The Annals of Statistics
1986, Vol. 14, No. 3, 896–906

CONSERVATIVE CONFIDENCE BANDS IN CURVILINEAR REGRESSION[1]

BY DANIEL Q. NAIMAN

The Johns Hopkins University

This paper gives a method for constructing conservative Scheffé-type simultaneous confidence bands for curvilinear regression functions over finite intervals. The method is based on the use of a geometric inequality giving an upper bound for the uniform measure of the set of points within a given distance from γ, an arbitrary piecewise differentiable path with finite length in S^{k-1}, the unit sphere in R^k. The upper bound is obtained by "straightening" the path so that it lies in a great circle in S^{k-1}.

1. Introduction. This paper gives a method for constructing conservative Scheffé-type simultaneous confidence bands for curvilinear regression functions over intervals. The method is based on the use of a geometric inequality (Theorem 3.1) giving an upper bound for the uniform measure of the set of points within a given distance from γ, an arbitrary piecewise differentiable path with finite length in S^{k-1}, the unit sphere in R^k.

Consider the curvilinear regression model in which we observe

$$Y_i = \sum_{j=1}^{k} b_j f_j(x_i) + e_i, \qquad i = 1, \ldots, n, \tag{1.1}$$

where the regression coefficients b_j are unknown, the f_j are known functions, the design points x_i are known, and the random variables e_i are i.i.d. normal with mean 0 and variance σ^2, with σ^2 unknown. For example, if $f_j(x) = x^{j-1}$ for $j = 1, \ldots, k$, (1.1) is the usual polynomial regression model of degree $k - 1$. More generally, all of the results of this paper have obvious analogues when the random vector $\mathbf{e} = (e_1, \ldots, e_n)'$ has a spherically symmetric distribution about $\mathbf{0}$.

We use $\mathbf{f}(x)$ to denote the vector $(f_1(x), \ldots, f_k(x))'$ for $x \in R$. Let $I \subset R$ be a closed (not necessarily finite) interval fixed for the remainder of this paper. $\mathbf{f}$ defines a function from I to R^k which we assume to be continuous, bounded away from the origin, and piecewise differentiable with $\int_I \|\mathbf{f}'(x)\|^2\,dx$ finite. $\hat{\mathbf{b}}$ is used to denote the least-squares estimator of $\mathbf{b} = (b_1, \ldots, b_k)'$ and s^2 denotes the usual unbiased estimator of σ^2. We assume the design matrix is of full rank so that $\hat{\mathbf{b}} \sim N_k(\mathbf{b}, \sigma^2\Sigma)$ for some known positive definite matrix Σ, $\nu s^2/\sigma^2 \sim \chi^2_\nu$, where $\nu = n - k$, and $\hat{\mathbf{b}}$ and s^2 are independent. Let P be a $k \times k$ matrix such that $P'P = \Sigma$.

Received October 1984; revised September 1985.

[1]Research supported in part by National Science Foundation Grant No. DMS-8403646, and by Office of Naval Research Contract No. N00014-79-C-0801.

AMS 1980 *subject classifications*. Primary 60E15, 62J02; secondary 62J05, 62F25, 60D05.

Key words and phrases. Curvilinear regression, confidence band.

Suppose we wish to construct a simultaneous confidence band for the regression function $\mathbf{b}'\mathbf{f}(x)$ for $x \in I$. We consider Scheffé-type bands, that is, bands of the form

$$\hat{\mathbf{b}}'\mathbf{f}(x) \pm \mathrm{c}sp(x) \quad \text{for } x \in I, \tag{1.2}$$

where $p(x) = \{\mathbf{f}(x)'\Sigma\mathbf{f}(x)\}^{1/2} = \|P\mathbf{f}(x)\|$, and $c \geq 0$. The coverage probability of the band is defined to be the probability that all of the intervals (1.2) cover $\mathbf{b}'\mathbf{f}(x)$ simultaneously, as x ranges throughout I. The main result of this paper (Theorem 4.1) gives a lower bound for this probability, which allows for the construction of a conservative band.

The problem of obtaining simultaneous confidence bands in regression has received a considerable amount of attention recently. Results for multiple regression functions under restrictions on the predictor variables have been obtained by Casella and Strawderman (1980), Uusipaikka (1984) and Naiman (1984). Wynn and Bloomfield (1971) considered quadratic regression over the real line. Knafl, Sacks and Ylvisaker (1985) presented a numerical method for estimating the coverage probability of Scheffé-type bands in polynomial regression, based on the use of an inequality for the distribution of the maximum of a Gaussian process. Wynn (1984) obtained results in the polynomial regression context for a different class of bands, using results from the theory of quadrature.

2. Expression for the coverage probability. Before giving an expression for the coverage probability of the band (1.2) we introduce some notation and definitions. S^{k-1} denotes the unit sphere centered at the origin in R^k. Throughout this paper $\mathbf{U}$ denotes a random vector with a uniform distribution on S^{k-1}. μ is used to denote the uniform probability measure on S^{k-1}. $F_{i,j}$ denotes the F distribution function with i numerator degrees of freedom and j denominator degrees of freedom.

DEFINITION 2.1. A *path* in S^{k-1} is a piecewise differentiable function γ mapping I into S^{k-1} such that $\Lambda(\gamma) = \int_I \|\gamma'(x)\|\, dx$, which we refer to as the *length* of γ, is finite. The *image* of the path is defined by $\Gamma(\gamma) = \{\gamma(x): x \in I\}$.

Note that it is possible for a path to overlap itself so that the length of the path is not necessarily the same as the length of the curve that $\Gamma(\gamma)$ defines.

For any closed subset Γ of S^{k-1} and $\mathbf{u} \in S^{k-1}$ define

$$c_\Gamma(\mathbf{u}) = \sup\{\mathbf{u}'\mathbf{v}: \mathbf{v} \in \Gamma\}.$$

For $r \in [0,1]$ define

$$\begin{aligned}\Gamma_{(r)} &= \{\mathbf{u} \in S^{k-1}: c_\Gamma(\mathbf{u}) \geq r\}\\ &= \{\mathbf{u} \in S^{k-1}: \|\mathbf{v} - \mathbf{u}\|^2 \leq 2(1-r) \text{ for some } \mathbf{v} \in \Gamma\},\end{aligned}$$

so that $\Gamma_{(r)}$ is the set of points in S^{k-1} which are within $\{2(1-r)\}^{1/2}$ of Γ. Note that $\Gamma_{(r)}$ is empty for $r > 1$.

We define

$$\gamma(x) = \|P\mathbf{f}(x)\|^{-1}P\mathbf{f}(x) \quad \text{for } x \in I, \tag{2.1}$$

where p and $\mathbf{f}$ are defined in Section 1. Using the assumptions given about $\mathbf{f}$ in Section 1 it is easy to verify that γ and $-\gamma$ are paths in S^{k-1}.

Lemma 2.1, which is due to Uusipaikka (1984), expresses the probability that coverage fails for the band (1.2) as the probability that a random vector $\mathbf{U}$ distributed uniformly on the unit sphere lies within the (random) distance $\{2(1 - cT)\}^{1/2}$ of $\Gamma(\gamma) \cup -\Gamma(\gamma)$, where $T = s/\|\mathbf{B}\|$.

LEMMA 2.1. *The coverage probability of the band* (1.2) *is given by*

$$1 - \int_0^{1/c} \mu\left[\{\Gamma(\gamma) \cup -\Gamma(\gamma)\}_{(ct)}\right] f_T(t)\, dt,$$

where f_T denotes the density function of $T = s/\|\mathbf{B}\|$, so that $kT^2 \sim F_{\nu, k}$.

PROOF. Let $\mathbf{B} = P'^{-1}(\mathbf{b} - \hat{\mathbf{b}})$ and $\mathbf{U} = \|\mathbf{B}\|^{-1}\mathbf{B}$. $\mathbf{B}$ has a k-variate normal distribution with zero mean vector and covariance matrix $\sigma^2 I_k$, independent of s, and $\mathbf{U}$ has a uniform distribution on S^{k-1} independent of $(\|\mathbf{B}\|, s)$. If $T = s/\|\mathbf{B}\|$ it follows that $kT^2 \sim F_{\nu, k}$, and T is independent of $\mathbf{U}$.

The probability that simultaneous coverage for the intervals (1.2) fails is given by

$$\begin{aligned} &P\left[\sup_I \left\{p(x)^{-1}(\mathbf{f}(x)'(\mathbf{b} - \hat{\mathbf{b}}))\right\} \geq cs \quad \text{or} \ \sup_I \left\{p(x)^{-1}(-\mathbf{f}(x)'(\mathbf{b} - \hat{\mathbf{b}}))\right\} \geq cs\right] \\ &\qquad = P\left[\sup_I \{\gamma(x)'\mathbf{U}\} \geq cT \,\text{or}\, \sup_I \{-\gamma(x)'\mathbf{U}\} \geq cT\right] \\ &\qquad = P\left[\mathbf{U} \in \{\Gamma(\gamma) \cup -\Gamma(\gamma)\}_{(cT)}\right]. \end{aligned}$$

Conditioning on T and using the independence of $\mathbf{U}$ and T leads to the desired result. □

3. Upper bound for $\mu\{\Gamma(\gamma)_{(r)}\}$.

For a given path γ in S^{k-1} the main result of this section, Theorem 3.1, gives an upper bound for $\mu\{\Gamma(\gamma)_{(r)}\}$, the uniform measure of the set of points within a given distance of the image of γ. The upper bound may be interpreted as follows. If we replace γ by γ^*, a path of the same length but whose image lies on a great circle, then the bound may be thought of as $\mu\{\Gamma(\gamma^*)_{(r)}\}$, except that we calculate this by ignoring overlap and instead of counting points in $\Gamma(\gamma^*)_{(r)}$ which are closest to multiple points in $\Gamma(\gamma)$ once, we count them according to their multiplicities. Thus, we obtain a bound which depends only on the length of the path and consists of two terms. The first term is proportional to the length of the path and corresponds to the "tube" of points in $\Gamma(\gamma^*)_{(r)}$ which are closest to points in the interior of $\Gamma(\gamma^*)$. The second term is the sum of the measures of two half spherical caps of angular radius $\cos^{-1} r$ corresponding to the points in $\Gamma(\gamma^*)_{(r)}$ which are closest to one of the endpoints of γ^*.

The proof of Theorem 3.1 may be sketched as follows. It suffices to consider the case when $\Gamma(\gamma)$ is piecewise composed of great circular arcs, since $\Gamma(\gamma)$ can be approximated by curves of this type. In Lemmas 3.3 and 3.4 we prove the inequality described above for a path whose image is composed of great circular arcs by induction on the number of arcs. The piecewise great circular curve is replaced by a curve of equal length on a single great circle by "straightening out" the curve at each point where the circular arcs are joined. For a path whose image is on a great circle the geometry of the problem is simple and exact formulas are obtained in Lemmas 3.1 and 3.2.

We first state some definitions and give some lemmas used in the proof of the theorem. Let $r \in (0,1)$ be fixed throughout this section.

DEFINITION 3.1. A *great circular arc* in S^{k-1} with endpoints $\mathbf{a}$ and $\mathbf{b}$ is a set of points of the form

$$T\{\mathbf{x}: \mathbf{x} \in S^{k-1},\ x_1^2 + x_2^2 = 1,\ x_1 \geq t,\ x_2 \geq 0\},$$

for some $t \in (0,1)$, and some orthogonal transformation T, where $\mathbf{a} = T((1,0,0,\dots,0)')$ and $\mathbf{b} = T((t,\{1-t^2\}^{1/2},0,\dots,0)')$. The length of the arc is $\cos^{-1}t$.

For a given great circular arc A in S^{k-1} with endpoints $\mathbf{a}$ and $\mathbf{b}$ we define the following sets:

$$C(A) = \{\mathbf{u} \in A_{(r)}: c_A(\mathbf{u}) > \max\{\mathbf{u}'\mathbf{a}, \mathbf{u}'\mathbf{b}\}\},$$
$$D(A,\mathbf{a},\mathbf{b}) = \{\mathbf{u} \in A_{(r)}: c_A(\mathbf{u}) = \mathbf{u}'\mathbf{a}\},$$
$$E(A,\mathbf{a},\mathbf{b}) = \{\mathbf{u} \in A_{(r)}: c_A(\mathbf{u}) = \mathbf{u}'\mathbf{b}\}.$$

REMARK 3.1. If A is any closed subset of S^{k-1} it is easy to verify that $(TA)_{(r)} = T(A_{(r)})$, so $\mu\{(TA)_{(r)}\} = \mu(A_{(r)})$, for any orthogonal transformation T.

REMARK 3.2. It follows easily from the above definitions that $C(TA) = T\{C(A)\}$, $D(TA, T\mathbf{a}, T\mathbf{b}\} = T\{D(A,\mathbf{a},\mathbf{b})\}$, and $E(TA, T\mathbf{a}, T\mathbf{b}) = T\{E(A,\mathbf{a},\mathbf{b})\}$, for any great circular arc A and for any orthogonal transformation T.

LEMMA 3.1. *Let A be the great circular arc given in Definition* 3.1 *with T being the identity map. Then*

(i) $C(A) = \left\{\mathbf{u} \in S^{k-1}: u_1^2 + u_2^2 \geq r^2,\ u_1/\{u_1^2+u_2^2\}^{1/2} > t,\ u_2 > 0\right\}$,

(ii) $D(A,\mathbf{a},\mathbf{b}) = \{\mathbf{u} \in S^{k-1}: u_1 \geq r,\ u_2 \leq 0\}$,

(iii) $E(A,\mathbf{a},\mathbf{b}) = \left\{\mathbf{u} \in S^{k-1}: tu_1 + \{1-t^2\}^{1/2}u_2 \geq r,\right.$
$\left. u_1/\{u_1^2+u_2^2\}^{1/2} \leq t\right\}$.

PROOF. Fix $\mathbf{u} \in S^{k-1}$ and let $c = c_A(\mathbf{u})$. Define $h(s) = (s,\{1-s^2\}^{1/2}, 0,\dots,0)\mathbf{u}$ for $s \in [t,1]$, so that $c = \sup\{h(s): s \in [t,1]\}$.

To prove (i), let C' denote the set on the right-hand side of the equality in (i). If $\mathbf{u} \in C'$ let $\mathbf{v} = (u_1/\{u_1^2+u_2^2\}^{1/2}, u_2/\{u_1^2+u_2^2\}^{1/2}, 0,\dots,0)'$. Clearly $\mathbf{v} \in A$, hence $c \geq \mathbf{u}'\mathbf{v} = \{u_1^2+u_2^2\}^{1/2} \geq r$, so $\mathbf{u} \in A_{(r)}$. Using the fact that $t < u_1/\{u_1^2+u_2^2\}^{1/2}$ and $u_2 > 0$ it is easy to show $h(1^-) > h(1)$, hence $c > h(1) = \mathbf{u}'\mathbf{a}$, and $h(t^+) > h(t)$, hence $c > h(t) = \mathbf{u}'\mathbf{b}$. This proves $\mathbf{u} \in C(A)$.

Now suppose $\mathbf{u} \in C(A)$ so that $c \geq r$ and h is maximized at some $s \in (t,1)$, where $h'(s) = u_1 - su_2/\{1-s^2\}^{1/2} = 0$, and $h(s) > \max\{h(t), h(1)\}$. u_2 must be nonzero since $u_2 = 0$ and $h'(s) = 0$ together imply $u_1 = 0$ hence $h(s) = h(1)$. It follows that $s/\{1 - s^2\}^{1/2} = u_1/u_2$ and $h(s) = \{u_1^2 + u_2^2\}^{1/2} \geq r$. Also $h''(s) = -u_2/\{1 - s^2\}^{3/2} < 0$ so $u_2 > 0$. We have

$$0 < t/\{1 - t^2\}^{1/2} < s/\{1 - s^2\}^{1/2} = u_1/u_2,$$

so $u_1 > 0$, and if we apply the monotone increasing function $g(x) = x/\{1 + x^2\}^{1/2}$ we obtain $t < u_1/\{u_1^2 + u_2^2\}^{1/2}$ so $\mathbf{u} \in C'$ and the proof of (i) is complete.

To prove (ii), let D' denote the set on the right-hand side of the equality in (ii). If $\mathbf{u} \in D$ then $u_1 = \mathbf{u}'\mathbf{a} = c \geq r$ and $h(s)$ is maximized at $s = 1$. Since $h(1^-) \leq h(1)$, it follows that $u_2 \leq 0$, so $\mathbf{u} \in D'$.

If $\mathbf{u} \in D'$ then $h'(s) = u_1 - su_2/\{1 - s^2\}^{1/2} \geq 0$ for $s \in [t, 1]$. It follows that h is maximized at $s = 1$, so $c = \mathbf{u}'\mathbf{a}$. Furthermore, $\mathbf{u}'\mathbf{a} = u_1 \geq r$, so $\mathbf{u} \in D$.

For (iii), let T be the orthogonal transformation on S^{k-1} defined by

$$T(\mathbf{u}) = \left(tu_1 + \{1 - t^2\}^{1/2}u_2, \{1 - t^2\}^{1/2}u_1 - tu_2, 0, \dots, 0\right)'.$$

Then $T(A) = A$, $T(\mathbf{a}) = \mathbf{b}$, and $T(\mathbf{b}) = \mathbf{a}$. It follows that $T(D(A, \mathbf{a}, \mathbf{b})) = E(A, \mathbf{a}, \mathbf{b})$, and using (ii) it is easy to verify that $T(D(A, \mathbf{a}, \mathbf{b}))$ is the set given on the right-hand side of (iii). □

REMARK 3.3. From Lemma 3.1 it follows easily that $A_{(r)}$ is the disjoint union of the sets $C(A)$, $D(A, \mathbf{a}, \mathbf{b})$, and $E(A, \mathbf{a}, \mathbf{b})$.

LEMMA 3.2. *If A is a great circular arc in S^{k-1} with endpoints* $\mathbf{a}$ *and* $\mathbf{b}$ *and length L, then*

$$\mu\{C(A)\} = F_{k-2,2}\big(2(r^{-2} - 1)/(k - 2)\big) \times L/(2\pi)$$

and

$$\mu\{D(A, \mathbf{a}, \mathbf{b})\} = \mu\{E(A, \mathbf{a}, \mathbf{b})\} = F_{k-1,1}\big((r^{-2} - 1)/(k - 1)\big)/4.$$

PROOF. By Remarks 3.1 and 3.2 it suffices to consider the case when A is as given in Definition 3.1 with T being the identity map.

Write $\mathbf{U} = \|\mathbf{X}\|^{-1}\mathbf{X}$ where $\mathbf{X} \sim N_k(\mathbf{0}, I_k)$. Then $\|(U_1, U_2)'\|^{-1}(U_1, U_2)' = \|(X_1, X_2)'\|^{-1}(X_1, X_2)'$ is independent of $(X_1^2 + X_2^2, X_3, \dots, X_k)'$ so if we let $F = \{(\sum_{i=3}^k X_i^2)/k - 2\}/\{(X_1^2 + X_2^2)/2\}$, then $F \sim F_{k-2,2}$ independent of $\|(X_1, X_2)'\|^{-1}(X_1, X_2)'$.

Using Lemma 3.1(i) we obtain

$$\begin{aligned} P(\mathbf{U} \in C(A)) &= P\Big(\{X_1^2 + X_2^2\}/\|\mathbf{X}\|^2 \geq r^2, X_1/\{X_1^2 + X_2^2\}^{1/2} > t, \\ &\qquad X_2/\{X_1^2 + X_2^2\}^{1/2} > 0\Big) \\ &= P\big(1/\{1 + (k - 2)F/2\} \geq r^2\big) P\Big(X_1/\{X_1^2 + X_2^2\}^{1/2} > t, \\ &\qquad X_2/\{X_1^2 + X_2^2\}^{1/2} > 0\Big) \\ &= P\big(F \leq 2(r^{-2} - 1)/(k - 2)\big) \times L/(2\pi), \end{aligned}$$

and the proof of the first equality is complete.

To prove the second equalities, note that by symmetry

$$P(\mathbf{U} \in D(A, \mathbf{a}, \mathbf{b})) = P(\mathbf{U} \in E(A, \mathbf{a}, \mathbf{b})).$$

Let $\mathbf{X}$ be as above and note that $F' = \sum_{i=2}^{k} X_i^2/((k-1)X_1^2) \sim F_{k-1,1}$. Using Lemma 3.1 and the independence of $X_2/|X_2|$ and $(X_1, X_2^2, X_3, \dots, X_k)'$ we obtain

$$\begin{aligned} P(\mathbf{U} \in D(A, \mathbf{a}, \mathbf{b})) &= P(X_1/\|\mathbf{X}\| \geq r,\ X_2/|X_2| \leq 0) \\ &= P(X_1/\|\mathbf{X}\| \geq r) P(X_2/|X_2| \leq 0) = P(X_1/\|\mathbf{X}\| \geq r)/2 \\ &= P\big(X_1^2/\|\mathbf{X}\|^2 \geq r^2\big)/4 = P\big(1/\{1 + (k-1)F'\} \geq r^2\big)/4 \\ &= P\big(F' \leq (r^{-2} - 1)/(k-1)\big)/4, \end{aligned}$$

and the proof is complete. □

LEMMA 3.3. *Let* A_1 *and* A_2 *be great circular arcs in* S^{k-1} *with common endpoint* $(1, 0, \dots, 0)'$ *so that*

$$A_i = \left\{ \left(\{1 - s^2\}^{1/2}, s\mathbf{v}_i' \right)' : 0 \leq s \leq s_i \right\}$$

for some $\mathbf{v}_i \in S^{k-2}$, *and* $s_i \in (0,1)$ *for* $i = 1, 2$. *Thus* A_1 *has endpoints* $\mathbf{a}_1 = (\{1 - s_1^2\}^{1/2}, s_1\mathbf{v}_1')'$ *and* $\mathbf{b}_1 = (1, 0, \dots, 0)'$, *and* A_2 *has endpoints* $\mathbf{a}_2 = (1, 0, \dots, 0)'$ *and* $\mathbf{b}_2 = (\{1 - s_2^2\}^{1/2}, s_2\mathbf{v}_2')'$. *Set* $B_i = (A_i)_{(r)}$, $C_i = C(A_i)$, $D_i = D(A_i, \mathbf{a}_i, \mathbf{b}_i)$, *and* $E_i = E(A_i, \mathbf{a}_i, \mathbf{b}_i)$ *for* $i = 1, 2$.

Define

$$F_1 = \left\{ (t, \mathbf{x}')' \in S^{k-1} : \mathbf{x}'\mathbf{v}_1 > 0, \mathbf{x}'\mathbf{v}_2 > 0,\ r \leq t \leq 1 \right\},$$

$$F_2 = \left\{ (t, \mathbf{x}')' \in S^{k-1} : \mathbf{x}'\mathbf{v}_1 \leq 0, \mathbf{x}'\mathbf{v}_2 \leq 0,\ r \leq t \leq 1 \right\}.$$

Then we have the following:

(i) $(B_1 \cup B_2) - (C_1 \cup D_1 \cup C_2 \cup E_2) \subset F_2$;
(ii) $F_1 \subset (C_1 \cup D_1) \cap (C_2 \cup E_2)$;
(iii) $\mu(F_2) = \mu(F_1)$.

REMARK 3.4. Since Lemma 3.3 is the basic tool used to prove Lemma 3.4 some discussion is appropriate. Consider the set $A = A_1 \cup A_2$. $A_{(r)}$ is composed of four pieces, namely, $C_1 \cup D_1$, $C_2 \cup E_2$, S, the set on the left-hand side of (i), and T, the set on the right-hand side of (ii). Using (i)–(iii) we see that $\mu(C_1 \cup D_1) + \mu(C_2 \cup E_2)$ is an upper bound for $\mu(A_{(r)})$. We can interpret this upper bound as $\mu(A_{(r)}^*)$, where A^* is A "straightened out," that is, A^* has the same length as A but its components all lie on the same great circle.

PROOF. We use c_i to denote c_A for $A = A_i$, $i = 1, 2$. To prove (i), define

$$G_i = \left\{ (t, \mathbf{x}')' \in S^{k-1} : \mathbf{x}'\mathbf{v}_i \leq 0,\ r \leq t \leq 1 \right\}, \qquad i = 1, 2.$$

We will show below that

(a) $E_1 - (C_2 \cup E_2) \subset D_2$ and $D_2 - (C_1 \cup D_1) \subset E_1$,
(b) $E_1 \subset G_1$ and $D_2 \subset G_2$.

To see that (i) follows, we have

$$\begin{aligned}(B_1 \cup B_2) - (C_1 \cup D_1 \cup C_2 \cup E_2) &\subset [B_1 - (C_1 \cup D_1 \cup C_2 \cup E_2)] \\ &\quad \cup [B_2 - (C_1 \cup D_1 \cup C_2 \cup E_2)] \\ &= [E_1 - (C_2 \cup E_2)] \cup [D_2 - (C_1 \cup D_1)] \\ &\subset E_1 \cap D_2 \subset G_1 \cap G_2 = F_2,\end{aligned}$$

where the last line uses (a) and (b).

To prove (a), suppose $\mathbf{u} \in E_1 - (C_2 \cup E_2)$. Then $\mathbf{u} \in E_1$ so $c_1(\mathbf{u}) = \mathbf{u}'\mathbf{b}_1 \geq r$. But $\mathbf{b}_1 \in A_2$, so $c_2(\mathbf{u}) \geq \mathbf{u}'\mathbf{b}_1 \geq r$, thus $\mathbf{u} \in B_2$ and it follows that $\mathbf{u} \in D_2$. The second claim in (a) is proved in exactly the same manner.

To prove (b), note that if $\mathbf{u} = (t, \mathbf{x}')' \in E_1$ then $c_1(\mathbf{u}) \geq r$ and $c_1(\mathbf{u}) = \mathbf{u}'\mathbf{b}_1 = t$ so $t \geq r$. Let $h(s) = (t, \mathbf{x}')(\{1 - s^2\}^{1/2}, s\mathbf{v}_1')'$ for $s \in [0, s_1]$. Since the supremum of $h(s)$ is attained at $s = 0$, $0 \geq h'(0) = \mathbf{x}'\mathbf{v}_1$, thus $\mathbf{u} \in G_1$. The proof of the second claim is the same.

To prove (ii), let $\mathbf{u} = (t, \mathbf{x}')' \in F_1$ so that $r \leq t \leq 1$ and $\mathbf{x}'\mathbf{v}_i > 0$ for $i = 1, 2$. Then $c_1(\mathbf{u}) \geq \mathbf{u}'\mathbf{b}_1 = t \geq r$, so $\mathbf{u} \in B_1$. If h is the function defined above, then $h'(0) = \mathbf{x}'\mathbf{v}_1 > 0$, so $h(s)$ is not maximized at $s = 0$, and hence $c_1(\mathbf{u}) > \mathbf{u}'\mathbf{b}_1$. Thus $\mathbf{u} \in B_1 - E_1 = C_1 \cup D_1$. This proves the first claim. The proof of the second claim is the same.

For (iii), if T is the orthogonal transformation defined by $T((t, \mathbf{x}')) = (t, -\mathbf{x}')'$, then $\mu\{F_2\} = \mu\{TF_2\} = \mu\{F_1\}$. □

LEMMA 3.4. *Let A_i be a great circular arc with endpoints $\mathbf{a}_i$ and $\mathbf{b}_i$ for $i = 1, \ldots, m$ and assume $\mathbf{a}_i = \mathbf{b}_{i-1}$ for $i = 2, \ldots, m$. Set $B_i = (A_i)_{(r)}$, $C_i = C(A_i)$, $D_i = D(A_i, \mathbf{a}_i, \mathbf{b}_i)$, and $E_i = E(A_i, \mathbf{a}_i, \mathbf{b}_i)$. Then*

$$\mu\left(\bigcup_{i=1}^{m} B_i\right) \leq \sum_{i=1}^{m} \mu(C_i) + \mu(D_1) + \mu(E_m).$$

PROOF. The proof is by induction on m. For $m = 1$ the result follows from the fact that $B_1 = C_1 \cup D_1 \cup E_1$. Assume the result holds for $m = M \geq 1$ and consider the case when $m = M + 1$. We can assume without loss of generality that $\mathbf{b}_1 = \mathbf{a}_2 = (1, 0, \ldots, 0)'$, since if necessary, we can apply an orthogonal transformation and use Remarks 3.1 and 3.2. Thus, we may assume that A_1 and A_2 are given as in the statement of Lemma 3.3. Let F_1, F_2 be as in the statement of Lemma 3.3 and define

$$H_1 = C_1 \cup D_1$$

and

$$H_2 = \bigcup_{i=2}^{M+1} B_i - D_2.$$

We will show the following:

(i) $\bigcup_{i=1}^{M+1} B_i - (H_1 \cup H_2) \subset F_2$;
(ii) $F_1 \subset H_1 \cap H_2$.

To prove (i), note that for $i > 2$,

$$B_i - H_2 \subset D_2 = B_2 - (C_2 \cup E_2).$$

Thus

$$(B_i - H_2) - H_1 \subset (B_2 - (C_2 \cup E_2)) - (C_1 \cup D_1),$$

so we obtain

$$B_i - (H_1 \cup H_2) \subset (B_1 \cup B_2) - (D_1 \cup C_1 \cup C_2 \cup E_2). \tag{3.1}$$

Since $D_1 \cup C_1 \cup C_2 \cup E_2 \subset H_1 \cup H_2$ it follows that (3.1) holds for $i = 1$ and 2. Thus,

$$\bigcup_{i=1}^{M+1} B_i - (H_1 \cup H_2) \subset (B_1 \cup B_2) - (D_1 \cup C_1 \cup C_2 \cup E_2)$$

and the result follows from Lemma 3.3(i).

To prove (ii), note that

$$C_2 \cup E_2 = B_2 - D_2 \subset H_2,$$

so

$$(C_1 \cup D_1) \cap (C_2 \cup E_2) \subset H_1 \cap H_2$$

and the result follows from Lemma 3.3(ii).

Using (ii) and Lemma 3.3(iii)

$$\mu(H_1 \cup H_2) \le \mu(H_1) + \mu(H_2) - \mu(F_1) = \mu(H_1) + \mu(H_2) - \mu(F_2). \tag{3.2}$$

By the induction hypothesis

$$\begin{aligned} \mu(H_2) &= \mu\left(\bigcup_{i=2}^{M+1} B_i\right) - \mu(D_2) \\ &\le \sum_{i=2}^{M+1} \mu(C_i) + \mu(D_2) + \mu(E_{M+1}) - \mu(D_2) \qquad (3.3) \\ &= \sum_{i=2}^{M+1} \mu(C_i) + \mu(E_{M+1}). \end{aligned}$$

It follows from (i) and inequalities (3.2) and (3.3) that

$$\begin{aligned} \mu\left(\bigcup_{i=1}^{M+1} B_i\right) &\le \mu\left(\bigcup_{i=1}^{M+1} B_i - (H_1 \cup H_2)\right) + \mu(H_1 \cup H_2) \\ &\le \mu(F_2) + \mu(H_1 \cup H_2) \le \mu(H_1) + \mu(H_2) \\ &\le \mu(C_1) + \mu(D_1) + \sum_{i=2}^{M+1} \mu(C_i) + \mu(E_{M+1}), \end{aligned}$$

and the result holds for $m = M + 1$, so the induction is complete. □

Theorem 3.1. *If γ is a path in S^{k-1} then*

$$\mu\{\Gamma(\gamma)_{(r)}\} \le \min\{F_{k-2,2}[2(r^{-2} - 1)/(k-2)] \times \Lambda(\gamma)/(2\pi) + F_{k-1,1}[(r^{-2} - 1)/(k-1)]/2, 1\}.$$

Proof. If γ is a path such that $\Gamma(\gamma) = \bigcup_{i=1}^{m} A_i$ where the sets A_i are as given in the statement of Lemma 3.4, then the inequality for $\mu\{\Gamma(\gamma)_{(r)}\}$ follows from Lemmas 3.2 and 3.4, and the fact that μ is a probability measure. The general case follows from the fact that we can approximate $\Gamma(\gamma)$ by sets of this form. □

REMARK 3.5. A result due to Hotelling (1939) gives equality in Theorem 3.1 under two conditions. The first condition is that there be no "local self-overlapping" of the "tube" $\Gamma(\gamma)_{(r)}$, which amounts to the condition that the path be twice differentiable and that the radius of curvature of the path at each point be at least $(1 - r^2)^{1/2}$. The second condition is that there be no global overlapping of the tube, which occurs when points lie within parts of the tube corresponding to nonneighboring arcs of $\Gamma(\gamma)$.

A generalization of Hotelling's result due to Weyl (1939) gives an exact expression for $\mu(\Gamma_{(r)})$, for Γ a manifold contained in S^{k-1}, when r is sufficiently large and no global overlap occurs.

4. Conservative confidence bands. We now apply Theorem 3.1 to define conservative confidence bands in the context of Section 1.

THEOREM 4.1. *The following is a lower bound for the coverage probability of the band* (1.2),

$$1 - \int_0^{1/c} \min\Big\{F_{k-2,2}\big[2\big((ct)^{-2} - 1\big)/(k-2)\big] \times \Lambda(\gamma)/\pi + F_{k-1,1}\big[\big((ct)^{-2} - 1\big)/(k-1)\big], 1\Big\} f_T(t)\, dt,$$

where f_T denotes the density function of T, a random variable such that $kT^2 \sim F_{\nu,k}$. Thus, if c is such that the above expression is at least $1 - \alpha \in (0,1)$ *then the band* (1.2) *is a* $100(1-\alpha)\%$ *conservative confidence band.*

PROOF. Since $\mu\{\Gamma \cup -\Gamma\} \le 2\mu(\Gamma)$ the result follows from Lemma 2.1 and Theorem 3.1. □

Some applications call for a one-sided confidence band. For upper (or lower) Scheffé-type bands, i.e., bands of the form

$$\hat{\mathbf{b}}'\mathbf{f}(x) + c\mathrm{sp}(x) \quad \text{or} \quad \hat{\mathbf{b}}'\mathbf{f}(x) - c\mathrm{sp}(x) \quad \text{for } x \in I, \tag{4.1}$$

a proof similar to the proof of Lemma 2.1 yields the following expression for the coverage probability,

$$1 - \int_0^{1/c} \mu\big(\{\Gamma(\gamma)\}_{(ct)}\big) f_T(t)\, dt,$$

and this leads to the following result.

THEOREM 4.2. *Under the assumptions given in Section* 1 *the following expression is a lower bound for the coverage probability of the upper* (*or lower*) *simultaneous confidence band* (4.1),

$$1 - \int_0^{1/c} \min\Big\{F_{k-2,2}\big[2\big((ct)^{-2} - 1\big)/(k-2)\big] \times \Lambda(\gamma)/(2\pi) + F_{k-1,1}\big[\big((ct)^{-2} - 1\big)/(k-1)\big]/2, 1\Big\} f_T(t)\, dt.$$

Thus, if c is such that the above expression is at least $1 - \alpha \in (0,1)$ *then the band* (4.1) *is a* $100(1 - \alpha)\%$ *conservative confidence band.*

REMARK 4.1. Theorems 4.1 and 4.2 are easily seen to give strict improvements over the Scheffé method, for which the critical constant in (1.2) is given by $c_S = \{kF_{k,\nu;\alpha}\}^{1/2}$. This is because the integrands in Theorems 4.1 and 4.2 are bounded by $f_T(t)$, with strict inequality for values of t sufficiently close to $1/c$. See Section 5 for numerical comparisons in the case of quadratic regression.

5. Example. Theorem 4.1 can be used to construct conservative two-sided confidence bands for quadratic regression over I, an interval subset of R. In this section we compare bands constructed using this method to those constructed by other methods. The results are summarized in Table 1, which gives ratios of critical points, or equivalently, ratios of band widths.

For quadratic regression in (1.1) we take $k = 3$, and $f_j(x) = x^{j-1}$ for $j = 1, 2, 3$. For the case when $I = R$, Wynn and Bloomfield (1971) (Section 3.4) show that the image of the path in (2.1) is the intersection of the cone

$$(\delta/(1+\delta))x_1^2 + (1/(1+\delta))x_2^2 = x_3^2 \tag{5.1}$$

in R^3 with the unit sphere S^2, where $\delta \in [0,1]$ is a constant depending on the design. They tabulate the constant c_{WB} for which the band (1.2) has an exactly prescribed coverage probability for $\delta = 0$, 0.5, and 1.

For the purpose of comparing band widths, Table 1 gives ratios c/c_{WB}, where c is the constant obtained using Theorem 4.1. Note that the ratios are fairly close

TABLE 1
Ratios of critical points for quadratic regression

α	δ	ν	c/c_{WB}	c/c_S	c_0/c_{WB}
0.01	0.0	10	1.026	0.996	0.66
	0.0	$+\infty$	1.027	0.996	0.68
	0.5	10	1.046	0.980	0.65
	0.5	$+\infty$	1.047	0.980	0.67
	1.0	10	1.015	0.951	0.67
	1.0	$+\infty$	1.016	0.951	0.69
0.05	0.0	10	1.025	0.964	0.69
	0.0	$+\infty$	1.023	0.964	0.73
	0.5	10	1.045	0.979	0.68
	0.5	$+\infty$	1.042	0.979	0.72
	1.0	10	1.014	0.950	0.70
	1.0	$+\infty$	1.013	0.951	0.74
0.10	0.0	10	1.022	0.967	0.74
	0.0	$+\infty$	1.015	0.967	0.79
	0.5	10	1.043	0.982	0.73
	0.5	$+\infty$	1.036	0.981	0.78
	1.0	10	1.014	0.954	0.75
	1.0	$+\infty$	1.009	0.955	0.80

to unity. This is in spite of the fact that we should not expect the inequality in Theorem 4.1 to be very sharp when $\Gamma(\gamma)$ forms a closed curve, due to the fact that the $F_{k-1,1}$ term in Theorem 4.1 is unnecessary when the Hotelling (1939) result applies. The Scheffé (1953, 1959) method can be used to give simultaneous confidence intervals for all linear combinations of the three unknown parameters, and restriction to quadratic regression leads to a conservative confidence band. Table 1 gives the ratio c/c_S, where $c_S = \{3F_{\alpha;3,\nu}\}^{1/2}$, the critical point for the Scheffé method.

For the case when I is a proper subset of R use of c_{WB} leads to conservative bands and for sufficiently small intervals, one would expect to obtain narrower bands by using Theorem 4.1. In order to indicate the greatest potential improvement we give the limiting ratio c/c_{WB}, as I shrinks to a point. Clearly, this equals c_0/c_{WB}, where c_0 is the critical point obtained by using Theorem 4.1 with $\Lambda(\gamma) = 0$. Note that use of Theorem 4.1 can lead to considerable savings in band width.

In preliminary calculations for compiling Table 1, little variation was found in ratios of critical points as a function of ν, for $\nu = 10$, 20, 40, and $+\infty$. For this reason, only the values for $\nu = 10$ and $+\infty$ were included.

Acknowledgments. I am grateful to Esa Uusipaikka for showing me Lemma 2.1. I wish to thank Søren Johansen for the references to Hotelling (1939) and Weyl (1939), and for his many suggestions on writing this paper. I would also like to thank the referees for their many comments which led to a much improved version of this paper.

REFERENCES

CASELLA, G. and STRAWDERMAN, W. E. (1980). Confidence bands for linear regression with restricted predictor variables. *J. Amer. Statist. Assoc.* **75** 862–868.

HOTELLING, H. (1939). Tubes and spheres in n-spaces, and a class of statistical problems. *Amer. J. Math.* **61** 440–460.

KNAFL, G., SACKS, J. and YLVISAKER, D. (1985). Confidence bands for regression functions. *J. Amer. Statist. Assoc.* **80** 683–691.

NAIMAN, D. Q. (1984). Simultaneous confidence bounds for multiple regression functions over regions defined by constraints on the predictor variables. Unpublished manuscript.

SCHEFFÉ, H. (1953). A method for judging all contrasts in the analysis of variance. *Biometrika* **40** 87–104.

SCHEFFÉ, H. (1959). *The Analysis of Variance.* Wiley, New York.

UUSIPAIKKA, E. (1984). Exact confidence bands on certain restricted sets in the general linear model. Unpublished manuscript.

WEYL, H. (1939). On the volume of tubes. *Amer. J. Math.* **61** 461–472.

WYNN, H. P. (1984). An exact confidence band for one-dimensional polynomial regression. *Biometrika* **71** 375–380.

WYNN, H. P. and BLOOMFIELD, P. (1971). Simultaneous confidence bounds in regression analysis. *J. Roy. Statist. Soc. Ser. B* **33** 202–217.

DEPARTMENT OF MATHEMATICAL SCIENCES
THE JOHNS HOPKINS UNIVERSITY
BALTIMORE, MARYLAND 21218