Volume 5, No. 3 (1996), pp. 343–356 Allerton Press, Inc.

MATHEMATICAL METHODS OF STATISTICS

ASYMPTOTIC NORMALITY OF THE ESTIMATOR OF AN INFINITE-DIMENSIONAL PARAMETER IN THE MODEL WITH A SMOOTH REGRESSION FUNCTION

A. G. Kukush

Kiev University
Vladimirskaya str. 64, 252017 Kiev, Ukraine

A model with C^1-smooth regression function in a Hilbert space is considered when the parameter space is an infinite-dimensional compact set. We construct an analogue of the least squares estimator for the regression parameter. Under some geometric conditions on the compact set we prove the asymptotic normality of the estimator, which generalizes finite-dimensional results due to B. L. S. Prakasa Rao. The convergence of finite-dimensional projections of the normalized estimator is proved using the weak convergence of a random field generated by the minimized functional.

Key words: Key words: nonlinear regression, infinite-dimensional parameter, least squares method, asymptotic normality

AMS 1991 Subject Classification: Primary 62G07; secondary 62G20

1. Introduction

We consider a nonlinear regression model

$$y_n = f_n(\theta_0) + \xi_n, \qquad 1 \le n \le N.$$

where parameter θ_0 belongs to a compact set in a Hilbert space, f_n are known regression functions with values in another Hilbert space H, and $\{\xi_n\}$ are i.i.d. H-valued random vectors. We study the least squares estimator for θ_0. Under certain regularity conditions the convergence in distribution of the normalized estimator to a Gaussian element as $N \to \infty$ is proven. Our conditions do not require the regression function to be twice differentiable.

There are many papers on estimation of finite-dimensional regression parameters motivated by diverse applications (see [1–4] and references therein). Estimation problems with infinite-dimensional parameters are also of interest for applications. These include estimation of a transfer function [5], estimation of a signal from noisy observations of its transforms [6], estimation of the intensity of an nonhomogeneous Poisson process [16], etc.

Let us discuss difficulties arising in asymptotic study of such estimators. When estimating a finite-dimensional parameter in nonlinear regression based on N observations, one introduces a certain random functional $Q_N(\theta)$ defined for θ from the parameter set Θ. If this set is compact and the functional Q_N is continuous a.s., then the least squares estimator (LSE) $\hat{\theta}_N$ is defined by

$$\min_{\theta \in \Theta} Q_N(\theta) = Q_N(\hat{\theta}_N). \tag{1.1}$$

For a Gaussian noise the maximum likelihood estimator is defined in a similar manner, by maximization of the corresponding functional. Unlike the linear model, none of the estimators can be expressed in an explicit form, which complicates their study.

The standard approach to the analysis of the LSE is as follows. Under the assumptions of identifiability and of stabilization in the average the estimator $\hat{\theta}_N$ is consistent. If the true parameter value is known to be an interior point of Θ and the regression functions are sufficiently smooth, consistency implies that with probability tending to 1

$$\operatorname{grad} Q_N(\hat{\theta}_N) = 0. \tag{1.2}$$

Then the asymptotic normality of $\hat{\theta}_N$ is deduced from (1.2) (see [1]). In a similar way the asymptotic normality of the MLE is proven [7].

When the parameter set is an infinite-dimensional compact set and the regression functions are continuous, (1.1) still determines an estimator which is consistent under suitable conditions [2]. However an infinite-dimensional compact set contains no interior points, and (1.1) does not imply (1.2). In [8, 9] convex compact parameter sets of some special form in a Hilbert space are considered. First from the convexity assumption an inequality is deduced, which is used to prove the relative compactness for distributions of the normalized estimators. Then for an everywhere dense set of directions h, equations

$$\langle Q_N'(\hat{\theta}_N), h\rangle = 0 \tag{1.3}$$

for the directional derivatives are considered. By utilizing the geometric properties of the compact sets under consideration and the consistency of the estimator, it is shown that (1.3) holds with probability tending to 1 as $N \to \infty$ for each h from the everywhere dense set. From (1.3) the asymptotic normality of linear functionals of the normalized estimator is inferred, which combined with relative compactness proves the asymptotic normality of $\hat{\theta}_N$.

In [12] these results are carried over to arbitrary convex compact sets in a Hilbert space. The geometric properties of these compact sets, such as, e.g., ellipsoids, do not allow obtaining equations like (1.3). In [12] the true parameter point θ_0 is supposed to possess a dense set of directions of shifts which do not take θ_0 away from Θ. Then certain directional derivatives of Q_N are shown to satisfy inequalities which enable one to establish the convergence for the distributions of linear functionals of the normalized estimator.

Let us also mention the paper [5] where Θ is a convex, closed, and bounded set in a separable reflexive Banach space, and the functional Q_N is assumed to be continuous in a stronger sense than continuity in the norm. In this setup, consistency of linear functionals of the LSE, asymptotic normality of certain functionals of LSE and weak convergence of LSE normalized by a Hilbert–Schmidt operator are proved. In [11] the parameter set is an ellipsoid in a Hilbert space and Q_N is determined through projections onto finite-dimensional subspaces whose dimension is growing with N. The orthogonal series LSE is introduced, with explicit Lagrange functions, and the asymptotics of the corresponding Lagrange multipliers is studied. The asymptotic normality of the estimators is derived with the aid of their representation through these multipliers.

In all the papers mentioned above, to obtain the asymptotic normality of estimators, the regression function is assumed to be at least twice differentiable. A finite-dimensional model with C^1-smooth regression function is considered in [12]. Without conditions on the second derivatives of the regression function the asymptotic normality cannot be deduced from (1.2). The goal is achieved by studying the random field

$$J_N(\varphi) := Q_N\Big(\theta_0 + \frac{\varphi}{\sqrt{n}}\Big) - Q_N(\theta_0), \qquad \|\varphi\| \le \tau, \tag{1.4}$$

with an arbitrary $\tau > 0$. This field is shown to weakly converge in the space of continuous functions to a random field $J(\varphi)$, which is a quadratic function of φ. From this, the convergence

$$\arg\min_{\|\varphi\|\le\tau} J_N(\varphi) \to \arg\min_{\|\varphi\|\le\tau} J(\varphi), \qquad N \to \infty,$$

in distribution in the Euclidean space is inferred, and this entails the asymptotic normality of the estimators.

In this paper we also require only C^1-smoothness of the regression functions. The parameter set Θ is a compact set in a Hilbert space satisfying certain geometric restrictions, which are not met, e.g., by an ellipsoid. The results nevertheless are applicable to a fairly broad class of compact sets. The weak convergence of linear functionals of estimators is deduced from that of certain finite-dimensional random fields generated by functionals Q_N. These fields resemble (1.4), but their form is adapted to the infinite-dimensional setup in that they are expressed through only some coordinates of $\hat{\theta}_N$.

We are not concerned here with the numerical computation of the estimator; the corresponding recurrent algorithms are given in [13]. The main results of the paper were announced in [14]. We refer to [17] for background in functional analysis to be used. Concerning Frechét derivatives of functions on linear normed spaces, see [18].

2. Construction of the Estimator

Let H be a real separable Hilbert space with inner product $(\cdot,\cdot)$ and norm $\|\cdot\|$, and let B be a real separable Hilbert space with inner product $\langle\cdot,\cdot\rangle$ and norm $|||\cdot|||$. With the corresponding Borel σ-algebras, H and B will be regarded as measurable spaces. We assume the following

Conditions:

(i) Θ is a compact subset of B.

(ii) For each $n \in \mathbb{N}$ a continuous function $f_n\colon \Theta \to H$ is given.

(iii) On a complete probability space $(\Omega, \mathcal{F}, P)$ a sequence $\{\xi_n, n \geq 1\}$ of i.i.d. H-valued centered random elements is given with $\sigma^2 := \mathbf{E}\|\xi_1\|^2 < \infty$.

Let for an unkonwn $\theta_0 \in \Theta$ the first N terms of the sequence

$$y_n := f_n(\theta_n) + \xi_n, \qquad n \geq 1,$$

be observed, and we have to estimate θ_0 based on these observations. Put

$$Q_N(\theta) := \frac{1}{N} \sum_{n=1}^{N} \|y_n - f_n(\theta)\|^2, \qquad \theta \in \Theta, \quad N \geq 1. \tag{2.1}$$

According to [8] under the conditions (i)–(iii) for each $N \in \mathbb{N}$ there exists a Θ-valued random element $\hat{\theta}_N$ such that

$$Q_N(\hat{\theta}_N) = \min_{\theta \in \Theta} Q_N(\theta)$$

for each $\omega \in \Omega$. Since it may be not unique, we choose and fix some version of $\hat{\theta}_N$. This element $\hat{\theta}_N$ will be called the LSE for θ_0.

The following two conditions ensure consistency of the LSE.

(iv) The limit

$$\varphi(\alpha, \beta) := \lim_{N \to \infty} \frac{1}{N} \sum_{n=1}^{N} \big(f_n(\alpha), f_n(\beta)\big)$$

exists, with convergence uniform in $(\alpha, \beta) \in \Theta \times \Theta$.

(v) $\Phi(\theta, \theta_0) > 0$ for all $\theta \neq \theta_0$, where

$$\Phi(\alpha, \beta) := \varphi(\alpha, \alpha) - 2\varphi(\alpha, \beta) + \varphi(\beta, \beta), \qquad (\alpha, \beta) \in \Theta \times \Theta.$$

Lemma 1 ([9]). *Let the conditions* (i)–(v) *be fulfilled. Then the estimator* $\hat{\theta}_N$ *is strongly consistent in the sense that* $|||\hat{\theta}_N - \theta_0||| \to 0$ *a.s.*

3. Main Conditions

Let $B(\theta, r)$ denote the open ball in B with center θ and radius r; let $\mathbf{0}$ stand for the zero element of B. The space H will be identified with H^* and so will be B and B^*. The operator norm of a linear operator A will be denoted by $\|A\|$. Let S be the correlation operator of ξ_1. We impose the following restrictions on Θ.

(vi) For a neighborhood $U(\theta_0)$ of θ_0, the set $U(\theta_0) \cap \Theta$ is convex.

(vii) There exists a linear subset of L everywhere dense in B such that

$$\forall\, \ell \in L \quad \exists \varepsilon > 0 \colon \theta_0 + \varepsilon \ell \in \Theta \cap U(\theta_0),$$

where $U(\theta_0)$ is the neighborhood as in (vi).

We will also need a stronger condition

(viii) There exists a linear subset of L everywhere dense in B such that

$$\forall\, \ell \in L \quad \exists \varepsilon > 0 \quad \exists r_\ell > 0 \quad \forall\, x \in B(\theta_0, r_\ell) \cap \Theta \colon x + \varepsilon \ell \in \Theta \cap U(\theta_0),$$

where $U(\theta_0)$ is the neighborhood as in (vi).

Condition (vii) means that for θ_0 there exists a dense set of directions of admissible shifts which do not take θ_0 away from Θ. Condition (viii) requires such directions to exist for all points from a neighborhood of θ_0 in Θ.

To exemplify these conditions, consider three classes of convex compact sets in a Hilbert space. Let us fix a basis $\{e_i, i \geq 1\}$ in B.

(a) Hilbert cube. Let $\{a_i, i \geq 1\}$ be a sequence of positive numbers such that $\sum_{i=1}^{\infty} a_i^2 < \infty$. Put

$$K := \{x \in B \mid \forall\, i \geq 1\colon |\langle x, e_i\rangle| \leq a_i\}. \tag{3.1}$$

(b) Ellipsoid. Let A be a positive compact operator in B and let $\{e_i\}$ be the basis of its eigenvectors. Put

$$E := \{x \in R(A)\colon |||A^{-1}x||| \leq 1\}.$$

(c) Let $\{r_i, i \geq 1\}$ be a nonincreasing sequence of positive numbers tending to zero. Put

$$C := \Big\{x \in B \mid \forall\, m \geq 1 : \sum_{i=m}^{\infty} \langle x, e_i\rangle^2 \leq r_m^2\Big\}. \tag{3.2}$$

Moreover, define the following analogues of interiors of these compact sets:

$$K_0 := \{x \in B \mid \forall\, i \geq 1\colon |\langle x, e_i\rangle| < a_i\};$$
$$E_0 := \{x \in R(A)\colon |||A^{-1}x||| < 1\};$$
$$C_0 := \Big\{x \in B \mid \forall\, m \geq 1\colon \sum_{i=m}^{\infty} \langle x, e_i\rangle^2 < r_m^2\Big\}.$$

If θ_0 belongs to K_0, E_0, or C_0, then all three compact sets satisfy condition (vii) with L being the set of all finite linear combinations of $\{e_i\}$. However only compact sets (3.1) and (3.2) satisfy (viii) (with the same L), whereas the ellipsoid does not satisfy (viii) for any $\theta_0 \in E$.

Next we impose smoothness conditions on functions $\{f_n\}$, which substitute for regularity conditions on regression functions in a finite-dimensional case.

(ix) For some set $U \supset \Theta$ open in B, the functions f_n, $n \geq 1$, can be extended to U so that their extensions are Frechét differentiable on U and moreover

$$\frac{1}{N} \sum_{n=1}^{N} f_n'^{*}(\theta_0) f_n'(\theta_0) \to V = V(\theta_0), \qquad N \to \infty,$$

in the uniform operator convergence sense, with $N(V) = \{\mathbf{0}\}$, $R(V) = B$.

Henceforth we assume that the basis $\{e_i\}$ lies in L as in (vii) or (viii). Put

$$V_m x := \sum_{i=m+1}^{\infty} \langle x, e_i\rangle e_i, \qquad m \geq 0, \quad x \in B.$$

Recall that σ^2 is defined in condition (iii).

(x) A. For some nonnegative numbers $\{\alpha_n(\theta_0), n \in \mathbb{N}\}$ such that

$$\limsup_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} \alpha_n^2(\theta_0) < \frac{1}{\sigma^2 \|V^{-1}\|^2},$$

the following inequality holds:

$$\exists r_0 > 0, \quad \forall n \geq 1, \quad \forall \theta \in B(\theta_0, r_0) \colon \|f_n'(\theta) - f_n'(\theta_0)\| \leq \alpha_n(\theta_0) |||\theta - \theta_0|||. \tag{3.3}$$

B. For some nonnegative numbers $\{\alpha_{mn}(\theta_0), m, n \in \mathbb{N}\}$ such that

$$\liminf_{m\to\infty} \limsup_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} \alpha_{mn}^2(\theta_0) < \frac{1}{\sigma^2 \|V^{-1}\|^2},$$

the following inequality holds:

$$\exists m_0 \geq 1, \quad \forall m \geq m_0, \quad \exists r_m, N(m), \quad \forall n \geq N(m) \quad \forall v \in \Theta - \theta_0, |||v||| < r_m :$$

$$\big\| \big(f_n'(\theta_0 + v) - f_n'(\theta_0)\big) V_m v \big\| \leq \alpha_{mn}(\theta_0) |||V_m v|||. \tag{3.4}$$

(xi) $\|V V_m - V_m V\| \to 0 \quad \text{as} \quad m \to \infty$.

Let us point out that the requirement A in (x) concerns the behavior of the derivatives f_n' not only in Θ, but also in a neighborhood of this parameter set.

We introduce two more conditions.

(xii) In the sense of operator convergence in the trace norm (see, e.g., [19])

$$\frac{1}{N} \sum_{n=1}^{N} f_n'^*(\theta_0) S f_n'(\theta_0) \to T = T(\theta_0), \qquad N \to \infty.$$

(xiii) For a linear subset M everywhere dense in B

$$\forall C > 0, h \in M \colon \frac{1}{N} \sum_{n=1}^{N} \int_{|\eta_n(h)| > C\sqrt{N}} |\eta_n(h)|^2 \, dP \to 0, \quad N \to \infty,$$

where $\eta_n(h) = (\xi_1, f_n'(\theta_0) h)$.

The last condition is an analogue of Lindeberg's condition which together with (iii) and (xii) ensures the convergence of B-valued random elements

$$\gamma_N := \frac{1}{\sqrt{N}} \sum_{n=1}^{N} f_n'^*(\theta_0) \xi_n, \qquad N \geq 1, \tag{3.5}$$

in distribution to a centered Gaussian element γ with the correlation operator T.

4. Tightness of Distributions of Normalized Estimators

Theorem 1. *Let conditions* (i), (iii)–(vi), (ix), (xii) *and the requirement* A *of* (x) *be fulfilled. Then the sequence* $\{\sqrt{N}|||\hat{\theta}_N - \theta_0|||, N \geq 1\}$ *is bounded in probability.*

We give a sketch of the proof. By Lemma 1 $\hat{\theta}_N(\omega) \to \theta_0$ as $N \to \infty$ in the norm of B for all $\omega \in \Omega_0$, where $P(\Omega_0) = 1$. Fix some $\omega \in \Omega_0$. There exists a number $N_0 = N_0(\omega)$ such that $\hat{\theta}_N(\omega) \in U(\theta_0)$ for $N \geq N_0$, where $U(\theta_0)$ is the neighborhood from condition (vi). Since the set $U(\theta_0) \cap \Theta$ is convex, for $N \geq N_0$ the function

$$q_N(t) := Q_N\big(\hat{\theta}_N + t(\hat{\theta}_0 - \hat{\theta}_N)\big), \quad t \in [0,1],$$

is well defined. It attains the minimum at zero, hence

$$q'_N(0) = \langle Q'_N(\hat{\theta}_N), \theta_0 - \hat{\theta}_N \rangle \geq 0. \tag{4.1}$$

Making use of conditions of Theorem 1 and inequality (4.1) one can estimate $\sqrt{N}|||\hat{\theta}_N(\omega) - \theta_0|||$ through $|||\gamma_N(\omega)|||$. But the sequence of random variables $\{|||\gamma_N|||, N \geq 1\}$ is bounded in probability by condition (xii), which implies the theorem. $\boxdot$

Theorem 2. *Let conditions* (i), (iii)–(vii), (ix)–(xii) *be fulfilled. Then the sequence of distributions of the normalized estimators* $\{\delta_N := \sqrt{N}(\hat{\theta}_N - \theta_0), N \geq 1\}$ *is relatively compact.*

Sketch of the proof. On account of Theorem 1, it suffices to show (see [10]) that for any $\varepsilon > 0$

$$\lim_{m\to\infty} \limsup_{N\to\infty} \mathbf{P}\{|||V_m \delta_N||| > \varepsilon\} = 0. \tag{4.2}$$

The requirement B of condition (x) implies that from any subsequence $\{m'\} \subset \{m\}$ one can select a further subsequence $\{p\} \subset \{m'\}$ such that

$$\lim_{p\to\infty} \limsup_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} \alpha^2_{p,n}(\theta_0) < \frac{1}{\sigma^2 \|V^{-1}\|}.$$

To establish (4.2) it suffices to show that along this subsequence

$$\lim_{p\to\infty} \limsup_{N\to\infty} \mathbf{P}\{|||V_p \delta_N||| > \varepsilon\} = 0. \tag{4.3}$$

Consider the random element $\hat{\theta}_N + V_p(\theta_0 - \hat{\theta}_N) = \theta_0 + \pi_p(\hat{\theta}_N - \theta_0)$, where $\pi_p := I - V_p$ with I denoting the identity operator in B. By Lemma 1 $\pi_p(\hat{\theta}_N - \theta_0) \to \mathbf{0}$ a.s. as $N \to \infty$. The basis $\{e_i\}$ lies in the set L from condition (vii), therefore $\pi_p(\hat{\theta}_N - \theta_0) \in L(e_1, \ldots, e_p) \subset L$. Hence by condition (vii) there exists a set $\Omega_0 \subset \Omega$ of probability 1 such that $\hat{\theta}_N + V_p(\theta_0 - \hat{\theta}_N) \in \Theta \cap U(\theta_0)$ for all $N \geq N_p(\omega)$, $\omega \in \Omega_0$. For a fixed $\omega \in \Omega_0$, let $\hat{\theta}_N \in U(\theta_0)$ for $N \geq N_p(\omega)$. Since $\Theta \cap U(\theta_0)$ is convex, the function

$$q_p(t) := Q_N\big(\hat{\theta}_N + tV_p(\theta_0 - \hat{\theta}_N)\big), \quad t \in [0,1],$$

is well defined for $\omega \in \Omega_0$, $N \geq N_p(\omega)$. It attains the minimum at the left endpoint $t = 0$, hence

$$q_p'(0) = \langle Q_N'(\hat{\theta}_N), V_p(\theta_0 - \hat{\theta}_N)\rangle \geq 0. \tag{4.4}$$

Making use of conditions of Theorem 2 and inequality (4.4) one can estimate $|||V_p\delta_N(\omega)|||$ through $|||V_p\gamma_N|||$ so that

$$\lim_{p\to\infty} \limsup_{N\to\infty} P\{|||V_p\delta_N||| > \varepsilon\} \leq \lim_{p\to\infty} \sup_{N\geq 1} P\{|||V_p\gamma_N||| > \varepsilon_1\} \tag{4.5}$$

for an ε_1 depending on ε. By conditions (iii), (xii) the right hand side of (4.5) is equal to zero, which implies (4.3) and hence the theorem. ⊡

5. Asymptotic Normality of the Estimators

In the proof of asymptotic normality the nontrivial part is the convergence of finite-dimensional distributions of normalized estimators. To that end we have to strengthen the requirements on the parameter set.

Theorem 3. *Let the conditions* (i), (iii)–(vi), (viii)–(xiii) *be fulfilled. Then*

$$\delta_N = \sqrt{N}(\hat{\theta}_N - \theta_0) \to V^{-1}\gamma \quad as \quad N \to \infty$$

in distribution in B, where γ is a centered Gaussian element with correlation operator T.

Proof. 1°. By Theorem 2 the sequence of distributions of $\{\delta_N\}$ is relatively compact. Moreover under the conditions (iii), (xii) and (xiii) γ_N (see (3.5)) converges in distribution to γ. Then the sequence of joint distributions of $(\delta_N; \gamma_N)$ is relatively compact in $B \times B$. Select a subsequence $\{(\delta_{N'}; \gamma_{N'})\}$ such that the corresponding distributions weakly converge to some measure μ on $B \times B$. Then on the same probablity space there exists a $B \times B$-valued random element $(\delta_\infty; \gamma_\infty)$ with distribution μ [15]. Note that γ_∞ and γ are identically distributed. For the proof of the theorem it suffices to show that δ_∞ and $V^{-1}\gamma$ are also identically distributed.

2°. For notational convenience assume that $(\delta_N; \gamma_N) \to (\delta_\infty; \gamma_\infty)$ in distribution for the entire sequence. Further, the operator V is invertible, and by transformation $\varkappa = V^{1/2}\theta$ one can introduce new regression functions $g_n(\varkappa) := f_n(V^{-1/2}\varkappa)$ with the new parameter space $V^{1/2}\Theta$ and the set of admissible shifts $V^{1/2}L$. The functions g_n satisfy all the conditions of the theorem with V replaced by the identity operator I, except for part B of condition (x). However this assumption has been used for the proof of relative compactness and will not be needed any longer. Hence without loss of generality we will assume the conditions of the theorem to hold with $V = I$.

For the convergence $\delta_N \to V^{-1}\gamma$ in distribution it suffices to prove that for any $m \in \mathbb{N}$

$$\pi_m\delta_N \to \pi_m V^{-1}\gamma = \pi_m\gamma \qquad \text{as} \quad N \to \infty \tag{5.1}$$

in distribution in $L_m := L(e_1, e_2, \dots, e_m)$.

3°. Put $\Delta\hat{\theta}_N := \hat{\theta}_N - \theta_0$ and fix some $\tau > 0$. For any $N \in \mathbb{N}$, $m \in \mathbb{N}$ define the random field $J_N(\varphi)$ for $\varphi \in B_m(\tau) := \{\varphi \in L_m\colon |||\varphi||| \leq \tau\}$ by

$$J_N(\varphi) := N\Big(Q_N\Big(\pi_m\theta_0 + \frac{\varphi}{\sqrt{N}} + V_m\hat{\theta}_N\Big) - Q_N(\theta_0)\Big), \tag{5.2}$$

if $\pi_m\theta_0 + \frac{\varphi}{\sqrt{N}} + V_m\hat{\theta}_N \in \Theta$ for all $\varphi \in B_m(\tau)$, and let $J_N(\varphi) := 0$ otherwise. In view of the strong consistency of the LSE, there is a set Ω_0 of probability 1 such that given a neighborhood of θ_0 the vector $\pi_m\theta_0 + V_m\hat{\theta}_N(\omega) = \theta_0 + V_m\Delta\hat{\theta}_N(\omega)$ belongs to this neighborhood for all $N \geq N_1(\omega)$. Therefore by conditions (vi), (viii) for all $N \geq N_2(\omega)$, one has $\pi_m\theta_0 + \frac{\varphi}{\sqrt{N}} + V_m\hat{\theta}_N(\omega) \in \Theta$ for any $\varphi \in B_m(\tau)$, in which case J_N is given by (5.2). Thus (5.2) holds with probability tending to 1 as $N \to \infty$, so that the convergence of distributions of J_N in $C(B_m(\tau))$ can be derived from the analysis of the representation (5.2).

4°. Expansion for J_N. By the definition (2.1) of Q_N one has

$$J_N(\varphi) = -2Z_N(\varphi) + D_N(\varphi), \tag{5.3}$$

where

$$Z_N(\varphi) := \sum_{n=1}^{N} \Big(f_n\Big(\theta_0 + \frac{\varphi}{\sqrt{N}} + V_m\Delta\hat{\theta}_N\Big) - f_n(\theta_0), \xi_n \Big), \tag{5.4}$$

$$D_N(\varphi) := \sum_{n=1}^{N} \Big\| f_n\Big(\theta_0 + \frac{\varphi}{\sqrt{N}} + V_m\Delta\hat{\theta}_N\Big) - f_n(\theta_0) \Big\|^2. \tag{5.5}$$

5°. Asymptotic behavior of D_N. Denote $f_n^0 := f_n(\theta_0)$, $f_n^{0\prime} := f_n'(\theta_0)$. By conditions (vi), (x) one has for sufficiently large N

$$\Big\| f_n\Big(\theta_0 + \frac{\varphi}{\sqrt{N}} + V_m\Delta\hat{\theta}_N\Big) - f_n(\theta_0) \Big\| = \Big\| f_n^{0\prime}\Big(\frac{\varphi}{\sqrt{N}} + V_m\Delta\hat{\theta}_N\Big) \Big\| + r_n(\varphi),$$

where

$$|r_n(\varphi)| \leq \alpha_n(\theta_0) \Big\| \frac{\varphi}{\sqrt{N}} + V_m\Delta\hat{\theta}_N \Big\|^2 \leq \frac{2\alpha_n(\theta_0)}{N}(\tau^2 + \||\delta_N|\|^2).$$

Put

$$D_N^1(\varphi) := \sum_{n=1}^{N} \Big\| f_n^{0\prime}\Big(\frac{\varphi}{\sqrt{N}} + V_m\Delta\hat{\theta}_N\Big) \Big\|^2.$$

Then (5.5) implies that

$$\max_{\varphi\in B_m(\tau)} \left|\sqrt{D_N(\varphi)}-\sqrt{D_N^1(\varphi)}\right| \le \left[\sum_{n=1}^N \frac{4\alpha_n^2(\theta_0)}{N^2}\right]^{1/2} (\tau^2+|||\delta_N|||^2) \xrightarrow{P} 0 \quad \text{as } N\to\infty.$$

Next, let

$$I_N := \frac{1}{N}\sum_{n=1}^N f_n'^*(\theta_0) f_n'(\theta_0).$$

Recall that by condition (ix) and by the construction in subsection 2° $I_N \to I$ in the operator norm. One has

$$\begin{aligned} D_N^1(\varphi) &= \langle I_N\varphi,\varphi\rangle + \langle I_N V_m\delta_N, V_m\delta_N\rangle + 2\langle I_N\varphi, V_m\delta_N\rangle \\ &= |||\varphi|||^2 + |||V_m\delta_N|||^2 + R_N^1(\varphi), \end{aligned}$$

with

$$\max_{\varphi\in B_m(\tau)} |R_N^1(\varphi)| \le \|I_N - I\|\big(\tau^2 + |||\delta_N|||^2 + 2\tau|||\delta_N|||\big) \xrightarrow{P} 0.$$

Thus

$$\sqrt{D_N(\varphi)} = \sqrt{D_N^1(\varphi)} + R_N^2(\varphi),$$

where $\max_{\varphi\in B_m(\tau)} |R_N^2(\varphi)| \to 0$, $N\to\infty$. Moreover, this representation implies

$$D_N(\varphi) = D_N^1(\varphi) + R_N^3(\varphi) = |||\varphi|||^2 + |||V_m\delta_N|||^2 + R_N^3(\varphi), \tag{5.6}$$

where also $\max_{\varphi\in B_m(\tau)} |R_N^3(\varphi)| \to 0$ as $N\to\infty$.

6°. Expansion for Z_N. The field (5.4) can be represented as

$$Z_N(\varphi) = Z_N^1(\varphi) + Z_N^2(\varphi),$$

where

$$\begin{aligned} Z_N^1(\varphi) &:= \sum_{n=1}^N \left(f_n^{0\prime}\Big(\frac{\varphi+V_m\delta_N}{\sqrt N}\Big), \xi_n\right) = \langle \gamma_N, \varphi + V_m\delta_N\rangle, \\ Z_N^2(\varphi) &:= \sum_{n=1}^N \left(f_n\Big(\theta_0 + \frac{\varphi+V_m\delta_N}{\sqrt N}\Big) - f_n^0 - f_n^{0\prime}\Big(\frac{\varphi+V_m\delta_N}{\sqrt N}\Big), \xi_n\right). \end{aligned}$$

7°. An auxiliary field. Let K be a compact set such that $\theta_0 + N^{-1/2}K \subset B(\theta_0, r_0)$ for any $N\in\mathbb{N}$, where r_0 is as in (3.3). Define the random field

$$W_N(\alpha) := \sum_{n=1}^N \left(f_n\Big(\theta_0 + \frac{\alpha}{\sqrt N}\Big) - f_n(\theta_0) - f_n^{0\prime}\frac{\alpha}{\sqrt N}, \xi_n\right), \qquad \alpha\in K, \quad N\in\mathbb{N}.$$

Let us show that

$$\max_{\alpha\in K} |W_N(\alpha)| \xrightarrow{P} 0, \qquad N\to\infty. \tag{5.7}$$

a) One has $W_N(\alpha) \xrightarrow{P} 0$ for any $\alpha \in K$, since in view of condition (x)

$$\operatorname{Var} W_N(\alpha) \le \sigma^2 \sum_{n=1}^{N} \Big\| f_n\Big(\theta_0 + \frac{\alpha}{\sqrt{N}}\Big) - f_n(\theta_0) - f_n^{0\prime} \frac{\alpha}{\sqrt{N}} \Big\|^2 \to 0 \quad \text{as} \quad N \to \infty.$$

b) Let us estimate the increment of W_N. One has for $\alpha, \beta \in K$

$$\begin{aligned} W_N(\alpha) - W_N(\beta) &= \sum_{n=1}^{N} \Big(f_n\Big(\theta_0 + \frac{\alpha}{\sqrt{N}}\Big) - f_n\Big(\theta_0 + \frac{\beta}{\sqrt{N}}\Big) - f_n^{0\prime} \frac{\alpha - \beta}{\sqrt{N}}, \xi_n \Big) \\ &= \sum_{n=1}^{N} \Big((\bar f_n' - f_n^{0\prime}) \frac{\alpha - \beta}{\sqrt{N}}, \xi_n \Big). \end{aligned}$$

Here $\bar f_n'$ is the value of f_n' at a point in the interval $[\theta_0 + \frac{\alpha}{\sqrt{N}}, \theta_0 + \frac{\beta}{\sqrt{N}}]$. Let $M := \max_{\alpha \in K} |||\alpha|||$. One has by condition (x)

$$\begin{aligned} |W_N(\alpha) - W_N(\beta)| &\le \frac{M|||\alpha - \beta|||}{N} \sum_{n=1}^{N} \alpha_n(\theta_0) \|\xi_n\| \\ &\le M|||\alpha - \beta||| \sqrt{\frac{1}{N} \sum_{n=1}^{N} \alpha_n^2(\theta_0)} \sqrt{\frac{1}{N} \sum_{n=1}^{N} \|\xi_n\|^2} \\ &= |||\alpha - \beta||| \cdot O_p(1) \end{aligned}$$

with $O_p(1)$ denoting a sequence of random variables bounded in probability.

c) By choosing an ε-net in K and using the relations obtained at the above steps, (5.7) is established by a routine argument.

8°. Asymptotics of Z_N^2. Since the sequence of random elements $\{V_m \delta_N, N \in \mathbb{N}\}$ is relatively compact, by Prohorov's theorem it is tight, i.e., for any $\varepsilon > 0$ there exists a compact set $K_\varepsilon \subset B$ such that for any N, $V_m \delta_N \in K_\varepsilon$ with probability at least $1 - \varepsilon$. Let $K := B_m(\tau) + K_\varepsilon$ to be understood as the set of elementwise sums. This set K is compact, and $\theta_0 + N^{-1/2} K \subset B(\theta_0, r_0)$ for sufficiently large N. Then for an arbitrary $\delta > 0$ one obtains from the results of subsection 7° that

$$\mathbf{P}\{\max_{\varphi \in B_m(\tau)} |Z_N^2(\varphi)| > \delta\} \le \varepsilon + \mathbf{P}\{\max_{\alpha \in K} |W_N(\alpha)| > \delta\}.$$

Hence

$$\limsup_{N \to \infty} \mathbf{P}\{\max_{\varphi \in B_m(\tau)} |Z_N^2(\varphi)| > \delta\} \le \varepsilon,$$

and, since $\varepsilon > 0$ is arbitrary,

(5.8) $$\max_{\varphi \in B_m(\tau)} |Z_N^2(\varphi)| \xrightarrow{P} 0 \quad \text{as} \quad N \to \infty.$$

Taking into account subsection 6°, we have

(5.9) $$Z_N(\varphi) = \langle \varphi \pi_m \gamma_N \rangle + \langle V_m \gamma_N, V_m \delta_N \rangle + Z_N^2(\varphi).$$

9°. Asymptotics of J_N. Making use of (5.3) and the representations (5.6), (5.8), (5.9), we obtain

$$J_N(\varphi) = -2\langle\varphi, \pi_m\gamma_N\rangle - 2\langle V_m\gamma_N, V_m\delta_N\rangle + |||\varphi|||^2 + |||V_m\delta_N|||^2 + R_N^4(\varphi), \tag{5.10}$$

where $\max_{\varphi\in B_m(\tau)} |R_N^4(\varphi)| \to 0$ as $N \to \infty$. In view of convergence of the joint distributions of $(\delta_N; \gamma_N)$ the finite-dimensional distributions of J_N converge to those of the random field

$$J_\infty(\varphi) := |||\varphi|||^2 - 2\langle\varphi, \pi_m\gamma_\infty\rangle + |||V_m\delta_\infty|||^2 - 2\langle V_m\gamma_\infty, V_m\delta_\infty\rangle \tag{5.11}$$

for $\varphi \in B_m(\tau)$.

Moreover, the distributions of each term in the right hand side of (5.10) are relatively compact in $C(B_m(\tau))$. We explain it in more detail for the first term. Let

$$\eta_N(\varphi) := \langle\varphi, \pi_m\gamma_N\rangle, \quad \varphi \in B_m(\tau), \quad N \in \mathbb{N}.$$

The finite-dimensional distributions of $\eta_N(\varphi)$ converge; next, since $\{|||\gamma_N|||\}$ is bounded in probability,

$$|\eta_N(\varphi_1) - \eta_N(\varphi_2)| \le |||\varphi_1 - \varphi_2||| \cdot O_p(1).$$

Hence the distributions of η_N are relatively compact in $C(B_m(\tau))$ (see [20]).

Thus it follows from (5.7) that the distributions of J_N are also relatively compact in $C(B_m(\tau))$. Therefore

$$J_N(\varphi) \to J_\infty(\varphi) \tag{5.12}$$

in distribution in $C(B_m(\tau))$.

10°. Convergence of the points of minimum. The last step of the proof uses standard methods of weak convergence theory (see, e.g., [20]). By (5.11) the limiting random field can be written in the form

$$J_\infty(\varphi) = |||\varphi - \pi_m\gamma_\infty|||^2 + \Delta_m,$$

where the random variable Δ_m does not depend on φ. This field attains its minimun at a single point $\varphi_\infty(\tau)$, which is the point of the ball $B_m(\tau)$ closest to the projection $\pi_m\gamma_\infty$. For large τ the point φ_∞ coincides with $\pi_m\gamma_\infty$ with large probability, so that

$$\varphi_\infty(\tau) = \pi_m\gamma_\infty + \beta_1(\tau),$$

where $\mathbf{P}\{\beta_1(\tau) \neq 0\} \to 0$ as $\tau \to +\infty$.

The convergence (5.12) and the uniqueness of the minimum point of the limiting field entail

$$\arg\min_{\varphi\in B_m(\tau)} J_N(\varphi) \to \varphi_\infty(\tau) \quad \text{as} \quad N \to \infty \tag{5.13}$$

in distribution. But when $\varphi = \pi_m\delta_N$, the argument of Q_N in (5.2), $\pi_m\theta_0 + \frac{\varphi}{\sqrt{N}} + V_m\hat\theta_N$, equals $\hat\theta_N$. Hence whenever $\pi_m\delta_N \in B_m(\tau)$ and $\pi_m\theta_0 + \frac{1}{\sqrt{N}}B_m(\tau) + V_m\hat\theta_N \subset$

Θ, the minimum of J_N is attained for $\varphi = \pi_m \delta_n$. Thus the minimum point of J_N can be represented as

$$\varphi_{\min}(N;\tau) = \pi_m \delta_N + \beta_2(N;\tau),$$

where

$$\lim_{\tau\to+\infty} \limsup_{N\to\infty} \mathbf{P}\{\beta_2(N,\tau) \neq \mathbf{0}\} = 0.$$

Finally, for any $\tau > 0$ (5.13) implies the convergence

$$\pi_m \delta_N + \beta_2(N;\tau) \to \pi_m \gamma_\infty + \beta_1(\tau) \quad \text{as} \quad N \to \infty.$$

Hence we conclude that for any bounded continuous function $f\colon B \to \mathbb{R}$

$$\mathbf{E} f(\pi_m \delta_N) \to \mathbf{E} f(\pi_m \gamma_\infty) \qquad \text{as} \quad N \to \infty,$$

i.e., $\pi_m \delta_N \to \pi_m \gamma_\infty$ as $N \to \infty$ in distribution in B. This implies (5.1), which proves the theorem. $\boxdot$

References

[1] R. I. Jennrich, *Asymptotic properties of nonlinear least squares estimators*, Ann. Math. Statist., 40 (1969), 633–643.

[2] A. Ya. Dorogovtsev, *Theory of Parameter Estimation for Random Processes*, Vishcha shkola, Kiev, 1982. (In Russian.)

[3] G. A. F. Seber and C. J. Wild, *Nonlinear Regression*, Wiley, New York, 1989.

[4] A. V. Ivanov and N. N. Leonenko, *Statistical Analysis of Ranodm Fields*, Kluwer, Dordrecht–Boston–London, 1989.

[5] A. Ya. Dorogovtsev, *Asymptotic normality of the least squares estimators of an infinite-dimensional parameter*, Ukrain. Mat. Zh., 45 (1993), 44–53. (In Russian.)

[6] A. S. Nemirovskii and R. Z. Khasmin'skii, *Nonparameteric estimation of functionals of signal derivatives observed in white noise*, Probl. Peredachi Inform., 23 (1987), 194–202.

[7] H. Cramér, *Mathematical Methods of Statistics*, Princeton Univ. Press, Princeton, 1974.

[8] A. Ya. Dorogovtsev, N. Zerek, and A. G. Kukush, *Asymptotic properties of nonlinear regression estimators in Hilbert space*, Theory Probab. Math. Statist., 35 (1987), 37–44.

[9] A. Ya. Dorogovtsev, N. Zerek, and A. G. Kukush, *Weak convergence of an infinite-dimensional parameter estimator to a normal distribution*, Theory Probab. Math. Statist., 37 (1988), 45–51.

[10] A. G. Kukush, *Convergence in distribution of a normalized estimate of an infinite-dimensional regression parameter*, Dokl. Akad. Nauk Ukrain. SSR, ser. A, No. 5, (1988), 11–14. (In Russian.)

[11] A. G. Kukush, *Asymptotic normality of the orthogonal series estimator of an infinite-dimensional parameter of nonlinear regression*, Ukrain. Mat. Zh., 45 (1993), 1213–1222. (In Russian.)

[12] B. L. S. Prakasa Rao, *Weak convergence of the least squares random field in the smooth case*, Statist. Decisions, 4 (1986), 363–377.

[13] J.-P. Aubin and I. Ekeland, *Applied Nonlinear Analysis*, Wiley, New York, 1984.

[14] A. G. Kukush, *Asymptotic normality of the estimator of infinite-dimensional parameter in the model with C^1-smooth regression function.* In: *Exploring the Stochastic Laws*, Festschrift in honour of the 70th birthday of Acad. V. S. Korolyuk; A. V. Skorokhod and Yu. V. Borovskikh, eds., The Netherlands, VSP, 1995, pp. 251–256.

[15] H. Engl and A. Wakolbienger, *On weak limits of probability distributions on Polish spaces*, Stochastic Anal. and Appl., 1 (1983), 197–203.

[16] A. Ya. Dorogovtsev and A. G. Kukush, *Asymptotic properties of the nonparametric estimator of intensity of a nonhomogeneous Poisson process*, Kibernetika i Sistemny Analiz, 1 (1996), 91–104.

[17] Y. M. Berezansky, Z. G. Sheftel, and G. F. Us, *Functional Ananlysis*, Vol. I, Operator Theory Advances and Applications, Vol. 85, Birkhäuser, Basel, 1996.
[18] H. Cartan, *Calcul Différentiel. Formes Différentielles*, Collection Méthods, Hermann, Paris, 1967.
[19] H. H. Kuo, *Gaussian Measures in Banach Spaces*, Lecture Notes in Math., Vol. 463, Springer, Berlin, 1975.

[*Received May* 1994, *revised June* 1995]