Accepted Manuscript

Multivariate goodness-of-fit on flat and curved spaces via nearest neighbor distances

Bruno Ebner, Norbert Henze, Joseph E. Yukich

PII: S0047-259X(16)30258-5
DOI: https://doi.org/10.1016/j.jmva.2017.12.009
Reference: YJMVA 4318

To appear in: *Journal of Multivariate Analysis*

Received date : 21 December 2016

Multivariate goodness-of-fit on flat and curved spaces via nearest neighbor distances

Bruno Ebner[a,*], Norbert Henze[a], Joseph E. Yukich[b,1]

[a] *Institut für Stochastik, Karlsruher Institut für Technologie, Karlsruhe, Germany*
[b] *Lehigh University, Department of Mathematics, Bethlehem, USA*

Abstract

We present a unified approach to goodness-of-fit testing in $\mathbb{R}^d$ and on lower-dimensional manifolds embedded in $\mathbb{R}^d$ based on sums of powers of weighted volumes of kth nearest neighbor spheres. We prove asymptotic normality of a class of test statistics under the null hypothesis and under fixed alternatives. Under such alternatives, scaled versions of the test statistics converge to the α-entropy between probability distributions. A simulation study shows that the procedures are serious competitors to established goodness-of-fit tests. The tests are applied to two data sets of gamma-ray bursts in astronomy.

Keywords: Multivariate goodness-of-fit test, nearest neighbors, α-entropy, manifold, test for uniformity on a circle or a sphere, Gamma-ray burst data
2000 MSC: Primary 62H15 Secondary 60F05, 60D05

1. Introduction and summary

Nearest neighbor methods have been successfully applied in a variety of fields, such as classification [15], density and regression function estimation [6, 11], and multivariate two-sample testing [18, 33, 39]. Moreover, nearest neighbor methods have also been employed in the context of testing the goodness-of-fit of given data with a distributional model; see [7, 17, 21].

This paper is devoted to a class of universally consistent goodness-of-fit tests based on nearest neighbors. These tests can be applied not only to test for uniformity on a compact domain in $\mathbb{R}^d$, but also to test for a specified density on a m-dimensional manifold embedded in $\mathbb{R}^d$, where $m \le d$. The problem of testing uniformity on manifolds has been considered in [16, 23]. Here, prominent special cases are testing for uniformity on a circle or on a sphere. For an overview of existing methods and modern techniques; see Section 6 of each of the monographs [29, 31]. Regarding related literature to statistics on manifolds, see [5, 12], as well as the references therein.

To be specific, let $\mathcal{M}$ denote a C^1 m-dimensional manifold embedded in $\mathbb{R}^d$, where $m \le d$. $\mathcal{M}$ is endowed with the subset topology and is a closed subset of $\mathbb{R}^d$. Let $\mathrm{d}x$ be the Riemannian volume element on $\mathcal{M}$. A probability density function on $\mathcal{M}$ is a measurable non-negative real-valued function f on $\mathcal{M}$ satisfying $\int_{\mathcal{M}} f(x)\,\mathrm{d}x = 1$. The support $\mathcal{K}(f)$ of f is the smallest closed set $K \subset \mathcal{M}$ such that $\int_K f(x)\,\mathrm{d}x = 1$.

Let $\mathcal{P}(\mathcal{M})$ denote the class of bounded probability density functions f on $\mathcal{M}$, and write $\mathcal{P}_b(\mathcal{M}) \subset \mathcal{P}(\mathcal{M})$ for the subset of probability density functions f such that $\mathcal{K}(f)$ is compact and either (i) $\mathcal{K}(f)$ has no boundary or (ii) $\mathcal{K}(f)$ is a C^1 submanifold-with-boundary of $\mathcal{M}$; we refer to Section 2 of [36] for details. Notice that $\mathcal{K}(f)$ could be an m-sphere (or any ellipsoid) embedded in $\mathbb{R}^d$. Let $\mathcal{P}_c(\mathcal{M})$ denote those probability density functions $f \in \mathcal{P}_b(\mathcal{M})$ which are bounded away from zero on their support.

*Corresponding author
Email addresses: bruno.ebner@kit.edu (Bruno Ebner), norbert.henze@kit.edu (Norbert Henze), joseph.yukich@lehigh.edu (Joseph E. Yukich)
[1] Research supported in part by NSF grant DMS-1406410