Robust Statistical Modeling for Nonnormality, Skewness, and Multimodality using Generalized t-Distributions

Olcay Arslan

[1] St. Cloud State University, Department of Statistics, 720 4th Avenue South, St. Cloud, MN USA, 56301
E-mail:oarslan@stcloudstate.edu

Keywords: Robust estimation, Asymmetry, Heavy-tailed distribution, Multimodality, GT Distribution, GET Distribution.

Abstract

Since statistical inference based on the normal distribution is very sensitive to outliers, the detection and treatment of outliers in data has become an important research area of statistics. One of the approaches that has been used involves assuming that outliers create a fat-tailed error distribution and using a fat-tiled symmetric model for the error term to take this into account. A number of symmetric fat-tailed distributions have been used as the alternative to the normal distribution for statistical modeling of data sets involving errors with longer than normal tails. But the restriction that the distribution is symmetric implies that no consideration is given to the fact that outliers not only affect the fatness of the tails but also may affect the skewness of the distribution. Data with sort-tails, asymmetric, and multimodal distributions may not be well modeled using fat-tailed symmetric distributions.

The families of generalized t-distributions (there are two different families called generalized t distribution) introduced by McDonald and Newey (1988), Lye and Martin (1993), and Arslan (2000) provide a great flexibility in modeling not only symmetric fat-tail distributions but also sorter-than-normal tails, asymmetric, and even multimodal distributions. One important feature of the generalized t-distributions is that they nest a number of well-known distributions including the normal, student t, power exponential, and double exponential.

This paper illustrates the ability of models based on the generalized t distributions to handle nonnormality, outliers, skewness, kurtosis, and even multimodality in a wide range of settings. The strategy is based on the maximum likelihood estimation for a general model with generalized t distributed errors. This strategy is applied to a variate of problems, including simple location and scale estimation of a univariate data set, location and scatter estimation of a multivariate data set, regression estimation, and empirical density estimation.

In the paper by Lange et al. (1989) it has been pointed out that modeling based on the student t distribution is not the solution for all robustness problems. In particular, data with shorter-than-normal tails, or asymmetric error distributions are not well modeled by the t distribution. We believe that generalized t-distributions can provide robust alternatives to the classical distributions such as normal and the student t for statistical modeling of data sets involving errors with longer-than-normal tails, sorter-than-normal tails, asymmetric, or even multimodal, and hence can handle some of the problems pointed out by Lange et al. (1989).

References

Arslan, O. (2000) "Family of Multivariate Generalized t Distributions" Unpublished Manuscript (Submitted).

Arslan, O. and Genc A.I. (2001) " Robust location and scale estimation based on the univariate generalized t (GT) distribution", Unpublished Manuscript (Submitted).

Lange, K.L., Little, R.J.A., and Taylor, J.M.G. (1989) "Robust Statistical Modeling

Using the t Distribution." *Journal of the American Statistical Association,* Vol.84, No.408. pp.881-896.

Lye, J.N. and Martin, V.L. (1993) "Robust Estimation, Nonnormalities, and Generalized Exponential Distributions." , *Journal of the American Statistical Association,* Vol.88, No.421. pp.261-267.

McDonald, J.B. and Newey, W. K. (1988) "Partially Adaptive Estimation of regression Models Via the Generalized T Distribution", *Econometric Theory*, 4, pp. 428-467.

Theodossius, P. (1998) "Financial Data and the Skewed Generalized T Distribution", *Management Science,* Vol.44, N0.12, pp.1650-1661.