The Annals of Statistics
1989, Vol. 17, No. 1, 293–306

SPHERICAL REGRESSION WITH ERRORS IN VARIABLES

BY TED CHANG

University of Kansas and Kansas Geological Survey

Suppose $u_1, \ldots, u_n, v_1, \ldots, v_n$ are random points on the sphere such that for unknown points $\xi_1, \ldots, \xi_n$ and unknown rotation A_0, the distribution of u_i depends only on $u_i^t \xi_i$ and that of v_i on $v_i^t A_0 \xi_i$. This paper provides asymptotic tests and confidence regions for A_0 and for its axis of rotation. Results are given in arbitrary dimension.

In a previous paper [Chang (1986)], the author studied the asymptotic properties of the least squares estimator of an unknown rotation A_0 on the unit sphere S^p in Euclidean p-space. The probabilistic model was that $u_1, \ldots, u_n$ are fixed points on $S^p, v_1, \ldots, v_n$ are random points with each v_i symmetrically distributed around $A_0 u_i$.

A numerical example was discussed which involved the motion of two rigid bodies, once coincident, on the surface of the Earth. In that example the u_i are points on one body, the v_i corresponding points on the second body and A_0 is the unknown rotation which describes the motion of one body relative to the other. It would seem for that example a preferable model would allow both the u_i and the v_i to be random. The author proposes to study "random u" models in this paper.

More precisely we assume that $u_1, \ldots, u_n, v_1, \ldots, v_n$ are points on S^p (written as column vectors) satisfying:

(i) $u_1, \ldots, u_n, v_1, \ldots, v_n$ are independent.

(ii) For unknown $\xi_1, \ldots, \xi_n$ on S^p the density of u_i with respect to uniform measure on S^p is of the form $g(u_i^t \xi_i)$ for some real valued function g satisfying $E[u_i^t \xi_i] > 0$ and $1 > E[(u_i^t \xi_i)^2] \geq 1/p$.

(iii) For some unknown orthogonal matrix A_0 (that is A_0 satisfies $A_0 A_0^t = I$) the density of v_i is of the form $g(v_i^t A_0 \xi_i)$.

(iv) $(1/n)\Sigma_i \xi_i \xi_i^t \to \Sigma$ where Σ is a positive definite symmetric matrix.

The assumption that $E[(u^t \xi)^2] \geq 1/p$ always holds for the Fisher distribution $d(\kappa)e^{\kappa u^t \xi}$ as long as $\kappa > 0$. For a uniform distribution $E[(u^t \xi)^2] = 1/p$. The assumption $E[(u_i^t \xi_i)^2] \geq 1/p$ attempts to guarantee that the u_i really do cluster around ξ_i and excludes distributions which would, for example, cluster the u_i around a small circle that is too far from the pole ξ_i.

If A_0 is known to have determinant 1, for example, if A_0 represents a rotation in 3-space, then Σ can be positive semidefinite as long as its rank is at least $p - 1$. This point is further discussed in the Appendix.

Not surprisingly the random u model is significantly more difficult to analyze that the fixed u model. For example, the least squares estimate would choose

Received March 1986; revised October 1987.
AMS 1980 *subject classifications*. 62J99, 86A60.
Key words and phrases. Estimated rotations on spheres.

$\hat{A}, \hat{\xi}_1, \ldots, \hat{\xi}_n$ to minimize

$$SSE = \sum_i \left[\|u_i - \xi_i\|^2 + \|v_i - A\xi_i\|^2 \right]$$
$$= 4n - 2\sum_i \left(u_i^t \xi_i + v_i^t A \xi_i \right).$$

It follows that $\hat{\xi}_i = (u_i + \hat{A}^t v_i)/\|u_i + \hat{A}^t v_i\|$ and that $\hat{A}$ maximizes

$$(1/n)\sum_i \|u_i + A^t v_i\| = (1/n)\sum_i \sqrt{2 + 2v_i^t A u_i}\,.$$

Unfortunately, the author is unable to prove any theorems about the statistical properties of this $\hat{A}$.

Rather the author proposes that $\hat{A}$ be chosen to maximize the vector correlation first defined by Stephens (1979),

$$r(A) = \frac{1}{n}\sum_i v_i^t A u_i.$$

For G a closed subgroup of the orthogonal matrices $O(p)$ write $\hat{A}_n(G)$ for the element of G which maximizes r and write $r(G)$ for $r(\hat{A}_n(G))$. In Section 1 of this paper the author finds the large sample asymptotic distributions of $\hat{A}_n(G)$, $r(G)$ and for $G' \subseteq G$ of $\hat{A}_n(G')^t \hat{A}_n(G)$ and $r(G) - r(G')$. The latter two can be used to test $A_0 \in G'$ given $A_0 \in G$. Using $\hat{A}_n(G')^t \hat{A}_n(G)$ leads to a complicated test statistic with a simple distribution ($\chi^2(\dim G - \dim G')$). Using $r(G) - r(G')$ leads to a simple test statistic with a complicated distribution [weighted sum of independent $\chi^2(1)$].

Rivest (1989) studied the asymptotic behavior of the fixed u model with an underlying Fisher distribution with concentration parameter $\kappa \to \infty$ with n fixed. In Section 2, we extend his results to the random u model. Roughly speaking a random u model with large κ behaves like a fixed u model with $\kappa/2$. In a Fisher distribution the mean square distance of a vector from its modal direction is, for large κ, proportional to κ^{-1} and it is quite reasonable that for concentrated error distributions a random u model is indistinguishable from a fixed u model with twice as much error placed in the points v_i.

In Section 3, the Gulf of Aden data set from Chang (1986) is reanalyzed with a random u model. An Appendix contains many of the details omitted from Section 1.

1. Large sample asymptotics. For the density $g(u^t\xi)$ on S^p, define constants c_0, c_1 and c_2 by

$$E(u) = c_0\xi,$$
$$E\left[(u - c_0\xi)(u - c_0\xi)^t\right] = c_1\xi\xi^t + c_2 I.$$

We assume $c_0 > 0$ and that $1 > E[(u^t\xi)^2] \geq 1/p$.

LEMMA 1.

$$c_0^2 + c_1 + pc_2 = 1, \qquad 0 < c_2 \le \frac{1}{p}.$$

Let $X_n = (1/n)\Sigma_i u_i v_i^t$. Then $X_n \to c_0^2 \Sigma A_0^t$.

PROOF. $E(uu^t) = (c_0^2 + c_1)\xi\xi^t + c_2 I$ and the first identity follows by taking the trace.

$E[(u^t\xi)^2] = c_0^2 + c_1 + c_2$ and the inequality follows from $1/p \le E[(u^t\xi)^2] < 1$. □

Let G be a closed subgroup of $O(p)$ and assume $A_0 \in G$. The proof of Chang (1986) provides a simple and direct proof that $\hat{A}_n(G) \to A_0$ (strong convergence). $\hat{A}_n(G)$ is the quasimaximum likelihood estimator of A_0 using the mistaken log likelihood $nr(A)$ instead of the true log likelihood $\Sigma_i \log g(u_i^t\xi_i) + \Sigma_i \log g(v_i^t A\xi_i)$. Kent (1982) has made a general study of the asymptotic properties of the MLE and the likelihood ratio statistic when a mistaken likelihood is used. In fact, the presence of the nuisance parameters ξ_i implies that the errors in variables spherical regression model does not quite fit the hypotheses of his paper. Nevertheless, the proof of his Theorem 3.1 can easily be modified for the errors in variables spherical regression model and hence we will content ourselves here with the calculations needed to use his theorem. Alternatively, the calculations here are sufficient to adapt the "from basic principles" proofs of Chang (1986) to the present situation.

The hypotheses of the tests discussed in Kent are phrased in terms of the parameter value A nearest to the density g. A is chosen to maximize

$$F(A) = E\left(\sum_i v_i^t A u_i\right) = c_0^2 \operatorname{tr}\left(A \sum_i \xi_i \xi_i^t A_0^t\right).$$

If n is sufficiently large and Σ is positive definite, then $\Sigma_i \xi_i \xi_i^t$ is positive definite and it then follows that F is maximized at A_0. On the other hand if n is sufficiently large, Σ has rank $p - 1$, and A_0 is known to be in $SO(p)$ (the $p \times p$ orthogonal matrices of determinant $+1$), then $\Sigma_i \xi_i \xi_i^t$ is positive indefinite with rank at least $p - 1$ and the maximum of F over $SO(p)$ is A_0.

If H is a $p \times p$ skew symmetric matrix ($H + H^t = 0$),

$$\Phi(H) = \sum_{r=0}^{\infty} H^r/r!$$

lies in $SO(p)$. For the closed subgroup G, define $L(G)$ to be the vector subspace consisting of those skew symmetric H such that $\Phi(tH)$ is in G for all t. The vector space dimension of $L(G)$ is the dimension of G. Since $\hat{A}_n(G) \to A_0$ we can write (for large enough n) $A_0^t \hat{A}_n(G) = \Phi(H_n(G))$, where $H_n(G) \in L(G)$ is chosen to have smallest magnitude [this is possible as soon as $\hat{A}_n(G)$ is in the same connected component of G as A_0].

By replacing each v_i with $A_0^t v_i$, we can assume without loss of generality that $A_0 = I$.

For each $H \in L(G)$ the total score function $U.(H)$ is the linear transformation on $L(G)$ defined by

$$U.(H)B = \frac{d}{dt}\bigg|_{t=0} nr(\Phi(H + tB)).$$

We have $U.(0)B = n\,\mathrm{tr}(BX_n)$.

The expected score derivative matrix is the quadratic form on $L(G)$ defined by

$$\begin{aligned} Q(B_1, B_2) &= \lim_{n\to\infty} -\frac{1}{n}E(U.(0)(B_1, B_2)) \\ &= \lim_{n\to\infty} -E\left(\frac{d}{ds}\bigg|_{s=0}\frac{d}{dt}\bigg|_{t=0} r(\Phi(sB_1 + tB_2))\right) \\ &= -c_0^2\,\mathrm{tr}(B_1\Sigma B_2). \end{aligned}$$

The usual Taylor series argument yields

$$n^{-1/2}U.(0)B = Q\big(\sqrt{n}\,H_n(G), B\big) + o_p(\|B\|). \tag{1}$$

The expected squared score matrix is the quadratic form on $L(G)$ defined by

$$Q_J(B_1, B_2) = \lim_{n\to\infty} \mathrm{Cov}\big(n^{-1/2}U.(0)B_1, n^{-1/2}U.(0)B_2\big).$$

PROPOSITION 1.

$$Q_J(B_1, B_2) = -2(1 - pc_2)c_2\,\mathrm{tr}(B_2\Sigma B_1) - c_2^2\,\mathrm{tr}(B_2B_1) \quad \textit{for } B_1, B_2 \in L(G).$$

PROOF.

$$\begin{aligned} \mathrm{Cov}\big(n^{-1/2}U.(0)\big)(B_1, B_2) &= nE\left[\mathrm{tr}(B_1X_n)\mathrm{tr}(B_2X_n)\right] \\ &= E\left[\sum_{i,j}\left(v_i^tB_1u_i\right)\left(v_j^tB_2u_j\right)^t\right]\Big/ n \\ &= \sum_{i,j} E\left[\mathrm{tr}\left(B_2^tv_jv_i^tB_1u_iu_j^t\right)\right]\Big/ n. \end{aligned}$$

Now if $i \neq j$,

$$E\left[\mathrm{tr}\left(B_2^tv_jv_i^tB_1u_iu_j^t\right)\right] = c_0^4\,\mathrm{tr}\left(B_2^t\xi_j\xi_i^tB_1\xi_i\xi_j^t\right) = 0,$$

since $\xi_i^tB_1\xi_i = 0$. Also

$$\begin{aligned} E\left[\mathrm{tr}\left(B_2^tv_iv_i^tB_1u_iu_i^t\right)\right] &= \mathrm{tr}\left[B_2^t\left(\left(c_0^2 + c_1\right)\xi_i\xi_i^t + c_2I\right)B_1\left(\left(c_0^2 + c_1\right)\xi_i\xi_i^t + c_2I\right)\right] \\ &= \left(c_0^2 + c_1\right)c_2\,\mathrm{tr}\left[\left(B_1B_2^t + B_2^tB_1\right)\xi_i\xi_i^t\right] + c_2^2\,\mathrm{tr}\,B_2^tB_1 \\ &= -2\left(c_0^2 + c_1\right)c_2\,\mathrm{tr}\left(B_1B_2\xi_i\xi_i^t\right) - c_2^2\,\mathrm{tr}\,B_2B_1. \end{aligned}$$

Since $c_0^2 + c_1 = 1 - pc_2$, we get the desired form of Q_J. Using Lemma 1, we note that Q_J is nonsingular. □

In the fixed u case, the expected score derivative matrix and the expected squared score matrix are multiples of each other and this leads to the simple structure of the analysis of the fixed u model. Kent (1982) discusses asymptotics when Q and Q_J are different. Following his approach, we will first express the analysis of the random u model using matrices. Then we will rewrite his results into basis free form using orthogonal projections under Q and Q_J. This reformulation does not depend upon the spherical regression model and the same principles can be used to rewrite much of the Kent paper.

Let $G' \subseteq G$ be a closed subgroup with $A_0 \in G'$. Let $g' = \dim G'$ and $g = \dim G$. Pick any basis of $L(G')$ and extend it to a basis of $L(G)$. In terms of this basis write the matrices of Q and Q_J as

$$M = \begin{bmatrix} M_{11} & M_{12} \\ M_{21} & M_{22} \end{bmatrix}$$

and

$$J = \begin{bmatrix} J_{11} & J_{12} \\ J_{21} & J_{22} \end{bmatrix},$$

respectively, where the square $g \times g$ matrices M and J have been split after their g'th rows and columns. Let $h_n(G)$ be the g-vector representation of $H_n(G)$.

THEOREM 1. (a) *$\sqrt{n}\, h_n(G)$ is asymptotically normal with mean* 0 *and covariance matrix* $M^{-1}JM^{-1}$.

(b) $2n(r(G) - r(G'))$ *is asymptotically the weighted sum of* $g - g'$ *independent* $\chi^2(1)$ *variates* $\Sigma\mu_i\chi^2(1)$, *where* μ_i *are the eigenvalues of the matrix* $(M_{22} - M_{21}M_{11}^{-1}M_{12})(M^{-1}JM^{-1})_{22}$.

(c) *Write* $\hat{A}_n(G) = \hat{A}_n(G')\Phi(H_n(G', G))$ *and let* $h_n(G', G)$ *be the vector representation of* $H_n(G', G)$. *Let* L *be the* $g \times (g - g')$ *matrix* $L = [-M_{21}M_{11}^{-1} \quad I]^t$. *Then*

$$\tilde{r}(H, G) = nh_n(G', G)^t ML(L^tJL)^{-1}L^tMh_n(G', G) \tag{2}$$

is asymptotically $\chi^2(g - g')$.

(d) $r(G)$ *is asymptotically normal with mean* c_0^2 *and variance*

$$\left(1 - 2(p-1)c_2 + p(p-1)c_2^2 - c_0^4\right)/n.$$

Part (a) of Theorem 1 follows from (1) and Proposition 1 [or equivalently Kent's equations (3.8) and (3.9)]. Part (b) is Kent's equation (3.3). We will show in the Appendix that (c) is asymptotically equivalent to Kent's equation (4.1). Part (d) is a reasonably straightforward calculation which is also presented in the Appendix.

To use Theorem 1 in most cases, we will need consistent estimates of c_0, c_2 and Σ.

PROPOSITION 2. *Suppose $A_0 \in G$. Let*

$$\hat{c}_0 = r(G)^{1/2},$$

$$\hat{\Sigma} = \left(X_n \hat{A}_n(G) + \hat{A}_n(G)^t X_n^t\right)/r(G),$$

$$\hat{c}_2 = \frac{1}{p}\left[1 - \sqrt{\frac{p\sum\left(v_i^t \hat{A}_n(G) u_i\right)^2 - n}{pn - n}}\right].$$

Then $\hat{c}_0$, $\hat{\Sigma}$ and $\hat{c}_2$ consistently estimate c_0, Σ and c_2.

REMARK. Equivalent forms of these estimates are

$$1 - r(G) = \frac{1}{2n}\sum_i \|v_i - \hat{A}_n(G)u_i\|^2,$$

$$\hat{c}_2' = 1 - r(G) - \frac{1}{8n}\sum_i \|v_i - \hat{A}_n(G)u_i\|^4,$$

$$\hat{c}_2 = \frac{2\hat{c}_2'}{p-1}\Bigg/\left(1 + \sqrt{1 - \frac{2p\hat{c}_2'}{p-1}}\right).$$

These forms are more computationally stable for concentrated error distributions.

The proof of Proposition 2 is deferred until the Appendix.

To avoid always having to work in a basis which extends a basis of $L(G')$, we rewrite the results of Theorem 1 into an equivalent basis free form.

Let P_1 be orthogonal projection under Q_J of $L(G)$ onto $L(G')^\perp$ [the orthogonal complement under Q of $L(G')$] and let ρ: $L(G) \to L(G)$ be the linear isomorphism defined by $\rho(H)$ is the unique element of $L(G)$ so that $Q_J(\rho(H), B) = Q(H, B)$ for all $B \in L(G)$. The matrix of ρ is $J^{-1}M$.

THEOREM 2. (a) *$H_n(G)$ is asymptotically normal with mean* 0 *and density proportional to* $\exp(-(n/2)Q_J(\rho(H_n(G)), \rho(H_n(G))))$.

(b) *$P_1\rho$ has exactly $g - g'$ nonzero eigenvalues $\lambda_1, \dots, \lambda_{g-g'}$. Each λ_i is real and positive and $2n(r(G) - r(G'))$ is asymptotically a sum of independent $(1/\lambda_i)\chi^2(1)$ variates.*

(c) *$\tilde{r}(G', G) = nQ_J(P_1\rho H_n(G', G), P_1\rho H_n(G', G))$ is asymptotically $\chi^2(g - g')$.*

Since $L(G)$ is a vector space, a choice of basis of $L(G)$ gives it a Lebesgue measure. Any two Lebesgue measures defined in this way agree up to a multiplicative factor. Theorem 2(a) should be interpreted as a density with respect to one of these Lebesgue measures. The equivalence of Theorems 1(a) and 2(a) is then immediate. The equivalence of Theorems 1(b) and 2(b) or 1(c) and 2(c) is discussed in the Appendix. It is also shown there that $\mu_i = 1/\lambda_i$.

We conclude with the following lemma which is often useful for calculating P_1.

LEMMA 2. *If P_2 is orthogonal projection under Q_J of $L(G)$ onto $\rho(L(G'))$, then $P_2 = I - P_1$.*

PROOF. If $B_1 \in L(G')^{\perp}$ and $B_2 \in L(G')$,

$$Q_J(B_1, \rho(B_2)) = Q(B_1, B_2) = 0,$$

and the lemma follows. □

EXAMPLE (Rotations in three dimensions). The only connected subgroups of $O(3)$ are $SO(3)$, the rotations around a fixed axis ξ_0 [which form a group isomorphic to $SO(2)$] and the identity subgroup. If G' is any subgroup of $O(3)$ and $\tilde{G}'$ is its connected component containing the identity, then $\tilde{G}'$ is a normal subgroup and $G/\tilde{G}'$ is finite. Thus we will only describe large sample asymptotic tests for the subgroups $G' = \{\text{rotations around a fixed axis } \xi_0\}$ or $G' = \{I\}$.

Let α: $R^3 \to L(O(3))$ be the map,

$$\alpha([t_1 t_2 t_3]^t) = \begin{bmatrix} 0 & -t_3 & t_2 \\ t_3 & 0 & -t_1 \\ -t_2 & t_1 & 0 \end{bmatrix}.$$

Then for $x \in R^3$, $\psi(x) = \Phi(\alpha(x))$ is right-hand rule rotation of $|x|$ radians around the axis $x/|x|$. If Σ is a 3×3 symmetric matrix

$$x^t((\operatorname{tr}\Sigma)I - \Sigma)y = -\operatorname{tr}(\alpha(x)\Sigma\alpha(y))$$

and hence the matrices M and J become

$$M = c_0^2(I - \Sigma),$$
$$J = 2c_2((1 - 2c_2)I - (1 - 3c_2)\Sigma).$$

Hence, using Theorem 1(a), a test for H_0: $A = A_0$ can be based upon the test statistic $\tilde{r} = nh^t\tilde{\Sigma}h$, where $\tilde{\Sigma} = MJ^{-1}M$ and $\psi(h) = A_0^t\hat{A}(O(3))$. $\tilde{r}$ has an asymptotic $\chi^2(3)$ distribution.

Theorem 2(c) can be used to test if the axis of rotation is a fixed unit vector ξ_0. Let G' be the group of rotations around ξ_0. Let $L = \alpha(\xi_0)$ and $a_r = (1/n)\Sigma_i v_i^t L^r u_i$. Then $\hat{A}(G')$ is a rotation of $\hat{\theta}$ radians where $\sin\hat{\theta} = a_1/(a_1^2 + a_2^2)^{1/2}$ and $\cos\hat{\theta} = -a_2/(a_1^2 + a_2^2)^{1/2}$. We recall that the matrix of ρ is $J^{-1}M$ and using Lemma 2 we calculate that the matrix of P_1 is $I - J^{-1}M\xi_0(\xi_0^t\tilde{\Sigma}\xi_0)^{-1}\xi_0^t M$. Thus if $\psi(h_1) = \hat{A}(O(3))\psi(-\hat{\theta}\xi_0)$,

$$\begin{aligned}\tilde{r} &= nQ_J(P_1\rho h_1, P_1\rho h_1) = nQ_J(P_1\rho h_1, \rho h_1) = nQ(h_1, P_1\rho h_1)\\ &= nh_1^t M\left[I - c_0^4 J^{-1}M\xi_0\left(\xi_0^t\tilde{\Sigma}\xi_0\right)^{-1}\xi_0^t M\right]J^{-1}Mh_1\\ &= nh_1^t\left[\tilde{\Sigma} - \left(\tilde{\Sigma}\xi_0\right)\left(\tilde{\Sigma}\xi_0\right)^t/\left(\xi_0^t\tilde{\Sigma}\xi_0\right)\right]h_1.\end{aligned}$$

Under H_0, $\tilde{r}$ has an asymptotic $\chi^2(2)$ distribution. For later use, we note that $r(G') = a_0 + a_2 + (a_1^2 + a_2^2)^{1/2}$.

Both of these tests remain asymptotically valid if c_0, c_2 and Σ are estimated using Proposition 2.

2. Asymptotic results for Fisher distributions with large κ. We study in this section the asymptotic behavior of spherical regressions when the underlying error distribution is Fisher with concentration parameter $\kappa \to \infty$, n fixed. As in Section 1, it suffices to consider the case where $A_0 = I$. Let

$$x_i = u_i - (u_i^t \xi_i)\xi_i,$$
$$y_i = v_i - (v_i^t \xi_i)\xi_i.$$

Then the distributions of $\kappa^{1/2}x_i$ and $\kappa^{1/2}y_i$ approach the normal distribution $N_{p-1}(0, I)$ as $\kappa \to \infty$.

Let

$$\Sigma_n = \frac{1}{n}\sum_i \xi_i \xi_i^t,$$
$$\hat{\kappa} = (p-1)(1 - r(G))^{-1}.$$

THEOREM 3. *Let $\hat{A}(G) = A_0\Phi(H(G))$ and $g = \dim G$. As $\kappa \to \infty$:*

(a) *$H(G)$ is asymptotically normal with a density proportional to*

$$\exp\left(\tfrac{1}{4}n\kappa \operatorname{tr}(H\Sigma_n H)\right).$$

(b) *$(\kappa n(p-1)/\hat{\kappa})$ is asymptotically $\chi^2(n(p-1) - g)$.*

(c) *$\hat{\kappa}$ and $H(G)$ are asymptotically independent.*

If $G' \subseteq G$ is a subgroup of dimension g':

(d) *$n\kappa(r(G) - r(G'))$ is asymptotically $\chi^2(g - g')$.*

(e) *$(r(G) - r(G'))/(1 - r(G))((p-1)n - g)/(g - g')$ is asymptotically $F(g - g', (p-1)n - g)$.*

PROOF. Let X_n be as in Section 1. Since

$$X_n = \frac{1}{n}\sum\left[(u_i^t\xi_i)(v_i^t\xi_i)\xi_i\xi_i^t + (u_i^t\xi_i)\xi_i y_i^t + (v_i^t\xi_i)x_i\xi_i^t + x_i y_i^t\right]$$

and since $u_i^t\xi_i = 1 - \|x_i\|^2/2 + o_p(\kappa^{-1})$ the usual Taylor series yields that for all $B \in L(G)$,

$$\operatorname{tr}\left(\kappa^{1/2}BX_n\right) = -\operatorname{tr}\left(\kappa^{1/2}H(G)\Sigma_n B\right) + o_p(1).$$

Defining $\beta(B)$ to be the left-hand side of this equation, routine calculations show

$$\operatorname{Cov}(\beta(B_1), \beta(B_2)) = +\frac{2}{n}\operatorname{tr}\left(B_1^t B_2 \Sigma_n\right) + o(1)$$

and this implies part (a).

Now if $B \in L(G)$ is $O(\kappa^{-1/2})$,

$$2n\kappa(1 - r(\Phi(B))) = \sum \|\kappa^{1/2}y_i - \kappa^{1/2}x_i - \kappa^{1/2}B\xi_i\|^2 + o_p(1),$$

so usual least squares theory implies that $\kappa^{1/2}H(G)$ and $\kappa(1 - r(G))$ are asymptotically independent. Furthermore since $\kappa^{1/2}y_i - \kappa^{1/2}x_i$ is asymptotically $N_{p-1}(0, 2I)$, $n\kappa(1 - r(G))$ is asymptotically $\chi^2(n(p - 1) - g)$. This proves parts (b) and (c).

Parts (d) and (e) follows similarly. □

REMARK. Rivest (1989) gives an argument that shows the subgroup conditions in Theorem 3 can be replaced by the weaker requirement that G and G' be topologically imbedded submanifolds of $O(p)$. Using his notation we define $V(A_0, G)$ to be the tangent plane at O to the submanifold $\Phi^{-1}(A_0^{-1}G)$ of $L(O(p))$. If $L(G)$ and $L(G')$ are always replaced by $V(A_0, G)$ and $V(A_0, G')$, the theorems of Section 1 and Chang (1986) remain valid. Their usefulness, however, is severely limited by the fact that if G and G' are not subgroups, $V(A_0, G)$ and $V(A_0, G')$ depend strongly on A_0. The only theorems that appear to have useful extensions on these lines are Theorem 3 of the present paper and Theorem 2 [on the distribution of $r(G) - r(G')$ in the fixed u model] of Chang (1986).

3. A numerical example. We consider in this section the Gulf of Aden data set previously discussed in Chang (1986) and Rivest (1989). This example arises from the motion of Arabia relative to Somalia. For this data set, $n = 11$ and $\hat{A}(SO(3))$ is a rotation of 2.38° around an axis at 25.31°N latitude, 24.29°E longitude. We note that $r(SO(3)) = 1 - 5.812 \times 10^{-7}$.

The small sample size and closeness of $r(SO(3))$ to 1 indicate that large κ approximations are more reasonable. Indeed the errors thought to underlie tectonic data of this type are miniscule (relative to the circumference of the Earth) and so one would usually use large κ approximations for tectonic data. Theorem 3(d) for κ known as Theorem 3(e) for κ unknown can be used to test if the unknown rotation A is a specific A_0 or if it has a specific axis ξ_0. Numerically, these tests are identical in the random u and fixed u cases and so the reader is referred to Rivest (1989) for specific numerical calculations. We also get

$$\hat{\kappa} = 3.4 \times 10^6$$

and using Theorem 3(c), a 95% confidence interval for κ is

$$1.4 \times 10^6 < \kappa < 5.1 \times 10^6.$$

These values are double the values obtained in a fixed u analysis. Indeed, comparing Theorem 3 with the corresponding results in Rivest (1989) one observes that as $\kappa \to \infty$ a random u model with concentration parameter κ behaves like a fixed u model with concentration parameter $\kappa/2$.

The original paper by MacKenzie (1957) on estimating a rotation A from data (u_i, v_i) which in the absence of error would satisfy $v_i = Au_i$, was motivated by a problem arising in crystallography and *Science Citations Index* lists several references to it from the engineering literature. Thus although large sample approximations are not as relevant to tectonic data as large κ approximations, it is likely that they will be relevant to applications elsewhere. Thus, for instructional purposes, we will illustrate an analysis based upon large sample approximation with the Gulf of Aden data set.

For this data set

$$\hat{\Sigma} = \begin{bmatrix} 0.3568 & 0.4531 & 0.1325 \\ 0.4531 & 0.5924 & 0.1733 \\ 0.1325 & 0.1733 & 0.0508 \end{bmatrix},$$

$$\hat{c}_0 = 1 - 2.906 \times 10^{-7},$$

$$\hat{c}_2 = 2.906 \times 10^{-7},$$

$$\hat{\hat{\Sigma}} = \begin{bmatrix} 1.1065 \times 10^6 & -0.7796 \times 10^6 & -0.2279 \times 10^6 \\ -0.7796 \times 10^6 & 0.7013 \times 10^6 & -0.2982 \times 10^6 \\ -0.2279 \times 10^6 & -0.2982 \times 10^6 & 1.6331 \times 10^6 \end{bmatrix}.$$

Let A_0 be a rotation of 2.04° around 26.5°N, 21.5°E. In a test of H_0: $A = A_0$, we get

$$h = [0.004572 \quad 0.003749 \quad 0.001826]^t$$

and hence $\tilde{r} = 42.02$, which needs to be compared with a $\chi^2(3)$ distribution. Alternatively, if we desire to use Theorem 2(b), $r(A_0) = 1 - 1.6915 \times 10^{-6}$ so $2n(r(SO(3)) - r(A_0)) = 2.442 \times 10^{-5}$. The estimated eigenvalues of ρ are $\hat{\lambda}_1 = 1.72044 \times 10^6$, $\hat{\lambda}_2 = 1.72044 \times 10^6$, $\hat{\lambda}_3 = 1.72037 \times 10^6$. Thus $2n(r(SO(3)) - r(A_0))$ must be compared with

$$\frac{1}{\hat{\lambda}_1}\chi^2(1) + \frac{1}{\hat{\lambda}_2}\chi^2(1) + \frac{1}{\hat{\lambda}_3}\chi^2(1).$$

In this case $\hat{\lambda}_1$, $\hat{\lambda}_2$ and $\hat{\lambda}_3$ are all approximately equal and $\hat{\lambda}_i 2n(r(SO(3)) - r(A_0)) \simeq 42.02$.

To test that the axis of A is 26.5°N, 21.5°E, we use

$$h_1 = [0.002318 \quad 0.002860 \quad 0.0006159]^t \quad \text{and} \quad \tilde{r} = 2.903,$$

which needs to be compared with a $\chi^2(2)$ distribution. Alternatively, to use Theorem 2(b), we note that the matrix of $P_1\rho$ is

$$M^{-1}\left[\tilde{\Sigma} - \left(\tilde{\Sigma}\xi_0\right)\left(\tilde{\Sigma}\xi_0\right)^t / \left(\xi_0^t\tilde{\Sigma}\xi_0\right)\right].$$

The eigenvalues of $P_1\rho$ are 0, 1.72038×10^6 and 1.72044×10^6 and $2n(r(G) - r(H)) = 1.687 \times 10^{-6}$. Again the two nonzero eigenvalues λ_1 and λ_2 of $P_1\rho$ are approximately equal and $\lambda_i 2n(r(G) - r(H)) \simeq 2.902$.

Using Theorem 1(d), with approximately 95% confidence $1 - c_0^2 = (5.812 \pm 3.675) \times 10^{-7}$. Assuming a Fisher error distribution, this corresponds to $2.1 \times 10^6 < \kappa < 9.4 \times 10^6$. The close agreement between the tests using Theorems 2(b) and 2(c) and their agreement with the corresponding procedures in the fixed u analysis is due to this extremely concentrated error distribution.

Geophysicists generally believe that data of this type have an error of around 20 km which corresponds to $\kappa = 2 \times 10^5$. This value of κ is much below the lower limits of the 95% confidence interval for κ constructed using either large sample or large κ approximation. The plate boundaries are, to good approximation, piecewise linear (in a spherical sense). The points u_i and v_i represent the vertices of the boundaries on opposing plates. Since the boundaries are underwater the u_i and v_i are never never measured directly but are "interpreted" on a map of estimated boundary crossings. The author is now able to analyze a model of the estimated boundary crossings and hopes to report on this at a later date. Based upon this analysis the author believes that the geophysicists' estimates of their errors are reasonable for the raw data, that is the estimated boundary crossings, but that the process of interpretation has drastically reduced the errors in the interpreted u_i and v_i. There remains the nagging possibility that the process of interpretation introduces a stochastic dependence among the estimated vertices on each side.

Theorem 1(a) can be used to produce a large sample asymptotic confidence region for A of the form

$$\left\{\hat{A}\psi(h) | nh^t\tilde{\Sigma}h < \chi^2_{1-\alpha}(3)\right\}.$$

Alternatively, Theorem 3 implies that

$$\left\{\hat{A}\psi(h) || \frac{2n-3}{4}\hat{\kappa}h^t(I - \Sigma_n)h < 3F_{1-\alpha}(3, 2n-3)\right\}$$

is a large κ asymptotic confidence region for A. The author suggest that such regions be displayed using procedures identical with those discussed in Chang (1986).

Defining

$$SSE(A) = \sum_i \|v_i - Au_i\|^2 = 2n(1 - r(A)),$$

a naive approach to spherical regression would be to ignore the nuisance variables ξ_i and to assume an asymptotic $3F(3, 2n-3)$ distribution for

$$\chi = \frac{SSE(A) - SSE(\hat{A})}{SSE(\hat{A})/(2n-3)}.$$

Theorem 3(e) for the random u model and its corresponding fixed u result in Rivest (1989) imply that as $\kappa \to \infty$ the naive approach is asymptotically correct.

In the fixed u case, as $n \to \infty$, χ has an asymptotic $c_2/c_0(1 - c_0)\chi^2(3)$ distribution. Roughly speaking, no increase in sample size can eliminate the effect of the sphericity. Nevertheless, for a Fisher distribution with $\kappa > 5$, $1 < c_2/c_0(1 - c_0) < 1.0001$ and hence the naive approach provides an excellent approximation to the true asymptotic distribution.

In the random u_i case χ has a limiting $1/(1 - c_0^2)\Sigma_i(1/\lambda_i)\chi^2(1)$ distribution where λ_1, λ_2 and λ_3 are the eigenvalues of ρ. If σ_1, σ_2 and σ_3 are the eigenvalues of Σ,

$$\lambda_i = \frac{c_0^2}{2c_2} \frac{1 - \sigma_i}{(1 - 2c_2) - (1 - 3c_2)\sigma_i}.$$

We know that $\sigma_i > 0$ and $\sigma_1 + \sigma_2 + \sigma_3 = 1$. Notice that as $\sigma_i \to 0$, $\lambda_i \to c_0^2/2c_2(1 - 2c_2)$, but as $\sigma_i \to 1$, $\lambda_i \to 0$. Thus in the random u case, the discrepancy between the naive approach and the true large sample asymptotic distribution depends both on the concentration of the error distribution and the conditioning of the matrix Σ. Assuming a Fisher distribution with reasonably large κ, $c_0 \sim 1 - 1/\kappa$ and hence $(1 - c_0^2)\lambda_i \sim (1 + 1/2\kappa)(1 - \delta)$, where $0 < \delta < (\kappa(1 - \sigma_i))^{-1}$. A σ_i close to 1 indicates that the ξ_i are extremely tightly clustered around a point. The author believes for tectonic data the naive approach provides an adequate approximation to the true large sample asymptotic distribution unless the data points on each plate are extremely tightly clustered around a point.

APPENDIX

We prove here Theorems 1(c), 1(d), Proposition 2 and the equivalence of Theorems 1(b) and 2(b) and Theorems 1(c) and 2(c). Alternatively, one could prove Theorem 2 without the use of a basis of $L(G)$ using only the definitions of Q, Q_J, ρ and P_1 and the well known properties of the normal distribution.

If G' is a subgroup of G, let $L(G')^\perp$ be the orthogonal complement under Q of $L(G')$ in $L(G)$. Let P: $L(G) \to L(G)$ be orthogonal projection under Q of $L(G)$ onto $L(G')^\perp$.

PROPOSITION 3. *Suppose $A_0 \in G' \subseteq G$.*

(a) $H_n(G) - H_n(G') = P(H_n(G)) + o_p(1/\sqrt{n})$.
(b) *Let $H_n(G', G)$ be defined by*

$$\hat{A}_n(G) = \hat{A}_n(G')\Phi(H_n(G', G)).$$

Then $H_n(G', G) = P(H_n(G)) + o_p(1/\sqrt{n})$.

PROOF. If $B \in L(G')$, using (1),

$$Q(\sqrt{n}H_n(G), B) = Q(\sqrt{n}H_n(G'), B) + o_p(\|B\|),$$

from which it follows that $H_n(G')$ is the projection of $H_n(G)$ onto $L(G')$ modulo lower order terms. Part (a) follows. Part (b) follows by applying Φ to part (a). □

PROOF OF THEOREM 2(b) [given Theorem 1(b)]. Let Q_2 be the quadratic form defined on $L(G)$ by

$$\begin{aligned} Q_2(B_1, B_2) &= Q_J(P_1\rho(B_1), P_1\rho(B_2)) = Q_J(\rho(B_1), P_1\rho(B_2)) \\ &= Q(B_1, P_1\rho(B_2)). \end{aligned}$$

If we choose a basis of $L(G)$ which is orthonormal with respect to Q, the matrix of Q_2 is the matrix of $P_1\rho$. Furthermore, if $B_1 \in L(G')$, $Q_2(B_1, B_2) = 0$ for all B_2. On the other hand if $B \in L(G')^{\perp}$, $Q_2(B, B) = Q(B, P_1\rho(B)) = Q(B, \rho(B)) = Q_J(\rho(B), \rho(B)) > 0$. It follows that $P_1\rho$ has exactly $g - g'$ nonzero eigenvalues, all positive.

Since the columns of the matrix L of Theorem 1(c) form a basis of $L(G')^{\perp}$, the matrix of $P_1\rho$ is $L(L^tJL)^{-1}L^tM$. If $P_1\rho$ is restricted to $L(G')^{\perp}$ and then expressed as a matrix in terms of the basis consisting of the columns of L, the resulting matrix is $(L^tJL)^{-1}(L^tML)$. Tedious matrix algebra using Rao [(1973), Exercise 2.7], yields that the matrix of Theorem 1(b) is $(L^tJL)(L^tML)^{-1}$. Theorem 2(b) follows. □

PROPOSITION 4. *If T_e is the statistic defined by Kent* [(1982), (4.1)], $\tilde{r}(G', G) = T_e + o_p(1/\sqrt{n})$.

PROOF. In Kent's notation, $\hat{\psi} - \psi_0$ is the last $g - g'$ rows of $h_n(G)$. Let $E = M_{22} - M_{21}M_{11}^{-1}M_{12}$. Tedious matrix arithmetic yields $E(M^{-1}JM^{-1})_{22}E = L^tJL$. Thus

$$T_e = n(\hat{\psi} - \psi_0)^t E(L^tJL)^{-1}E(\hat{\psi} - \psi_0).$$

As a $g \times g$ matrix, the matrix of P is $[0 \quad L]$ where 0 is a $g \times g'$ matrix of 0's. Thus using Proposition 3,

$$\begin{aligned} L^tMh_n(G', G) &= L^tM[0 \quad L]h_n(G) + o_p\left(\frac{1}{\sqrt{n}}\right) \\ &= L^tML(\hat{\psi} - \psi_0) + o_p\left(\frac{1}{\sqrt{n}}\right) \\ &= E(\hat{\psi} - \psi_0) + o_p\left(\frac{1}{\sqrt{n}}\right) \end{aligned}$$

and the proposition follows. □

Proposition 4 implies Theorem 1(c). To reconcile the forms of $\tilde{r}(G', G)$ given in Theorems 1(c) and 2(c), we note again that $P_1\rho$ has matrix $L(L^tJL)^{-1}L^tM$.

PROOF OF THEOREM 1(d). We assume, as usual, $A_0 = I$. Then

$$\begin{aligned} r(G) &= \operatorname{tr}(\phi(H_n(G))X_n) = \operatorname{tr}(X_n) + \operatorname{tr}(H_n(G)X_n) + o_p\left(\frac{1}{\sqrt{n}}\right) \\ &= \frac{1}{n}\sum_i v_i^t u_i + c_0^2 \operatorname{tr}(H_n(G)\Sigma) + o_p\left(\frac{1}{\sqrt{n}}\right) \\ &= \frac{1}{n}\sum_i v_i^t u_i + o_p\left(\frac{1}{\sqrt{n}}\right), \end{aligned}$$

since $H_n(G)$ is skew symmetric and Σ is symmetric.

Now $E(v_i^t u_i) = c_0^2$ and

$$E\left(v_i^t u_i\right)^2 = E\left(\operatorname{tr} u_i u_i^t v_i v_i^t\right) = \operatorname{tr}\left[\left((1 - pc_2)\xi_i \xi_i^t + c_2 I\right)^2\right]$$
$$= 1 - 2(p-1)c_2 + p(p-1)c_2^2. \qquad \square$$

Proof of Proposition 2. From the proof of Theorem 1(d),

$$E\left(v_i^t A_0 u_i\right)^2 = 1 - 2(p-1)c_2 + p(p-1)c_2^2.$$

Since $\hat{A}_n(G) = A_0 + o_p(1)$,

$$\frac{1}{n}\Sigma_i\left(v_i^t \hat{A}_n(G) u_i\right)^2 \to 1 - 2(p-1)c_2 + p(p-1)c_2^2.$$

The proposition follows. $\square$

Remark. That Σ has rank p is used in two places: to ensure the consistency of $H_n(G)$ and the nonsingularity of the quadratic form Q. For the latter, it suffices that Σ have rank $p-1$. For the former, using the approach of Chang (1986), it suffices that $F(A) = \operatorname{tr}(A\Sigma)$ be uniquely maximized at $A = I$ over G. If Σ has rank $p-1$ and G is a subgroup of $SO(p)$, this will still be true. Thus Theorems 1 and 2 of Section 1 will hold if Σ has rank p or if Σ has rank $p-1$ and $G \subseteq SO(p)$. The same remarks apply to the fixed u theorems of Chang (1986) and to Theorem 3 of Section 2 if these rank conditions are true for Σ_n.

Acknowledgment. The author would like to thank the referee for pointing out the relevance of the Kent (1982) paper.

REFERENCES

Chang, T. (1986). Spherical regression. *Ann. Statist.* **14** 907–924.

Kent, J. (1982). Robust properties of likelihood ratio tests. *Biometrika* **69** 19–27.

MacKenzie, J. K. (1957). The estimation of an orientation relationship. *Acta Cryst.* **10** 61–62.

Rao, C. R. (1973). *Linear Statistical Inference and Its Applications*, 2nd ed. Wiley, New York.

Rivest, L.-P. (1989). Spherical regression for concentrated Fisher–von Mises distributions. *Ann. Statist.* **17** 307–317.

Stephens, M. A. (1979). Vector correlation. *Biometrika* **66** 41–48.

Department of Mathematics
University of Virginia
Charlottesville, Virginia 22903-3199