Stat 315C, Spring 1997/98

Sampling/Resampling/Simulating/Experimenting

Art Owen

This course surveys those statistical topics in which one can or must choose the data to generate or use. While the course covers a lot of topics, there is considerable overlap between them. Many of the same ideas re-appear under different names. The selection of topics is geared towards those of use in modern statistical methods.

Sampling Whether it is people in a county or entries in a data warehouse, one must often choose a reduced data set to work with. Sampling methods to be considered: Simple random sampling, Stratification and post-stratification, Cluster sampling, Ratio estimation.

Resampling Sample re-use methods allow one to assess a statistical method under weak assumptions. They range from straightforward resampling to more complicated methods. General topics: Jackknife, Bootstrap, bootstrap-t, BCa. For regression: resampling cases, resampling residuals, Wild bootstrap. For classification: .632 bootstrap. Cross-validation: leave one out, k-fold.

Simulation Monte Carlo simulation provides a valuable alternative to asymptotic theory for studying and comparing methods. Topics: non-uniform random variable generation, variance reduction, quasi-Monte Carlo.

Experimental design Factorial and fractional factorial designs over high dimensions. Designs for fitting response surfaces (regressions). Design for computer experiments (approximation). Design for binary responses. Active learning.

Details

The course will present the classical material on these topics. Where possible, the implications for modern computationally intensive statistical and data mining applications will be explored. It is expected that this course will expose lots of holes in the available literature and suggest research topics.

There is no required text. Some recommended texts will be placed on reserve at the Math/CS library. Recent papers and preprints will also be used.

Evaluation will be based on a small number of problem sets and possibly a project. These will be primarily computational.

The prerequisites are as follows: Competence in some computational environment such as Splus, matlab, mathematica, maple, SAS. Doctoral standing or consent of the instructor. Stat 305 or Stat 315A or Stat 315B or experience with modeling or machine learning.