coordinates. The result is

$$\Pi_{1\alpha}(\xi_p)=F(C^TD_\alpha C)\mathbf{P}\{\chi^2_{k+2}\leqq\xi_p\},$$

where $F(C^TD_\alpha C)$ does not depend on ξ_p. The expressions for $\Pi_{i\alpha}(\infty)$, $i=1, 2, 3$, can be obtained by twice differentiating the identity

$$\frac{\sqrt{|A|}}{(2\pi)^{k/2}}\int e^{-Y^TAY/2}\,dY_1\cdots dY_k=1,$$

where one must take G^{-1} or $DG^{-1}D$ as the matrix A, and then make the substitution $Y=A^{-1}Z$ and set $u_1=1,\cdots,u_r=1$. The result is

$$\Pi_{1\alpha}(\infty)=-T_{1\alpha},\quad \Pi_{2\alpha}(\infty)=3T_{1\alpha}^2-4T_{1\alpha}-2T_{2\alpha},\quad \Pi_{3\alpha}(\infty)=T_{1\alpha}^2-2T_{1\alpha}-T_{2\alpha},$$

where $T_{1\alpha}$ and $T_{2\alpha}$ are defined in the formulation of the theorem.

We now write the final result:

$$\begin{aligned}\left(\frac{\partial^2L}{\partial u_\alpha^2}\right)_0&=(7T_{1\alpha}^2/4-T_{1\alpha}-3T_{2\alpha}/2)\mathbf{P}\{\chi^2_k\leqq\xi_p\}\\&\quad-(T_{1\alpha}^2/2)\mathbf{P}\{\chi^2_{k+2}\leqq\xi_p\}+(3T_{1\alpha}^2/4-T_{1\alpha}-T_{2\alpha}/2)\mathbf{P}\{\chi^2_{k+4}\leqq\xi_p\}\\&\quad-2T_{1\alpha}\mathbf{P}\{\chi^2_{k+2}\leqq\xi_p\}+(2T_{2\alpha}+4T_{1\alpha}-2T_{1\alpha}^2)\mathbf{P}\{\chi^2_{k+2}\leqq\xi_p\}\\&=\frac{\xi_p^{k/2}e^{-\xi_p/2}}{2^{k/2}\Gamma(k/2+1)}[(7T_{1\alpha}^2/4-T_{1\alpha}-3T_{2\alpha}/2)-\xi_p(3T_{1\alpha}^2-4T_{1\alpha}-2T_{2\alpha})/(4(k+2))].\end{aligned}$$

The expression obtained and the representation for $h(\xi_p,\sigma)$ indicated in the formulation of the theorem prove the relation (11). The theorem is proved.

A particular case of the result obtained for $k=1$ and $A_\alpha=(1,\cdots,1)^T$ with an investigation of the maximum and minimum of the difference

$$\mathbf{P}\{V(s)\leqq\xi_p[1+h(\xi_p,s)]\}-P$$

over all values of σ_α was considered in [2]. In concluding the author thanks the referee for his remarks.

Received by the editors
August 12, 1978

REFERENCES

[1] O. V. SHALAEVSKII, *Smoothing of observations with unknown weights by the method of least squares*, Proc. Steklov Inst. Math., 104 (1968), pp. 229–260. (In Russian.)

[2] V. I. PAGUROVA AND V. V. GURSKII, *A confidence interval for the common mean of several normal distributions*, Theory Prob. Appl., 24 (1979), pp. 882–888.

A GEOMETRICAL APPROACH TO MAXIMUM LIKELIHOOD ESTIMATION FOR INFINITE-DIMENSIONAL GAUSSIAN LOCATION. I

B. S. TSIREL'SON

(*Translated by A. B. Aries*)

In this paper we discuss the estimation of a shift of an infinite-dimensional Gaussian measure or, what is the same, estimation of the mean of a Gaussian process (a nonstationary process, with known covariance function), the estimated parameter belonging to a (known) infinite-dimensional set V. We consider here only a *maximum likelihood* estimator. Other estimators, more appropriate for an "extensive" set V, have been studied by I. A. Ibragimov and R. Z. Khas'minskii [1], [2]. We give necessary and sufficient conditions for existence, uniqueness and consistency of the maximum likelihood estimator. We formulate these conditions in terms of geometric characteristics of the set V, namely, mean width; V. N. Sudakov established its connection with a Gaussian measure, see for example [3]. All the geometrical notions used in this paper concern a "canonical" metric on V, uniquely defined by the given Gaussian measure. Making the most general assumptions, we obtained for the distance between the true value of the parameter and the maximum likelihood estimator inequalities which, in particular, imply that this distance has finite-order moments. As an example, we consider the estimation of a signal satisfying a Hölder condition of order α, in additive white noise; it turns out that the maximum likelihood estimator is applicable for $\alpha > 1/2$ but not for $\alpha \leqq 1/2$.

Assume given a linear space with a centered Gaussian measure (E, ν) and its kernel $E_0 \subset E$; for definitions see, for example, [3], pp. 13–14. As is well known, E_0 is a Hilbert space; for $\theta, \eta \in E_0$, a scalar product $\langle \theta, \eta \rangle$ and norm $\|\theta\| = \langle \theta, \theta \rangle^{1/2}$ are given and uniquely defined by the measure γ. For each $\theta \in E_0$ the linear functional $\langle \theta, \eta \rangle$ which is continuous in $\eta \in E_0$ has a unique (up to equality almost everywhere) extension to a linear functional $\langle \theta, x \rangle$, measurable in $x \in E$; here $\int \langle \theta, x \rangle^2 \gamma(dx) = \|\theta\|^2$. The measure γ_0 obtained from the measure γ by a shift vector θ has a density with respect to the measure γ if (and only if) $\theta \in E_0$; this density is given by the formula

$$\frac{\gamma_\theta(dx)}{\gamma(dx)} = \mathscr{L}(\theta, x) = \exp\left(\langle \theta, x \rangle - \tfrac{1}{2}\|\theta\|^2\right). \tag{1}$$

We shall estimate the shift parameter θ on the basis of one observation of $x \in E$. But our next statements can be applied to estimation by a sample if by x we mean the sample mean; the distribution of this mean is the Gaussian measure homothetic to the initial one. To simplify our discussion of consistency, we introduce a parameter $\sigma > 0$. In addition, let $\gamma_{\theta,\sigma}$ denote the image of the measure γ under the homothetic mapping $x \to \sigma x + \theta$. We assume that θ runs through a known set $V \subset E_0$ and σ is known, i.e., we consider the statistical structure $\mathscr{P}_{V,\sigma} = \{\gamma_{\theta,\sigma} : \theta \in V\}$ with the corresponding likelihood function

$$\mathscr{L}_\sigma(\theta, x) = \exp\left(\frac{1}{\sigma^2}\left(\langle \theta, x \rangle - \tfrac{1}{2}\|\theta\|^2\right)\right).$$

In the infinite-dimensional case, γ-almost every x does not lie in E_0 and hence $\langle \theta, x \rangle$ cannot be regarded as continuous in $\theta \in E_0$. We cannot even define $\langle \theta, x \rangle$ in a natural way for all $\theta \in E_0$ simultaneously; whether this can be done for all $\theta \in V$ simultaneously depends on the "value" of the set V (see [4]). It is sufficient now to go over to a separable modification of the process $\langle \theta, x \rangle$ and therefore of the process $\mathscr{L}_\sigma(\theta, x)$, x running through (E, γ) and θ running through V (equipped with a metric from E_0). Recall that V is called a *GB*-set if

$$\gamma\{x \in E : \sup_{\theta \in V} \langle \theta, x \rangle < +\infty\} = 1.$$

In this case, but only this case is

$$h_1(V) = (2\pi)^{1/2} \int \left(\sup_{\theta \in V} \langle \theta, x \rangle\right) \gamma(dx) \tag{2}$$

is finite; see, for example, [3], Chapter 2, Theorem 1 and Proposition 14. In order that the likelihood function $\mathscr{L}_\sigma(\theta, x)$ be bounded from above in $\theta \in V$, it is sufficient that V be a *GB*-set and necessary that any bounded part of the set V be *GB*. It is well known that a closed *GB*-set is compact. Even if V is *GB*-compact, the likelihood function may not attain its maximum on V. If (and only if) the compact set V has the *GC*-property (for the definition see, for instance, [3], p. 28), the likelihood function is continuous on V

with probability 1 and therefore there exists $\hat{\theta}(x)\in V$ such that

(3) $$\mathscr{L}_\sigma(\hat{\theta}(x), x)=\sup_{\theta\in V}\mathscr{L}_\sigma(\theta, x).$$

Is the point $\hat{\theta}(x)$ unique? Is it a consistent estimator? To what extent is the GC-property and the boundedness of the set V essential? The following theorem provides answers to these questions. First we introduce the quantity (possibly, equal to $+\infty$)

(4) $$C_1(V)=\overline{\lim_{r\to+\infty}}\frac{1}{r^2}(2\pi)^{-1/2}h_1(V\cap B(\theta, r));$$

here and below $B(\theta, r)=\{\eta\in E_0\colon \|\eta-\theta\|\leqq r\}$, where θ denotes a point of E_0; it is easy to show that $C_1(V)$ does not depend on θ.

Theorem 1. *Let $V\subset E_0$ be a closed set,*

$$C_1(V)<+\infty,\qquad \sigma>0,\qquad 4\sigma C_1(V)<1,\qquad \theta\in E_0;$$

then, for $\gamma_{\theta,\sigma}$-almost all $x\in E$,

(a) $\sup_{\eta\in V}\mathscr{L}_\sigma(\eta, x)<+\infty$;

(b) *$\hat{\theta}(x)\in V$ exists, is unique and is such that, for any $r>0$,*

(5) $$\sup_{\eta\in V\setminus B(\hat{\theta}(x), r)}\mathscr{L}_\sigma(\eta, x)<\sup_{\eta\in V}\mathscr{L}_\sigma(\eta, x).$$

Furthermore, for any $\theta\in V$ and any $r>0$,

(c) $\lim_{\sigma\to0+}\gamma_{\theta,\sigma}\{x\in E\colon \|\hat{\theta}(x)-\theta\|\leqq r\}=1$.

If, however, $C_1(V)=+\infty$, then, for any $\sigma>0$, $\theta\in E_0$, for $\gamma_{\theta,\sigma}$-almost all $x\in E$,

$$\sup_{\theta\in V}\mathscr{L}_\sigma(\theta, x)=+\infty.$$

REMARK 1. From (5) we have the following: if $\theta_n\in V$, $\lim_{n\to\infty}\mathscr{L}_\sigma(\theta_n, x)=\sup_{\theta\in V}\mathscr{L}_\sigma(\theta, x)$, then $\|\theta_n-\hat{\theta}(x)\|\to0$, $n\to\infty$.

REMARK 2. The point $\hat{\theta}(x)$ does not depend on σ; this can easily be deduced from the fact that $\sigma^2\log\mathscr{L}_\sigma(\theta, x)$ does not depend on σ.

We say that $\hat{\theta}(x)$ mentioned in the theorem is a maximum likelihood estimator in spite of the fact that (3) is, perhaps, violated with positive probability. (For the correct consideration of the probability of this event separable modification of the process $\langle\theta, x\rangle$ is not sufficient, a natural modification in the sense defined in [4] is required.) This term is appropriate if we mean by it a loss function which is continuous in $\hat{\theta}\in V$. Where necessary, we shall denote $\hat{\theta}(x)$ by the more precise $\hat{\theta}(x, V)$.

REMARK 3. If $V, V_1, V_2, \cdots$ are closed sets in E_0, $V_1\subset V_2\subset\cdots\subset V$, $V_1\cup V_2\cup\cdots$ is dense in V and $C_1(V)<+\infty$, then $\|\hat{\theta}(x, V_n)-\hat{\theta}(x, V)\|\to0$, $n\to\infty$, for $\lambda_{\theta,\sigma}$-almost all x, for any $\theta\in E_0$, $\sigma<1/(4C_1(V))$. This follows from Remark 1. There also are situations when $\lim_{n\to\infty}C_1(V_n)<+\infty$, but $C_1(V)=+\infty$, $\hat{\theta}(x, V_n)$ converging as $n\to\infty$ to a limit which can be used as an estimator of $\theta\in V$. This resembles somewhat the principal value of an improper integral. The sets V_n must be chosen in a special way, otherwise estimation can be very poor; see Example 3 below. For $C_1(V)=+\infty$ one should not use as an estimator the point near which the maximum likelihood function goes to infinity, even if such a point exists and is unique; one can show that the set of these points is defined only by the geometry of the set V and does not depend on either x, θ, or σ.

We turn now to inequalities. Since $C_1(V)$ gives only "asymptotic" information concerning the set V, we need the characteristic

$$C_2(V, \theta)=\sup_{r\geq1}\frac{1}{r^2}(2\pi)^{-1/2}h_1(V\cap B(\theta, r)).$$

It is clear that $C_2(V, \theta)<+\infty$ for some (or any) θ if and only if $C_1(V)<+\infty$. Furthermore, for a bounded V we set $R(V, \theta)=\sup_{\eta\in V}\|\eta-\theta\|$.

Theorem 2. (a) *Let $V \subset E_0$ be a closed set,*

$$C_1(V) < +\infty, \quad \theta \in V, \quad \sigma > 0, \quad 4\sigma C_2(V, \theta) < 1;$$

then

$$\gamma_{\theta,\sigma}\{x \in E : \|\hat{\theta}(x) - \theta\| \geqq r\} \leqq e^{-u}$$

for any positive r, u such that

$$r^2 = 32u\sigma^2 + 3.2C_2(V, \theta)\sigma + 81\sigma^2.$$

(b) *Let $V \subset E_0$ be GB-compact, $\theta \in V$, $\sigma > 0$; then*

$$\gamma_{\theta,\sigma}\{x \in E : \|\hat{\theta}(x) - \theta\| \geqq r\} \leqq e^{-u}$$

for any positive r, u such that

$$r^2 = 2\sqrt{2}\,\sqrt{u}R(V, \theta)\sigma + 2(2\pi)^{-1/2}h_1(V)\sigma.$$

REMARK 4. Under the conditions (b) of Theorem 2 we have: $R(V, \theta) \leqq h_1(V)$; hence we can replace the relation between r and u given there by

$$r^2 = 2\sqrt{2}\left(\sqrt{u} + \frac{1}{2\sqrt{\pi}}\right)h_1(V)\sigma.$$

Let us now consider several examples connected with the estimation of a signal in additive white noise; this problem has been studied in its geometrical aspect in [1], [2]. Let the measure γ correspond to white noise on [0, 1], and let E be a suitable space of generalized functions, for example, $(C_1[0, 1])^*$; then $E_0 = L_2[0, 1]$ with the usual scalar product. The maximal likelihood estimator makes sense if it is *a priori* known that the estimated signal θ (which is a function of $t \in [0, 1]$) belongs to a definite closed set $V \subset L_2[0, 1]$, $C_1(V) < +\infty$. The last assertion is clearly not satisfied if V contains all continuous (and therefore all measurable) functions θ such that $|\theta(t)| \leqq 1$ for all t; different conditions need to be imposed on θ.

EXAMPLE 1. The set V consists of all functions θ satisfying the Hölder condition $|\theta(s) - \theta(t)| \leqq M|s - t|^\alpha$ (α and M are fixed). We can show that $C_1(V) = +\infty$ for $\alpha \leqq \frac{1}{2}$, $C_1(V) < +\infty$ for $\frac{1}{2} < \alpha \leqq 1$; in what follows we consider only the second case. The set V is a sum of the one-dimensional space of constants and the set $V_1 = \{\theta \in V : \int_0^1 \theta(t)\,dt = 0\}$; it is easy to show that $\hat{\theta}(x, V) = \hat{\theta}(x, V_1) + 1_{[0,1]}\int_0^1 x(t)\,dt$, hence it is sufficient to consider $\hat{\theta}(x, V_1)$. The set V_1 is *GC*-compact; the maximal likelihood function is continuous on V_1 and attains a maximum exactly at one point $\hat{\theta}(x)$. One can show that

$$\frac{0.09}{2\alpha - 1}M \leqq (2\pi)^{-1/2}h_1(V_1) \leqq \left(\frac{0.94}{2\alpha - 1} - 0.16\right)M.$$

For $\alpha = 1$ one can find the exact value

$$(2\pi)^{-1/2}h_1(V_1) = \sqrt{\frac{\pi}{32}}M \approx 0.31M.$$

EXAMPLE 2. To estimate a discontinuous signal one can use functions of bounded variation. Let V consist of all functions on (0, 1) whose variation does not exceed M. As in the previous example, we go over from V to the set V_1 of functions, orthogonal to 1. We can show that V_1 is *GC*-compact, and $(2\pi)^{-1/2}h_1(V_1) = \sqrt{\pi/8}M \approx 0.63M$. In our next paper we shall prove that $\hat{\theta}(x)$ is a step function on (0, 1) with probability 1, independent of the properties of the "true" function θ.

EXAMPLE 3. Let V consist of all increasing square-summable functions on (0, 1). Then $C_1(V) = +\infty$; the likelihood function is not bounded from above on V. Consider the set $V_{a,b}$ of all functions $\theta \in V$ such that $a \leqq \theta(t) \leqq b$ for all $t \in (0, 1)$. We can show that $V_{a,b} \in GB$ and that the limit $\lim_{a \to -\infty, b \to +\infty} \hat{\theta}(x; V_{a,b})$ exists (in the metric $L_2[0, 1]$)

with probability 1; we denote this limit by $\hat{\theta}(x, V)$. In fact, the function $\hat{\theta}(x, V)$ is neither bounded from below nor from above on $(0, 1)$, and the function $\hat{\theta}(x, V_{a,b})$ is obtained from it by replacing all values lying beyond the interval $[a, b]$ by the values of this interval nearest to a or b. Note that the function $\hat{\theta}(x, V)$ can be constructed on the basis of a given observation x in a very simple way, namely, by the "method of stretched thread"; $\hat{\theta}(x, V)$ is a derivative (in t) of the lower convex enveloping function of $X(t)$, primitive for $x(t)$; see [5]. This enveloping function is a polygon whose vertices cluster toward the end-points of the interval $(0, 1)$.

The proof of Theorem 1 used the property of a Gaussian process to assume the maximum at a single point, which is of interest in itself (see, for example, [6], [7]).

Theorem 3. (a) *Let T be a metric compact set and let $\xi(t, \omega)$ be a Gaussian random process with realizations which are continuous in $t \in T$, $\mathbf{E}|\xi(t_1)-\xi(t_2)|^2>0$ for any distinct $t_1, t_2 \in T$. Then for almost each ω there exists a unique $t_\omega \in T$ such that $\xi(t_\omega, \omega) = \sup_{t\in T}\xi(t, \omega)$.*

(b) *Let (T, ρ) be a separable metric space; let $\xi(t, \omega)$ be a separable modification of a stochastically continuous Gaussian process, $\sup_{t\in T}\xi(t, \omega)<+\infty$ with probability 1. We say that the points $t_1, t_2 \in T$ are equivalent with respect to ξ, if $\xi(t_1, \omega)=\xi(t_2, \omega)$ for almost all ω. We define $T_{\max}(\omega)$ as the set of all points $t \in T$ such that for any neighborhood U of the point t the equality $\sup_{s\in U}\xi(s, \omega)=\sup_{s\in T}\xi(s, \omega)$ is satisfied. Then for almost each ω the set $T_{\max}(\omega)$ is such that all its points are equivalent with respect to ξ. (It is possible that $T_{\max}(\omega)$ is empty.)*

PROOF. Assertion (a) follows from (b); let us prove (b). We expand the given Gaussian process as a random function series of random variables:

$$\xi(t, \omega)=f_0(t)+\sum_{n=1}^{\infty}\zeta_n(\omega)f_n(t), \tag{6}$$

where f_n denote continuous functions on T and ζ_n denote independent Gaussian standard random variables; for each $t \in T$ the equality (6) is satisfied with probability 1. (It is well known (cf. [4]) that the series (6) converges with probability 1 at all points $t \in T$ simultaneously, but we do not employ this fact here.) Assume that for a given ω the set $T_{\max}(\omega)$ contains points t_1, t_2, not equivalent with respect to ξ. Then $f_n(t_1)\neq f_n(t_2)$ at least for one n. We assume that $f_n(t_1)<f_n(t_2)$; let rational a, b be such that $f_n(t_1)<a<b<f_n(t_2)$. Since $t_1, t_2 \in T_{\max}(\omega)$, we have

$$\sup_{t\in T, f_n(t)<a}\xi(t, \omega)=\sup_{t\in T}\xi(t, \omega)=\sup_{t\in T, f_n(t)>b}\xi(t, \omega).$$

Thus, it suffices to prove that for any n and any rational $a<b$

$$\sup_{t\in T, f_n(t)<a}\xi(t, \omega)=\sup_{t\in T, f_n(t)>b}\xi(t, \omega) \tag{7}$$

has zero probability. Invoking separability we now assume that T is at most countable. The space of elementary outcomes Ω can be identified with the space $\mathbf{R}^\infty$ of all numerical sequences with standard Gaussian measure, ζ_k are coordinate functionals on $\mathbf{R}^\infty$, $\xi(t, \omega)$ is given by (6). We have

$$\xi(t, \omega)=f_0(t)+\zeta_n(f_n(t)-a)+a\zeta_n+\sum_{m\neq n}\zeta_m f_m(t);$$

showing that, for fixed ζ_m with $m\neq n$,

$$\sup_{t\in T, f_n(t)<a}\xi(t, \omega)=a\zeta_n+(\text{a decreasing function of } \zeta_n);$$

similarly,

$$\sup_{t\in T, f_n(t)>b}\xi(t, \omega)=b\zeta_n+(\text{an increasing function of } \zeta_n);$$

it is clear that these suprema can coincide only for one value of ζ_n, that is with zero probability. Note that these considerations have something in common with those used in [8].

Lemma 1. *Let $V \subset E_0$ be GB-compact, $\theta \in E_0$, $\sigma > 0$, $r \geqq (2\sigma(2\pi)^{-1/2}h_1(V))^{1/2}$; then*

$$\gamma_{\theta,\sigma}\{x \in E: \sup_{\eta \in V \setminus B(\theta,r)} \mathcal{L}(\eta, x) \geqq \mathcal{L}(\theta, x)\}$$
$$\leqq \exp\left(-\frac{1}{2R^2(V,\theta)}\left(\frac{r^2}{2\sigma} - (2\pi)^{-1/2}h_1(V)\right)^2\right).$$

PROOF. We have

$$\log \mathcal{L}(\eta, \sigma y + \theta) - \log \mathcal{L}(\theta, \sigma y + \theta) = \langle \eta, \sigma y + \theta \rangle - \tfrac{1}{2}\|\eta\|^2$$
$$-\langle \theta, \sigma y + \theta \rangle + \tfrac{1}{2}\|\theta\|^2 = \sigma\langle \eta - \theta, y\rangle - \tfrac{1}{2}\|\eta - \theta\|^2 \leqq \sigma\langle \eta - \theta, y \rangle - \tfrac{1}{2}r^2$$

for any $\eta \in V \setminus B(\theta, r)$. Let $\varphi(y) = \sup_{\eta \in V\setminus B(\theta,r)} \langle \eta - \theta, y\rangle$; then

$$\gamma_{\theta,\sigma}\{x \in E: \sup_{\eta \in V \setminus B(\theta,r)} \mathcal{L}(\eta, x) \geqq \mathcal{L}(\theta, x)\}$$
$$= \gamma\{y \in E: \sup_{\eta \in V\setminus B(\theta,r)} \mathcal{L}(\eta, \sigma y + \theta) \geqq \mathcal{L}(\theta, \sigma y + \theta)\}$$
$$\leqq \gamma\{y \in E: \sigma\varphi(y) - \tfrac{1}{2}r^2 \geqq 0\}.$$

The function φ belongs to the class Lip $(\gamma, R(V, \theta))$, see [9], Example 3, on page 23. Hence, for any $a \geqq 0$,

$$\gamma\{y \in E: \varphi(y) - m \geqq aR(V,\theta)\} \leqq 2\int_a^\infty (2\pi)^{-1/2}\exp\left(-\frac{s^2}{2}\right) ds;$$

see [9], Corollary 1, p. 26; here $m = \int_E \varphi(y)\gamma(dy)$. It remains only to note that $m = (2\pi)^{-1/2} \times h_1(V \setminus B(\theta, r)) \leqq (2\pi)^{-1/2}h_1(V)$, take

$$a = \frac{1}{R(V,\theta)}\left(\frac{r^2}{2\sigma} - (2\pi)^{-1/2}h_1(V)\right)$$

and use the elementary inequality

$$2\int_a^\infty (2\pi)^{-1/2}\exp\left(-\frac{s^2}{2}\right) ds \leqq \exp\left(-\frac{a^2}{2}\right) \quad \text{for } a \geqq 0.$$

Lemma 2. *Let $V (V \subset E_0)$ be a closed set, $C_1(V) < +\infty$, $\theta \in E_0$, $\sigma > 0$. We partition V into "rings" $V_n = \{\eta \in V: \sqrt{8(n-1)\sigma(c \vee \sigma)} \leqq \|\eta - \theta\| \leqq \sqrt{8n\sigma(c\vee\sigma)}\}$, $n = 1, 2, \cdots$; here and below $c \vee \sigma$ denotes the largest among the numbers c, σ; we choose c later.*

(a) *If $4\sigma C_2(V,\theta) < 1$ and $c = \frac{1}{10}C_2(V,\theta)$, then, for any n,*

$$\gamma_{\theta,\sigma}\{x \in F: \sup_{\eta \in V_n} \mathcal{L}(\eta, x) \geqq \mathcal{L}(\theta, x)\} \leqq \exp\left(-\left(\frac{n}{4} - 1\right)\left(\frac{c}{\sigma} \vee 1\right)\right).$$

(b) *If $4\sigma C_1(V) < 1$, then there exist $c > 0$, $\varepsilon > 0$ and N such that, for any $n \geqq N$,*

$$\gamma_{\theta,\sigma}\{x \in E: (1+\varepsilon)\sup_{\eta \in V_n} \mathcal{L}(\eta, x) \geqq \mathcal{L}(\theta, x)\} \leqq 2\exp\left(-\left(\frac{n}{4} - 1\right)\left(\frac{c}{\theta} \vee 1\right)\right).$$

PROOF. First we prove (a). Application of Lemma 1 to the set V_n yields

$$\gamma_{\theta,\sigma}\{x \in E: \sup_{\eta \in V_n} \mathcal{L}(\eta, x) \geqq \mathcal{L}(\theta, x)\}$$
$$\leqq \exp -\frac{1}{16n\sigma(c\vee\sigma)}\left(\frac{8(n-1)\sigma(c\vee\sigma)}{2\sigma} - (2\pi)^{-1/2}h_1(V_n)\right)^2)$$

(the fact that the difference in the parentheses is positive will become clearer somewhat later). But

$$(2\pi)^{-1/2}h_1(V_n)\leqq C_2(V,\theta)(1\vee 8n\sigma(c\vee\sigma))$$
$$\leqq C_2(V,\theta)\left(1\vee\frac{8n(c\vee\sigma)}{4C_2(V,\theta)}\right)$$
$$=C_2(V,\theta)\vee 2n(c\vee\sigma)=10c\vee 2n(c\vee\sigma).$$

We assume that $n\geqq 5$; otherwise the inequality we are proving is trivial. We have: $(2\pi)^{-1/2}h_1(V_n)\leqq 2n(c\vee\sigma)$; hence

$$\gamma_{\theta,\sigma}\{x\in E:\sup_{\eta\in V_n}\mathscr{L}(\eta,x)\geqq\mathscr{L}(\theta,x)\}$$
$$\leqq\exp\left(-\frac{1}{16n\sigma(c\vee\sigma)}(4(n-1)(c\vee\sigma)-2n(c\vee\sigma))^2\right)$$
$$=\exp\left(-\frac{n^2-4n+4}{4n}\frac{c\vee\sigma}{\sigma}\right)\leqq\exp\left(-\left(\frac{n}{4}-1\right)\left(\frac{c}{\sigma}\vee 1\right)\right).$$

To prove (b) we note that under the conditions of Lemma 1, for any $\varepsilon>0$, the inequality

$$\gamma_{\theta,\sigma}\{x\in E:(1+\varepsilon)\sup_{\eta\in V\setminus B(\theta,r)}\mathscr{L}(\eta,x)\geqq\mathscr{L}(\theta,x)\}$$
$$\leqq\exp\left(-\frac{1}{2R^2(V,\theta)}\left(\frac{r^2}{2\sigma}-(2\pi)^{-1/2}h_1(V)-\frac{\varepsilon}{\sigma}\right)^2\right)$$

is satisfied; the proof of this fact is only slightly different from the proof of Lemma 1. Let c be such that $4\sigma C_1(V)<40\sigma c<1$. Then for sufficiently large R, we have: $(2\pi)^{-1/2}h_1(V\cap B(\theta,R))\leqq 10cR^2$; for sufficiently large n,

$$(2\pi)^{-1/2}h_1(V_n)\leqq(10c)8n\sigma(c\vee\sigma)\leqq 2n(c\vee\sigma).$$

We proceed in the same way as in proving (a) and obtain

$$\gamma_{\theta,\sigma}\{x\in E:(1+\varepsilon)\sup_{\eta\in V_n}\mathscr{L}(\eta,x)\geqq\mathscr{L}(\theta,x)\}\leqq\exp\left(-\left(\frac{n}{4}-1\right)\left(\frac{c}{\sigma}\vee 1\right)+\frac{\varepsilon}{4\sigma^2}\right);$$

now it is easy to choose the appropriate ε.

PROOF OF THEOREM 1. We can assume that $\theta\in V$. From Lemma 2(b) it follows that

$$\sum_{n=1}^{\infty}\gamma_{\theta,\sigma}\{x\in E:(1+\varepsilon)\sup_{\eta\in V_n}\mathscr{L}(\eta,x)\geqq\mathscr{L}(\theta,x)\}<+\infty;$$

hence for $\gamma_{\theta,\sigma}$-almost each $x\in E$ there is $r_x<+\infty$ such that $\sup_{\eta\in V\setminus B(\theta,r_x)}\mathscr{L}(\eta,x)<\mathscr{L}(\theta,x)$. But $V\cap B(\theta,r)$ is GB-compact for any r, and hence $\sup_{\eta\in V\cap(B,\theta,r_x)}\mathscr{L}(\theta,x)<+\infty$ for $\gamma_{\theta,\sigma}$-almost all x; (a) is proved. We prove (b) first for the case when V is bounded. In such a case V is compact. Let $\theta_k(x)\in V$ be such that

$$\sup_{\eta\in V}\mathscr{L}(\eta,x)=\lim_{k\to\infty}\mathscr{L}(\theta_k(x),x);$$

the sequence $\theta_k(x)$ has the limit point $\hat\theta(x,V)$. Theorem 3(b) shows that this limit point is the only one and that it does not depend on the choice of $\theta_k(x)$. We have thereby proved (b) for bounded V. If V is unbounded, then it suffices to set $\hat\theta(x,V)=\hat\theta(x,V\cap B(\theta,r_x))$ and take advantage of facts already proven. Assertion (c) of Theorem 1 follows from Theorem 2, which is proved below.

PROOF OF THEOREM 2. We start with (a). Let n be such that

$$8(n-1)\sigma(c\vee\sigma)<r^2<8n\sigma(c\vee\sigma),$$

where $c = \frac{1}{10}C_2(V, \theta)$, and apply Lemma 2(a):

$$\begin{aligned}
&\gamma_{\theta,\sigma}\{x \in E: \|\hat{\theta}(x) - \theta\| \geqq r\} \\
&\quad \leqq \gamma_{\theta,\sigma}\{x \in E: \sup_{\eta \in V_n \cup V_{n+1} \cup \cdots} \mathcal{L}(\eta, x) > \mathcal{L}(\theta, x)\} \\
&\quad \leqq \sum_{m=n}^{\infty} \exp\left(-\left(\frac{m}{4} - 1\right)\left(\frac{c}{\sigma} \vee 1\right)\right) \\
&\quad \leqq (1 - e^{-1/4})^{-1} \exp\left(-\left(\frac{n}{4} - 1\right)\left(\frac{c}{\sigma} \vee 1\right)\right) \\
&\quad \leqq 4.53 \exp\left(-\left(\frac{r^2}{32\sigma(c \vee \sigma)} - 1\right)\frac{c \vee \sigma}{\sigma}\right) \\
&\quad = 4.53 \exp\left(-\frac{r^2}{32\sigma^2} + \frac{c \vee \sigma}{\sigma}\right) \\
&\quad \leqq 4.53 \exp\left(-\frac{32u\sigma^2 + 32c\sigma + 81\sigma^2}{32\sigma^2} + \frac{c + \sigma}{\sigma}\right) \\
&\quad = 4.53 \exp\left(-u - \frac{81}{32} + 1\right) \leqq \exp(-u).
\end{aligned}$$

Assertion (b) follows from Lemma 1:

$$\begin{aligned}
&\gamma_{\theta,\sigma}\{x \in E: \|\hat{\theta}(x) - \theta\| > r\} \\
&\quad \leqq \gamma_{\theta,\sigma}\{x \in E: \sup_{\eta \in V \setminus B(\theta, r)} \mathcal{L}(\eta, x) \geqq \mathcal{L}(\theta, x)\} \\
&\quad \leqq \exp\left(-\frac{1}{2R^2(V, \theta)}\left(\frac{r^2}{2\sigma} - (2\pi)^{-1/2}h_1(V)\right)^2\right) = \exp(-u).
\end{aligned}$$

The author expresses his gratitude to I. A. Ibragimov for useful discussions.

Received by the editors
December 5, 1978

REFERENCES

[1] I. A. Ibragimov and R. Z. Khas'minskii, *On the estimation of an infinite-dimensional parameter in Gaussian white noise*, Soviet Math. Doklady, 18, 5 (1977), pp. 1307–1309.

[2] I. A. Ibragimov and R. Z. Khas'minskii, *A problem of statistical estimation in Gaussian white noise*, Soviet Math. Doklady, 18, 5 (1977), pp. 1351–1354.

[3] V. N. Sudakov, *Geometrical problems in the theory of infinite-dimensional probability distributions*, Proc. Steklov Inst. Math., 1979.

[4] B. S. Tsirel'son, *A natural modification of a random process and its application to random function series and Gaussian measures*, Zap. Nauch. Sem. LOMI, 55 (1976), pp. 35–63. (In Russian.)

[5] B. L. Rao, *Estimation of a unimodal density*, Sankhyā, ser. A (1969), 31, 1, pp. 23–36.

[6] I. A. Ibragimov and R. Z. Khas'minskii, *Asymptotic behavior of statistical estimators of a shift parameter for samples with intrinsic continuous density*, Zap. Nauch. Sem. LOMI 41 (1974), pp. 67–93. (In Russian.)

[7] M. S. Ermakov, *On sufficient statistics for families of densities with "moving carrier"*, Theory Prob. Appl., 21 (1976), pp. 657–661.

[8] Yu. A. Davydov, *On the absolute continuity of distributions of functionals of random processes*, Theory Prob. Appl., 23 (1978), pp. 218–219.

[9] B. S. Tsirel'son, I. A. Ibragimov and V. N. Sudakov, *Norms of Gaussian sample functions*, Lecture Notes in Mathematics 550, Springer, Berlin, 1976, pp. 20–41.