

# **Principal component analysis for zero-inflated compositional data**

**Sungkyu Jung**

**(Institute for Data Innovation in Science, Seoul National University)**

**Abstract:** Recent advances in DNA sequencing technology have heightened interest in microbiome data, which is often high-dimensional and presents challenges due to its compositional nature and zero-inflation. In this talk, I will introduce new PCA methods for zero-inflated compositional data, based on a framework called principal compositional subspace. These methods aim to identify both the principal compositional subspace and corresponding principal scores that best approximate the data while maintaining its compositional properties. Theoretical properties such as existence and consistency of the principal compositional subspace are investigated. Simulation studies show these methods achieve lower reconstruction errors than existing log-ratio PCA methods in linear patterns and perform comparably in curved patterns. The methods successfully uncover the low-rank structure in four microbiome compositional datasets with excessive zeros.