

COMBINATORICS SEMINAR

Inference Functions and Sequence Alignment

Sergi Elizalde
MSRI

February 25, 2005

Abstract:

Statistical models are used to solve certain problems in computational biology, such as determining what parts of the genome will be translated into proteins, or how a DNA sequence evolved into another via a series of mutations, insertions and deletions. Each answer has a certain probability depending on the model parameters. When these are given, the most probable answer, called explanation, is obtained by solving a combinatorial optimization problem. The map sending each observation to its explanation is called an inference function. Bounding the number of inference functions is important because it indicates how the parameters may affect the solution.

Using some theory about lattice polytopes, I will prove that the number of inference functions of any graphical model is polynomial in the size of the model. Then I will give applications to optimal sequence alignment, and discuss some open combinatorial problems that arise.