APPLIED MATHEMATICS COLLOQUIUM

EIGENVALUES OF LARGE DIMENSIONAL SAMPLE COVARIANCE MATRICES

JACK SILVERSTEIN
North Carolina State University

ABSTRACT:

The talk will consist of a review of the latest information on the eigenvalues of matrices of the form $B_n = (1/N)SXX^*S^*$, where X is an $n \times N$ matrix containing i.i.d. standardized entries, and S is $n \times n$, independent of X. This matrix can be viewed as the sample covariance matrix of N samples of the random vector $SX_{.1}$ having population matrix $T = SS^*$ (which includes all Wishart matrices). It is assumed that $n \to \infty$ with $n/N \to c > 0$ and the empirical distribution function (d.f.) of the eigenvalues of T converges to a nonrandom d.f. H. The results are important to multivariate inference for situations when the dimension is large but the sample size needed to apply standard multivariate methods is unattainable. They show that viable information can be obtained when the sample size is on the same order of magnitude as the vector dimension.

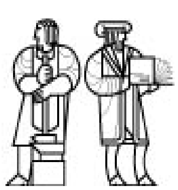
Strong (almost sure) limiting behavior of the eigenvalues of B_n is known. The limiting d.f. F is nonrandom, depends on c and H, and has a continuous derivative. Moreover, for n large, individual eigenvalues behave exactly as one would intuitively expect from the shape of F'. An example of this can be seen in the two graphs at my website www.math.ncsu.edu/~jack. Here $n = 200$, $N = 4000$, and T has 3 distinct eigenvalues: 1, 3, and 10, with respective multiplicities 40, 80, and 80. The limit theorem on the empirical d.f. explains the shape of the histogram. The scatter plot of the eigenvalues of B_n reveals what has also been mathematically verified: the exact number of eigenvalues of B_n appearing in each interval in the support of F, 40 in the leftmost interval and 80 in each of the other two.

Essentially, the only question remaining is how fast F_n, the empirical spectral d.f. of B_n, converges to F. Evidence will be given to suggest that the rate of convergence is $1/n$, due to results on the distributional behavior of linear functionals (spectral statistics) defined by the eigenvalues (joint work with Z.D. Bai at National University of Singapore).

MONDAY, MARCH 8, 2004
4:15 PM
Building 2, Room 105

Refreshments at 3:30 PM in Building 2, Room 349

Applied Math Colloquium: http://www-math.mit.edu/amc/spring04
Math Department: http://www-math.mit.edu



Massachusetts Institute of Technology
Department of Mathematics
Cambridge, MA 02139