Combinatorics from biology: inference functions and sequence alignment

Sergi Elizalde

Dartmouth College

Thursday, January 12, 2006
L02 Carson Hall, 4:00 pm
(Tea 3:30 pm Math Lounge)

Abstract

Statistical models are used in computational biology to draw conclusions from data. They give rise to several combinatorial problems. Sometimes the goal is to find the optimal alignment of a pair of DNA sequences, or to determine what parts of the genome will be translated into proteins. The functions that map each observation to its most probable explanation are called inference functions. They depend on the model parameters.

Even though the number of maps from the set of observations to possible values of the hidden data is doubly exponential, it turns out that most of these maps are not inference functions for any value of the parameters. I will show that for any graphical model with a fixed number of parameters, the number of inference functions is polynomial in the size of the model. The proof reduces the enumeration of inference functions to counting the vertices of a certain lattice polytope.

Then I will give applications of this result to optimal sequence alignment, and discus some open combinatorial problems that arise.

No previous knowledge of combinatorics, biology, or statistics will be assumed.

This talk should be accessible to undergraduates.   1136236714