Clustering algorithm and uniformity testing for circular data

Chuo University Yoshitomo Akimoto
Toshinari Kamakura

Introduction

Directional data have often appeared in the fields of geology, ecology and so on. When we investigate the directional data, we often do the statistical tests and clustering techniques. In this report, we will introduce several uniformity tests and propose the angular clustering method including line segments data with uniformity testing.

Tests of Uniformity

Mardia (1972) described the simple score statistics: Rayleigh test when the mean direction parameter μ is unknown based on von Mises distribution and V-test (μ is given). As a improvement, Akimoto, Sakumura & Kamakura(2016)[6] proposed LM test statistic of angular data. Hamilton (1976)[3] reported that Rayleigh unimodal tests will often fail to reject the null hypothesis of a uniform distribution even though distinct modes are apparent:bimodal, unimodal. Rao proposed a test which is powerful for both unimodal and multimodal alternatives.

Rao's space test

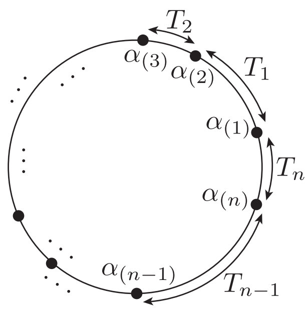

Figure 1: $\alpha_{(i)}$ and T_i

Let $\alpha_1, \ldots, \alpha_n$ denotes the sample that size of n $0 \le \alpha_i \le 2\pi$, the sorted sample $\alpha_{(1)}, \alpha_{(2)}, \ldots, \alpha_{(n)}$. The length of between each $\alpha_{(i)}$ is follows,

$$T_1 = \alpha_{(2)} - \alpha_{(1)}, T_2 = \alpha_{(3)} - \alpha_{(2)}, \cdots,$$
$$T_{n-1} = \alpha_{(n)} - \alpha_{(n-1)}, T_n = \alpha_{(1)} - \alpha_{(n)} + 2\pi.$$

Then the sum of the deviations is U_n.

$$U_n = \frac{1}{2}\sum_{i=1}^{n}\left|T_i - \frac{2\pi}{n}\right|.$$

The distribution of U_n is follows,

$$f_n(u) = (n-1)!\sum_{j=1}^{n-1}\binom{n}{j}\left(\frac{u}{2\pi}\right)^{n-j-1}\frac{\phi_j(nu)}{(n-j-1)!n^{j-1}} \quad \left(0 \le u \le 2\pi\left[1-\frac{1}{n}\right]\right),$$

where $\phi_j(x)$ is noted that,

$$\phi_j(x) = \frac{1}{2\pi\,(j-1)!}\sum_{k=0}^{\infty}(-1)^k\binom{j}{k}\left[\left(\frac{x}{2\pi}-k\right)^+\right]^{j-1}.$$

The U_n is computationally expensive so that we should use arbitrary-precision floating point numeric. We had calculated the moment of order r about the origin and used the normal approximation.

References

[1] Mardia, K. V., & Jupp, P. E. (2009). *Directional statistics*. John Wiley & Sons.
[2] Rao, J. S., & Sengupta, S. (1972). Mathematical techniques for paleocurrent analysis: treatment of directional data. *Journal of the International Association for Mathematical Geology*, 4(3), 235-248.
[3] Hamilton, P. V. (1976). A FORTRAN IV program for Rao's test of uniformity in circular distributions. *Behavior Research Methods*, 8(5), 467-467.
[4] Russell, G. S., & Levitin, D. J. (1995). An expanded table of probability values for Rao's Spacing Test. *Communications in Statistics-Simulation and Computation*, 24(4), 879-888.
[5] Gould, H. W. (1960). Stirling number representation problems. *Proceedings of the American Mathematical Society*, 11(3), 447-451.
[6] Akimoto, Y., Sakumura, T., and Kamakura, T.(2016). Testing for Uniformity of Line Segments. *IAENG International Journal of Computer Science*, vol. 43, no.2, pp192-197, 2016.