

Stanford University
Department of Statistics

DEPARTMENTAL SEMINAR

4:30pm, Tuesday, October 2, 2018
Sloan Mathematics Center Room 380C

Refreshments served at 4pm in Sequoia Lounge.

Speaker: David Donoho, *Stanford Statistics*

Title: **The Statistical Significance of Perfect Linear Separation**

Abstract:

Suppose we have a set of n iid normally distributed points in dimension d . A pre-specified subset of k points happens to be perfectly linearly separable from the other $n - k$. What is the exact probability of such an occurrence? With $k = 1$ and $d = 2$ or 3 , this was solved in Brad Efron's 1965 *Biometrika* paper, who evaluated the expected number of vertices of the convex hull of a random set of n points in dimensions $d = 2$ and $d = 3$.

We will describe exact formulas for evaluating such quantities in arbitrary n , k , and d . These formulas involve a new distribution, which we call the underdispersed Binomial, and a novel problem in integral geometry, involving the intersection of a hyperplane with a certain highly symmetric but seemingly little-studied cone.

Exact formulas for general problems of this type were derived by Affentranger and Schneider and by Vershik and Sporyshev. Such formulas are typically not easy to work with, but in our case we are able to get explicit results and produce software. We are also able to complete a large-deviations analysis and identify a number of interesting phase transitions.

This is joint work with Hatef Monajemi (Stanford) and Jared Tanner (Oxford).