enables us to make inferences about properties of an unknown population. For this reason statistical methods have been increasingly used in business as well. The element common to all problems faced by business managers is the need to make decisions at the presence of uncertainty, and as we have seen, the essence of modern statistics lies in the development of general principles for dealing intelligently with uncertainty. It is no surprise that statistical methods are widely applicable in nearly all areas of managerial decisions.

2. CONCEPT OF INFORMATION

The word statistics is derived from state. The first article of the Constitution of the United States provides that the government shall collect statistics – a decennial census to serve as a basis for representation of the states in Congress. However, sheer volume and complexity of data collected or available to most organization, whether government or private, has created an imposing barrier to its effective use. These challenges have propelled data mining to the forefront of making profitable and effective use of data. Statistical data analysis provides a process that uses a variety of data analysis and modeling techniques to discover patterns and relationships in data. Subsequently, these patterns and relationships in data may be used to make precise predictions.

The concept of information in statistics was first defined by Fisher [Fisher 1925] in his pioneering research on the theory of estimation. As such, estimation theory can be considered as a branch of the theory of probability and statistics. Estimation theory is fundamental to statistical inference. Fisher defined information in statistics as a function of the density function $f(x, \theta)$, where $\theta$ is a parameter as follows:

$$I_f = \int f(x, \theta) \left( \frac{\partial}{\partial \theta} \log f(x, \theta) \right)^2 dx$$

$$= E\left[ \left( \frac{\partial}{\partial \theta} \log f(X, \theta) \right)^2 \right]$$

The above equation plays an important role in point estimation and is part of the famous Cramer-Rao lower bound. It is the amount of information contained in the data about the unknown parameter $\theta$ for regular density $f(x, \theta)$. 