NOTES

A SIMPLE SOLUTION FOR OPTIMAL CHEBYSHEV REGRESSION EXTRAPOLATION[1]

By Paul G. Hoel

University of California, Los Angeles

1. Summary. A simplified solution is presented for the problem of finding a set of points and corresponding weights that will minimize the variance of the estimated value of a Chebyshev regression function at a point outside the interval of observations. This problem, among others, was solved by Kiefer and Wolfowitz [3] by means of game-theoretic methods. The solution here is based on a simple theorem in [2] and well known properties of Chebyshev systems of functions.

2. Introduction. Let $f_0(x), f_1(x), \cdots, f_k(x)$ be a set of continuous functions that are linearly independent and constitute a Chebyshev system on the two intervals $[-1, 1]$ and $[-1, t]$, where $t > 1$. Let

$$E[y(x)] = \beta_0 f_0(x) + \cdots + \beta_k f_k(x) \tag{1}$$

denote a regression function whose value is to be estimated at the point $x = t$ by means of observations $y_1, y_2, \cdots, y_n$ to be taken at the points $x_1, x_2, \cdots, x_n$ in the interval $[-1, 1]$. It is assumed that the y's are uncorrelated random variables possessing a common variance σ^2. The x's need not be distinct.

The technique introduced in [3] to reduce the preceding estimation problem to one of estimating a single regression coefficient will be employed here; therefore consider the problem of estimating the coefficient β_k in an optimum manner. It is assumed that $f_k(t) \neq 0$, otherwise the coefficient of a non-vanishing $f_j(t)$ is selected instead. Following the notation of [2], let

$$g_k(x) = f_k(x) - \textstyle\sum_{j=0}^{k-1} h_j f_j(x), \tag{2}$$

where the h's are numbers such that $g_k(x)$ is orthogonal to $f_0(x), \cdots, f_{k-1}(x)$ with respect to the x's. That is,

$$\textstyle\sum_{i=1}^{n} g_k(x_i) f_j(x_i) = 0. \qquad j < k. \tag{3}$$

Then (1) may be expressed in the form

$$E[y(x)] = \alpha_0 f_0(x) + \cdots + \alpha_{k-1} f_{k-1}(x) + \beta_k g_k(x).$$

Because of the orthogonality property in (3), the least squares estimate of the

Received 26 July 1965.

[1] This research was supported by the Air Force Office of Scientific Research Grant No. AF-AFOSR-459-63.

parameter β_k, denoted by $\hat{\beta}_k$, is given by the formula

$$\hat{\beta}_k = \sum_{i=1}^n g_k(x_i)y_i / \sum_{i=1}^n g_k^2(x_i).$$

The variance of $\hat{\beta}_k$ then becomes

$$V(\hat{\beta}_k) = \sigma^2 / \sum_{i=1}^n g_k^2(x_i). \tag{4}$$

Now consider the vectors $(f_j(x_1), \cdots, f_j(x_n)), j = 0, 1, \cdots, k-1$, and the vector $(g_k(x_1), \cdots, g_k(x_n))$ in n-space. Since g_k has the form (2) and is orthogonal to $f_0, \cdots, f_{k-1}$, the quantity $\sum_{i=1}^n g_k^2(x_i)$ represents the square of the distance of the vector $(f_k(x_1), \cdots, f_k(x_n))$ from the linear space spanned by the vectors $(f_j(x_1), \cdots, f_j(x_n)), j = 0, 1, \cdots, k-1$. Symbolically, this means that

$$\sum_{i=1}^n g_k^2(x_i) = \min_c \sum_{i=1}^n [f_k(x_i) - \sum_{j=0}^{k-1} c_j f_j(x_i)]^2. \tag{5}$$

Thus, (4) will be minimized if (5) is maximized. An optimum design for estimating β_k may therefore be defined as a discrete probability measure ξ^* which satisfies

$$\min_c \int [f_k(x) - \sum_{j=0}^{k-1} c_j f_j(x)]^2 \xi^*(dx)$$
$$= \max_\xi \min_c \int [f_k(x) - \sum_{j=0}^{k-1} c_j f_j(x)]^2 \xi(dx).$$

Here ξ denotes any discrete probability measure on the interval $[-1, 1]$.

Next, let $c_j^*, j = 0, \cdots, k-1$, denote a set of coefficients that yields a best Chebyshev approximation of $f_k(x)$ in the interval $[-1, 1]$ by means of the functions $f_0(x), \cdots, f_{k-1}(x)$. That is, the c_j^* satisfy

$$\max_x |f_k(x) - \sum_{j=0}^{k-1} c_j^* f_j(x)| = \min_c \max_x |f_k(x) - \sum_{j=0}^{k-1} c_j f_j(x)|. \tag{6}$$

In terms of the preceding notation, the theorem of [2] that will be used here may be stated as follows:

THEOREM 1. *If c^* is Chebyshev and $\xi(B(c^*)) = 1$, where $B(c^*)$ denotes the set of points where the left side of* (6) *is attained, and if*

$$\int [f_k(x) - \sum_{j=0}^{k-1} c_j^* f_j(x)] f_i(x) \xi(dx) = 0 \tag{7}$$

for $i < k$, then ξ is optimum.

This theorem is very easy to demonstrate but the demonstration will not be duplicated here.

3. Construction of an optimum design. Assume that the functions $f_0(x)$, $f_1(x), \cdots, f_{k-1}(x)$ form a Chebyshev system on $[-1, 1]$. It is known [1] that there exists a unique best Chebyshev approximation of $f_k(x)$ by means of $f_0(x)$, $\cdots, f_{k-1}(x)$ and that it is characterized by the fact that the difference

$$f_k(x) - \sum_{j=0}^{k-1} c_j^* f_j(x) \tag{8}$$

assumes its maximum absolute value at no less than $k + 1$ points in the interval $[-1, 1]$, the sign of this difference at consecutive points being taken alternately plus and minus. But it is not possible for this difference to assume its maximum absolute value at more than $k + 1$ points with such alternations of sign because if it did the function given by (8) would then, because of continuity, possess at least $k + 1$ zeros. However, no linear combination of the $k + 1$ functions of a Chebyshev system can possess more than k zeros, and it was assumed originally that $f_0(x), \cdots, f_k(x)$ formed such a system.

Let $z_0 < z_1 < \cdots < z_k$ denote the $k + 1$ points where the difference (8) assumes its maximum absolute value. Then Condition (7) for finding an optimum measure reduces to finding a set of positive weights ξ_r that will satisfy the equations

$$\sum_{r=0}^{k} [f_k(z_r) - \sum_{j=0}^{k-1} c_j^* f_j(z_r)] f_i(z_r) \xi_r = 0, \qquad i < k.$$

But since the quantity in brackets is, except possibly for sign, the maximum value of the difference (8), these equations reduce to

$$\sum_{r=0}^{k} (-1)^r f_i(z_r) \xi_r = 0, \qquad i < k. \tag{9}$$

The solution to the problem of the optimum estmation of β_k therefore consists in finding the points $z_0, \cdots, z_k$ corresponding to the best Chebyshev approximation in (8) and then finding a set of weights ξ_r satisfying (9).

Now return to the original problem of finding an optimum design for estimating the regression function value $\beta_0 f_0(t) + \cdots + \beta_k f_k(t)$. As in [3], introduce the following quantities:

$$\begin{aligned} \gamma_i &= \beta_i, & i < k \\ \gamma_k &= \sum_{i=0}^{k} \beta_i f_i(t)/f_k(t) & \\ h_i(x) &= f_i(x) - [f_i(t)/f_k(t)] f_k(x), & i < k \\ h_k(x) &= f_k(x). & \end{aligned} \tag{10}$$

It will be observed that

$$\sum_{i=0}^{k} \gamma_i h_i(x) = \sum_{i=0}^{k} \beta_i f_i(x). \tag{11}$$

Consequently, the problem of estimating $\sum_{i=0}^{k} \beta_i f_i(t)$ when the regression function is the right side of (11) is equivalent to the problem of estimating γ_k when the regression function is the left side of (11). This means that the problem of the optimum estimation of the regression value $\sum_{i=0}^{k} \beta_i f_i(t)$ can be reduced to the corresponding problem of the optimum estimation of γ_k for the regression function $\sum_{i=0}^{k} \gamma_i h_i(x)$. But the solution of the latter problem is available in (8) and (9) with h's replacing the f's, provided it can be shown that the h's satisfy the necessary properties of the f's needed for the preceding theory to hold.

To demonstrate that the functions $h_0(x), \cdots, h_{k-1}(x)$ constitute a Chebyshev system on $[-1, 1]$ it suffices to show that the determinant $|h_j(x_i)|$, $-1 \leqq x_0 < x_1 < \cdots < x_{k-1} \leqq 1$, $j = 0, 1, \cdots, k - 1$, is non-zero for all such sets of x's. In this connection, consider the expansion of the following determinant, which may be taken to have a positive value, because of the Chebyshev assumptions on the f's.

$$\begin{vmatrix} f_0(x_0) & f_0(x_1) & \cdots & f_0(x_{k-1}) & f_0(t) \\ f_1(x_0) & f_1(x_1) & \cdots & f_1(x_{k-1}) & f_1(t) \\ \vdots & \vdots & & \vdots & \vdots \\ f_{k-1}(x_0) & f_{k-1}(x_1) & \cdots & f_{k-1}(x_{k-1}) & f_{k-1}(t) \\ f_k(x_0) & f_k(x_1) & \cdots & f_k(x_{k-1}) & f_k(t) \end{vmatrix}$$

If the last column is multiplied by $f_k(x_i)/f_k(t)$ and subtracted from the ith column for each $i < k$, it will be found that this determinant reduces to

$$\begin{vmatrix} h_0(x_0) & h_0(x_1) & \cdots & h_0(x_{k-1}) & f_0(t) \\ h_1(x_0) & h_1(x_1) & \cdots & h_1(x_{k-1}) & f_1(t) \\ \vdots & \vdots & & \vdots & \vdots \\ h_{k-1}(x_0) & h_{k-1}(x_1) & \cdots & h_{k-1}(x_{k-1}) & f_{k-1}(t) \\ 0 & 0 & \cdots & 0 & f_k(t) \end{vmatrix}$$

Since this equals $f_k(t)|h_j(x_i)|$ and $f_k(t)$ has a non-zero value, it follows that the functions $h_0(x), \cdots, h_{k-1}(x)$ do form a Chebyshev system. The same type of technique shows that the functions $h_0(x), \cdots, h_k(x)$ form a Chebyshev system over both intervals $[-1, 1]$ and $[-1, t]$.

From the theory related to (8), it follows that there exists a unique set of coefficients c_j yielding a best Chebyshev approximation to $h_k(x)$ by means of $h_0(x), \cdots, h_{k-1}(x)$ and that

$$h_k(x) - \textstyle\sum_{j=0}^{k-1} c_j h_j(x) \tag{12}$$

assumes its maximum absolute value at exactly $k + 1$ points with alternating signs for this difference. But from (10) it follows that (12) is equivalent to

$$c[f_k(x) - \textstyle\sum_{j=0}^{k-1} d_j f_j(x)], \tag{13}$$

where $d_j = c_j/c$ and $c = 1 + \sum_{j=0}^{k-1} c_j f_j(t)/f_k(t)$. Since (13) will possess the same properties as (12), it follows that these two functions will have the same $k + 1$ points where their maximum absolute values are assumed with alternating signs and that these are the Chebyshev points for the best Chebyshev approximation of (8).

When $f_i(z_r)$ is replaced by $h_i(z_r)$ in (9), those equations become

$$\textstyle\sum_{r=0}^{k} (-1)^r [f_i(z_r) - [f_i(t)/f_k(t)] f_k(z_r)] \xi_r = 0, \qquad i < k.$$

Or,

$$\textstyle\sum_{r=0}^{k} (-1)^r f_i(z_r) \xi_r = [f_i(t)/f_k(t)] \sum_{r=0}^{k} (-1)^r f_k(z_r) \xi_r, \qquad i < k. \tag{14}$$

These equations can be shown to be satisfied by choosing ξ_r by means of the formula

$$(15) \qquad \xi_r = \begin{vmatrix} f_0(z_0) & \cdots & f_0(z_{r-1}) & f_0(z_{r+1}) & \cdots & f_0(z_k) & f_0(t) \\ \vdots & & \vdots & \vdots & & \vdots & \vdots \\ f_k(z_0) & \cdots & f_k(z_{r-1}) & f_k(z_{r+1}) & \cdots & f_k(z_k) & f_k(t) \end{vmatrix}$$

Because of the assumption that $f_0, \cdots, f_k$ constitute a Chebyshev system on $[-1, t]$ with positive determinant, it follows that this determinant is positive for each r. To demonstrate that these positive weights do satisfy (14), it suffices to look at the expansion of the following determinant by minors of elements of the first row. For notational convenience let $z_{r+1} = t$ and $\xi_{r+1} = |f_j(z_i)|$.

$$\begin{vmatrix} f_i(z_0) & \cdots & f_i(z_k) & f_i(t) \\ f_0(z_0) & \cdots & f_0(z_k) & f_0(t) \\ \vdots & & \vdots & \vdots \\ f_k(z_0) & \cdots & f_k(z_k) & f_k(t) \end{vmatrix} = \sum_{r=0}^{k+1} (-1)^r f_i(z_r)\xi_r .$$

Since the value of this determinant is zero for $i = 0, 1, \cdots, k$, it follows that for each such value of i

$$\textstyle\sum_{r=0}^{k} (-1)^r f_i(z_r)\xi_r = (-1)^k f_i(t)\xi_{r+1} .$$

When $i = k$, this becomes

$$\textstyle\sum_{r=0}^{k} (-1)^r f_k(z_r)\xi_r = (-1)^k f_k(t)\xi_{r+1} .$$

Eliminating ξ_{r+1} from these two equations shows that (14) is satisfied with this choice of ξ_r.

In view of the equivalence of the optimum estimation of γ_k and of $\sum_{i=0}^{k} \beta_i f_i(z)$, the preceding results demonstrate that the optimum estimate of the latter quantity is obtained by finding the $k + 1$ values of x at which the difference (8) assumes its maximum absolute value and then taking observations at those points in the proportions given by (15).

The assumption that the functions $f_0(x), \cdots, f_k(x)$ constitute a Chebyshev system in the two intervals $[-1, 1]$ and $[-1, t]$ and that $f_0(x), \cdots, f_{k-1}(x)$ do so on $[-1, 1]$ is slightly weaker than the assumptions made in [3] to solve this same problem. As in [3], the assumption that these functions constitute a Chebyshev system on $[-1, t]$ can be weakened by merely requiring that the determinant in (15) be positive for all sets of x's satisfying $-1 \leqq x_0 < \cdots < x_k \leqq 1$ but with fixed $t > 1$. The explicit formula for the weights given by (15) is another advantage of this formulation over that given in [3] where it is necessary to solve equations (14). This was possible only because of the form of the particular parametric function being estimated here and is not applicable to the more general parametric functions studied in [3].

4. Acknowledgment. I wish to express my appreciation to Professors Kiefer and Wolfowitz for sending me an early pre-publication copy of their paper [3] and for their numerous neat formulations of problems in [3] and [2] which I have been able to exploit to advantage in this and other papers.

REFERENCES

[1] ACHIESER, N. I. (1956), *Theory of Approximation*, Ungar, New York.

[2] KIEFER, J. and WOLFOWITZ, J. (1959). Optimum designs in regression problems. *Ann. Math. Statist.*, **30** 275–276.

[3] KIEFER, J. and WOLFOWITZ, J. (1965). On a theorem of Hoel and Levine on extrapolation designs. *Ann. Math. Statist.* **36** 1627–1655.