ABSTRACT

The objective of this research is to illustrate the use of multistage sampling techniques in fixed precision point estimation problems using an asymmetric loss function.

Various tools of statistical inference are extensively used in social and pure sciences where the aim is to estimate the parameters of a system under investigation so as to draw meaningful conclusions about the system. But these inference processes also result in an error, which is reflected in a ′loss function′. One of the objectives in inference problems is to find estimators having desirable statistical properties by suitably controlling this loss. For example, among several competitive estimators in a given situation, one may try to find the best estimator which controls the average loss (called the ′risk′) to be below a threshold level. Such a procedure is called a ′bounded risk estimation′ problem. Among the different forms of loss function, the use of asymmetric loss function is practically very appealing, as overestimation and underestimation are generally not equally penalized. One form of these types of asymmetric loss functions is the Linear Exponential (LINEX) loss function.

On the other hand, multistage sampling or sequential analysis is the method of statistical inference whose characteristic feature is that the number of observations required by the procedure is not predetermined at the beginning of the experiment. The decision to terminate the experiment depends, at each stage on the result of the observations previously made. The advantage of this method over fixed sampling

techniques is that the former requires substantially smaller number of observations than equally reliable fixed sampling techniques.

In the first part of our research we use different multistage or sequential sampling techniques under a known bounded risk condition (for LINEX loss), to find out the estimator for the mean of a normal distribution with unknown variance. Simulation runs are used to validate the asymptotic expansion of the risk function for different multistage sampling methodologies. We also use a live data set to validate our results.

In the second part we extend our work by estimating a linear function of means of $'k'$ normal distributions for which the variances are unknown. A known bound for the associated risk results in the use of sequential sampling techniques to determine the estimate of the linear function. As in the first part, here also we find the asymptotic expansion of the risk functions for the different multi stage sampling methodologies and validate our results using simulation runs.

Finally in the last part of our work, using similar concept of bounded risk we apply sequential sampling techniques for estimating a linear parametric function and also for calculating the optimal forecasts of a future value of a dependent variable under a multiple linear regression setup. We compare the results obtained for the estimation problem by carrying out runs on a live data set. Similarly we also carry out runs using a live data set for the prediction problem also.

Finally we propose some interesting research areas worth pursuing in the future.