The Annals of Statistics
2006, Vol. 34, No. 6, 2657–2663
DOI: 10.1214/009053606000001028
Main article DOI: 10.1214/009053606000001019

DISCUSSION: LOCAL RADEMACHER COMPLEXITIES AND ORACLE INEQUALITIES IN RISK MINIMIZATION

BY PETER L. BARTLETT AND SHAHAR MENDELSON

University of California, Berkeley and Australian National University

1. Relating empirical and real structures: additive and multiplicative results. The key issue investigated in Vladimir Koltchinskii's paper is the behavior of an empirical minimizer $\hat{f} \in F$, that is, a function f in F with minimal sample average,

$$P_n f = \frac{1}{n}\sum_{i=1}^{n} f(X_i),$$

where $X_1, \dots, X_n$ are drawn i.i.d. from a probability measure P on $\mathcal{X}$ and F is a class of real-valued functions defined on $\mathcal{X}$. The study of bounds on the expectation $P\hat{f}$ arises in many applied areas, including the analysis of randomized optimization methods involving Monte Carlo estimates of integrals. Motivated by prediction problems that arise in machine learning and nonparametric statistics, the paper makes an important contribution to the study of these bounds, and to the development of model selection methods that exploit the bounds.

The broad approach taken in this paper, and in much earlier work, is to show that the empirical structure (i.e., the collection of sample averages, $P_n f$) is close to the real structure (i.e., the collection of expectations, Pf). If they are close in the additive sense that $\|P_n - P\|_F$ decreases at some rate, then it is clear that $P\hat{f}$ approaches $\inf_{f\in F} Pf$ at that rate. As the paper recalls, there is a tight relationship between the Rademacher process indexed by coordinate projections of the class F and this additive notion of closeness of empirical and real structures. Also, it can be advantageous to consider these properties only locally, that is, in the set $F(\delta) \subset F$ of near-minimizers of Pf. In particular, if the variance of elements of $F(\delta)$ goes to zero with δ, then faster rates are possible through the study of these local properties.

An alternative, developed in the paper, is closeness in the multiplicative sense that for $0 < \varepsilon < 1$, for all functions f in F that have expectations not too small,

$$(1-\varepsilon)P_n f \le Pf \le (1+\varepsilon)P_n f.$$

Again, these results rely on the variance of an element of F decreasing as its expectation decreases. Let us call a class that has this property a Bernstein class.

Received March 2006.