

2018 – 01

Discrete circular distributions with
applications to bacterial genomes

Tomoaki Imoto, Grace S. Shieh and Kunio Shimizu

July 25, 2018

Discrete circular distributions with applications to bacterial genomes

Tomoaki Imoto, Grace S. Shieh and Kunio Shimizu

University of Shizuoka, Academia Sinica and

The Institute of Statistical Mathematics

Abstract: We propose a novel method to construct a discrete circular distribution (DCD) from a base continuous circular distribution (CCD). The probability mass function (pmf) is defined to take the normalized values of the probability density function (pdf) at some pre-fixed equidistant points on the circle. When the pdf is represented by a Fourier series, the normalizing constant of the constructed pmf is concisely expressed by the cosine moments of the CCD. Six families of constructed DCDs are presented, and among them, some have normalizing constants and trigonometric moments in closed form. Simulation studies show that DCDs outperform the corresponding CCDs in modelling grouped (discrete) circular data, and that minimum chi-square estimation is better than maximum likelihood estimation when the number of groups on the circle is not large (when discrete patterns of data are more apparent). DCDs are applied to compare the structures of paired bacteria, an important topic in synthetic and evolutionary biology; discrete cardioid (nonnegative trigonometric sums) distribution is shown to model uni-modal (multi-modal) shared orthologous genes in *Thermotoga* and *Sulfolobus* well.