2. Title: **Large-scale multiple testing**

Supervisor: Nicolai Meinshausen

Brief Description:
In many modern fields of statistics, the number of simultaneously tested hypotheses is very large, possibly in the thousands or (for the astronomy data that can be used in this project) up to 10^(12) tests. The high multiplicity means that individual testing of hypotheses would potentially incur many false rejections. There are several relevant error rates that can be controlled to mitigate the multiplicity of the testing problem, like the False Discovery Rate (FDR) or the (generalized) Family-wise error rate (FWER). The first part of the project is to give a brief overview of the literature. Control of these error rates requires often knowledge of the distribution of the test statistics (under the null) far out in the tail, which is completely unrealistic. An alternative approach would be to split the data and use part of the data to select a small interesting subset of hypotheses which can then be studied and tested rigorously with the remaining part of the data. The advantage of such a scheme is that much less knowledge is required about the distribution of the test statistics. The goal of the project is to simulate this data-splitting procedure and/or derive theoretical insights about how much power is lost with this data-splitting approach (as opposed to the case where no data-splitting is used). Time permitting, it can be explored if power can actually be gained if the original test statistics are not independent. The methods can be applied to data from the Taiwanese-American Occultation survey, which is searching for rocks in the outer solar system.

Prerequisite courses: none
this can be both a simulation project and/or a theoretical project

Donoho, DL,Compressed Sensing, *IEEE Transactions on Information Theory*, (2006), Vol 52 (4), pp 1289-1306

Candes, E. and Romberg, J, Practical signal recovery from random projections, IEEE Trans. Signal Processing (2005)

Efron, B., Hastie, T., Johnstone, I. and Tibshirani, R., Least angle regression, *Annals of Statistics*, Vol 32 (2), pp 407-499, (2004)