

ON A NONPARAMETRIC APPROACH TO PROBLEMS OF DESIGNING REGRESSION EXPERIMENTS

B. L. Granovskii, S. M. Ermakov

1968

SovietRxiv

View the original and related papers at <https://sovietrxiv.org/items/ru-196801.02913>

Source: Math-Net.Ru and CyberLeninka. Machine translation. Verify with the original.

Abstract

Full Text

UDC 519.281

MATHEMATICS

B. L. GRANOVSKII, S. M. ERMAKOV

ON A NONPARAMETRIC APPROACH TO PROBLEMS OF DESIGNING REGRESSION EXPERIMENTS

(Presented by Academician Yu. V. Linnik on 8 VII 1967)

Let $\{U, \sigma, P\}$ be a probability space and let $g(Q, u) \equiv \xi(Q)$ be a random function from a class H of measurable Hilbert random functions, defined for all Q from some measurable space $\{X, A, \mu\}$ with σ -finite measure μ , such that $Eg(Q, u) = f(Q) \in L_2(X, A, \mu)$, and let a set of orthonormalized functions $\{\varphi_i(Q)\}_0^n$ be fixed in $L_2(X, A, \mu)$. It is further assumed that, for any $Q \in X$, the result of observing $g(Q, u)$ of the unknown function $f(Q)$ can be obtained. In papers by J. Kiefer and J. Wolfowitz ⁽¹⁾, under the assumption that $f(Q)$ is a linear combination of $\varphi_i(Q)$

$$\left(f(Q) = \sum_{i=0}^n a_i \varphi_i(Q) \right),$$

the problem was considered of choosing observation points $Q_0, Q_1, \dots, Q_N$ (an experimental design) with the aim of optimally estimating, in some sense, the regression coefficients a_i . The solution of this problem, for a given N , is a fixed set of points $Q_0^*, Q_1^*, \dots, Q_N^*$, called an optimal design.

In contrast to this parametric problem, below we consider the problem of designing an experiment under the assumption that $f(Q)$ is an arbitrary function from $L_2(X, A, \mu)$. In this case it is natural to estimate the polynomial of best approximation to the function $f(Q)$ with respect to the system $\{\varphi_i(Q)\}_0^n$, which leads to a nonparametric problem of mathematical statistics. This new approach necessitates the consideration of randomized designs. The problems of their selection were initially considered in connection with minimizing variances in the Monte Carlo method ⁽²⁻⁴⁾.

1. Denote by

$$a_j[f] = \int_{(X)} f(Q) \varphi_j(Q) \mu(dQ)$$

the Fourier coefficient of the function f ; set $N = n$ and compute $\hat{a}_j$ by solving the system of linear algebraic equations

$$\sum_{i=0}^n \hat{a}_i \varphi_i(Q_j) = \xi(Q_j), \quad j = 0, \dots, n,$$

i.e., according to the interpolation formula,

$$\hat{a}_j = \hat{a}_j(\xi) = \frac{(-1)^j}{\Delta} \det \|\varphi_0(Q_i), \dots, \varphi_{j-1}(Q_i), \xi(Q_i), \varphi_{j+1}(Q_i), \dots, \varphi_n(Q_i)\|_0^n,$$

where $\Delta = \det \|\varphi_i(Q_j)\|_0^n$ is the determinant of the indicated system—assumed to be nonzero. In what follows we assume that, for all $g(Q, u)$ from H ,

$$\int_{(X)} E_u g^2(Q, u) \mu(dQ) < \infty.$$

Definition 1. We shall call a function $G(Q_0, \dots, Q_n)$ of the joint probability distribution of $Q_0, \dots, Q_n$, which for all $g(Q, u)$ from H and the corresponding $f(Q)$ satisfies the following conditions, a **randomized experimental design** (r.e.d.):

A. $E \hat{a}_j^{(G)}(g) = a_j(f).$

B. $D \hat{a}_j^{(G)}(g) < +\infty, \quad j = 0, 1, \dots, n.$

Here: 1) $G(Q_0, \dots, Q_n)$ is regarded as given on $\{X^{(n+1)}, A^{(n+1)}, \mu^{(n+1)}\}$ —the $(n+1)$ -fold product of $\{X, A, \mu\}$ with itself; 2) $\hat{a}_j^{(G)}(\xi)$ denotes $\hat{a}_j(\xi)$ corresponding to the r.p.e. G ; 3) $G(Q_0, \dots, Q_n)$ is assumed symmetric with respect to its arguments, which does not diminish generality; 4) the symbol E everywhere denotes the symbol of full mathematical expectation, i.e. $E = E_u E_G$, where E_G denotes averaging over $\{X^{(n+1)}, A^{(n+1)}, \mu^{(n+1)}\}$ with distribution function G , and E_u is averaging over the space $\{U, \sigma, P\}$ of realizations $\xi(Q)$.

Theorem 1. *In order that a distribution function G having density V be an r.p.e., it is necessary and sufficient that:*

$$1) \quad (n+1) \int_{(X^{(n)})} \frac{1}{\Delta} \Delta_0^j(Q_0, \dots, Q_n) V(Q_0, \dots, Q_n) \mu(dQ_1) \cdots \mu(dQ_n) = \varphi_j(Q_0)$$

for all $Q_0 \in X(\text{mod } \mu)$.

$$2) \quad \frac{1}{\Delta^2} V(Q_0, \dots, Q_n) < \infty \quad \text{for all } Q_0, \dots, Q_n \in X^{(n+1)}(\text{mod } \mu^{(n+1)}),$$

where Δ_0^j is the algebraic complement of the element $\varphi_j(Q_0)$ of the determinant Δ .

The proof of the theorem follows from the assumption of symmetry of G .

Corollary. *The probability density for the marginal distribution of the points $Q_0, \dots, Q_n$ of an r.p.e. is equal to*

$$\frac{1}{n+1} \sum_{i=0}^n \varphi_i^2(Q).$$

Definition 2. As in (4), we shall call the system $\{\varphi_i(Q)\}_0^n$ **regular** in $\{X, A, \mu\}$ if $\varphi_i(Q)$ ($i = 0, \dots, n$) are linearly independent on every measurable subset of X of nonzero measure, and **irregular** in the contrary case.

It is easy to indicate an r.p.e. for a given system $\{\varphi_i(Q)\}_0^n$. In particular, the following holds.

Theorem 2 (a generalization of the results of (2)). *The function $W(Q_0, \dots, Q_n)$, defined by the equality*

$$dW = \frac{\Delta^2}{(n+1)!} \mu(dQ_0) \cdots \mu(dQ_n),$$

is an r.p.e. Moreover:

$$\begin{aligned} 1) \quad D\hat{a}_j^{(W)}(\xi) &\leq \int_{(X)} \left[f(Q) - \sum_{i=0}^n a_i[f] \varphi_i(Q) \right]^2 \mu(dQ) + \int_{(X)} \sigma^2(Q) \mu(dQ) - \\ &- \sum_{i=0}^n \int_{(X)} \int_{(X)} R(P, Q) \varphi_i(Q) \varphi_i(P) \mu(dQ) \mu(dP), \quad j = 0, 1, \dots, n, \end{aligned}$$

where the equality sign holds for regular systems,

$$\sigma^2(Q) = E_u |g(Q, u) - f(Q)|^2,$$

$R(P, Q)$ is the correlation function of $\xi(Q)$;

2) for regular systems the correlation coefficient

$$\rho[\hat{a}_j^{(W)}(\xi), \hat{a}_i^{(W)}(\xi)] = 0, \quad i, j = 0, 1, \dots, n; \quad i \neq j.$$

The proof of the theorem follows from the generalization to the case of a space with an arbitrary measure of the known identity (5)

$$\det \left\| \int_a^b f_i(x) g_j(x) dx \right\|_{i,j=1}^n = \frac{1}{n!} \int_a^b \cdots \int_a^b \det \|f_i(x_j)\|_{i,j=1}^n \times \\ \times \det \|g_i(x_j)\|_{i,j=1}^n dx_1 \cdots dx_n.$$

2. We next define the notion of admissibility of an r.p.e.

Definition 3. An r.p.e. G^* is called **admissible** in H if in H there does not exist an r.p.e. $\hat{G}$ dominating it, i.e. such that for all $\xi \in H$

$$D\hat{a}_j^{(G^*)}(\xi) \leq D\hat{a}_j^{(\hat{G})}(\xi), \quad j = 0, \dots, n,$$

and, moreover, for at least one ξ from H and at least one j , a strict inequality holds.

Lemma. In order that an r.p.e. be admissible in H , it is necessary and sufficient that it be admissible in $L_2(X, A, \mu)$.

The proof of the lemma follows from the fact that any realization $g(Q, u)$ belongs to $L_2(X, A, \mu)$.

From the lemma and the results of [4] it follows that

Theorem 3. If the system $\{\varphi_i(Q)\}_0^n$ is regular and the functions $\varphi_i(Q)\varphi_j(Q)$ do not form a complete system in $\{X, A, \mu\}$ for $i, j = 0, \dots, n$, then the r.p.e. W is not admissible in H .

In the case of a nonregular system $\{\varphi_i(Q)\}_0^n$ —a Haar system of functions, which is constructed in $L_2\{X, A, \mu\}$ quite analogously to the way it was constructed in [4] on a set with Lebesgue measure—the following is valid:

Theorem 4. For a Haar system of functions, every r.p.e. is admissible.

The proof can be obtained by transforming the expression for $D\hat{a}_j(\xi)$, taking into account the results of Theorem 1.

If the measure μ is concentrated on a finite set of M points, then the problem of constructing the probability density of an admissible r.p.e. can in a number of cases be reduced to a problem of linear programming. Indeed, here the unbiasedness conditions represent $M(n+1)$ linear equalities with C_M^{n+1} unknowns $V(Q_{i_1}, \dots, Q_{i_{n+1}})$, while $D\hat{a}_j(\xi)$ is a linear form with respect to these same unknowns. In addition, the inequalities

$$V(Q_{i_1}, \dots, Q_{i_{n+1}}) \geq 0$$

must be satisfied. Minimizing $D\hat{a}_j(\xi)$ for some j at fixed $\xi_0 = \xi$, we obtain the desired density, if it is unique. Otherwise the method of finding an admissible r.p.e. is more complicated (it may reduce to a sequence of linear programming problems). In the simplest case, when $M = n+2$, it follows from Theorem 1

that the r.p.e. W is the only possible one and, consequently, is admissible. In the general case, however, for a given system $\{\varphi_i(Q)\}_0^n$ the admissible r.p.e. is not unique.

The use of r.p.e.'s appears reasonable in setting up extensive experiments, when a priori information concerning $f(Q)$ is practically absent. The problems of describing all admissible r.p.e.'s and of constructing r.p.e.'s satisfying certain conditions (minimizing the variance of the random component of the function $g(Q, u)$, or dominating the r.p.e. W) are of obvious interest. For a measure concentrated on a finite set of points, these problems can be solved with the aid of computers.

In this paper the case $N = n$ was considered. For $N > n$, analogous problems may be posed in the same way as was done in [3].

The authors consider it their pleasant duty to express their gratitude to Yu. V. Linnik, who pointed out the applicability of the results of [2] to random processes, and also to A. M. Kagan and O. V. Shalaevsky for discussing the results obtained.

Leningrad Civil Engineering Institute
Leningrad State University named after A. A. Zhdanov

Received
16 VI 1967

REFERENCES

1. J. Kiefer, J. Wolfowitz, *Ann. Math. Stat.*, No. 2, **30** (1959).
2. S. M. Ermakov, V. G. Zolotukhin, *Theory of Probability and Its Applications*, **5**, No. 4 (1960).
3. S. M. Ermakov, *Trudy Mat. Inst. im. V. A. Steklova, Academy of Sciences of the USSR*, **79** (1965).
4. S. M. Ermakov, *DAN*, **172**, No. 2 (1967).
5. C. Andreief, *Mém. de la Soc. Sci. Bordeaux* (3), **2** (1883).

Note: Figure translations are in progress. See original paper for figures.

Source: Math-Net.Ru and CyberLeninka. Machine translation. Verify with the original.