

19a5-Advances in Computational Biology
December 19 (Monday), 9:00 a.m. - 10:30 a.m., AC 2nd Conference Room
Organizer: Shwu-Rong Grace Shieh
Chair: Shwu-Rong Grace Shieh

19a5-1 On Automated Flow Cytometric Data Analysis

Geoff McLachlan
Department of Mathematics and Institute for Molecular Bioscience,
University of Queensland, Brisbane, Australia

Flow cytometry is widely used for single cell interrogation of surface and intracellular protein expression by measuring fluorescence intensity of fluorophore-conjugated reagents. We report some extensions to our previous work (Pyne et al., 2009, PNAS 106) that provides a procedure called FLAME (FLow analysis with Automated Multivariate Estimation) for automated high-dimensional flow cytometric analysis. The FLAME procedure, which has been incorporated with the GenePattern package of the Broad Institute, uses finite mixture models of heavy-tailed and asymmetric distributions in the form of the skew t-component distributions to identify and model cell populations in a flow cytometric sample. In our more recent work, we extend the component distributions in FLAME to include skew t-distributions without any restrictions on the form of skewness. A program is given using the EM algorithm that is able to fit these models exactly without need for Monte Carlo computation. It thus broadens the applications of flow cytometry to new biological and clinical problems.

[Geoff McLachlan, Department of Mathematics and Institute for Molecular Bioscience, University of Queensland, Brisbane, Australia; g.mclachlan@uq.edu.au]

19a5-2 Statistical Learning Based on Distributions of Oligonucleotides in DNA Sequences

Probal Chaudhuri
Theoretical Statistics & Mathematics Unit, Indian Statistical Institute,
India

I will start with some examples from evolutionary biology to demonstrate how statistical analysis of distributions of various oligonucleotides in DNA sequences can lead to important biological discoveries. The problems considered can be formulated as statistical learning problems that may be supervised, unsupervised or partially supervised in nature. The statistical analysis involves an interesting variable selection problem. This motivated developing certain probabilistic models for DNA sequences, and those will be discussed.

[Probal Chaudhuri, Theoretical Statistics & Mathematics Unit, Indian Statistical Institute, 203 B. T. Road, Kolkata700108, India; probal@isical.ac.in,

probalchaudhuri@gmail.com]

19a5-3 **Sequential Lasso for Feature Selection with Ultra-high Dimensional Feature Space**

Zehua Chen
Shan Luo

Department of Statistics and Applied Probability, National University of Singapore

We propose a novel approach, Sequential Lasso, for feature selection in linear regression models with ultra-high dimensional feature spaces. In this talk, we discuss the asymptotic properties of Sequential Lasso, especially, its selection consistency. Like other sequential methods, the implementation of Sequential Lasso is not limited by the dimensionality of the feature space. However, it has advantages over other sequential methods. The simulation studies comparing Sequential Lasso with other sequential methods and demonstrating the advantages of Sequential Lasso will be reported.

[Zehua Chen, Department of Statistics and Applied Probability, National University of Singapore; stachenz@nus.edu.sg]