THEORY OF PROBABILITY AND ITS APPLICATIONS

Volume XXVI — Number 3

1981

NECESSARY AND SUFFICIENT CONDITIONS FOR THE UNIFORM CONVERGENCE OF MEANS TO THEIR EXPECTATIONS

V. N. VAPNIK AND A. YA. CHERVONENKIS

(*Translated by W. U. Sirk*)

Introduction

In [1] we found necessary and sufficient conditions for the uniform convergence of the relative frequencies of events to their probabilities with respect to the class of events, i.e., necessary and sufficient conditions for the fact that with increasing size of a random independent sample

$$x_1, \cdots, x_l$$

the equality

$$\lim_{l\to\infty} \mathbf{P}\left\{\sup_\alpha \left| \mathbf{E}\theta[F(x,\alpha)] - \frac{1}{l}\sum_{i=1}^{l} \theta[F(x_i,\alpha)]\right| > \varepsilon\right\} = 0$$

is satisfied. Here

$$\theta[z] = \begin{cases} 1 & \text{if } z \geqq 0, \\ 0 & \text{if } z < 0, \end{cases}$$

$F(x, \alpha)$ is a family of functions which is parametric with respect to α.

In [2], on the basis of the conditions found in [1], sufficient conditions were obtained for the uniform convergence of means to their expectations for a uniformly bounded family

$$0 \leqq F(x, \alpha) \leqq C; \tag{1}$$

i.e., sufficient conditions were found for the fact that with increasing size of the sample, the equality

$$\lim_{l\to\infty} \mathbf{P}\left\{\sup_\alpha \left| \mathbf{E}F(x,\alpha) - \frac{1}{l}\sum_{i=1}^{l} F(x_i,\alpha)\right| > \varepsilon\right\} = 0. \tag{2}$$

is satisfied.

In the present paper necessary and sufficient conditions will be obtained for the uniform convergence of means to their expectations for the uniformly bounded parametric family (1).

Below, without loss of generality, we shall assume that $C = 1$.

For the formulation of these conditions we shall introduce a few concepts.

1. ε-entropy

Let A be a bounded set in E_l. A finite set $T \subset E_l$ such that for every $x \in A$ there exists an element $y \in T$, less than $\varepsilon\,(\rho(x, y) < \varepsilon)$ away from x, is called the *relative ε-net A in E_l*. Below, we shall assume that the metric is given in the form

$$\rho(x, y) = \max_{i=1,\cdots,l} |x_i - y_i|,$$

and the norm of the vector x is

$$|x| = \max_i |x_i|.$$

If the ε-net T of the set A is such that $T \subset A$, we shall say that it is a *proper* ε-net of the set A.

The minimum number of elements of an ε-net of the set A relative to E_l is denoted by $N(\varepsilon, A)$; the minimum number of elements of a proper ε-net is denoted by $N_0(\varepsilon, A)$. It is not difficult to see that

$$N_0(\varepsilon, A) \geqq N(\varepsilon, A). \tag{3}$$

On the other hand,

$$N_0(2\varepsilon, A) \leqq N(\varepsilon, A). \tag{4}$$

Indeed, let T be the minimum ε-net A relative to E_l. To each element $y \in T$ we let correspond an element $x \in A$ such that $\rho(x, y) < \varepsilon$ (such an element x can always be found, since otherwise the ε-net could be contracted). The collection T_0 of elements x thus found forms a proper 2ε-net in A (for every $z \in A$ there exists $y \in T$ such that $\rho(y, z) < \varepsilon$ and for this $y \in T$ there exists $x_0 \in T_0$ such that $\rho(x, y) < \varepsilon$; thus, $\rho(x, z) < 2\varepsilon$).

Let $F(x, \alpha)$ be a certain class of numerical functions of the variable $x \in X$, depending on the parameter $\alpha \in \Lambda$. Let $x_1, \cdots, x_l$ be a sample. We consider in the Euclidean space E_l a set A of vectors z with coordinates $z_i = F(x_i, \alpha)$, $i = 1, 2, \cdots, l$, formed by all $\alpha \in \Lambda$.

If the condition $0 \leqq F(x, \alpha) \leqq 1$ is fulfilled, then the set $A(x_1, \cdots, x_l)$ belongs to an l-dimensional cube $0 \leqq z_i \leqq 1$ and, consequently, is bounded and has a finite ε-net. The number of elements of the minimum relative ε-net A in E_l is $N(\varepsilon, A(x_1, \cdots, x_l)) = N^\Lambda(x_1, \cdots, x_l, \theta)$. The number of elements of the minimum proper ε-net is $N_0^\Lambda(x_1, \cdots, x_l; \varepsilon)$. If a probability measure P_X is specified on X and $x_1, \cdots, x_l$ is an independent random sample, while $N^\Lambda(\cdot)$ is measurable relative to the measure on the sequences $x_1, \cdots, x_l$, then there exists a mean ε-entropy (or simply ε-entropy)

$$H^\Lambda(\varepsilon, l) = \mathbf{E} \log_2 N^\Lambda(x_1, \cdots, x_l; \varepsilon).$$

It is not difficult to see that the following inequality holds for a minimum relative ε-net:

$$N^\Lambda(x_1, \cdots, x_l, \cdots, x_{l+k}; \varepsilon) \leqq N^\Lambda(x_1, \cdots, x_l; \varepsilon) N^\Lambda(x_{l+1}, \cdots, x_{l+k}; \varepsilon), \tag{5}$$

if

$$\rho(x, y) = \max_i |x_i - y_i|.$$

Indeed, in this case a direct product of relative ε-nets will again be a relative ε-net. Then

$$H^{\Lambda}(\varepsilon, l+k) \leqq H^{\Lambda}(\varepsilon, l) + H^{\Lambda}(\varepsilon, k). \tag{6}$$

From (4) and (6) it can be shown (cf. [3]) that the limit $c(\varepsilon)$ exists:

$$c(\varepsilon) = \lim_{l\to\infty} \frac{H^{\Lambda}(\varepsilon, l)}{l}, \qquad 0 \leqq c(\varepsilon) \leqq \log\left[1 + \frac{1}{\varepsilon}\right].$$

It is obvious that $c(\varepsilon)$ does not decrease with a decrease in ε. Analogously to [1] we can show that

$$\frac{\log_2 N^{\Lambda}(x_1, \cdots, x_l, \varepsilon)}{l} \xrightarrow{P} c(\varepsilon) \quad \text{as } l \to \infty. \tag{7}$$

We shall consider two cases.

1. $\lim_{l\to\infty} (H^{\Lambda}(\varepsilon, l)/l) = c(\varepsilon) = 0$ for all $\varepsilon > 0$.

2. There exists $\varepsilon_0 > 0$ such that $c(\varepsilon_0) > 0$ (then for all $\varepsilon < \varepsilon_0$ we also have $c(\varepsilon) > 0$).

In the first case, it follows from (4) and (7) that

$$\frac{\log_2 N_2^{\Lambda}(x_1, \cdots, x_l; \varepsilon)}{l} \xrightarrow{P} 0, \qquad l \to \infty. \tag{8}$$

In the second case, it follows from (3) and (7) that

$$\lim_{l\to\infty} \mathbf{P}\left\{\frac{\log_2 N_0^{\Lambda}(x_1, \cdots, x_l; \varepsilon)}{l} > c(\varepsilon) - \delta\right\} = 1 \tag{9}$$

for all $\varepsilon \leqq \varepsilon_0$ and $\delta > 0$.

In this article we shall show that uniform convergence of means to their expectations follows from (8), while no such convergence exists in the case of the condition (9). Thus, necessary and sufficient conditions for uniform convergence are given by the following theorem.

Theorem. *For the uniform convergence of means to their expectations, with respect to a uniformly bounded class of functions $F(x, \alpha)$, it is necessary and sufficient that, for every $\varepsilon > 0$, the equality*

$$\lim_{l\to\infty} \frac{H^{\Lambda}(\varepsilon, l)}{l} = 0$$

be satisfied. The next section is devoted to the proof of this theorem and some of its corollaries.

2. Quasi-cube

We define by induction an n-dimensional quasi-cube with edge a.

DEFINITION. A set Q of the space E_1 is called a *one-dimensional quasi-cube* with edge a, if Q is a segment $[c, c+a]$.

A set Q of the space E_n is called an *n-dimensional quasi-cube* with edge a if there exists a coordinate subspace E_{n-1} (in the following, for the sake of simplicity we shall assume that it is formed by the first $n-1$ coordinates) such that the projection Q' of the set Q onto this subspace is an $(n-1)$-dimensional quasi-cube with edge a, and for each point $x^* \in Q'$ $(x^* = x_1^*, \cdots, x_{n-1}^*)$ the set of numerical values x_n such that $(x_1^*, \cdots, x_{n-1}^*, x_n) \in Q$ forms a segment $[c, c+a]$, where, generally speaking, c depends on x^*.

The space E_{n-1} will be called an $(n-1)$-dimensional *canonical* space. For it, in turn, an $(n-2)$-dimensional canonical space E_{n-2} can be determined, and so on.

The collection of subspaces $E_1, \cdots, E_{n-1}$ is called the canonical structure.

The following lemma is valid.

Lemma 1. *Let a convex set A belong to an l-dimensional cube the coordinates of which satisfy the condition*

$$0 \leqq x_i \leqq 1, \qquad i = 1, 2, \cdots, l.$$

Let $V(A)$ be an l-dimensional volume of the set A.

Then, if for certain $1 \leqq n \leqq l$, $0 \leqq a \leqq 1$, $l > 1$ the condition

$$V(A) > C_l^n a^{l-n} \tag{10}$$

is fulfilled, there exists a coordinate n-dimensional subspace such that a projection of the set A onto that subspace contains a quasi-cube with edge a.

PROOF. We shall prove the lemma by induction.

1. For $n = l$ condition (10) is

$$V(A) > C_l^n = 1. \tag{11}$$

On the other hand,

$$V(A) \leqq 1. \tag{12}$$

Therefore condition (10) is never fulfilled and the statement of the lemma is trivially true.

2. For $n = 1$ and any l we shall prove the lemma by contradiction. Let there exist no homogeneous coordinate space, a projection of the set A onto which contains a segment $[c, c+a]$.

The projection of a bounded convex set onto a one-dimensional axis is always an interval, a segment or a half-segment. Consequently, according to the assumption, the length of this straight line is not greater than a.

But then the set A itself is included in the (ordinary) cube with edge a. Hence it follows that

$$V(A) \leqq a^l,$$

and, taking into account the fact that $a \leqq 1$, we obtain

$$V(A) \leqq a^l \leqq l a^{l-1},$$

which contradicts condition (10) of the lemma.

3. We consider now the general inductive step. Let the lemma be true for all $n \leqq n_0$ in the case of all l, and for $n = n_0 + 1$ in the case of all l such that $n \leqq l \leqq l_0$. We shall show that it is true for $n = n_0 + 1$, $l = l_0 + 1$.

We consider a coordinate subspace E_{l_0} of dimension l_0, consisting of vectors $x = x_1, \cdots, x_{l_0}$.

Let A^{I} be the projection of A onto this subspace (A^{I} obviously is convex). If

$$V(A^{\mathrm{I}}) > C_{l_0}^n a^{l_0-n}, \tag{13}$$

then in view of the inductive assumption there exists a subspace of dimension n such that a projection of the set A^{I} onto that subspace contains a quasi-cube with edge a. Thus, in case (13) holds the lemma is proved.

Let

$$V(A^{\mathrm{I}}) \leqq C_l^n a^{l_0-n}. \tag{14}$$

We consider two functions

$$\varphi_1(x_1, \cdots, x_{l_0}) = \sup_x \{x : (x_1, \cdots, x_{l_0}, x) \in A\},$$

$$\varphi_2(x_1, \cdots, x_{l_0}) = \inf_x \{x : (x_1, \cdots, x_{l_0}, x) \in A.$$

These functions are convex upwards and downwards, respectively. Therefore the function

$$\varphi_3(x_1, \cdots, x_{l_0}) = \varphi_1(x_1, \cdots, x_{l_0}) - \varphi_2(x_1, \cdots, x_{l_0})$$

is convex upwards.

We consider the set

$$A^{\mathrm{II}} = \{x_1, \cdots, x_{l_0}: \varphi_3(x_1, \cdots, x_{l_0}) > a\}. \tag{15}$$

This set is convex and lies in E_{l_0}.

For the set A^{II} one of the following two inequalities is fulfilled:

$$V(A^{\mathrm{II}}) < C_{l_0}^{n-1} a^{l_0-n+1}, \tag{16}$$

$$V(A^{\mathrm{II}}) \leqq C_{l_0}^{n-1} a^{l_0-n+1}. \tag{17}$$

We assume that the inequality (16) is satisfied.

Then, in view of the inductive assumption, there exists a coordinate subspace E_{n-1} of the space E_{l_0} such that a projection A^{III} of the set A^{II} onto it contains an $(n-1)$-dimensional quasi-cube Ω_{n-1} with edge a. We now consider the n-dimensional coordinate subspace E_n formed by E_{n-1} and the coordinate $x = x_{l_0+1}$. Let, further, A^{IV} be the projection of the set A onto the space E_n.

For a given point $x_1, \cdots, x_{n-1} \in A^{\mathrm{III}}$ we consider a set $d(x_1, \cdots, x_{n-1})$ of values z such that $x_1, \cdots, x_{n-1}, z \in A^{\mathrm{IV}}$.

It is not difficult to see that the set $d(\cdot)$ includes the interval with the endpoints

$$r_1(x_1, \cdots, x_{n-1}) = \sup_{x \in A^{\mathrm{II}}}{}' \varphi_1(x_1, \cdots, x_{l_0}),$$

$$r_2(x_1, \cdots, x_{n-1}) = \inf_{x \in A^{\mathrm{II}}}{}' \varphi_2(x_1, \cdots, x_{l_0}),$$

where sup′ and inf′ are taken over the points $x \in A^{\mathrm{II}}$ which are projected into the given point $x_1, \cdots, x_{n-1}$. It is obvious that $r_1 - r_2 > a$ in view of (15). To each point $x_1, \cdots, x_{n-1} \in A^{\mathrm{III}}$ we let correspond a segment $c(x_1, \cdots, x_{n-1})$ of length a on the x_{l_0+1}-axis

$$\left[\tfrac{1}{2}(r_1(x_1, \cdots, x_{n-1}) + r_2(x_1, \cdots, x_{n-1})) - \frac{a}{2},\right.$$

$$\left.\tfrac{1}{2}(r_1(x_1, \cdots, x_{n-1}) + r_2(x_1, \cdots, x_{n-1})) + \frac{a}{2}\right].$$

Obviously, $c(x_1, \cdots, x_{n-1}) \subset d(x_1, \cdots, x_{n-1})$.

We now consider the set $Q \subset E_n$ consisting of points $(x_1, \cdots, x_{n-1}, x_{l_0+1})$ such that

$$(x_1, \cdots, x_{n-1}) \in \Omega_{n-1}. \tag{18}$$

$$x_{l_0+1} \in c(x_1, \cdots, x_{n-1}). \tag{19}$$

This set is in fact the required quasi-cube Ω_n. Indeed, in view of (18) and (19), the set Q satisfies the definition of an n-dimensional quasi-cube with edge a. At the same time, by construction, $Q \in A^{\mathrm{IV}}$.

For the proof of the lemma it remains for us to consider the case where the inequality (17) is fulfilled, i.e.,

$$V(A^{\mathrm{II}}) \leqq C_{l_0}^{n-1} a^{l_0-n+1}.$$

Then

$$V(A) = \int \varphi_3(x_1, \cdots, x_{l_0})\, dx_1, \cdots, dx_{l_0}$$

$$= \int_{A^{\mathrm{I}}-A^{\mathrm{II}}} \varphi_3(x_1, \cdots, x_{l_0})\, dx_1, \cdots, dx_{l_0} + \int_{A^{\mathrm{II}}} \varphi_3(x_1, \cdots, x_{l_0})\, dx_1, \cdots, dx_{l_0}$$

$$\leqq aV(A^{\mathrm{I}}) + V(A^{\mathrm{II}}).$$

Now, in view of (14) and (17), we obtain

$$V(A) \leqq C_{l_0}^{n} a^{l_0-n+1} + C_{l_0}^{n-1} a^{l_0-n+1} = C_{l_0+1}^{n} a^{(l_0+1)-n}$$

in contradiction to the condition of the lemma.

3. ε-Extension of a Set

Let A be a convex set bounded in E_l. To each point $x \in A$ we let correspond an open cube $\Omega(x)$ with center at x and side ε, oriented along the coordinate axes.

Along with the set A we consider the set

$$A_\varepsilon = \bigcup_{x \in A} \Omega(x),$$

which we shall call the ε-extension of the set A. The set A_ε is the set of points $y = (y_1, \cdots, y_l)$, for each of which there exists a point $x \in A$ such that

$$\rho(x, y) < \frac{\varepsilon}{2}.$$

It is easy to show that the ε-extension A_ε of the convex set A is convex.

We now isolate on the set A the minimum proper ε-net. Let the minimum number of elements of the proper ε-net of the set A be $N_0(\varepsilon, A)$. We denote by $V(A_\varepsilon)$ the volume of the set A_ε.

Lemma 2. *The inequality*

$$N_0(1.5\varepsilon, A)\varepsilon^l \leqq V(A_\varepsilon)$$

is valid.

PROOF. Let T be a certain proper $\frac{1}{2}\varepsilon$-net of the set A. We select from T a subset T' according to the following rule.

(i) As the first point x^1 of the set T' we select any point from T.

(ii) Let there already be selected m different points $x^1, \cdots, x^m$. As the $(m+1)$th point we take any point x from T such that

$$\min_{i=1,2,\cdots,m} \rho(x, x^i) \geqq \varepsilon.$$

(iii) If there is no such point in T, or T is already exhausted, then the construction is terminated.

The set T' thus constructed is a 1.5ε-net in A. Indeed, for every $x \in A$ there exists $y \in T$ such that $\rho(x, y) < \varepsilon/2$. For this y there exists $z \in T'$ such that $\rho(y, z) < \varepsilon$.

Consequently, $\rho(x, z) < 1.5\varepsilon$ and the number of elements in T' is not less than $N_0(1.5\varepsilon, A)$.

Furthermore, the union of open cubes with side ε and center at points of T' is included in A_ε. At the same time these cubes do not intersect. (Otherwise, there would exist $z \in \Omega(x^1)$ and $z \in \Omega(x^2)$, x^1, $x^2 \in T'$ and, consequently, $\rho(x^1, z) < \varepsilon/2$ and $\rho(x^2, z) < \varepsilon/2$. Hence $\rho(x^1, x^2) < \varepsilon$, and thus $x^1 = x^2$.)

Consequently,

$$V(A_\varepsilon) \geqq N_0(1.5\varepsilon, A)\varepsilon^l.$$

The lemma is proved.

Lemma 3. *Let a convex set A belong to a unit cube in E_l, and A_ε be its ε-extension, $0 < \varepsilon \leqq 1$, and let, for a certain $\gamma > \log(1+\varepsilon)$, the inequality*

$$N_0(1.5\varepsilon, A) > e^{\gamma l} \tag{20}$$

be satisfied. Then there exist $t_0(\varepsilon, \gamma)$ and $a(\varepsilon, \gamma)$ such that if $n = [t_0 l] > 0$, then there exists a coordinate subspace of dimension $n = [t_0 l]$, the projection of the set A_ε onto which contains an n-dimensional quasi-cube with edge a.

PROOF. According to Lemmas 1 and 2 and condition (20) of the present lemma, in order that there exist an n-dimensional coordinate subspace, the projection of A_ε onto which would contain an n-dimensional quasi-cube with edge a, it is sufficient that

$$C_l^n b^{l-n} < e^{\gamma l}\varepsilon^l(1+\varepsilon)^{-l}$$

holds, where $b = a/(1+\varepsilon)$.

In turn, from Stirling's formula it follows that for this it is sufficient for

$$b^{l-n}\frac{l^n e^n}{n^n} < e^{\gamma_1 l}\varepsilon^l,$$

to hold, where $\gamma_1 = \gamma - \log(1+\varepsilon)$.

Setting $t = n/l$ and assuming that $0 < t < 1/3$, we obtain, by an equivalent transformation,

$$-\frac{t(\log t - 1)}{1-t} + \log b < \frac{\log \varepsilon + \gamma_1}{1-t}.$$

Under our constraints, this inequality is satisfied if the inequality

(21) $$-\tfrac{3}{2}t(\log t - 1) + \log b < \log \varepsilon(1+2t) + \tfrac{2}{3}\gamma_1$$

is fulfilled. We now choose $t_0(\gamma, \varepsilon)$ so that the conditions

$$0 < t_0(\varepsilon, \gamma) \leqq \frac{1}{3}, \quad \frac{-3}{2}t_0(\log t_0 - 1) < \frac{\gamma_1}{6}, \quad -2t_0 \log \varepsilon < \frac{\gamma_1}{6}$$

are satisfied.

This can always be done since, by condition, $\gamma_1 > 0$. It is obvious that for $0 < t \leqq t_0$ these conditions are also fulfilled, and in such a case, (21) is satisfied for

$$\log b = \log \varepsilon + \frac{\gamma_1}{3},$$

or

(22) $$a = (1+\varepsilon)\varepsilon\, e^{(\gamma - \log \varepsilon)/3}.$$

The lemma is proved.

4. Auxiliary Lemma

We now consider a class of functions $\Phi = \{F(x, \alpha)\}$, parametrized with respect to $\alpha \in \Lambda$, which is defined on X. We assume the class to be convex in the sense that if

(23) $$F(x, \alpha_1), \cdots, F(x, \alpha_2) \in \Phi,$$

then

$$\sum_{i=1}^{r} \tau_i F(x, \alpha_i) \in \Phi, \quad \sum_{i=1}^{r} \tau_i = 1, \quad \tau_i \geqq 0.$$

We define two sequences: the sequence

$$x_1, \cdots, x_l, \qquad x_i \in X,$$

and the random independent sequence

$$y_1, \cdots, y_l,$$

possessing the property

(24) $$y_i = \begin{cases} 1 & \text{with probability } \tfrac{1}{2}, \\ -1 & \text{with probability } \tfrac{1}{2}. \end{cases}$$

With respect to these sequences, we determine the quantity

$$Q(\Phi)=\mathbf{E}\sup_{F(x,\alpha)\in\Phi}\frac{1}{l}\left|\sum_{i=1}^{l}F(x_i,\alpha)y_i\right|.$$

(The expectation is computed with respect to the random sequences (24).

In Section 1, we denoted by A a set of l-dimensional vectors with coordinates $z_i=F(x_i,\alpha)$, $i=1,2,\cdots,l$, for all possible $\alpha\in\Lambda$. It is obvious that A belongs to a unit l-dimensional cube in E_l and is convex.

We rewrite the function $Q(\Phi)$ in the form

$$Q(\Phi)=\mathbf{E}\sup_{t\in A}\frac{1}{l}\left|\sum_{i=1}^{l}z_iy_i\right|.$$

Lemma 4. *If, for $\varepsilon>0$, the inequality*

$$N_0(1.5\varepsilon,A)>e^{\gamma l},\qquad \gamma>\log(1+\varepsilon)$$

is fulfilled for the set A, then the inequality

$$Q(\Phi)\geqq\varepsilon(e^{(\gamma-\log(1+\varepsilon))/3}-1)\left(\frac{t}{2}-\frac{1}{2l}\right)$$

holds, where $t>0$ does not depend on l.

PROOF. As was shown in the previous section, if the conditions of the lemma are satisfied, there exist $t(\varepsilon,\gamma)$ and $a(\varepsilon,\gamma)$ such that there exists a coordinate subspace of dimensionality $n=[tl]$, the projection of the set A_ε onto which contains an n-dimensional quasi-cube with edge a. Without loss of generality, we assume that this subspace is formed by the first n coordinates, while the corresponding n-dimensional subspace is a canonical subspace of this quasi-cube.

We determine the vertices of the quasi-cube by means of the following iterative rule.

(i) The ends of the segment, c and $c+a$, are the vertices of the one-dimensional quasi-cube.

(ii) Let us determine the vertices of the n-dimensional quasi-cube. Let the vertices of the $(n-1)$-dimensional quasi-cube be determined in the $(n-1)$-dimensional canonical space. With each such vertex $x_1^k,\cdots,x_{n-1}^k$ (k is the number of the vertices) we match the segment

$$\left[\varphi^{n-1}(x_1^k,\cdots,x_{n-1}^k)-\frac{a}{2},\varphi^{n-1}(x_1^k,\cdots,x_{n-1}^k)+\frac{a}{2}\right],$$

where we have denoted:

$$\varphi^{n-1}(x_1^k,\cdots,x_{n-1}^k)=\tfrac{1}{2}(\varphi_1(x_1^k,\cdots,x_{n-1}^k)+\varphi_2(x_1^k,\cdots,x_{n-1}^k)),$$

$$\varphi_1(x_1,\cdots,x_{n-1})=\max_{x_n}\{x_n\colon(x_1,\cdots,x_n)\in\Omega_n\},$$

$$\varphi_2(x_1,\cdots,x_{n-1})=\min_{x_n}\{x_n\colon(x_1,\cdots,x_n)\in\Omega_n\},$$

Ω_n being an n-dimensional quasi-cube. This segment is formed by the intersection of the straight line $(x_1^k, \cdots, x_{n-1}^k, x)$ with the quasi-cube. The ends of the segment form the vertices of the quasi-cube.

Thus, if

$$(x_1^k, \cdots, x_{n-1}^k) \in E_{n-1}$$

is the kth vertex of the $(n-1)$-dimensional quasi-cube, then

$$\left(x_1^k, \cdots, x_{n-1}^k, \varphi^{n-1}(x_1^k, \cdots, x_{n-1}^k) - \frac{a}{2}\right),$$

$$\left(x_1^k, \cdots, x_{n-1}^k, \varphi^{n-1}(x_1^k, \cdots, x_{n-1}^k) + \frac{a}{2}\right)$$

are, respectively, the $(2k-1)$th and $2k$th vertices of the n-dimensional quasi-cube.

We now match with every sequence

$$y_1, \cdots, y_n, \qquad y_i = -1, 1,$$

the vertex $\bar{x}$ of the quasi-cube, given by

$$\bar{x}_1 = \left(c + \frac{a}{2}\right) + \frac{a}{2} y_1, \qquad \bar{x}_k = \varphi^{k-1}(\bar{x}_1, \cdots, \bar{x}_{k-1}) + \frac{a}{2} y_k,$$

$$\bar{x}_n = \varphi^{n-1}(\bar{x}_1, \cdots, \bar{x}_{n-1}) + \frac{a}{2} y_n.$$

In turn, the vertex $\bar{x}$ of the quasi-cube in E_n is matched with a point $\bar{z} = (\bar{z}_1, \cdots, z_l) \in A$ such that the projection $\hat{z} = (\bar{z}_1, \cdots, \bar{z}_n)$ of this point in E_n is not further than $\varepsilon/2$ away from $\bar{x}$, i.e.,

$$|\bar{x} - \bar{z}_k| < \frac{\varepsilon}{2}, \qquad k = 1, 2, \cdots, n.$$

This is possible, since $\bar{x} \in \Pr A_\varepsilon$ onto E_n. Thus we have introduced two functions:

$$\bar{x} = \bar{x}(y_1, \cdots, y_n), \qquad \bar{z} = \bar{z}(\bar{x}_1, \cdots, \bar{x}_n).$$

We denote $\delta_i = \bar{z}_i - \bar{x}_i$, $|\delta_i| \leqq \varepsilon/2$, $i = 1, \cdots, n$. We now evaluate the quantity

$$Q(\Phi) = \mathbf{E} \sup_{z \in A} \frac{1}{l} \left| \sum_{i=1}^{l} z_i y_i \right| \geqq \frac{1}{l} \mathbf{E} \sum_{i=1}^{l} \bar{z}_i y_i$$

$$= \frac{1}{l} \sum_{i=1}^{n} \mathbf{E} y_i(\bar{x}_i + \delta_i) + \frac{1}{l} \sum_{i=n+1}^{l} \mathbf{E} y$$

We note that the second term of the sum is equal to zero, since each term of the sum is the product of two independent random quantities y_i and $\bar{z}_i$, $i > n$, one of which (y_i) has mean zero.

We evaluate the first summand. For this we consider the first term of the first summand

$$\frac{1}{l}\mathbf{E}\left[y_1\left(c+\frac{a}{2}+\frac{a}{2}y_1+\delta_1\right)\right]$$

$$=\frac{1}{l}\left[\left(c+\frac{a}{2}\right)\mathbf{E}y_1+\frac{a}{2}\mathbf{E}y_1^2+\mathbf{E}(y_1\delta_1)\right]\geqq\left(\frac{a}{2l}-\frac{\varepsilon}{2l}\right).$$

For the evaluation of the k-th term

$$I_k=\frac{1}{l}\mathbf{E}\left[y_k\left(\varphi^{k-1}(\bar{x}_1,\cdots,\bar{x}_{k-1})+\frac{a}{2}y_k+\delta_k\right)\right]$$

we note that the vertex $\bar{x}_1,\cdots,\bar{x}_{k-1}$ was chosen so that it depended on $y_1,\cdots,y_{k-1}$, but not on y_k. Therefore,

$$I_k=\frac{1}{l}\left[\frac{a}{2}+\mathbf{E}y_k\delta_k\right]\geqq\frac{1}{2l}(a-\varepsilon).$$

Thus, we obtain

$$Q(\Phi)\geqq\mathbf{E}\sup_{z\in A}\frac{1}{l}\sum_{i=1}^{l}z_iy_i\geqq\frac{n}{2l}(a-\varepsilon)\geqq(a-\varepsilon)\left(\frac{t}{2}-\frac{1}{2l}\right).$$

If we choose the quantity a in accordance with (22), we obtain

$$Q(\Phi)\geqq\varepsilon\left(\exp\left(\frac{\gamma-\log(1+\varepsilon)}{3}\right)-1\right)\left(\frac{t}{2}-\frac{1}{2l}\right).$$

The lemma is proved.

5. Necessary and Sufficient Conditions for the Uniform Convergence of Means to Their Expectations

Let $A(x_1,\cdots,x_l)$ be a set of vectors z with coordinates $z_1,\cdots,z_l$, $z_i=F(x_i,\alpha)$, $i=1,2,\cdots,l$, formed by all $\alpha\in\Lambda$.

Let

$$H^\Lambda(\varepsilon,l)=\mathbf{E}\log_2 N^\Lambda(x_1,\cdots,x_l;\varepsilon)$$

be the ε-entropy of the set A relative to E_l.

We proceed to the proof of the theorem, according to which necessary and sufficient conditions for uniform convergence of means to the expectations with respect to a uniformly bounded class of functions are given by the condition

$$\lim_{l\to\infty}\frac{H^\Lambda(\varepsilon,l)}{l}=0\quad\text{for all }\varepsilon>0. \tag{25}$$

When proving necessity, we shall assume without loss of generality that the class $F(x,\alpha)$ is convex in the sense (23), since the uniform convergence of means to the expectations for an arbitrary class implies such a convergence for its convex closure, while the condition (25) for the convex closure implies such a convergence for the original class of functions.

PROOF OF NECESSITY. We assume the contrary. Let for some $\varepsilon_0>0$ the equality

$$\lim_{l\to\infty}\frac{H^{\Lambda}(\varepsilon_0 l)}{l}=c>0 \tag{26}$$

be fulfilled, and at the same time assume uniform convergence, i.e., for all $\varepsilon>0$ the relation

$$\lim_{l\to\infty}\mathbf{P}\left\{\sup_{\alpha}\left|\mathbf{E}F(x,\alpha)-\frac{1}{l}\sum_{i=1}^{l}F(x_i,\alpha)\right|>\varepsilon\right\}=0. \tag{27}$$

is fulfilled. We shall obtain a contradiction.

Since the functions $\mathbf{E}F(x,\alpha)$ and $l^{-1}\sum_{i=1}^{l}F(x_i,\alpha)$ are uniformly bounded by the quantity 1 for all l, it follows from (27) that

$$\lim_{l\to\infty}\mathbf{E}\left(\sup_{\alpha}\left|\mathbf{E}F(x,\alpha)-\frac{1}{l}\sum_{i=1}^{l}F(x_i,\alpha)\right|\right\}=0.$$

Hence it follows that if $l_1\to\infty$ and $l-l_1\to\infty$, then the equation

$$\lim_{l_1,l-l_1\to\infty}\mathbf{E}\left\{\sup_{\alpha}\left|\frac{1}{l_1}\sum_{i=1}^{l_1}F(x_i,\alpha)-\frac{1}{l-l_1}\sum_{i=l_1+1}^{l}F(x_i,\alpha)\right|\right\}=0 \tag{28}$$

is satisfied. We consider the expression

$$I(x_1,\cdots,x_l)=\sum_{n=0}^{l}\sup_{\alpha}\left\{\frac{C_l^n}{2^l}\frac{1}{l}\left|\sum_{i=1}^{n}F(x_i,\alpha)-\sum_{i=n+1}^{l}F(x_i,\alpha)\right|\right\}.$$

We divide the region of summation over n into two:

$$\text{I: }\left|n-\frac{l}{2}\right|<l^{2/3},$$

$$\text{II: }\left|n-\frac{l}{2}\right|\geqq l^{2/3}.$$

Then, taking into account the fact that

$$\frac{1}{l}\left|\sum_{i=1}^{n}F(x_i,\alpha)-\sum_{i=n+1}^{l}F(x_i,\alpha)\right|\leqq 1,$$

we obtain

$$I(x_1,\cdots,x_l)\leqq\sum_{n\in\text{II}}\frac{C_l^n}{2^l}$$
$$+\sum_{n\in\text{I}}\frac{C_l^n}{2^l}\sup_{\alpha}\left|\frac{n}{l}\left(\frac{1}{n}\sum_{i=1}^{n}F(x_i,\alpha)\right)-\frac{l-n}{l}\left(\frac{1}{l-n}\sum_{i=n+1}^{l}F(x_i,\alpha)\right)\right|.$$

We note that in the region I, $\frac{1}{2}-l^{-1/3}<n/l<\frac{1}{2}+l^{-1/3}$,

$$\sum_{n=\text{I}}\frac{C_l^n}{2^l}\to 1\quad\text{as } l\to\infty,$$

while in the region II,

$$\sum_{n\in \mathrm{II}} \frac{C^n}{2^l} \to 0 \quad \text{as } l\to\infty. \tag{29}$$

We evaluate the quantity

$$\begin{aligned}&\lim_{l\to\infty} \mathbf{E} I(x_1,\cdots,x_l)\\ &\leqq \lim_{l\to\infty}\left(\sum_{n\in \mathrm{II}} \frac{C_l^n}{2^l}+\frac{1}{2}\max_{n\in \mathrm{I}} \mathbf{E}\sup_\alpha \left|\frac{1}{n}\sum_{i=1}^n F(x_i,\alpha)\right.\right.\\ &\qquad\left.\left.-\frac{1}{l-n}\sum_{i=n+1}^l F(x_i,\alpha)\right|\sum_{n\in \mathrm{I}}\frac{C_l^n}{2^l}\right).\end{aligned}$$

From (28) it follows that

$$\max_{n\in \mathrm{I}} \mathbf{E}\sup_\alpha \left|\frac{1}{n}\sum_{i=1}^n F(x_i,\alpha)-\frac{1}{l-n}\sum_{i=n+1}^l F(x_i,\alpha)\right|\to 0 \quad \text{as } l\to\infty.$$

Therefore, taking into account (29), we obtain

$$\lim_{l\to\infty} \mathbf{E} I(x_1,\cdots,x_l)=0. \tag{30}$$

On the other hand,

$$\mathbf{E} I(x_1,\cdots,x_l)=\mathbf{E}\frac{1}{l!}\sum_{k=1}^{l!} I(T_k(x_1,\cdots,x_l)),$$

where T_k, $k=1,\cdots,l!$, are all permutations of the sequence. We transform the right side of the expression

$$\begin{aligned}&\mathbf{E}\frac{1}{l!}\sum_{k=1}^{l!} I(T_k(x_1,\cdots,x_l))\\ &=\mathbf{E}\frac{1}{l!}\sum_{k=1}^{l!}\sum_{n=0}^{l}\sup_\alpha\left[\frac{C_l^n}{2^l}\frac{1}{l}\left|\sum_{i=1}^n F(x_{j(i,k)},\alpha)-\sum_{i=n+1}^l F(x_{j(i,k)},\alpha)\right|\right]\\ &=\mathbf{E}\sum_{n=0}^l \frac{1}{C_l^n}\sum_{y_1,\cdots,y_l}{}' \sup_\alpha \frac{C_l^n}{2^l}\frac{1}{l}\left|\sum_{i=1}^l y_i F(x_i,\alpha)\right|.\end{aligned}$$

Here by $j(i,k)$ we have denoted the index into which the index i is transformed in the permutation T_k. In the last expression the summation is carried out over all sequences

$$y_1,\cdots,y_l,\qquad y_i=\{-1,1\},$$

for which the number of positive values equals n. We further obtain

$$\mathbf{E} I(x_1,\cdots,x_l)=\mathbf{E}\left\{\frac{1}{2^l}\sum_{y_1,\cdots,y_l}\sup_\alpha \frac{1}{l}\left|\sum_{i=1}^l g_i F(x_i,\alpha)\right|\right\}. \tag{31}$$

In (31) the summation is carried out over all sequences

$$y_1,\cdots,y_l,\qquad y_i=\{-1,1\}$$

As $\varepsilon_0 > 0$ we select a number such that

$$\lim_{l\to\infty} \frac{H^\Lambda(\varepsilon_0, l)}{l} = c(\varepsilon_0) > 0.$$

Since $c(\varepsilon)$ does not decrease as ε decreases, we can choose ε so that the relations

$$0 < 1.5\varepsilon \leqq \varepsilon_0, \quad \log(1+\varepsilon) < \frac{c(\varepsilon_0)\log 2}{2}, \quad c(1.5\varepsilon) \geqq c(\varepsilon_0),$$

are fulfilled. Then in view of (9) the probability of fulfilment of the inequality

$$N_0^\Lambda(x_1, \cdots, x_l; 1.5\varepsilon) > \exp\left[c(\varepsilon_0)\frac{\log 2}{2}\right] l \tag{32}$$

tends to 1. According to Lemma 4, when (32) is satisfied, the expression in curly brackets in (31) exceeds the quantity

$$\left(\frac{t}{2} - \frac{1}{2l}\right)\varepsilon(e^{\gamma/3} - 1),$$

where

$$\gamma = \frac{c(\varepsilon_0)\log 2}{2} - \log(1+\varepsilon)$$

and $t > 0$ does not depend on l. Hence we conclude that

$$\lim_{l\to\infty} I(x_1, \cdots, x_l) > \lim_{l\to\infty} \varepsilon(e^{\gamma/3} - 1)\left(\frac{t}{2} - \frac{1}{2l}\right) > 0.$$

This inequality contradicts the statement (30). The contradiction thus obtained proves the first part of the theorem.

6. Necessary and Sufficient Conditions for the Uniform Convergence of Means to Their Expectations (Proof of Sufficiency)

The following lemma is valid.

Lemma 5. *If, for every $\varepsilon > 0$, the relation*

$$\mathbf{P}\left\{\sup_\alpha \left|\frac{1}{l}\sum_{i=1}^{l} F(x_i, \alpha) - \frac{1}{l}\sum_{i=l+1}^{2l} F(x_i, \alpha)\right| > \varepsilon\right\} \to 0 \quad \text{as } l \to \infty,$$

holds, then for every $\varepsilon > 0$ the convergence

$$\mathbf{P}\left\{\sup_\alpha \left|\frac{1}{l}\sum_{i=1}^{l} F(x_i, \alpha) - \mathbf{E}F(x, \alpha)\right| > \varepsilon\right\} \to 0 \quad \text{as } l \to \infty,$$

is valid.

Proof. We assume the contrary. Let, for $\varepsilon_0 > 0$,

$$\lim_{l\to\infty} \mathbf{P}\left\{\sup_\alpha \left|\frac{1}{l}\sum_{i=1}^{l} F(x_i, \alpha) - \mathbf{E}F(x, \alpha)\right| > \varepsilon_0\right\} \neq 0.$$

We denote by R_l the event

$$\sup_\alpha \left| \frac{1}{l} \sum_{i=1}^{l} F(x_i, \alpha) - \mathbf{E}F(x, \alpha) \right| > \varepsilon_0.$$

Then for sufficiently large l the inequality

$$\mathbf{P}(R_l) > \eta > 0$$

is fulfilled. We introduce the notation

$$\frac{1}{l} \left| \sum_{i=1}^{l} F(x_i, \alpha) - \sum_{i=l+1}^{2l} F(x_i, \alpha) \right| = S(x_1, \cdots, x_{2l}, \alpha).$$

We consider the quantity

$$P_{2l} = \mathbf{P}\left\{ \sup_\alpha S(x_1, \cdots, x_{2l}, \alpha) > \frac{\varepsilon_0}{3} \right\}$$

$$= \underset{x_1, \cdots, x_{2l}}{\int \cdots \int} \theta\left[\sup_\alpha S(x_1, \cdots, x_{2l}, \alpha) - \frac{\varepsilon_0}{3} \right] dP(x_1) \cdots dP(x_{2l}),$$

where we have denoted

$$\theta(z) = \begin{cases} 1 & \text{if } z \geqq 0, \\ 0 & \text{if } z < 0. \end{cases}$$

Further we have

$$P_{2l} \geqq \int_{R_l} \left\{ \int_{x_{l+1}} \cdots \int_{x_{2l}} \theta\left[\sup_\alpha S(x_1, \cdots, x_{2l}, \alpha) - \frac{\varepsilon_0}{3} \right] dP(x_{l+1}), \cdots dP(x_{2l}) \right\}$$
$$\cdot\, dP(x_1) \cdots dP(x_l).$$

Each point $x_1, \cdots, x_l$ from R_l can be put into correspondence with a value $\alpha^*(x_1, \cdots, x_l)$ such that

$$\left| \frac{1}{l} \sum_{i=1}^{l} F(x_i, \alpha^*) - \mathbf{E}F(x, \alpha^*) \right| > \frac{\varepsilon_0}{3}.$$

We denote by $\bar{R}_l$ an event in $X = \{x_{l+1}, \cdots, x_{2l}\}$:

$$\left| \frac{1}{l} \sum_{i=l+1}^{2l} F(x_i, \alpha^*) - \mathbf{E}F(x, \alpha^*) \right| < \frac{\varepsilon_0}{3}.$$

Since the function $F(x, \alpha)$ is uniformly bounded, we have

$$\mathbf{P}\{\bar{R}_l\} \to 1 \quad \text{as } l \to \infty.$$

Further,

$$P_{2l} \geqq \int_{R_l} \left\{ \int_{\bar{R}_l} \theta\left[S(x_1, \cdots, x_{2l}, \alpha^*(x_1, \cdots, x_l)) - \frac{\varepsilon_0}{3} \right] dP(x_{l+1}) \cdots dP(x_{2l}) \right\}$$
$$\cdot\, dP(x_1) \cdots dP(x_l).$$

But if $x_1, \cdots, x_l \in R_l$, while $x_{l+1}, \cdots, x_{2l} \in \bar{R}_l$, then the expression under the integral sign equals one. Assuming l to be so large that $\mathbf{P}\{\bar{R}_l\} > \frac{1}{2}$, we obtain

$$P_{2l} > \tfrac{1}{2} \int_{R_l} dP(x_1) \cdots dP(x_l) = \tfrac{1}{2} P(R_l),$$

and, consequently, $\lim_{l\to\infty} \mathbf{P}_{2l} \neq 0$, which contradicts the condition of the lemma.

PROOF. We shall prove that under the conditions of the theorem the relation

$$\mathbf{P}\{\sup_\alpha S(x_1, \cdots, x_{2l}, \alpha) > \varepsilon\} \to 0 \quad \text{as } l \to \infty, \tag{33}$$

holds. According to Lemma 5, from the condition (33) it follows that the statement of the theorem is fulfilled, i.e.,

$$\mathbf{P}\left\{\sup_\alpha \left|\frac{1}{l} \sum_{i=1}^{l} F(x_i, \alpha) - \mathbf{E}F(x, \alpha)\right| > \varepsilon\right\} \to 0 \quad \text{as } l \to \infty.$$

We shall show the validity of (33).

For this we note that in view of symmetry of the definition of the measure, the equation

$$\begin{aligned} &\mathbf{P}\{\sup_\alpha S(x_1, \cdots, x_{2l}, \alpha) > \varepsilon\} \\ &= \frac{1}{(2l)!} \sum_{j=1}^{(2l)!} \mathbf{P}\{\sup_\alpha S(T_j(x_1, \cdots, x_{2l}), \alpha) > \varepsilon\} \\ &= \int \left\{\frac{1}{(2l)!} \sum_{j=1}^{(2l)!} \theta[\sup_\alpha S(T_j(x_1, \cdots, x_{2l}), \alpha) - \varepsilon]\right\} dP(x_1) \cdots dP(x_{2l}) \end{aligned} \tag{34}$$

holds. Here T_j, $j = 1, 2, \cdots, (2l)!$, are all possible permutations of the indices $1, 2, \cdots, 2l$, $T_j(x_1, \cdots, x_{2l})$ is a sequence of arguments which is obtained by the permuation T_j from the sequence $x_1, \cdots, x_{2l}$.

We investigate the expression (34) under the integral sign:

$$K = \frac{1}{(2l)!} \sum_{j=1}^{(2l)!} \theta[\sup_\alpha S(T_j(x_1, \cdots, x_{2l}), \alpha) - \varepsilon].$$

Let A be a set of points in E_l with coordinates $z_i = F(x_i, \alpha)$ for all possible $\alpha \in \Lambda$.

Let $z(1), \cdots, z(N_0)$ be the minimum proper $\frac{1}{3}\varepsilon$-net in A, and $\alpha(1), \cdots, \alpha(N_0)$ values of α such that

$$z_i(k) = F(x_i, \alpha(k)), \qquad i = 1, \cdots, 2l, \quad k = 1, \cdots, N_0.$$

We shall show that if the relation

$$\max_k S(x_1, \cdots, x_{2l}, \alpha(k)) < \frac{\varepsilon}{3}$$

is satisfied, then

$$\sup_{\alpha \in \Lambda} S(x_1, \cdots, x_{2l}, \alpha) < \varepsilon.$$

Indeed, for every α we find $\alpha(k)$ such that

$$|F(x_i, \alpha)-F(x_i, \alpha(k))|<\frac{\varepsilon}{3}, \qquad i=1, 2, \cdots, 2l.$$

Therefore,

$$\begin{aligned}
&\left|\frac{1}{l}\sum_{i=1}^{l} F(x_i, \alpha)-\frac{1}{l}\sum_{i=l+1}^{2l} F(x_i, \alpha)\right| \\
&=\left|\frac{1}{l}\left(\sum_{i=1}^{l} F(x_i, \alpha)-\sum_{i=1}^{l} F(x_i, \alpha(k))\right)-\frac{1}{l}\left(\sum_{i=l+1}^{2l} F(x_i, \alpha)-\sum_{i=l+1}^{2l} F(x_i, \alpha(k))\right)\right. \\
&\qquad\left.+\frac{1}{l}\left(\sum_{i=1}^{l} F(x_i, \alpha(k))-\sum_{i=l+1}^{2l} F(x_i, \alpha(k))\right)\right| \\
&\leqq\frac{2}{3}\varepsilon+\frac{1}{l}\left|\sum_{i=1}^{l} F(x_i, \alpha(k))-\sum_{i=l+1}^{2l} F(x_i, \alpha(k))\right|<\varepsilon.
\end{aligned}$$

Analogous estimates hold for $S(T_j(x_1, \cdots, x_{2l})\alpha)$. Therefore,

$$\begin{aligned}
K&\leqq\frac{1}{(2l)!}\sum_{j=1}^{(2l)!}\theta\left[\max_k S(T_j(x_1, \cdots, x_{2l}), \alpha(k))-\frac{\varepsilon}{3}\right] \\
&\leqq\frac{1}{(2l)!}\sum_{j=1}^{(2l)!}\sum_{k=1}^{N_0}\theta\left[S(T_j(x_1, \cdots, x_{2l}), \alpha(k))-\frac{\varepsilon}{3}\right] \\
&=\sum_{k=1}^{N_0}\left\{\frac{1}{(2l)!}\sum_{j=1}^{(2l)!}\theta\left[S(T_j(x_1, \cdots, x_{2l}), \alpha(k))-\frac{\varepsilon}{3}\right]\right\}.
\end{aligned}$$

We evaluate the expression in the curly brackets:

$$R_1=\frac{1}{(2l)!}\sum_{j=1}^{(2l)!}\theta\left[\frac{1}{l}\left|\sum_{i=1}^{l} F(x_{t(i,j)}, \alpha(k))-\sum_{i=l+1}^{2l} F(x_{t(i,j)}, \alpha(k))\right|\right],$$

where $t(i, j)$ is the index into which the index i is transformed in the permutation T_j. We order the values $F(x_i, \alpha(k))$ by magnitude:

$$F(x_{i_1}, \alpha(k))\leqq F(x_{i_2}, \alpha(k))\leqq\cdots\leqq F(x_{i_{2l}}, \alpha(k)),$$

and denote $z_p=F(x_{i_p}, \alpha(k))$.

We further denote

$$\Delta_1=z_1, \qquad \Delta_p=z_p-z_{p-1},$$

$$\delta_{ip}=\begin{cases}1 & \text{if } F(x_i, \alpha(k))\leqq z_p,\\ 0 & \text{if } F(x_i, \alpha(k))>z_p,\end{cases}$$

$$r_i^j=\begin{cases}1 & \text{if } t^{-1}(i, j)\leqq l,\\ 0 & \text{if } t^{-1}(i, j)>l,\end{cases}$$

where $t^{-1}(i, j)$ is the index which is transformed into i in the permutation T_j. Then

$$\frac{1}{l}\left|\sum_{i=1}^{l} F(x_{t(i,j)}, \alpha(k)) - \sum_{i=l+1}^{2l} F(x_{t(i,j)}, \alpha(k))\right|$$
$$= \frac{1}{l}\left|\sum_p \Delta_p \sum_{i=1}^{2l} \delta_{ip} r_i^j - \sum_p \Delta_p \sum_{i=1}^{2l} \delta_{ip}(1 - r_i^j)\right|$$
$$= \sum_p \Delta_p \left|\frac{1}{l}\sum_{i=1}^{2l} \delta_{ip}(2r_i^j - 1)\right|.$$

Further, if the inequality

$$\max_p \frac{1}{l}\left|\sum_{i=1}^{2l} \delta_{ip}(2r_i^j - 1)\right| < \frac{\varepsilon}{3} \tag{35}$$

is valid, then the inequality

$$\sum_p \Delta_p \frac{1}{l}\left|\sum_{i=1}^{2l} \delta_{ip}(2r_i^j - 1)\right| < \frac{\varepsilon}{3}\sum_p \Delta_p \leqq \frac{\varepsilon}{3}$$

is fulfilled. The condition (35) is equivalent to the following:

$$\max_p \theta\left[\frac{1}{l}\left|\sum_{i=1}^{2l} \delta_{ip}(2r_i^j - 1)\right| - \frac{\varepsilon}{3}\right] = 0. \tag{36}$$

Thus, we obtain

$$\begin{aligned} R_1 &< \frac{1}{(2l)!}\sum_{j=1}^{(2l)!} \max_p \theta\left[\frac{1}{l}\left|\sum_{i=1}^{2l} \delta_{ip}(2r_i^j - 1)\right| - \frac{\varepsilon}{3}\right] \\ &\leqq \sum_p \left\{\sum_{j=1}^{(2l)!} \frac{1}{(2l)!}\theta\left[\frac{1}{l}\left|\sum_{i=1}^{2l} \delta_{ip}(2r_i^j - 1)\right| - \frac{\varepsilon}{3}\right]\right\}. \end{aligned} \tag{37}$$

In sampling without replacement from an urn let there be $2l$ balls of which $\sum_{i=1}^{2l} \delta_{1p} = m$ are black. From the urn we take l balls (without replacing them). Then the expression in the curly brackets in (37) is the probability of the fact that the number of black balls taken from the urn differs from the number of remaining black balls by not less than $\varepsilon l/3$.

This quantity equals

$$\Gamma = \sum_k \frac{C_m^k C_{2l-m}^{l-k}}{C_{2l}^l},$$

where k runs through all values for which

$$\left|\frac{k}{l} - \frac{m-k}{l}\right| \geqq \frac{\varepsilon}{3}.$$

It is known (cf. [1]) that the quantity Γ is evaluated as

$$\Gamma < 3 \exp\left[-\frac{\varepsilon^2(l-1)}{9}\right].$$

Thus,

$$R_1 < \sum_{p=1}^{2l} 3 \exp\left[-\frac{\varepsilon^2(l-1)}{9}\right] = 6l \exp\left[-\frac{\varepsilon^2(l-1)}{9}\right].$$

Returning to the estimate of K, we obtain

$$K < 6lN_0\left(x_1, \cdots, x_{2l}, \frac{\varepsilon}{3}\right) \exp\left[-\frac{\varepsilon^2(l-1)}{9}\right].$$

Finally, for any $c > 0$ the inequalities

$$\mathbf{P}\left\{\sup_\alpha \frac{1}{l}\left|\sum_{i=1}^{l} F(x_i, \alpha) - \sum_{i=l+1}^{2l} F(x_i, \alpha)\right| > \varepsilon\right\}$$

$$\leqq \int_{\log N_0(x_1,\cdots,x_{2l},\varepsilon/3)>cl} dP(x_1) \cdots dP(x_{2l})$$

$$+ \int_{\log N_0(x_1,\cdots,x_{2l},\varepsilon/3)\leqq cl} K\, dP(x_1) \cdots dP(x_{2l})$$

$$\leqq \mathbf{P}\left(\frac{\log N_0(x_1, \cdots, x_{2l}, \varepsilon/3)}{l} > c\right\} + 6l \exp\left[-\frac{\varepsilon^2(l+1)}{9} + cl\right]$$

hold. Setting $c < \varepsilon^2/10$, we find that the second term on the right side tends to zero as l increases. The first term tends to zero in view of the condition of the theorem and the relation (8) of Section 1.

7. Corollaries

Corollary 1. *For the uniform convergence of means to their expectations, it is necessary and sufficient that, for every $\varepsilon > 0$, the relation*

$$\lim_{l\to\infty} \frac{1}{l} \mathbf{E} \log V(A_\varepsilon(x_1, \cdots, x_l)) = \log \varepsilon,$$

is satisfied, where A_ε is the ε-extension of the set A.

Proof of necessity. Let $\varepsilon, \delta < 0, \delta < \varepsilon$, and let T_0 be the minimum proper δ-net of A with number of elements $N_0(x_1, \cdots, x_l, \delta)$. To each point from T_0 we let correspond a cube with side $\varepsilon + 2\delta$ and center at this point, orientated along the coordinate axes. The union of these cubes includes A_ε and, consequently,

$$V(A_\varepsilon) < N_0(x_1, \cdots, x_l, \delta)(\varepsilon + 2\delta)^l,$$

whence we obtain

$$\mathbf{E}\frac{1}{l} \log V(A_\varepsilon) \leqq \mathbf{E}\frac{H^\Lambda(\frac{1}{2}\delta, l)}{l} + \log(\varepsilon + 2\delta),$$

and, in view of the fundamental theorem,

$$\lim_{l\to\infty} \mathbf{E}\frac{1}{l} \log V(A_\varepsilon) \leqq \log(\varepsilon + 2\delta).$$

Since $V(A_\varepsilon) \geqq \varepsilon^l$ and δ are arbitrary, we obtain the sought statement.

Sufficiency is obtained from the following considerations. We assume that there is no uniform convergence. Then, for some $\varepsilon > 0$,

$$\lim_{l\to\infty} \mathbf{E} \log N_0(x_1, \cdots, x_l, 1.5\varepsilon) = \gamma > 0,$$

and, according to Lemma 2,

$$\lim_{l\to\infty} \frac{1}{l} M \log V(A_\varepsilon) \geqq \gamma + \log \varepsilon,$$

which was to be proved.

Corollary 2. *If uniform convergence holds in the class of functions* $F(x, \alpha)$, *then it also holds in the class* $|F(x, \alpha)|$.

PROOF. The transformation

$$F(x, \alpha) \to |F(x, \alpha)|$$

does not increase the distance

$$\rho(\alpha_1, \alpha_2) = \max_{i=1,\cdots,l} |F(x_i, \alpha_1) - F(x_i, \alpha_2)|.$$

Therefore,

$$N_0(x_1, \cdots, x_l, \varepsilon) \geqq N_0'(x_1, \cdots, x_l, \varepsilon),$$

where N_0 and N_0' are the minimum numbers of elements of the ε-net, respectively, in the sets A and A' generated by the classes $F(x, \alpha)$ and $|F(x, \alpha)|$.

Consequently, from the condition

$$\lim_{l\to\infty} \mathbf{P}\left\{\frac{\log N_0(x_1, \cdots, x_l, \varepsilon)}{l} > \delta\right\} = 0$$

it follows that

$$\lim_{l\to\infty} \mathbf{P}\left\{\frac{\log N_0'(x_1, \cdots, x_l, \varepsilon)}{l} > \delta\right\} = 0.$$

Along with the class of functions $F(x, \alpha)$ we consider the two-parameter class of functions

$$f(x, \alpha_1, \alpha_2) = |F(x, \alpha_1) - F(x, \alpha_2)|.$$

Corollary 3. *Uniform convergence in the class* $F(x, \alpha)$ *gives rise to uniform convergence in* $f(x, \alpha_1, \alpha_2)$.

PROOF. Clearly, uniform convergence in $F(x, \alpha)$ inplies such convergence in $F(x, \alpha_1) - F(x, \alpha_2)$.

Indeed, from the condition

$$\sup_\alpha \left| \mathbf{E}F(x, \alpha) - \frac{1}{l} \sum_{i=1}^{l} F(x_i, \alpha) \right| \leqq \varepsilon$$

and the condition

$$\left|\mathbf{E}F(x,\alpha_1)-\mathbf{E}F(x,\alpha_2)-\frac{1}{l}\sum_{i=1}^{l}F(x_i,\alpha_1)+\frac{1}{l}\sum_{i=1}^{l}F(x_i,\alpha_2)\right|$$
$$\leqq\left|\mathbf{E}F(x,\alpha_1)-\frac{1}{l}\sum_{i=1}^{l}F(x_i,\alpha_1)\right|+\left|\mathbf{E}F(x,\alpha_2)-\frac{1}{l}\sum_{i=1}^{l}F(x_i,\alpha_2)\right.$$

it follows that

$$\sup_{\alpha_1\alpha_2}\left|\mathbf{E}[f(x,\alpha_1)-F(x,\alpha_2)]+\frac{1}{l}\sum_{i+1}^{l}[F(x_i,\alpha_1)-F(x_i,\alpha_2)]\right|\leqq 2\varepsilon.$$

Applying Corollary 2, we obtain the required result.

We denote by $L(x_1,\cdots,x_l,\varepsilon)$ the minimum ε-net of the set $A(x_1,\cdots,x_l)$ in the metric

$$\hat{\rho}(z^1,z^2)=\frac{1}{l}\sum_{i=1}^{l}|z_i^1-z_i^2|.$$

Corollary 4. *For the uniform convergence of means to their expectations, it is necessary and sufficient that a function $T(\varepsilon)$ exists such that*

$$\lim_{l\to\infty}\mathbf{P}\{L(x_1,\cdots,x_l,\varepsilon)>T(\varepsilon)\}=0.$$

NECESSITY. From uniform convergence of $F(x,\alpha)$ follows uniform convergence for the functions

$$f(x,\alpha_1,\alpha_2)=|F(x,\alpha_1)-F(x,\alpha_2)|,$$

that is,

$$\sup_{\alpha_1,\alpha_2}\left|\frac{1}{l}\sum_{i=1}^{l}|F(x_i,\alpha_1)-F(x_i,\alpha_2)|-\mathbf{E}|F(x,\alpha_1)-F(x,\alpha_2)|\right|\xrightarrow{P}0\quad\text{as } l\to\infty. \tag{38}$$

Consequently, for a finite l_0 and a given ε there exists a sequence $x_1^*,\cdots,x_{l_0}^*$ such that the left side of (38) will be less than ε.

This means that the distance

$$\hat{\rho}(\alpha_1,\alpha_2)=\frac{1}{l_0}\sum_{i=1}^{l_0}|F(x_i^*,\alpha_1)-F(x_i^*,\alpha_2)| \tag{39}$$

approximates uniformly with respect to α_1 and α_2 and with an accuracy ε, the distance

$$\rho_0(\alpha_1,\alpha_2)=\int|F(x,\alpha_1)-F(x,\alpha_2)|\,dP(x) \tag{40}$$

in the space of functions Λ. But in the metric (39) on the set Λ there exists a finite ε-net S with number of elements $L(x_1^*,\cdots,x_{l_0}^*,\varepsilon)$

The same net forms a 2ε-net in the space Λ with metric (40). Thus the space Λ in the metric (40) is quasi-compact.

We use next the uniform convergence of $\hat{\rho}(\alpha_1,\alpha_2)$ to $\rho_0(\alpha_1,\alpha_2)$ and find that the same net S, with probability tending to one as $l\to\infty$, forms a 3ε-net on the set $A(x_1^*, \cdots, x_l^*)$. Having put $T(\varepsilon) = L(x_1^*, \cdots, x_{l_0}^*, \varepsilon/3)$, we obtain the statement of the theorem.

The proof of sufficiency is analogous to the proof of sufficiency for the fundamental theorem. In passing, we have proved the following fact.

Corollary 5. *For the uniform convergence of means to their expectations, it is necessary for the space of functions in the metric* (40) *to be quasi-compact.*

Corollary 6. *For the uniform convergence of means to their expectations, it is necessary and sufficient for the quantity*

$$\int_0^1 |P\{F(x, \alpha) > c\} - \nu^l\{F(x, \alpha) > c\}| dx$$

to converge to zero uniformly with respect to α, *or for*

$$\left| \int_c^1 \mathbf{P}\{F(x, \alpha) > q\} \, dq - \int_c^1 \nu^l\{F(x, \alpha) > q\} \, dq \right|$$

to converge to zero uniformly with respect to α *and* c, *where* $\nu^l\{F(x, \alpha) > q\}$ *is the frequency of the event* $\{F(x, \alpha) > q\}$ *found from a sample of length* l.

In other words, for the uniform convergence of means to the expectations it is necessary and sufficient that the empirical distribution functions converge in the mean to the true values, uniformly with respect to α. (We point out that the sufficient conditions of [2] consist of the convergence of empirical distribution functions to the true values, uniformly with respect to α and c.)

Received by the editors
July 28, 1978

REFERENCES

[1] V. N. Vapnik and A. Ya. Chervonenkis, *On the uniform convergence of relative frequencies of events to their probabilities*, Theory Prob. Appl., 16, (1971), pp. 264–280.

[2] V. N. Vapnik and A. Ya. Chervonenkis, *On the method of ordered risk minimization*, I, II, Avtom. i Telemekh., 8 (1974), pp. 21–30; 9 (1974), pp. 29–39. (In Russian.)

[3] A. Ya. Khinchin, *On the fundamental theorems of information theory*, Uspekhi Matem. Nauk, 11, 1 (1956), pp. 17–75. (In Russian.)