

# STATISTICS COLLOQUIUM

MONDAY, May 6, 2013

TALK: 4:15 PM — SCIENCE CENTER RM. 705

RECEPTION: 3:50 PM — SCIENCE CENTER, 7TH FLOOR

## **“Efficient Sampling Methods for Protein Structure Refinement”**

**Samuel Wong**  
**Department of Statistics**  
**Harvard University**

### ABSTRACT

In protein folding, scientists are interested in the prediction of the three-dimensional structure, based on the amino acid sequence. Initial structures of new proteins are often built by finding templates from databases of proteins with known structure; this procedure is called homology modeling in the bioinformatics literature. The goal of refinement is to generate a structure prediction that improves upon a given homology model, especially in regions where a good template is not available.

The conformational space of a protein is too large for an exhaustive search to be computationally feasible. An efficient sampling algorithm is therefore crucial to the success of any refinement procedure. In this talk, we first focus on segments that are not well modeled by matching templates; the prediction of these segments is known as loop modeling. We present a new loop sampling method, based on sequentially guiding the sampling distributions of each amino acid to obey geometric and energetic constraints. We then describe extensions to the general refinement problem, where Monte Carlo optimization of the energy function on the entire protein is desired. Our method is compared with existing ones on real protein datasets, and encouraging results are obtained thus far.