Goodness-of-fit for nested spheres

Eduardo García-Portugués • Department of Statistics, UC3M
Andrea Meilán-Vila • Department of Statistics, UC3M

Principal Nested Spheres (PNS; Jung et al. (2012)) is arguably the best analogue of Principal Component Analysis (PCA) on the d-dimensional sphere $\mathbb{S}^d = \{\boldsymbol{x} \in \mathbb{R}^{d+1} : \|\boldsymbol{x}\| = 1\}$ and $d \geq 1$. The sphere $\mathbb{S}^d$, or more generally the polysphere $(\mathbb{S}^d)^r = \mathbb{S}^d \times \overset{r}{\cdots} \times \mathbb{S}^d$, appears naturally in a variety of fields where one or r directions, in $\mathbb{R}^{d+1}$ (usually $\mathbb{R}^2$ or $\mathbb{R}^3$), need to be analyzed. Some examples include bioinformatics (e.g., protein or RNA structure), environmental sciences (e.g., sea currents), or medical imaging (e.g., hippocampi shapes). Analyzing directions is not trivial due to its non-Euclidean support, which dynamites classic statistical analyses, even the simplest ones (the average between the angles 2° and 358° is not 180°). 'Directional statistics' is the branch of statistics that has been devoted to proposing methods tailored for spherical-like supports (see Mardia & Jupp (1999), for an introduction) and is within a current larger trend in statistics focused on the analysis of non-Euclidean data (Marron & Dryden, 2021).

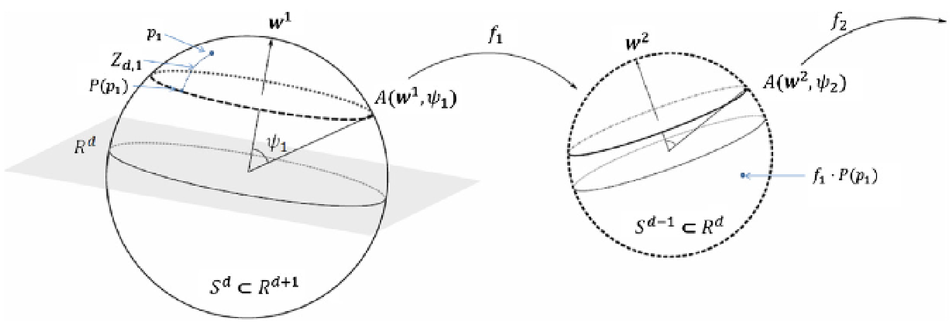
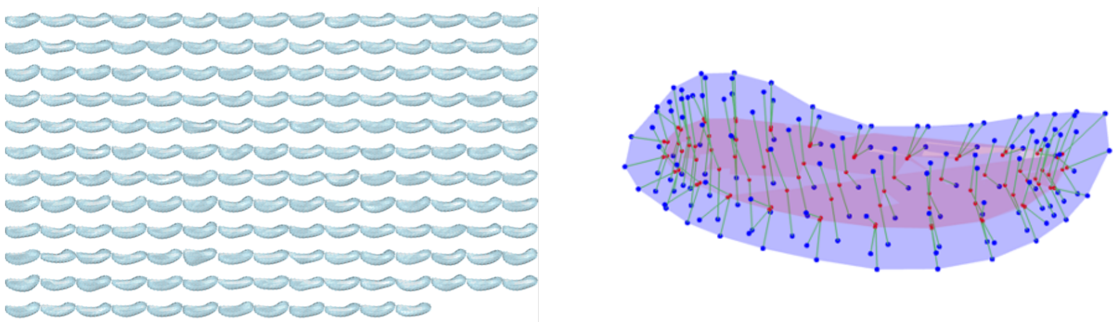


Figure 1: Sketch of principal nested spheres. Extracted from Jung et al. (2012).

PNS considers a sequence of nested subspheres (instead of planes, as PCA does), to perform dimension reduction (see Figure 1). It has been successfully applied to various types of data, including the statistical analysis of shapes based on skeletal representations known as 's-reps' (Pizer et al., 2022). s-reps belong to the space $\mathbb{R}^m \times (\mathbb{S}^2)^r$ and are convenient parametrizations of three-dimensional shapes. They are constructed from a skeleton (red points in Figure 2b) and a series of spokes (green lines) that determine the boundary of the object (blue surface). In particular, s-reps have been used to analyze the shapes of samples of hippocampi (Figure 2a). Precisely, the analysis of s-reps can be done with Composite PNS (CPNS; Pizer et al. (2012)) which roughly consists of (1) doing PNS on each of the r $\mathbb{S}^2$; (2) making PCA on the vectors that concatenate the scores from each PNS on (1) and the data on $\mathbb{R}^m$. In this construction, (1) is performed upon the belief that PNS, which contains modes of variation given by subspheres, is an appropriate model. The present work is designed to precisely test this assumption and to formally validate the adequacy of PNS and CPNS models in practice.

The objective of this thesis proposal is to formally test for the adequacy of (C)PNS. For that, we will resort to a kernel density estimator-based statistic on $(\mathbb{S}^d)^r$ that evaluates the distance between the data and a suitable parametric model that induces a PNS-like variation on $(\mathbb{S}^d)^r$. Therefore, the statistic will compare the kernel density estimator with a smooth parametric estimate according to a given distance between both functions on $(\mathbb{S}^d)^r$. A parametric bootstrap resampling algorithm will be proposed to yield usable p-values in practice. Simulations will be conducted to validate the testing proposal. Finally, the developed test will be used to test the (C)PNS-adequacy for analyzing a dataset comprised by $n = 177$ hippocampi (see Figure 2a).



(a) A dataset of 177 hippocampi.

(b) The s-rep construction.

Figure 2: Hippocampi and s-reps.

References

Jung, S., Dryden, I. L., & Marron, J. S. (2012). Analysis of principal nested spheres. *Biometrika*, *99*(3), 551–568. https://doi.org/10.1093/biomet/ass022

Mardia, K. V., & Jupp, P. E. (1999). *Directional statistics*. Wiley. https://doi.org/10.1002/9780470316979

Marron, J. S., & Dryden, I. L. (2021). *Object oriented data analysis* (Vol. 169). CRC Press. https://doi.org/10.1201/9781351189675

Pizer, S. M., Jung, S., Goswami, D., Vicory, J., Zhao, X., Chaudhuri, R., Damon, J. N., Huckemann, S., & Marron, J. S. (2012). Nested sphere statistics of skeletal models. In M. Breuß, A. Bruckstein, & P. Maragos (Eds.), *Innovations for shape analysis* (pp. 93–115). Springer. https://doi.org/10.1007/978-3-642-34141-0_5

Pizer, S. M., Marron, J. S., Damon, J. N., Vicory, J., Krishna, A., Liu, Z., & Taheri, M. (2022). Skeletons, object shape, statistics. *Frontiers in Computer Science*, *4*, 842637. https://doi.org/10.3389/fcomp.2022.842637