

# Stanford University

## Department of Statistics

### DEPARTMENTAL SEMINAR

4:30pm, Tuesday, February 4, 2020  
Sloan Mathematics Center Room 380C

Refreshments served at 4pm in Sequoia Lounge.

**Speaker:** Paromita Dubey, *UC Davis*

**Title:** **Fréchet Change Point Detection**

#### **Abstract:**

Change point detection is a popular tool for identifying locations in a data sequence where an abrupt change occurs in the data distribution and has been widely studied for Euclidean data. Modern data very often is non-Euclidean, for example, distribution valued data or network data. Change point detection is a challenging problem when the underlying data space is a metric space where one does not have basic algebraic operations like addition of the data points and scalar multiplication.

In this talk, I propose a method to infer the presence and location of change points in the distribution of a sequence of independent data taking values in a general metric space. Change points are viewed as locations at which the distribution of the data sequence changes abruptly in terms of either its Fréchet mean or Fréchet variance or both. The proposed method is based on comparisons of Fréchet variances before and after putative change point locations. First, I will establish that under the null hypothesis of no change point the limit distribution of the proposed scan function is the square of a standardized Brownian Bridge. It is well known that such convergence is rather slow in moderate to high dimensions. For more accurate results in finite sample applications, I will provide a theoretically justified bootstrap-based scheme for testing the presence of change points. Next, I will show that when a change point exists, (1) the proposed test is consistent under contiguous alternatives and (2) the estimated location of the change point is consistent. All of the above results hold for a broad class of metric spaces under mild entropy conditions. Examples include the space of univariate probability distributions and the space of graph Laplacians for networks. I will illustrate the efficacy of the proposed approach in empirical studies and in real-data applications with sequences of maternal fertility distributions. Finally, I will talk about some future extensions and other related research directions, for instance, when one has samples of dynamic metric space data.

This talk is based on joint work with Professor Hans-Georg Müller.