# BIOSTATISTICS SEMINAR

**Allison J.B. Chaney, PhD**

Assistant Professor
of Marketing,
Fuqua School of
Business,
Duke University

## Nonparametric Deconvolution Models

In this talk, I will describe nonparametric deconvolution models (NDMs), a family of Bayesian nonparametric models for collections of data in which each observation is the average over the features from heterogeneous particles. For example, these types of data are found in elections, where we observe precinct-level vote tallies (observations) of individual citizens' votes (particles) across each of the candidates or ballot measures (features), where each voter is part of a specific voter cohort or demographic (factor); similar structure exists in data collected from bulk RNA sequencing.  Like the hierarchical Dirichlet process, NDMs rely on two tiers of Dirichlet processes to explain the data with an unknown number of latent factors; each observation is modeled as a weighted average of these latent factors. Unlike existing models, NDMs recover how factor distributions vary locally for each observation. This uniquely allows NDMs both to deconvolve each observation into its constituent factors, and also to describe how the factor distributions specific to each observation vary across observations and deviate from the corresponding global factors. I will present variational inference techniques for this family of models and describe its performance on simulated data and voting data from California to show that including local factors improves estimates of global factors and provides a novel scaffold for exploring data.

**Thursday April 9, 2020**                    **3:30 pm - 4:30 pm**
**Zoom meeting:  Please also find a link in the email invite.**
**https://uncsph.zoom.us/j/893328432?pwd=MnRUWitYaEZva3ZpYzhGMkJ0VktBQ**
**T09&_ga=2.141183730.1187571104.1585657287-1769805331.1548686859**

UNC
GILLINGS SCHOOL OF
GLOBAL PUBLIC HEALTH