

Nonparametric Statistical Inference via Metric Distribution Function in Metric Spaces

潘文亮 中国科学院数学与系统科学研究院

The distribution function is essential in statistical inference and connected with samples to form a directed closed loop by the correspondence theorem in measure theory and the Glivenko-Cantelli and Donsker properties. This connection creates a paradigm for statistical inference. However, existing distribution functions are defined in Euclidean spaces and are no longer convenient to use in rapidly evolving data objects of complex nature. It is imperative to develop the concept of the distribution function in a more general space to meet emerging needs. Note that the linearity allows us to use hypercubes to define the distribution function in a Euclidean space. Still, without the linearity in a metric space, we must work with the metric to investigate the probability measure. We introduce a class of metric distribution functions through the metric only.

We overcome this challenging step by proving the correspondence theorem and the Glivenko-Cantelli theorem for metric distribution functions in metric spaces, laying the foundation for conducting rational statistical inference for metric space-valued data. Then, we develop a homogeneity test and a mutual independence test for non-Euclidean random objects and present comprehensive empirical evidence to support the performance of our proposed methods.

报告人简介：潘文亮，现任中国科学院数学与系统科学研究院副研究员及博士生导师，专注于统计学习算法、医学图像数据分析和度量空间的非参数方法等领域研究。在 *Annals of Statistics*、*Journal of the American Statistical Association* 等统计学顶级杂志上发表了 20 篇以上学术论文，获得 2022 年教育部高等学校科学研究优秀成果自然科学类二等奖（排名第二）。主持的科研项目涵盖国家自然科学基金委优秀青年基金、面上项目、青年基金等。同时，担任北京生物医学统计与数据管理研究会副理事长，以及中国现场统计研究会统计交叉科学研究分会副秘书长。

近年来，科学技术的快速发展导致我们面对着越来越复杂的数据形式，包括神经影像和网络结构等复杂结构的数据。本报告以神经影像为例，详细介绍了应对这些复杂数据的新技术。我们深入研究了度量空间上的分布函数，不仅证明了其与随机元概率测度存在 1-1 对应关系，还阐明了相关的 Cantelli 和 Donsker 性质。此外，我们引入了两样本检验和独立性检验等新方法，以更全面地满足不同数据类型的分析需求。同时，我们特别关注影像高维问题和配准误差处理方法，进一步提高了对神经影像复杂数据的精确分析水平。这些创新方法和技术的应用，不仅在神经科学领域具有显著价值，同时为应对不断增长的多样化数据类型挑战提供了新的思路。