The Annals of Statistics
1981, Vol. 9, No. 4, 909–912

REGRESSION SYSTEMS FOR WHICH OPTIMAL EXTRAPOLATION DESIGNS REQUIRE EXACTLY $k + 1$ POINTS

By Paul G. Hoel

University of California, Los Angeles

The problem of optimal design for linear regression extrapolation is discussed. A restriction is introduced which will ensure that the number of optimal design points is equal to the number of independent regressors.

It is known Hoel (1966) that an optimal design for estimating an extrapolated value of a random variable $y(x)$, based on a regression function of the form $E\{y(x)\} = \beta_0 f_0(x) + \cdots + \beta_k f_k(x)$, where the f's constitute a restricted Chebyshev system of functions, can be obtained by using only $k + 1$ observation points in an interval $[a, b]$. However, examples of regression functions satisfying the restrictions needed to ensure the existence of a $k + 1$ point optimal design can be constructed that also possess an optimal design based on more than $k + 1$ points. It is the purpose of this note to introduce an additional restriction on the Chebyshev system that will ensure that no optimal design can involve more than $k + 1$ points, and hence ensures that exactly $k + 1$ points must be used in the design.

The additional restriction is the following:

If $f_k(x)$ is not a constant function on $[a, b]$, then for any constants $\alpha_0, \cdots, \alpha_{k-1}, c$, the equation

$$f_k(x) - \sum_{j=0}^{k-1} \alpha_j f_j(x) = c$$

may possess at most $k + 1$ roots in $[a, b]$.

A different restriction was introduced in Karlin and Studden (1966b) and Kiefer and Wolfowitz (1965) to solve this problem; however that restriction is quite severe and is not satisfied by many natural regression systems. Examples are given at the end of this paper.

In the following discussion it is assumed that the variance of $y(x)$ is independent of x and that the continuous functions $f_0(x), \cdots, f_{k-1}(x)$ are a Chebyshev system of functions on $[a, b]$. For later use, it is also assumed that the continuous functions $f_0(x), \cdots, f_k(x)$ are a Chebyshev system on $[a, t]$, where $t > b$ is the extrapolation point. These are the assumptions made in Hoel (1966) to ensure that an optimal design exists that requires only $k + 1$ points.

The problem of the optimal estimation of $E\{y(t)\}$ can be reduced to the problem of the optimal estimation of a single regression coefficient of a related regression function. Therefore it will suffice to first study the problem of how to estimate β_k in an optimal manner. If estimation is by least squares and n observations of $y(x)$ are to be taken at the points $x_1, \cdots, x_n$ in $[a, b]$, then it is shown in Kiefer and Wolfowitz (1959) that the variance of the least squares estimator, $\hat{\beta}_k$, will be minimized if the function

$$G = \min_c \sum_{i=1}^{n} \{f_k(x_i) - \sum_{j=0}^{k-1} c_j f_j(x_i)\}^2$$

is maximized. If some of the x_i are the same, G may be written in the form

$$G = \min_c \sum_{i=1}^{k+s} \{f_k(x_i) - \sum_{j=0}^{k-1} c_j f_j(x_i)\}^2 \xi_i,$$

where there are now $k + s$ distinct points and the ξ_i, $\sum_{i=1}^{k+s} \xi_i = n$, are the weights assigned to those points.

Received December, 1979; revised October, 1980.

AMS 1970 *subject classifications.* Primary 62K05; secondary 62J05.

Key words and phrases. Optimal regression designs, Chebyshev systems.

Let c_j^* be the coefficients of the best Chebyshev approximation of $f_k(x)$ by means of $\sum_{j=0}^{k-1} c_j f_j(x)$ over $[a, b]$. Then, for any set of design points and weights,

$$G \leq \sum_{i=1}^{k+s} \{f_k(x_i) - \sum_{j=0}^{k-1} c_j^* f_j(x_i)\}^2 \xi_i.$$

Let m denote the maximum absolute value over $[a, b]$ of the function

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

It then follows that

$$\sum_{i=1}^{k+s} \{f_k(x_i) - \sum_{j=0}^{k-1} c_j^* f_j(x_i)\}^2 \xi_i \leq m^2 \xi + \delta^2(n - \xi),$$

where ξ is the sum of the ξ_i where $|D(x_i)| = m$ and $\delta < m$ is the maximum value of $|D(x_i)|$ for those points where $|D(x_i)| < m$. Hence, unless $|D(x_i)| = m$ at all $k + s$ points, it is certain that

$$\max G < nm^2.$$

But it is known (Hoel (1966)) that the maximum value of G can be attained by using only $k + 1$ points, if the proper weights are chosen and the x_i are chosen to make $|D(x_i)| = m$ with $D(x_{i+1}) = -D(x_i)$, and that this maximum value is nm^2. In order to attain optimality using $k + s$ points, where $s > 1$, it is therefore necessary that $|D(x_i)| = m$ at all $k + s$ points.

Since $D(x)$ is a linear combination of the $f_j(x)$ and they constitute a Chebyshev system, $D(x)$ cannot posses more than k zeros in $[a, b]$. Furthermore, since the c_j^* produced a best Chebyshev approximation of $f_k(x)$ on $[a, b]$, the function $D(x)$ must possess at least $k + 1$ points where $|D(x_i)| = m$ with $D(x_{i+1}) = -D(x_i)$. But if $D(x)$ possessed more than $k + 1$ points where it alternately assumes the values m and $-m$, it would necessarily have more than k zeros, which is not possible. This contradiction shows that if $D(x)$ is to have more than $k + 1$ points where $|D(x)| = m$, there must be at least two consecutive points among the $k + s$ where the sign alternation does not occur. This possibility for $k = 3$ is illustrated in the sketch.

Although, as required here, $D(x)$ possesses only 3 zeros, and 4 points that satisfy $|D(x_i)| = m$ with $D(x_{i+1}) = -D(x_i)$, it possesses 5 points where $|D(x_i)| = m$. By choosing $c < m$ sufficiently large, the equation $D(x) = c$ will, because of continuity, possess 5 roots. But from the additional restriction placed on the f's in paragraph two, such an equation is permitted to have at most 4 roots; therefore two such consecutive points cannot occur and $s = 1$.

The preceding geometrical type of argument is applicable for general k, provided that $|D(x)| = m$ does not hold for an interval of values in $[a, b]$. If $|D(x)| = m$ for an interval of values, it will suffice to choose $c = m$ or $c = -m$ to arrive at a contradiction concerning the assumption made in paragraph two. In either case, the contradiction proves that an optimal design for this class of functions cannot contain more than $k + 1$ points.

It is known (Hoel (1966)) that under the assumptions of paragraph three the points that yield an optimal design for estimating β_k are also the optimizing points for estimating $E\{y(t)\}$, where t is the extrapolation point outside the interval $[a, b]$. Furthermore, a set of points that is optimal for estimating $E\{y(t)\}$ must also be optimal for estimating β_k. This equivalence requires that $f_k(t) \neq 0$. However any one of the $f_j(x)$ that does not vanish at $x = t$ may play the role of $f_k(x)$. Consequently, an optimal solution for the extrapolation problem cannot use more than $k + 1$ points in the design if the continuous functions $f_j(x)$ satisfy the following restrictions:

(a) $f_0(x), \cdots, f_{k-1}(x)$ are a Chebyshev system on $[a, b]$;

(b) $f_0(x), \cdots, f_k(x)$ are a Chebyshev system on $[a, t]$, $t > b$;

(c) $f_k(x) - \sum_{j=0}^{k-1} \alpha_j f_j(x) = c$ may possess at most $k + 1$ roots in $[a, b]$ for any constants $\alpha_0, \cdots, \alpha_{k-1}, c$, where $f_k(x)$ is assumed to be a nonconstant function such that $f_k(t) \neq 0$.

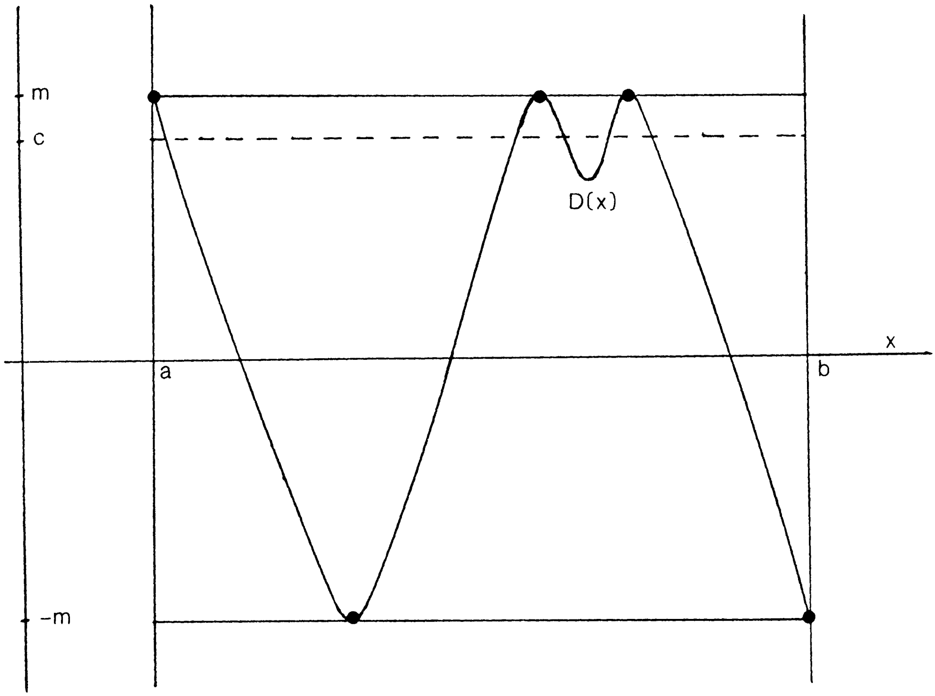

FIG. 1

The restrictions (a) and (b) are those that were used in Hoel (1966) and constitute sufficient conditions to ensure the existence of a $k + 1$ point optimal solution. Restriction (c) is an additional sufficient condition to ensure the impossibility of finding an optimal solution based on more than $k + 1$ points.

The preceding result also applies to the situation in which the standard deviation, $\sigma(x)$, of $y(x)$ is a known function of x, provided that the functions $F_j(x) = f_j(x)\sigma(x)$ satisfy these restrictions.

The following sets of functions are examples of Chebyshev systems to which the preceding theory applies.

(1) A Chebyshev system that contains the function $f_0(x) = 1$. The polynomial functions $1, x, \cdots, x^k$ are a special case of this.

(2) A Chebyshev system of functions whose derivatives also form a Chebyshev system. The exponential functions $\exp(a_0 x), \cdots, \exp(a_k x)$ are a special case of this.

(3) Weighted polynomials of the type $w(x), xw(x), \cdots, x^k w(x)$, where $R(x) = \dfrac{1}{w(x)} > 0$ and $R^{(k+1)}(x) = 0$ has no roots in $[a, b]$.

The fact that the exponential functions in (2) satisfy the Chebyshev systems assumptions follows from the discussion on page 10 of "Tchebycheff Systems" by Karlin and Studden (1966).

The fact that the weighted polynomials in (3) satisfy the assumptions made on the $f_j(x)$ can be demonstrated by calculating $k + 1$ derivatives of the function

$$g(x) = \alpha_0 + \alpha_1 x + \cdots + \alpha_k x^k - cR(x)$$

and then, under the assumption that $g(x)$ can possess more than $k + 1$ zeros, applying Rolle's theorem repeatedly to arrive at a contradiction concerning the roots of $R^{(k+1)}(x) = 0$.

The restriction that was imposed in the two papers Karlin and Studden (1966b) and Kiefer and Wolfowitz (1965) referred to earlier, does not hold for systems (2) and (3) but it does hold for system (1).

REFERENCES

HOEL, P. G. (1966). A simple solution for Chebyshev regression extrapolation. *Ann. Math. Statist.* **37** 720-725.

KARLIN, S. and STUDDEN, W. (1966a). *Tchebycheff Systems.* Interscience, New York.

KARLIN, S. and STUDDEN, W. (1966b). Optimal experimental designs. *Ann. Math. Statist.* **37** 783-815.

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

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

DEPARTMENT OF MATHEMATICS
UNIVERSITY OF CALIFORNIA, LOS ANGELES
LOS ANGELES, CALIFORNIA 90024