Representative Points of Elliptically Symmetric Distributions

Kai-Tai Fang

BNU-HKBU United International College, Hong Kong, China

Abstract

The problem of selecting a given number of representative points (RPs) to retain as much information of the population as possible arises in many situations. One approach is proposed by Cox (1957) who proposed the mean square error (MSE) criterion and gave a table of RPs of the univariate normal distribution for $k \leq 6$. In general, this approach is definedas follows: a set of k RPs for the distribution of a random vector $\in \mathbf{R^p}$ is set of k points minimizing the expected squared distance between and the nearest point in the set.

There are different motivation for defining representative points. Max (1960) seeks to quantize the univariate normal distribution, Bofinger (1970) studied the question of grouping a continuous bivariate distribution by intervals on the marginals thereby obtaining a discrete bivariate distribution. In order to the standardize clothes, suppose taking p measurements of the body of each n individuals (in general, n is sufficiently large), and project these p dimensional data onto a $q(q < p)$ dimensional space by principal components analysis or by some other method. They wish to select k points that best represent the data in the q-dimensional space (see Fang (1976)).Motivated by this problem Fang and He (1982) proposed the question based on the standardize clothes how to choose k points under MSE. A similar background, Flury (1990) studied a project of the Swiss Army which wanted to design new protection masks. To put the construction of the new protection masks on a good empirical grounds, a group of anthropologists was hired to measure the heads of 900 Swiss soldiers. He and his coauthors found that when $k = 2, p > 2$ the representative points of an elliptical symmetric distribution on the direction of the eigenvector associated the largest eigenvalue of the covariance of . Therefore, they propose the name "principal points". In this talk I first review historical development of the representative points for univariate and multivariate cases and applications in resampling and density estimation.

Some new results for the elliptical symmetric distributions will be presented. Some comparisons among the representative points for different dimension p, the number of representative points k and different subclasses of elliptically symmetric distributions: normal, Kotz type, Pearson Type II and Pearson Type VII are given.

Keywords

Representative Points, Elliptically Symmetric Distributions, Principal points