Traditional statistical methods and results typically rely on linearity of the data. However, in many applications, data lie on nonlinear geometric structures such as Riemannian manifolds or more generally stratified spaces. Examples include corpus callosum shape data encoded as a Riemannian manifold, the evolutionary data encoded in phylogenetic trees, the space of persistence diagrams in topological data analysis and clustering problems in genotype data. These examples naturally lead to a problem of sampling from stratified spaces. However, statistical theory on stratified spaces has only been studied on specific examples of particular stratified spaces; examples include graphs, open books, and hyperbolic flat 2-dimensional spaces with an isolated singularity. Therefore, there is considerable demand for establishing a mathematical foundation of statistical theory on stratified spaces before statistics can be confidently tested.



Do Tran PhD Thesis

## **Sampling from Stratified Spaces**

Duke MATH



My dissertation studies Central Limit Theorems (CLTs) of Fréchet means on stratified spaces. The broad goal of the work is to answer the following question: What information one should expect to get by sampling from a stratified space? In particular, my thesis explores relationship between geometry and different forms of CLTs, namely classical, smeary and sticky. The work starts with explicit forms of CLTs for spaces of constant sectional curvature. As a consequence, we explain the effect of sectional curvature on behaviors of Fréchet means. We then give a sufficient condition for a smeary CLT to occur on spheres. The most important part of my thesis is a conjectural general form of CLT for starshaped Riemannian stratified spaces. The general CLT we propose is universal in the sense that it contains all of the aforementioned forms of CLTs for Fréchet means. The proposed CLT is verified on manifolds and on hyperbolic flat 2-dimensional spaces with an isolated singularity.