Partha Dey Talk

Date: Friday, January 18, 2013

Time: 4:00 pm in MATH 350

Energy Landscape for large average submatrix detection problems in Gaussian random matrices

ABSTRACT

Combinatorial optimization problems such as finding submatrices with large average value within a large data matrix arise in a wide array of fields, ranging from statistical genetics, bioinformatics, computer science to various social sciences. We analyze asymptotics for such problems in an idealized setting where the underlying matrix is a large Gaussian random matrix and provide detailed asymptotics for various characteristics of the energy landscape for such problems. For fixed k we provide a structure theorem for the $k\times k$ submatrix with the largest average. We also show that for any given $\tau> 0$, the size of the largest square sub-matrix with average bigger than τ satisfies a two point concentration phenomena. Finding such submatrices for a fixed k is a computationally intensive problem. We study the natural algorithm that attempts to find submatrices with large average; such algorithms typically converge to a local optimum. We prove a structure theorem for such locally optimal sub-matrices and derive refined asymptotics for the mean and the variance for $L_n(k):=$ number of such local optima. In particular for $k=2$ and $k=3$, the order of the means are n^2 and n^3, while the variances are $n^{8/3}$ and $n^{9/2}$, respectively, with logarithmic corrections. We develop a new variant of Stein's method to prove a Gaussian Central Limit Theorem for $L_n(k)$ for all finite k which has a highly complex dependency structure. Partly based on joint work with Shankar Bhamidi and Andrew Nobel.