

STATISTICS COLLOQUIUM

WEDNESDAY, AUGUST 28, 2013
TALK: 11:00 AM — SCIENCE CENTER RM. 705

“How Different People Are: A Biclustering Analysis of Human Genome”

Lei Guo
Department of Statistics
Harvard University

ABSTRACT

People are different. However, any two individuals in the world share more than 99% of their DNA. It's the less than 1% differences that make everyone unique. Among the many forms of DNA variations, single-nucleotide polymorphism (SNP) is the most common type of sequence variations, which occurs at nucleotide level. The international HapMap project has made available the SNP data of thousands of individuals across the world. We describe a model-based Biclustering method for using multilocus genotype data to infer the similarities and differences between human populations. In contrast to existing methods, our method can locate the SNPs that are specific to given subpopulation groups. We show that the method can produce highly accurate classification of populations using individual genotype data and locate the differences between population groups. The Biclustering process can be used as a variable selection step prior to existing population inference procedures. The algorithm can also provide insight to the genome-wide association study (GWAS) by finding SNPs that are common to different ethnic groups.