COLLOQUIUM

Raoul LePage

Department of Statistics and Probability, Michigan State University

A New Sampling Based Approach to Multiple Linear Regression

Tuesday, October 25, 2011
A405 Wells Hall
10:20-11:10 a.m.
Refreshments at 10:00 a.m.

Abstract

A sampling approach to ordinary least squares multiple linear regression is proposed in which there is no "true model." Dependent variable or variables are observed at sample units whose values on the independent variables are pre-randomized. In this way nearly homogeneous, independent, normal, random errors are imposed by the experimenter on models of interest. Ordinary least squares is made nearly optimal for estimating the parameters of a great range of multiple linear regression models linking dependent and independent variables in the usual way.

Special pre-randomizations able to achieve this become our reference source of randomness, leading to an inferential model differing from linear models with random errors. Models are judged according to their estimated ability to provide useful statistical descriptions over the support of the distribution of sampling units at which the dependent variables might be observed. Figures below convey the power of these results, in the setting of straight line regression with sample size n = 60:

a. (left) plot of all values of (x, y) that could in principle be observed;
b. (middle) bi-variate plot of the centered joint bootstrap distribution of the least squares intercept and slope (grey) and the corresponding Monte-Carlo relative to imposed randomization, but the vertical axis flipped about zero to facilitate comparison (black);
c. (right) plots of the cumulative centered bootstrap distribution of slope (grey) overlaid with the sampling distribution centered counterpart (black).