ZASTOSOWANIA MATEMATYKI
APPLICATIONES MATHEMATICAE
XIX, 1 (1987), pp. 93–107

T. RYCHLIK (Warszawa)

ROBUST EXPERIMENTAL DESIGN
A COMMENT ON HUBER'S RESULT

0. This paper deals with a problem considered by P. J. Huber in Chapter 9 of his monograph [2] and earlier in [1]. He obtained a solution, which is a design in the form of a density with regard to a Lebesgue measure. It is not straightforwardly applicable in practice.

Here the problem is investigated under some additional, but typical, restrictions and the results are compared with the solution of Huber.

1. The statement of the design problem and Huber's continuous solution ([1], [2]). Let us start with considering the simple linear model

$$y_i = \alpha+\beta x_i+\varepsilon_i, \quad i = 1, \ldots, n,$$

where ε_i are independent random errors with known normal distribution $N(0, \sigma^2)$ and the design $X = (x_1, \ldots, x_n)$ is an n-tuple of points in the interval $I = [-\frac{1}{2}, +\frac{1}{2}]$ (or equivalently, the probability measure over I

$$\xi = \xi_X = \frac{1}{n}\sum_{i=1}^{n} \delta_{x_i} \tag{1}$$

which is concentrated at n points with identical weights). Following Huber, we assume throughout the paper that the design is symmetric. We want to estimate the unknown regression coefficients α and β so as to minimize the integrated mean square error

$$Q = \mathrm{E}\int_I (\alpha+\beta x-\hat{\alpha}-\hat{\beta}x)^2\, dx.$$

Then the best estimators are the least squares estimators

$$\hat{\alpha} = \frac{1}{n}\sum_{i=1}^{n} y_i, \quad \hat{\beta} = \sum_{i=1}^{n} x_i y_i \Big/ \sum_{i=1}^{n} x_i^2. \tag{2}$$

Now, suppose that the regression function is only approximately linear. In the case

$$y_i = f(x_i)+\varepsilon_i, \quad i = 1, \ldots, n,$$

when the function $f(x)$ insignificantly differs from the linear one, the use of the least squares estimator $\hat{\alpha}+\hat{\beta}x$ may be justified, if its integrated mean square error for the violated model

$$Q = Q(f, X) = \mathrm{E}\int_I [f(x)-\hat{\alpha}-\hat{\beta}x]^2\,dx \tag{3}$$

does not increase too much. The regret can be reduced, if we choose an appropriate design. The error (3) takes the form

$$Q(f, X) = \int_I [f(x)-\alpha_f-\beta_f x]^2\,dx+\left[\frac{1}{n}\sum_{i=1}^{n} f(x_i)-\alpha_f\right]^2$$
$$+\frac{1}{12}\Big[\sum_{i=1}^{n} x_i f(x_i)/\sum_{i=1}^{n} x_i^2-\beta_f\Big]^2+\frac{\sigma^2}{n}\left[1+\left(12\cdot\frac{1}{n}\sum_{i=1}^{n} x_i^2\right)^{-1}\right],$$

where $\alpha_f = \int f$ and $\beta_f = \frac{1}{12}\int xf$ are coefficients of the L^2-projection of f on the space of linear functions. From now on we put $\alpha_f = \beta_f = 0$. It is obvious that without loss of generality we can take into account only nonlinear disturbances of the regression function. Then

$$Q(f, X) = \int [f(x)]^2\,dx+\left[\frac{1}{n}\sum_{i=1}^{n} f(x_i)\right]^2 \tag{4}$$
$$+\frac{1}{12}\Big[\sum_{i=1}^{n} x_i f(x_i)/\sum_{i=1}^{n} x_i^2\Big]^2+\frac{\sigma^2}{n}\left[1+\left(12\cdot\frac{1}{n}\sum_{i=1}^{n} x_i^2\right)^{-1}\right].$$

Roughly, the problem is to protect by means of design against the increase in estimation error (4), when the function f belongs to a set of small deviations from the linear regression model and the least squares estimator $f(x) = \hat{\alpha}+\hat{\beta}x$ with (2) is used.

Huber proposes the following approach, which consists in continuous approximation of the problem. Continuous symmetric designs (i.e. any probability measures over $[-\frac{1}{2}, +\frac{1}{2}]$) are allowed and the definition of the risk function Q is extended for them

$$Q(f, \xi) = \int f^2+\Big[\int f\,d\xi\Big]^2+\frac{1}{12}\Big[\int xf\,d\xi/\int x^2\,d\xi\Big]^2+\frac{\sigma^2}{n}\Big[1+\Big(12\int x^2\,d\xi\Big)^{-1}\Big]. \tag{5}$$

The values (4) and (5) are identical for discrete designs. Possible regression functions are such that their distance in L^2-norm to the subspace of linear functions is not greater than a small positive number $\sqrt{\eta}$. Thus, we consider the class of disturbances

(6) $$F = \{f\colon \int f = \int xf = 0,\ \int f^2 \leqslant \eta\}.$$

Huber has found a design which minimizes the value $\sup_{f \in F} Q(f, \xi)$ for n, σ^2 and η fixed in advance. It has a density and depends on the ratio $\eta/(\sigma^2/n)$.

If $\eta \geqslant \eta_0 = \dfrac{25}{162}\dfrac{\sigma^2}{n}$, then for $t = t(\eta, \sigma^2, n) \in (1, 9/5]$ such that

(7) $$\tfrac{5}{2} t^2 (t-1) = \sigma^2/n\eta$$

the density is

(8) $$m_t(x) = 1 + \tfrac{5}{4}(t-1)(12x^2 - 1).$$

If $\eta < \dfrac{25}{162}\dfrac{\sigma^2}{n}$, then for $c = c(\eta, \sigma^2, n) \in (0, 1)$ such that

(9) $$\frac{18}{25}\frac{(3+6c+4c^2+2c^3)^2}{(1+2c)^3(1-c)^2} = \frac{\sigma^2}{n\eta}$$

the density is

(10) $$m_c(x) = \frac{3}{(1+2c)(1-c)^2}(4x^2 - c^2)^+ .$$

The parameters t and c, which are given in implicit formulas (7) and (9), are uniquely determined.

Let us return to the original problem with discrete sampling and the criterion function (4). The minimax design is essentially continuous and cannot be simply applied in practice. Unfortunately, it appears also that no discrete approximation of it implies a coinciding value of the risk function. The class of disturbances (6) is so wide that it contains elements which attain arbitrarily large values at the sampling points of any discrete design. This fact implies the unboundedness of the error (4) and therefore, for each design X, $\sup_{f \in F} Q(f, X) = \infty$. In practice, however, we never consider such extraordinary functions. It seems to be reasonable to restrict the class of regression functions. We propose to take into account a finite-dimensional set of disturbances. This assumption concerns the situation in which the real problem is described by a linear regression model, but the model is incorrectly identified. So, the design should be robust on such mistakes.

2. Finite-dimensional set of disturbances. A sufficient optimality condition. The purpose is to find a discrete design minimizing the supremum of (4) over a set F_k, which is the intersection of (6) and the k-dimensional subspace of $L^2(-\frac{1}{2}, +\frac{1}{2})$ with basis $f_1, \ldots, f_k$. It is assumed that it contains a certain quadratic function (let, for instance, f_1 be quadratic)

(11) $$F_k = \{f = \sum_{j=1}^{k} \alpha_j f_j\colon \int f = \int xf = 0,\ \int f^2 \leqslant \eta,\ f_1(x) = ax^2+bx+c\}.$$

This technical assumption is made because of the form of the measurement error

$$\frac{\sigma^2}{n}\left[1+\left(12\cdot\frac{1}{n}\sum_{i=1}^{n} x_i^2\right)^{-1}\right],$$

but fortunately, it often appears acceptable.

As we shall see in Theorem 1, an experimental design will be minimax if its expectations of basis functions are identical to the corresponding expectations for the continuous design m_t (8). Then it is the best in the class of all designs, not necessarily concentrated on the interval $[-\frac{1}{2}, +\frac{1}{2}]$ and discrete.

THEOREM 1. *If a plan $H = (h_1, \ldots, h_n)$ satisfies*

(12) $$\begin{aligned} \mathrm{E}_{\xi_H} f_j(x) &= \mathrm{E}_{m_t} f_j(x), \\ \mathrm{E}_{\xi_H} xf_j(x) &= \mathrm{E}_{m_t} xf_j(x), \quad j = 1, \ldots, k, \end{aligned}$$

where ξ_H, $t > 1$ and m_t are as in (1), (7) *and* (8), *respectively, then*

$$\max_{f\in F_k} Q(f, H) = \min_{X\in \mathbf{R}^n} \max_{f\in F_k} Q(f, X).$$

Proof. There is no loss of generality if we assume that $\{f_1, \ldots, f_k\}$ is orthonormal. Then

$$f_1(x) = \frac{\sqrt{5}}{2}(12x^2-1).$$

The other f_j can be written in the form

$$f_j = u_j + v_j, \quad j = 2, \ldots, k,$$

where $u_j(x) = \frac{1}{2}[f_j(x)+f_j(-x)]$ are even functions, $v_j(x) = \frac{1}{2}[f_j(x)-f_j(-x)]$ are odd ones. The set $\{1, x, f_1, u_2, v_2, \ldots, u_k, v_k\}$ is an orthogonal system. The following relations hold:

$$m_t(x) = 1+\frac{\sqrt{5}}{2}(t-1)f_1(x), \quad xm_t(x) = tx+\frac{3\sqrt{7}}{28}(t-1)w(x),$$

where $w(x) = \sqrt{7}(20x^3-3x)$ is orthogonal to $\{1, x, x^2\}$ and $\int w^2 = 1$. Due to the facts that the design is symmetric and the considered functions are odd, we have

$$\frac{1}{n}\sum_i h_i f_1(h_i) = \frac{1}{n}\sum_i h_i u_j(h_i) = \frac{1}{n}\sum_i v_j(h_i) = 0.$$

Therefore, formulas (12) take the form

$$\frac{1}{n}\sum_i f_1(h_i) = \frac{\sqrt{5}}{2}(t-1),$$

$$\frac{1}{n}\sum_i u_j(h_i) = \int u_j\left[1+\frac{\sqrt{5}}{2}(t-1)f_1\right] = 0, \quad j = 2, \ldots, k,$$

$$\frac{1}{n}\sum_i h_i v_j(h_i) = \int v_j\left[tx+\frac{3\sqrt{7}}{28}(t-1)w\right] = \frac{3\sqrt{7}}{28}(t-1)\int v_j w, \quad j = 2, \ldots, k.$$

Furthermore,

$$\frac{1}{n}\sum_i h_i^2 = \frac{1}{12}t.$$

For each $f \in F_k$ having the form

$$f = \sqrt{\eta}\sum_{j=1}^{k} \alpha_j f_j \quad \text{with} \quad \sum_{j=1}^{k} \alpha_j^2 \leqslant 1$$

we have

$$Q(f, X) = \eta\left\{\sum_{j=1}^{k} \alpha_j^2 + \left[\alpha_1\frac{1}{n}\sum_i f_1(x_i) + \sum_{j=2}^{k} \alpha_j\frac{1}{n}\sum_i u_j(x_i)\right]^2\right.$$
$$\left.+\frac{1}{12}\left[\sum_{j=2}^{k} \alpha_j\frac{1}{n}\sum_i x_i v_j(x_i)\right]^2\Big/\left(\frac{1}{n}\sum_i x_i^2\right)^2\right\} + \frac{\sigma^2}{n}\left[1+\left(12\frac{1}{n}\sum_i x_i^2\right)^{-1}\right].$$

It suffices to show that H together with $\sqrt{\eta}f_1$ is a saddle point of Q, i.e. that

(13) $$Q(f, H) \leqslant Q(\sqrt{\eta}f_1, H) \leqslant Q(\sqrt{\eta}f_1, X)$$

for every f and X. The left-hand inequality is obtained in the following manner:

$$Q(f, H) = \eta\left\{\sum_{j=1}^{k} \alpha_j^2 + \alpha_1^2\frac{5}{4}(t-1)^2 + \frac{27}{28t^2}(t-1)^2\left[\sum_{j=2}^{k} \alpha_j\int v_j w\right]^2\right\} + \frac{\sigma^2}{n}\left(1+\frac{1}{t}\right)$$
$$\leqslant \eta\left[1+\max\left\{\frac{5}{4}(t-1)^2, \frac{27}{28t^2}(t-1)^2\sum_{j=2}^{k} \alpha_j^2\int v_j^2\right\}\right] + \frac{\sigma^2}{n}\left(1+\frac{1}{t}\right)$$

and because $\int v_j^2 \leqslant 1$, $\sum_{j=2}^{k} \alpha_j^2 \leqslant 1$ and $5/4 > 27/(28t^2)$ for each $t > 1$, we have

$$Q(f, H) \leqslant \eta\left[1+\frac{5}{4}(t-1)^2\right]+\frac{\sigma^2}{n}\left(1+\frac{1}{t}\right) = Q(\sqrt{\eta}f_1, H).$$

On the other hand,

$$Q(\sqrt{\eta}f_1, X) = \eta\left[1+\frac{5}{4}\left(12\frac{1}{n}\sum_i x_i^2-1\right)^2\right]+\frac{\sigma^2}{n}\left[1+\left(12\frac{1}{n}\sum_i x_i^2\right)^{-1}\right]$$

as a function of the nonnegative argument $\tau = 12(1/n)\sum_i x_i^2$ attains its unique minimum at a point $t > 1$ such that $\frac{5}{2}t^2(t-1) = \sigma^2/n\eta$, i.e. for $t = 12(1/n)\sum_i h_i^2$. Therefore, for each symmetric $X = (x_1, \ldots, x_n) \in \boldsymbol{R}^n$

$$Q(\sqrt{\eta}f_1, X) \geqslant Q(\sqrt{\eta}f_1, H),$$

which establishes the statement (13).

Remarks: (i) The uniquely determined function $m_{t(\eta,\sigma^2,n)}$ need not be nonnegative as in the original Huber's solution.

(ii) It is easily seen that the theorem remains true if we consider continuous designs.

3. Polynomial regression function. Now we pay our attention to the case when the nonlinear regression function is a polynomial with degree not greater than a known integer k. This restriction is justified by the following reasons:

(a) Real relations in nature are commonly characterized by polynomials. Moreover, it is the most natural way to generalize linear relations.

(b) The geometry of the classical moment space is best elaborated and the assumptions of Section 2 on the family of possible nonlinearities (11) are naturally satisfied.

3.1. General properties of optimal designs. Let

$$W_k = \{f(x) = \sum_{j=0}^{k} \alpha_j x^j \colon \int f = \int xf = 0,\ \int f^2 \leqslant \eta\}.$$

At first we present the explicit formula of $\max_{f \in W_k} Q(f, X)$. The set $\{e_j(x) = \sqrt{2j+1}\, P_j(2x),\ j = 2, \ldots, k\}$, where

(14) $$P_j(x) = \sum_{l=0}^{j} p_{j,l}\, x^l$$

denote Legendre polynomials, is an orthonormal basis of the space spanned on W_k. For $f = \sqrt{\eta} \sum_{j=2}^{k} \alpha_j e_j$, $\sum_{j=2}^{k} \alpha_j^2 \leqslant 1$,

$$(15)\quad Q(f, X) = \eta \left\{ \sum_j \alpha_j^2 + \left[\sum_j \alpha_{2j} \frac{1}{n} \sum_i e_{2j}(x_i) \right]^2 + \right.$$
$$\left. + \frac{1}{12} \left[\sum_j \alpha_{2j+1} \frac{1}{n} \sum_i x_i e_{2j+1}(x_i) \right]^2 \Big/ \left(\frac{1}{n} \sum_i x_i^2 \right)^2 \right\} +$$
$$+ \frac{\sigma^2}{n} \left[1 + \left(12 \frac{1}{n} \sum_i x_i^2 \right)^{-1} \right].$$

To maximize Q over W_k, $\sum_{j=2}^{k} \alpha_j^2$ must obviously be equal to 1. We see that the second and third terms of (15) depend on different variables α_i. For any arbitrary value of the sum $\sum_j \alpha_{2j}^2 = \theta \in [0, 1]$ they are maximal, if each of the α_j is proportional to its coefficient:

$$\alpha_{2j} \sim \frac{1}{n} \sum_i e_{2j}(x_i), \qquad \sum_j \alpha_{2j}^2 = \theta,$$
$$\alpha_{2j+1} \sim \frac{1}{n} \sum_i x_i e_{2j+1}(x_i), \qquad \sum_j \alpha_{2j+1}^2 = 1 - \theta.$$

Therefore

$$\max_{f \in W_k} Q(f, X) = \eta \max_{0 \leqslant \theta \leqslant 1} \left\{ 1 + \theta \sum_j \left[\frac{1}{n} \sum_i e_{2j}(x_i) \right]^2 + \right.$$
$$\left. + (1-\theta) \frac{1}{12} \sum_j \left[\frac{1}{n} \sum_i x_i e_{2j+1}(x_i) \right]^2 \Big/ \left(\frac{1}{n} \sum_i x_i^2 \right)^2 \right\} +$$
$$+ \frac{\sigma^2}{n} \left[1 + \left(12 \frac{1}{n} \sum_i x_i^2 \right)^{-1} \right]$$

and eventually

$$(16)\quad \max_{f \in W_k} Q(f, X) = \eta \left[1 + \max \left\{ \sum_j \left[\frac{1}{n} \sum_i e_{2j}(x_i) \right]^2, \right. \right.$$
$$\left. \left. \frac{1}{12} \sum_j \left[\frac{1}{n} \sum_i x_i e_{2j+1}(x_i) \right]^2 \Big/ \left(\frac{1}{n} \sum_i x_i^2 \right)^2 \right\} \right] +$$
$$+ \frac{\sigma^2}{n} \left[1 + \left(12 \frac{1}{n} \sum_i x_i^2 \right)^{-1} \right].$$

We notice that the regret depends only on the values

$$\frac{1}{n}\sum_i x_i^{2j}, \quad j = 1, \ldots, s = [(k+1)/2].$$

For simplicity we normalize them

$$\mu_j = \frac{1}{n}\sum_i (2x_i)^{2j}, \quad j = 1, \ldots, s.$$

Let $\mu_0 = 1$ and $\mu = \mu(X) = (\mu_0, \ldots, \mu_s)$. When the constant terms are omitted, the criterion function (16) takes the form

$$(17) \qquad \Phi_k(\mu) = \eta \max\Big\{\sum_{2j\leqslant k}(4j+1)\Big(\sum_{i=0}^{j} p_{2j,2i}\,\mu_i\Big)^2, \\ \frac{4}{3\mu_1^2}\sum_{2j+1\leqslant k}(4j+3)\Big(\sum_{i=0}^{j} p_{2j+1,2i+1}\,\mu_{i+1}\Big)^2\Big\} + \frac{\sigma^2}{n}\frac{1}{3\mu_1},$$

where, as in (14), the $p_{j,i}$ are coefficients of Legendre polynomials. Unfortunately, the solution of the minimization problem depends strongly on the number of experiments n through constraints on the set of moment vectors $M(n) = \{\mu(X)\colon\ -\frac{1}{2} \leqslant x_i \leqslant \frac{1}{2},\ i = 1, \ldots, n\}$. The trouble is avoided, if one considers all continuous designs whose moments form the classical moment space of polynomials on the unit interval

$$(18) \qquad M = \{\mu \in \boldsymbol{R}^{s+1}\colon\ \mu_j = \mathrm{E}_\zeta x^j,\ 0 \leqslant j \leqslant s,\ \zeta \text{ is a probability on } [0, 1]\}.$$

For this problem we use some well-known properties of M. For the sake of completeness they are collected in Theorem 2. Then we shall try to answer, whether the minimal point $\hat{\mu}$ of (17) over (18) belongs to the set $M(n)$. Because $M(n) \subset M(l\cdot n)$, $l = 1, 2, \ldots,$ we are also interested in the determination of small integers n such that $\hat{\mu} \in M(n)$. The design $X = (x_1, \ldots, x_n)$ such that $\mu(X) = \hat{\mu}$ can be applied in the construction of optimal strategies of experiments, the sample size of which is a multiple of n.

THEOREM 2 (see [3], Chapter 4). (i) *M from (18) is bounded, closed and convex.*

(ii) *For every $\mu \in M$ there exists a distribution with the support consisting of at most $[s/2]+1$ points such that it generates μ.*

(iii) *$\mu \in \partial M$ iff it is generated by exactly one distribution.*

(iv) *$\mu \in \operatorname{int} M$ iff the matrices*

$$(19) \qquad \begin{aligned} \underline{\Delta}_s &= [\mu_{s-2[s/2]+i+j-2}]_{1\leqslant i,j\leqslant [s/2]+1}, \\ \bar{\Delta}_s &= [\mu_{s-2[(s+1)/2]+i+j-1} - \mu_{s-2[(s+1)/2]+i+j}]_{1\leqslant i,j\leqslant [(s+1)/2]} \end{aligned}$$

are positive definite.

So, the best continuous design exists for every n, σ^2 and η. In the polynomial case Theorem 1 can be expressed in terms of moments of design distribution.

COROLLARY 1. *If a design ξ satisfies*

$$E_\xi x^{2j} = E_{m_t} x^{2j}, \quad j = 1, \ldots, s = [(k+1)/2], \tag{20}$$

where $t = t(\eta, \sigma^2, n)$ is as in (7), then, for the class of disturbances W_k, ξ is minimax with respect to Q in the set of all symmetric continuous designs on R^1.

Observe that the number of conditions (20) guaranteeing global optimality appears about four times less than in the general situation. Let us consider now when they are fulfilled. As before, k is the greatest possible degree of polynomial disturbances and $\eta_0 = \frac{25}{162}\frac{\sigma^2}{n}$ is a certain particular point connected with Huber's design.

THEOREM 3. (i) *For each $k = 2, 3, \ldots$ there is $\bar{\eta}_k \in (0, \eta_0)$ such that for every $\eta \geqslant \bar{\eta}_k$ and for m_t defined in (7) and (8) there exists a design satisfying (20) (and for $\eta < \bar{\eta}_k$ there is not).*

(ii) *If $\eta > \bar{\eta}_k$, then there are an integer n and a discrete design $X = (x_1, \ldots, x_n)$ as good as the best continuous one.*

(iii) *The sequence $\{\bar{\eta}_k, k \geqslant 2\}$ is nondecreasing and tends to η_0.*

Proof. (i) Of course, for $\eta \geqslant \eta_0$ the conclusion holds for m_t. The optimal moment point $\mu(t)$ with coordinates

$$\mu_j(t) = \frac{5jt - 3(j-1)}{(2j+1)(2j+3)}, \quad j = 0, 1, \ldots, s \tag{21}$$

for $t = t(\eta_0, \sigma^2, n) = 9/5$ belongs to the interior of M, because it is generated by the measure with infinite support. Therefore, by continuity

$$\mu\big(t(\eta, \sigma^2, n)\big) \in M \quad \text{for some } \eta < \eta_0. \tag{22}$$

The intersection of the convex set M and the straight line $\mu(t)$ is a segment. Thus, (22) holds if and only if $\eta \geqslant \bar{\eta}_k$ for some $\bar{\eta}_k$.

(ii) Let us fix any point of M. Its coordinates, according to Theorem 2 (ii), may be presented in explicit form

$$\sum_{i=1}^{r} \alpha_i (2x_i')^{2j} \quad \text{for some } r, \alpha_i, x_i', \ i = 1, \ldots, r.$$

So, there are integer sequences $\{p_{i,n}, n \geqslant 1\}$, $i = 1, \ldots, r$, $\sum_{i=1}^{r} p_{i,n} = n$ such that

$$\sum_{i=1}^{r} \frac{1}{n} p_{i,n} (2x_i')^{2j} \to \sum_{i=1}^{r} \alpha_i (2x_i')^{2j}, \quad j = 0, \ldots, s.$$

Hence, in particular, the increasing sequences of sets

$$M(2^q) = \{(1, 2^{-q} \sum_{i=1}^{2^q} (2x_i)^2, \ldots, 2^{-q} \sum_{i=1}^{2^q} (2x_i)^{2s}): -\tfrac{1}{2} \leqslant x \leqslant \tfrac{1}{2}\}, \quad q \geqslant 1,$$

tends to M and for large q contains any inner point of M.

The remaining statement (iii) is also obvious.

Remark. $\bar{\eta}_k$ can be calculated by means of the recurrent formula

$$\eta_k = \frac{5}{2} t_s^2 (t_s - 1) \frac{\sigma^2}{n},$$

where $s = [(k+1)/2]$, $t_1 = 3$ and t_s, $s \geqslant 2$, is the greatest, less than t_{s-1}, zero of the polynomial $\det \underline{\Delta}_s \det \bar{\Delta}_s$ with $\underline{\Delta}_s$, $\bar{\Delta}_s$ and μ_j as in (19) and (21), respectively.

For small η there are also discrete designs not worse than optimal continuous strategies.

THEOREM 4. *For every $k \geqslant 2$ there exists $\underline{\eta}_k \in (0, \eta_0)$ such that for $\eta \leqslant \underline{\eta}_k$ a symmetric plan concentrated at the pair $\pm 1/2$ is minimax. Furthermore,*

$$\underline{\eta}_{2k-1} = \underline{\eta}_{2k} \leqslant \frac{2}{3k(k+1)(2k+1)(2k+3)} \frac{\sigma^2}{n}.$$

Proof. Let k be fixed and $s = [(k+1)/2]$. The loss function (17) may be written shortly

$$\eta \max\{f_1(\mu), f_2(\mu)\} + \frac{\sigma^2}{n} \frac{1}{3\mu_1}.$$

We construct first a one-dimensional, smooth minorant of $f_1(\mu)$ on the neighbourhood of $\mathbf{1} = (1, \ldots, 1) \in \boldsymbol{R}^{s+1}$. Let

$$\varphi_{2j}(\mu) = \sum_{i=0}^{j} p_{2j,2i} \mu_i, \quad 2j \leqslant k,$$

and

$$\underline{\varphi}_{2j}(\mu_1) = \varphi_{2j}(\nu),$$

where $\nu_0 = 1$, $\nu_1 = \mu_1$,

$$\nu_i = \begin{cases} \mu_1^i & \text{if } p_{2j,2i} > 0, \\ \mu_1 & \text{otherwise,} \end{cases} \quad i = 2, \ldots, s.$$

We notice that for every j and $\mu \in M$

$$\underline{\varphi}_{2j}(\mu_1) \leqslant \varphi_{2j}(\mu) \quad \text{and} \quad \underline{\varphi}_{2j}(\mathbf{1}) = \varphi_{2j}(\mathbf{1}) = P_{2j}(1) = 1.$$

Since $\underline{\varphi}_{2j}$ are continuous, there exists $\mu_1^* < 1$ such that for every j and all $\mu_1 \geqslant \mu_1^*$

$$0 \leqslant \underline{\varphi}_{2j}(\mu_1) \leqslant \varphi_{2j}(\mu),$$

which implies that for each $\mu \in M$ with $\mu_1 \geqslant \mu_1^*$

$$f_1(\mu) = \sum_j (4j+1)\varphi_{2j}^2(\mu) \geqslant \sum_j (4j+1)\underline{\varphi}_{2j}^2(\mu_1).$$

The right-hand side is a polynomial and has a bounded, less than a certain $A > 0$, derivative.

If $\eta < \eta(A) = \frac{1}{3A}\frac{\sigma^2}{n}$, then for every $\mu_1 \in (\mu_1^*, 1)$

$$\frac{d}{d\mu_1}\left[\eta \sum_j (4j+1)\underline{\varphi}_{2j}^2(\mu_1) + \frac{\sigma^2}{n}\frac{1}{3\mu_1}\right] < 0$$

and, for all $\mu \in M$ with $\mu_1 \in (\mu_1^*, 1)$, we have

$$\begin{aligned}\eta f_1(\mu) + \frac{\sigma^2}{n}\frac{1}{3\mu_1} &\geqslant \eta \sum_j (4j+1)\underline{\varphi}_{2j}^2(\mu_1) + \frac{\sigma^2}{n}\frac{1}{3\mu_1} \\ &> \eta \sum_j (4j+1)\underline{\varphi}_{2j}^2(1) + \frac{\sigma^2}{n}\frac{1}{3} = \eta f_1(1) + \frac{\sigma^2}{n}\frac{1}{3}.\end{aligned}$$

If $\eta \leqslant \eta(\mu_1^*) = \frac{1-\mu_1^*}{3\mu_1^* f_1(1)}\frac{\sigma^2}{n}$, then for all $\mu \in M$, such that $\mu_1 \leqslant \mu_1^*$, we have

$$\eta f_1(\mu) + \frac{\sigma^2}{n}\frac{1}{3\mu_1} > \frac{\sigma^2}{n}\frac{1}{3\mu_1} \geqslant \eta f_1(1) + \frac{\sigma^2}{n}\frac{1}{3}.$$

Therefore, if $\eta \leqslant \underline{\eta}_k = \min\{\eta(A), \eta(\mu_1^*)\}$, then $\eta f_1(\mu) + \frac{\sigma^2}{n}\frac{1}{3\mu_1}$ is minimized at the point 1.

It remains to show that $f_2(1) \leqslant f_1(1)$. Indeed,

$$f_1(1) = j(2j+3) \quad \text{for} \quad k = 2j \quad \text{and} \quad k = 2j+1,$$

$$f_2(1) = \begin{cases} \frac{1}{3}(j-1)(2j+3) & \text{if} \quad k = 2j, \\ \frac{1}{3}j(2j+5) & \text{if} \quad k = 2j+1. \end{cases}$$

Finally, if $\underline{\eta}_{2k-1}$ or $\underline{\eta}_{2k}$ were less than $\frac{2}{3k(k+1)(2k+1)(2k+3)}\frac{\sigma^2}{n}$, then

$$\frac{d}{d\mu_1}\left[\eta f_1(1, \mu_1, \ldots, \mu_1^k) + \frac{\sigma^2}{n}\frac{1}{3\mu_1}\right]\Bigg|_{\mu_1 = 1} > 0$$

and for some $\mu_1 < 1$ a design with the whole mass at $\pm\frac{1}{2}\sqrt{\mu_1}$ would be better.

Several elements of the sequences $\{\bar{\eta}_k\}$ and $\{\underline{\eta}_k\}$ are presented in Table 1.

TABLE 1

k	2	4	6	8	...	∞
$\bar{\eta}_k \Big/ \frac{\sigma^2}{n}$	0.0222	0.0660	0.0892	0.1075	...	0.1543
$\underline{\eta}_k \Big/ \frac{\sigma^2}{n}$	0.0222	$3.17 \cdot 10^{-3}$	$8.82 \cdot 10^{-4}$	$? \leqslant 3.37 \cdot 10^{-4}$	...	0

For odd k the values are the same as for $k+1$.

The above proofs suggest that Theorems 3 and 4 hold for more general functions than classical polynomials. If η belongs to the interval $(\underline{\eta}_k, \bar{\eta}_k)$, then the solution lies on the border of the moment space and the design distribution is uniquely determined. One can then hardly expect that it might be represented as a discrete design. For small k, however, we show below that the best design is a symmetric pair close to $\pm\frac{1}{2}$, when η is slightly greater than $\underline{\eta}_k$.

3.2. Small degree polynomial cases. The exact solutions for polynomial regression of low degree $k \leqslant 6$ are shown below. Elementary methods were used here, so only final results are presented. Unfortunately, for $k > 4$ all needed formulae cannot be described in an exact analytic way.

Case of quadratic disturbances $k = 2$. The regret function (17) depends only on one variable μ_1 and, as usual, on the ratio $\eta/(\sigma^2/n)$. Let $t(\eta, \sigma^2, n)$ be defined implicitly, as in (7).

The best choice is $\mu_1 = \min\{\frac{1}{3}t, 1\}$ and a design, whose second moment $(1/n)\sum_i x_i^2$ equals $\frac{1}{4}\mu_1$, is minimax. The most simple one concentrates at two points

$$x_i = \begin{cases} \pm\dfrac{1}{2} & \text{if } \eta \leqslant \dfrac{1}{45}\dfrac{\sigma^2}{n}, \\ \pm\sqrt{t/12} & \text{otherwise.} \end{cases}$$

The relation between the parameters of the problem and this design (the positive part, in fact) describes the thin curve in the Figure 1.

Cubic disturbances $k = 3$. The solution is the same as for $k = 4$, but there are other ones, too. This is caused by the form of regret (17) expressed by the maximum of two functions. The smaller of them (at the optimal point) has one more variable than the other. This free value can vary as long as its function remains smaller than the other.

The remark concerns also other odd k.

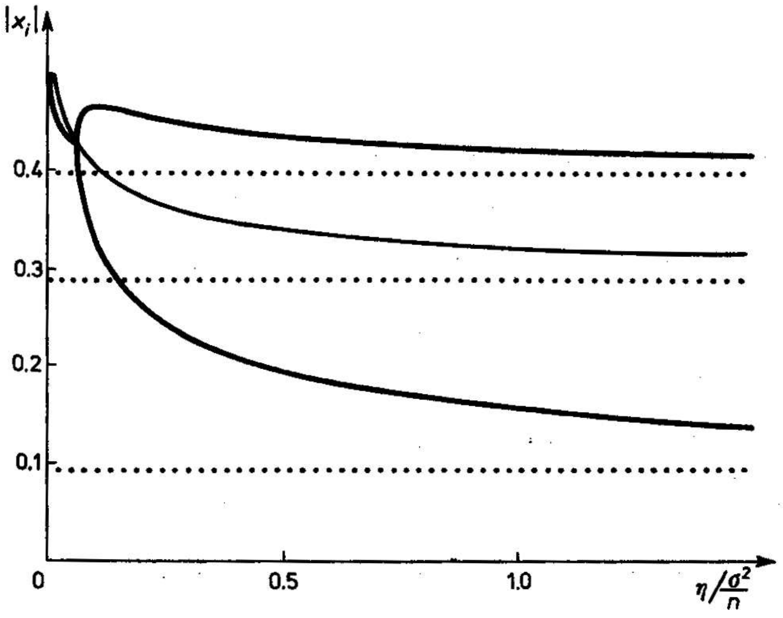

Fig. 1. The robust design for the polynomial regression functions of degree $k = 2$ (thin line) and $k = 4$ (thick lines) as a function of the ratio $\eta/(\sigma^2/n)$. The dotted lines represent asymptotic values for $\sigma^2/n = 0$

Case $k = 4$. The quality of design is determined by two factors μ_1 and μ_2 (or equivalently, $(1/n)\sum_i x_i^2$ and $(1/n)\sum_i x_i^4$). They are optimal when

$$(\mu_1, \mu_2) = \begin{cases} (1, 1) & \text{if} \quad \eta \leqslant \underline{\eta}_4, \\ (a, a^2) & \text{if} \quad \underline{\eta}_4 \leqslant \eta \leqslant \bar{\eta}_4, \\ \left(\dfrac{t}{3}, \dfrac{10t-3}{35}\right) & \text{if} \quad \eta \geqslant \bar{\eta}_4, \end{cases}$$

where $a \in [0.742, 1]$ is the root of the equation

$$\frac{315}{16} a^2 (105a^3 - 135a^2 + 51a - 5) = \frac{\alpha^2}{n\eta},$$

t is defined by relation (7) and $\underline{\eta}_4$, $\bar{\eta}_4$ are as in Table 1. The best design with 'minimal support' consists of one or two (for $\eta > \bar{\eta}_4$) pairs of sample points

$$x_i = \begin{cases} \pm\dfrac{1}{2} & \text{if} \quad \eta \leqslant 3.17 \cdot 10^{-3} \dfrac{\sigma^2}{n}, \\ \pm\dfrac{1}{2}\sqrt{a} & \text{if} \quad 3.17 \cdot 10^{-3} \dfrac{\sigma^2}{n} < \eta \leqslant 0.066 \dfrac{\sigma^2}{n}, \\ \pm\dfrac{1}{2}\sqrt{\dfrac{t}{3} \pm \sqrt{-t^2 + \dfrac{18}{7}t - \dfrac{3}{35}}} & \text{if} \quad \eta > 0.066 \dfrac{\sigma}{n}. \end{cases}$$

The thick lines in Figure 1 illustrate them.

Studden [4], who compared the extra-risk for information matrices $M_k(\xi) = [\mathrm{E}_\xi x^{i+j}]_{0\leqslant i,j\leqslant k}$ of the asymptotic solution $\hat{\xi}$ to the best $M_k(\hat{\xi}_k)$ with respect to the D-optimality criterion function $|\det M_k(\xi)|$, where $\hat{\xi}$ is a weak limit of $\hat{\xi}_k$. Let us do the same for the function Φ_k of vector $\mu \in M \subset \mathbf{R}^{s+1}$. Suppose that $\hat{\mu}$ minimizes $\Phi_k(\mu)$ over M and $\mu_t = (1, \mathrm{E}_{m_t}(2x)^2, \ldots, \mathrm{E}_{m_t}(2x)^{2s})$ for $\eta \geqslant \eta_0$ or $\mu_c = (1, \mathrm{E}_{m_c}(2x)^2, \ldots, \mathrm{E}_{m_c}(2x)^{2s})$ otherwise represents the 'asymptotic' solution. One should not expect it to be as good as the strategies specially constructed with regard to Φ_k. Nevertheless, for $\eta \geqslant \eta_0$ it is. In the consequence of Corollary 1, $\Phi_k(\mu_t) = \Phi_k(\hat{\mu})$ for each k. So, looking for a design with moments of m_t is a rewarding effort.

The statement is never true for $\eta < \eta_0$ and the corresponding design m_c. Namely, $\Phi_k(\mu_c) > \Phi_k(\hat{\mu})$. It turns out, however, that the relative extra-risk $[\Phi_k(\mu_c) - \Phi_k(\hat{\mu})]/\Phi_k(\hat{\mu})$ is for $\eta < \eta_0$ and even small k almost negligible (see Table 2).

TABLE 2. The relative extra-risk $[\Phi_k(\mu_c) - \Phi_k(\hat{\mu})]/\Phi_k(\hat{\mu})$

$\eta/(\sigma^2/n)$	0.01	0.02	0.05	0.10	0.15	0.1543
$k = 2$	$2.56 \cdot 10^{-2}$	$1.98 \cdot 10^{-2}$	$3.69 \cdot 10^{-3}$	$1.87 \cdot 10^{-4}$	$5.5 \cdot 10^{-8}$	0
$k = 4$	$2.49 \cdot 10^{-3}$	$4.44 \cdot 10^{-3}$	$5.51 \cdot 10^{-3}$	$9.61 \cdot 10^{-3}$	$1.1 \cdot 10^{-7}$	0

Acknowledgement. The author is very grateful to Ryszard Zieliński for suggesting the problem and for his valuable advice during preparation of this paper.

References

[1] P. J. Huber, *Robustness and design*; in: J. N. Srivastava (Ed.), *A survey of statistical design and linear models*, North-Holland, Amsterdam 1975, p. 287-303.

[2] —, *Robust statistics*, J. Wiley, New York 1981.

[3] S. Karlin and W. J. Studden, *Tchebycheff systems: with applications in analysis and statistics*, J. Wiley, New York 1966.

[4] J. Kiefer and W. J. Studden, *Optimal designs for large degree polynomial regression*, Ann. Statist. 4 (1976), p. 1113-1123.

MATHEMATICAL INSTITUTE
POLISH ACADEMY OF SCIENCES
00-656 WARSZAWA

Received on 1984.08.08