RUTGERS UNIVERSITY
DEPARTMENT OF STATISTICS AND BIOSTATISTICS
HILL CENTER #501, BUSCH CAMPUS, PISCATAWAY

www.stat.rutgers.edu

Seminar

Speaker: Dr. Michael Chernick, Director of Biostatistical Services, Lankenau Institute for Medical Research

Title: Bootstrapping Variances: Difficulties with Small Sample Sizes and Highly Skewed or Heavy-tailed Distributions

Date: Wednesday, March 10, 2010

Time: 3:20 p.m.

Place: 552 Hill Center

Abstract

In a rather important paper titled "Qualms about bootstrap confidence intervals", Schenker (1985) used a particular chi-square distribution to show that the percentile method bootstrap and even the BC bootstrap break down for very practical sample sizes when estimating a variance. This caused Brad Efron (1987) to think hard about the problem and he devised the BCa method in a very well-known paper titled "Better bootstrap confidence intervals."

In this talk I present results from a paper that will soon appear in the American Journal of Mathematical and Management Sciences. We revisit the issues surrounding bootstrap confidence intervals by looking at a particularly difficult problem, estimating variances in a nonparametric setting. We show by simulation methods that for certain heavy-tailed and skewed distributions, the convergence of even the second order accurate bootstrap methods BCa, and ABC, must be slow because even at a sample size of nSize=200 the confidence level is not close to the advertised and correct asymptotic level (e.g. for Student's t with 5 degrees of freedom the methods compared are between 4 and 5% below their nominal levels). At nSize =25 all of the methods provide true confidence levels that are at least 5% below their nominal confidence levels. We illustrate this by using confidence levels of 50%, 75% and 90% among others. To investigate more deeply into the convergence properties, we looked at sample sizes nSize=1000 or more (i.e. number of observations in the original data set). To adequately show the pattern of convergence we need the standard deviation of the Monte Carlo approximation of the proportion to be around 0.005. But to achieve this requires something like nRepl =10,000 iterations (i.e. Monte Carlo replications of the bootstrap estimates) of the simulated results!

It is interesting though a bit surprising to us that the first order percentile method bootstrap did nearly as well for these simulations and sometimes better than the higher order bootstraps.