



BIO
MATH

Statistical and Computational Methods for Ancestry Estimation and Variable Selection in Genome-Scale Datasets



David Alexander

Doctoral Graduate Student
Department of Biomathematics
UCLA

**Friday, May 6, 2011
11:00 AM**

14-214U Center for the Health Sciences (CHS)

ABSTRACT: As genotyping and sequencing technologies reach higher and higher throughput levels, genetic datasets are becoming ever larger, creating a growing need for highly efficient algorithms for routine analyses. Our work on the efficient individual ancestry estimation program ADMIXTURE has shown that easily-implemented and stable EM algorithms, widely believed to be a good choice for estimation in large datasets, can sometimes prove vastly inferior to more intricate coordinate- and block-relaxation approaches. Furthermore, our work on a novel quasi-Newton convergence acceleration procedure shows that the efficiency of existing iterative optimization algorithms can be greatly improved with no change to the statistical model and only minor implementation effort. For secondary analyses, computationally intensive methods are more tolerable. In this vein, we explore the application of some new developments in bootstrap aggregation and high-dimensional variable selection to genome-wide association, to see whether more complex models can be used to wring more information from previously studied association data.

Doctoral Committee: Dr. Kenneth Lange, Dr. John Novembre,
Dr. Janet Sinsheimer, Dr. Elliot M. Landaw