NISS Affiliates Meeting at JSM - July 31, 2016

Guest Speaker Abstracts

David Higdon, PhD
Social Decision Analytics Laboratory at the Biocomplexity Institute of Virginia Tech

A small, biased sample of big data experiences
David M. Higdon (Virginia Tech)

Beginning with some of my first experiences as a faculty affiliate at NISS, most applications I've worked with have had to adapt the statistical methods to deal with computational challenges. Almost any computational bottleneck can be traced back to the same basic problem - there's too much data to carry out the needed computations in the usual way.

This talk will cover three different examples that grapple with big data and computational issues in statistical inference:
- computer model calibration for cosmological inference;
- response surface modeling in big data settings;
- combining varieties of automatically collected data to better manage a supply chain with an industrial partner.

If I manage to explain these efforts before I'm thrown off the speaking podium, I'll share thoughts I've gleaned from these experiences.

Linda J. Young, PhD
Chief Mathematical Statistician and Director of Research and Development of USDA's National Agricultural Statistics Service

Web Scraping: An Application to List Building for the 2015 NASS Local Foods Marketing Survey

The USDA's National Agricultural Statistics Service (NASS) maintains a list frame of all known or potential farms in the U.S. However, the list is not complete, and the lack of coverage is greatest for small farms. For segments of the agricultural economy dominated by small operations, such as local foods farms, estimates of the number of farms and the production of farms can be biased downward if no adjustment is made for under-coverage. NASS has collaborated with the Multi-Agency Collaboration Environment (MACE) to build a list of local foods operations using web scraping that is independent of the NASS list frame. Independent samples from the two frames are being used in a capture-recapture framework to estimate the number of local foods farms and their production, adjusting for under-coverage, non-response, and misclassification. Some of the technical challenges are highlighted.