COLLOQUIUM

Department of Statistics and Probability
Michigan State University

Heng Wang
Michigan State University

Copy Number Variation Detection Using Next Generation Sequencing Read Counts

Thursday, January 24, 2013
A405 Wells Hall
10:20 a.m. - 11:10 a.m.
Refreshments: 10:00 a.m.

Abstract

A copy number variation (CNV) is a difference between genotypes in the number of copies of a genomic region. Next generation sequencing (NGS) technologies provide sensitive and accurate tools for detecting genomic variations that include CNVs. We propose a new methodology for detecting CNVs using NGS data. This method (henceforth denoted by m-HMM) is based on a hidden Markov model with emission probabilities that are governed by mixture distributions. We use the Expectation-Maximization (EM) algorithm to estimate the parameters in the model. A simulation study demonstrates that our proposed m-HMM approach improves upon existing methods. We apply the m-HMM method to NGS data from the two maize inbred lines B73 and Mo17 to identify CNVs that may play a role in creating phenotypic differences between these inbred lines. We show that the results of our m-HMM analysis is concordant with previous array-based efforts to identify CNVs.

Key Words: Count data; Gamma-Poisson mixture; Hidden Markov model; Plant genomics; Poisson mixture model.

To request an interpreter or other accomodations for people with disabilities, please call the Department of Statistics and Probability at 517-355-9589.