[7] S. A. KOVYAZIN, *A variant of the law of large numbers for random subsets of a space with an atomless measure*, Preprint VTs SO Akad. Nauk SSSR, Krasnoyarsk, 1983, no. 1, pp. 8–9. (In Russian.)
[8] A. I. ORLOV, *Random sets: laws of large numbers and the testing of statistical hypotheses*, Theory Prob. Appl., 23 (1978), pp.
[9] Z. ARTSTEIN AND R. A. VITALE, *A strong law of large numbers for random compact sets*, Ann. Probab., 3 (1975), pp. 879–882.

A GEOMETRIC APPROACH TO MAXIMUM LIKELIHOOD ESTIMATION FOR INFINITE-DIMENSIONAL GAUSSIAN LOCATION. II.

B. S. TSIREL'SON

(*Translated by K. Durr*)

Part I of this paper [1] examined necessary and sufficient conditions for the existence, uniqueness and consistency of the MLE for an infinite-dimensional location parameter of a Gaussian measure. Regarding the infinite-dimensional set V to which the parameter to be estimated was known to belong, nothing was assumed beforehand except closedness. The geometric tool of the study was the mean width $h_1(V)$ introduced by Sudakov. This second part considers the more special question of the stability of the MLE with respect to sampling fluctuations. The set V is assumed to be convex. The geometric tool will be the k-th thickness $h_k(V)$ introduced by Simone Chevet (however, the statements of the main probabilistic results contain only h_1). Roughly speaking, it is found that in the MLE almost all degrees of freedom (except a finite number) inherent in a sample point are "frozen". This fact which is fairly natural in itself, occasionally produces unexpected and even curious effects in applications. For instance, in some cases of estimating a signal in additive white noise, the MLE turns out to be a step function regardless of the properties of the signal.

We continue to use the notation and things introduced in [1]: E, γ, E_0, $\langle\theta, x\rangle$, $\|\theta\|$, γ_θ, $\gamma_{\theta,\sigma}$, $\mathcal{L}_\sigma(\theta, x)$, V, as well as $B(\theta, r)$. We shall also consider the finite-dimensional case. If E is finite-dimensional, then without loss of generality we can assume that $E = E_0 = \mathbf{R}^n$; $\langle\theta, \eta\rangle$ and $\|\theta\|$ are the usual Euclidean scalar product and norm in $\mathbf{R}^n$; $\gamma = \gamma^n$ is the standard Gaussian measure in $\mathbf{R}^n$ with density

$$(2\pi)^{-n/2}\exp\left(-\tfrac{1}{2}\|x\|^2\right)$$

with respect to Lebesgue measure. The logarithm likelihood can be written as

$$\log \mathcal{L}'_\sigma(\theta, x) = (\|x\|^2 - \|x-\theta\|^2)/2\sigma^2,$$

from which it is evident that the MLE is simply the closest point in V to x (however, in the infinite-dimensional case the vectors x and $x-\theta$ are "infinitely long" in the considered norm).

In Theorems 1–2 below, $V \subset \mathbf{R}^n$ is a convex closed set containing the point θ with coordinates $\theta_1, \cdots, \theta_n$, σ is a positive number and $\hat{\theta}(x)$ denotes the closest point in V to x.

Theorem 1. *Let V be a convex polyhedron, i.e., the intersection of a finite number of halfspaces, and F_k be the set of all points lying on a k-dimensional face of the polyhedron V but not lying on a face of lower dimension (here F_0 is the set of vertices and F_n the interior of V). Let ξ be a Gaussian random vector in $\mathbf{R}^n$ whose coordinates ξ_k are independent random variables with the means θ_k and the same variance σ^2.*

(a) *If V is bounded, then*

$$\mathbf{P}\{\hat{\theta}(\xi)\in F_k\}\leqq\frac{1}{k!}\left(\frac{C}{\sigma}\right)^k$$

for all $k=0,1,\cdots,n$, *where* $C=(2\pi)^{-1/2}h_1(V)$;

(b)

$$\mathbf{P}\left\{\hat{\theta}(\xi)\in\bigcup_{i=k}^{n}F_i\right\}\leqq 10\left(\frac{e^2C_2^2}{k}\log\frac{k}{C_2^2}\right)^{k/2}$$

for $k>C_2^2$ *and* $\sigma^2k\log(k/C_2^2)\geqq 1$; *here* $C_2=(2\pi)^{-1/2}h_1(V\cap B(\theta,1))$.

Theorem 2. *Suppose that* $\xi_1=(\xi_{11},\cdots,\xi_{1n})$ *and* $\xi_2=(\xi_{21},\cdots,\xi_{2n})$ *are Gaussian random vectors in* $\mathbf{R}^n$ *and for each k the variables* ξ_{1k} *and* ξ_{2k} *both have the mean* θ_k *and the variance* σ^2, *and their correlation coefficient is equal to* ρ; *and suppose that all the other correlations vanish.*

(a) *If V is bounded,* $C=(2\pi)^{-1/2}h_1(V)$ *and* $\rho\geqq 1-\min(\sigma/5C,\frac{1}{2})$, *then for any* $u\geqq 0$

$$\mathbf{P}\left\{\frac{\|\hat{\theta}(\xi_1)-\hat{\theta}(\xi_2)\|}{\sqrt{\sigma(1-\rho)}}\geqq 3\sqrt{u\sigma+2C}\right\}\leqq e^{-u};$$

(b) *if* $C_2=(2\pi)^{-1/2}h_1(V\cap B(\theta,1))$ *and*

$$\rho\geqq 1-\min\left((5C_2(\sqrt{2(u+3)}+C_2))^{-1},\sigma/5C_2,\tfrac{1}{2}\right),$$

then for any $u\geqq 0$

$$\mathbf{P}\{\|\hat{\theta}(\xi_1)-\hat{\theta}(\xi_2)\|/\sqrt{\sigma(1-\rho)}\geqq q\}\leqq e^{-u},$$

where

$$q=3\max\left(\sqrt{\sigma}(\sqrt{u+3}+\sqrt{2}\,C_2),(\sigma(u+3)+2C_2)^{1/2}\right).$$

The inequalities in Theorems 1–2 do not involve the dimension n of the space E; they can of course be carried over to the infinite-dimensional case. To avoid talking about faces of a polyhedron, we introduce the following definition. For every $\theta\in V$ we define $K(\theta)$ to be the largest k for which there are linearly independent vectors $\eta_1,\cdots,\eta_k\in E_0$ such that $\theta+a_1\eta_1+\cdots+a_k\eta_k\in V$ for any $a_1,\cdots,a_k\in[-1,+1]$. If θ is an extreme point of V, then $K(\theta)=0$. If any amount of such η_i exist, then $K(\theta)=+\infty$. Under the conditions of Theorem 1, obviously $K(\theta)=k$ for $\theta\in F_k$. We introduce yet another definition to avoid the correlation coefficients of the components. Let ξ_1,ξ_2 be random elements of the space E each with the distribution $\gamma_{\theta,\sigma}$ and let $\rho\in(-1,+1)$. We say that ξ_1 and ξ_2 are ρ-correlated if they are representable as

$$\xi_1=\theta+\sigma(((1+\rho)/2)^{1/2}\zeta_1+((1-\rho)/2)^{1/2}\zeta_2),$$

$$\xi_2=\theta+\sigma(((1+\rho)/2)^{1/2}\zeta_1+((1-\rho)/2)^{1/2}\zeta_2),$$

where ζ_1,ζ_2 are independent random elements of E each with the distribution γ. Under the conditions of Theorem 2 the random vectors ξ_1 and ξ_2 are obviously ρ-correlated.

In Theorems 3–4 below, E can be both finite-dimensional and infinite-dimensional; $V\subset E_0$ is a convex closed set. According to [1] the MLE of the parameter θ ranging over V is well-defined if and only if the characteristic $C_1(V)$ introduced in [1] is finite. In contrast to [1], V is assumed here to be convex, and in this case $C_1(V)$ is either infinite or 0; and if it is 0, then the other characteristic $C_2(V,\theta)$ for $\theta\in V$ introduced in [1] becomes

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

Indeed, for convex V we have

$$(V\cap B(\theta,r))-\theta\subset r((V\cap B(\theta,1))-\theta),$$

and so also

(1) $$h_1(V\cap B(\theta,r))\leqq rh_1(V\cap B(\theta,1))$$

for $r\geqq 1$, $\theta\in V$.

It is assumed in Theorems 3-4 that $C_1(V)<+\infty$; θ is a point in V, σ is a positive number; and $\hat{\theta}(x)$ is defined as in [1].

Theorem 3. *Let ξ be a random element in E having the distribution $\gamma_{\theta,\sigma}$.*
(a) *If V is bounded, then for any $a \geqq 0$*

$$\mathbf{E}\exp(aK(\hat{\theta}(\xi))) \leqq \exp(Ce^a/\sigma),$$

where $C=(2\pi)^{-1/2}h_1(V)$;
(b)

$$\mathbf{P}\{K(\hat{\theta}(\xi)) \geqq k\} \leqq 10(e^2C_2^2k^{-1}\log(k/C_2^2))^{k/2}$$

for $k>C_2^2$ and $\sigma^2k\log(k/C_2^2)\geqq 1$; here $C_2=(2\pi)^{-1/2}h_1(V\cap B(\theta,1))$.

Theorem 4. *Suppose that ξ_1, ξ_2 are random elements in E both having the distribution $\gamma_{\theta,\sigma}$ and that they are ρ-correlated. Then parts* (a) *and* (b) *of Theorem 2 hold.*

REMARK 1. The constant "3" in front of the radical in the inequality in part (a) of Theorem 2 can be improved. But this is all that can be done to strengthen this inequality if ρ is close to 1 and C/σ and u are not close to 0. More precisely, let the function B of four arguments be such that in the conditions of part (a) of Theorem 2 (or Theorem 4—it makes no difference)

$$\lim_{\rho\to 1-}\mathbf{P}\left\{\frac{\|\hat{\theta}(\xi_1)-\hat{\theta}(\xi_2)\|}{\sqrt{\sigma(1-\rho)}} \geqq B(C,u,\sigma,\rho)\right\} \leqq e^{-u}.$$

Then it can be shown that the expression

$$\overline{\lim_{\rho\to 1-}}\frac{\sqrt{u\sigma}+2C}{B(C,u,\sigma,\rho)}$$

is bounded for any range of the parameters C, u and σ for which σ/C and $1/u$ are bounded.

We continue the analysis begun in [1] of some examples linked with estimation of a signal in an additive white noise.

EXAMPLE 1. The set V consists of all functions on (0, 1) whose variation does not exceed M (see [1], example 2). We apply part (a) of Theorem 3 to the set V_1 of functions in V orthogonal to the unit element. It is not hard to show that $K(\theta)$ is finite only for step functions $\theta\in V_1$ with variation exactly M; for such functions $K(\theta)=J(\theta)-2$, where $J(\theta)$ is the number of steps. Part (a) of Theorem 3 yields the following proposition.

Let $dX(t)=S(t)\,dt+\sigma\,dw(t)$, where w is a Wiener process and S is a function of bounded variation; $\mathrm{Var}_{t\in(0,1)}S(t)\leqq M$. Then the MLE $\hat{S}$ for S in the indicated class of functions based on the observation X is a step function with probability, its variation is equal to M, and the random number J of steps of $\hat{S}$ satisfies

$$\mathbf{E}\exp(aJ) \leqq \exp(\sqrt{\pi/8}\,e^a\sigma^{-1}M+2a)$$

for any $a\geqq 0$.

EXAMPLE 2. The set V consists of all increasing functions in $L_2[0,1]$; as in Example 3 in [1], we pass over to $V_{a,b}$. It is again clear that $K(\theta)$ is finite only for step functions θ; from part (a) of Theorem 3 we conclude that $\hat{\theta}$ is a step function with probability 1. However, after passing to the limit in a, b the number of steps becomes infinite (they cluster at the ends of the interval).

REMARK 2. Considering V to be the set of all L_2 functions on the square $[0,1]\times[0,1]$ that are increasing in both arguments, we conclude that the ML-estimation of such functions is impossible in additive two-dimensional white noise; indeed, even limiting ourselves to functions with two values 0 and 1, we obtain a compact set V not possessing the GB-property, as Dudley showed [5].

EXAMPLE 3. The set V consists of all functions θ satisfying a Lipschitz condition $|\theta(s)-\theta(t)|\leqq M|s-t|$; see Example 1 for $\alpha=1$ in [1]. Part (a) of Theorem 3 applied to the set V_1 of functions in V orthogonal to the unit element makes it possible to draw only the following conclusion: $d\hat{S}(t)/dt = \pm M$ for almost all t with probability 1. The fact is that any function in V_1 (not only piecewise linear) with a derivative $\pm M$ is an extreme point of V_1. In effect, with probability 1 the function $\hat{S}$ is not piecewise linear; the intervals where it is linear make up the complement to a Cantor set.

REMARK 3. If so desired, Examples 1-2 can be generalized in the following direction. The condition $\mathrm{Var}_{t\in(0,1)} S(t) \leqq M$ can be written in the form

$$\int_0^1 \left| \frac{d}{dt} S(t) \right| dt \leqq M,$$

interpret as follows: the derivative of S is a finite mass in the sense of generalized function theory, and the norm of this mass does not exceed M. Similarly, we can examine the more general condition $\int_0^1 |LS(t)|\, dt \leqq M$, where L is some linear differential operator. In that case the MLE is a piecewise smooth function, satisfying the equation $LS(t) = 0$ on each piece. The condition that S is increasing can be generalized to the condition $LS(t) \geqq 0$.

The proofs of Theorems 1-4 will be given in part 3 of this paper [2]. For convenience of orientation, we point out what topics comprise this paper (parts 2 and 3) and how these topics are interrelated.

Topics 1-3: the relationship between the probability and geometric considerations. Here there are three formally independent topics: (1) the thickness of an infinite-dimensional GB-compact set as the volume of the joint spectrum averaged with respect to Gaussian measure (Theorem 6, as well as Lemma 2); (2) geometric-probabilistic analyses for the case where V is a finite-dimensional polyhedron (Lemma 1; Theorem 5 is also relevant here); (3) the same where V is a finite-dimensional convex solid with smooth boundary ([2], Lemmas 1, 2). Topics 2 and 3 deal essentially with the same thing but in entirely different languages; it is easier to develop them independently rather than derive one from the other.

Topic 4: the distance of the MLE from the true value of the parameter ([2], Lemmas 3, 4) depends on topic 2.

Topics 5-7: probabilistic results on the behavior of the MLE. Here again there are three formally independent topics; all three rely on topic 4; moreover, topic 5 relies on topic 2, and topics 6 and 7 on topic 3. Topic 5: the probability that the MLE occurs on a face of given dimension if V is a finite-dimensional polyhedron (Theorem 1); topic 6: the same if V is infinite-dimensional (Theorem 3 and also Lemmas 5, 6 in [2]); topic 7: the distance between the MLE's for two strongly correlated sample points ([2], Theorems 2 and 4 and Lemmas 7, 8). It would appear that topic 6 ought to depend on topic 5 but it was found to be easier to develop it independently. In a certain sense topic 7 is also about the same thing, but in an entirely different language, in order to give a nontrivial result when V has no nontrivial faces.

Chevet [3] introduced a scale of geometric characteristics h_k for convex GB-compact sets in a Hilbert space; $h_k(V)$ has come to be called the k-thickness of the set V. Starting out from the well-known integral cross-sectional measures W_k^n, she observed that under suitable numbering and norming they cease to depend on the dimension of the space containing the given finite-dimensional set; this allowed the h_k to be defined for finite-dimensional convex compact sets in Hilbert space; then an arbitrary convex compact set was approximated from within by finite-dimensional ones; here it turned out that $h_k(V) < +\infty$ if and only if $h_1(V) < +\infty$, i.e., for $V \in GB$. A relationship was found between the k-thickness and the moments of the supremum of a Gaussian process. Finally, Chevet carried over to the infinite-dimensional case the classical inequalities of Fenchel-Alexandrov for mixed volumes. These inequalities are a basic tool for proving results of this paper. See a modern introduction to the theory of cross-sectional measures in [4], Chapter 4; Section 9.9 of the cited chapter considers the infinite-dimensional case following Chevet.

We give two new definitions for h_k; in contrast to Chevet's definition, they are related directly to a Gaussian measure and do not use finite-dimensional approximation. The normalization of the thickness used by Chevet (and for $k=1$ by Sudakov) seems to us to be not the most convenient; it is more convenient to use the expression

$$\mathcal{M}_k(V)=(2\pi)^{-k/2}k!\,h_k(V), \qquad k=0,1,\cdots, \tag{2}$$

cf. [3], (3.6.2). We shall also call the quantity $\mathcal{M}_k(V)$ the k-thickness of the set V. Inequalities (4.2.1) and (4.2.2) in [3] can now be written as

$$\mathcal{M}_{k-1}(V)\mathcal{M}_{k+1}(V)\leqq \mathcal{M}_k^2(V), \tag{3}$$

$$\mathcal{M}_k(V)\leqq \mathcal{M}_1^k(V), \qquad k=0,1,\cdots. \tag{4}$$

For every GB-set $V\subset E_0$ the collection of probability measures $\{\gamma_\theta: \theta\in V\}$ has a least upper bound γ_V in the set of all (not only probability) finite measures on E; it is clear that the measure γ_V has the following density with respect to the measure γ:

$$\frac{\gamma_V(dx)}{\gamma(dx)}=\sup_{\theta\in V}\mathcal{L}(\theta,x)=\exp\left(\sup_{\theta\in V}\left(\langle\theta,x\rangle-\frac{1}{2}\|\theta\|^2\right)\right). \tag{5}$$

Theorem 5. $\gamma_V(E)=\sum_{k=0}^{\infty}\mathcal{M}_k(V)/k!$.
From (4) and this theorem we obtain

Corollary 1. $\gamma_V(E)\leqq\exp(\mathcal{M}_1(V))$.
Noting that the functionals $\mathcal{M}_K$ are homogeneous, we find that

$$\gamma_{aV}(E)=\sum_{k=0}^{\infty}\mathcal{M}_k(V)a^k/k!$$

for any $a>0$, and hence we obtain an equivalent definition for k-thickness.

Corollary 2.

$$\mathcal{M}_k(V)=\frac{d^k}{da^k}\gamma_{aV}(E)\Big|_{a=0+} =\frac{d^k}{da^k}\left(\int\exp\left(\sup_{\theta\in V}\left(a\langle\theta,x\rangle-\frac{1}{2}a^2\|\theta\|^2\right)\right)\gamma(dx)\right)\Big|_{a=0+}, \qquad k=0,1,\cdots.$$

One further equivalent definition of k-thickness will be given below in terms of the "joint spectrum" of several realizations of a random process. Here special care must be given to choosing a modification of the process. For GC-sets there is no problem since the realizations can be assumed to be continuous. However, in the general case the separable modification is found to be insufficient; the natural modification introduced in [6] is required. To every set $V\subset E_0$, a Gaussian random process $\langle\theta,x\rangle$ is defined where θ ranges over V and x ranges over the space E equipped with the Gaussian measure γ. This process has a natural modification if and only if $V\in GB\sigma$ (a countable union of GB-sets) [6]. It is understood below that $\langle\theta,x\rangle$ denotes precisely the natural modification of the above-mentioned process.

Let $V\in GB$ (or $GB\sigma$) and $x_1,\cdots,x_k\in E$; we call the set $\operatorname{spec}(x_1,\cdots,x_k|V)=\{(\langle\theta,x_1\rangle,\cdots,\langle\theta,x_k\rangle): \theta\in V\}$ lying in $\mathbf{R}^k$ the joint spectrum for $x_1,\cdots,x_k$ on V. If V is a convex GB-compact set, then the joint spectrum is (a.s.) a convex bounded set; if $V\in GC$, it is closed and if $V\notin GC$, this is not necessarily so. We point out that a separable modification would not guarantee the convexity of the joint spectrum. The natural modification does guarantee it since the mapping $\theta\to(\langle\theta,x_1\rangle,\cdots,\langle\theta,x_k\rangle)$ is algebraically linear on the linear span of the set V even if it is not continuous on V.

Theorem 6. *For any convex GB-compact set $V\subset E_0$ and $k=0,1,\cdots$,*

$$\mathcal{M}_k(V)=\frac{1}{\pi_k}\int\cdots\int \operatorname{mes}_k\operatorname{spec}(x_1,\cdots,x_k|V)\gamma(dx_1)\cdots\gamma(dx_k);$$

here mes_k *is Lebesgue measure in* $\mathbf{R}^k$; $\pi_k = \pi^{k/2}/\Gamma(k/2+1)$ *is the volume of the k-dimensional unit sphere.*

Lemma 1. *Let V be a finite-dimensional convex polyhedron in* $\mathbf{R}^n$, *and let* F_k *be defined as in Theorem* 1. *Then*

$$\gamma_V\{x: \hat{\theta}(x) \in F_k\} = \mathcal{M}_k(V)/k!, \qquad k = 0, 1, \cdots, n.$$

PROOF. Consider a k-dimensional face F of the polyhedron V (F includes no points lying on faces of lower dimension), the k-dimensional subspace $E_F \subset E_0$ parallel to it and its annihilator $E_F^\perp = \{x \in E: \langle \eta, x\rangle = 0 \ \forall \eta \in E_F\}$, as well as the affine subspace $\tilde{F} = F + E_F$ containing F. In accordance with the decomposition $E = \tilde{F} \oplus E_F^\perp$ each measure γ_θ splits into the product of its projections $\gamma_\theta = (\gamma_\theta|\tilde{F}) \otimes (\gamma_\theta|E_F^\perp)$. It is easy to see that for $\theta \in F$ the set $\{x - \hat{\theta}(x): x \in E, \hat{\theta}(x) = \theta\}$ is a convex cone K_F^0 in $E_F^\perp$ not depending on θ. Therefore, $\{x \in E: \hat{\theta}(x) \in F\} = F + K_F^0$. The measure γ_V clearly coincides on this set with the measure

$$\gamma_F = \sup_{\theta \in F} \gamma_\theta = \sup_{\theta \in F} ((\gamma_\theta|\tilde{F}) \otimes (\gamma_\theta|E_F^\perp)) = (\sup_{\theta \in F} (\gamma_\theta|\tilde{F})) \otimes (\gamma_{\theta_0}|E_F^\perp),$$

where θ_0 is the unique point in $\tilde{F} \cap E_F^\perp$; however, the first factor coincides on the set F with the measure $(2\pi)^{-k/2}\,\mathrm{mes}_k$. Thus

$$\gamma_V\{x \in E: \hat{\theta}(x) \in F\} = ((2\pi)^{-k/2}\,\mathrm{mes}_k F)((\gamma|E_F^\perp)(K_F^0)) = (2\pi)^{-k/2}(\mathrm{mes}_k F)\gamma(K_F),$$

where $K_F = K_F^0 + E_F$. Summing over all k-dimensional faces, we find that

$$\gamma_V\{x \in E: \hat{\theta}(x) \in F_k\} = (2\pi)^{-k/2} \sum_F (\mathrm{mes}_k F)\gamma(K_F).$$

It remains to apply Lemma 3.5 in [3], according to which

$$h_k(V) = \sum_F (\mathrm{mes}_k F)\gamma(K_F), \qquad k = 0, 1, \cdots, n.$$

PROOF OF THEOREM 5. If V is a finite-dimensional polyhedron $\subset \mathbf{R}^n$, applying Lemma 1, we find that

$$\gamma_V(E) = \sum_k \gamma_V\{x: \hat{\theta}(x) \in F_k\} = \sum_{k=0}^n \frac{1}{k!}\mathcal{M}_k(V).$$

In the general case we approximate V from within by finite-dimensional polyhedra V_n, $V_1 \subset V_2 \subset \cdots \subset V$, with $\bigcup_{n=1}^\infty V_n$ dense in V; here $\mathcal{M}_k(V) = \lim_{n\to\infty} \mathcal{M}_k(V_n)$, $k = 0, 1, \cdots$, according to Proposition 3.9 in [3], and

$$\gamma_V = \sup_{\theta \in V} \gamma_\theta = \sup_n \sup_{\theta \in V} \gamma_\theta = \sup_n \gamma_{V_n},$$

so

$$\gamma_V(E) = \lim_{n\to\infty} \gamma_{V_n}(E) = \lim_{n\to\infty} \sum_{k=0}^n \frac{1}{k!}\mathcal{M}_k(V_n) = \sum_{k=0}^\infty \frac{1}{k!}\mathcal{M}_k(V).$$

The following general property of a natural modification will be used in the proof of Theorem 6.

Lemma 2. *Let* $\xi(\omega, t)$ *be a natural modification of some random process,* $\omega \in \Omega$, $t \in T$, *and let S be dense in T in the following sense: for any* $t \in T$ *there are* $s_n \in S$, $n = 1, 2, \cdots$, *such that* $\xi(\omega, s_n) \to \xi(\omega, t)$, $n \to \infty$, *for almost all* ω (*the corresponding set of probability* 1 *may depend on t*). *Then there is a set* $\Omega_1 \subset \Omega$ *of probability* 1 *possessing the following property: for any* $t \in T$ *there exist* $s_n' \in S$, $n = 1, 2, \cdots$, *such that* $\xi(\omega, s_n') \to \xi(\omega, t)$, $n \to \infty$, *for all* $\omega \in \Omega_1$.

PROOF. By the definition of a natural modification, we can introduce a metrix ρ on T such that (T, ρ) is a separable metric space and $\xi(\omega, t)$ is continuous in t on (T, ρ) for all ω in some $\Omega_2 \subset \Omega$ of probability 1. Let $\{t_m\}_{m=1}^\infty$ be a fixed countable set dense in (T, ρ);

for each m, fix $s_{m,n} \in S$ such that

$$\mathbf{P}\{|\xi(\omega, s_{m,n}) - \xi(\omega, t_m)| > 1/n\} \leqq 2^{-m-n}.$$

The sum of these probabilities over all m and n is finite. Therefore there is a set $\Omega_3 \subset \Omega$ of probability 1 possessing the following property: $|\xi(\omega, s_{m,n}) - \xi(\omega, t_m)| \leqq 1/n$ for all $\omega \in \Omega_3$ and all pairs (m, n) except a finite (depending on ω) number of pairs. Put $\Omega_1 = \Omega_2 \cap \Omega_3$. Let $t \in T$ be given. Choose $m_1, m_2, \cdots$, such that $t_{m_n} \to t$, $n \to \infty$, in (T, ρ). Put $s'_n = s_{m_n,n}$. Then for large n we have

$$\begin{aligned}|\xi(\omega, s'_n) - \xi(\omega, t)| &\leqq |\xi(\omega, s_{m_n,n}) - \xi(\omega, t_{m_n})| + |\xi(\omega, t_{m_n}) - \xi(\omega, t)| \\ &\leqq 1/n + |\xi(\omega, t_{m_n}) - \xi(\omega, t)| \to 0, \qquad n \to \infty,\end{aligned}$$

for all $\omega \in \Omega_1$, as required.

Corollary 3. *Let $V \subset E_0$ be a GB-set and let $V_0 \subset V$. If V_0 is dense in V, then* $\operatorname{spec}(x_1, \cdots, x_k | V_0)$ *is dense in* $\operatorname{spec}(x_1, \cdots, x_k | V)$ *for almost all* $(x_1, \cdots, x_k)$.

Note that the assertion is trivial for GC-sets.

PROOF OF THEOREM 6. We first reduce the general case to the finite-dimensional one. To do this, we approximate V from within by finite-dimensional convex compact sets V_n, $n = 1, 2, \cdots$, where $\mathcal{M}_k(V) = \lim_{n\to\infty} \mathcal{M}_k(V_n)$, as already noted in the proof of Theorem 5. By Corollary 3 the union (over n) of the convex sets $\operatorname{spec}(x_1, \cdots, x_k | V_n)$ is dense in the convex set $\operatorname{spec}(x_1, \cdots, x_n | V)$ a.s., and therefore

$$\begin{aligned}&\int \cdots \int \operatorname{mes}_k \operatorname{spec}(x_1, \cdots, x_k | V) \gamma(dx_1) \cdots \gamma(dx_k) \\ &\quad = \int \cdots \int \lim_{n\to\infty} \operatorname{mes}_k \operatorname{spec}(x_1, \cdots, x_k | V_n) \gamma(dx_1) \cdots \gamma(dx_k) \\ &\quad = \lim_{n\to\infty} \int \cdots \int \operatorname{mes}_k \operatorname{spec}(x_1, \cdots, x_k | V_n) \gamma(dx_1) \cdots \gamma(dx_k).\end{aligned}$$

Hence, it suffices to prove Theorem 6 for the case where V is finite-dimensional. The space E can then also be assumed to be finite-dimensional of dimension n. The set $\operatorname{spec}(x_1, \cdots, x_k | V)$ is the image of V under the linear mapping $A_x: E \to \mathbf{R}^k$ which is defined by $A_x(\theta) = (\langle \theta, x_1 \rangle, \cdots, \langle \theta, x_k \rangle)$; here and elsewhere $x = (x_1, \cdots, x_k)$. Choose a fixed orthonormal basis in E; the matrix of the mapping A_x in this basis (denote it also by A_x) has k rows and n columns. Each element of this matrix depends on the point x of the probability space $(E \oplus \cdots \oplus E, \gamma \otimes \cdots \otimes \gamma)$ and in this sense can be viewed as a random variable. It is easy to see that all the elements of A_x are independent standard Gaussian random variables. Orthogonalizing the rows of A_x, we obtain a factorization $A_x = B_x C_x$, where B_x is a lower triangular square matrix of order k and C_x is a matrix of k orthonormal rows of n elements. Examining the orthogonalization process, we can easily show that all the elements of B_x are independent, all the off-diagonal elements have a standard normal distribution and all diagonal elements have the following distributions: $\chi_n, \chi_{n-1}, \cdots, \chi_{n-k+1}$ (χ_l is the distribution of the square root of the sum of the squares of l independent standard normal variables). Therefore,

$$\begin{aligned}\mathbf{E} \det B_x &= (\mathbf{E}\chi_n)(\mathbf{E}\chi_{n-1}) \cdots (\mathbf{E}\chi_{n-k+1}) \\ &= \frac{\sqrt{2}\,\Gamma\left(\frac{n+1}{2}\right)}{\Gamma\left(\frac{n}{2}\right)} \frac{\sqrt{2}\,\Gamma\left(\frac{n}{2}\right)}{\Gamma\left(\frac{n-1}{2}\right)} \cdots \frac{\sqrt{2}\,\Gamma\left(\frac{n-k+2}{2}\right)}{\Gamma\left(\frac{n-k+1}{2}\right)} = 2^{k/2} \frac{\Gamma\left(\frac{n+1}{2}\right)}{\Gamma\left(\frac{n-k+1}{2}\right)}.\end{aligned}$$

Further, it is not hard to show that B_x and C_x are independent and that the matrix C_x has a unique probability distribution invariant under rotation or more precisely under right-multiplication by any orthogonal matrix of order n; it is essentially a question of a uniform distribution on the Stiefel manifold $V_k(\mathbf{R}^n)$, and just below of a uniform distribution on the Grassman manifold $G_k(\mathbf{R}^n)$. We have

$$\begin{aligned}\mathbf{E}\,\mathrm{mes}_k\,\mathrm{spec}\,(x|V) &= \mathbf{E}\,\mathrm{mes}_k\,B_x C_x V\\ &= (\mathbf{E}\det B_x)(\mathbf{E}\,\mathrm{mes}_k\,C_x V)\\ &= 2^{k/2}\frac{\Gamma((n+1)/2)}{\Gamma((n-k+1)/2)}\mathbf{E}\,\mathrm{mes}_k\,C_x V.\end{aligned}$$

The set $C_x V$ is isometric to the orthogonal projection of V on the k-dimensional subspace spanned by the rows of the matrix C_x. Therefore, $\mathbf{E}\,\mathrm{mes}_k\,C_x V$ is the mean volume of the k-dimensional projection of V. We use a consequence of Kubota's formula [3], formula 4.4.4″; it says that the mean volume of the k-dimensional projection of V is equal to

$$\Gamma\left(\frac{n}{2}+1\right)\left[\binom{n}{k}\Gamma\left(\frac{k}{2}+1\right)\Gamma\left(\frac{n-k}{2}+1\right)\right]^{-1} h_k(V);$$

by means of the duplication formula for the Γ-function

$$\Gamma(2z) = \pi^{-1/2}2^{2z-1}\Gamma(z)\Gamma(z+\tfrac{1}{2})$$

we now find that

$$\begin{aligned}\frac{1}{\pi_k}\mathbf{E}\,\mathrm{mes}_k\,\mathrm{spec}\,(x|V) &= \frac{\Gamma(k/2+1)}{\pi^{k/2}}2^{k/2}\frac{\Gamma((n+1)/2)}{\Gamma((n-k+1)/2)}\,\frac{k!(n-k)!}{n!}\\ &\quad\cdot\frac{\Gamma(n/2+1)}{\Gamma(k/2+1)\Gamma((n-k)/2+1)}\,\frac{(2\pi)^{k/2}}{k!}\mathcal{M}_k(V)\\ &= 2^k\frac{\Gamma((n+1)/2)\Gamma(n/2+1)}{\Gamma(n+1)}\\ &\quad\cdot\frac{\Gamma(n-k+1)}{\Gamma((n-k+1)/2)\Gamma((n-k)/2+1)}\mathcal{M}_k(V)\\ &= 2^k\frac{\sqrt{\pi}}{2^n}\,\frac{2^{n-k}}{\sqrt{\pi}}\mathcal{M}_k(V) = \mathcal{M}_k(V).\end{aligned}$$

REMARK 4. Theorem 6 can be proved differently for the finite-dimensional case by introducing a large number of "extraneous" measurements, as was done in the proof of Theorem 3 in [7].

Received by the editors
December 9, 1983

REFERENCES

[1] B. S. TSIREL'SON, *A geometric approach to maximum likelihood estimation for infinite-dimensional Gaussian location.* I, Theory Prob. Appl., 27 (1982), pp. 411–418.

[2] B. S. TSIREL'SON, *A geometric approach to maximum likelihood estimation for infinite-dimensional Gaussian location.* III, Theory Prob. Appl., (in press).

[3] S. CHEVET, *Processus gaussiens et volumes mixtes,* Z. Wahrsch Verw. Gebiete, 36 (1976), pp. 47–65.

[4] YU. D. BURAGO AND V. A. ZALGALLER, *Geometric Inequalities,* Nauka, Leningrad, 1980. (In Russian.)

[5] R. M. DUDLEY, *Lower layers in* $\mathbf{R}^2$ *and convex sets in* $\mathbf{R}^3$ *are not GB classes,* Lect. Notes Math., 709 (1979), pp. 97–102.

[6] B. S. TSIREL'SON, *A natural modification of a random process and its applications to random series of functions and Gaussian measures,* Zap. Nauchr. Sem. Leningrad Otdel. Mat. Inst. Steklon, 55 (1976), pp. 35–63. (In Russian.)

[7] V. N. SUDAKOV AND B. S. TSIREL'SON, *Extremal properties of halfspaces for spherically invariant measures*, Zap. Nauchr. Sem. Leningrad Otdel. Mat. Inst. Steklon, 41 (1974), pp. 14-24. (In Russian.)

THE UNIFORM DISTRIBUTION ON COMPACT HOMOGENEOUS SPACES AND THE KANTOROVICH-RUBINSHTEIN METRIC

V. A. KAIMANOVICH

(*Translated by Yona Ellis*)

Let X be a compact G-space with a Haar measure m (e.g., G is a compact group, $X=G$). A number of papers (see [1]-[4]) have studied the problem of the *uniform distribution* on X, i.e., the weak convergence of the measures $\mu_1 \cdots \mu_n \nu$ or $\mu^n \nu$ to m (μ_i and ν are probability measures on G and X, respectively and μ^n is the n-fold convolution of μ). Below we apply a new technique to the investigation of this problem based on the metrization of the weak topology by the Kantorovich-Rubinshtein (KR) metric and we generalize old results of Arnol'd-Krylov and Ullrich-Urbanik.

1. The KR Metric

Let (X, ρ) be a compact metric space, $M(X)$ the space of Borel probability measures on X and $\operatorname{Lip}(X, \rho)$ the space of functions $u: X \to \mathbf{R}$ such that $u(s)-u(t) \leqq \rho(s, t) (\forall s, t \in X)$. The KR-metric

$$d(\nu_1, \nu_2) = \inf \int \rho(s, t)\, d\Psi(s, t),$$

where the inf is taken over all measures $\Psi \in M(X \times X)$ whose projections onto the first and second factors are ν_1 and ν_2, metrizes the weak topology in $M(X)$, and the inf is always attained for a measure Ψ if and only if $u(s)-u(t)=\rho(s, t)$ $(\forall (s, t) \in \operatorname{supp} \Psi)$ for some $u \in \operatorname{Lip}(X, \rho)$. The dual definition is $d(\nu_1, \nu_2) = \sup \{(f, \nu_1 - \nu_2): f \in \operatorname{Lip}(X, \rho)\}$ [5].

Everywhere below (except in Section 6) we assume that the *topological* group G *acts continuously and minimally on* X *by isometries*, i.e., G is embedded in the compact group $\operatorname{Iso}(X)$ of all isometries of X equipped with the invariant metric $\rho^*(g_1, g_2) = \sup_x \rho(g_1 x, g_2 x)$. Obviously, $d(g\nu_1, g\nu_2) = d(\nu_1, \nu_2)$, $d(\mu\nu_1, \mu\nu_2) \leqq d(\nu_1, \nu_2)$. We call a measure μ on G *transitive* on X if $\rho(s, \operatorname{supp} \mu^n \delta_t) \to 0$, $n \to \infty$ $(\forall s, t \in X)$. For $X = G$ this is equivalent to μ being *strictly aperiodic* [1].

Lemma 1. *If* $d(\mu\nu_1, \mu\nu_2) = d(\nu_1, \nu_2)$ *and* $\operatorname{supp} \mu' \subset \operatorname{supp} \mu$, *then* $d(\mu'\nu_1, \mu'\nu_2) = d(\nu_1, \nu_2)$.

Lemma 2. *If the measure* μ *is transitive on* X *and* $\nu_1 \neq \nu_2 \in M(X)$, *then* $d(\mu^n \nu_1, \mu^n \nu_2) < d(\nu_1, \nu_2)$ *for some* n *(depending on* $\nu_{1,2}$*)*.

PROOF. Let Ψ be a fixed measure realizing $d(\nu_1, \nu_2)$, and $(s, t) \in \operatorname{supp} \Psi$, $\rho(s, t) = \varepsilon > 0$. If $d(\mu^n \nu_1, \mu^n \nu_2) = d(\nu_1, \nu_2)$, then the measure $\mu^n \Psi$ realizes $d(\mu^n \nu_1, \mu^n \nu_2)$, i.e., there is a $u_n \in \operatorname{Lip}(X, \rho)$ such that

$$u_n(gs) - u_n(gt) = \rho(gs, gt) = \varepsilon (\forall g \in \operatorname{supp} \mu^n).$$

The values of $u_n(gs)$ for $g \in \operatorname{supp} \mu^n$ for large n can be arbitrarily close to the smallest value of u_n on X, whence $\varepsilon = 0$ and $\nu_1 = \nu_2$.