

ON LINEAR ESTIMATES OF REGRESSION

A. A. Tempel'man

1970

SovietRxiv

View the original and related papers at <https://sovietrxiv.org/items/ru-197001.26483>

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

Abstract

Full Text

UDC 519.24

MATHEMATICS

A. A. TEMPEL' MAN

ON LINEAR ESTIMATES OF REGRESSION

(Presented by Academician Yu. V. Linnik on 19 VIII 1969)

The note is devoted to estimation, from observed realizations of a random function, of the unknown "regression coefficients" entering the expression for its mathematical expectation. In § 1 the connection is studied between Hilbert spaces generated by the random function and its extended correlation kernel. In §§ 2 and 3 the structure of regression estimates, their convergence and consistency are studied. In § 4 these results are applied to the study of regression estimates constructed by the method of least squares.

1. Let $\xi(t)$, $t \in T$, be a real random function (r.f.) on a probability space $(\Omega, \mathcal{B}, P)$ with mathematical expectation $M_P \xi(t) = m(t)$ and finite correlation kernel

$$R(s, t) = M_P(\xi(s) - m(s))(\xi(t) - m(t));$$

$H(\xi) = H_P(\xi)$ is the Hilbert space of real random variables (r.v.'s) with scalar product $(\eta, \zeta) = M_P \eta \zeta$, generated by the r.v.'s $\xi(t)$, $t \in T$; $H(R)$ is the Hilbert space with reproducing kernel $R(s, t)$ (see ^(1-4,8)); the scalar product in $H(R)$ will be denoted by $\langle f, g \rangle_{H(R)}$.

Lemma 1. *The condition $m(t) \in H(R)$ is necessary and sufficient for $H(\xi)$ to contain no nonzero nonrandom scalars.*

If $m(t) \in H(R)$, then in $H(\xi)$ one can introduce a new scalar product

$$(\eta, \zeta)^* = M_P(\eta - M\eta)(\zeta - M\zeta);$$

$H(\xi)$ with this scalar product will be denoted by $H^*(\xi)$. We shall say that the mapping of $H(R)$ onto $H(\xi)$ is canonical if, for every $t_0 \in T$, the function $R(s, t_0)$ corresponds to the r.v. $\xi(t_0) \in H(\xi)$. It is known that between $H(R)$ and $H^*(\xi)$ there exists a unique canonical isometry i ; for any $\varphi \in H(R)$ we denote

$$i(\varphi) = \langle \xi(t), \varphi(t) \rangle_{H(R)}^{**}.$$

Let $m \notin H(R)$ and

$$m(t) = \psi(t) + \vartheta(t),$$

where $\psi \in H(R)$ and $\vartheta \notin H(R)$. Consider the extended kernel

$$\widetilde{R}(s, t) = R(s, t) + \theta(s, t),$$

where $\theta(s, t) = \vartheta(s)\vartheta(t)$. We have

$$H(\widetilde{R}) = H(R) \oplus H(\theta).$$

Denote

$$\xi_\psi(t) = \xi(t) - \vartheta(t), \quad \xi_\vartheta(t) = \xi(t) - \psi(t);$$

obviously,

$$M_P \xi_\psi(t) = \psi(t), \quad M_P \xi_\vartheta(t) = \vartheta(t).$$

Let P_ϑ be a measure on $\mathcal{B}$ such that the P_ϑ -distribution of the r.f. $\xi(t)$ coincides with the P -distribution of the r.f. $\xi_\vartheta(t)$; $H^*(\xi)$ is the space $H(\xi)$ with the new scalar product

$$(\eta, \zeta)^* = M_P(\eta - M\eta)(\zeta - M\zeta) + M_{P_\vartheta} \eta M_{P_\vartheta} \zeta.$$

It is not difficult to verify that

$$H^*(\xi) = H^*(\xi_\psi) \oplus E^1,$$

where E^1 is the space of scalars. Let $\Pi_Q^{H(\widetilde{R})}$ be the projector in $H(\widetilde{R})$ onto the subspace Q ; for any function $\varphi \in H(\widetilde{R})$ put

$$\langle \xi, \varphi \rangle_{H(\widetilde{R})} = \langle \xi_\psi, \Pi_{H(R)}^{H(\widetilde{R})} \varphi \rangle_{H(R)} + \langle \vartheta(t), \Pi_{H(\theta)}^{H(\widetilde{R})} \varphi \rangle_{H(\theta)}. \quad (1)$$

Lemma 2. *The correspondence*

$$\varphi \leftrightarrow \langle \xi, \varphi \rangle_{H(\widetilde{R})}$$

is a canonical isometry between $H(\widetilde{R})$ and $H^*(\xi)$, under which $H(R)$ corresponds to $H(\xi_\psi)$ and $H(\theta)$ corresponds to E^1 .

Lemma 3. *For any function $\varphi \in H(\widetilde{R})$ we have*

$$M_P \langle \xi, \varphi \rangle_{H(\widetilde{R})} = \langle m, \varphi \rangle_{H(\widetilde{R})}.$$

* By kernels we mean nonnegative definite functions.

** The meaning of this symbol is discussed in ^(3,8).

2. Suppose now that the correlation function $R(s, t)$ is known, while the mathematical expectation

$$m(t) = \sum_{i=1}^n a_i \varphi_i(t),$$

where $n < \infty$ and $\varphi_i(t)$, $i = 1, \dots, n$, are known linearly independent functions, and a_i are unknown constants that must be estimated from the observed realization of the random function $\xi(t)$; in other words, it is known only that $m(t) \in M$, an n -dimensional linear space spanned by the functions $\varphi_1, \dots, \varphi_n$.

Let $\mathcal{B}(\xi)$ be the smallest σ -algebra of events with respect to which all random variables $\xi(t)$, $t \in T$, are measurable; $\mathcal{P}_{R,M}$ the family of all probability measures P on $\mathcal{B}(\xi)$ with respect to which the correlation kernel of the random function $\xi(t)$ is $R(s, t)$ and $M_P \xi(t) = m(t) \in M$; $\mathcal{G}_{R,M}$ the family of all probability measures P on $\mathcal{B}(\xi)$ with respect to which the correlation kernel of the random function $\xi(t)$ is $R(s, t)$ and $M_P \xi(t) = m(t) \in M$; $\mathcal{G}_{R,M}$ the family of all measures from $\mathcal{P}_{R,M}$ with respect to which the random function $\xi(t)$ is Gaussian, and $\mathcal{P}$ an arbitrary family of measures from $\mathcal{P}_{R,M}$. A random function $\hat{m}(t)$ is called the uniformly, with respect to $\mathcal{P}$, best unbiased regression estimate ($\mathcal{P}$ -u.b.u.r.e.) from the realization $\xi(t)$ on T , if:

- 1) $\hat{m}(t) \in \mathcal{B}(\xi)$ for every $t \in T$;
- 2) $M_P \hat{m}(t) = M_P \xi(t)$, if $P \in \mathcal{P}$;
- 3) $D_P \hat{m}(t) \leq D_P n(t)$ for all $t \in T$, $P \in \mathcal{P}$, and all random functions $n(t)$ with properties 1) and 2).

If here we replace property 1) by property 1') $\hat{m}(t) \in H_P(\xi)$ for every $t \in T$, for every measure $P \in \mathcal{P}$, then we obtain the definition of the uniformly best unbiased linear regression estimate ($\mathcal{P}$ -u.b.u.l.r.e.). Estimates with properties 1') and 2) are called unbiased linear regression estimates ($\mathcal{P}$ -u.l.r.e.).

Consider the linear space $\bar{H} = H(R) + M$ (the linear sum of the spaces $H(R)$ and M) and introduce in it a scalar product which coincides with $\langle \varphi, \psi \rangle_{H(R)}$ if $\varphi, \psi \in H(R)$. Put $M_\psi = H(R) \cap M$, $M_\vartheta = \bar{H} \ominus H(R)$. Let $\psi_1, \dots, \psi_k$ and $\vartheta_1, \dots, \vartheta_l$ be orthonormal bases in M_ψ and M_ϑ . In view of the decomposition $M = M_\psi \oplus M_\vartheta$, we have for $m(t)$ the representation

$$m(t) = \sum_{i=1}^k b_i \psi_i(t) + \sum_{i=1}^l c_i \vartheta_i(t), \quad (2)$$

where b_i and c_i are unknown constants and $k + l = n$. Denote:

$$\theta(s, t) = \sum_{i=1}^l \vartheta_i(s) \vartheta_i(t); \quad \widetilde{R}(s, t) = R(s, t) + \theta(s, t);$$

$$\xi_0(t) = \xi(t) - m(t); \quad \xi_\psi(t) = \xi_0(t) + \sum_{i=1}^k b_i \psi_i(t); \quad \xi_\vartheta(t) = \xi_0(t) + \sum_{i=1}^l c_i \vartheta_i(t).$$

It is easy to see that $M_\vartheta = H(\theta)$, $\bar{H} = H(\tilde{R}) = H(R) \oplus H(\theta)$ and $M = H(\tilde{R}_M^R)$, where

$$\tilde{R}_M^R(s, t_0) = \Pi_M^{H(R)} \tilde{R}(s, t_0)$$

for every $t_0 \in T$. For any function $\varphi \in H(\tilde{R})$ denote:

$$\langle \xi, \varphi \rangle_{H(\tilde{R})} = \langle \xi_0, \Pi_{H(R)}^{H(\tilde{R})} \varphi \rangle_{H(R)} + \langle m, \varphi \rangle_{H(\tilde{R})} = \langle \xi_\psi, \Pi_{H(R)}^{H(\tilde{R})} \varphi \rangle_{H(R)} + \langle \vartheta, \Pi_{H(\theta)}^{H(\tilde{R})} \varphi \rangle_{H(\theta)}.$$

For $l = 1$, $c_1 = 1$, the mapping

$$\varphi \rightarrow \langle \xi, \varphi \rangle_{H(\tilde{R})}$$

coincides with the canonical isometry (1). For $l > 1$ it is also canonical, but not one-to-one, since the l -dimensional space $H(\theta)$ goes into the subspace E^1 . However, Lemma 3 remains valid in this case as well.

Theorem 1. The random function

$$\hat{m}_{R,M}(t) = \left\langle \xi(s), \Pi_M^{H(\tilde{R})} \tilde{R}(s, t) \right\rangle_{H(\tilde{R})}$$

is a $\mathcal{G}_{R,M}$ -u.b.u.r.e. and a $\mathcal{P}_{R,M}$ -u.b.u.l.r.e. It can be represented in the form

$$\hat{m}_{R,M}(t) = \sum_{i=1}^n \hat{a}_i \varphi_i(t)$$

or

$$\hat{m}_{R,M}(t) = \sum_{i=1}^k \hat{b}_i \psi_i(t) + \sum_{i=1}^l \hat{c}_i \vartheta_i(t),$$

where

$$\hat{b}_i = \langle \xi, \psi_i \rangle_{H(\tilde{R})} = \langle \xi_\psi, \psi_i \rangle_{H(R)}, \quad i = 1, \dots, k;$$

$$\hat{c}_i = \langle \xi, \vartheta_i \rangle_{H(\tilde{R})} = \langle m, \vartheta_i \rangle_{H(\tilde{R})} = c_i, \quad i = 1, \dots, l,$$

and the coefficients $\hat{a}_i$ satisfy the system of equations;

$$\sum_{j=1}^n \langle \varphi_i, \varphi_j \rangle_{H(\hat{R})} \hat{a}_j = \langle \xi, \varphi_i \rangle_{H(\hat{R})}, \quad i = 1, \dots, n;$$

$$D\hat{b}_i > 0 \quad \text{for } i = 1, \dots, k; \quad D\hat{c}_i = 0 \quad \text{for } i = 1, \dots, l^*.$$

This theorem generalizes a well-known result of Ya. Gaek⁽²⁾ and E. Parzen⁽⁸⁾, who considered the case where all $\varphi_i \in H(R)$.

A $\mathcal{P}_{R,M}$ -n.l.r.e. $\hat{n}(t)$ will be called **simple** if, for any measure $P \in \mathcal{P}_{R,M}$, with P -probability 1,

with P -probability 1,

$$\hat{n}(t) \equiv \sum_{i=1}^n \eta_i \varphi_i(t),$$

where $\eta_i \in H_P(\xi)$, $\varphi_i \in M$; obviously, if

$$M_P \xi(t) = \sum_{i=1}^n a_i \varphi_i(t),$$

then $M_P \eta_i = a_i$, i.e. η_i are unbiased estimates of the coefficients a_i . We shall say that a $\mathcal{P}_{R,M}$ -n.l.r.e. $\hat{n}(t)$ is **determined by the kernel** B if, for all $P \in \mathcal{P}_{R,M}$, with P -probability 1,

$$\hat{n}(t) \equiv \hat{m}_{B,M}(t)$$

($\mathcal{P}_{B,M}$ -r.n.n.l.r.e.)^{**}.

Theorem 2. $\mathcal{P}_{R,M}$ -n.l.r.e. is simple if and only if it is determined by some kernel.

3. We shall now determine which kernels determine $\mathcal{P}_{R,M}$ -n.l.r. estimates and what are the conditions for convergence and consistency of identical (defined by one and the same kernel) simple n.l.r.e. as the volume of observations increases.

Let A be a directed set; $T = \bigcup_{\alpha \in A} T_\alpha = T_{\alpha_T}$, where $T_\alpha, \alpha \in A$, is an increasing generalized sequence of sets: $T_{\alpha_1} \subseteq T_{\alpha_2}$ if $\alpha_1 < \alpha_2$; $\xi^\alpha(t)$, $R^\alpha(s, t)$, $m^\alpha(t)$ are the restrictions to the set T_α of the functions $\xi(t)$, $R(s, t)$, $m(t)$; $M^\alpha = \{\varphi : \varphi(t) = m^\alpha(t), m(t) \in M\}$. Denote by

$$\hat{m}_{R,M}(t)$$

the $\mathcal{P}_{R^\alpha, M^\alpha}$ -r.n.p.l.r.e. for the realization $\xi(t)$, $t \in T_\alpha$.

A sequence of $\mathcal{P}_{R^\alpha, M^\alpha}$ -n.l.r.e. $\hat{n}_\alpha(t)$ for the realization $\xi(t)$ on T_α will be called $\mathcal{P}_{R,M}$ -**consistent** if, for any measure $P \in \mathcal{P}_{R,M}$, there exists the limit in quadratic mean (P)

$$\text{l.i.m.}_{\alpha \in A} \hat{n}_\alpha(t) \equiv M_P \xi(t)$$

with P -probability 1.

Let $R(s, t)$ and $B(s, t)$ be two kernels on T . We shall say that the kernel $R(s, t)$ is **majorized** by the kernel $B(s, t)$ ($R \triangleleft B$) if $B(s, t) - R(s, t)$ is a kernel; the kernel $R(s, t)$ is **dominated** by the kernel $B(s, t)$ ($R \ll B$) if

$$R(t, t) \equiv \lim_{\Lambda \in \mathcal{U}_R^B} \Lambda(t, t),$$

where

$$\mathcal{U}_R^B = \{\Lambda : \Lambda \triangleleft R, \Lambda \triangleleft k_\Lambda B\}$$

is a directed set of kernels (with respect to the order relation $\triangleleft$).

Theorem 3. The condition $R \triangleleft kB$ (for some $k > 0$) is necessary and sufficient in order that, for any finite-dimensional space M and for any increasing sequence of sets $T_\alpha, \alpha \in A$ ($\bigcup_{\alpha \in A} T_\alpha = T_{\alpha_T} = T$), each of the following assertions be valid:

- 1) the estimates $\hat{m}_{B,M}^\alpha(t)$ are $\mathcal{P}_{R^\alpha, M^\alpha}$ -n.l.r.e. for the realization $\xi(t)$ on $T_\alpha (\alpha \in A)$;
- 2) if $P \in \mathcal{P}_{R,M}$, then, for any $t \in T$, there exists the limit

$$(P) \text{ l.i.m. } \hat{n}_{B,M}^\alpha(t) \equiv \hat{m}_{B,M}(t)$$

with P -probability 1;

- 3) if $M \cap H(B) = \{0\}$, then the sequence $\hat{m}_{B,M}^\alpha(t)$ is $\mathcal{P}_{R,M}$ -consistent;
- 4) if $B \ll R$ and the sequence $\hat{m}_{B,M}^\alpha(t)$ is $\mathcal{P}_{R,M}$ -consistent, then $M \cap H(B) = \{0\}$.

* The expansion (2) in which $\hat{b}_i$ and $\hat{c}_i$ have these properties was obtained by Yu. A. Rozanov ⁽⁵⁾ (this result was also independently obtained by A. V. Skorokhod and the author of these lines, who reported on it at the seminar on the statistics of random processes of I. G. Ignalin, June 1968).

** Estimates determined by kernels were introduced by Yu. A. Rozanov ⁽⁶⁾ under the name pseudo-best estimates.

A number of results in this direction have been obtained in works ⁽⁵⁻⁷⁾.

4. Let $T_\alpha, \alpha \in A$, be an increasing sequence of measurable sets in E^r ;

$$T = T_{\alpha_T} = \bigcup_{\alpha \in A} T_\alpha;$$

φ_i are measurable functions; $B(s, t) = \delta(t - s)$ is the Dirac function. The estimates defined by this kernel coincide with the estimates of $m(t)$ by the method of least squares (l.s.). We shall give several corollaries of Theorem 3, obtained by reformulating the conditions $R/k\delta, \delta \ll R$ and $M \cap H(\delta) = \{0\}$.

Corollary 1. *Let $\xi(t)$ be a continuous in the quadratic mean homogeneous random field with bounded spectral density. Then*

$$\text{l. i. m.}_{\alpha \in A} \hat{m}_{\delta, M}^{\alpha}(t) \equiv \hat{m}_{\delta, M}(t);$$

for the consistency of the sequence of l.s. estimates $\hat{m}_{\delta, M}^{\alpha}(t)$, it is sufficient that

$$\int_T |\varphi(t)|^2 dt = \infty$$

for any function $\varphi \in M$, $\varphi \neq 0$; if the spectral density is positive almost everywhere, this condition is also necessary (cf. (7), Theorem 1).

Now put $T = (a, b)$, $-\infty \leq a, b \leq +\infty$, and let $\xi(t)$ be a continuous in the quadratic mean Markov process in the broad sense; the correlation function of such a process has the form

$$R(s, t) = \sigma(s)\sigma(t) \min \tau(t)/\sigma(t),$$

where $\tau(t)/\sigma(t)$ is a nondecreasing nonnegative function.

Corollary 2. *If*

$$\sup_{a < t < b} \left\{ |\sigma(t)| \int_a^t |\tau(s)| ds + |\tau(t)| \int_t^b |\sigma(s)| ds \right\} < \infty,$$

then the sequence of l.s. estimates $\hat{m}_{\delta, M}^{\alpha}(t)$ converges in the quadratic mean for any finite-dimensional space M ; if, moreover,

$$\int_a^b |\varphi(t)|^2 dt = \infty$$

for all $\varphi \in M$, $\varphi \neq 0$, then this sequence is consistent.

Finally, consider the case of a discrete set of observations:

$$T = \{t_1, t_2, \dots\}, \quad T_{\alpha} = \{t_1, \dots, t_{\alpha}\}, \quad \alpha = 1, 2, \dots,$$

where the t_i are elements of an arbitrary set S . Let

$$R^{\alpha} = \|R(t_i, t_j)\|_{i, j=1}^{\alpha};$$

δ is the identity matrix; again $\hat{m}_{\delta, M}^{\alpha}$ is the l.s. estimate.

Corollary 3. *In order that, for any finite-dimensional space M , the limit*

$$\text{l. i. m.}_{\alpha \rightarrow \infty} \hat{m}_{\delta, M}^{\alpha}$$

exist, it is necessary and sufficient that the sequence of norms of the matrices R^{α} be bounded; if, in addition,

$$\sum_{i=1}^{\infty} |\varphi(t_i)|^2 = \infty$$

for $\varphi \in M$, $\varphi \neq 0$, then the sequence $\hat{m}_{\delta, M}^{\alpha}$ is consistent; if the matrices R^{α} are nonsingular, this condition is necessary for the consistency of l.s. estimates.

The apparatus of operator reproducing kernels (see ⁽³⁾) makes it possible, without changing the formulations, automatically to transfer all that has been said to vector-valued random functions and, in particular, to obtain estimates of regression coefficients from several (dependent or independent) realizations of the random function $\xi(t)$.

Institute of Physics and Mathematics
Academy of Sciences of the Lithuanian SSR
Vilnius

Received
7 VII 1969

CITED LITERATURE

1. H. Aronshajn, *Sborn. per. Matematika*, **7**, 2, 67 (1963).
2. J. Hájek, *Sborn. per. Matematika*, **7**, 3, 97 (1963).
3. Yu. I. Golosov, A. A. Tempelman, *DAN*, **184**, No. 6, 1271 (1969).
4. T. I. Mirskaya, A. S. Pabedinskaite, A. A. Tempelman, *Lit. matem. sborn.*, **7**, 3, 469 (1967).
5. Yu. A. Rozanov, *Gaussian Infinite-Dimensional Distributions*, "Nauka," 1968.
6. Yu. A. Rozanov, *Collection: The Soviet-Japanese Symposium on Probability Theory*, Khabarovsk, August, 1969, Publishing House of the Academy of Sciences of the USSR, 1969, p. 231.
7. A. S. Kholevo, *Theory of Probability and Its Applications*, **14**, 1, 78 (1969).

8. E. Parzen, *Proc. IV Berkeley Sympos.*, **1**, 469 (1961).

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

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