Statistical Engineering Division Seminar

Thursday January 26, 1995 1:30pm
Adminstration Bldg., Lecture Room C
NIST
Gaithersburg, MD

Estimating a DNA Sequence From a Collection of Noisy Copies

Herman Chernoff

Professor of Statistics
Harvard University

The Human Genome project proposes to sequence the DNA molecules that make up the chromosomes that constitute the blueprint for a human. These chromosomes are long double helical chains, each of which can be described by a sequence, each element of which is one of four nucleotides. It is estimated that there are roughly three billion nucleotides in the human genome and, with current technology, it costs several dollars per nucleotide to list a section of the chromosomes.

There is a proposal to build a device which can sequence a section or subsequence of about 40,000 nucleotides in a short time. However, it is anticipated that the sequences derived from this device will be noisy copies of the original section, and several noisy copies and their analysis will be required for accuracy.

An algorithm is proposed for estimating the sequence from noisy copies. Properties of this algorithm are estimated, based on a simple probability model which includes assumptions about the rates of errors due to substitution, transposition, deletion of small sections and insertion of small false sections. There is also a problem of registration, since unknown parts of the subsequence may be chopped off at both ends.

Herman Chernoff received a B.S. degree in Mathematics from the City College of New York in 1943, and Sc.M. and Ph.D. degrees from Brown University in 1945 and 1948, respectively. He wrote his Ph.D. dissertation in absentia at Columbia University under the supervision of Abraham Wald. In addition to numerous visiting positions, he has taught statistics at the University of Illinois, Stanford University, M.I.T., and is presently Professor of statistics at Harvard University. During his distinguished career, Professor Chernoff has made major contributions in many areas of statistics, including large deviations, asymptotic theory, sequential analysis, decision theory, statistical graphics, classification theory, and optimal experimental design.

Professor Chernoff has been the recipient of numerous honors. He is a fellow of the American Statistical Association and of the Institute of Mathematical Statistics, a past president of the Institute of Mathematical Statistics, an elected member of the American Academy of Arts and Sciences, the National Academy of Sciences, and the International Statistical Institute, and a recipient of the Samuel S. Wilks award of the American Statistical Association.

[NIST contact: Mark Vangel, 301-975-3208]