方向統計学における最近の話題

統計数理研究所 統計思考院　清水 邦夫

1. What is Directional Statistics?

Directional Statistics (DS) deals with data which include angles. Statistics which treats angles only is called Circular Statistics, since an angle corresponds to a point on the unit circle. Like this we use geometrical terminologies in DS, and thus DS refers to statistics on various manifolds such as circle, sphere, torus, cylinder and disc (disk). A point $\mathrm{P}(x_1, x_2, x_3)$ on the unit sphere with center at the origin is expressible as $x_1 = \sin\theta\cos\phi,\ x_2 = \sin\theta\sin\phi,\ x_3 = \cos\theta$ by using polar coordinates, where $0 \le \theta < \pi$ and $0 \le \phi < 2\pi$ with colatitude θ and longitude ϕ. A torus may be viewed as the Cartesian product of two circles, and a cylinder as the product of a line and a circle. A unit disc is the region bounded by a unit circle and a point in the disc is representable in the form (r, θ), where $0 \le r < 1$ and $0 \le \theta < 2\pi$.

2. Examples

Examples of circular data include angles such as wind direction, ocean current direction, direction taken by turtles after treatment and turning angles between successive movements of butterflies, and the times on a 24-hour clock such as arrival times at an intensive care unit and spawning time of a particular fish. Pairs of latitude and longitude are spherical data, and pairs of wind directions such as (wind direction observed at one monitoring site A, wind direction observed at another monitoring site B) and (wind direction at 6:00, wind direction at 12:00) are toroidal data. Pairs of (wind direction, wind speed) and (wave direction, wave height) are cylindrical data, and pairs of (concentration of pollutant, wind direction) are data on the disc.

3. Circular Distributions

Cardioid, wrapped Cauchy and von Mises distributions are well known as probability distributions on the circle. A family of symmetric distributions including these distributions has been proposed by Jones and Pewsey (2005, JASA). The family is an extension of the t-distribution on the circle given by Shimizu and Iida (2002, CIS). Extended wrapped Cauchy distributions are presented by Kato (2016, JJSS, Japanese Issue) based on his several papers.

Some other distributions are: Umbach and Jammalamadaka (2009, SPL) and Abe and Pewsey (2011, Statistical Papers) for asymmetric distributions and sine-skewing as a particular case, Abe, Pewsey and Shimizu (2013, AISM) and references therein for Batschelet–Papakonstantinou transformation, Abe, Shimizu and Pewsey (2010, JJSS) for an extension of the family by Jones and Pewsey (2005, JASA), Jones and Pewsey (2012, Biometrics) for transformation of scale, and many others.

4. Miscellaneous

Models on the hyper sphere, torus, cylinder and disc are studied by many authors including Scealy and Welsh (2011, JRSS; 2017, JASA) on the hyper sphere for compositional data, Mardia and Voss (2014, CIS) and references therein for multivariate von Mises distributions, Kim, SenGupta and Arnold (2016, JMVA) for a multivariate circular distribution which extends Wehrly and Johnson's (1980, Biometrika) model, and Uesu, Shimizu and SenGupta (2015, JMVA) on the hyper disc for a possibly asymmetric multivariate generalization of the Möbius distribution given by Jones (2004, AISM).

Applications can be seen in Shieh et al. (2011, Bioinformatics) for circular genomes, Mardia et al. (2012, JAS) and Kim, SenGupta and Arnold (2016, JMVA) for dihedral angles in protein bioinformatics, and Lagona, Picone and Maruotti (2015, Environmetrics) and Wang, Gelfand and Jona-Lasinio (2015, Statistica Sinica) for wave direction and wave height, among others. Studies of $\mathrm{PM}_{2.5}$ using nonparametric regression with wind direction as a variable include Liang et al. (2015, Proc. R. Soc. A; 2016, JGR).