Available online at www.sciencedirect.com

ScienceDirect

Stochastic Processes and their Applications 126 (2016) 3808–3827

stochastic processes and their applications

www.elsevier.com/locate/spa

Nonparametric estimation of trend in directional data

Rudolf Beran

Department of Statistics, University of California, Davis, One Shields Avenue, Davis, CA 95616-8705, USA

Received 12 July 2014; accepted 11 February 2015
Available online 28 April 2016

Abstract

Consider measured positions of the paleomagnetic north pole over time. Each measured position may be viewed as a direction, expressed as a unit vector in three dimensions and incorporating some error. In this sequence, the true directions are expected to be close to one another at nearby times. A simple trend estimator that respects the geometry of the sphere is to compute a running average over the time-ordered observed direction vectors, then normalize these average vectors to unit length. This paper treats a considerably richer class of competing directional trend estimators that respect spherical geometry. The analysis relies on a nonparametric error model for directional data in R^q that imposes no symmetry or other shape restrictions on the error distributions. Good trend estimators are selected by comparing estimated risks of competing estimators under the error model. Uniform laws of large numbers, from empirical process theory, establish when these estimated risks are trustworthy surrogates for the corresponding unknown risks.

Keywords: Random directions; Directional trend model; Projected linear estimator; Uniform law of large numbers; Minimizing estimated risk

1. Introduction

1.1. Preliminaries

Consider measurements on the position of the Earth's north magnetic pole, derived from rock samples collected at various sites. Each observed position, usually reported as latitude and longitude, may be represented as a unit vector in R^3 that specifies the direction from the center

E-mail address: rjberan@ucdavis.edu.

http://dx.doi.org/10.1016/j.spa.2016.04.018
0304-4149/

of the Earth to the point on the Earth's surface with that latitude and longitude. Associated with each such direction vector is the geological dating of the corresponding rock sample. Substantial measurement errors are to be expected in the data. The problem is to extract trend in the position of the north magnetic pole as a function of time.

Consider in R^3 the orthonormal basis (j_1, j_2, j_3) in which j_3 is the unit vector pointing to the Earth's geographical north pole and j_1 is the unit vector orthogonal to j_3 that points to longitude 0. Relative to this basis, an observed direction has polar coordinates (θ, ϕ). Here $\theta \in [0, \pi]$ is the angle, in radians, between j_3 and the observed direction. The angle $\phi \in [0, 2\pi)$ specifies, in radians, the counterclockwise rotation angle in the j_1–j_2 plane from j_1 to the longitude of the observed direction.

Relative to the same basis, the unit vector with polar coordinates (θ, ϕ) has Cartesian coordinates

$$x_1 = \sin(\theta)\cos(\phi), \qquad x_2 = \sin(\theta)\sin(\phi), \qquad x_3 = \cos(\theta). \tag{1.1}$$

Cartesian coordinates prove useful in defining trend estimators that operate on directional data. From the Cartesian coordinates of a direction, whether observed or fitted, the polar coordinates may be recovered as

$$(\theta, \phi) = \begin{cases} (\arccos(x_3), \operatorname{atan2}(x_2, x_1)) & \text{if } \operatorname{atan2}(x_2, x_1) \geq 0 \\ (\arccos(x_3), \operatorname{atan2}(x_2, x_1) + 2\pi) & \text{otherwise.} \end{cases} \tag{1.2}$$

The function atan2, which has domain $R^2 - \{0, 0\}$ and range $(-\pi, \pi]$ is defined by

$$\operatorname{atan2}(v, u) = \begin{cases} \arctan(v/u) & \text{if } u > 0 \\ \arctan(v/u) + \pi & \text{if } u < 0, v \geq 0 \\ \arctan(v/u) - \pi & \text{if } u < 0, v < 0 \\ \pi/2 & \text{if } u = 0, v > 0 \\ -\pi/2 & \text{if } u = 0, v < 0. \end{cases} \tag{1.3}$$

Mathematical programming languages generally provide this function.

Important for visualizing directions in R^3 is the Lambert azimuthal projection of a hemisphere. If a direction has polar coordinates (θ, ϕ), let

$$(\rho, \psi) = \begin{cases} (2\sin(\theta/2), \phi) & \text{if } \theta \leq \pi/2 \\ (2\sin(\pi - \theta)/2, \phi) & \text{if } \theta > \pi/2. \end{cases} \tag{1.4}$$

Then plot the direction as the projected point $(\rho\cos(\psi), \rho\sin(\psi))$ in R^2, using different plotting symbols according to whether $\theta \leq \pi/2$ (the northern hemisphere) or $\theta > \pi/2$ (the southern hemisphere). The first case, where $\theta \leq \pi/2$, gives an area preserving projection of the northern hemisphere into a disk of radius $\sqrt{2}$. The center of the disk represents the north pole and the perimeter of the disk corresponds to the equator. The second case does likewise for the southern hemisphere, the center of the disk now representing the south pole. See [15] for a brief derivation of the Lambert (or equal area) projection and a discussion of its use in directional statistics.

Jupp and Kent [9, pp. 42–45] reported 25 positions of the paleomagnetic north pole, measured in rock specimens from various sites in Antarctica. Each rock specimen was dated, so that the time sequence of the measured positions is known. The left-hand Lambert plot in Fig. 1 displays the measured magnetic pole positions, expressed in polar coordinates and plotted according to (1.4). Line segments join positions adjacent in time. Little pattern emerges. Insight is gained by: (a) using (1.1) to express each observed direction as a unit vector in Cartesian coordinates;