

REU Proposal

Topic: Statistical Analysis of High Frequency Data

Student: Nicholas Humphreys

Faculty Advisor: Lajos Horváth

In several applications (magnetic storms, stock prices, foreign exchange rates) observations are coming like one in every second. Thus, we have an extremely large amount of observations and standard statistical methods cannot be used. We use the model when the observations are treated like a discretized version of a continuous curve. We propose the study of the modifications of the principle component analysis for continuous curves. We hope to describe the data set using only less than five summary data points.

The basic idea is that stochastic processes can be written as infinite sums using an orthonormal system in L_2 . There are several choices to use orthogonal systems in these decompositions. According to the general principle component analysis, the orthonormal system determined by the correlations between the observations should be used. In this case the first three or four terms in the infinite representation can be used and a very large percentage of randomness in the data can be explained.

The mathematical theory of the suggested method is based on the L_2 of stochastic processes and orthogonal functions.