IMS Public Lecture

Data Mining with Modeling: Managing Diabetes

Speaker: Professor Larry Shepp
Rutgers University

Date: Thursday, 24 April 2008

Time: 6:30pm - 7:30pm

Venue: LT31, Block S16, Science Drive 1
National University of Singapore, Singapore 117543

FREE ADMISSION

About the Speaker

Professor Larry Shepp is renowned for his pioneering and fundamental contributions to discrete tomography and for his work on applications of probability, statistics and mathematics to physics, engineering, communications, genetics and mathematical finance. His work in tomography has a profound influence on biomedical imaging with important applications in medical X-ray and nuclear magnetic resonance technology.

He is a member of the U.S. National academy of Sciences, National Academy of Medicine (Institute of Medicine) and the American Academy of Arts and Sciences. For his research in stochastic processes and computer tomography, he has won awards and recognition from major scientific and professional bodies such as IEEE and Institute of Mathematical Statistics. He is actively involved in editorial work and services for leading journals in probability, imaging sciences and computer assisted tomography.

He was professor of statistics in Stanford University and Columbia University before joining Rutgers University in 1997 and has been the Board of Governor's Professor of Statistics since 2004. Before joining academia, he worked in Bell Laboratories from 1962 to 1980. After joining academia, he continues to contribute his expertise in the service of the medical and engineering industries.

Abstract

Should one allow data to "speak for itself" or should one inject one's preconceptions of the data set at hand with a mathematical model? In the '60's, John Tukey and his followers brought exploratory data analysis into statistics, partly as a revolt against what was then perceived as an overly rigid and brittle mathematical modeling philosophy that held sway at that time. Some problems seem to demand such a purely data-driven approach. Tukey did not want to be biased by preconceived ideas about the origin of the data by formulating a model. Instead, he wanted to allow the data to "speak for itself'", via graphical methods alone.

I will argue that Tukey's approach, as he stated it, does not permit the solution to a problem to depend on the problem; and thereby inhibits statistics to grow and interact with the rest of science.

I will illustrate my point with data-mining examples, in particular discussing a new large data set composed of glucose levels of blood of a large number of diabetics at 5 minute intervals over a period of a year to study the important problem of how to make algorithmic use of these readings for closed-loop control of an insulin pump.

Organized by

**Institute for Mathematical Sciences
National University of Singapore**