Regression Modeling for Spherical Data via Non-parametric and Least Square Methods

Meisam Moghimbeygi, Mousa Golalizadeh[*];
Department of Statistics, Tarbiat Modares University

Extended Abstract

Introduction

Statistical analysis of the data on the Earth's surface was a favorite subject among many researchers. Such data can be related to animal's migration from a region to another position. Then, statistical modeling of their paths helps biological researchers to predict their movements and estimate the areas that are most likely to constitute the presence of the animals.

From a geometrical view, spherical data are points that take their values on the surface of a unit sphere. There are many methods to fit a curve, especially regression curves to spherical data. For example, Gould [1] used the corresponding angles of spherical data coordinates to introduce regression models. He considered Fisher distribution as a candidate density for the error in his analysis. A non-parametric version of his model was proposed by Thompson and Clark [2]. Usually, the data that are close to the North or South pole have different behavior. Hence, their proposed model was failing to work there, and so they tried to keep the data somewhat away from the pole via adopting their model. They advocated overcoming this problem by using the tangent plane and suggested the use of splines there [3]. On the other hand, Fisher et al. [4] proposed two families of the spherical spline for spherical data. They introduced two families of curves using differential geometry suitable for fitting the splines.

One of the methods to predict statistics is to utilize non-parametric regression models. Another strategy is to consider some forms of smooth models. Both of these procedures, along with other approaches in non-Euclidean statistics context are somewhat an initiative method in analyzing the spherical data. It worths mentioning that the benefits of using the spline path by employing the rotation parameters were of interest in directional statistics in [5], albeit for circular data. One of the interesting techniques to construct the non-parametric regression model was to minimize the Euclidean risk function, first proposed in [6]. We also follow the same procedure in this paper. In particular, the primary objective of this paper is to introduce a non-parametric regression model based on minimizing the mean square errors risk function for spherical data. To apply this idea, we used the suggested method in [6] for data on the circle. We initiate our model by considering two separate models for two common angles on the sphere. Then, we impose a correlation among these angles using an appropriate risk function. The proposed models will be evaluated using simulated and real-life data.

Material and methods

In this paper, we presented two methods for modeling spherical data. The first one considers, separately, a regression model for each angle on the sphere. To construct a feasible model, a risk

function is then suggested for modeling spherical data using Haversine distance. A non-parametric longitudinal model is derived by minimizing the proposed risk function. Hence, a parametric longitudinal model for spherical data, as the second method, is built. The estimates of the parameters in the latter model are done using the quadratic risk function.

Results and discussion

Some of the data sets are intrinsically on the surface of a sphere in many scientific disciplines. For example, the location of quakes on earth can be considered a point with a constant norm on a unit sphere. Many researchers paid attention to construct a proper model to analyze such data. Regression models are among popular forms of treating spherical data, statistically. In this paper, we also attempted to provide an efficient model to analyze spherical data. To aim this, we first adopted a regression model for each angle on the sphere, independently. Our methods included two different approaches; a non-parametric longitudinal regression modeling and minimizing a least square error framework to construct a parametric longitudinal model. In the first method, the Haversine distanced, and its minimization were considered. The validity of this approach was studied using simulated and real-life data. Then, regression modeling was proposed using the least-square error approach with an appropriate link function. Although the efficiency of this latter method in comparison with the former was in doubt, it was able to provide a suitable smooth paths prediction on the sphere. Moreover, the proposed method was more appropriate while using Haversine distance. The idea to increase the efficiency of the current model is using other distances having a secure connection with the least square method suitable for spherical data.

Conclusion

The following conclusions were drawn from this research.

- A non-parametric model inspired by previous models and a generalized version of it from circle to sphere was introduced.
- A risk function was proposed based on the Haversine distance on sphere.
- Two separated longitudinal models were suggested for the angles on the sphere and then a correlation was imposed using the least square risk function.
- Although the non-parametric method was more accurate in analyzing real data, the parametric method predicts more smooth paths.

*Corresponding author: golalizadeh@modares.ac.ir