The University of Chicago

Department of Statistics

Seminar Series

RUDOLF BERAN

Department of Statistics
University of California, Davis

"A Minimalist's Approach to Fitting
and Extrapolating a Discrete Incomplete Multi-way Layout"

MONDAY October 17, 2005 at 4:00 PM
133 Eckhart Hall, 5734 S. University Avenue
Refreshments following the seminar in Eckhart 110.

ABSTRACT

Reliance on probability models in statistical writing spans the spectrum from none in exploratory data analysis, to fixed effects models, to random effects models, to the broadband randomness of Bayesian reasoning. Statisticians have thereby learned to distinguish among data, probability models, pseudorandom numbers, and algorithms. Data is not certifiably random.

The discrete multi-way layout is an abstract data-type that is associated with regression, experimental designs, digital images or videos, spatial statistics, gene or protein chips, and more. The factor levels are finitely discrete and can be either nominal or ordinal. This talk identifies fitting an incomplete multi-way layout with two tasks: estimating the means of a fixed effects model at the factor level combinations where response data is observed; and extrapolating the estimated means to a larger complete discrete layout.

The approach is through candidate penalized least squares (PLS) estimators for the unknown means. Multiple quadratic penalties are used to express prior conjectures, not assumptions, about each of the main effects and interactions in the ANOVA decomposition of the means, taking into account the nominal or ordinal character of each factor. The candidate PLS estimator that has smallest estimated quadratic risk as the penalty weights vary is shown to attain, asymptotically, the smallest risk attainable over all estimators in the candidate PLS class. All risk calculations are under a minimalist fixed effects model for the observed incomplete layout, a model that makes no assumptions about the unknown means.

The candidate PLS estimators are equivalent to candidate Bayes estimators for a large class of prior distributions. Estimated risks enable the data to comment on the quality of each candidate estimator. Case studies support the pertinence of the asymptotic theory.