Department of Applied Mathematics and Statistics
The Johns Hopkins University

STUDENT SEMINAR

Elizabeth (Libby) Beer
Dept. of Applied Mathematics & Statistics
The Johns Hopkins University

Tuesday, October 18, 2005
304 Whitehead Hall
4:00 p.m.

SCAN STATISTICS ON GRAPHS: THEOREMS, CONJECTURES, AND INVESTIGATIONS

Abstract

In their original context of image analysis, scan statistics—better known as moving window analyses—are a generalization of the quadrat counts of Fisher. A "window" of some sort is moved across a data field (an image, for example), and a statistic is calculated at each window position (in Fisher's example, the number of bacterial clumps in the window). The "moving window" idea can also be used to understand scan statistics on graphs. In general, given a graph $G = (V(G), E(G))$ and a local statistic ψ that can be applied to each node of the graph (for example, a function evaluated on a neighborhood of each node), the *scan statistic* $M(G)$ is the maximum, over all nodes in the graph, of this local statistic:

$$M(G) := \max_{v \in V(G)} \psi(v).$$

In many cases, the $\arg\max$ of the locality statistic is also significant.

Our goal, in the analysis of communication graphs, is to detect "local subregions of excessive activity". We wish to understand the distribution of our first-order scan statistic under the null hypothesis–no such excessively active subregions—in order to create effective, principled tests that reject this null hypothesis when a graph does contain such subregions. We investigate two null hypotheses: one in which a "quiet" null-hypothesis graph is modeled as an Erdős-Rényi random graph, and one in which a null-hypothesis graph is modeled using the random dot-product graphs of Scheinerman.

Priebe and Kapur conjecture that, under the Erdős-Rényi null hypothesis, the distribution of M_1 (the first-order scan statistic) is also Gumbel. They conjecture, furthermore, that the location parameter a_1 of this Gumbel distribution (in its leading term) is asymptotically $r_{a_1} n^{s_{a_1}}$, where n is the number of vertices in the graph and r and s are constants. More specifically, they have surmised that $s = 2$. Monte Carlo simulations do not contradict the first two conjectures, but suggest that s is a (non-constant) function of p, the edge probability for the Erdős-Rényi random graph.

We will present the results of our Monte Carlo investigation of these and related conjectures.