Topics on the Eigenvalues of Large Dimensional Sample Covariance Matrices

Jack W. SILVERSTEIN, *Department of Mathematics, North Carolina State University, USA*

Key words: Sample covariance matrices, Sample correlation matrices, Spearman's rank correlation matrices

Mathematical Subject Classification: 60H15, 62H99

Abstract:

The talk will outline recent work on spectral properties of random matrices of sample covariance type. One topic to be covered concerns the extreme eigenvalues of a subclass of matrices which have been extensively studied. The general class is defined as $B_p = (1/n)T_p^{1/2}X_pX_p^*T_p^{1/2}$ where $X_p = (X_{ij})$ is $p \times n$ with i.i.d. complex standardized entries, and $T_p^{1/2}$ is a Hermitian square root of the nonnegative definite Hermitian matrix T_p. This matrix can be viewed as the sample covariance matrix of n i.i.d. samples of the p dimensional random vector $T_p^{1/2}(X_p)_{\cdot 1}$. It is known that if $p/n \to c > 0$ and the empirical distribution function (e.d.f.) of the eigenvalues of T_p converge as $p \to \infty$, then the e.d.f. of the eigenvalues of B_p converges a.s. to a nonrandom limit. This result is relevant in situations in multivariate analysis where the vector dimension is large, but the number of samples to adequately approximate the population matrix (required in standard statistical procedures) cannot be attained.

The subclass consists of matrices B_p where all but a finite number of eigenvalues of T_p are 1, which has been called the "spike population model". Results are obtained for the limiting behavior of those eigenvalues of B_p which correspond to the population ones which deviate from 1 (joint work with Jinho Baik at University of Michigan).

Another topic is on the class of matrices of the form $C_p = (1/n)(R_p + \sigma X_p)(R_p + \sigma X_p)^*$ where X_p is as in B_p, $\sigma > 0$, and R_p is $p \times n$ random, independent of X_p with the spectral e.d.f. of $(1/n)R_pR_p$ converging to a nonrandom limit. These matrices model situations where information is contained in the sampling of the vectors $R_{\cdot 1} \cdots R_{\cdot n}$, but the received vector is contaminated by additive noise (the columns of σX_p). The e.d.f. of the eigenvalues of C_p also converges a.s. as $p \to \infty$ (with $p/n \to c > 0$), the analysis of the limiting distribution being currently under investigation. An example of where C_p can be used will be discussed. It is the detection problem in array signal processing: determining the number of sources (presumed large) impinging on a bank of sensors in a noise filled environment (joint work with Brent Dosier at NCSU).