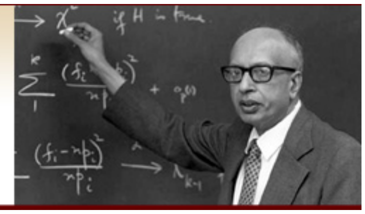
BAHADUR
MEMORIAL LECTURES

In honor of Raj Bahadur's fundamental contributions to statistics and to our department.



The University of Chicago, **Department of Statistics, presents the**

Twenty-first Annual Bahadur Memorial Lectures

Sara van de Geer

Department of Mathematics, ETH Zürich

Monday, April 27, 2020
Location: TBD

"Total Variation Regularization, Part 1"

Let $Y \in \mathbb{R}^n$ be a vector of independent observations with unknown mean $f^0 := \mathbb{E}Y$. We consider the estimator $\hat{f} = \hat{f}_D$ that solves the "analysis problem"

$$\min_{f \in \mathbb{R}^n} \left\{ \|Y - f\|_2^2/n + 2\lambda \|Df\|_1 \right\},$$

where $D \in \mathbb{R}^{m \times n}$ is a given matrix and $\lambda > 0$ is a tuning parameter. An example for D is the difference operator

$$(Df)_i = f_i - f_{i-1},\ i \in [2:n]$$

in which case $\|Df\|_1 = \text{TV}(f)$ is the total variation of f. Other examples include higher order discrete derivatives, total variation on graphs and total variation in higher dimensions. Our aim is to show that the estimator $\hat{f}$ is adaptive. For example, when f^0 is a piecewise linear function, we show that the analysis estimator $\hat{f}_D$, with D the second differences operator, adapts to the number of kinks of f^0. As is the case with the Lasso, the theory for the analysis estimator $\hat{f}$ requires a form of "restricted eigenvalue" condition. We will show that this can be established using interpolating vectors. We will illustrate this (with drawings) for the various examples.

Thursday, April 30th, 2020
3:30 PM, Stevanovich Center,
MS 112, 5727 S. University Avenue

"Total Variation Regularization, Part 2"

Let $Y \in \mathbb{R}^n$ be a vector of independent observations with unknown mean $f^0 := \mathbb{E}Y$. For $f \in \mathbb{R}^n$, let $\text{TV}(f) := \sum_{i=2}^n |f_i - f_{i-1}|$ or, more generally, let $\text{TV}(f)$ be total variation over a graph. It may also be a version of total variation in higher dimensions. We consider the total variation regularized estimator

$$\hat{f} := \arg\min_{f \in \mathbb{R}^n} \left\{ \|Y - f\|_2^2/n + 2\lambda \text{TV}(f) \right\},$$

where $\lambda > 0$ is a tuning parameter. We aim at showing that $\hat{f}$ adapts to the number of constant pieces of f^0. To this end, we will introduce the concept "effective sparsity" (which plays a role similar to "restricted eigenvalue") and show how it can be derived. This requires some bounds from empirical process theory. We will moreover use combinatorial arguments such as counting paths on graphs.

The results can be extended to other loss functions, for instance logistic loss when Y is a vector of binary observations. In that case we place the total variation penalty on the log-odds ratio.