

**SOMS Friday Seminar**  
**January 25, 2008**  
**SMC 305**  
**2:30-3:25**

**Are there patterns in lottery numbers?**  
**Tests of uniformity in multidimensional datasets**

**Adam Petrie**

Assistant Professor

Department of Statistics, Operations & Management Science  
University of Tennessee

**ABSTRACT**

An important yet not often studied problem in high-dimensional data analysis is determining whether data points are uniformly distributed over some compact domain, or instead possess some underlying structure (e.g., clumps, clusters, or other non-homogeneities). We use the total length of the minimum spanning tree to test for uniformity in the unit hypercube. We also consider the number of runs and maximum run length in the sequence of segment lengths of a Hamiltonian path (or “snake”) constructed through the data to confirm or reject uniformity of more arbitrarily shaped compact domains. We compare the power of these new statistics for testing uniformity with three others from the literature, both inside the unit hypercube and on other possibly non-convex domains, against a variety of non-uniform alternatives. The snake has superior power when the support of the data is unknown. Further, we discuss how the snake can estimate the number and locations of clustering when uniform noise is prevalent in the dataset. We conclude by examining several real-world datasets for the presence of underlying structure or patterns, including the Minnesota Hot Lotto and Powerball winning numbers over the years.

**Refreshments provided after seminar**