

STANFORD UNIVERSITY
DEPARTMENT OF STATISTICS
BERKELEY-STANFORD SYMPOSIUM

University of California, Berkeley
4:10 p.m., Tuesday, April 16, 2002
Evans Hall, Room 1011

Art B. Owen
Stanford University

Identifying important variables and interactions in black box functions

Quasi-Monte Carlo methods have been successfully used to integrate some very high dimensional functions. Inspection always seems to reveal that those functions are especially simple, having functional anova decompositions dominated by main effects and low dimensional interactions.

This talk will present a method for inspecting black box functions to determine which variables and interactions are most important. The method is constructive, in that provides a computable approximation to the function under study. The approach uses a multivariate orthogonal series expansion of square integrable functions on the unit cube. The coefficients are estimated by quasi-regression: a combination of Monte Carlo sampling and shrinkage.

Though speculative, it is interesting to apply the method to black box functions from machine learning. Such functions can be much simpler than their functional forms suggest. The challenge, still not solved, is to reconcile variable importance determined with respect to a product measure and data decidedly not from a product measure.

(This talk is based on joint work with Tao Jiang.)