1. Title: **Compressive sampling for variable selection**

Supervisor: Nicolai Meinshausen

In recent years, compressive sampling (or compressed sensing), invented by Emanuel Candes, Terence Tao and David Donoho, has attracted a lot of attention, mostly in the signal processing community. The basic idea is that a set of n measurements / samples of a response variables (e.g. n pixels of a picture) can be projected into a lower-dimensional space with dimension m<<n, by taking for example random Gaussian projections of the original samples, while the original image can still be re-constructed with these m<<n measurements if and only if the original signal (e.g. the image) has a sparse representation in terms of a suitable basis (e.g. wavelet basis for images).

In this project, it is of interest to apply the idea of compressive sampling to variable selection in regression problems with many predictor variables, using the standard framework of compressive sampling, but examining the effect of the projected dimension m (the regularisation) on predictive performance of the chosen model and on the reliability with which a correct sparse regression model can be identified. Data from biological and physical problems are available.

Prerequisites: no specific course required Type of work: involves simulation; can involve (if desired) data analysis and/or theoretical developments

Dudoit, S, Shaffer, J.P. and Boldrick, J.C, Multiple hypothesis testing in microarray experiments, *Statistical Science*, 2003, vol 18 (1), pp 71—103

Liang, C.L, Rice, J.A, de Pater, I, Alcock, C., Axelrod, T., Wang, A. and Marshall, S., Statistical Methods for Detecting Stellar Occultations by Kuiper Belt Objects: the Taiwanese-American Occultation Survey, *Statistical Science*, 2004, vol 19 (2), pp 265- 274,

Benjamini, Y. and Hochberg, Y., Controlling the false discovery rate: a practical and powerful approach to multiple testing, *Journal of the Royal Statistical Society, Series B,* 1995, vol 57 (1), pp 289—300

Romano, J.P. and Wolf, M., Control of Generalized Error Rates in Multiple Testing, Annals of Statistics, 2007, vol 35 (4), pp 1378-1408