



THE UNIVERSITY OF CHICAGO

Department of Statistics

STATISTICS COLLOQUIUM

PO-RU LOH

Department of Epidemiology
Harvard School of Public Health

Bayesian Mixed Model Association Statistics in Linear Time

MONDAY, May 19, 2014, at 4:00 PM

Eckhart 133, 5734 S. University Avenue

Refreshments following the seminar in Eckhart 110.

ABSTRACT

Linear mixed models (LMM) are a powerful statistical tool for identifying loci associated to phenotypes and avoiding confounding. Mixed model analysis is computationally demanding, however, and is becoming infeasible as study sizes reach the scale of 100,000 samples. Existing algorithms rely on spectral analysis of a genetic relationship matrix (GRM) at total time cost $O(MN^2)$, where M is the number of markers and N is the sample size. Additionally, these methods implicitly assume an infinitesimal genetic architecture in which all markers are causal. I will present a fast $O(MN)$ -time mixed model association algorithm, BOLT-LMM, which increases power by generalizing the LMM to model non-infinitesimal (sparse) genetic architectures via a Bayesian mixture prior on marker effect sizes, used within a retrospective hypothesis testing framework. BOLT-LMM performs a variational iteration that circumvents computing the GRM by operating directly on raw genotypes stored compactly in memory. When specialized to the infinitesimal model, BOLT-LMM achieves additional speedup, matching existing methods at dramatically reduced time and memory cost. I will describe preliminary results of applying BOLT-LMM to analyze 60,000 samples from the recently released Genetic Epidemiology Research on Aging (GERA) data set.

For further information and about building access for persons with disabilities, please contact Kirsten Wellman at 773.702.8333 or send email (kwellman@galton.uchicago.edu). If you wish to subscribe to our email list, please visit the following website: <https://lists.uchicago.edu/web/arc/statseminars>.