Sequential Analysis

Sequential Analysis is a Subfield of statistics and machine learning.

First developed by Abraham Wald with Jacob Wolfowitz as a tool for more efficient industrial quality control during world war II.

In applying statistics to a scientific, industrial or social problem, one begins with a process or population to be studied. This might be a population of a people in a country, crystal grains in a rock or of goods manufactured by a particular factory during a given period. It may instead be a process observed at various times : data collect about this kind of population constitute what is called time series.

For practical reasons, rather than compiling data about an entire population, one usually studies a chosen subset of a population, called sample. Data are collected about the sample in an observational or experimental setting. The data are then subjected to statistical analysis, which serves two related purposes; description and inference.

Descriptive Statistics : can be used to summarize the data, either numerically or graphically to describe the sample. Basic examples of numerical descriptors include the mean and standard deviation.

Inferential statistics : Inferential statistics is used to model patterns in the data, accounting for randomness and drawing inferences about the larger population. These inferences may take the form of answers to yes / no questions (hypothesis testing) estimates of numerical characteristics (estimation), descriptions of association (correlation) or modelling of relationships (regression). Other modelling techniques include ANOVA, time series and data mining.

In very general terms there are two reasons for introducing sequential methods into statistical analysis. One is to solve more efficiently a problem which has a fixed sample solution. The other is to deal with problems for which no fixed sample solution exists.

Some problems are intrinsically sequential and cannot be discussed without considering their sequential aspects. An important example is a control system with unknown dynamics, about which something can be learned as the system operates. Dynamic programming is one method for dealing with problems of this sort.

Another intrinsically sequential problem is the fixed precision estimation of a parameter in the presence of an unknown nuisance parameter. It is almost obvious

that one cannot give a confidence interval of prescribed length for the mean of a normal distribution based on a sample of some fixed size n if one does not know the variance of the distribution. However, by taking data sequentially one can use the data to estimate the variance and the estimated variance to determine a (random) sample size which will permit the mean to be estimated by a fixed length confidence interval.

In statistics, sequential analysis or sequential hypothesis testing is a statistical analysis where sample is not fixed in advance. Instead data is evaluated as it is collected, and further sampling is stopped in accordance with a pre-defined stopping rule as soon as significant results are observed. It allows to decide the object class using a smaller number of measurements (features) than other methods,

The way the analysis is carried out depends in some way on the results of previous steps:

1. selection of the experiment, measurement, or test that should be carried out next,
2. terminate/continue with the analysis

Sequential analysis has plenty of applications, especially where

- the time to decision must be minimized, and / or
- individual measurements are not cheap

Advantages:

Testing can be shorter than using the classic analysis of the whole sample.

The individual tests do not have to be of the same type!

The diagnostic plan may be modified depending on the results of preceding steps

Applications:

■ Sequential sampling plans: - measurements are homogeneous, their ordering does not matter.

■ Clinical studies of drug effectivity: sequential testing by groups – if a measurement on a sample of n people is sufficient, the study is terminated, otherwise another batch of n people is used . . .

Real-time face detection in photos and videos

The term "sequential" is occasionally extended to cover also investigations in which various aspects of the design may be changed according to the observations made. For example, preliminary experience in an experiment may suggest changes in the treatments being compared; in a social survey a small pilot survey may lead to modifications in the design of the main investigation.

The most appropriate design and method of analysis of a sequential investigation depend on the purpose of the investigation. The statistical formulation of that purpose may take one of a number of forms, usually either estimation of some quantity to a given degree of precision or testing a hypothesis with given size and given power against a given alternative hypothesis. Economy in number of observations is typically important for sequential design.

Sequential hypothesis testing. Suppose that one wishes to test a specific hypothesis, H_0, in such a way that if H_0 is indeed true it will usually be accepted and that if an alternative hypothesis, H_1, is true H_0 will usually be rejected. In the most elementary case H_0 and H_1 are simple hypotheses; that is, each specifies completely the probability distribution of the generic random variable, X. Suppose $f_0(x)$ and $f_i(x)$ are the probabilities (or probability densities) that X takes the value x when H_0 and H_t are true, respectively.