A MULTIVARIATE CIRCULAR DISTRIBUTION WITH APPLICATIONS TO THE PROTEIN STRUCTURE PREDICTION PROBLEM

SUNGSU KIM
Assistant Professor od Statistics
University of Louisiana at Lafayette
sungsu@louisiana.edu

Apr. 22, 2019

One of the major unsolved problems in Molecular Biology today is the protein folding problem: given an amino acid sequence, predict the overall three dimensional structure of the corresponding protein. It is often called the holy grail of Structural Bioinformatics. Three dimensional structure of a protein is defined by four dihedral angles ϕ, ψ, ω, χ, and Circular Statistics is an indispensable tool in the three dimensional protein structure prediction problem. In this talk, I present a multivariate circular distribution in order to study functional relationships among dihedral angles occurring in amino acid sequences in the same protein as well as in different proteins. The new family of k-variate circular distributions and inferential methods are applied to tri-variate circular data set of (ϕ, ψ, χ) arising from 334 γ turns consisting of Glycine-Phenylalanine-Threonine sequence. Our current research on diagnostic analysis for the proposed distribution will be discussed.