For Favour of Posting

THE CHINESE UNIVERSITY OF HONG KONG

Department of Statistics

will present a seminar entitled

Estimation of Covariance Matrices Through Penalisation

by

Mr Clifford Lam
Princeton University

on

Tuesday, 12 February 2008
2:00pm – 3:00pm

in

Lady Shaw Building C5
The Chinese University of Hong Kong

Abstract:

Estimation of covariance or inverse covariance (precision) matrix has always been a fundamental problem in statistical inferences and multivariate analyses, from risk management and portfolio allocation in finance, to input in various statistical procedures like the Principal Component Analysis (PCA) or the Linear/Quadratic Discriminant Analysis (LDA/QDA) in classification, among many others. Also, with the advance of technologies, data with number of variables comparable to or even larger than the sample size is common nowadays.

In this talk, we first present a general penalisation framework for estimation of covariance or inverse covariance matrices, and argue that a non-convex, unbiased penalty like the SCAD or hard-thresholding penalty, performs better than the L-1 penalty (the LASSO) in sparse structure exploration under mild sparsity condition of the true underlying covariance or precision matrix.

In the second part of the talk, we focus on the estimation of inverse covariance matrices for data with a natural ordering or a notion of distance between variables. Through the introduction of a block penalty on the blocks of off-diagonals of the modified Cholesky factor for the covariance matrix of interest, we show a notion of sign consistency for the resulting estimator when number of variables is much larger than the sample size, where blocks of zero off-diagonals are estimated to be zero, and blocks of non-zero off-diagonals are estimated to be non-zero, all with probability goes to one. We also prove a rate of convergence in the matrix operator norm, which links to the non-sparsity rate of the true inverse covariance matrix, for the resulting estimator under certain tail assumptions of the variables. Hence the data is not required to be normally distributed for the results to hold. Various simulation settings demonstrate the effectiveness and flexibility of the method.

All are Welcome