

Mathematics Colloquium

Mathematical Models in Error-Prone Data Regimes

Mario Banuelos
Department of Mathematics
Cal State, Fresno

Friday, October 26, 2018
4:10 – 5 p.m.
Building 53 Room 206

Abstract

Every time a cell divides, its entire DNA sequence is duplicated. Each time this information is copied, there is an opportunity for the introduction of errors such as insertions and deletions. Both the introduction and accumulation of genomic variation may lead to speciation events or to development of genetic diseases. As the costs of genome sequencing decreases, the volume of sequencing data has resulted in the need for advanced mathematical and computational methods. Unlike highly annotated reference genomes available in public repositories, sequenced genomes from large public repositories tend to suffer from errors in both sequencing and mapping. In this talk, I will discuss mathematical models in error-prone data regimes to improve predictions of genomic variation in organisms. We predict genomic variation between members of the same species by developing a constrained-optimization framework using gradient-based methods and constrain our solution with a sparsity-promoting ℓ_1 penalty to detect structural variants (SVs) in family lineages. As machine and deep learning models become more common, we investigate the role of such methods in this genomics context by considering the reconstruction of images affected by similar noise.

About the speaker: I was born and raised in the small, agricultural town of Delano, California. I went on to earn my B.A. in Mathematics from California State University, Fresno (Fresno State) and I obtained my Ph.D. in Applied Mathematics (2018) from the University of California, Merced under the guidance of my advisor Prof. Suzanne Sindi. My research interests include mathematical biology, optimization, statistical models for genome evolution, and data science. Most data I consider in my work is low quality or corrupted by noise, and I am currently working on developing and applying machine learning methods for arbitrary architectures with applications in signal reconstruction. Instead of relying on high quality and often expensive data to improve results, I am applying these machine learning models to a multitude of low-quality biological (e.g., genomics) data of related individuals or species.

Cookies will be provided before the talk at 4 p.m.
in the same room as the talk, Building 53 Room 206.