The Annals of Statistics
2006, Vol. 34, No. 6, 2697–2706
DOI: 10.1214/009053606000001082
Main article DOI: 10.1214/009053606000001019

REJOINDER: LOCAL RADEMACHER COMPLEXITIES AND ORACLE INEQUALITIES IN RISK MINIMIZATION

BY VLADIMIR KOLTCHINSKII

Georgia Institute of Technology

I would like to thank the discussants for a number of deep and interesting comments and for their inspiring work on the subject over the years. I will not be able to address all the issues raised in the discussion; I will concentrate just on several of them.

1. Local complexities and excess risk bounds. The first question is about possible ways to define distribution- and data-dependent complexities (such as local Rademacher complexities). The approach taken in my paper is based on geometric and probabilistic properties of the δ-minimal set

$$\mathcal{F}(\delta) := \Big\{ f \in \mathcal{F} : Pf - \inf_{g \in \mathcal{F}} Pg \le \delta \Big\}$$

of the true risk function $\mathcal{F} \ni f \mapsto Pf$. The first quantity of interest is the L_2-diameter of this set, $D(\mathcal{F};\delta)$, and the second one is the function $\phi_n(\mathcal{F};\delta)$ that is equal to the expected supremum of the empirical process indexed by the differences $f-g$, $f,g \in \mathcal{F}(\delta)$. These two functions are then combined in the expression $\bar{U}_n(\delta;t)$ that has its roots in Talagrand's concentration inequalities for empirical processes. The $\sharp$-transform of $\bar{U}_n(\cdot;t)$ (which is just a way to write solutions of fixed point-type equations) is then used to define the localized complexities that provide upper bounds on the excess risk. Under further assumptions, such as mean-variance relationships discussed in detail by Shen and Wang (Bartlett and Mendelson also discuss this and call the function classes satisfying these relationships "Bernstein classes"), these complexities can be redefined in terms of the local L_2-continuity modulus of empirical processes. Since the Rademacher process can be used as a data-dependent bootstrap-type "estimate" of the empirical process, this approach also leads to data-dependent local Rademacher complexities. The use of the whole δ-minimal set is not the only possibility. One can also look at its "slices" $\mathcal{F}(\delta_1,\delta_2] := \mathcal{F}(\delta_2) \setminus \mathcal{F}(\delta_1)$ and define the excess risk bounds in terms of the accuracy of empirical approximation on the slices. One can even make the slices really thin and look at $\{f \in \mathcal{F} : Pf - \inf_{g\in\mathcal{F}} Pg = \delta\}$. This was the approach taken by Peter Bartlett and Shahar Mendelson. Under an additional (and relatively innocent) assumption that the class $\mathcal{F}$ is star-shaped, they established excess risk bounds (and also ratio-type bounds) in terms of complexities of such "thin slices."

Received May 2006.