OPTIMAL SPACING AND WEIGHTING IN POLYNOMIAL PREDICTION

By P. G. Hoel[1] and A. Levine

University of California, Los Angeles

1. Summary. A solution is given to the problem of how to determine at which points in the interval $[-1, 1]$ observations should be taken and what proportion of the observations should be taken at each such point so as to minimize the variance of the predicted value of a polynomial regression curve at a specified point beyong the interval observations. The solution obtained states that the points are to be chosen to be Chebychev points and the number of observations are to be selected proportional to the absolute value of the corresponding Lagrange polynomial at the specified point. The preceding Chebychev solution becomes the minimax solution for the interval $(-1, t)$, provided $t > t_1 > 1$ where t_1 is a value satisfying a certain equation. Under the customary normality assumptions, the Chebychev solution to the prediction problem is used to construct a confidence band for a polynomial curve that will possess minimum width at any specified point beyond the interval of observations.

2. Optimum prediction. Let $-1 \leqq x_i \leqq 1, i = 1, 2, \cdots, n$, denote the selected values of a variable x at which observations are to be made on a related variable y corresponding to those selected values. If $y(x_i)$ denotes the observed value of y corresponding to x_i, it will be assumed that the variables $y(x_i)$, $i = 1, 2, \cdots, n$, are uncorrelated random variables with a common variance σ^2. It will also be assumed that the means of the y's lie on a polynomial curve of known degree k, that is, that $E[y(x_i)] = \beta_0 + \beta_1 x_i + \cdots + \beta_k x_i^k$.

The basic problem that will be considered here is how to determine where the x_i should be chosen in the interval $[-1, 1]$ so as to minimize the variance of the predicted value of y at some point x beyond the interval of observations. If the predicted value of $E[y(x)]$, which will be denoted by $\hat{y}(x)$, for a given set of x_i values is chosen to be the traditional least squares estimator of $E[y(x)]$, which is also its minimum variance unbiased linear estimator, then the problem is to choose the x_i so as to minimize $V[\hat{y}(x)]$. Now it is well known that this variance is given by the matrix formula $V[\hat{y}(x)] = x'(X'S^{-1}X)^{-1}x$ where $x' = (1, x, x^2, \cdots, x^k)$, S is the covariance matrix of the y's, and X is the spacing matrix

$$X = \begin{bmatrix} 1 & x_1 & x_1^2 & \cdots & x_1^k \\ 1 & x_2 & x_2^2 & \cdots & x_2^k \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & x_n & x_n^2 & \cdots & x_n^k \end{bmatrix}.$$

Received 7 May 1964.

[1] This research was supported by the Air Force Office of Scientific Research Grant #AF-AFOSR-62-158.

It is also well known [6] that it is possible to choose at most $k + 1$ distinct points x_i in the interval $[-1, 1]$ and distribute the observations at those points in proportions p_i in such a way as to obtain the same information matrix $X'S^{-1}X$ as for the original choice of x_i values. If the total number of observations to be taken, n, is fixed, these proportions may not yield integer values for the number of observations to be taken at the various points; therefore the theory that is about to be presented must be considered as an approximate theory only. Thus, if an optimum design required $n_i = np_i$ observations to be taken at x_i and n_i is not an integer, it would be necessary to choose an integer close to n_i for the actual number of observations.

When only $k + 1$ observation points are chosen for estimating a polynomial of degree k, the least squares estimator passes through the $k + 1$ mean points $(x_i, \bar{y}_i)$, $i = 0, 1, \cdots, k$; therefore its equation can be written in the form

$$\hat{y}(x) = \sum_{i=0}^{k} L_i(x)\bar{y}_i, \tag{1}$$

where $L_i(x)$ is the Lagrange polynomial given by

$$L_i(x) = \frac{(x - x_0) \cdots (x - x_{i-1})(x - x_{i+1}) \cdots (x - x_k)}{(x_i - x_0) \cdots (x_i - x_{i-1})(x_i - x_{i+1}) \cdots (x_i - x_k)}. \tag{2}$$

It therefore follows (1) that

$$V[\hat{y}(x)] = (\sigma^2/n) \sum_{i=0}^{k} L_i^2(x)/p_i. \tag{3}$$

The problem is now reduced to choosing the $k + 1$ values of x_i, $i = 0, 1, \cdots, k$, and the corresponding values of p_i to minimize (3). This will be accomplished by first selecting the minimizing p's corresponding to any chosen set of x's, and then selecting the minimizing x's. The following lemma yields the desired solution for the p's.

LEMMA 1. *If the p_i, $i = 0, 1, \cdots, k$, are allowed to vary continuously in $(0, 1)$ under the restriction $\sum_{i=0}^{k} p_i = 1$, then for a fixed $x \neq x_i$, $i = 0, 1, \cdots, k$, the choice*

$$p_i = \frac{|L_i(x)|}{\sum |L_i(x)|}, \qquad i = 0, 1, \cdots, k \tag{4}$$

will minimize $V[\hat{y}(x)]$.

The proof is carried out by using standard calculus techniques and verifying by means of second derivatives that the critical point yields a relative minimum which is also an abolute minimum. Although x is assumed to be a given point outside the interval $[-1, 1]$, and therefore it should be unnecessary to insert the restriction $x \neq x_i$, this is done to point out the fact that these p's yield a minimum whether x is inside or outside $[-1, 1]$, as long as x is not chosen at an observation point.

The next lemma is needed for determining the minimizing x's.

LEMMA 2. *Let $x_0 < x_1 < \cdots < x_k$ be $k + 1$ distinct points in $[-1, 1]$ and let*

$G(x)$ be a kth degree polynomial possessing the property that $G(x_0) = 0$ and the property that $G(x_i) = 0$ or sign *$G(x_i) = (-1)^i$, $i = 1, 2, \cdots, k$. Then $G(x)$ possesses at least j roots, counting multiplicities, in the closed interval $[x_0, x_j]$.*

PROOF. Consider any interval (x_{i-1}, x_i), $i > 1$. Assume, for convenience, that i is odd. Then $G(x_{i-1}) = 0$, or $G(x_{i-1}) > 0$ and $G(x_i) = 0$, or $G(x_i) < 0$. There are four possibilities that need to be considered:

(a) $G(x_{i-1}) > 0$, $G(x_i) < 0$.
Here there is obviously at least one root in the interval (x_{i-1}, x_i).

(b) $G(x_{i-1}) > 0$, $G(x_i) = 0$.
There is at least one root in the interval if the right end point is included in the interval.

(c) $G(x_{i-1}) = 0$, $G(x_i) < 0$.
There is at least one root in the interval if the left end point is included in the interval.

(d) $G(x_{i-1}) = 0$, $G(x_i) = 0$.
There is at least one root in the interval if either end point is included in the interval.

Assume that only the left end point will be considered as always belonging to an interval. Then (a), (c), and (d) all qualify as yielding at least one root in the interval. It remains, therefore, to consider case (b). There are three possibilities to be considered here.

$$(\alpha) \qquad G'(x_i) = 0, \qquad G(x_{i+1}) \geqq 0.$$

Here x_i is a double root; hence one root may be assigned to the left interval and one to the right interval; thus yielding at least one root to each interval.

$$(\beta) \qquad G'(x_i) > 0, \qquad G(x_{i+1}) \geqq 0.$$

Here there must exist a root between x_{i-1} and x_i; hence the root at x_i may be assigned to the right interval, thereby yielding at least one root in each interval.

$$(\gamma) \qquad G'(x_i) < 0, \qquad G(x_{i+1}) \geqq 0.$$

If $G(x_{i+1}) > 0$, there will exist a root between x_i and x_{i+1}; therefore the root at x_i may be assigned to the left interval and the other root to the right interval. If $G(x_{i+1}) = 0$, there is either a root between x_i and x_{i+1}, in which case the root at x_i may be assigned to the interval (x_{i-1}, x_i) and the other root to (x_i, x_{i+1}), or the graph of $G(x)$ will lie below the x axis in that interval. Since $G(x_{i+2}) \leqq 0$, if $G(x_{i+2}) < 0$ it will follow that there must be a root between x_{i+1} and x_{i+2} or a double root at x_{i+1}; hence the earlier roots at x_i and x_{i+1} can be assigned to the left intervals and the remaining root to the right interval, and thus each of the three intervals will be assigned a root. If $G(x_{i+2}) = 0$, the same type of situation arises as for $G(x_{i+1}) = 0$ and the same type of argument can be applied repeatedly until a non-zero value of G is obtained. If no non-zero value is obtained, it must be that $G(x_j) = 0$. But then x_j may be assigned to the interval on its left, and thus each preceding interval can be assigned a root.

Hence it is always possible to assign one root to each interval (x_{i-1}, x_i), $i > 1$.

For the interval (x_0, x_1) there is the root x_0. Thus, there must exist at least j roots of $G(x)$ in $[x_0, x_j]$. The lemma is obviously valid also if the sign $G(x_i) = (-1)^{i+1}$; it is merely necessary that the signs alternate.

This lemma will now be used to prove the following theorem which shows how to choose the minimizing x_i values.

THEOREM. *If the minimizing p's given by formula* (4) *are used, the* $k + 1$ *observation points that will minimize* $V[\hat{y}(x)]$ *at* $x > 1$ *are given by the Chebychev points*

$$x_i = -\cos i\pi/k, \qquad i = 0, 1, \cdots, k. \tag{5}$$

PROOF. If the p's of formula (4) are substituted in (3), the variance of $\hat{y}(x)$ will reduce to

$$V[\hat{y}(x)] = (\sigma^2/n)\left(\sum_{i=0}^{k} |L_i(x)|\right)^2.$$

It sufficies therefore to find the set $\{x_i\}$ that will minimize $\sum_{i=0}^{k} |L_i(x)|$.

It will be observed that when $x > 1$ the value of $|L_i(x)|$ will decrease as x_0 decreases, or as x_k increases; consequently the same property will hold for $V[\hat{y}(x)]$. In seeking for a set $\{x_i\}$ to minimize $V[\hat{y}(x)]$ it therefore suffices to consider only those sets for which $x_0 = -1$ and $x_k = 1$.

For $x > 1$ it is readily seen that

$$\sum_{i=0}^{k} |L_i(x)| = \sum_{i=0}^{k} (-1)^{k-i} L_i(x). \tag{6}$$

In view of this relationship, the problem is reduced to finding the set of values $\{x_i\}$ that will minimize the value of a polynomial of the type on the right side of (6) at the fixed point x, where $x > 1$. A polynomial of this type assumes the value $(-1)^{k-i}$ at the point x_i associated with it. Thus, the problem is to find a polynomial $C(x)$, if it exists, that passes through $(-1)^{k-i}$ at the x_i associated with it and such that $C(x) < Q(x)$ for $x > 1$, where $Q(x)$ is any other polynomial which passes through $(-1)^{k-i}$ at the x_i associated with it.

Now it will be shown that there exists a polynomial $C(x)$ of degree k that possesses the properties

$$C(x_i) = (-1)^{k-i}, \qquad i = 0, 1, \cdots, k$$

$$C'(x_i) = 0, \qquad i = 1, \cdots, k - 1$$

which is unique among polynomials possessing these properties, and which satisfies the inequality $C(x) < Q(x)$ for $x > 1$, where $Q(x)$ is any polynomial satisfying $Q(x_i') = (-1)^{k-i}$, $i = 0, 1, \cdots, k$ and where $\{x_i'\}$ is any set of observation points other than the set $\{x_i\}$ associated with $C(x)$.

The existence and uniqueness of $C(x)$ follows from the fact that $C(x)$ is the kth degree Chebychev polynomial whose properties are well known [4]. The Chebychev polynomial oscillates between the limits ± 1 in the interval $[-1, 1]$ and has its points of tangency to these limits located at the points given by formula (5).

It remains to be shown that $C(x) < Q(x)$ for $x > 1$. Toward this end, consider the polynomial $D(x) = Q(x) - C(x)$. The degree of $D(x)$ is at most k. Assume, temporarily, that its degree is d, with $0 < d < k$. Since $|C(x)| \leqq 1$, it follows that $D(x_i') = (-1)^{k-i} - C(x_i')$ will possess the sign of $(-1)^{k-i}$, or the value 0. The latter may occur if $x_i' = x_j$ for some j. Further, $D(-1) = 0$ because $x_0 = x_0' = -1$ and $C(-1) = Q(-1) = (-1)^k$. Thus, $D(x)$ satisfies the conditions of Lemma 2 and therefore possesses d roots in the interval $[x_0, x_d']$. Since $D(x)$ is assumed to be of degree d it cannot have any roots outside $[x_0, x_d']$. But $D(x_k) = 0$ because $x_k = x_k' = 1$ and $C(1) = D(1) = 1$. This contradiction proves that $D(x)$ must be of degree k. But now it follows from Lemma 2 that $D(x)$ possesses k roots in $[-1, 1]$. As a result $D(x)$ cannot have any roots outside $[-1, 1]$. Hence, when $x > 1$.

$$C(x) > Q(x), \text{ or } C(x) < Q(x).$$

To show that the latter inequality holds it suffices to study the sign changes of $D(x)$. Since the sign $D(x_i') =$ sign of $(-1)^{k-i}$ if $x_i' \neq x_j$ for any j and $D(x_i') = 0$ if $x_i' = x_j$ for some j with $i + j$ even, it follows that $D(x_{k-1}') < 0$ unless $x_{k-1}' = x_j$ for some j for which $k + j$ is odd, in which case $D(x_{k-1}') = 0$. Since x_i' cannot be equal to some x_j for $i + j$ even for every i, let $k - m$ denote the largest value of i for which $x_i' \neq x_j$ for some j with $i + j$ even. Then

$$D(x_{k-i}') = 0 \qquad \text{for } i = 0, 1, \cdots, m - 1$$

and

$$\text{sign } D(x_{k-m}') = \text{sign of } (-1)^m.$$

Since by Lemma 2 there are at least $k - m$ roots of $D(x)$ in the interval $[x_0, x_{k-m}']$ and there are m roots at the points x_{k-i}', $i = 0, 1, \cdots, m - 1$, all k roots are accounted for. Thus, there cannot be any double roots at the points x_{k-i}', $i = 0, 1, \cdots, m - 1$, and therefore the graph of $D(x)$ must cross at those points. Since $D(x_{k-m}')$ has the sign of $(-1)^m$, $D(x)$ will possess this sign inside the open interval (x_{k-m}', x_{k-m+1}'), the opposite sign inside the interval (x_{k-m+1}', x_{k-m+2}'), etc. It will therefore possess the sign of $(-1)^{2m-1}$ in the open interval (x_{k-1}', x_k). Thus, $D(x) < 0$ in this interval and therefore $D'(x_k) > 0$, otherwise a double root would occur at x_k. By continuity it follows that $D(x) > 0$ for x sufficiently close to 1 on the right; consequently $D(x) > 0$ for all $x > 1$. This demonstrates that $C(x)$ possess the optimum property claimed for it.

As an illustration of the advantage of using the optimum spacing and weighting given by this theorem, consider the problem of predicting the value of a third degree polynomial at the point $x = 2$ if 52 observations are to be taken in the interval $[-1, 1]$ and if $\sigma = 1$. Under the traditional approach of using equal spacing and weighting, one would choose

$$x_0 = -1, x_1 = -\tfrac{1}{3}, x_2 = \tfrac{1}{3}, x_3 = 1$$

$$n_0 = 13, n_1 = 13, n_2 = 13, n_3 = 13.$$

The corresponding values for the Chebychev solution given by formulas (4) and (5) will be found to be

$$x_0 = -1, x_1 = -\tfrac{1}{2}, x_2 = \tfrac{1}{2}, x_3 = 1$$

$$n_0 = 5, n_1 = 12, n_2 = 20, n_3 = 15.$$

Calculations will show that the corresponding values of $V[\hat{y}(2)]$ are, to the nearest integer, 20 and 13, respectively. Thus, the traditional method yields a value of the variance that is over 50 percent larger than that of the optimum solution.

3. Minimax solutions. If one desires a solution that will be appropriate for more than a single fixed value of x, it will be necessary to consider some other criterion, such as that of minimax. Since $V[\hat{y}(x)]$ is an increasing function of x for $x > 1$, it follows that $\max_{1 \leqq x \leqq t} V[\hat{y}(x)] = V[\hat{y}(t)]$. As a consequence, the solution to the minimax problem, namely, $\min_{\{p_i\},\{x_i\}} \max_{1 \leqq x \leqq t} V[\hat{y}(x)]$ is given by the preceding solution to the problem of minimizing $V[\hat{y}(t)]$.

Next, consider the problem of finding the minimax solution for the interval $(-1, t)$, where $t \geqq 1$. For $t = 1$ the solution is known [1] to be given by using equal weights and choosing the x_i, $i = 0, 1, \cdots, k$, to be the roots of the polynomial which is the integral from -1 to x of the Legendre polynomial of degree k. This solution will be called the Legendre solution. For $t > 1$ the Legendre solution cannot hold for the following reasons. It is known [1] that for this design $V[\hat{y}(x)]$ assumes its maximum value in $[-1, 1]$ at the internal Legendre points and at the end points of the interval. Since $V[\hat{y}(x)]$ is an increasing function of x for $x \geqq 1$, it follows that $V[\hat{y}(t)]$ exceeds its maximum value inside the interval for every $t > 1$. To show that this is not possible for an optimum solution, assume the contrary. Choose two of the functions $L_i(x)$ that are not equal at $x = t$. Let them be denoted by $L_\alpha(x)$ and $L_\beta(x)$ with $L_\alpha(t) < L_\beta(t)$. Now write

$$V[\hat{y}(x)] = \sum_{i \neq \alpha,\beta} \frac{L_i^2(x)}{p_i} + \frac{L_\alpha^2(x)}{p_\alpha} + \frac{L_\beta^2(x)}{p_\beta}.$$

For the Legendre solution the p's are equal; therefore consider changing this solution slightly by altering p_α and p_β under the restriction that $p_\alpha + p_\beta = 2/(k+1)$. It is easily shown that $[L_\alpha^2(t)/p_\alpha] + [L_\beta^2(t)/p_\beta]$ is a decreasing function of p_β in a sufficiently small interval to the right of $p_\beta = 1/(k+1)$. Thus, by increasing p_β slightly from its Legendre value, the value of $V[\hat{y}(t)]$ can be decreased slightly. From continuity considerations this can increase the maximum value of $V[\hat{y}(x)]$ inside the interval only slightly and therefore the net result is to produce a smaller overall maximum value, contradicting the assumption that the Legendre solution could be a minimax solution for the interval $[-1, t]$ where $t > 1$.

For t sufficiently large, the Chebychev solution will also be a minimax solution for $[-1, t]$ for the following reasons. Consider the Chebychev design based on the point $x = t > 1$ and let $V_c[\hat{y}(x)]$ denote the corresponding variance function. If t is such that $\max_{-1 \leqq x \leqq 1} V_c[\hat{y}(x)] < V_c[\hat{y}(t)]$ the maximum value of $V_c[\hat{y}(x)]$

for $-1 \leqq x \leqq t$ will be attained at $x = t$ because $V_c[\hat{y}(x)]$ is an increasing function of x for $x > 1$. Since no other design can have $V[\hat{y}(t)]$ smaller than $V_c[\hat{y}(t)]$, it follows that the Chebychev design will be a minimax design for the interval $[-1, t]$. It remains to be shown that there exists a value of t satisfying the preceding inequality.

For a given t, let $R(t)$ denote the ratio

$$R(t) = \max_{-1 \leqq x \leqq 1} V_c[\hat{y}(x)]/V_c[\hat{y}(t)].$$

When the values of the p_i given by (4) with $x = t$ are substituted in (3), $R(t)$ will assume the form

$$R(t) = [\max_{-1 \leqq x \leqq 1} \sum L_i^2(x)/|L_i(t)|]/\sum |L_i(t)|.$$

Each $|L_i(t)|$ is an increasing function of t for $t > 1$; therefore $R(t)$ is a decreasing function of t for $t > 1$. Since each $L_i(t) \to \infty$ as $t \to \infty$, $R(t)$ will approach 0 as $t \to \infty$. Since $L_i(t) \to \delta_{ik}$ as $t \to 1$, $R(t)$ will become infinite as $t \to 1$. But $R(t)$ is a continuous function of t for $t > 1$; consequently there will exist a unique value of t, denoted by t_1, satisfying $R(t_1) = 1$. For $t > t_1$, $R(t) < 1$; therefore the earlier desired inequality is satisfied to insure that the Chebychev design is a minimax design for any interval $[-1, t]$ for which $t > t_1$.

It follows from the same type of reasoning as that just used and that employed to reject the possibility of a Legendre design for $t > 1$ that the Chebychev design cannot be optimum for $t < t_1$. From continuity considerations one would expect the optimum design to gradually change from the Legendre spacing and weighting to the Chebychev spacing and weighting as t increases from 1 to t_1.

4. Confidence bands. The preceding results can be used to construct a confidence band for a polynomial curve that will possess minimum width at any desired external point. First it is necessary to formulate the problem in terms of the earlier notation.

The equation of a kth degree polynomial curve can be written in the vector form

$$\mu(x) = z'\alpha, \tag{7}$$

where $\alpha' = (\alpha_0, \alpha_1, \cdots, \alpha_k)$ is the vector of unknown coefficients and $z' = (z_0, z_1, \cdots, z_k)$ is the vector of Lagrange polynomial components given by formula (2) with $z_i = L_i(x)$.

If it is assumed, as before, that n_i observations are taken at x_i and that the corresponding y's are uncorrelated with a common variance σ^2, then it follows from (1) that the standard least squares estimator of (7) is given by $\hat{\mu}(x) = z'\hat{\alpha} = z'\bar{y}$.

Now if one inspects the formulas [2], [5] for a confidence band for a polynomial curve and expresses them in terms of the preceding notation, he will find that they assume the form

$$z'\hat{\alpha} \pm (c\sum z_i^2/n_i)^{\frac{1}{2}},$$

where $c = \sigma^2\chi_0^2$, based on $k + 1$ degrees of freedom, if σ^2 is known, and $c = F_{j;k+1,n-k-1}[(k + 1)/(n - k - 1)]\sum_{i=0}^{k} \sum_{j=1}^{n_i}(y_{ij} - \bar{y}_j)^2$ if σ^2 is unknown. The quantity $\sum z_i^2/n_i$ is the same as the quantity $\sum L_i^2(x)/p_i$ in (3); consequently the problem of minimizing the width, or the expected width, of this confidence band at any desired point is equivalent to the earlier problem of minimizing the variance of a predicted value at a specified point. Thus, if the x_i are chosen to be the Chebychev points and the n_i to be the corresponding optimum weights, the band width will be minimized at the selected x value. This same property obviously holds also for the confidence interval of a single ordinate of a polynomial regression curve.

As an illustration of the improvement in band width, consider the two confidences bands obtained for a third degree polynomial when equal spacing and weighting is used and when the optimizing technique is employed for $x = 2$.

Under equal spacing and weighting, the quantity $\sum z_i^2/n_i$ becomes

$$\frac{4}{n}\sum_{i=0}^{3} z_i^2, \tag{8}$$

where z_i is calculated using $x_0 = -1, x_1 = -\frac{1}{3}, x_2 = \frac{1}{3}, x_3 = 1$. Under optimum spacing and weighting it becomes

$$\frac{52}{n}\left(\frac{z_0^2}{5} + \frac{z_1^2}{12} + \frac{z_2^2}{20} + \frac{z_3^2}{15}\right), \tag{9}$$

where z_i is calculated using $x_0 = -1, x_1 = -\frac{1}{2}, x_2 = \frac{1}{2}, x_3 = 1$. Since it is the square root of $\sum z_i^2/n_i$ that enters as a factor in the width of these confidence bands, calculations were made of the square root of this quantity in the two cases with $n = 52$, but with the common factor $13^{-\frac{1}{2}}$ omitted, with the following results:

x	-1	$-\frac{1}{2}$	$-\frac{1}{3}$	0	$\frac{1}{3}$	$\frac{1}{2}$	1	1.5	2
(8)	1	1.08	1	.80	1	1.08	1	5.5	16.1
(9)	1.61	1.04	1.06	.93	.84	.81	.93	4.6	13.0

The optimum band is considerably narrower for most positive values of x, and of course is much narrower at $x = 2$.

REFERENCES

[1] Guest, P. G. (1958). The spacing of observations in polynomial regression. *Ann. Math. Statist.* **29** 294–299.

[2] Hoel, P. G. (1954). Confidence bands for polynomial curves. *Ann. Math. Statist.* **25** 534–542.

[3] Kiefer, J. and Wolfowitz, J. (1959). Optimum designs in regression problems. *Ann. Math. Statist.* **30** 271–294.

[4] *Nat. Bur. Standards, Applied Math. Ser.*, No. 9. Tables of Chebychev Polynomials.

[5] Rao, C. R. (1959). Some problems involving linear hypotheses in multivariate analysis. *Biometrika* **46** 49–58.

[6] Shohat, J. A. and Tamarkin, J. D. (1943). The Problem of Moments, Math. Surveys, No. 1, *Amer. Math. Soc.*, New York.