The Annals of Statistics
1988, Vol. 16, No. 4, 1345–1368

POLYNOMIAL ESTIMATION OF REGRESSION FUNCTIONS WITH THE SUPREMUM NORM ERROR

BY VÁCLAV FABIAN

Michigan State University

Regression with the error measured by the supremum norm is considered. Analytic functions on $[0, 1]$ and functions with a bounded rth derivative are considered as functions to be estimated. It is assumed that the experimenter chooses the points at which the observations are taken. Polynomial and piecewise polynomial estimates are considered. Asymptotic and nonasymptotic bounds for the error are obtained.

1. Introduction. In this paper we obtain results concerning nonparametric regression functions on $[0, 1]$, with the error measured by the supremum norm $\| \ \|$. As estimates we use polynomial interpolants through observations taken at expanded Chebyshev points; by a polynomial we mean an algebraic polynomial unless specified otherwise.

In numerical analysis and in approximation theory, measuring the error by $\| \ \|$ plays a central role. Kiefer and Wolfowitz (1959) remark on the desirability of using $\| \ \|$ in regression analysis. We will refer to results obtained in this area and toward the end of this introduction we restate the advantages of defining the error by $\| \ \|$.

Approximants studied in approximation theory are polynomials and trigonometric polynomials, and piecewise polynomials with varying smoothness, i.e., splines. For analytic functions, polynomial approximations give very good approximations [cf. (2.5.1)], better than piecewise approximations [see Hart, Cheney, Lawson, Maehly, Mesztenyi, Rice, Thacher and Witzgall (1968), Section 3.5], but for functions that are not smooth enough (e.g., $\sqrt{x}$ on $[0, 1]$), the relation is reversed [see de Boor (1978), end of Chapter II].

The accuracy of polynomial and rational approximations derived for computer use is remarkable [see Hasting (1955) and Hart et al. (1968)]. Often the degree of the approximating polynomial is large compared to the usual choice in polynomial regression. The best polynomial approximations can be (approximately) obtained by the Remes algorithm [see Remes (1934) and, e.g., Hart et al. (1968), Section 3.2]. A recent result by Brutman (1978) makes it possible to obtain easily a polynomial approximant with the error close to minimal error. Such a polynomial is obtained by interpolating the function at expanded Chebyshev points (cf. Section 2.3). The penalty paid for using this suboptimal approximation is small: The error is at most $(1 + \Lambda_k)$ times the minimal error, with k the order of the polynomials considered and Λ_k tabulated in Table 1. Thus, for $k \leq 15$

Received August 1987.

AMS 1980 *subject classifications*. Primary 62G99, 62J99; secondary 62K05.

Key words and phrases. Nonparametric regression, supremum norm, polynomial, interpolation, expanded Chebyshev points.

($k > 15$ have not been much used) the penalty is at most $1 + \Lambda_{15} = 3.27$, rather negligible in situations when the minimal error is quite small. For asymptotic considerations with $k \to \infty$, this is not so good, since $1 + \Lambda_k \geq (2/\pi)\log k + 1.5 \to \infty$.

In (nonparametric) regression analysis, several authors constructed estimates with asymptotically optimal rates. For functions with bounded rth derivatives, Halász (1978) uses piecewise polynomials as estimates and Stone (1982) uses moving polynomials (as in moving averages); Stone allows multidimensional domains of the functions. Ibragimov and Has'minskii (1980, 1981, 1982) consider classes of 2π-periodic functions and estimates that are trigonometric functions—a standard argument replaces an $f(t)$ on $[-1, 1]$ by $f(\cos x)$ and makes the results applicable to nonperiodic functions as well.

The estimates have errors of the order $O((\log n/n)^{r/(2r+1)})$ for functions with bounded rth derivatives and $O((\log n \log_2 n/n)^{1/2})$ for analytic functions. However important from the theoretical point of view, these results are difficult to use in applications because of unknown constants involved in the statement and because of the asymptotic nature of the statement.

A related result is due to Walk (1987), who, improving on previous work by Pantel (1979), adapted a stochastic approximation method, obtained a sequential estimate of the best polynomial approximant and described asymptotic properties of the estimate (not restricted to algebraic polynomials).

Improving the applicability of the estimates by specifying the constants in the order of convergence or strengthening the asymptotic results for the estimates seems difficult for the estimates considered up to now, but we have obtained such a strengthening for the estimates we study here.

We consider here estimates obtained by interpolating the estimates of the function values at expanded Chebyshev points. For our estimate we obtain almost immediately nonasymptotic bounds for the error in Theorem 2.8 with more details and tables in Sections 6 and 7. The polynomial estimates may be preferred in some situations to the less smooth estimates proposed by Halász (1978) and Stone (1982) and to the smooth but somewhat complicated estimates described by Ibragimov and Has'minskii (1980, 1982). In addition, approximations by polynomials are considered important in approximation theory and (with order kept low) also in classical regression theory.

The organization of this paper and the results in more detail are described next:

Section 2 is introductory and reviews some results from approximation theory.

Section 3 gives asymptotic error bounds for polynomial estimates of analytic functions and of functions with bounded rth derivatives. The order k_n of the polynomial estimate when n observations are available is considered and a choice that minimizes the asymptotic bound for the error is obtained in Theorems 3.6 and 3.7. The asymptotic results are of the type $\limsup c_n^{-1}\|Y_n - f\| \leq C$ with numbers c_n and C specified rather than of the type $\|Y_n - f\| = O(c_n)$. The asymptotic considerations confirm that k_n, the order of the polynomial approximant, should be rather small.

The simplicity of the estimate imposes a penalty: The asymptotic bounds are larger by a factor of $C \log k_n$ than the optimal error rates, with C an unknown number. This does not say much about finite n.

The suitable choice of k_n depends on the class to which the estimated f belongs. An adaptive estimate is described in Section 4. Under the assumption that f belongs to a union of a finite number of classes, the estimate selects the best class that contains f and has an asymptotic error bound equal to that of the estimate that uses the knowledge of the class. This result parallels a result by Härdle and Marron (1985) in a related area.

Section 5 describes, for functions with bounded rth derivatives, a piecewise polynomial estimate which achieves the optimal rate. The result is an easy consequence of results from Section 3 and it adds to results by Halász (1978) and Stone (1982) who obtained a bound $O((\log n/n)^{r/(2r+1)})$: we obtain a more specific bound $\limsup(n/\log n)^{r/(2r+1)}\|Y_n - f\| \le C$ for an explicit constant C. This makes possible a comparison with the asymptotic properties of complete cubic spline estimates described in Fabian (1987); see Remark 5.6.

It should be understood that the asymptotic considerations are of limited importance although they are more elegant than the ploddingly produced partial tables of nonasymptotic bounds: they make possible comparisons with other asymptotic results; they indicate that k, the order of the estimating polynomial, should be rather small. That is pleasant since we really would not want to use large k. But the asymptotic bounds use $k \to \infty$ and are very inaccurate for small k. Thus results of Sections 4 and 5 would also need a nonasymptotic reformulation to be applicable (for the result in Section 5 this is partly done in Section 8). It would be possible to rewrite the asymptotic bounds to make them more accurate, but the most accurate are the nonasymptotic bounds discussed in Section 2 and in more detail in Sections 6 and 7.

Sections 6 discusses several bounds for the quantile of $\|Y - f_k\|$, where Y is a polynomial estimate and f_k a polynomial interpolant to f. This is also a simultaneous band considered in least squares regression. A bound N based on the Lebesgue constant is compared with the bound W due to Wynn (1984) and to the bound S of Scheffé (1953). A misleading asymptotic result obtains: Asymptotically, bound N is better than bound S, but bound S is better than N for $k = 3, \dots, 15$ (for $k = 2$ and for k larger than about 50, N is better than S). Bound W is closely connected to bound N, thus also worse than S, for our design and for $k = 3, \dots, 15$. Results by Knafl, Sacks and Ylvisaker (1985) and some preliminary computation by the author make it clear that bound S, in our context, can be improved with practical significance.

Section 7 returns to the nonasymptotic error bounds of Section 2 and gives tables of the quantiles of the error $\|Y - f\|$ for f analytic and Y a polynomial estimate. The tables also show the decomposition of the error $\|Y - f\|$ into several partial errors. For f in $B^{(4)}(1)$, quantiles of the error are obtained for both the polynomial and the piecewise polynomial estimates.

The author believes that the considerations and the results presented are important for applications.

The main difficulty in actually applying the results is in the assumption that f belongs to a certain class of functions. Of course, such a difficulty is present in applications of all statistical and, more generally, mathematical methods; Remark 7.2 is a more detailed discussion of this.

The definition of the error by the supremum norm seems considerably more appropriate in many applications than that of the least squares regression. It leads to a more difficult theoretical problem. Even so, the results we obtain can be used in actual applications. We shall talk about these two points in some more detail.

The definition of the error by the supremum norm seems considerably closer to the needs of applications than other definitions. Thus, in particular, the often used definition of error as the sum of squared deviations at the points of observations completely neglects deviations at the other points, compensates for large errors at some points by small errors at other points and weighs the errors at individual points by the multiplicity of observations. All of these are undesirable in most applications (and unacceptable in numerical analysis). An estimate that minimizes the error may be unsatisfactory from another important point of view. Additional ad hoc rules are used to eliminate such "optimal but undesirable" estimates. In particular, when using a polynomial estimate, increasing the order (i.e., the allowed degree) enlarges the class of the estimates and thus can only diminish the error of the best estimate, yet this is strongly discouraged because of the potentially bad behavior of the estimate, behavior unaffecting the error. Proposals to add to the error a penalty for the complexity or lack of smoothness usually involve a large amount of arbitrariness. It is meaningless to study the effect of the design on the error because the definition of the error changes with the design (of course the effect of the design on, e.g., the estimate of the leading coefficient can be studied).

The definition of the error by the supremum norm has none of the preceding disadvantages except it is more difficult to study.

One result which is qualitative more than quantitative is that a high order (say 10) polynomial estimate gives a useful estimate when a suitable design is used. The asymptotic results give some insight on the behavior of the error and on a suitable choice of the order, and are accompanied in Section 7 by non-asymptotic numerical results. For all these results it is assumed that the estimated function lies in a certain class. This is perhaps unpleasant for applications, but cannot be improved upon. However, the classes considered are much wider than those considered in classical parametric regression theory.

The present results reemphasize the importance of suitable design. It is pleasant that expanded Chebyshev points are suitable in our context since close to them Chebyshev points have been found optimal in a related context of estimating the leading coefficient of a polynomial regression by Kiefer and Wolfowitz (1959). It seems that most practically oriented texts on regression analysis miss completely the overwhelming importance of the design in general and the results by Kiefer and Wolfowitz and related results by other authors in particular.

With the design properly chosen, it is possible (and may be advantageous, see Section 7) to choose the order k of the polynomial rather large. (Practical considerations and the practice of computer approximations seem to favor k not exceeding 10; Section 7 considers $k = 2, \dots, 15$). Of course, disasters occur for large k if the design is not chosen suitably, e.g., if the equidistant design is used.

The asymptotic results and the nonasymptotic results in Section 7 give insight on the bounds obtained. For concrete applications the bound for quantiles of the error can be determined as demonstrated in Section 7.

2. Some approximation theory results.

2.1. *Notation.* A function means a real valued function unless the range is specified differently. By B we mean the family of all bounded functions on $[0,1]$. By $B^{(r)}$, for a positive integer r, we mean the family of all functions f on $[0,1]$ such that the $(r-1)$st derivative $f^{(r-1)}$ exists and is absolutely continuous. If f is in $B^{(r)}$, then, for a set A of (Lebesgue) measure 1, f has a derivative $f^{(r)}(x)$ at x for all x in A. We denote B also by $B^{(0)}$.

By $\|\ \|$ on B we mean the supremum $\|f\| = \sup\{|f(x)|: x \in [0,1]\}$; for an f in $B^{(r)}$ we let $\|f^{(r)}\| = \operatorname{vrai\,sup}|f^{(r)}|$.

For positive M and $r \geq 0$, we set

$$B^{(r)}(M) = \{f: f \in B^{(r)}, \|f^{(r)}\| \leq M\}.$$

$\mathscr{P}_k$ is the family of all polynomials of order k, i.e., of degree at most $k-1$. For $f \in B$, $d_k(f)$ is the distance of f from $\mathscr{P}_k$, i.e., $d_k(f) = \inf\{\|f-g\|; g \in \mathscr{P}_k\}$.

2.2. *Interpolation.* Suppose X_k is a size k subset of $[0,1]$ and for every function f with a domain containing X_k, f_k is the interpolant of f through X_k, i.e., the unique g in $\mathscr{P}_k$ that agrees with f on X_k. Denote by $\mathscr{I}_k$ the function $f \to f_k$. Note that $\mathscr{I}_k$ is linear and idempotent. The Lebesgue constant Λ_k corresponding to X_k is the norm of $\mathscr{I}_k$, i.e.,

$$\Lambda_k = \sup\{\|\mathscr{I}_k f\|; f \in B(1)\}. \tag{2.2.1}$$

An easy result is then [see, e.g., de Boor (1978), Chapter 2, (9)]: For every $f \in B$,

$$\|f - f_k\| \leq (1 + \Lambda_k) d_k(f). \tag{2.2.2}$$

Indeed, for every g in $\mathscr{P}_k$,

$$\|\mathscr{I}_k f - g\| = \|\mathscr{I}_k f - \mathscr{I}_k g\| \leq \Lambda_k \|f - g\|;$$

thus

$$\|f - f_k\| \leq \|f - g\| + \|f_k - g\| \leq (1 + \Lambda_k)\|f - g\|.$$

Taking the infimum with respect to g gives (2.2.2).

Relation (2.2.2) says that f_k is an approximation in $\mathscr{P}_k$ to f which is worse than the best approximation by at most a factor of $1 + \Lambda_k$. It is desirable to have Λ_k small.

TABLE 1
The Lebesgue norm for k expanded Chebyshev points (EC), for the equidistant points (ED) and values m_k

k	Λ_k EC	Λ_k ED	m_k
2	1.0000	1.0000	1.0000
3	1.2500	1.2500	1.0000
4	1.4299	1.6311	1.0070
5	1.5702	2.2078	1.0154
6	1.6851	3.1063	1.0221
7	1.7825	4.5493	1.0272
8	1.8670	6.9297	1.0312
9	1.9416	10.9456	1.0344
10	2.0083	17.8486	1.0370
11	2.0687	29.9000	1.0392
12	2.1239	51.2142	1.0410
13	2.1747	89.3249	1.0426
14	2.2217	158.1023	1.0439
15	2.2655	283.2107	1.0451

2.3. *The expanded Chebyshev points.* For an integer $k > 1$, the expanded Chebyshev points $c_{k1}, c_{k2}, \ldots, c_{kk}$ are given by

$$c_{ki} = \frac{1}{2}\left(1 + \frac{\cos((2i-1)a)}{\cos a}\right), \quad \text{with } a = \frac{\pi}{2k}. \tag{2.3.1}$$

Set $X_{k0} = \{c_{k1}, \ldots, c_{kk}\}$.

If $X_k = X_{k0}$, i.e., when the interpolation is through the expanded Chebyshev points, then, by a result of Brutman (1978),

$$\frac{2}{\pi}\log k + 0.5 \le \Lambda_k < \frac{2}{\pi}\log k + 0.73. \tag{2.3.2}$$

de Boor [(1978), page 27] states that numerical evidence strongly suggests that the choice of expanded Chebyshev points comes within 0.02 of minimizing Λ_k.

For the Chebyshev points, the second inequality in (2.3.2) holds with 0.73 increased to 4; for equidistant division of $[0,1]$, $\Lambda_k >= Ce^{k/2}$ for a constant C [see de Boor (1978), page 26]. For $k = 2, \ldots 15$, Table 1 gives Λ_k for the expanded Chebyshev points and for the equidistant choice. The equidistant choice seems intuitively very appealing, the expanded Chebyshev points do not seem to be so much different from the equidistant points and thus the difference in their Lebesgue constants Λ_k is at first surprising. But the difference explains why interpolation (and similarly least square approximation) with high degree polynomials and without a suitable design, leads to disasters; the situation is different with the expanded Chebyshev points design.

The norm Λ_k can be expressed more explicitly as follows: Denote by l_{ki} the Lagrange polynomials, i.e., l_{ki} is the interpolant through $X_k = \{x_{k1}, \ldots, x_{kk}\}$ of the function which is equal to 1 at x_{ki} and 0 at the other points in X_k. Denote by l_k the vector function with components l_{ki}. If y is the k-tuple of values of f at the points x_{ki}, then $f_k = y^t l_k$ and it follows easily that

$$\Lambda_k = \|\lambda_k\|, \tag{2.3.3}$$

where

$$\lambda_k = \sum_{i=1}^{k} |l_{ki}| \tag{2.3.4}$$

is called the Lebesgue function.

2.4. *Computation of maxima.* Table 1 gives the Lebesgue constants Λ_k for the expanded Chebyshev points, for the equidistant choice and also values of

$$m_k = \max_t \sqrt{\sum_{i=1}^{k} l_{ki}^2(t)} \tag{2.4.1}$$

for the expanded Chebyshev points. The values m_k will be needed in Sections 7 and 8. Powell (1967) computed the Lebesgue constants for the Chebyshev points.

A bound for the error of the values in Table 1 is desirable. An obvious way is to determine a bound for the derivative of the Lebesgue function λ_k and use it to bound the difference between Λ_k and the maximum of $\lambda_k(t_i)$ for $i = 1, \dots, n$ and t_i suitably selected points in [0,1]. Unfortunately, direct calculations of the coefficients of $\sum_{i=1}^{k} l_{ki}$ lead to bounds too large to be useful. However, there is a theoretical bound for a derivative of any polynomial g of order k:

$$|g'(x)| \le (k-1)\|g\| \min\left\{2(k-1), \frac{1}{\sqrt{x(1-x)}}\right\} \tag{2.4.2}$$

[see Timan (1963), Section 4.8, (33)]. The calculation proceeded in two steps. First a lower bound c for Λ_k was determined by finding the maximum at 5001 equidistant points in $[0, 1/2]$. Then λ_k was evaluated at points $x_1, x_2, \dots, x_N$, updating lower bounds to $b_1, b_2, \dots, b_N$ and with the increment $x_{i+1} - x_i$ at point x_i chosen in such a way that $\lambda_k \le \max\{b_i, \lambda_k(x_{i+1})\} + \kappa$ on the interval $[x_i, x_{i+1}]$ with $\kappa = 10^{-4}$ the error allowed.

The same calculations were performed to find m_k. (In each of the two cases, the maximum found in the second step was equal to the lower bound for the maximum found in the first step.)

The values of Λ_k for the equidistant choice are lower bounds, determined as maxima of λ_k over $\{(i-1)/10000;\ i = 1, \dots, 10001\}$.

2.5. *The distance $d_k(f)$.* An $f \in B$ is called analytic if it is a restriction of an analytic function F of a complex variable. In such a case, the domain of F contains the interior of an ellipse with foci 0 and 1 and sum of the axes equal to a number R; for a given number R, $A(R, M)$ denotes the family of all such functions f with $|F|$ bounded by M on the interior of the ellipse. For every f in

$A(R, M)$ and for all $k = 2, 3, \ldots,$

$$(2.5.1) \qquad d_k(f) \le (2\pi)^{1/2} M \frac{R}{R-1} R^{-k}.$$

This follows from the proof in Section 5.4.1 in Timan (1963).

For f in $B^{(r)}(M)$ we obtain, for every $k = r+1, r+2, \ldots,$

$$(2.5.2) \qquad d_k(f) \le (\pi/4)^r M \frac{(k-r)!}{k!}.$$

This is a slight strengthening of Jackson's Theorem V, Part (iii), as formulated in Chenney [(1982), Section 4.6]. For f with a continuous rth derivative, (2.5.2) is obtained from there by setting Chenney's n, k to our $k-1, r$ and then multiplying the bound by 2^{-r}, accounting for the fact that our f is defined on $[0,1]$ and not on $[-1,1]$. If $f^{(r)}$ is not continuous and ε is a positive number, then there is a continuous g_0 in $L_1(\lambda)$ (λ the Lebesgue measure on $[0,1]$) with $\|g_0 - f^{(r)}\|_1 < \varepsilon$ and $\| \ \|_1$ the $L_1(\lambda)$ norm. Take $g = \max\{-M, \min[g_0, M]\}$. Again $\|g - f^{(r)}\| < \varepsilon$. But there is a function G with $G^{(r)} = g$ and $\|G - f\| < \varepsilon$. Since $d_k(G)$ is bounded by the right-hand side of (2.5.2), $d_k(f)$ also must satisfy (2.5.2).

2.6 Example. Consider, as an example of (2.5.1), the function $f(x) = \sqrt{x}$ on $[a, b]$ with $a > 0$. f can be extended to an analytic function F on the interior of the ellipse that has loci at a, b and passes through 0. We obtain the sum of the axes $R_0 = (\sqrt{a} + \sqrt{b})^2$, $M = \sqrt{a+b}$. Since (2.5.1) was stated for f on $[0,1]$, we obtain by a simple argument that for our f, (2.5.1) holds with

$$R = \frac{R_0}{b-a} = \frac{\sqrt{b/a} + 1}{\sqrt{b/a} - 1}, \qquad M = \sqrt{a+b}.$$

For $b = 1$ and $a = 0.1, 0.5$ we give R, M and the right-hand side D_k of (2.5.1) in Table 2.

Table 2
The bounds D_k and d_k^* for $d_k(f)$ for $f(x) = \sqrt{x}$ on $[a, 1]$

	$a = 0.1$	$a = 0.5$
R	1.92495	5.82843
M	1.04881	1.22474
D_2	1.47655	0.10909
D_3	0.76706	0.01872
D_4	0.39848	0.00321
d_2^*	0.07943	0.00759
d_3^*	0.02188	0.00065
d_4^*	0.00741	0.00007

The actual values of $d_k(f)$ are considerably smaller and from properties of the algorithms SQRT 0030, SQRT 0031, SQRT 0032, SQRT 0150, SQRT 0130 and SQRT 0131 in Hart et al. (1968) we obtain much smaller bounds d_k^* for $d_k(f)$. We also give these in Table 2 .

2.7 ASSUMPTION. f is in B, Y is the interpolant through X_k to the function with values $f(x_i) + \varepsilon_i$ for $x_i \in X_k$ and ε_i random variables. $Z = \max_{i=1,\dots,k}|\varepsilon_i|$.

2.8 THEOREM. *Under Assumption* 2.7,

$$\|Y - f\| \le (1 + \Lambda_k)d_k(f) + \|Y - f_k\| \tag{2.8.1}$$

and

$$\|Y - f_k\| \le \Lambda_k Z. \tag{2.8.2}$$

PROOF. If ω is an elementary event, take a function ε in $B(Z(\omega))$ with $\varepsilon(x_i) = \varepsilon_i(\omega)$ and obtain

$$Y(\omega) = \mathscr{I}_k f + \mathscr{I}_k \varepsilon, \qquad |Y(\omega) - f\| \le \|Y(\omega) - f_k\| + \| f_k - f\,|.$$

Apply (2.2.2) to obtain (2.8.1). Relation (2.8.2) holds since $\|Y(\omega) - f_k\| = \|\mathscr{I}_k\varepsilon\| \le \Lambda_k Z(\omega)$. □

3. Polynomial approximation, asymptotic case. We shall consider the case now in which an increasing number n of observations are available, these divided into k groups. Averages of these are used to estimate the k function values. Usually then n/k will be a large number and it is not a strong assumption that the k estimates be unbiased independent normal with variances k/n. This corresponds to the case of the individual observations having variance 1; the general case of variance σ^2 reduces to this special case by rescaling. Estimating σ^2 from the observations will usually be a minor problem compared to the estimation of f.

Thus the following will be assumed in this section.

3.1 ASSUMPTION. For each $n = 1, 2, \dots$ let $k = k_n$ be an integer, $k_n \ge 2$, $k_n \to \infty$ as $n \to \infty$. Let X_k be the set of the k expanded Chebyshev points $x_{ki} = c_{ki}$ defined by (2.3.1) and let $\varepsilon_{n1}, \dots, \varepsilon_{nk}$ be independent normal $(0, k/n)$ random variables. For any function f on $[0, 1]$, consider the estimate Y_n that is the interpolant through X_k of the function with values $f(x_{ki}) + \varepsilon_{ni}$. Set $Z_n = \max_{i=1,\dots,k}|\varepsilon_{ni}|$.

3.2 LEMMA. *With* $k = k_n$,

$$Z_n = n^{-1/2}A_k B_k(1 + V_n/B_k), \tag{3.2.1}$$

where

$$A_k = \left[\frac{k}{2\log k}\right]^{1/2}, \qquad B_k = \log\frac{k^2}{(4\pi\log k)^{1/2}}. \tag{3.2.2}$$

The distribution of the random variables V_n *depends only on* k_n *and, as* $n \to \infty$, V_n *converges in distribution to* $\max\{-\log\eta, -\log\xi\}$, *where* ξ *and* η *are independent gamma*(1) *random variables.*

PROOF. We have

$$Z_n = \max\{\xi_k, -\xi_1\} \tag{3.2.3}$$

with $\xi_k = \max_i \varepsilon_{ni}$, $\xi_1 = \min_i \varepsilon_{ni}$. By Theorem 4.7.22 in Fabian and Hannan (1985), the random variables defined by

$$V_n = a_k(n/k)^{1/2} Z_n - B_k \tag{3.2.4}$$

and

$$a_k = (2\log k)^{1/2} \tag{3.2.5}$$

have the limit in distribution as asserted and it is easy to see that they satisfy (3.2.1). □

3.3 REMARK. Write $X_n \sim Y_n$ if $X_n/Y_n \to 1$ in probability for random variables X_n, Y_n, including also number sequences. Write $X_n \lesssim Y_n$ or $Y_n \gtrsim X_n$ if $X_n + c_n \sim Y_n$ for positive numbers c_n.

For A_k, B_k and V_n in (3.2.1) we obtain

$$1 + \frac{V_n}{B_k} \sim 1 \tag{3.3.1}$$

and

$$Z_n \sim \sqrt{\frac{2k\log k}{n}}. \tag{3.3.2}$$

We shall consider essentially two choices of k_n: In case (i) of $k_n = \alpha_n \log n$ with $\log\alpha_n/\log_2 n \to 0$, we obtain from (2.3.2) and (3.3.2) that

$$\Lambda_k \sim \frac{2}{\pi}\log_2 n, \qquad Z_n \sim \sqrt{\frac{2\alpha_n(\log n)(\log_2 n)}{n}}, \tag{3.3.3}$$

where $\log_2 n$ denotes $\log\log n$.

In case (ii) of $k_n = \alpha_n(n/\log n)^{1/(2r+1)}$ with $\log\alpha_n/\log n \to 0$, we obtain similarly

$$\Lambda_k \sim \frac{2}{\pi(2r+1)}\log n, \qquad Z_n \sim \sqrt{\frac{2\alpha_n}{2r+1}}\left(\frac{\log n}{n}\right)^{r/(2r+1)}. \tag{3.3.4}$$

3.4. *The error bound.* We shall assume that f is in $A(R, M)$ or in $B^{(r)}(M)$ and denote by D_k the corresponding bound for $d_k(f)$, i.e., the right-hand side of (2.5.1) or (2.5.2). By Theorem 2.8, a bound $C(n, k)$ for $\|Y_n - f\|$ is given by

$$C(n,k) = (1 + \Lambda_k)D_k + \Lambda_k Z_n. \tag{3.4.1}$$

We will derive asymptotic properties of $C(n, k)$ and will find asymptotically optimal choices of k_n in the sense of making $C(n, k_n)$ the smallest possible asymptotically. We should keep in mind that $C(n, k)$ is only a bound for the error and not even a sharp bound. Already the bound (2.2.2), when restricted to functions in $B^{(r)}(M)$ or $A(R, M)$, is not sharp. On the other hand, notice that $C(n, k)$ is a common bound for all f in the class considered.

3.5 LEMMA. *Assume $f \in A(R, M)$ for some R and M, α_n are positive numbers,* $\log \alpha_n / \log_2 n \to 0$ *and*

$$c_n = \frac{(2\log_2 n)^{3/2}}{\pi} \sqrt{\frac{\alpha_n \log n}{n}}. \tag{3.5.1}$$

Suppose $\langle n_i \rangle$ is a subsequence of $\langle n \rangle$.

Then: If $k_{n_i} = \alpha_{n_i} \log n_i$ and $\alpha_{n_i} \geq (2\log R)^{-1}$ eventually, then

$$C(n_i, k_{n_i}) \sim c_{n_i}; \tag{3.5.2}$$

if $k_{n_i} \geq \alpha_{n_i} \log n_i$, then

$$C(n_i, k_{n_i}) \gtrsim c_{n_i}; \tag{3.5.3}$$

and, if k and n satisfy $k \leq (a/\log R)\log n$, then

$$C(n, k) \geq \sqrt{2\pi} M \frac{R}{R-1} n^{-a}. \tag{3.5.4}$$

PROOF. It is enough to consider the case $\langle n_i \rangle = \langle n \rangle$. By (3.4.1) we have $C(n, k) \geq D_k$ and (3.5.4) obtains from the expression for D_k, i.e., from the right-hand side of (2.5.1).

For $k_n = \alpha_n \log n$ we obtain

$$\Lambda_{k_n} Z_n \sim c_n \tag{3.5.5}$$

by (3.3.3). For $k_n \geq \alpha_n \log n$ we then obtain $\Lambda_{k_n} Z_n \gtrsim c_n$ by (3.3.3) and the isotoneity of the asymptotic behavior of Λ_k and Z_n [cf. (2.3.2) and (3.3.2)]. This proves (3.5.3). If $\alpha_n \geq 1/(2\log R)$ eventually, we have $D_{k_n} = O(n^{-1/2})$ by (2.5.1). This and (3.5.5) give (3.5.2). □

3.6 THEOREM. *Assume $f \in A(M, R)$ for some M and R. Assume $k_n^* = \alpha_n^* \log n$ with $\alpha_n^* \geq 1/(2\log R)$ eventually and $\alpha_n^* \to 1/(2\log R)$. Then*

$$C(n, k_n^*) \sim \frac{2}{\pi\sqrt{\log R}} (\log_2 n)^{3/2} \sqrt{\frac{\log n}{n}} \tag{3.6.1}$$

and $\langle k_n^* \rangle$ is optimal in the sense that for any other sequence $\langle k_n \rangle$,

$$C(n, k_n^*) \lesssim C(n, k_n). \tag{3.6.2}$$

PROOF. Relation (3.6.1) obtains from Lemma 3.5 applied with $\alpha_n = \alpha_n^*$.

Write $k_n = \alpha_n \log n$. If $\alpha_{n_i} \geq \alpha_{n_i}^*$ or $\alpha_{n_i} \sim \alpha_{n_i}^*$, we obtain (3.6.2) for the subsequence from (3.5.3). If $\alpha_{n_i} \leq a(1/\log R)$ with an $a < 1/2$, then (3.6.2) holds because of (3.5.4). This completes the proof. □

3.7 THEOREM. *Assume $f \in B^{(r)}(M)$ for some r and M and set*

$$\kappa_n = (n/\log n)^{1/(2r+1)}. \tag{3.7.1}$$

Then:

(i) *If α_n are positive numbers such that* $\log \alpha_n/\log n \to 0$ *and* $k_n = \alpha_n \kappa_n$, *then*

$$C(n, k_n) \sim \frac{2 \log n}{\pi(2r+1)} \left[\left(\frac{\pi}{4}\right)^r M\alpha_n^{-r} + \sqrt{\frac{2\alpha_n}{2r+1}} \right] \left(\frac{\log n}{n}\right)^{r/(2r+1)}. \tag{3.7.2}$$

(ii) *If α_n^* are positive, $k_n^* = \alpha_n^* \kappa_n$ and*

$$\alpha_n^* \to \left[2r^2(2r+1)\left(\frac{\pi}{4}\right)^{2r} M^2\right]^{1/(2r+1)}, \tag{3.7.3}$$

then

$$C(n, k_n^*) \sim \frac{2}{\pi}\left(\frac{\pi}{8(2r+1)r^2}\right)^{r/(2r+1)} M^{1/(2r+1)} (\log n) \left(\frac{\log n}{n}\right)^{r/(2r+1)} \tag{3.7.4}$$

and $\langle k_n^* \rangle$ is optimal in the sense that for any other sequence $\langle k_n \rangle$,

$$C(n, k_n^*) \lesssim C(n, k_n). \tag{3.7.5}$$

PROOF. By (2.5.2),

$$D_k = \left(\frac{\pi}{4}\right)^r M \frac{(k-r)!}{k!} \sim \left(\frac{\pi}{4}\right)^r M k^{-r}, \tag{3.7.6}$$

the last relation a consequence of the Stirling formula [see Cramér (1946), the third displayed formula past (12.5.3)]. In case (i), we obtain from (3.3.4) and (3.7.6) that

$$\Lambda_{k_n} \sim \frac{2 \log n}{\pi(2r+1)}, \qquad Z_n \sim \sqrt{\frac{2\alpha_n}{2r+1}} \left(\frac{\log n}{n}\right)^{r/(2r+1)} \tag{3.7.7}$$

and

$$D_{k_n} \sim \left(\frac{\pi}{4}\right)^r M\alpha_n^{-r} \left(\frac{\log n}{n}\right)^{r/(2r+1)} \tag{3.7.8}$$

and (3.7.2) follows from (3.4.1).

Denote by $h(\alpha_n)$ the expression in brackets on the right-hand side of (3.7.2) and by α the right-hand side in (3.7.3). It is not difficult to verify that h has a minimum at α. [$g(\alpha) = A\alpha^{-r} + B\sqrt{\alpha}$ has a minimum

$$A^{1/(2r+1)} [B/(2r+1)]^{2r/(2r+1)} (1 + 2r)$$

at $\alpha = (2rA/B)^{2/(2r+1)}$.] Relation (3.7.4) is then a special case of (3.7.2) with $h(\alpha_n)$ replaced by $h(\alpha)$.

To obtain (3.7.5), we must treat in a special way k_n to which (3.7.2) does not apply because of the restriction $\log \alpha_n / \log n \to 0$. Suppose $k_{n_i} \geq \alpha_{n_i} \kappa_{n_i}$ with $\alpha_{n_i} = \log_2 n_i$; set $k_n^0 = \alpha_n \kappa_n$. Then $C(n_i, k_{n_i}) \geq \Lambda_{k_{n_i}} Z_{n_i}$ and from the asymptotic isotoneity of Λ_k and from (3.7.7) applied to k_n^0, we obtain

$$C(n_i, k_{n_i}) \gtrsim \frac{2\log n}{\pi(2r+1)} \sqrt{\frac{2\log_2 n}{2r+1}} \left(\frac{\log n}{n} \right)^{r/(2r+1)}$$

and (3.7.5) holds for this subsequence. If

$$k_{n_i} = \alpha_{n_i} \kappa_{n_i} \leq (\log_2 n_i) \kappa_{n_i},$$

then (3.7.2) and (3.7.4) imply (3.7.5) for this subsequence. This proves (3.7.5). □

4. Adaptive estimates.

4.1 REMARK. The asymptotically optimal choice of $\langle k_n \rangle$ depends on R for f in $A(R, M)$ in Theorem 3.6 and on r and M for f in $B^{(r)}(M)$ in Theorem 3.7. In a related situation of estimating an f in Θ_r, a set similar to $B^{(r)}(M)$ but with an L_2 type norm, Härdle and Marron (1985) described a method that achieves an optimal rate of convergence for f in Θ_r without knowing the r, but assuming r is in a known finite set. The difference in the norms seems to make the method used by Härdle and Marron inapplicable in our context. However, another construction makes it possible to obtain similar results. This will be pursued in detail only for f in $A(R, M)$, but the result can be extended easily to the situation of Theorem 3.6.

4.2 ASSUMPTIONS. We shall assume that F is a family of functions on $[0, 1]$ and $R_1, R_2, \ldots, R_K$ are numbers such that $1 < R_1 < \cdots < R_K$ and that every f in F satisfies

$$d_k(f) \sim C_f R(k, f)^{-k} \tag{4.2.1}$$

for a constant C_f and numbers $R(k, f)$ in $\{R_1, \ldots, R_K\}$.

We assume that at step n we can use n independent estimates of the function values, with errors independent and standard normal.

4.3. *The adaptive estimate.* Consider a sequence $\langle \eta_n \rangle$ of positive numbers such that

$$\eta_n \to 0, \qquad \log(n\eta_n) \sim \log n \tag{4.3.1}$$

(for example, $\eta_n = 1/\log n$). Form estimates Y_{ni} of the form described in Assumption 2.7 using $m_n = \eta_n n K^{-1}$ observations at k_{ni} expanded Chebyshev points with

$$k_{ni} \sim \frac{\log n}{2\log R_i}, \qquad k_{ni} \geq \frac{\log n}{2\log R_i}. \tag{4.3.2}$$

Select i_n as the largest i for which $i = 1$ or

$$\|Y_{n,i} - Y_{n,i-1}\| \le m_n^{-1/2} \log m_n.$$

Use then $(1 - \eta_n)n$ observations and $k_n = k_{ni_n}$ to construct the estimate Y_n.

Set $r_n(f) = \max_i R(k_{ni}, f)$.

Theorem 4.4 asserts the adaptivity of Y_n in the following sense. If the $R(k, f)$ in (4.2.1) do not depend on k, then the bound (4.4.1) for the error of Y_n is the optimal bound given in Theorem 3.6 [the assumption $f \in A(M, R)$ there is used only to obtain $d_k(f) \lesssim CR^{-k}$]. If $R(k, f)$ depends on k, the same bound (4.4.1) holds, but we cannot claim its optimality.

4.4 THEOREM. *If $f \in F$, then the estimate Y_n defined above satisfies*

$$\|Y_n - f\| \lesssim \frac{2}{\pi\sqrt{\log r_n(f)}} (\log_2 n)^{3/2} \left(\frac{\log n}{n}\right). \tag{4.4.1}$$

PROOF. Note that

$$k_{ni} = q_{ni} \frac{\log m_n}{2 \log R_i}$$

with $q_{ni} \ge 1$ and $q_{ni} \to 1$. Consider a given f and denote by I_n the index i such that $R_i = r_n(f)$. For each $i = 1, \dots, K$,

$$\|Y_{ni} - f\| \ge d_{k_{ni}}(f) - \|Y_{ni} - f_{k_{ni}}\|, \tag{4.4.2}$$

where

$$d_{k_{ni}}(f) \sim C_f m_n^{-[\log R(k_{ni}, f)/(2 \log R_i)] q_{ni}} \tag{4.4.3}$$

by (4.2.1) and

$$\|Y_{ni} - f_{k_{ni}}\| \lesssim m_n^{-1/2} \log m_n \tag{4.4.4}$$

by (2.8.2) and (3.3.3).

From (4.4.2)–(4.4.4), we obtain for a positive ε and every $i = 2, \dots, K$ that

$$\|Y_{v,i} - f\| \gtrsim C_f m_v^{-1/2+\varepsilon} \tag{4.4.5}$$

with $\langle v \rangle$ the subsequence of all n such that $i > I_n$.

On the other hand, by (3.5.2) in Lemma 3.5 [we can replace the requirement $f \in A(M, R)$ there by $d_k(f) \le CR^{-k}$] we obtain for every i and the subsequence w of all n such that $i \le I_n$,

$$m_w^{1/2} (\log m_w)^{-1} \|Y_{w,i} - f\| \to 0. \tag{4.4.6}$$

Relations (4.4.5) and (4.4.6) imply

$$P\{i_n = I\} \to 1. \tag{4.4.7}$$

The result now follows by Lemma 3.5 applied to subsequences for which I_n is constant. The fact that Y_n is constructed differently on the complement of $\{i_n = I_n\}$ and that it uses $(1 - \eta_n)n$ observations can be neglected because of (4.4.7) and because $((1 - \eta_n)n)^{-1/2} \sim n^{-1/2}$. □

5. Approximation by piecewise polynomials.

5.1 REMARK. For f in $B^{(r)}(M)$, Theorem 3.7 gives a bound for the polynomial estimate that is larger by a factor of $\log n$ than the best achievable rate established by Halász (1978) and Stone (1982) for estimates not necessarily polynomial.

We describe a piecewise polynomial estimate and show it has the best achievable rate. This has been done already by Halász (1978) and, for another estimate, by Stone (1982), but our result is more detailed in that it specifies the constant involved in the order $O(\log n/n)^{r/(2r+1)}$; also of some interest is the proof, which shows that the property is an easy consequence of approximation theory results reviewed in Section 2.

5.2. *Notation.* Suppose A is a closed interval of finite positive length. The concepts of expanded Chebyshev points and the interpolant transformation $\mathscr{I}_k$, introduced for the interval $[0,1]$ in Sections 2.2 and 2.3, are easily redefined for the interval A; $\mathscr{I}_k$ so modified will be devoted by $\mathscr{I}_{k,A}$ and we assume the interpolation is through the expanded Chebyshev points for A. In what follows, the order of the polynomial approximation will be denoted by s rather than by k.

For an integer m, set $A_i = [(i-1)/m, i/m]$ for $i = 1, \dots, m$ and abbreviate $\mathscr{I}_{s,A_i}$ to $\mathscr{I}_{si}$.

For f in B and f_{A_i} the restriction of f to A_i denote $\mathscr{I}_{si} f_{A_i}$ simply by $\mathscr{I}_{si} f$ and set $\mathscr{T}_{sm} f$ to be the function that agrees with $\mathscr{I}_{si} f$ on A_i. Since at the endpoints a of A_i we have $(\mathscr{I}_{si} f)(a) = f(a)$, there is no ambiguity in the definition of $\mathscr{T}_{sm}$.

$\mathscr{T}_{sm}$ depends on f at s points in each of the subintervals $A_1, \dots, A_m$. These are $k = sm - m + 1$ points x_{mi} in $[0,1]$.

5.3 ASSUMPTION. Assume s is a positive integer and, for each n, $m = m_n$ is a positive integer, $k = (s-1)m + 1$ and $\varepsilon_{n1}, \dots, \varepsilon_{nk}$ are independent normal $(0, k/n)$ random variables. For an $f \in B$, Y_n is the estimate obtained by applying $\mathscr{T}_{sm}$ to the function with values $f(x_{mi}) + \varepsilon_{ni}$ at x_{mi} for $i = 1, \dots, k$.

5.4 THEOREM. *Suppose M is a positive number, r a positive integer and $\sup_f$ means the supremum over f in $B^{(r)}(M)$. Suppose Assumption* 5.3 *holds and* $s \ge r+1$. *Suppose*

$$k_n = \alpha_n \kappa_n \tag{5.4.1}$$

with $\log \alpha_n / \log n \to 0$ *and* κ_n *as in* (3.7.1). *Then*

$$\sup_f \|Y_n - f\| \lesssim \left[(1+\Lambda_s)\left(\frac{\pi}{4}\right)^r M \frac{(s-r)!}{s!}\left(\frac{s-1}{\alpha_n}\right)^r + \Lambda_s \sqrt{\frac{2\alpha_n}{2r+1}}\right] \times \left(\frac{\log n}{n}\right)^{r/(2r+1)}. \tag{5.4.2}$$

PROOF. Set $Z_n = \max_{i=1,\dots,k}|\varepsilon_{ni}|$. From (3.3.4),

$$Z_n \sim \sqrt{\frac{2\alpha_n}{2r+1}}\left(\frac{\log n}{n}\right)^{r/(2r+1)}. \tag{5.4.3}$$

Relation (2.5.2), when applied to an interval of length m^{-1}, allows a multiplication of the right-hand side by m^{-r}. This and (2.2.2) then give

$$\sup_f \|f - \mathcal{T}_{ms} f\| \le (1 + \Lambda_s)\left(\frac{\pi}{4}\right)^r M \frac{(s-r)!}{s!} m^{-r} \tag{5.4.4}$$

and (2.8.2) gives

$$\sup_f \|\mathcal{T}_{ms} f - Y_n\| \le \Lambda_s Z_n. \tag{5.4.5}$$

Since $m \sim \alpha_n \kappa_n/(s-1)$, (5.4.2) follows from (5.4.4), (5.4.5) and (5.4.3). □

5.5 REMARK. If, in Theorem 5.4, $r = 4$, $s = 5$, then (5.4.2) becomes, with four decimal digit accuracy,

$$\sup_f \|Y_n - f\| \lesssim \left[2.0863 M \alpha_n^{-4} + 0.7402\sqrt{\alpha_n}\right]\left(\frac{\log n}{n}\right)^{4/9}. \tag{5.5.1}$$

The right-hand side is minimized by

$$\alpha_n = 1.9985 M^{2/9}; \tag{5.5.2}$$

for this α_n, (5.5.1) becomes

$$\sup_f \|Y_n - f\| \lesssim 1.1772 M^{1/9}\left(\frac{\log n}{n}\right)^{4/9}. \tag{5.5.3}$$

5.6 REMARK. The asymptotic bound (5.5.3) can be compared to that obtained in Fabian [(1987), Remark 9] for the complete cubic spline estimate T_n:

$$\sup_f \|T_n - f\| \lesssim 2.869 M^{1/9}\left(\frac{\log n}{n}\right)^{4/9}. \tag{5.6.1}$$

Notice that the bound in (5.5.3) is smaller than that in (5.6.1). The same relation holds between the left-hand sides of (5.5.3) and (5.6.1) because in (5.6.1) the estimate is selected from a smaller class of estimates. For the polynomial estimate Y_n we obtain from (3.7.4) with $r = 4$ that

$$\sup_f \|Y_n - f\| \lesssim 0.0462 M^{1/9} \log n \left(\frac{\log n}{n}\right)^{4/9}; \tag{5.6.2}$$

the right-hand side of (5.6.2) is larger than that of (5.5.3) for $n > (1.16)10^{11}$. That is, however, due to the inaccuracy of the asymptotic bounds for finite n; Table 8 gives quite different results.

6. Bounds for $\|Y - f_k\|$.

6.1. *Introduction.* We shall assume Assumption 2.7 with ε_i for $i = 1, \dots, k$ independent normal random variables $(0, 1)$. Remark 6.5 discusses the case $\operatorname{Var} \varepsilon_i \neq 1$.

In Sections 3–5 we used a bound

$$\|Y - Y_k\| \leq \Lambda_k Z_k \tag{6.1.1}$$

[see (2.8.2)]. We obtain easily a bound for the β quantile $K_\beta \|y - f_k\|$:

$$K_\beta \|Y - f_k\| \leq \Lambda_k K_\beta Z_k. \tag{6.1.2}$$

Call this bound N (it is obtained from simple properties of the norm Λ_k).

We shall compare bound N with a bound S due to Scheffé (1953) and a bound W due to Wynn (1985). We shall mention some results on obtaining not bounds, but the quantile $K_\beta \|Y - f_k\|$ by numerical methods.

We shall find that, for large k, bound N is considerably better than bound S, but the reverse relation is true for all $k > 2$ and likely to be used. We shall find that bound W is close and related to bound N.

Results by Knafl, Sacks and Ylvisaker (1985) and some other limited computation show that bound S overestimates the quantile $K_\beta \|Y - f_k\|$. Notice that

$$Y - f_k = \sum_{i=1}^{k} l_i \varepsilon_i = l' \varepsilon. \tag{6.1.3}$$

6.2. *Bound S.* The estimate was originally proposed for another use, but applied in our context, it provides a conservative bound.

Set

$$z = K_\beta \chi(k), \tag{6.2.1}$$

i.e., set z equal to the β quantile of the square root of a χ^2 random variable with k degrees of freedom.

z gives an upper bound for the β quantile of $s^{-1} \|(Y_k - f_k)\|$, where $s(t)$ is the standard deviation of $Y_k(t)$. In our case, $s^2 = \sum_{i=1}^{k} l_i^2$. Table 1 gives values of m_k, the maximum of s, for $k = 2, \dots, 15$.

It follows that $z m_k$ is a bound for $K_\beta \|Y_k - f_k\|$ and we shall consider this weaker assertion since we are interested in the $\| \ \|$ norm. We might also keep the original stronger estimate. However, we should note that, with exception for $k = 2$, the values $s(t)$ vary considerably with t (see Figure 1) and replacing them by the maximum may be an appealing modification even if the norm $\| \ \|$ is not of exclusive interest.

As $k \to \infty$, it follows from the asymptotic behavior of the $\chi^2(k)$ distribution that bound S satisfies $S \gtrsim \sqrt{k}$. On the other hand, $N \sim (1/\pi)\sqrt{2}(\log k)^{3/2}$ by (2.3.2) and Lemma 3.2 [notice that $\operatorname{Var}(\varepsilon_i)$ is 1 here and k/n there]. Thus, for k large, N is a better bound than S, but for $k = 3, \dots, 15$ (in fact for k up to about 50) the relation is reversed; see Table 3.

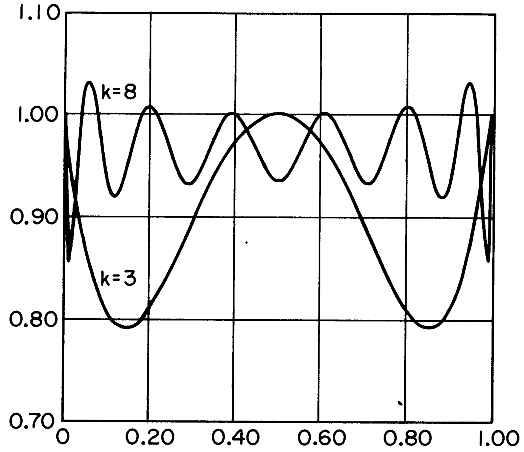

FIG. 1. *The standard deviation s for $k = 3$ and $k = 8$.*

6.3. *Bound W.* In (6.1.3), replacing ε_i by the maximum Z_k of their absolute values gives a bound

$$P\left\{|Y_k - f_k|(t) \le (K_\beta Z_k) \sum_{i=1}^{k} |l_i(t)|, \forall t\right\} \ge \beta. \tag{6.3.1}$$

This is bound W in our context. Changing it by making the bound a constant and with similar justifications (see Figure 2) as for bound S, we replace $\sum_{i=1}^{k}|l_i|$ by its maximum Λ_k. But this gives bound N.

Wynn addresses a more general problem than that of interpolation, but shows that the more general problem can be reinterpreted as the special problem.

6.4. *Exact bounds.* Better bounds than bound S might be obtained, but not easily. The method developed by Knafl, Sacks and Ylvisaker (1985) might be used, but their numerical results are not applicable in our context because of a difference in the design. On the other hand, their numerical results indicate that

TABLE 3
The bounds S and N

	$\beta = 0.95$		$\beta = 0.99$		$\beta = 0.999$	
	S	N	S	N	S	N
2	2.448	2.236	3.035	2.806	3.717	3.481
3	2.795	2.985	3.368	3.668	4.033	4.485
5	3.378	4.033	3.944	4.850	4.599	5.839
10	4.437	5.623	4.996	6.606	5.641	7.813
15	5.225	6.633	5.779	7.706	6.417	9.034

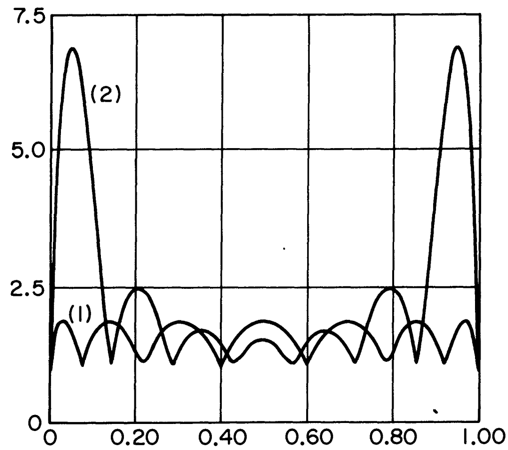

FIG. 2. *The Lebesgue function* λ_8 *for* (1) *expanded Chebyshev knots*, (2) *equidistant knots.*

bound S can be improved; that is also indicated by the asymptotic relation of N to S (cf. Section 6.2).

The author computed $K_\beta \|Y - f_k\|$, approximately, by direct numerical integration of the distribution function for several values of β and for $k = 2, 3, 4$. These are given in Table 4 and compared with bound S. The computation was performed as follows: For $k \leq 4$ an explicit expression is possible and was used to determine the maximum on $[0, 1]$ of $\sum_{i=1}^{k} l_i \varepsilon_i$ for any given ε. Then numerical integration with respect to ε_i gave the distribution function of $\|Y_k - f\|$ and from there the quantiles were determined. The precision seems to be that of the last digit displayed.

The difficulty in direct computation for $k > 4$ is that numerical integration is more difficult or impossible and, perhaps more importantly, the maximum on $[0, 1]$ of $\sum_{i=1}^{k} l_i \varepsilon_i$, with ε_i given, is difficult to determine. (The method described in Section 2.4 might be used to determine the maximum.)

6.5 REMARK. The case of a known variance σ^2 of the ε_i is treated easily by premultiplying by σ the bounds already obtained for the special case of variance 1.

TABLE 4
The β *quantiles* K *of* $\|Y - f_k\|$ *for* $\sigma = 1$, *compared with bound* S

	$k = 3$		$k = 4$	
β	K	S	K	S
0.95	2.4	2.80	2.6	3.10
0.99	3.0	3.37	3.2	3.67

If σ^2 is unknown and V is a $\sigma^2\chi^2(f)$ random variable independent of ε, then bound S can be again used, with z given by

$$z = \left(k\frac{V}{f}K_{1-\alpha}F(k, f)\right)^{1/2} \tag{6.5.1}$$

instead of by (4.3.1).

7. Nonasymptotic considerations.

7.1. *Introduction.* We shall discuss here the nonasymptotic properties of the polynomial estimates. Extensions to piecewise polynomials are almost immediate.

Suppose Assumption 2.7 holds with ε_i independent normal $(0, \sigma^2)$. To reduce the complexity of the tabulation, we assume σ known.

Under Assumption 2.7, Theorem 2.8 yields the following bound for the β quantile $K_\beta\|Y_k - f\|$ of the error $\|Y_k - f\|$ of the kth order polynomial estimate Y_k:

$$K_\beta\|Y_k - f\| \le (1 + \Lambda_k)d_k(f) + B_k \tag{7.1.1}$$

with B_k a bound for $K_\beta\|Y_k - f_k\|$. We use bound N for $k = 2$ and the conservative but useful bound S for $k > 2$.

If f is known to be in a class of functions with a known bound D_k for $d_k(f)$ (as discussed in Section 2) we obtain a bound for $K_\beta\|Y_k - f\|$, independent of f.

A bound D_k may overestimate $d_k(f)$ considerably as we have shown in Example 2.6, yet the bounds we discussed do not admit easy improvements.

In the next section we discuss situations in which k is chosen and a (temporary) assumption is made that $d_k(f)$ is negligible. In Section 7.3 we discuss a more conservative approach.

7.2. *Case of k chosen, $d_k(f)$ negligible.* It seems that in some practical problems, one may be justified to select a k (say 5 or 8) and assume—temporarily at least—that $d_k(f)$ is practically negligible. This can be done on the basis of bounds for $d_k(f)$ described in Section 2, but often also by analogies with many functions $\tilde{f}$ for which $d_k(\tilde{f})$ has been approximately determined as in Hart et al. (1968).

This would be more justified in applications in which the estimate is quickly put into use that further checks the accuracy of the estimate. (On the other side of the spectrum, guessing k is also harmless if the estimate is never put into use.)

In other situations it may be that the interpolant f_k of f itself is of a primary concern as a smoother and simpler approximation to a possibly more complicated f.

Every statistical method requires, in applications, an assumption of varying strength, e.g., independence, normality, zero interactions and so on. Regression applications of the classical theory often assume that the regression function is or is approximately a polynomial of order 2 or 3.

TABLE 5

The relative cost of estimating c1 = k: *one of k means*; c2: *simultaneously k means*; c3: *the interpolating polynomial*; c: *overall cost of estimating the interpolating polynomial*

	β = 0.95			β = 0.99			β = 0.999		
k	*c2*	*c3*	*c*	*c2*	*c3*	*c*	*c2*	*c3*	*c*
2	1.30	1.00	2.60	1.19	1.00	2.37	1.12	1.00	2.24
3	1.48	1.37	6.10	1.30	1.32	5.13	1.19	1.26	4.51
4	1.62	1.55	10.02	1.38	1.47	8.12	1.24	1.40	6.92
5	1.72	1.73	14.86	1.44	1.63	11.72	1.28	1.53	9.77
6	1.80	1.90	20.55	1.49	1.78	15.88	1.31	1.66	13.00
7	1.87	2.06	27.05	1.53	1.92	20.57	1.34	1.77	16.59
8	1.94	2.22	34.34	1.57	2.05	25.76	1.36	1.89	20.53
9	1.99	2.37	42.41	1.60	2.18	31.45	1.38	2.00	24.79
10	2.04	2.51	51.25	1.63	2.31	37.62	1.40	2.10	29.39
11	2.09	2.65	60.84	1.66	2.43	44.27	1.41	2.20	34.30
12	2.13	2.79	71.18	1.68	2.55	51.38	1.43	2.30	39.53
13	2.16	2.92	82.26	1.70	2.66	58.97	1.44	2.40	45.06
14	2.20	3.06	94.06	1.72	2.78	67.01	1.46	2.50	50.90
15	2.23	3.18	106.61	1.74	2.89	75.51	1.47	2.59	57.04

Suppose now k is chosen and $d_k(f)$ is negligible. Then an approximate bound for $K_\beta\|Y_k - f\|$ is B_k [see (7.1.1)] with B_k/σ independent of σ.

Consider n independent normal observations available, with the standard deviation σ. If all are used to estimate one value, the β quantile of the error is $d_0 = \sigma\varphi_{1-\alpha/2}/\sqrt{n}$ with $\alpha = 1 - \beta$ and φ_γ the standard normal γ quantile. If n/k observations are used to estimate by Y_i the value $f(x_i)$, the β quantile $K_\beta|Y_i - f(x_i)|$ of the error is $d_1 = d_0\sqrt{k}$. The β quantile of the maximum error, $K_\beta \max_{i=1,\dots,k}|Y_i - f(x_i)|$ is $d_2 = d_1\varphi_{\beta_0}/\varphi_{1-\alpha/2}$ with $\beta_0 = (1 + \beta^{1/k})/2$. Finally $d_3 = B_k$ is a bound for $K_\beta\|Y_k - f_k\|$.

We can keep d_i equal to d_{i-1} if we increase the number n of observations by a factor of $c_i = (d_i/d_{i-1})^2$. These relative costs are tabulated in Table 5 including the cost $c = (d_3/d_0)^2$. It is apparent from Table 5 that the most costly step is that from estimating one value to estimating k values. For $\beta = 0.99$ and $k = 10$, that step has cost 10. The next step has only cost 1.63 and the last step only costs 2.31. The overall cost is 37.62; it is 37.62 times more expensive to estimate the interpolating polynomial f_{10} than to estimate one value. Table 5 shows such costs for other α and other k.

7.3. *The overall error for f in $A(R, M)$.* Sometimes it may be known, from the theory of the applied field (e.g., from chemical kinetics) that the estimated function is in a class $A(R, M)$ [or in a class $B^{(r)}(M)$] or such an assumption may be accepted as a weakening of the assumption that $d_k(f)$ is negligible.

We shall consider here f in $A(R, M)$ and polynomial estimates. By rescaling we may reduce the considerations to the case $M = 1$.

TABLE 6
Upper bound for the error $K_\beta \|Y_k - f\|$ of Y_k for f in $A(R, 1)$ and $\beta = 0.95$

		$R = 2$			$R = 4$	
k e:	0.05	0.01	0.001	0.05	0.01	0.001
2	2.5873	2.5228	2.5082	0.4985	0.4339	0.4194
3	1.5335	1.4347	1.4124	0.2410	0.1422	0.1200
4	0.9196	0.7930	0.7645	0.1900	0.0634	0.0349
5	0.5954	0.4412	0.4065	0.2011	0.0469	0.0122
6	0.4370	0.2557	0.2149	0.2288	0.0475	0.0067
7	0.3690	0.1610	0.1142	0.2606	0.0526	0.0058
8	0.3492	0.1147	0.0620	0.2932	0.0587	0.0060
9	0.3544	0.0939	0.0353	0.3257	0.0652	0.0066
10	0.3727	0.0863	0.0219	0.3579	0.0716	0.0072
11	0.3975	0.0855	0.0153	0.3900	0.0780	0.0078
12	0.4257	0.0882	0.0123	0.4218	0.0844	0.0084
13	0.4554	0.0926	0.0110	0.4535	0.0907	0.0091
14	0.4859	0.0980	0.0107	0.4849	0.0970	0.0097
15	0.5167	0.1037	0.0108	0.5162	0.1032	0.0103

		$R = 6$			$R = 8$	
k e:	0.05	0.01	0.001	0.05	0.01	0.001
2	0.2478	0.1832	0.1687	0.1702	0.1057	0.0911
3	0.1549	0.0560	0.0338	0.1361	0.0373	0.0151
4	0.1639	0.0373	0.0088	0.1600	0.0334	0.0049
5	0.1937	0.0395	0.0048	0.1929	0.0388	0.0041
6	0.2268	0.0455	0.0047	0.2267	0.0454	0.0046
7	0.2601	0.0520	0.0052	0.2600	0.0520	0.0052
8	0.2930	0.0586	0.0059	0.2930	0.0586	0.0059
9	0.3256	0.0651	0.0065	0.3256	0.0651	0.0065
10	0.3579	0.0716	0.0072	0.3579	0.0716	0.0072
11	0.3900	0.0780	0.0078	0.3900	0.0780	0.0078
12	0.4218	0.0844	0.0084	0.4218	0.0844	0.0084
13	0.4535	0.0907	0.0091	0.4535	0.0907	0.0091
14	0.4849	0.0970	0.0097	0.4849	0.0970	0.0097
15	0.5162	0.1032	0.0103	0.5162	0.1032	0.0103

In Tables 6 and 7, the values depend on σ; we do not label them by σ, however, but by the half-width e of the β interval estimate of the mean based on a normal random variable with variance σ^2, i.e., we compare all other error bounds to a corresponding bound one would get when estimating one function value, instead of k.

Table 6 shows the dependence of the error bound on R, k and e for $\beta = 0.95$. Table 7 shows the dependence of the error bound on R, e and β in case k is selected to minimize the bound.

Table 8 shows bounds for the 0.95 quantile of the error for polynomial estimates and for the piecewise polynomial estimate described in Section 5 for f

TABLE 7
Upper bound for the error $K_\beta \|Y_k - f\|$ of the estimate Y_k for f in $A(R, 1)$ and optimal k

		$R = 2$		$R = 4$		$R = 6$		$R = 8$	
e	β	k	Bound	k	Bound	k	Bound	k	Bound
0.050	0.950	8	0.3492	4	0.1900	3	0.1549	3	0.1361
0.050	0.990	9	0.3092	4	0.1742	3	0.1446	3	0.1258
0.050	0.999	9	0.2778	4	0.1632	4	0.1372	3	0.1187
0.010	0.950	11	0.0855	5	0.0469	4	0.0373	4	0.0334
0.010	0.990	11	0.0740	6	0.0420	4	0.0341	4	0.0302
0.010	0.999	11	0.0661	6	0.0382	4	0.0319	4	0.0280
0.001	0.950	14	0.0107	7	0.0058	6	0.0047	5	0.0041
0.001	0.990	14	0.0092	7	0.0051	6	0.0042	5	0.0036
0.001	0.999	15	0.0081	7	0.0046	6	0.0038	5	0.0034

TABLE 8
Upper bounds for the 0.95 quantile of the error for polynomial and piecewise polynomial estimates for f in $B^{(4)}(1)$; optimal k and m

	Polynomial		Piecewise Polynomial	
e	k	Bound	m	Bound
0.0500	5	0.238229	1	0.238229
0.0100	5	0.054165	1	0.054165
0.0010	7	0.007716	2	0.007156
0.0001	10	0.001134	3	0.000933

in $B^{(4)}(1)$ and for different e with the same meaning as in Tables 6 and 7. Bound N was used. The k and m were chosen optimally, $s = 5$.

REFERENCES

BRUTMAN, L. (1978). On the Lebesgue function for polynomial interpolation. *SIAM J. Numer. Anal.* **15** 694–704.

CHENNEY, E. W. (1982). *Introduction to Approximation Theory*. Chelsea, New York.

DE BOOR, C. (1978). *A Practical Guide to Splines*. Springer, New York.

FABIAN, V. (1987). Complete cubic spline estimation of non-parametric regression functions. Dept. Statistics and Probability, Michigan State Univ.

FABIAN, V. and HANNAN, J. (1985). *Introduction to Probability and Mathematical Statistics*. Wiley, New York.

HALÁSZ, G. (1978). Statistical interpolation. In *Fourier Analysis and Approximation Theory* (G. Alexits and P. Turán, eds.) **1** 403–410. North-Holland, Amsterdam.

HÄRDLE, W. and MARRON, J. S. (1985). Optimal bandwidth selection in nonparametric regression function estimation. *Ann. Statist.* **13** 1465–1481.

HART, J. F., CHENEY, E. W., LAWSON, C. L., MAEHLY, H. J., MESZTENYI, C. K., RICE, J. R., THACHER, H. G., JR. and WITZGALL, C. (1968). *Computer Approximations*. Wiley, New

York.

HASTINGS, C., JR. (assisted by J. T. Hayward and J. P. Wong, Jr.) (1955). *Approximations for Digital Computers*. Princeton Univ. Press, Princeton, N.J.

IBRAGIMOV, I. A. and HAS'MINSKII, R. Z. (1980). On nonparametric estimation of regression. *Soviet Math. Dokl.* **21** 810–814.

IBRAGIMOV, I. A. and HAS'MINSKII, R. Z. (1981). Asymptotic bounds on the quality of the nonparametric regression estimation in L_p. *Zap. Naučn. Sem. Leningrad. Otdel. Mat. Inst. Steklov. (LOMI)* **97** 88–101. [Translated as *J. Soviet Math.* **24** 540–550 (1984).]

IBRAGIMOV, I. A. and HAS'MINSKII, R. Z. (1982). Bounds for the risks of nonparametric estimates of the regression. *Theory Probab. Appl.* **27** 84–99.

KIEFER, J. (1980). Designs for extrapolation when bias is present. In *Multivariate Analysis V* (P. R. Krishnaiah, ed.) 79–93. North-Holland, Amsterdam.

KIEFER, J. and WOLFOWITZ, J. (1959). Optimum designs in regression problems. *Ann. Math. Statist.* **30** 271–294.

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

PANTEL, M. (1979). Adaptive Verfahren der Stochastischen Approximation. Dissertation, Universität Essen.

POWELL, M. J. D. (1967). On the maximum errors of polynomial approximations defined by interpolation and by least squares criteria. *Comput. J.* **9** 404–407.

REMES, E. JA. (1934). Sur le calcul effectif des polynomes d'approximation de Tchebichef. *C. R. Acad. Sci. Paris* **199** 337–340.

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

STONE, C. J. (1982). Optimal global rates of convergence for nonparametric regression. *Ann. Statist.* **10** 1040–1053.

TIMAN, A. F. (1963). *Theory of Approximation of Functions of a Real Variable*. Macmillan, New York.

WALK, H. (1987). A stochastic Remes algorithm. *J. Approx. Theory* **49** 79–92.

WYNN, H. P. (1984). An exact confidence band for one-dimensional polynomial regression. *Biometrika* **71** 375–379.

DEPARTMENT OF STATISTICS AND PROBABILITY
WELLS HALL
MICHIGAN STATE UNIVERSITY
EAST LANSING, MICHIGAN 48824-1027