LSD Plots: Some Aspects of Their Implementation in R

Ivan Mizera, University of Alberta

In the contribution, we discuss our experience with the implementation of a new graphical tool for exploratory data analysis, the LSD plot proposed by Mizera and Muller (2004). LSD plots are plots of the contours of location-scale depth, a data-analytic construct in the vein of general theory of Mizera (2002), whose origins may be traced back to Tukey (1975) and Rousseeuw and Hubert (1999). The Lobachevski geometry structure of its most feasible variant, the Student depth, creates a link to multivariate location halfspace depth, which enables to utilize the recent algorithmic advances in the field - like those of Struyf and Rousseeuw (2000) or Miller et al. (2003). The LSD plots can be used for checking distributional assumptions about the data, in a fashion similar to that of quantile-quantile plots; they exhibit the similar incisive nature of the latter as well.

While the computational experiments with the new methodology were done predominantly in MATLAB, a transition to more user-oriented implementation in R posed several technical problems. In particular, the availability of (somewhat) interactive statistical environment, like iPlots developed by Urbanek and Theus (2003), seems to be paramount for the routine use of the technique.

Miller, K., Ramaswami, S., Rousseeuw, P., Sellares, A., Souvaine, D., Streinu, I., and Struyf, A. (2003). Efficient computation of location depth contours by methods of computational geometry, Statist. and Comp. in press.

Mizera, I. (2002). On depth and deep points: A calculus. Ann. of Statist. 30, 1681-1736.

Mizera, I. and Muller, Ch. H. (2004). Location-scale depth, J. Amer. Statist. Assoc. in revision.

Rousseeuw, P. J. and Hubert, P. (1999) Regression depth (with discussion). J. Amer. Statist. Assoc. 94, 388-402.

Struyf. A. and Rousseeuw, P. J. (2000). High-dimensional computation of the deepest location. Comput. Statist. and Data Anal. 34, 415-426.

Tukey, J. W. (1975). Mathematics and the picturing of data. Proceedings of the International Congress of Mathematicians, Vol 2., Vancouver, B. C., 1974, Canad. Math. Congress, Quebec, 523-531.

Urbanek, S. and Theus M. (2003). iPlots: High interaction graphics in R. Proceedings of the 3rd International Workshop in Distributed Statistical Computing, March 20-22, Vienna, Austria (Kurt Hornik, Friedrich Leisch, and Achim Zeileis, eds.).