Statistics & Probability Letters 54 (2001) 61–65

STATISTICS & PROBABILITY LETTERS

www.elsevier.nl/locate/stapro

Uniqueness, consistency and optimality in spherical regression experiments☆

Hwashin H. Shin[a], Glen K. Takahara[a,*], Duncan J. Murdoch[b]

[a] Department of Mathematics and Statistics, Queen's University, Kingston, Canada K7L 3N6
[b] Department of Statistical and Actuarial Sciences, University of Western Ontario, London, Canada N6A 5B7

Received February 2000; received in revised form January 2001

Abstract

For designed experiments based on the spherical regression model of Chang (Ann. Statist. 14 (1986) 907) we provide results on the minimum number of covariate directions that are necessary and sufficient for uniqueness and consistency of least squares estimates and on minimizing confidence regions. © 2001 Elsevier Science B.V. All rights reserved

Keywords: Directional data; Spherical regression; Design of experiments; Uniqueness; Consistency; Optimality

1. Introduction

For paired observations (u_i, v_i), $i = 1, \dots, n$, of directions in p-dimensional Euclidean space, which can be thought of as points on the surface of the p-dimensional unit hypersphere, Chang (1986) developed a nonparametric regression model for responses v_i and fixed covariates u_i. Given an orientation $A \in \mathrm{SO}(p)$, the space of $p \times p$ orthogonal matrices with unit determinant, the v_i's are assumed to be independent with density of the form $g(v_i^t A u_i)$, for fixed u_i's.

This work was motivated by the need to understand spherical regression design issues in calibration experiments for an electromagnetic motion-tracking system, used to track the orientation as well as the position of a sensor as the sensor moves in 3-dimensional space. While a complete description of an orientation in 3-space requires three Euler angles, a rotation matrix, a unit quaternion, or other equivalent representation (a problem considered by Prentice (1986)), often the orientations of interest are described by just two angles (or even one angle), in which case Chang's model for directions in 3-space (or 2-space) is appropriate. Such orientations occur, for example, when

☆ This research was supported by NSERC.

* Corresponding author.
E-mail address: takahara@shannon.mast.queensu.ca (G.K. Takahara).

PII: S0167-7152(01)00053-0

1. The sensor is attached to the back of a human subject performing a lifting task such that the movement is essentially restricted to a plane (1 angle), or also involves some side to side motion (2 angles), or
2. The sensor is attached to the end of a robotic arm changing its direction in 3-space (2 angles).

When one is able to control the values of the covariates u_i, the following design questions arise in the context of Chang's model:

1. What is the minimum number of distinct covariate values needed to ensure estimation of A that is (a) unique, (b) consistent, and (c) optimal?
2. For a given number of distinct u_i's, what should their values be for optimal estimation of A?

Chang (1986) showed that p linearly independent u_i's are sufficient for uniqueness and consistency. In Section 2, we strengthen this by showing that $p-1$ linearly independent u_i's are sufficient and at least $p-1$ linearly independent u_i's are necessary for unique and consistent least squares estimation of A. For optimality, we consider minimization of the volume of a large sample confidence region for the $p(p-1)/2$-dimensional component vector of A. For $p-1$ linearly independent u_i values, we show in Section 3 that the volume of this confidence region is minimized if these $p-1$ directions are mutually orthogonal. In Section 4 we provide a numerical example.

2. Uniqueness and consistency

The least squares estimate $\hat{A}_n$ of A minimizes the least squares criterion $\sum_{i=1}^n \|v_i - Au_i\|^2 = \sum_{i=1}^n (v_i - Au_i)^{\mathrm{t}}(v_i - Au_i)$, which is equivalent to maximizing the *vector correlation* (Mackenzie, 1957; Stephens, 1979)

$$\sup_{A\in \mathrm{SO}(p)} \frac{1}{n}\sum_{i=1}^n v_i^{\mathrm{t}} A u_i. \tag{1}$$

Lemma 1. *With probability* 1, *for the vector correlation* (1) *to be uniquely maximized it is necessary and sufficient that at least* $p-1$ *of the* u_i*'s be linearly independent.*

Proof. Let $D = \mathrm{span}\{u_1, u_2, \ldots, u_n\}$, $q = \dim(D)$, and $b_1, \ldots, b_q$ be an orthonormal basis for D. Then we can write the vector correlation as

$$\sup_{A\in \mathrm{SO}(p)} \sum_{j=1}^q a_j^{\mathrm{t}} A b_j \tag{2}$$

for some $a_1, \ldots, a_q$. Suppose A_1 and A_2 both maximize the vector correlation, so that

$$\sum_{j=1}^q a_j^{\mathrm{t}} A_1 b_j = \sum_{j=1}^q a_j^{\mathrm{t}} A_2 b_j. \tag{3}$$

Then postmultiplying both sides of (3) by b_k gives that

$$a_k^{\mathrm{t}} A_1 = a_k^{\mathrm{t}} A_2 \quad \text{for } k = 1, \ldots, q \quad \text{or} \quad B a_k = a_k \quad \text{for } k = 1, \ldots, q, \tag{4}$$

where $B = A_1 A_2^{\mathrm{t}}$. Since $\dim(\mathrm{span}\{a_1, \ldots, a_q\}) = q$ with probability 1 (assuming an absolutely continuous density g), B a.s. has an eigenvalue of 1 with multiplicity at least q. Since $B \in \mathrm{SO}(p)$, it follows immediately from the Spectral Theorem for $\mathrm{SO}(p)$ (Lawson, 1996) that if $q \geqslant p-1$ then $B = I$, so that $A_1 = A_2$. This proves sufficiency. For necessity, we note that any A_1 and A_2 that satisfy (4) also satisfy (3). If $1 \leqslant q \leqslant p-2$, then there are infinitely many $B \neq I$ that satisfy (4) and hence infinitely many A_1 and A_2 that satisfy (3). □

Since the density of v_i is assumed to be of the form $g(v_i^t A u_i)$ it is spherically symmetric about Au_i, and it follows that $E[v_i] = c_0 A u_i$, for some constant $c_0 \subset [0,1]$ which does not depend on i. Let A_0 be the true value of A. Our next result is an extension of Lemma 2 in Chang (1986).

Lemma 2. *Provided $c_0 > 0$, $\hat{A}_n$ is strongly consistent for A_0 if and only if $(1/n)\sum_i u_i u_i^t$ converges to a positive semi-definite symmetric matrix Σ of rank at least $(p-1)$.*

Proof. We have that $\hat{A}_n$ maximizes $\mathrm{tr}(A(1/n)U_n V_n^t)$, where $U_n = [u_1, \ldots, u_n]$ and $V_n = [v_1, \ldots, v_n]$. Also, $(1/n)$ $U_n V_n^t \to c_0 \Sigma A_0^t$ with probability 1 by Lemma 1 of Chang (1986). Since Σ is symmetric, we may write $\Sigma = PDP^t$, where P is orthonormal and D is diagonal with entries given by the eigenvalues of Σ. Letting $B = P^t A_0^t A P = ((b_{ij}))_{i,j=1}^p$ and $D = \mathrm{diag}(\lambda_1, \ldots, \lambda_p)$, we have

$$\begin{aligned} \max_{A \in \mathrm{SO}(p)} \mathrm{tr}\left(A \frac{1}{n} U_n V_n^t\right) &\to \max_{A \in \mathrm{SO}(p)} \mathrm{tr}(c_0 A \Sigma A_0^t) \\ &= c_0 \max_{A \in \mathrm{SO}(p)} \mathrm{tr}(P^t A_0^t A P D) \\ &= c_0 \max_{B \in \mathrm{SO}(p)} \mathrm{tr}(BD) \\ &= c_0 \max_{b_{ii}} (b_{11}\lambda_1 + \cdots + b_{pp}\lambda_p). \end{aligned} \qquad (5)$$

Assume $\lambda_i > 0$ for $i = 1, \ldots q$, where $q \leqslant p$, and $\lambda_i = 0$ for $i = q+1, \ldots, p$. Then the maximum of (5) is obtained when $b_{11} = b_{22} = \cdots = b_{qq} = 1$ since $B \in \mathrm{SO}(p)$. Moreover, $q = p$ or $q = p-1$ implies $B = I$, so that $A = A_0$.

However, when $q < p-1$, any $B \in \mathrm{SO}(p)$ of the form

$$B = \begin{bmatrix} I_{p-2} & 0 \\ 0 & R \end{bmatrix} \quad \text{with} \quad R = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix}$$

will maximize (5) for any θ. Then A will not be uniquely determined and $\hat{A}_n$ will not be a consistent estimate of A_0. This completes the proof. □

We note that the condition in Lemma 2 that $\mathrm{rank}(\Sigma) \geqslant p-1$ is equivalent to the condition that the number of linearly independent u_i's is at least $p-1$, since $\mathrm{rank}(\Sigma) = \dim(\mathrm{span}(\{u_i\}))$.

3. Optimal designs

In this section we consider optimal designs for experiments based on Chang's spherical regression model, using the criterion of minimizing the volume of the large sample confidence region for the component vector of $A_0^t A$, defined as follows. For any $A \in \mathrm{SO}(p)$, let H be the $p \times p$ skew-symmetric matrix (i.e. satisfying $H + H^t = 0$) in the tangent space at the identity of $\mathrm{SO}(p)$ such that $A = \phi(H)$, where ϕ is the exponential map

$$\phi(H) = I + H + \frac{H^2}{2!} + \frac{H^3}{3!} + \cdots.$$

The entries of H above the diagonal give the entries of the component vector $h \in \mathscr{R}^{\dim(\mathrm{SO}(p))} = \mathscr{R}^{p(p-1)/2}$ of A. Chang (1986) showed that a large sample $1-\alpha$ confidence region for h is given by

$$\{h \mid -\mathrm{tr}(H^2\Sigma) < c/n\}, \tag{6}$$

where c is a constant determined from the data and the $1-\alpha$ quantile of the χ^2 distribution with $p(p-1)/2$ degrees of freedom, and $\Sigma = (1/n)\sum_{i=1}^n u_i u_i^{\mathrm{t}}$. For our purposes the exact value of c is not important. Chang (1986) stated that this confidence region will have minimum volume if the u_i's are chosen so that $\Sigma = (1/p)I$, which can be achieved by taking r observations at each of p mutually orthogonal design directions $u_1,\dots,u_p$, so that $n = pr$.

We are interested in minimizing the volume when we have only $p-1$ linearly independent u_i's which, from Section 2, is the minimum number of linearly independent u_i's sufficient for unique and consistent least squares estimation of A. Such a case is of interest, for example, when calibrating orientation measurements in a large 3-dimensional experimental space, where fixing just two, instead of three, orientations of a sensor at many predetermined locations of the experimental space can yield a significant savings in time and cost. Intuitively, the optimal design in p-space would be to take r observations at each of $p-1$ mutually orthogonal design directions $u_1,\dots,u_{p-1}$.

Since Σ is symmetric and nonnegative definite, we may write $\Sigma = PDP^{\mathrm{t}}$, where $D = \mathrm{diag}(\lambda_1,\dots,\lambda_p)$, P is orthogonal, and $\lambda_1,\dots,\lambda_p$ are the nonnegative eigenvalues of Σ. Also, since $H = -H^{\mathrm{t}}$, we have

$$-\mathrm{tr}(H^2\Sigma) = \mathrm{tr}(H^{\mathrm{t}}HPDP^{\mathrm{t}}) = \mathrm{tr}(P^{\mathrm{t}}H^{\mathrm{t}}HPD) = \sum_{i=1}^{p} \lambda_i (P^{\mathrm{t}}H^{\mathrm{t}}HP)_{ii}, \tag{7}$$

which is nonnegative since $P^{\mathrm{t}}H^{\mathrm{t}}HP$ is nonnegative definite and so has nonnegative diagonal elements. Thus $-\mathrm{tr}(H^2\Sigma) = c$ is the equation of an ellipsoid in $p(p-1)/2$-space. To minimize the volume of this ellipsoid for a given c, we may, without loss of generality, take $P = I$. Then from (7), the equation of the ellipsoid $-\mathrm{tr}(H^2\Sigma) = c$, in terms of the component vector h, is given by

$$\sum_{1 \leqslant i < j \leqslant p} (\lambda_i + \lambda_j) h_{s(i,j)}^2 = c \quad \text{where } s(i,j) = (i-1)(p-i/2) + j - i. \tag{8}$$

The volume of this ellipsoid is proportional to $\{\prod_{1\leqslant i<j\leqslant p}(\lambda_i + \lambda_j)\}^{-1}$, and so we wish to choose $\lambda_1,\dots,\lambda_p$ to maximize $\prod_{1\leqslant i<j\leqslant p}(\lambda_i+\lambda_j)$, subject to the constraint that the λ's are obtained as the eigenvalues of Σ, which depends on the design directions $u_1,\dots,u_n$. Note that since $\mathrm{tr}(\Sigma)=1$, we must have $\lambda_1+\cdots+\lambda_p=1$. If we have p linearly independent u_i's then setting $\lambda_i = 1/p$ for all i in (8) gives the equation of a $p(p-1)/2$-dimensional sphere which minimizes the volume in (8), thus recovering Chang's statement. If we have $p-1$ linearly independent u_i's, then we may arbitrarily set $\lambda_1 = 0$ and we wish to maximize $\lambda_2\cdots\lambda_p\prod_{2\leqslant i<j\leqslant p}(\lambda_i+\lambda_j)$. Clearly choosing $\lambda_2=\cdots=\lambda_p=1/(p-1)$ simultaneously maximizes $\lambda_2\dots\lambda_p$ and $\prod_{2\leqslant i<j\leqslant p}(\lambda_i+\lambda_j)$ and hence their product. This solution can be obtained by taking r observations at each of $p-1$ mutually orthogonal directions $u_1,\dots,u_{p-1}$, thus verifying our intuition.

4. A numerical example

In this section we compare designs for experiments for $p=3$ using observations from simulation. For given $A \in \mathrm{SO}(p)$ and covariate direction u_i, we obtain an observed direction v_i by simulating the spherical coordinate angles θ_i between v_i and Au_i, and ϕ_i between the projection of v_i onto the plane perpendicular to Au_i and a fixed direction perpendicular to Au_i, for $i=1,\dots,n$, where n is the sample size. The distribution

Table 1
The mean angular distance (degrees) between $\hat{A}_n$ and the true A_0 for 3 different designs with $n=6$, as inferred from our simulation study

Design	$\alpha \pm$ s.e.
Optimal 2-point design	6.77 ± 0.02
Optimal 3-point design	6.46 ± 0.02
Naive 3-point design	7.04 ± 0.02

of θ_i is taken to be that of $|X_i|$, where X_i is normally distributed with mean 0 and variance σ^2. We take $\sigma = 0.2$ corresponding to a mean angle of roughly 9 degrees. To ensure spherical symmetry, we take ϕ_i to be uniformly distributed on the interval $[0, 2\pi]$. Without loss of generality we take $A = I_3$.

For fixed n, with n a multiple of 6, we compared the optimal 2-point design, the optimal 3-point design, and a naive 3-point design with design directions u_1, $-u_1$, and u_2 orthogonal to u_1. From the simulations we obtained the angular distance α between the least squares estimate $\hat{A}_n$ and the true A_0, where the angular distance is defined as

$$\alpha = \cos^{-1}\left(\frac{\mathrm{tr}(A_0^{\mathrm{t}}\hat{A}_n) - 1}{2}\right),$$

with $0 \leqslant \alpha \leqslant 180$ degrees. The angular distance is a measure of closeness between A_0 and $\hat{A}_n$ and equals 0 if and only if $\hat{A}_n = A_0$. We note that minimum volume confidence regions are obtained with the optimal 3-point design, but that, when $p=3$, the minimum volume confidence region obtained from the optimal 2-point design has volume just 1.08 times the volume of the corresponding confidence region from the optimal 3-point design, so we do not expect a large difference in the angular distances obtained from these two designs.

We simulated a range of values of n from 6 to 240. As one might expect, the mean of $\alpha\sqrt{n}$ appeared to be independent of n; we were thus able to pool all of the simulations to obtain the estimates of α shown in Table 1.

We note that the mean error with the optimal 2-point design is about 5% larger than that of the optimal 3-point design, and about 4% smaller than that of the naive 3-point design. These are not large differences, but the potential savings in time and effort afforded by implementation of the 2-point design, especially when conducting spherical calibration experiments at many locations in space, may make the 2-point design desirable.

References

Chang, T., 1986. Spherical regression. Ann. Statist. 14, 907–924.
Lawson, T., 1996. Linear Algebra. Wiley, New York.
Mackenzie, J.K., 1957. The estimation of an orientation relationship. Acta Cryst. 10, 61–62.
Prentice, M.J., 1986. Orientation statistics without parametric assumptions. J. Roy. Statist. Soc. B 48, 214–222.
Stephens, M.A., 1979. Vector correlation. Biometrika 66, 41–48.