The Annals of Statistics
1989, Vol. 17, No. 2, 589–605

ON POLYNOMIAL-BASED PROJECTION INDICES FOR EXPLORATORY PROJECTION PURSUIT

BY PETER HALL

Australian National University

We develop asymptotic theory for two polynomial-based methods of estimating orientation in projection pursuit density approximation. One of the techniques uses Legendre polynomials and has been proposed and implemented by Friedman [1]. The other employs Hermite functions. Issues of smoothing parameter choice and robustness are addressed. It is shown that each method can be used to construct $\sqrt{n}$-consistent estimates of the projection which maximizes distance from normality, although the former can only be employed in that manner when the underlying distribution has extremely light tails. The former can be used very generally to measure "low-frequency" departure from normality.

1. Introduction. Projection pursuit is a tool for finding the most interesting low-dimensional features of high-dimensional data sets. In exploratory projection pursuit, the focus of interest is the density of the population from which the data were drawn. The present article examines measures of interestingness based on orthogonal series density estimators. We study the influence of the smoothing parameter (i.e., number of terms in the series), and show that some interestingness measures are more robust than others against problems that occur with heavy-tailed densities.

The first step is to estimate that direction in which data are most interesting. If it is accepted that normal data are the *least* interesting, then a measure of departure from normality can be viewed as an index of interestingness. Indeed, this is the approach which is usually adopted. Sometimes, distance from normality is measured in terms of entropy [2, 6 and 8]. Among distributions with given variance, the normal distribution maximizes entropy. Therefore the projection in which entropy is minimized could be termed the "most interesting."

An alternative approach has recently been proposed and implemented by Friedman [1]. See also Jones and Sibson [8]. It is based on transforming the distribution of a projection to a distribution which would be uniform if the projection were normal; and measuring L^2 distance of the density of the transformed projection from the uniform density. In the present article we give asymptotic theory for this technique and propose an alternative approach. Section 2 discusses both Friedman's method and our own in broad terms, shedding light on advantages and disadvantages. We argue that as a measure of overall departure from normality ours is more robust against problems caused by heavy-tailed distributions. Nevertheless, Friedman's index would perform well as

Received July 1987; revised July 1988.
AMS 1980 *subject classifications*. Primary 62H99; secondary 62H05.
Key words and phrases. Hermite functions, Legendre polynomials, nonparametric density estimation, orthogonal series, projection pursuit.

a measure of "low-frequency" departure from normality. It should be stressed here that Friedman was not interested in finding heavy-tailed departures from normality—indeed, heavy-tailed departures are essentially a nuisance that frustrate the search for other kinds of structure. Friedman was most interested in clustering, low-dimensional relations and other "low-frequency" features.

Sections 3 and 4 develop theory for Friedman's index of interestingness and for our own, respectively. That theory gives concise advice on construction of those empiric indices of interestingness which yield $\sqrt{n}$-consistent orientation estimates. This amounts to prescribing the "optimal" number of terms in a certain orthogonal series density estimator. Unlike classical smoothing problems in density estimation, construction of an empiric index which produces $\sqrt{n}$-consistent orientation estimates is relatively insensitive to properties of the unknown distribution. All that is required is that the smoothing parameter (i.e., number of terms in the series) be chosen within a certain band of values. Our asymptotic theory gives a concise description of the band.

We assume throughout that the data have the distribution of a p-variate vector, Y, with zero mean and identity covariance matrix. In practical terms, this means that techniques are applied to data which have been empirically standardized for location and scale, exactly as done by Friedman [1]. Our main conclusions do not change if empiric standardization is used in place of theoretical standardization. All our results have generalizations to the case where projections are in q dimensions, for any $q < p$; we treat $q = 1$ for simplicity.

We write g for the p-variate density of Y; ϕ and Φ for density and distribution functions, respectively, of the univariate standard normal distribution; Ω for the set of all unit p-vectors; θ for a generic element of Ω; $x \cdot y$ for the scalar product of p-vectors x and y; g_θ for the univariate density of $V_\theta \equiv \theta \cdot Y$; $D_\theta^r g$ for the rth derivative of g in direction θ; and $\|x\| = (x \cdot x)^{1/2}$.

2. Indices of "interestingness." We begin by describing Friedman's [1] index $I(\theta)$. With $V_\theta \equiv \theta \cdot Y$, put

$$U_\theta \equiv 2\Phi(V_\theta) - 1,$$

and let f_θ denote the density of U_θ. Then V_θ is normal $N(0,1)$ if and only if U_θ is uniform on $(-1,1)$. Hence, the L_2 distance of f_θ from the uniform density on $(-1,1)$ may be used to index the departure of V_θ from normality. Formally, this index is

$$I(\theta) \equiv \int_{-1}^{1} \{ f_\theta(u) - \tfrac{1}{2} \}^2 \, du = \int_{-1}^{1} \{ f_\theta(u) \}^2 \, du - \tfrac{1}{2}.$$

The "most interesting" direction θ is that which maximizes $I(\theta)$.

Let $p_0, p_1, \ldots$ be a complete orthonormal basis for the space of square-integrable functions on $(-1,1)$, chosen so that p_0 is constant. (Orthonormality dictates that $p_0 \equiv 2^{-1/2}$.) The sequence selected by Friedman, and by ourselves in Section 3, is the normalized Legendre polynomial sequence, but there are many other possibilities. Write

$$a_i(\theta) \equiv \int_{-1}^{1} p_i(u) f_\theta(u)\, du, \qquad i \geq 0,$$

for generalized Fourier coefficients of f_θ. Then by Parseval's identity,

$$I(\theta) \equiv \sum_{i=1}^{\infty} a_i(\theta)^2.$$

We may easily estimate $a_i(\theta)$. An unbiased estimator is

$$\hat{a}_i(\theta) \equiv n^{-1} \sum_{j=1}^{n} p_i\{2\Phi(\theta \cdot Y_j) - 1\},$$

where $Y_1, \dots, Y_n$ is a sample from the distribution of Y. Then, for some (suitably chosen) $m \geq 1$, the empiric index

$$\hat{I}_m(\theta) \equiv \sum_{i=1}^{m} \hat{a}_i(\theta)^2$$

should be close to the true index $I(\theta)$. Interest centres on selection of m.

A difficulty with the population-based index in this approach is that it is unsuitable for all but extremely thin-tailed Y's. To appreciate why, let g_θ denote the density of $\theta \cdot Y$, and define $v = v(u)$ by $u = 2\Phi(v) - 1$, for $-1 < u < 1$. Then $f_\theta(u) = g_\theta(v)(dv/du)$, and $du/dv = 2\phi(v)$, so that

$$I(\theta) + \tfrac{1}{2} = \int_{-1}^{1} \{g_\theta(v)\}^2 (dv/du)^2\, du$$

$$= \tfrac{1}{2} \int_{-\infty}^{\infty} \{g_\theta(v)\}^2 \{\phi(v)\}^{-1}\, dv.$$

Therefore the tails of the density of $\theta \cdot Y$ must decrease at least as fast as $e^{-v^2/4}$ if $I(\theta)$ is not to be infinite. The index $I(\theta)$ can be infinite even if the tails of $\theta \cdot Y$ are exponentially small, like those of a gamma distribution. It will certainly be infinite if some algebraic moment of $\theta \cdot Y$ is infinite.

All of this means that for heavy-tailed distributions, $I(\theta)$ is not very useful as a measure of departure from normality. When $I(\theta)$ is infinite, there is not much point in thinking of $\hat{I}_m(\theta)$ as an approximation to $I(\theta)$. There is some virtue in studying $\hat{I}_m(\theta)$ for certain fixed, low values of m, as an empiric index of "low-frequency" departure from normality. (Low-frequency p_i's are those with low index i.) But this approach depends very much on choice of orthonormal basis. It is conceptually less satisfying than viewing $\hat{I}_m(\theta)$ as a measure of overall departure from normality.

Of course, our objections vanish if the distribution of Y is compactly supported. Section 3 will investigate this case in detail, giving concise information about our choice of m. It turns out that if the orientation which maximizes $\hat{I}_m(\theta)$ is to be $\sqrt{n}$-consistent for the orientation which maximizes $I(\theta)$, then m must increase no more rapidly than the cube root of sample size. Lower bounds to the rate at which m should increase depend on smoothness of the unknown density of Y, and will be discussed in Section 3.

Difficulties which we have with $I(\theta)$ as a measure of departure from normality are greatly alleviated if $I(\theta)$ is replaced by the L^2 distance $J^\dagger(\theta)$ between the density of $\theta \cdot Y$ and the standard normal density. This index is

$$J^\dagger(\theta) \equiv \int_{-\infty}^{\infty} \{g_\theta(u) - \phi(u)\}^2 \, du,$$

and may be expressed in terms of orthogonal functions as follows. Let $H_0, H_1, \ldots$ be Hermite polynomials, orthogonal on $(-\infty, \infty)$ with respect to the weight function ϕ^2, and standardized by the relation $\int H_i^2 \phi^2 = i! \pi^{-1/2} 2^{i-1}$ and by the requirement that the term of highest degree in H_i have positive coefficient. The Hermite functions,

$$h_i(u) \equiv (i!)^{-1/2} \pi^{1/4} 2^{-(i-1)/2} H_i(u) \phi(u), \qquad -\infty < u < \infty, \tag{2.1}$$

are orthonormal: $\int h_i h_j = \delta_{ij}$, the Kronecker delta. Fourier coefficients in a Hermite function expansion of g_θ are

$$a_i(\theta) \equiv E\{h_i(\theta \cdot Y)\}, \qquad i \geq 0.$$

If g_θ is square-integrable, then

$$\begin{aligned} J^\dagger(\theta) &= \int_{-\infty}^{\infty} \left\{ \sum_{i=0}^{\infty} a_i(\theta) h_i(u) - \pi^{-1/4} 2^{-1/2} h_0(u) \right\}^2 du \\ &= \sum_{i=0}^{\infty} a_i(\theta)^2 - (2^{1/2}/\pi^{1/4}) a_0(\theta) + (2\pi^{1/2})^{-1} \end{aligned}$$

Maximizing $J^\dagger$ is equivalent to maximizing

$$J(\theta) \equiv J^\dagger(\theta) - (2\pi^{1/2})^{-1} = \sum_{i=0}^{\infty} a_i(\theta)^2 - (2^{1/2}/\pi^{1/4}) a_0(\theta). \tag{2.2}$$

We might redefine the "most interesting" direction θ_1 to be that which maximizes J.

Next we construct an empiric version of J. Given a random sample $Y_1, \ldots, Y_n$ from the distribution of Y, put

$$\begin{aligned} \hat{a}_i(\theta) &\equiv n^{-1} \sum_{j=1}^{n} h_i(\theta \cdot Y_j), \\ \hat{J}_m(\theta) &\equiv \sum_{i=0}^{m} \hat{a}_i(\theta)^2 - (2^{1/2}/\pi^{1/4}) \hat{a}_0(\theta), \qquad m \geq 1. \end{aligned} \tag{2.3}$$

Our estimate of θ_1 is a value $\hat{\theta}_1$ which maximizes $\hat{J}_m(\theta)$.

We shall show in Section 4 that if m increases sufficiently quickly, yet more slowly than $n^{2/3}$, then there exists a $\hat{\theta}_1$ which gives at least a local maximum of $\hat{J}_m$ and is $\sqrt{n}$-consistent for θ_1. The most important aspect of our result is that we require only an *algebraic* moment condition on Y: $E(\|Y\|^t) < \infty$ for some $t > 0$. As we pointed out several paragraphs earlier, the latter tail condition is considerably weaker than that required for use of the indices I and $\hat{I}_m$: If $E(\|Y\|^{t'}) = \infty$ for some $t' > 0$ no matter how large, then $I(\theta)$ will be infinite for a range of values of θ.

The reason for this distinction of the Hermite function index is that Hermite functions are heavily weighted in the tails, by $e^{-x^2/2}$. That has the effect of alleviating pathological problems associated with the tail behaviour of Y. Johnstone [7] provides a succinct account of difficulties which can arise with nonweighted measures of "interestingness."

Another interesting feature is that the upper bound on the number, m, of terms appropriate for $\hat{J}_m(\theta)$, is roughly the square of the bound on the number appropriate for $\hat{I}_m(\theta)$ ($n^{2/3}$ vs. $n^{1/3}$). Likewise, the lower bound in the case of $\hat{J}_m$ is the square of that for $\hat{I}_m$. This happens because Hermite functions and Legendre polynomials are of quite different sizes: h_i decreases at rate $i^{-1/4}$ as i increases, whereas p_i does not decrease to 0.

In both Hermite and Legendre cases, $\sqrt{n}$-consistent orientation estimates have limit distributions of the same type. These limits may be expressed in terms of the unit vector which maximizes the square of a vector-indexed Gaussian process, as we shall show on each occasion. This type of limit also occurs when one is using empiric "interestingness" measures based on entropy and kernel density estimators [3]. It is possible to use kernel rather than orthogonal series density estimators to construct $\sqrt{n}$-consistent estimates of orientation, but the kernel estimators have to be substantially under-smoothed—for example, with nonnegative kernels we should use a window of size between $n^{-1/3}$ and $n^{-1/4}$, not $n^{-1/5}$; see [3].

3. Legendre polynomials. Let $P_0, P_1, \ldots$ be Legendre polynomials on the interval $(-1, 1)$. They are completely determined by orthogonality, by the fact that P_i is of degree i and by the relation $P_i(\pm 1) = (\pm 1)^i$. Orthonormal polynomials derived from the P_i's are

$$p_i(u) \equiv \left(i + \tfrac{1}{2}\right)^{1/2} P_i(u), \qquad -1 < u < 1,$$

and satisfy

$$\int_{-1<u<1} p_i(u) p_j(u)\, du = \delta_{ij}.$$

Put

$$q_i(u) \equiv p_i\{2\Phi(u) - 1\}, \qquad i \ge 0.$$

Fourier coefficients in a Legendre series expansion of the density f_θ of $U_\theta \equiv 2\Phi(\theta \cdot Y) - 1$ are $a_i(\theta) \equiv E\{q_i(\theta \cdot Y)\}$, $i \ge 0$. If f_θ is square-integrable, then

$$I(\theta) \equiv \int_{-1}^{1} \left\{ f_\theta(u) - \tfrac{1}{2} \right\}^2 du = \sum_{i=1}^{\infty} a_i(\theta)^2$$

is an index of the extent of departure from normality of the distribution of $\theta \cdot Y$. Our aim is to estimate the "most interesting" direction θ_1, which maximizes I.

Let $Y_1, \dots, Y_n$ be a random sample from the distribution of Y and put

$$\hat{a}_i(\theta) \equiv n^{-1} \sum_{j=1}^{n} q_i(\theta \cdot Y_j) \quad \text{and} \quad \hat{I}_m(\theta) \equiv \sum_{i=1}^{m} \hat{a}_i(\theta)^2, \qquad m \geq 1.$$

Choose $\hat{\theta}_1$ to maximize $\hat{I}_m$. Assume the following regularity condition on the p-variate density g of Y:

(3.1) g is compactly supported; for some $r \geq 2$, all rth-order directional derivatives of g are uniformly bounded; second-order directional derivatives are uniformly continuous in argument and in orientation; and the distribution associated with g has zero mean and identity covariance matrix.

THEOREM 3.1. *If* (3.1) *holds; if $m = m(n)$ diverges to infinity sufficiently slowly for $m/n^{1/3} \to 0$, yet sufficiently rapidly for $m/n^{1/\{4(r-1)\}} \to \infty$; and if all second-order derivatives of I at θ_1 are negative; then there exists a local maximum $\hat{\theta}_1$ of $\hat{I}_m$ such that $\hat{\theta}_1 - \theta_1 = O_p(n^{-1/2})$ as $n \to \infty$.*

In fact, $n^{1/2}(\hat{\theta}_1 - \theta_1)$ has a weak limit which may be defined as follows. There exists a continuous-path, zero-mean Gaussian process $\xi(\theta)$, indexed by $\theta \in \Omega$ with $\theta \perp \theta_1$, and a positive and continuous function $c(\theta)$, such that if θ^* maximizes $\xi(\theta)^2 c(\theta)$, then

$$n^{1/2}(\hat{\theta}_1 - \theta_1) \to \xi(\theta^*) c(\theta^*) \theta^* \tag{3.2}$$

in distribution. [Of course, θ^* is a random unit p-vector perpendicular to θ_1. Both $\xi(\theta^*)$ and $c(\theta^*)$ are random scalars.] The process ξ and function c are given in our proof of Theorem 3.1. Provided m satisfies the conditions in the theorem, ξ and c do not depend on the manner in which m diverges to infinity.

The $\sqrt{n}$-consistency claimed in Theorem 3.1 continues to be true in some, but not all, circumstances where $m \sim \text{const.}\, n^{1/3}$, although then the weak limit of $n^{1/2}(\hat{\theta}_1 - \theta_1)$ has a different form from that described in the previous paragraph. The $\sqrt{n}$-consistency fails if m diverges more rapidly than $n^{1/3}$. All this will become clear from our sketch proof of Theorem 3.1. Our assumption in Theorem 3.1 that all second-order derivatives of I at θ_1 are negative does no more than ensure that the maximum at θ_1 is attained in the usual quadratic manner. Note that since θ_1 does give a maximum then none of the second derivatives can be positive; our assumption only removes the possibility that one of them is 0.

In the remainder of this section we sketch a proof of Theorem 3.1. The reader interested in details should consult the proof of Theorem 4.1 in Section 4, which is very similar and given in detail. Let θ_0 be any fixed element of Ω; we have in mind $\theta_0 = \theta_1$. For θ close to θ_0, write

$$\theta = (1 - \eta^2)^{1/2} \theta_0 + \eta \theta_{00}, \tag{3.3}$$

where θ_{00} (perpendicular to θ_0) is in the same plane as θ and θ_0, and where $\eta \equiv \theta \cdot \theta_{00} \to 0$ as $\theta \to \theta_0$. We assume throughout that $|\eta| \leq n^{-1/2+\varepsilon}$, for some fixed $\varepsilon < 1/6$.

Since second-order directional derivatives of g are bounded and continuous, then

$$I(\theta) = I(\theta_0) + \eta I_1(\theta_0, \theta_{00}) + \tfrac{1}{2}\eta^2 I_2(\theta_0, \theta_{00}) + o(\eta^2), \tag{3.4}$$

where the continuous functions I_1 and I_2 do not depend on η. (Section 3 of [4] discusses results of this type.) Take $\theta_0 = \theta_1$. Since θ is a turning point of I, then $I_1(\theta_1, \theta_{00}) = 0$ for all $\theta_{00} \perp \theta_1$; and by hypothesis, $I_2(\theta_1, \theta_{00}) < 0$. Thus, for $\theta = \theta(\theta_1, \theta_{00}, \eta)$ given by (3.3),

$$I(\theta) = I(\theta_1) - \tfrac{1}{2}\eta^2 |I_2(\theta_1, \theta_{00})| + o(\eta^2). \tag{3.5}$$

Put $A_{ij} \equiv q_i(\theta_0 \cdot Y_j)$, $B_{ij} \equiv q_i(\theta \cdot Y_j) - q_i(\theta_0 \cdot Y_j)$, $b_i(\theta) \equiv a_i(\theta) - a_i(\theta_0)$, $\hat{b}_i(\theta) \equiv \hat{a}_i(\theta) - \hat{a}_i(\theta_0)$, $\hat{\alpha}_i(\theta) \equiv \hat{a}_i(\theta) - a_i(\theta)$, $\hat{\beta}_i(\theta) \equiv \hat{b}_i(\theta) - b_i(\theta)$. Then

$$\begin{aligned} &\hat{a}_i(\theta)^2 - \hat{a}_i(\theta_0)^2 - \{a_i(\theta)^2 - a_i(\theta_0)^2\} \\ &\quad = \hat{\beta}_i(\theta)^2 + 2\hat{\alpha}_i(\theta_0)\hat{\beta}_i(\theta) + 2a_i(\theta_0)\hat{\beta}_i(\theta) \\ &\quad\quad + 2\hat{\alpha}_i(\theta_0) b_i(\theta) + 2b_i(\theta)\hat{\beta}_i(\theta), \end{aligned} \tag{3.6}$$

whence

$$\begin{aligned} \hat{I}_m(\theta) &= \hat{I}_m(\theta_0) + I(\theta) - I(\theta_0) \\ &\quad - \sum_{i=m+1}^{\infty} \{a_i(\theta)^2 - a_i(\theta_0)^2\} + \sum_{k=1}^{5} S_k, \end{aligned} \tag{3.7}$$

where

$$\begin{aligned} &S_1 \equiv \sum_{i=1}^{m} \hat{\beta}_i(\theta)^2, \qquad S_2 \equiv 2\sum_{i=1}^{m} \hat{\alpha}_i(\theta_0)\hat{\beta}_i(\theta), \qquad S_3 \equiv \sum_{i=1}^{m} a_i(\theta_0)\hat{\beta}_i(\theta), \\ &S_4 \equiv 2\sum_{i=1}^{m} \hat{\alpha}_i(\theta_0) b_i(\theta), \qquad S_5 \equiv 2\sum_{i=1}^{m} b_i(\theta)\hat{\beta}_i(\theta). \end{aligned} \tag{3.8}$$

Derivatives of Legendre polynomials admit the following expansion: For $s, t \ge 0$,

$$\begin{aligned} p_i^{(s)}(\cos\psi) &= \sum_{j=0}^{t} c_{isj}(\sin\psi)^{-(s+j+1/2)} \\ &\qquad \times \cos\{(i - j + \tfrac{1}{2})\psi - (s + j + \tfrac{1}{2})(\pi/2)\} \\ &\quad + O(i^{s-t-1}) \end{aligned} \tag{3.9}$$

as $i \to \infty$, uniformly in $\delta < \psi < \pi - \delta$ for each $\delta > 0$, where c_{isj} is a constant satisfying

$$c_{isj} = i^{s-j}\Gamma\left(s + \frac{1}{2}\right)^{-1} 2^{1/2-j}\left(\frac{1}{2}\right)s\binom{j + s - \frac{1}{2}}{j}\left(\frac{1}{2} - s\right)j\{1 + O(i^{-1})\} = O(i^{s-j})$$

[9, pages 224 and 232]. From this formula, and r integrations by parts in expressions for $a_i(\theta)$ and $b_i(\theta)$, we obtain

$$\sup_\theta |a_i(\theta)| \le Ci^{-r}, \qquad \sup_\theta |b_i(\theta)| \le C\eta i^{1-r}, \tag{3.10}$$
$$|A_{ij}| \le C, \qquad |B_{ij}| \le C\eta i.$$

Therefore, since $n/m^{4(r-1)} \to 0$,

$$\sum_{i=m+1}^{\infty} \left|a_i(\theta)^2 - a_i(\theta_0)^2\right| \le C \sum_{i=m+1}^{\infty} \eta i^{1-r} i^{-r} = o(\eta^2 + n^{-1}). \tag{3.11}$$

It remains to elucidate properties of $S_1, \dots, S_5$, defined at (3.8). First we treat S_1 and S_2. Write (C_{ij}, c_i) for either $(A_{ij}, a_i(\theta_0))$ or (B_{ij}, b_i), where b_i denotes $b_i(\theta)$. Both S_1 and $\frac{1}{2}S_2$ have the form

$$S = n^{-2} \sum_{i=1}^{m} \sum_{j=1}^{n} \sum_{k=1}^{n} (B_{ij} - b_i)(C_{ik} - c_i) = T_1 + T_2,$$

where

$$T_1 \equiv n^{-2} \sum_{i=1}^{m} \sum_{j=1}^{n} (B_{ij} - b_i)(C_{ij} - c_i),$$

$$T_2 \equiv n^{-2} \sum\sum_{j<k} \sum_{i=1}^{m} \left\{(B_{ij} - b_i)(C_{ik} - c_i) + (B_{ik} - b_i)(C_{ij} - c_i)\right\}.$$

Now, T_1 is a sum of independent random variables with mean

$$n^{-1} \sum_{i=1}^{m} \{E(B_{i1}C_{i1}) - b_i c_i\} = n^{-1} \sum_{i=1}^{m} E(B_{i1}C_{i1}) + o(n^{-1}).$$

From this formula and extensive use of (3.9), we obtain

$$E(S_1 + S_2) = c_1 n^{-1}\eta m\{1 + o(1)\} + c_2 n^{-1}\eta^2 m^3\{1 + o(1)\} + o(n^{-1}),$$

where c_1 and $c_2 > 0$ denote continuous functions of θ_0 and θ_{00}.

The term $T_1 - ET_1$ is $o_p(\eta^2 + n^{-1})$. The series T_2 is a zero-mean martingale: $T_2 = n^{-2}\sum_{2 \le k \le n} Z_k$, where

$$Z_k \equiv \sum_{j=1}^{k-1} \sum_{i=1}^{m} \left\{(B_{ij} - b_i)(C_{ik} - c_i) + (B_{ik} - b_i)(C_{ij} - c_i)\right\}$$

and $E(Z_k|Y_1, \dots, Y_{k-1}) = 0$. In the case $(C_{ij}, c_i) \equiv (B_{ij}, b_i)$, two applications of Rosenthal's inequality [5, page 23] may be used to prove that $T_2 = o_p(\eta^2 + n^{-1})$. In the case $(C_{ij}, c_i) \equiv (A_{ij}, a_i(\theta_0))$, judicious use of (3.9) and (3.10) and a martingale central limit theorem [5, page 58] allow that $T_2 = n^{-1}\eta m^{3/2} Z_1$, where Z_1 is asymptotically normal $N(0, \sigma^2)$ for some $\sigma^2(\theta_{00}) > 0$. Thus,

$$S_1 + S_2 = c_2 n^{-1}\eta^2 m^3 + n^{-1}\eta m^{3/2} Z_1 + o_p\left(\eta^2 + n^{-1} + n^{-1}\eta^2 m^3 + n^{-1}\eta m^{3/2}\right). \tag{3.12}$$

Finally we turn our attention to S_3, S_4 and S_5. From the results on convergence of Jacobi polynomial series [12, page 244] and Fourier trigonometric series

[14, page 57], we may prove that the infinite-series analogues of S_3 and S_4 converge. In fact, each of S_3 and S_4 is $o_p(\eta^2 + n^{-1})$ away from its infinite-series counterpart. Arguing thus, we obtain

$$S_3 + S_4 = \eta n^{-1/2} Z(\theta_{00}) + o_p(\eta^2 + n^{-1}),$$

where Z is a sum of independent random variables with zero means and converges weakly to a continuous-path Gaussian process ξ. Furthermore, $S_5 = o_p(\eta^2 + n^{-1})$. From these results and (3.5), (3.7), (3.11) and (3.12), we see that, for $\theta = (1 - \eta^2)^{1/2}\theta_1 + \eta\theta_{00}$ and $\theta_{00} \perp \theta_1$,

$$\begin{aligned}\hat{I}_m(\theta) = \hat{I}_m(\theta_1) &+ \eta n^{-1/2}\left(n^{-1/2}m^{3/2}Z_1 + Z\right) - \tfrac{1}{2}\eta^2|I_2(\theta_1, \theta_{00})| \\ &+ c_2 n^{-1}\eta^2 m^3 + o_p\left(\eta^2 + n^{-1} + n^{-1}\eta^2 m^3 + n^{-1}\eta m^{3/2}\right).\end{aligned} \tag{3.13}$$

If $m/n^{1/3} \to \infty$, or if $m/n^{1/3} \to l$, where l is finite and sufficiently large, then the argument which we shall give in the next paragraph may be reworked to show that no local maximum of $\hat{I}_m(\theta)$ is $\sqrt{n}$-consistent for θ_1. If $m \sim ln^{1/3}$ for sufficiently small l, then $\sqrt{n}$-consistency is possible, but with a limit distribution different from that which we shall derive. Only when $m/n^{1/3} \to 0$ do the terms in Z_1 and c_2 make a negligible contribution to the right-hand side of (3.13), which becomes

$$\hat{I}_m(\theta) = \hat{I}_m(\theta_1) + \eta n^{-1/2} Z(\theta_{00}) - \tfrac{1}{2}\eta^2|I_2(\theta_1, \theta_{00})| + o_p\left(\eta^2 + n^{-1}\right). \tag{3.14}$$

Write $c(\theta_{00}) \equiv 1/|I_2(\theta_1, \theta_{00})|$. For fixed $\theta_{00} \perp \theta_1$, the nonremainder part of the right-hand side of (3.14) is maximized by taking $\eta \equiv n^{1/2}Z(\theta_{00})c(\theta_{00})$, and then $\hat{I}_m(\theta) = \hat{I}_m(\theta_1) + \frac{1}{2}n^{-1}Z(\theta_{00})^2 c(\theta_{00}) + o_p(n^{-1})$. This in turn is maximized by choosing θ_{00} to maximize $Z(\theta_{00})^2 c(\theta_{00})$. Arguing thus and remembering that $\theta - \theta_1 = \eta\theta_{00} + O(\eta^2)$, we obtain (3.2).

4. Hermite functions. Hermite functions and the "interestingness" indices J and $\hat{J}_m$ were defined at (2.1), (2.2) and (2.3), respectively. In the present section we show that, under smoothness conditions on g and moment conditions on Y, $\hat{J}_m(\theta)$ is a practical empiric measure of departure from normality.

Assume the following regularity conditions on the density g of Y:

(4.1) for some $r \geq 2$, all rth-order directional derivatives of g are uniformly bounded; second-order directional derivatives are uniformly continuous in argument and in orientation; $|g_\theta|$, $|g_\theta'|$ and $|g_\theta''|$ are bounded uniformly in θ and argument; for some $t_1, t_2 > 0$, chosen sufficiently large, $E(\|Y\|^{t_1}) < \infty$ and

$$\sup_{\theta \in \Omega} \int \left(1 + \|y\|^{t_2}\right)|D_\theta^r g(y)|\,dy < \infty;$$

and the distribution associated with g has zero mean and identity covariance matrix.

THEOREM 4.1. *If* (4.1) *holds; if* $m = m(n)$ *diverges to infinity sufficiently slowly for* $m/n^{2/3} \to 0$, *yet sufficiently rapidly for* $m/n^{1/\{2(r-1)\}} \to \infty$; *and if all* *second-order derivatives of J at θ_1 are negative; then there exists a local maximum $\hat{\theta}_1$ of $\hat{J}_m$ such that $\hat{\theta}_1 - \theta_1 = O_p(n^{-1/2})$ as $n \to \infty$.*

Formula (3.2) again describes the weak limit of $n^{1/2}(\hat{\theta}_1 - \theta_1)$. Our proof of Theorem 4.1 gives expressions for the Gaussian process ξ and the function c appearing in (3.2). Provided m satisfies the conditions in the theorem, ξ and c do not depend on the manner in which m diverges to infinity.

The $\sqrt{n}$-consistency in Theorem 4.1 continues to hold true if $m/n^{2/3}$ converges to a nonzero constant l, provided l is sufficiently small. But it fails if l is large, as our proof will show.

In the proof we follow a route which yields the theorem expediently. We do not attempt to determine economical values of t_1 and t_2. Our proof would follow a different, much more lengthy route if we sought the "best" t_1 and t_2; we do not know what those values are.

The remainder of this section is devoted to a proof of Theorem 4.1.

PROOF OF THEOREM 4.1. Let θ_0 denote any fixed element of Ω, such as θ_1, and for θ close to θ_0, express θ as in (3.3), where $\theta_{00} \perp \theta_0$ and $\eta \equiv \theta \cdot \theta_{00}$. The index $J(\theta)$ admits expansion (3.4), which may be simplified to (3.5) when $\theta_0 = \theta_1$,

$$J(\theta) = J(\theta_1) - \tfrac{1}{2}\eta^2 |J_2(\theta_1, \theta_{00})| + o(\eta^2) \tag{4.2}$$

as $\eta \to 0$, where $J_2(\theta_1, \cdot)$ is a continuous, strictly negative function. We assume throughout that $m \le Cn^{2/3}$, and $0 \le \eta \le Cn^{1/2+\varepsilon}$ for a small $\varepsilon > 0$.

Put $A_{ij} \equiv h_i(\theta_0 \cdot Y_j)$, $B_{ij} \equiv h_i(\theta \cdot Y_j) - h_i(\theta_0 \cdot Y_j)$, $b_i(\theta) \equiv a_i(\theta) - a_i(\theta_0)$, $\hat{b}_i(\theta) \equiv \hat{a}_i(\theta) - \hat{a}_i(\theta_0)$, $\hat{\alpha}_i(\theta) \equiv \hat{a}_i(\theta) - a_i(\theta)$, $\hat{\beta}_i(\theta) \equiv \hat{b}_i(\theta) - b_i(\theta)$. Result (3.6) continues to hold, and, in addition,

$$\hat{a}_0(\theta) - \hat{a}_0(\theta_0) - \{a_0(\theta) - a_0(\theta_0)\} = \hat{\beta}_0(\theta) = -\eta(\pi^{1/4}/2^{1/2}) S_6(\theta_{00}) + o_p(\eta^2),$$

uniformly in $\theta_{00} \perp \theta_0$, where

$$S_6(\theta_{00}) \equiv -(2^{1/2}/\pi^{1/4}) n^{-1} \sum_{j=1}^{n} \left[(\theta_{00} \cdot Y_j) h_0'(\theta_0 \cdot Y_j) - E\{(\theta_{00} \cdot Y) h_0'(\theta_0 \cdot Y)\} \right].$$

Define $\hat{J}_m(\theta)$ as in (2.3). Then we have the following analogue of (3.7):

$$\begin{aligned} \hat{J}_m(\theta) &= \hat{J}_m(\theta_0) + J(\theta) - J(\theta_0) \\ &\quad - \sum_{i=m+1}^{\infty} \{a_i(\theta)^2 - a_i(\theta_0)^2\} + \sum_{k=1}^{5} S_k + \eta S_6 \\ &\quad + o_p(\eta^2), \end{aligned} \tag{4.3}$$

uniformly in $\theta_{00} \perp \theta_0$, where $S_1, \dots, S_5$ are as at (3.8) but with the range of summation changed to $0 \le i \le m$.

The following lemma provides basic analytic properties of Hermite functions.

LEMMA 4.1.

(i) $H_i'(u) = 2iH_{i-1}(u)$;
(ii) $h_i'(u) = (2i)^{1/2}h_{i-1}(u) - uh_i(u)$;
(iii) $h_i'(u) = uh_i(u) - 2^{1/2}(i+1)^{1/2}h_{i+1}(u)$;
(iv) $h_i''(u) = (u^2 - 2i - 1)h_i(u)$;
(v) *for* $l \geq 0$, $\sup_u|h_i^{(l)}(u)| \leq C(i+1)^{l/2}$;
(vi) $h_i(u) = (2/\pi^2 i)^{1/4}\cos(N_i^{1/2}u - i\pi/2) + R_{1i}(u)$, *where* $|R_{1i}(u)| \leq C(i+1)^{-1/2}(1+|u|^{5/2})$ *and* (*here and below*) $N_i = 2i+1$;
(vii) *for each integer* $s \geq 1$,

$$h_i(u) = c_i\{\cos(N_i^{1/2}u - i\pi/2) + (u^3/6)N_i^{-1/2}\sin(N_i^{1/2}u - i\pi/2)\} + N_i^{-5/4}\{q_{i1}(u)\cos(N_i^{1/2}u) + q_{i2}(u)\sin(N_i^{1/2}u)\} + R_{2i}(u),$$

where the constant c_i *satisfies* $c_i = (2/\pi^2 i)^{1/4}\{1 + O(i^{-1})\}$ *as* $i \to \infty$, q_{i1} *and* q_{i2} *are polynomials of degree* $3s+3$ *with coefficients bounded uniformly in* i, *and* $|R_{2i}(u)| \leq C(i+1)^{-(2s+7)/4}(1+|u|^{(6s+11)/2})$.

Results (i)–(iv) follow from [9], page 252, results (v)–(vii) from [11], pages 324 and 332–333.

Taking $W \equiv (\theta - \theta_0)\cdot Y$, applying successively results (ii) and (vii) and integrating r times by parts to simplify trigonometric terms, we obtain

$$|b_i(\theta)| = \left|E\left[W\int_0^1\{(2i)^{1/2}h_{i-1}(\theta_0\cdot Y + tW) - (\theta_0\cdot Y + tW)h_i(\theta_0\cdot Y + tW)\}\,dt\right]\right| \leq C\eta i^{-(2r-1)/4}.$$

Likewise,

$$|a_i(\theta)| \leq Ci^{-(2r-1)/4}. \tag{4.4}$$

Therefore, since $n/m^{2(r-1)} \to 0$,

$$\sum_{i=m+1}^{\infty}\left|a_i(\theta)^2 - a_i(\theta_0)^2\right| \leq C\sum_{i=m+1}^{\infty}\eta i^{-r} = o(\eta^2 + n^{-1}). \tag{4.5}$$

Next we examine the terms S_1 and S_2 in (4.3). Let (C_{ij}, c_i) denote either $(A_{ij}, a_i(\theta_0))$ or (B_{ij}, b_i), where $B_{ij} = B_{ij}(\theta)$ and $b_i = b_i(\theta)$. Both S_1 and S_2 have the form $T_1 + T_2$, where

$$T_1(\theta) \equiv n^{-2}\sum_{j=1}^{n}\sum_{i=0}^{m}(B_{ij} - b_i)(C_{ij} - c_i),$$
$$T_2(\theta) \equiv n^{-2}\sum\sum_{j\neq k}\sum_{i=0}^{m}(B_{ij} - b_i)(C_{ik} - c_i). \tag{4.6}$$

We begin by treating T_2. Let $\sup^{(1)}$ denote the supremum over $\theta = \theta(\theta_{00}, \eta)$ [see (3.3)] with $\theta_{00} \perp \theta_0$, $0 \le \eta \le n^{-1/2+\varepsilon}$ and ε sufficiently small.

Lemma 4.2. $\sup^{(1)}(\eta^2 + n^{-1})^{-1}|T_2(\theta)| \to 0$ *in probability.*

Proof. Since $B_{ij}(\theta) = h_i(\theta \cdot Y_j) - h_i(\theta_0 \cdot Y_j)$, then if $\omega, \omega' \in \Omega$,

$$|B_{ij}(\omega) - B_{ij}(\omega')| \le C\|\omega - \omega'\|(i+1)^{1/4}\left(1 + \|Y_j\|^{7/2}\right)$$

[Lemma 4.1(ii) and (vi)], whence, since $|B_{ij}| + |C_{ij}| \le C$ [Lemma 4.1(v)],

$$|T_2(\omega) - T_2(\omega')| \le C\|\omega - \omega'\| m^{5/4} n^{-1} \sum_{j=1}^{n} \left(1 + \|Y_j\|^{7/2}\right). \tag{4.7}$$

Let $\{\omega_1, \dots, \omega_N\}$ be a collection of unit vectors such that, for each $\theta \in \Omega$, there exists $j = j(\theta)$ with $1 \le j \le N$ and $\|\theta - \omega_{j(\theta)}\| \le n^{-2}$. We may choose $N \le Cn^{2p}$. By (4.7) and since $m \le Cn^{2/3}$,

$$\sup_{\theta \in \Omega} |T_2(\theta) - T_2(\omega_{j(\theta)})| = O(n^{-7/6}) \quad \text{almost surely}.$$

Therefore, to prove Lemma 4.2, it suffices to show that, with $\Theta \equiv \{\omega_j \colon 1 \le j \le N$ and $\|\omega_j - \theta_0\| \le n^{-1/2+\varepsilon}\}$, we have for sufficiently small ε,

$$\sup_{\theta \in \Theta} \left(\eta^2 + n^{-1}\right)^{-1} |T_2(\theta)| = o_p(1). \tag{4.8}$$

We illustrate the technique by treating the case where $C_{ij} = A_{ij}$ and $c_i = a_i(\theta_0)$.
Observe from Lemma 4.1(v) that, with $W_j \equiv (\theta - \theta_0) \cdot Y_j$,

$$B_{ij} = \sum_{l=1}^{4} (1/l!) W_j^l h_i^{(l)}(\theta_0 \cdot Y_j) + R_{ij}, \tag{4.9}$$

where $|R_{ij}| \le C\eta^5 (i+1)^{5/2} \|Y_j\|^5$. Put $a_i = a_i(\theta_0)$,

$$d_i \equiv \sum_{l=1}^{4} (1/l!) E\left\{W_j^l h_i^{(l)}(\theta_0 \cdot Y_j)\right\},$$

$$T_3(\theta) \equiv n^{-2} \sum_{j \ne k} \sum \sum_{i=0}^{m} \left\{ \sum_{l=1}^{4} (1/l!) W_j^l h_i^{(l)}(\theta_0 \cdot Y_j) - d_i \right\} (A_{ik} - a_i),$$

(4.10)

$$Z_k \equiv \sum_{j=1}^{k-1} \sum_{i=0}^{m} \left\{ (R_{ij} - ER_{ij})(A_{ik} - a_i) + (R_{ik} - ER_{ik})(A_{ij} - a_i) \right\},$$

$$T_4(\theta) \equiv n^{-2} \sum_{j \ne k} \sum \sum_{i=0}^{m} (R_{ij} - ER_{ij})(A_{ik} - a_i) = n^{-2} \sum_{k=2}^{n} Z_k.$$

Notice that $T_2 = T_3 + T_4$ and $E(Z_k | Y_1, \dots, Y_{k-1}) = 0$. By Hölder's and

Rosenthal's inequalities [5, page 23],

$$E\left\{T_4(\theta)^{2s}\right\} \le Cn^{-4s}\left\{\sum_{k=2}^{n}\left(EZ_k^{2s}\right)^{1/s}\right\}^s.$$

Conditional on Y_k, Z_k is a sum of independent random variables, and so

$$\begin{aligned} E\left(Z_k^{2s}\right) &\le C_1 n^s E\left(\left[\sum_{i=0}^{m}\{(R_{i1}-ER_{i1})(A_{ik}-a_i)\right.\right. \\ &\qquad \left.\left. +(R_{ik}-ER_{ik})(A_{i1}-a_i)\}\right]^{2s}\right) \\ &\le C_2 n^s\left(\eta^5 m^{7/2}\right)^{2s}. \end{aligned}$$

Conversely, since $m \le Cn^{2/3}$ and $\eta \le n^{-1/2+\varepsilon}$, then

$$E\left\{T_4(\theta)^{2s}\right\} \le C_1\left(n^{-1}\eta^5 m^{7/2}\right)^{2s} \le C_2\left(\eta^2 n^{-(1/6)+3\varepsilon}\right)^{2s}.$$

Choosing $\varepsilon < 1/18$ and s large and using Markov's inequality, we deduce that for each $\xi > 0$,

$$P\left\{\sup_{\theta\in\Theta'}\eta^{-2}|T_4(\theta)| > \xi\right\} \le C(\xi)n^{2p}\left(n^{-(1/6)+3\varepsilon}\right)^{2s} \to 0,$$

from which it follows that

$$\sup_{\theta\in\Theta'}\left(\eta^2+n^{-1}\right)^{-1}|T_4(\theta)| = o_p(1). \tag{4.11}$$

Next we prove a similar result for $T_3(\theta)$, defined at (4.10),

$$\sup_{\theta\in\Theta'}\left(\eta^2+n^{-1}\right)^{-1}|T_3(\theta)| = o_p(1). \tag{4.12}$$

Observe that W_j^l equals a bounded linear combination of l-products of components of Y_j, in which the coefficient of each product is dominated by $C\eta^l$. There is only a bounded number of these products. Let $U_j = U_j(l)$ denote any one of them. Result (4.12) will follow if we show that for $1 \le l \le 4$, the random variable

$$T_5 \equiv n^{-2}\eta^l\sum_{j\ne k}\sum\sum_{i=0}^{m}\left\{U_j h_i^{(l)}(\theta_0\cdot Y_j) - EU_j h_i^{(l)}(\theta_0\cdot Y_j)\right\}(A_{ik}-a_i)$$

satisfies

$$T_5 = o_p\left(\eta^2+n^{-1}\right). \tag{4.13}$$

Put $e_i \equiv E\{U_1 h_i^{(l)}(\theta_0 \cdot Y_1)\}$. Then

$$E(T_5^2) = O\left(n^{-2}\eta^{2l} E\left[\sum_{i=0}^{m} \{U_1 h_i^{(l)}(\theta_0 \cdot Y_1) - e_i\}(A_{i2} - a_i)\right]^2\right)$$

$$= O\left[n^{-2}\eta^{2l} \sum_{i_1=0}^{m} \sum_{i_2=0}^{m} \{EU_1^2 h_{i_1}^{(l)}(\theta_0 \cdot Y_1) h_{i_2}^{(l)}(\theta_0 \cdot Y_1) - e_{i_1} e_{i_2}\}\right.$$

$$\left. \times \{E(A_{i_1 2} A_{i_2 2}) - a_{i_1} a_{i_2}\}\right].$$

Use Lemma 4.1(ii) to express $h_i^{(l)}$ in terms of $h_{i \pm k}$ for $0 \le k \le l$; use Lemma 4.1(vii) to expand $h_{i \pm k}$; use trigonometric formulae such as

$$2\cos(N_{i_1}^{1/2} u)\cos(N_{i_2}^{1/2} u) = \cos\{(N_{i_1}^{1/2} + N_{i_2}^{1/2})u\} + \cos\{(N_{i_1}^{1/2} - N_{i_2}^{1/2})u\}$$

to simplify products such as $h_{i_1}(u) h_{i_2}(u)$; and finally, integrate by parts r times, to deduce that

$$\left|E\{U_1^2 h_{i_1}^{(l)}(\theta_0 \cdot Y_1) h_{i_2}^{(l)}(\theta_0 \cdot Y_1)\} - e_{i_1} e_{i_2}\right|$$
$$\le C i_1^{(2l-1)/4} i_2^{(2l-1)/4} (1 + |i_1 - i_2|^{1/2})^{-r}.$$

Similarly,

$$\left|E(A_{i_1 2} A_{i_2 2}) - a_{i_1} a_{i_2}\right| \le C(1 + i_1)^{-1/4}(1 + i_2)^{-1/4}(1 + |i_1 - i_2|^{1/2})^{-r}.$$

Therefore, noting that $m \le Cn^{2/3}$ and $\eta \le n^{-1/2+\varepsilon}$,

$$E(T_5^2) = O\left\{n^{-2}\eta^{2l} \sum_{i_1=0}^{m} \sum_{i_2=0}^{m} (i_1 i_2)^{(l-1)/2}(1 + |i_1 - i_2|)^{-r}\right\}$$

$$= O(\eta^{2l} n^{-2} m^l) = O(\eta^2 n^{-2} m) = O\{(\eta^4 + n^{-2}) n^{-1} m\} = o(\eta^4 + n^{-1}),$$

from which follows (4.13). This completes the proof of (4.12).

Result (4.8) follows from (4.11) and (4.12). This completes the proof of Lemma 4.2. □

Next we examine $T_1(\theta)$, defined at (4.6). It may be proved from Lemma 4.1(ii) and (vii) that, with $W \equiv (\theta - \theta_0) \cdot Y$,

$$\beta_i \equiv E\{h_i(\theta \cdot Y) - h_i(\theta_0 \cdot Y)\}^2 \sim E\{W^2 h_i^1(\theta_0 \cdot Y)^2\}$$
$$\sim \eta^2 (2/\pi^2 i)^{1/2} 2iE\{(\theta_{00} \cdot Y)^2 \cos^2(N_i^{1/2}\theta_0 \cdot Y - i\pi/2)\}$$
$$\sim \eta^2 i^{1/2} 2^{1/2} \pi^{-1} E(\theta_{00} \cdot Y)^2.$$

Similarly, $\alpha_i \equiv E[\{h_i(\theta \cdot Y) - h_i(\theta_0 \cdot Y)\} h_i(\theta_0 \cdot Y)]$ satisfies

$$\alpha_i = -\eta^2 (i/2\pi^2)^{1/2} E(\theta_{00} \cdot Y)^2 + O(\eta i^{-1/2} + \eta^2 i^{1/4} + \eta^3 i^{1/2}).$$

Put $\gamma_i \equiv \alpha_i$ if in the definition of T_1, $(C_{ij}, c_i) = (A_{ij}, a_i(\theta_0))$, and $\gamma_i \equiv \beta_i$ if $(C_{ij}, c_i) = (B_{ij}, b_i)$. Then

$$E(T_1) = n^{-1} \sum_{i=0}^{m} (\gamma_i - b_i c_i) = n^{-1} \sum_{i=0}^{m} \gamma_i + o(n^{-1}),$$

and so the sum of $E(T_1)$ from both cases equals $\gamma n^{-1}\eta^2 m^{3/2}\{1 + o(1)\} + o(n^{-1})$, where $\gamma \equiv (2/3\pi)(2^{1/2} - 2^{-1/2}) > 0$. In both cases, $T_1 - E(T_1)$ is a sum of independent random variables with zero means, and a modification of the argument leading to Lemma 4.2 gives

$$\sup{}^{(1)}(\eta^2 + n^{-1})^{-1}|T_1(\theta)| \to 0$$

in probability. Combining Lemma 4.2 with the results in this paragraph, we conclude that

$$S_1 + S_2 = \gamma n^{-1}\eta^2 m^{3/2}\{1 + o(1)\} + o_p(\eta^2 + n^{-1}). \tag{4.14}$$

Next we analyse the terms S_3, S_4 and S_5 in (4.3). Put $\chi(u|\theta_0, \theta_{00}) \equiv E(\theta_{00} \cdot Y|\theta_0 \cdot Y = u)$, $K(u|\theta_0, \theta_{00}) \equiv -(d/du)\{\chi(u|\theta_0, \theta_{00})g_{\theta_0}(u)\}$ and

$$k_i(\theta_0, \theta_{00}) \equiv \int K(u|\theta_0, \theta_{00})h_i(u)\,du = E\{(\theta_{00} \cdot Y)h_i'(\theta_0 \cdot Y)\}.$$

By judicious use of Lemma 4.1(ii) and (iii) (both parts are needed), of Uspensky's theorem [11, page 381, and 13], and of a standard argument for estimating Hermite Fourier coefficients [11, page 369], we may prove that with

$$G_m(u) \equiv \sum_{i=0}^{m} a_i(\theta_0)h_i'(u) \quad \text{and} \quad K_m(u) \equiv \sum_{i=0}^{m} k_i(\theta_0, \theta_{00})h_i(u),$$

we have

$$\sup_{\theta_0 \in \Omega} \ \sup_{-\infty < u < \infty} \left(1 + |u|^{5/2}\right)^{-1} |g_{\theta_0}'(u) - G_m(u)| \to 0, \tag{4.15}$$

$$\sup_{\theta_0 \in \Omega;\ \theta_{00} \perp \theta_0} \ \sup_{-\infty < u < \infty} \left(1 + |u|^{5/2}\right)^{-1} |K(u) - K_m(u)| \to 0 \tag{4.16}$$

as $m \to \infty$.

To treat $S_3 \equiv 2\Sigma_{0 \le i \le m} a_i(\theta_0)\hat{\beta}_i(\theta)$, observe that $\hat{\beta}_i = n^{-1}\Sigma_{1 \le j \le n}(B_{ij} - EB_{ij})$. Expand B_{ij} as in (4.9) and adapt the argument succeeding (4.9) to prove that for fixed θ_0,

$$\sup{}^{(1)}(\eta^2 + n^{-1})^{-1}\left|S_3 - 2n^{-1} \sum_{i=0}^{m} a_i(\theta_0) \sum_{j=1}^{n} \{W_j h_i'(\theta_0 \cdot Y_j) - EW_j h_i'(\theta_0 \cdot Y_j)\}\right|$$
$$= o_p(1),$$

where $\sup^{(1)}$ denotes the supremum over $\theta = \theta(\theta_{00}, \eta)$ [see (2.3)] with $\theta_{00} \perp \theta_0$ and $0 \le \eta \le n^{1/2+\varepsilon}$. Put $\mu(\theta_0, \theta_{00}) \equiv E\{(\theta_{00} \cdot Y)g_{\theta_0}'(\theta_0 \cdot Y)\}$. Using the bound (4.4) on

$|a_i(\theta_0)|$ and the fact that $W_j = \eta\theta_{00} \cdot Y_j + \{(1-\eta^2)^{1/2} - 1\}\theta_0 \cdot Y_j$, we obtain

$$n^{-1} \sum_{i=0}^{m} a_i(\theta_0) \sum_{j=1}^{n} \{W_j h_i'(\theta_0 \cdot Y_j) - EW_j h_i'(\theta_0 \cdot Y_j)\}$$
$$= \eta n^{-1} \sum_{j=1}^{n} \{(\theta_{00} \cdot Y_j) G_m(\theta_0 \cdot Y_j) - E(\theta_{00} \cdot Y_j) G_m(\theta_0 \cdot Y_j)\} + o_p(\eta^2 + n^{-1})$$
$$= \eta n^{-1} \sum_{j=1}^{n} \{(\theta_{00} \cdot Y_j) g\theta_0'(\theta_0 \cdot Y_j) - \mu(\theta_0, \theta_{00})\} + o_p(\eta^2 + n^{-1}),$$

the last identity coming from (4.15). Thus for fixed θ_0,

$$\sup{}^{(1)}(\eta^2 + n^{-1})^{-1} \left| S_3 - 2\eta n^{-1} \sum_{j=1}^{n} \{(\theta_{00} \cdot Y_j) g_{\theta_0}'(\theta_0 \cdot Y_j) - \mu(\theta_0, \theta_{00})\} \right| = o_p(1). \tag{4.17}$$

A similar argument, using (4.16) in place of (4.15), gives

$$\sup{}^{(1)}(\eta^2 + n^{-1})^{-1} \left| S_4 - 2\eta n^{-1} \sum_{j=1}^{n} \{K(\theta_0 \cdot Y_j | \theta_0, \theta_{00}) - EK(\theta_0 \cdot Y | \theta_0, \theta_{00})\} \right| = o_p(1), \tag{4.18}$$

and also

$$\sup{}^{(1)}(\eta^2 + n^{-1})^{-1} |S_5| = o_p(1). \tag{4.19}$$

Put

$$A_1(y|\theta_0, \theta_{00}) \equiv 2(\theta_{00} \cdot y) g_{\theta_0}'(\theta_0 \cdot y) + 2K(\theta_0 \cdot y | \theta_0, \theta_{00}) - (2^{1/2}/\pi^{1/4})(\theta_{00} \cdot y) h_0'(\theta_0 \cdot y)$$

and

$$A(y|\theta_0, \theta_{00}) \equiv A_1(y|\theta_0, \theta_{00}) - EA_1(Y|\theta_0, \theta_{00}).$$

Take $\theta_0 = \theta_1$ and let $Z(\theta_{00}) \equiv n^{-1/2} \sum_{1 \le j \le n} A(Y_j | \theta_1, \theta_{00})$. Combining results (4.2), (4.3), (4.5), (4.14) and (4.17)–(4.19), we obtain

$$\hat{J}_m(\theta) = \hat{J}_m(\theta_0) + \eta n^{-1/2} Z(\theta_{00}) - \tfrac{1}{2}\eta^2 \{ |J_2(\theta_1, \theta_{00})| - 2\gamma n^{-1} m^{3/2} \} + o_p(\eta^2 + n^{-1}), \tag{4.20}$$

uniformly in $\theta_{00} \perp \theta_0$ and $0 \le \eta \le n^{-1/2+\varepsilon}$.

The function $A(y|\theta_1, \theta_{00})$ is linear in each component of θ_{00}, and so it is trivial that this function satisfies a Lipschitz condition in those components. This is sufficient to ensure that the Gaussian process $\xi(\cdot)$, which is the weak limit of $Z(\cdot)$, has continuous sample paths [10, pages 148, 150 and 157].

Put $c(\theta_{00}) \equiv 1/|J_2(\theta_1, \theta_{00})|$. If $m = o(n^{2/3})$, then the term $2\gamma n^{-1}m^{3/2}$ appearing in (4.20) is negligibly small. Then for fixed θ_{00}, the right-hand side of (4.20) is asymptotically maximized by taking $\eta = n^{-1/2}Z(\theta_{00})c(\theta_{00})$, and with that substitution, $\hat{J}_m(\theta)$ is asymptotically maximized by choosing θ_{00} to maximize $Z(\theta_{00})^2c(\theta_{00})$. This gives (3.2). If it should be the case that $n^{-1}m^{3/2} \to l(< \infty)$, then the above argument will continue to hold, provided l is so small that $c(\theta_{00}) \equiv |J_2(\theta_1, \theta_{00})| - 2\gamma l > 0$ for all $\theta_{00} \perp \theta_1$. But if l is so large that this function takes negative values, then it is clear from (4.20) that $\hat{J}_m(\theta)$ does not have a local maximum distant order $n^{-1/2}$ from θ_1. In this circumstance, result (3.2) fails.

This completes the proof of Theorem 4.1 □

Acknowledgments. I am grateful to two referees and an Associate Editor for their helpful comments.

REFERENCES

[1] FRIEDMAN, J. H. (1987). Exploratory projection pursuit. *J. Amer. Statist. Assoc.* **82** 249–266.

[2] FRIEDMAN, J. H., STUETZLE, W. and SCHROEDER, A. (1984). Projection pursuit density estimation. *J. Amer. Statist. Assoc.* **79** 599–608.

[3] HALL, P. (1988). Estimating the direction in which a data set is most interesting. *Probab. Theory Rel. Fields* **80** 51–78.

[4] HALL, P. (1989). On projection pursuit regression. *Ann. Statist.* **17** 573–588.

[5] HALL, P. and HEYDE, C. C. (1980). *Martingale Limit Theory and Its Application.* Academic, New York.

[6] HUBER, P. J. (1985). Projection pursuit (with discussion). *Ann. Statist.* **13** 435–525.

[7] JOHNSTONE, I. M. (1987). Discussion of Jones and Sibson [8]. *J. Roy. Statist. Soc. Ser. A* **150** 29–30.

[8] JONES, M. C. and SIBSON, R. (1987). What is projection pursuit? (with discussion). *J. Roy. Statist. Soc. Ser. A* **150** 1–36.

[9] MAGNUS, W., OBERHETTINGER, F. and SONI, R. P. (1966). *Formulas and Theorems for the Special Functions of Mathematical Physics.* Springer, New York.

[10] POLLARD, D. (1984). *Convergence of Stochastic Processes.* Springer, New York.

[11] SANSONE, G. (1959). *Orthogonal Functions.* Wiley, New York.

[12] SZEGÖ, G. (1959). *Orthogonal Polynomials. Amer. Math. Soc. Coll. Publ.* **23.** Amer. Math. Soc., New York.

[13] USPENSKY, J. V. (1927). On the development of arbitrary functions in series of Hermite and Laguerre polynomials. *Ann. of Math.* (2) **28** 593–619.

[14] ZYGMUND, A. (1968). *Trigonometric Series* **1.** Cambridge Univ. Press, London.

DEPARTMENT OF STATISTICS
FACULTY OF ECONOMICS AND COMMERCE
AUSTRALIAN NATIONAL UNIVERSITY
GPO BOX 4
CANBERRA, ACT 2601
AUSTRALIA