Z. Wahrscheinlichkeitstheorie verw. Geb. 14, 239–250 (1970)

On the Glivenko-Cantelli Theorem

FLEMMING TOPSØE

Summary. Various generalizations of the classical Glivenko-Cantelli theorem are proved. In particular, we have strived for as general results as possible for theoretical distributions on euclidean spaces, which are absolutely continuous with respect to Lebesgue measure.

Let $\{X_n\}$ be a sequence of independent identically distributed real random variables with distribution P and let $\{P_{n\omega}\}$ denote the empirical distributions. The classical Glivenko-Cantelli theorem asserts that

$$\lim_{n\to\infty} \sup_{x\in\dot{R}} |P_{n\omega}((-\infty, x]) - P((-\infty, x])| = 0 \qquad \text{a.e.}$$

There are several natural suggestions of generalizations:

1) Replace the range space $\dot{R}$ by a more general space.

2) Replace the class $\{(-\infty, x]\}_{x\in\dot{R}}$ by other classes of measurable sets or, more generally, measurable functions.

3) Try to study other random sequences $\{P_{n\omega}\}$ obeying the SLLN.

4) Investigate cases where one does not get 0 as the almost sure limit.

Another suggestion, perhaps more interesting than the previous ones, is to investigate the rate of convergence. However, with the methods of the present paper, nothing can be said about this question.

Denote by S a separable metrizable space, by $\mathscr{B}(S)$ the Borel σ-field on S and by $\mathscr{B}(S, \dot{R})$ the set of bounded real measurable functions on S. Given is a probability measure P on S and a subfamily $\mathscr{F}$ of $\mathscr{B}(S, \dot{R})$. We choose a canonical version of the problems and take as basic probability space $(\Omega, ., .)$ the completion of the countable product space $(S, \mathscr{B}(S), P)^{\dot{N}}$ and as random elements the coordinate mappings $\{X_n\}_{n\in\dot{N}}$ $(X_n(\omega) = x_n$ where $\omega = (x_1, x_2, \ldots))$. Define functions G_n of Ω into the extended real line by

$$G_n(\omega) = \sup_{f\in\mathscr{F}} |\textstyle\int f\, dP_{n\omega} - \int f\, dP|$$
$$= \sup_{f\in\mathscr{F}} |1/n(f(X_1\,\omega) + \cdots + f(X_n\,\omega)) - \textstyle\int f\, dP|.$$

Of course, these functions depend on $\mathscr{F}$ and P too. By the *upper Glivenko-Cantelli function* $\bar{G}(.)$ we shall understand the function

$$\bar{G}(\omega) = \limsup_{n\to\infty} G_n(\omega).$$

The *lower Glivenko-Cantelli function* is the function

$$\underline{G}(\omega)=\liminf_{n\to\infty} G_n(\omega).$$

We shall assume that the upper and lower Glivenko-Cantelli functions are measurable. This is a very weak assumption – indeed, I do not know whether it is always fulfilled. In all "naturally" occuring cases $\bar{G}(.)$ and $\underline{G}(.)$ will be measurable simply because the functions $G_n(.)$ are so, but even in cases where none of the $G_n(.)$ functions are measurable can it happen that $\bar{G}(.)$ and $\underline{G}(.)$ are measurable. (To see this, consider a non-atomic P, any subset A of S and the class of functions $\{f_a\}_{a\in A}$ where $f_a(x)=1$ for $x=a$, 0 for $x\neq a$; it will follow from results in this paper that $\bar{G}(.)$ and $\underline{G}(.)$ both are measurable and equal to 0 a.e.)

The functions $\bar{G}(.)$ and $\underline{G}(.)$ are symmetric and by assumption they are measurable. Applying a well known result of Hewitt and Savage and Kolmogorovs zero-one law (see [8]), it follows that $\bar{G}(.)$ and $\underline{G}(.)$ are constants a.e. It is natural to define the *upper* and *lower Glivenko-Cantelli constants* $\bar{G}$ and $\underline{G}$ by

$$\bar{G}(\omega)=\bar{G} \quad \text{a.e.},$$
$$\underline{G}(\omega)=\underline{G} \quad \text{a.e.}$$

If $\bar{G}=\underline{G}$, we denote the common value by G and call it the *Glivenko-Cantelli constant.*

The constants introduced depend on the probability-measure P and the class of functions $\mathscr{F}$; whenever convenient, we shall use notation such as $G(P,\mathscr{F})$.

How can we calculate the Glivenko-Cantelli constants? Are the upper and the lower Glivenko-Cantelli constants always equal? When is the Glivenko-Cantelli constant zero?

None of these questions we can answer completely. What I shall do below is to derive some sufficient conditions ensuring that G is zero. I believe that these conditions are close to beeing necessary as well, in fact I have been unable to find an example with $G=0$, which can not be handled by Lemma 2 and Theorem 1 below.

The method which is essentially due to Ranga Rao ([9]) consists in considering an associated problem on weak convergence. First, we note that for any $f\in\mathscr{B}(S,\grave{R})$ we have (by SLLN)

$$\int f\,dP_{n\omega}\to\int f\,dP \quad \text{a.e. } \omega.$$

By a result of Varadarajan (see [13]) one can find countably many functions in $\mathscr{B}(S,\grave{R})$ such that convergence for these implies weak convergence. Therefore, in our situation we can assert that:

For almost all ω, $P_{n\omega}$ *converges weakly* to P as $n\to\infty$.

The associated problem on weak convergence we in this way are led to consider is as follows: Given P, a probability-measure (or just a finite measure) and $\mathscr{F}$ a subclass of $\mathscr{B}(S,\grave{R})$, define a number $\alpha=\alpha(P,\mathscr{F})=\alpha(P,\mathscr{F};S)$ by

$$\alpha=\sup_{P_n\to P}\limsup_{n\to\infty}\sup_{f\in\mathscr{F}}\left|\int f\,dP_n-\int f\,dP\right|,$$

where the first supremum is taken over all sequences of probability-measures (or, just as well, finite measures) converging weakly to P; the problem is to calculate α. From the above remarks we obtain

Lemma 1. *Given $P, \mathscr{F}$. The inequality $\bar{G}(P, \mathscr{F}) \leqq \alpha(P, \mathscr{F})$ always holds. In particular, $\alpha(P, \mathscr{F})=0$ implies $G(P, \mathscr{F})=0$.*

The condition $\alpha=0$ is not necessary for G to be zero (it is easy to construct an example with P a point mass). However, we only have to split the space S in measurable parts and investigate what happens on the individual parts to arrive at a criterion which is much closer to being necessary.

Lemma 2. *Given $P, \mathscr{F}$ and assume that $\mathscr{F}$ is uniformly bounded. Let $S=\bigcup_i S_i$ be a decomposition of S in finitely or countably many Borel sets, denote by P_i the restriction of P to S_i and by $\mathscr{F}_i$ the subclass of $\mathscr{B}(S_i, \bar{R})$ obtained by restricting the functions in $\mathscr{F}$ to S_i. Then*

$$\bar{G}(P, \mathscr{F}) \leqq \sum_i \alpha(P_i, \mathscr{F}_i; S_i);$$

therefore, $\bar{G}$ is smaler than or equal to the infimum of all sums $\sum_i \alpha(\ldots)$ obtainable by decomposing S as described.

In particular, if there exists a decomposition with $\alpha(P_i, \mathscr{F}_i; S_i)=0$ for all i then $G(P, \mathscr{F})=0$.

The first step of the proof consists in removing a null set such that, for each i, the restriction of $P_{n\omega}$ to S_i converges weakly to P_i.

Consider a fixed ω among the remaining ones and a positive ε. Put $K=\sup\{\|f\|: f\in\mathscr{F}\}$. By assumption $K<\infty$. Assume that $\{S_i\}$ is infinite and indexed by the positive integers. Choose i_0 so large that $P(\bigcup_{i>i_0} S_i)<\varepsilon/K$. Since $P_{n\omega} S_i \to PS_i$ for each i, $P_{n\omega}(\bigcup_{i>i_0} S_i) \to P(\bigcup_{i>i_0} S_i)$.

Now we have

$$\begin{aligned}
&\limsup_{n\to\infty} \sup_{f\in\mathscr{F}} \left|\int f\,dP_{n\omega} - \int f\,dP\right| \\
&\leqq \limsup_{n\to\infty} \sum_i \sup_{f\in\mathscr{F}} \left|\int_{S_i} f\,dP_{n\omega} - \int_{S_i} f\,dP\right| \\
&\leqq \limsup_{n\to\infty} \sum_{i\leqq i_0} \sup_{f\in\mathscr{F}} \left|\int_{S_i} f\,dP_{n\omega} - \int_{S_i} f\,dP\right| + 2\varepsilon \\
&\leqq \sum_{i\leqq i_0} \limsup_{n\to\infty} \sup_{f\in\mathscr{F}} \left|\int_{S_i} f\,dP_{n\omega} - \int_{S_i} f\,dP\right| + 2\varepsilon \\
&\leqq \sum_i \alpha(P_i, \mathscr{F}_i; S_i) + 2\varepsilon
\end{aligned}$$

and we easily derive the desired inequality.

Remark. Note that the proof only depends on the independance via the validity of the SLLN. Accordingly, Lemma 2 (also Lemma 1 of course) remains valid if $\{P_{n\omega}\}$ is any sequence of random measures on S obeying the SLLN with respect to the measure P and defined on some complete probability space. This remark might turn out to be important when one is searching for criteria which are necessary.

Lemma 2 implies an useful extension of Lemma 1. Let $P=P_{at}+P_c$ be the decomposition of P in its atomic and non-atomic part. *If $\mathscr{F}$ is uniformly bounded and if $\alpha(P_c,\mathscr{F})=0$ then $G(P,\mathscr{F})=0$.* To see this, let $\{x_1,x_2,\dots\}$ be the set of atoms and consider the decomposition $(S_0,S_1,S_2,\dots)$ of S where $S_i=\{x_i\}$ for $i\geqq 1$ and $S_0=S\setminus\bigcup_{i\geqq 1}S_i$. Then $\alpha(P_i,\mathscr{F}_i;S_i)=0$ for $i\geqq 1$ and $\alpha(P_0,\mathscr{F}_0;S_0)\leqq\alpha(P_c,\mathscr{F})=0$. Hence, $G(P,\mathscr{F})=0$.

The simple lemmas we have obtained so far are useless if they are not supplied by information telling us how to calculate the α-quantities. One can in fact give a direct formula for $\alpha=\alpha(P,\mathscr{F})$, at least for a class of indicator functions, but we shall limit ourselves to give necessary and sufficient conditions for $\alpha(P,\mathscr{F})$ to be zero. Classes $\mathscr{F}$ with this property have been studied in [2] and [11] where they were called *P-uniformity classes.* I have decided to limit the discussion to the case where $\mathscr{F}$ is a class of indicator functions. It is left to the reader to deduce from [2] and [11] those "Glivenko-Cantelli theorems" on classes of functions he might find interesting. We borrow from [11] the following result:

Theorem 1. *Let $\mathfrak{A}$ be a subclass of $\mathscr{B}(S)$ and P a probability measure or just a finite measure. Then $\mathfrak{A}$ is a P-uniformity class (i.e. $\alpha(P,\mathfrak{A})=0$) if and only if*

$$P\left(\bigcap_{n=1}^{\infty}\partial_{\delta_n}A_n\right)=0$$

for every sequence $\{\delta_n\}\downarrow 0$ and every sequence $\{A_n\}$ of sets from $\mathfrak{A}$.

Here $\partial_\delta A$, the *δ-boundary* of A, is the set of those points within distance δ from A as well as from A^c (the complement of A). The ordinary boundary is $\partial A=\bigcap_{\delta>0}\partial_\delta A$.

I claim that all Glivenko-Cantelli theorems of the type considered which have been published so far are corollaries to Lemma 2 and Theorem 1, and in fact rather simpel corollaries. Perhaps I can provoke someone either to show me that my claim is unjustified or else to prove that the method leeds to essentially necessary and sufficient conditions.

Let us recall some steps in the development of Glivenko-Cantelli results. The classical result is due essentially to Glivenko and was proved in 1933. In 1953 Fortet and Mourier established a result for halfspaces in $\dot{R}^k$, assuming that P is absolutely continuous with respect to λ (Lebesgue measure). This assumption was removed in papers by Wolfowitz, 1954 and 1960. In 1955 Blum obtained a result for a class of sets in $\dot{R}^k$ much larger than the class of half-spaces and any P with $P\ll\lambda$. Tucker considered stationary sequences instead of independent ones (1959). Independantly of each other, Ranga Rao and Ahmad obtained a result for the class of convex sets and $P\ll\lambda$ (1962, 1961). In 1963 Sazonov published the only non-trivial result, I know of, giving an example where $G(P,\mathfrak{A})>0$; in that example $\mathfrak{A}$ is the class of halfspaces in an infinite-dimensional space.

Let us illustrate how our method works by proving one of the above mentioned results.

Theorem 2. *If S is an Euclidean space $\grave{R}^n$ and $\mathfrak{A}$ the class of all closed halfspaces then $G(P, \mathfrak{A})=0$ for all P.*

First assume that S is a measurable subset of $\grave{R}^n$ and that P is a probability measure, or just a finite measure on S such that $P(S \cap H)=0$ for every $((n-1)$-dimensional$)$ hyperplane H in $\grave{R}^n$. Let us prove that $S \cap \mathfrak{A}=\{S \cap A$: A closed halfspace in $\grave{R}^n\}$ is a P-uniformity class. Consider a set $D=\bigcap_{1}^{\infty} \partial_{\delta_k}(S \cap A_k)$ of the form appearing in Theorem 1. The δ-boundaries $\partial_{\delta_k}(S \cap A_k)$ are to be calculated in the space S. We claim that D is a subset of a set of the form $S \cap H$ with H a hyperplane in $\grave{R}^n$. If this were not so, we would be able to find $n+1$ points $x_1, \ldots, x_{n+1}$ in D such that no hyperplane passes through all of them. Choose k so large that no hyperplane intersects all the spheres $S(x_1, \delta_k), \ldots, S(x_{n+1}, \delta_k)$. Now we arrive at a contradiction since each of these spheres contains a point from the hyperplane ∂A_k.

We can now complete the proof. Let P be any probability measure in $\grave{R}^n$ and denote by $\mathfrak{A}$ the class of closed halfspaces. It is easy to see that one can construct a countable decomposition of $\grave{R}^n$ in measurable sets:

$$\grave{R}^n = \bigcup_{\nu=0}^{n} \Big(\bigcup_{i} S_{\nu,i}\Big)$$

where each $S_{\nu,i}$ is a subset of a ν-dimensional affine subspace in $\grave{R}^n$ (a ν-flat) such that every $(\nu-1)$-flat in $\grave{R}^n$ intersects $S_{\nu,i}$ in a set of measure 0. (The union $\bigcup_i S_{\nu,i}$ for $\nu=n$ can be chosen to consist of one set.) Each one of the sets $S_{\nu,i}$ can in an obvious way be considered as a subset of $\grave{R}^\nu$. When doing so we find that the class $S_{\nu,i} \cap \mathfrak{A}$ consists of the set $S_{\nu,i}$ itself and of all sets of the form $S_{\nu,i} \cap A$ with A a closed halfspace in $\grave{R}^\nu$. By the result in the first half of the proof,

$$\alpha(P_{\nu,i}, S_{\nu,i} \cap \mathfrak{A}; S_{\nu,i})=0,$$

where $P_{\nu,i}$ is the restriction of P to $S_{\nu,i}$. By Lemma 2, $G(P, \mathfrak{A})=0$ follows.

Glivenko-Cantelli Results for Classes of "Non Diffuse" Sets

The simplest classes to be studied in this section are those consisting of convex sets. If we insist on only proving theorems of a very concrete nature, where the assumptions are easy to check, I find that one must limit oneself to special spaces. I have chosen to give as general concrete results, I have been able to for euclidean spaces. Before starting this development, let me state the best result, I have been able to find, valid in infinite-dimensional spaces: *Let S be a separable Banach space, real or complex, and let $\mathfrak{A}$ be a class of closed convex sets such that the class $K \cap \mathfrak{A}=\{K \cap A$: $A \in \mathfrak{A}\}$ is closed in the notion of closed topological convergence for each compact and convex set K; if P (or just the non-atomic part of P) is $\mathfrak{A}$-continuous (i.e. if $P(\partial A)=0$ for all A in $\mathfrak{A}$) then $G(P, \mathfrak{A})=0$.* The proof is based on ideas in [2] and Lemma 2; I shall leave the details to the reader.

The Glivenko-Cantelli results we shall derive below will be based on the simple Lemma 1, thus we shall concentrate on finding P-uniformity classes. We noted in [11] that if the class $\mathfrak{A}$, intuitively speaking, is "closed" then $\mathfrak{A}$ is a P-uniformity class for every $\mathfrak{A}$-continuous P. Such classes are called *ideal uniformity classes.* (Notice that P must be $\mathfrak{A}$-continuous if $\mathfrak{A}$ is a P-uniformity class.) For an ideal uniformity class $\mathfrak{A}$ we have the Glivenko-Cantelli result that $G(P, \mathfrak{A})=0$ for every $\mathfrak{A}$-continuous P. In an euclidean space we shall agree only to find such a result interesting if we among the $\mathfrak{A}$-continuous measures find every measure absolutely continuous with respect to Lebesgue measure λ. What we shall search for then is ideal uniformity classes $\mathfrak{A}$, such that λ is $\mathfrak{A}$-continuous.

As in [11] we shall work with the notion of closed topological convergence. If $\{F_n\}$ is a sequence of closed subsets of S, we define lim sup F_n [lim inf F_n] as the set of those $x \in S$ for which every neighbourhood of x intersects F_n for infinitely many n [for all n sufficiently large]. By definition $\{F_n\}$ converges to F in the notion of closed topological convergence if $\liminf F_n = \limsup F_n = F$; we shall write $F_n \to F$ in this case. The most important general result on this notion of convergence is Hausdorffs selection theorem stating that any sequence of closed sets has a convergent subsequence; this is true in any second countable space ([7] p. 147).

We shall now define the classes of sets, we will investigate. To define these, we need a specific metric, call it d, on our separable metrizable space S. For a subset A of S and a positive δ we denote by $S(A, \delta)$ the δ-neighbourhood of A i.e. $S(A, \delta) = \{x: d(x, A) < \delta\}$; in particular, $S(x, \delta)$ denotes the open sphere

$$S(x, \delta) = \{y: d(y, x) < \delta\};$$

the closed sphere will be denoted by $S[x, \delta]$ i.e. $S[x, \delta] = \{y: d(y, x) \leqq \delta\}$. Although it is not absolutely necessary, we shall assume that all spheres $S(x, \delta)$ are connected; the convenience of this assumption lies in the fact that it allows us to use the formula $\partial_\delta A = S(\partial A, \delta)$.

Now, let η be a non-decreasing function defined for all positive real numbers, satisfying the inequalities $0 < \eta(\delta) \leqq \delta$; $\delta > 0$. By $\mathfrak{A}_\eta$ we denote the class of those Borel-sets A satisfying the condition:

$$\underset{x \in \partial A}{\forall} \ \underset{\delta > 0}{\forall} \ S[x, \delta] \setminus \partial_{\eta(\delta)}(A) \neq \emptyset. \tag{1}$$

Here are some useful equivalent forms of the defining property (1):

$$\underset{x \in \partial A}{\forall} \ \underset{\delta > 0}{\forall} \ S[x, \delta] \setminus S(\partial A, \eta(\delta)) \neq \emptyset, \tag{2}$$

$$\underset{x \in \partial A}{\forall} \ \underset{\delta > 0}{\forall} \ \underset{y \in S[x, \delta]}{\exists} \ S(y, \eta(\delta)) \cap \partial A = \emptyset, \tag{3}$$

$$\underset{x \in \partial A}{\forall} \ \underset{\delta > 0}{\forall} \ \underset{y \in S[x, \delta]}{\exists} \ S(y, \eta(\delta)) \subset A \vee S(y, \eta(\delta)) \cap A = \emptyset. \tag{4}$$

Notice that in (4) we allow a different alternative from point to point. If we had only allowed the second alternative in (4) then the class we obtained would be, loosely speaking, a class of sets without small holes. Perhaps one could say that the classes $\mathfrak{A}_\eta$ are classes consisting of not too "diffuse" sets.

Let us collect some immediate consequences of the definition:

Lemma 3. (i) *If* $\eta_1 \leqq \eta_2$ *then* $\mathfrak{A}_{\eta_1} \supset \mathfrak{A}_{\eta_2}$.

(ii) *If* $A \in \mathfrak{A}_\eta$ *then* ∂A *has empty interior.*

(iii) *If* $A \in \mathfrak{A}_\eta$, $B \in \mathscr{B}(S)$ *and if* $\partial B \subset \partial A$ *then* $B \in \mathfrak{A}_\eta$.

In particular, it follows that $\bar{A}$ *(the closure of* A*) and* $\mathring{A}$ *(the interior of* A*) are in* $\mathfrak{A}_\eta$ *if* A *is so.*

Now we shall prove a more substantial result by specializing the structure of S.

Lemma 4. *Let* (S, d) *be a locally compact, locally connected separable metric space such that all spheres are relatively compact and connected. Consider one of the classes* $\mathfrak{A}_\eta$ *and assume that* $\{A_n\}$ *is a sequence of sets in* $\mathfrak{A}_\eta$ *such that the sequence of closures* $\{\bar{A}_n\}$ *converges, say* $\bar{A}_n \to F$, *and such that the sequence of boundaries* $\{\partial A_n\}$ *converges, say* $\partial A_n \to D$. *Let* A *denote the set* $A = F \setminus (D \cap \mathring{F})$. *Then* $\partial A = D$ *and* $A \in \mathfrak{A}_\eta$.

Proof. Put $D \cap \mathring{F} = \Delta$. We find that $\partial A \subset \partial F \cup \partial \Delta \subset \partial F \cup D$. To prove $\partial A \subset D$ it is therefore enough to prove $\partial F \subset D$; this inclusion has nothing to do with the special character of our sets. To prove it, assume that $x \in \partial F$ and consider a sphere $S(x, \delta)$.

Since $x \in F$ and $\bar{A}_n \to F$, the sets $S(x, \delta) \cap \bar{A}_n$ are non-empty for all n sufficiently large, and the same then holds for the sets $S(x, \delta) \cap A_n$. $S(x, \delta)$ contains a point not in F and it is seen that the sets $S(x, \delta) \setminus A_n$ are non-empty for all n sufficiently large. We now use that $S(x, \delta)$ is connected and conclude that $S(x, \delta) \cap \partial A_n \neq \emptyset$ for all n sufficiently large. δ was arbitrary, so it follows that $x \in \lim \partial A_n = D$.

To prove $D \subset \partial A$ we treat the two inclusions $D \subset \bar{A}$ and $D \subset \overline{A^c}$ separately. The latter is easy since, if $x \in D$ and the sphere $S(x, \delta)$ intersects F^c then it also intersects A^c and if the sphere does not intersect F^c then $x \in \Delta \subset A^c$. The proof of the remaining inclusion $D \subset \bar{A}$ and of $A \in \mathfrak{A}_\eta$ is carried out simultaneously.

Let $x \in D$ and $\delta > 0$. There exists a sequence $\{x_{n_k}\}$ such that $n_1 < n_2 < \cdots$, such that $x_{n_k} \to x$ and such that $x_{n_k} \in \partial A_{n_k}$ for all k. Choose the sequence $\{y_{n_k}\}$ such that, for all k, $y_{n_k} \in S[x_{n_k}, \delta]$ and $S(y_{n_k}, \eta(\delta)) \cap \partial A_{n_k} = \emptyset$. By the compactness assumption, $\{y_{n_k}\}$ has a convergent subsequence. Assume for simplicity that $\{y_{n_k}\}$ itself converges, say $y_{n_k} \to y$. Clearly, $y \in S[x, \delta]$. It is also easy to see that for any $z \in S(y, \eta(\delta))$ there exists an $\varepsilon > 0$ such that the sets $S(z, \varepsilon) \cap \partial A_n$ are empty for infinitely many n. We see from this that $S(y, \eta(\delta)) \cap D = \emptyset$. What we have proved is the following:

$$\underset{x \in D}{\forall} \ \underset{\delta > 0}{\forall} \ \underset{y \in S[x, \delta]}{\exists} \ S(y, \eta(\delta)) \cap D = \emptyset. \tag{5}$$

Since A is measurable, it follows from (5) and the inclusion $\partial A \subset D$ that $A \in \mathfrak{A}_\eta$.

The last thing to prove is the inclusion $D \subset \bar{A}$. Let $x \in D$. In case $x \notin \Delta$ we have $x \in A$. In the other case $x \in \Delta$, we first choose $\delta_0 > 0$ such that $S(x, \delta_0) \subset F$. Then, if $0 < \delta < \delta_0$ and if we consider the point y appearing in (5), we see that $y \notin D$ and it follows that $y \in A$. The proof of $D \subset \bar{A}$ is complete. $\lrcorner$

Theorem 3. *Let again (S, d) be a separable metric space with relatively compact and connected spheres. Then all the classes $\mathfrak{A}_\eta$ are ideal uniformity classes and, furthermore, we have the Glivenko-Cantelli result that $G(P, \mathfrak{A}_\eta)=0$ if and only if the non-atomic part of P is $\mathfrak{A}_\eta$-continuous.*

Proof. For a sequence $\{A_n\}$ of sets in $\mathfrak{A}_\eta$ and a sequence $\{\delta_n\}\downarrow 0$, select a subsequence $\{A_{n_k}\}$ such that both sequences $\{\bar{A}_{n_k}\}$ and $\{\partial A_{n_k}\}$ converge, say $\bar{A}_{n_k}\to F$ and $\partial A_{n_k}\to D$. Put $A=F\setminus(D\cap \mathring{F})$. We find that

$$\bigcap_{n=1}^{\infty} \partial_{\delta_n} A_n \subset \bigcap_{k=1}^{\infty} \partial_{\delta_{n_k}} A_{n_k} = \bigcap_{k=1}^{\infty} S(\partial A_{n_k}, \delta_{n_k}) \subset D.$$

By Lemma 4, $D=\partial A$ and $A\in\mathfrak{A}_\eta$. By Theorem 1, $\mathfrak{A}_\eta$ is an ideal uniformity class.

Having seen that $\mathfrak{A}_\eta$ is an ideal uniformity class, we easily derive the more important part, viz. the "if" part of the Glivenko-Cantelli result in the theorem.

Lastly, assume that the non-atomic part of P is not $\mathfrak{A}_\eta$-continuous. Then for some set $A\in\mathfrak{A}_\eta$ we have $P(\partial A\setminus S_0)>0$, where S_0 denotes the set of atoms of P. By SLLN we can find an ω-set Ω_0 of probability 1 such that $P_{n\omega}(\partial A\setminus S_0)\to P(\partial A\setminus S_0)$ for every $\omega\in\Omega_0$. For each $n\geq 1$ and $\omega\in\Omega_0$ consider the set

$$A_{n\omega}=\{X_1(\omega), X_2(\omega), \ldots, X_n(\omega)\}\cap(\partial A\setminus S_0).$$

By property (iii) of Lemma 3 we have that $A_{n\omega}\in\mathfrak{A}_\eta$. Since $P_{n\omega}(A_{n\omega})=P_{n\omega}(\partial A\setminus S_0)$ and $P(A_{n\omega})=0$ we now see that $\underline{G}(P, \mathfrak{A}_\eta)\geq P(\partial A\setminus S_0)>0$. $\lrcorner$

Remark. The result in Theorem 3 can be extended as follows: consider any one of the classes $\mathfrak{A}_\eta$ and any P. Then $\underline{G}(P, \mathfrak{A}_\eta)=\bar{G}(P, \mathfrak{A}_\eta)$ and the Glivenko-Cantelli constant is given by

$$G(P, \mathfrak{A}_\eta)=\sup\{P_c(\partial A)\colon A\in\mathfrak{A}_\eta\},$$

where P_c denotes the non-atomic part of P. This is proved by the same method as above but one has to use a generalized version of Theorem 1.

We shall now examine the situation in an euclidean space.

Lemma 5. *Let (S, d) be the euclidean space R^N with the usual euclidean metric.*

(i) *If η is the identity: $\eta(\delta)=\delta$; $\delta>0$, then $\mathfrak{A}_\eta$ consists of those Borel sets A for which there exists a convex set C such that $\partial A\subset\partial C$.*

(ii) *If $\limsup_{\delta\to 0}\eta(\delta)/\delta>0$ then Lebesgue measure λ is $\mathfrak{A}_\eta$-continuous.*

Proof. (i) Let C be convex. If $x\in\partial C$ and $\delta>0$ then there exists a point in $S[x,\delta]\setminus S(C,\delta)$. It follows that the Borel set A lies in $\mathfrak{A}_\eta$ if $\partial A\subset\partial C$.

Assume now that $A\in\mathfrak{A}_\eta$. Consider a point $x\in\partial A$. Choose a sequence of points $\{y_n\}$ and a sequence $\{\delta_n\}$ such that $y_n\in S[x,\delta_n]\setminus S(\partial A,\delta_n)$, such that $\delta_n\to\infty$, and such that the unit vectors $d(x,y_n)^{-1}(y_n-x)$ converge, say to e. Denote by H_x the closed halfspace $H_x=\{y\colon \text{the inner product } \langle y-x, e\rangle\leq 0\}$. Then $\partial A\subset H_x$. When doing this for every $x\in\partial A$, we can define the convex set $C=\bigcap\{H_x\colon x\in\partial A\}$. It is easy to see that $\partial A\subset\partial C$.

(ii) Let A be a set in $\mathfrak{A}_\eta$. By the density theorem of Lebesgue, $\lim_{\delta\to 0} \lambda(S(x,\delta))^{-1} \cdot \lambda(S(x,\delta)\cap\partial A)=\chi_{\partial A}$, the characteristic function of ∂A, a.e. λ. By assumption, this limit is <1 for every $x\in\partial A$ (assuming that the limit exists). Hence $\chi_{\partial A}=0$ a.e. λ i.e. $\lambda(\partial A)=0$. Below the reader will see how this result can also be proved by very elementary considerations. ⌟

Combining Theorem 3 and property (ii) above we obtain a Glivenko-Cantelli result of the desired strength:

Theorem 4. *For a class $\mathfrak{A}_\eta$ in an euclidean space with $\limsup_{\delta\to 0} \eta(\delta)/\delta>0$ we have $G(P,\mathfrak{A}_\eta)=0$ for every P with a non-atomic part absolutely continuous with respect to Lebesgue measure.*

A few examples are on their place here. Blum [3] considers in $\dot{R}^N$ the class $\mathfrak{A}_1$ of Borel sets A possessing the following property: If $(x_1,\dots,x_N)\in A$ and $(y_1,\dots,y_N)$ is such that $y_i<x_i$ for $i=1,\dots,N$ then $(y_1,\dots,y_N)\in A$. If $\mathfrak{A}_j, j=2,\dots,2^N$ denotes the classes of sets obtained by reversing, one at a time, the N inequalities occuring in the definition of $\mathfrak{A}_1$, then Blum proves that $\mathfrak{A}=\bigcup_1^{2^N}\mathfrak{A}_j$ satisfies $G(P,\mathfrak{A})=0$ for every $P\ll\lambda$. Clearly, this result is a corollary to Theorem 4 (take $\eta(\delta)=\delta/\sqrt{N}$).

Consider the class $\mathfrak{A}=\{S(B,\Delta)$: B any subset, $\Delta\geqq\Delta_0\}$ where Δ_0 is a fixed positive number. Then $\mathfrak{A}\subset\mathfrak{A}_\eta$ with $\eta(\delta)=\min(\delta,\Delta_0)$ so that, by Theorem 4, $G(P,\mathfrak{A})=0$ for every $P\ll\lambda$. Another class of sets which can be dealt with using Theorem 4 is the class of open sets contained in a fixed bounded domain of the euclidean space and being (arbitrary) unions of open convex sets each of which contains a sphere of some fixed radius.

The last problem we shall investigate is to which extent one can generalize (ii), Lemma 5.

Theorem 5. *Consider one of the classes $\mathfrak{A}_\eta$ in $\dot{R}^N$.*

(i) *If $\int_0^1 \delta^{-(N+1)}\eta(\delta)^N\,d\delta=\infty$ then $\lambda(\partial A)=0$ for all $A\in\mathfrak{A}_\eta$.*

(ii) *If $\int_0^1 \delta^{-(N+1)}\eta(\delta)^N\,d\delta<\infty$ then there exists a $\delta_0>0$ and a set $A\in\mathfrak{A}_{\eta'}$ where $\eta'=\eta\wedge\eta(\delta_0)$ such that $\lambda(\partial A)>0$.*

For $n\geqq 0$ denote by $\mathfrak{N}_n$ the decomposition of $\dot{R}^N$ in cubes each of volume (λ-measure) 2^{-Nn} which is effectuated in the "natural" way by N systems of hyperplanes orthogonal to the N coordinate axes with each system containing a hyperplane through $(0,0,\dots,0)$. All cubes mentioned below are supposed to have faces orthogonal to the coordinate axes. The proof of (i), Theorem 5 will be based on the following lemma.

Lemma 6. *Let $\nu_0,\nu_1,\dots$ be integers with $\nu_n\geqq n+1\ \forall n\geqq 0$ and $\sum_0^\infty 2^{-N(\nu_n-n)}=\infty$. If D is a subset of $\dot{R}^N$ satisfying the condition*

$$\forall n\geqq 0\,\forall Q_n\in\mathfrak{N}_n\,\exists Q_n^*\in\mathfrak{N}_{\nu_n}\colon\ (Q_n^*\subset Q_n\wedge Q_n^*\cap D=\emptyset)$$

then $\lambda(D)=0$.

Proof. We may assume that $D \subset Q_0$ for some $Q_0 \in \mathfrak{N}_0$. Put $\mathfrak{N}_n^0 = \{Q_n \in \mathfrak{N}_n : Q_n \subset Q_0\}$. Define a sequence $\{\Delta_n\}_{n \geq 0}$ of pairwise disjoint sets by

$$\Delta_0 = Q_0^*,$$
$$\Delta_n = \bigcup \{Q_n^* : Q_n \in \mathfrak{N}_n^0 \wedge Q_n \cap A_{n-1} = \emptyset\}; \qquad n \geq 1,$$

where we have put $A_n = \bigcup_0^n \Delta_i$. Let r_n denote the number of $Q_n \in \mathfrak{N}_n^0$ with $Q_n \cap A_{n-1} = \emptyset$; then Δ_n is a union of r_n disjoint cubes from $\mathfrak{N}_{\nu_n}^0$. The set A_n consists of $r_0 + r_1 + \cdots + r_n$ cubes; notice that a cube in $\mathfrak{N}_{n+1}^0$ contains at most one of these cubes and that every cube in $\mathfrak{N}_n^0$ intersects A_n.

Put $B_n = Q_0 \setminus A_n$. Then $\lambda(D) \leq \lambda(B_n) = \lambda(B_0) \prod_0^{n-1} \lambda(B_{i+1})/\lambda(B_i)$. We now obtain

$$\begin{aligned}\lambda(D) &\leq \lambda(B_0) \prod_0^\infty \lambda(B_{n+1})/\lambda(B_n)\\ &= \lambda(B_0) \prod_0^\infty (\lambda B_n - \lambda \Delta_{n+1})/\lambda B_n\\ &= \lambda(B_0) \prod_0^\infty (1 - r_{n+1}\, 2^{-N\nu_{n+1}}/\lambda B_n).\end{aligned}$$

Denote by M_n the number of $Q_n \in \mathfrak{N}_n^0$ such that Q_n is not contained in A_n. Clearly, $r_{n+1} = (2^N - 1)\, M_n$; employing this and the inequality $\lambda B_n \leq M_n\, 2^{-Nn}$ we find

$$\lambda(D) \leq \lambda(B_0) \prod_0^\infty \left(1 - (2^N - 1)\, 2^{-N} \cdot 2^{-N(\nu_{n+1} - (n+1))}\right)$$

and this infinite product is zero-divergent by assumption. ⌟

For the proof of (i) as well as for the proof of (ii), Theorem 5 we find it convenient to write η in the form

$$\eta(\delta) = \delta \cdot f(\log 1/\delta); \qquad \delta > 0,$$

i.e. we we define f by

$$f(x) = e^x\, \eta(e^{-x}); \qquad -\infty < x < \infty.$$

f satisfies $0 < f \leq 1$. The assumption $\int_0^1 \delta^{-(N+1)}\, \eta(\delta)^N\, d\delta = \infty$ is equivalent to $\int_0^\infty f(x)^N\, dx = \infty$. Due to the fact that η is nondecreasing, f does not oscillate very much; therefore, the condition $\int_0^\infty f(x)^N\, dx = \infty$ is equivalent to $\sum_0^\infty f(a + nb) = \infty$ where a is any real number and b any positive real number.

Proof of (i), Theorem 5. Let D be any subset of $\dot{R}^N$ satisfying the condition:

$$\forall x \in D\ \ \forall \delta > 0\ \ \exists y \in S[x, \delta]:\ S(y, \eta(\delta)) \cap D = \emptyset.$$

We shall prove that $\lambda(\mathrm{D}) = 0$ $\left(\text{assuming } \int_0^\infty f^N = \infty\right)$.

For $n \geqq 0$ denote by v_n the integer uniquely determined by the requirements

$$2^{-v_n} \leqq \eta(3 \cdot 2^{-(n+4)})/\sqrt{N} < 2^{-v_n+1}.$$

Elementary manipulations shows that $v_n \geqq n+3 \, \forall n \geqq 0$ and that

$$\sum_0^\infty 2^{-N(v_n-n)} \geqq (N^{-\frac{1}{2}} \cdot 3 \cdot 2^{-5})^N \sum_0^\infty f(\log \tfrac{16}{3} + n \log 2)^N,$$

from which it follows that $\sum_0^\infty 2^{-N(v_n-n)} = \infty$.

To see that we can apply Lemma 6, we consider an $n \geqq 0$ and a $Q_n \in \mathfrak{N}_n$. Denote by q_n the cube with same center as Q_n and with volume $\lambda(q_n) = 2^{-2N} \cdot \lambda(Q_n)$. If $q_n \cap D = \emptyset$, Q_n contains a cube from $\mathfrak{N}_{n+3}$ disjoint with D and since $v_n \geqq n+3$ we can then find $Q_n^* \in \mathfrak{N}_{v_n}$ with $Q_n^* \subset Q_n$ and $Q_n^* \cap D = \emptyset$. Now consider the other alternative $q_n \cap D \neq \emptyset$. Choose $x \in q_n \cap D$ and then choose $y \in S[x, 3 \cdot 2^{-(n+4)}]$ such that $S(y, \eta(3 \cdot 2^{-(n+4)})) \cap D = \emptyset$. Since $3 \cdot 2^{-(n+4)} + \eta(3 \cdot 2^{-(n+4)}) \leqq 3 \cdot 2^{-(n+3)}$, we see that $S(y, \eta(3 \cdot 2^{-(n+4)})) \subset Q_n$. Elementary considerations shows that $S(y, \eta(3 \cdot 2^{-(n+4)}))$ contains a cube $Q_n^* \in \mathfrak{N}_{v_n}$; clearly, $Q_n^* \subset Q_n$ and $Q_n^* \cap D = \emptyset$. We can now apply Lemma 6 and find that $\lambda(D) = 0$. $\lrcorner$

Proof of (ii), Theorem 5. Since $\int_0^\infty f(x)\,dx < \infty$ we can choose a positive integer n_0 such that

$$\sum_0^\infty \{2\sqrt{N} f(\log(2^{n+n_0}/\sqrt{N}))\}^N < 1.$$

Choose $Q_{n_0} \in \mathfrak{N}_{n_0}$ and put $\mathfrak{N}_n^0 = \{Q_n \in \mathfrak{N}_n \colon Q_n \subset Q_{n_0}\}$ for $n \geqq n_0$. Define real numbers b_n; $n \geqq n_0$ by

$$b_n = 2^{-n+1} \sqrt{N} f(\log(2^n/\sqrt{N})).$$

For $n \geqq n_0$ and $Q_n \in \mathfrak{N}_n^0$, let Q_n^* denote the cube with same center as Q_n and with $\lambda(Q_n^*) = b_n^N$ (note that $b_n < 2^{-n}$). Define sets B_n; $n \geqq n_0$ by $B_n = \bigcup \{Q_n^* \colon Q_n \in \mathfrak{N}_n^0\}$ and consider the set

$$D = Q_{n_0} \setminus \bigcup_{n \geqq n_0} B_n.$$

We have $\lambda(D) \geqq 2^{-Nn_0} \left(1 - \sum_{n_0}^\infty 2^{Nn} \cdot b_n^N\right)$ and from the definition of the b_n's and the choice of n_0 it follows that $\lambda(D) > 0$.

For $n \geqq n_0$ put $(\delta_n, \eta_n) = (\frac{1}{2}\sqrt{N} \cdot 2^{-n}, \frac{1}{2} b_n)$ and then define the function η^* by

$$\eta^*(\delta) = \eta_{n_0} \quad \text{for } \delta \geqq \delta_{n_0},$$
$$\eta^*(\delta) = \eta_n \quad \text{for } \delta_n \leqq \delta < \delta_{n-1}; \quad n \geqq n_0 + 1.$$

From the construction of D it follows that

$$\forall x \in D \; \forall \delta > 0 \; \exists y \in S[x, \delta] \colon S(y, \eta^*(\delta)) \cap D = \emptyset.$$

We now note that $\eta_n = \eta(\delta_{n-1})$; $n \geqq n_0 + 1$. It follows that $\eta(\delta) \leqq \eta^*(\delta)$ for $\delta \leqq \delta_{n_0}$. If we put $\eta' = \eta \wedge \eta(\delta_{n_0})$, we see that

$$\forall x \in D \; \forall \delta > 0 \; \exists y \in S[x, \delta] \colon S(y, \eta'(\delta)) \cap D = \emptyset.$$

We have already seen that $\lambda(D) > 0$. $\lrcorner$

Acknowledgements. One of the basic ideas underlying the proof of (i), Theorem 5 we found in the paper [5] by P.E. Hansen with whom we had helpful discussions. However, the final results in Theorem 5 are both due to L. Mejlbro. The proofs given are, except for minor modifications, those of Mejlbro's and are reproduced with his permission.

References

1. Ahmad, S.: Sur le théorème de Glivenko-Cantelli. C.r. Acad Sci., Paris **252**, 1413–1417 (1961).
2. Billingsley, P., Topsøe, F.: Uniformity in weak convergence. Z. Wahrscheinlichkeitstheorie verw. Geb. **7**, 1–16 (1967).
3. Blum, J.R.: On the convergence of empiric distribution functions. Ann. math. Statistics **26**, 527–529 (1955).
4. Fortet, R., Mourier, E.: Convergence de la répartition empirique vers la répartition théorique. Ann. sci. École norm. sup., III. Sér. **60**, 267–285 (1953).
5. Glivenko, V.: Sulla determinazione empirica della leggi di probabilità. Giorn. Ist. Ital. Attuari Anno **IV**, 92–99 (1933).
6. Hansen, P.E.: Konvekse figures Jordan-indhold. Nordisk mat. Tidskr. **9**, 26–28 (1961) [in danish].
7. Hausdorff, F.: Mengenlehre, 2. Aufl. Berlin-Leipzig: Gruyter 1927.
8. Meyer, P.A.: Probability and potentials. Waltham Massachusetts: Blaisdell 1966.
9. Rango Rao, R.: Relations between weak and uniform convergence of measures with applications. Ann. math. Statistics **33**, 659–680 (1962).
10. Sazonov, V.V.: On the Glivenko-Cantelli theorem. Theor. Probab. Appl. **8**, 282–285 (1963).
11. Topsøe, F.: On the connection between P-continuity and P-uniformity in weak convergence. Theor. Probab. Appl. **12**, 281–290 (1967).
12. Tucker, A.G.: A generalization of the Glivenko-Cantelli theorem. Ann. math. Statistics **30**, 828–830 (1959).
13. Varadarajan, V.S.: On the convergence of probability distributions. Sankyā **19**, 23–26 (1958).
14. Wolfowitz, J.: Generalization of the theorem of Glivenko-Cantelli. Ann. math. Statistics **25**, 131–138 (1954).
15. – Convergence of the empiric distribution function on half-spaces. Contributions to probability and statistics 504–507 (1960).

F. Topsøe
Københavns Universitets Matematisk Institut
H.C. Ørsted Institutet, Universitetsparken 5
2100 København Ø, Danmark

(Received February 12, 1969)