Statistica Sinica **35** (2025), 1-23
doi:https://doi.org/10.5705/ss.202022.0347

POISSON KERNEL-BASED TESTS FOR UNIFORMITY ON THE D-DIMENSIONAL SPHERE

Yuxin Ding[1], Marianthi Markatou*[1] and Giovanni Saraceno[2]

[1] *University at Buffalo and* [2] *State University of New York*

Abstract: Tests for uniformity of distribution for data vectors on the d-dimensional hypersphere are proposed. The tests are U-statistic and V-statistic estimates of the quadratic distance between the hypothesized, under the null, uniform distribution on the sphere and the empirical cumulative distribution function. We introduce a class of *diffusion* kernels and study in detail a special member of this class, the Poisson kernel, on which our proposed tests of uniformity are based. We obtain the Karhunen-Loève decomposition of the kernel, connect it with its degrees of freedom, and hence with the power of the test via a tuning parameter, *the diffusion parameter*. We propose an algorithm that allows one to select the tuning parameter, and study the connection between the Poisson kernel-based tests and the Sobolev tests. We then study the performance of the proposed tests in terms of level and power, for a number of alternative distributions. Our simulations show that the proposed methods are powerful and outperform the Rayleigh, Giné, Ajne and Bingham test procedures in the case of multimodal alternatives. We apply the new methods to test uniformity of data on the orbits of comets obtained from the NASA website.

Key words and phrases: Diffusion kernels, directional data, exit on the sphere distribution, multimodal alternatives, Poisson kernel, spherical data, testing for uniformity.

1. Introduction

In many applications of interest, data are represented as unit vectors in a high-dimensional space or as points on a hyper-sphere. The area of directional statistics deals with data that belong to the unit hypersphere $S^{d-1} = \{\mathbf{x} \in \mathbb{R}^d : \|\mathbf{x}\|^2 = 1\}$ of $\mathbb{R}^d$. Many non-directional datasets can be usefully re-expressed in the form of directions and analyzed as such data (Golzy and Markatou, 2020). For example, in gene expression analysis, standardized gene expressions that have mean zero and variance 1 can be interpreted as directional data. This standardization is applicable when one is interested in gene expression variation under different conditions (Dortet-Bernadet and Wicker, 2008).

Assessing the presence of uniformity is one of the important initial modeling questions related to the analysis of data on the sphere. This question is formalized as a test of uniformity on S^{d-1}. There is a considerable amount of work on testing for uniformity and tests for dimensions $d \geq 2$ can be found in the literature.

*Corresponding author. E-mail: markatou@buffalo.edu