

## SEMINAR OF MATHEMATICS

**FRIDAY January 29, 2016 at 14h00, lecture hall [CM 5](#)**

**Dr. Stefan WAGER** (*Stanford University*) will present a seminar entitled:

### "Statistical Estimation with Random Forests"

#### Abstract:

Random forests, introduced by Breiman (2001), are among the most widely used machine learning algorithms today, with applications in fields as varied as ecology, genetics, and remote sensing. Random forests have been found empirically to fit complex interactions in high dimensions, all while remaining strikingly resilient to overfitting. In principle, these qualities also ought to make random forests good statistical estimators. However, our current understanding of the statistics of random forest predictions is not good enough to make random forests usable as a part of a standard applied statistics pipeline: in particular, we lack robust consistency guarantees and asymptotic inferential tools. In this talk, I will present some recent results that seek to overcome these limitations. The first half of the talk develops a Gaussian theory for random forests in low dimensions that allows for valid asymptotic inference, and applies the resulting methodology to the problem of heterogeneous treatment effect estimation. The second half of the talk then considers high-dimensional properties of regression trees and forests in a setting motivated by the work of Berk et al. (2013) on valid post-selection inference: at a high level, we find that the amount by which a random forest can overfit to training data scales only logarithmically in the ambient dimension of the problem.

This talk is based on joint work with Susan Athey, Bradley Efron, Trevor Hastie, and Guenther Walther.

Lausanne, January 14<sup>th</sup>, 2016 / mg