np – A Package for Nonparametric Kernel Smoothing with Mixed Datatypes

Jeff Racine

This package provides a variety of nonparametric kernel methods that seamlessly handle a mix of continuous, unordered, and ordered factor datatypes.

All estimation methods are fully multivariate, i.e., there are no limitations on the number of variables one can model (or number of observations for that matter).

Nonparametric methods include unconditional density (distribution), conditional density (distribution), regression, mode, and quantile estimators along with gradients where appropriate, while semiparametric methods include single index and partially linear models.

A number of tests are included such as consistent specification tests for parametric regression and regression quantile models along with tests of significance for nonparametric regression.

A variety of bootstrap methods for computing standard errors, nonparametric confidence bounds, and bias-corrected bounds are implemented.

A variety of bandwidth methods are implemented including fixed, nearest-neighbor, and adaptive nearest-neighbor.

A variety of data-driven methods of bandwidth selection are implemented, while the user can specify their own bandwidths should they so choose (either a raw bandwidth or scaling factor).

A flexible plotting utility, `np.plot()`, facilitates graphing of multivariate objects. An example for creating postscript graphs using the `np.plot()` utility and pulling this into a LaTeX document is provided.

The function `np.kernelsum()` allows users to create or implement their own kernel estimators or tests should they so desire.

The underlying functions are written in C for computational efficiency. Despite this, due to their nature, data-driven bandwidth selection methods involving multivariate numerical search can be time-consuming, particularly for large datasets. A version of this package using the **Rmpi** wrapper is under development that allows one to deploy this software in a clustered computing environment to facilitate computation involving large datasets.