

# Multi-group Gaussian Processes



**Didong Li, PhD**  
**Postdoctoral Research Associate**  
**Department of Computer Science**  
**Princeton University**

Gaussian processes (GPs) are pervasive in functional data analysis, machine learning, and spatial statistics for modeling complex dependencies. Modern scientific data sets are typically heterogeneous and often contain multiple known discrete subgroups of samples. For example, in genomics applications samples may be grouped according to tissue type or drug exposure. In the modeling process, it is desirable to leverage the similarity among groups while accounting for differences between them. While a substantial literature exists for GPs over Euclidean domains, GPs on domains suitable for multi-group data remain less explored. In this talk, Dr. Li will introduce multi-group Gaussian processes (MGGPs) defined on the product of a Euclidean space and a categorical set, where the categorical set represents the group label. General methods to construct valid (positive definite) covariance functions on this domain are provided, together with algorithms for inference, estimation, and prediction. The application to gene expression data illustrates the behavior and advantages of the MGGP in the joint modeling of continuous and categorical variables.

**Thursday February 17, 2022, 3:30-4:30 PM Eastern**

**McGavran-Greenberg Hall - Room 1301**

Virtual using link and info below.

Link: <https://unc.zoom.us/j/98412143955?pwd=a1p6c3hvZ28wSnk3dVVXQWl0dEpzdz09>

Meeting ID: 984 1214 3955 Passcode: 0375501630