

Background

The quality and volume of data from cosmological surveys means that the traditional two-point analyses is not able to exploit the full power of the observations. Motivated by this, the k-Nearest Neighbor (kNN) distributions of tracers has been proposed as a summary statistics that captures clustering at all orders, thereby extracting more information from the same data than 2-pt function approaches. In this project, we will build tools to measure kNN distributions from mock catalogs, and then employ them to infer cosmological parameters.

Project Goal

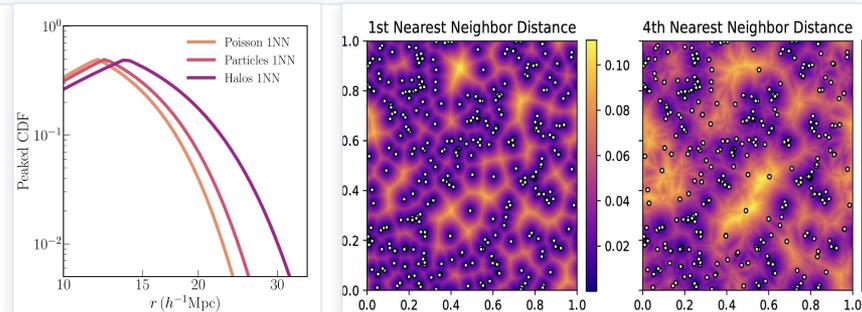
We will develop a pipeline to measure k-Nearest Neighbor (kNN) statistics from simulated and mock galaxy catalogs and use these measurements to perform cosmological parameter inference. Participants will gain hands-on experience working with large-scale structure simulations, constructing efficient algorithms to compute kNN distributions, and building likelihood frameworks that connect these statistics to cosmological models. Along the way, we will explore how kNN statistics capture non-Gaussian clustering information beyond the traditional two-point function, and investigate practical strategies for modeling them accurately.

Necessary Concepts

Basic cosmology, halo clustering, higher order correlation functions, Markov Chain Monte Carlo (MCMC) methods

Computing Need

Initial analysis can be run on a laptop. Might need cluster access towards the end of the module.



Left: Measurement of the CDF of nearest neighbor distributions for different tracers. Right: Creating continuous maps from discrete tracer locations where every pixel is painted by the distance to the nearest neighbor tracer point.

★ Background Resources

- Paper on kNN-CDFs: [2007.13342](#), [2102.01184](#)
- Implementations: [Github repo](#)
- Geometric connections: [2502.07713](#)
- Modeling kNN-CDFs: [2107.10287](#)



Arka Banerjee

Indian Institute of Science Education and Research Pune

Email: arka@iiserpune.ac.in