

**“Informative Priors for Clustering”**

**Amy H. Herring, Sc.D.**

**Sara & Charles Ayres Professor**

**Department of Statistical Science & Global Health Institute**

**Duke University**

# Abstract

There is a very rich literature proposing Bayesian approaches for clustering starting with a prior probability distribution on partitions. Most approaches assume exchangeability, leading to simple representations in terms of Exchangeable Partition Probability Functions (EPPF). Gibbs-type priors encompass a broad class of such cases, including Dirichlet and Pitman-Yor processes. Even though there have been some proposals to relax the exchangeability assumption, allowing covariate dependence and partial exchangeability, limited consideration has been given on how to include concrete prior knowledge on the partition. For example, we are motivated by an epidemiological application, in which we wish to cluster birth defects into groups and we have prior knowledge of an initial clustering provided by experts but that is not viewed as definitive.

As a general approach for including such prior knowledge, we propose a Centered Partition (CP) process that modifies the EPPF to favor partitions close to an initial one. Some properties of the CP prior are described, a general algorithm for posterior computation is developed, and we illustrate the methodology through an application to clustering birth defects.

# 12:00 Noon, Tuesday, February 4, 2020

# 47 College Street, 106B

# 11:45 AM Lunch served outside Rm. 106B

