Poster Snapshot

Robust inference by minimizing density power divergence with application to skewed data

Amarnath Nandy (ISI)

The conventional statistical distribution used in modeling different kinds of health data is the bell-shaped and symmetric normal distribution. Although it works for some health measurements like height or weight of patients, etc., most health data, specially those measured on some clinical metrics, often exhibit empirical skewness so that the conventional normal distribution can not be used to model them. Among several possible parametric distributions to model skewed data, possibly the most popular one is the \textit{Azzalini-type Skew Normal} (SN) distribution family which also covers the usual symmetric normal distribution as a special case. The usual method of estimation under SN model is the maximum likelihood estimator (MLE) which is asymptotically the most efficient at the model. But, a major drawback of the MLE is its extreme non-robust nature against data contaminations, outliers or model misspecifications; this further makes all the MLE based inference highly unstable yielding incorrect insights. However, it is not unusual to have some outlying observations in modern complex datasets due to several external or erroneous factors/activities. Hence, a robust inference procedure automatically taking care of the noises (outliers) in the data would be of great practical value to produce stable and more precise research insights leading to better policy formulation. Among several robust methods, we have applied minimum density power divergence estimation (MDPDE) for finding estimators of the parameters. In this lecture for the three Institutes meeting, we will discuss about the robust estimates of the parameters of the SN distribution in case of independently and identically distributed (IID) data along with its asymptotic properties. We also extend this technique to linear regression model with SN error. We show the numerical results which we have got by implementing our methods to some real life data sets.