

WEAK CONVERGENCE OF PROBABILITIES ON NONSEPARABLE METRIC SPACES AND EMPIRICAL MEASURES ON EUCLIDEAN SPACES

BY
R. M. DUDLEY¹

1. Introduction

It is known that under certain mild set-theoretic assumptions, a finite, countably additive measure defined on all Borel sets of a metric space is concentrated in a separable subspace (Marczewski and Sikorski [8]). However, there are interesting probability measures on metric spaces not concentrated in separable subspaces. In this paper, we consider countably additive probability measures on the smallest σ -field containing the open balls of a metric space. This σ -field is the Borel field for a separable space, but is smaller in general. A probability measure on it need not be confined to a separable subspace.

A sequence of such measures will be said to converge weak* to a Borel measure if the upper and lower integrals of each bounded continuous real function converge. Some abstract results on this convergence, similar to those in Prokhorov [9] for separable metric spaces, will be given in §2.

The rest of the paper deals with "empirical measures" on Euclidean spaces, whose study motivated the abstract results and provides an application of them. Two of the main results of Donsker [3], [4] for measures on the real line will be generalized to arbitrary Euclidean spaces. At the same time, his results are corrected by replacing some integrals, which may not be defined, by upper and lower integrals.

I discovered after writing most of the rest of this paper that a generalization of Donsker's work to multidimensional spaces was proved several years ago by L. LeCam, who is now revising a paper embodying his results for the *Illinois Journal of Mathematics*. I shall try to explain what seem to be the main differences between our approaches.

While my abstract results in §2 are for metric spaces and guided by those in Prokhorov [9] for the separable case, LeCam uses a more elaborate abstract apparatus involving the second dual spaces of topological linear spaces and nonmetric topologies; the place of upper and lower integrals is taken by integrals with respect to finitely additive extensions of a measure.

With regard to the more concrete equicontinuity properties of empirical distribution functions, my approach in §4 below uses a sort of Markov property for the random empirical measure μ_n , namely given $\mu_n(E)$ for a set E , the

Received September 28, 1964.

¹ Partially supported by a National Science Foundation Grant and by the Office of Naval Research.

values of μ_n on subsets of E are independent of its values on sets disjoint from E . LeCam instead lets n be a random variable $n(\cdot)$ with a Poisson distribution and obtains a random measure with independent values on disjoint sets.

Acknowledgments. I am grateful both to Prof. LeCam for providing the above information and to V. Strassen for earlier conversations about the Skorokhod topology, etc.

2. Weak*-convergence in nonseparable metric spaces

Suppose β is a measure on a σ -field \mathcal{S}_1 in a space S and let F be any real-valued function on S . Then the *lower integral*

$$\int_* F(s) d\beta(s)$$

is defined as the supremum of all integrals

$$\int f(s) d\beta(s)$$

where $f \leq F$ on S , f is \mathcal{S}_1 -measurable, and the integral of f is defined. Similarly, the upper integral

$$\int^* F d\beta$$

is defined as the infimum of $\int f d\beta$ for $f \geq F$ on S and $\int f d\beta$ defined. Clearly

$$\int_* F d\beta \leq \int^* F d\beta$$

for any F . If χ_A is the indicator function of a set A , let

$$\beta_*(A) = \int_* \chi_A d\beta, \quad \beta^*(A) = \int^* \chi_A d\beta;$$

β_* and β^* are clearly the usual inner and outer measures for β .

A *Baire set* in a topological space is a member of the smallest σ -field with respect to which all continuous functions are measurable. In a metric space, the Baire sets are precisely the Borel sets. Now if β_n are measures on a topological space S (not necessarily defined on all Baire sets), and β is a Baire measure on S , we say

$$\beta_n \rightarrow \beta \quad (\text{weak}^*)$$

if for every bounded continuous real function F on S ,

$$\lim_{n \rightarrow \infty} \int^* F d\beta_n = \lim_{n \rightarrow \infty} \int_* F d\beta_n = \int F d\beta.$$

Let (S, ρ) be a metric space, and let \mathcal{S} be the σ -field generated by the balls

$$[y \in S : \rho(x, y) < \varepsilon]$$

for any $x \in S$ and $\varepsilon > 0$. Let \mathfrak{B} be the Borel σ -field generated by the open sets of S . In the cases of interest here, S will be non-separable and \mathfrak{B} strictly larger than \mathfrak{S} .

Let $(\mathfrak{C}, \| \cdot \|)$ be the Banach space of bounded real continuous functions on S with supremum norm. For any subset A of S and $\varepsilon > 0$ let

$$A^\varepsilon = [x \in S : \rho(x, y) < \varepsilon \text{ for some } y \in A].$$

We need the following well-known fact:

LEMMA 1. *If F is a continuous real-valued function on a metric space (S, ρ) , K is a compact subset of S , and $\varepsilon > 0$, then there is a $\delta > 0$ such that if $x \in K$, $y \in S$, and $\rho(x, y) < \delta$ then*

$$|F(x) - F(y)| < \varepsilon.$$

Proof. If the conclusion is false there are $x_n \in K$ and $y_n \in S$, $n = 1, 2, \dots$, with $\rho(x_n, y_n) < 1/n$ and

$$|F(x_n) - F(y_n)| \geq \varepsilon.$$

A subsequence of the x_n converges to an $x \in K$ at which F is not continuous, a contradiction which completes the proof.

We call a set \mathfrak{K} of measures on S weak*-precompact if any sequence $[\beta_n]$ of distinct elements of \mathfrak{K} has a subsequence which is weak*-convergent (to a Borel measure on S).

THEOREM 1. *If (S, ρ) is a metric space and \mathfrak{K} is a set of probability measures each defined at least on \mathfrak{S} in S , then \mathfrak{K} is weak*-precompact if for every $\varepsilon > 0$ there is a compact set $K \subset S$ such that for every $\delta > 0$,*

$$\beta(K^\delta) \geq 1 - \varepsilon$$

for all but finitely many $\beta \in \mathfrak{K}$.

Proof. Note that K^δ is a countable union of open balls and hence is in \mathfrak{S} . For each positive integer N , let K_N be a compact set in S such that for any $\delta > 0$,

$$\beta(K_N^\delta) \geq 1 - 1/N$$

for all but finitely many $\beta \in \mathfrak{K}$. Let $\{F_n\}_{n=1}^\infty$ be a countable set of continuous functions on S with $\|F_n\| \leq 1$ for all n , uniformly dense on K_N for each N in the continuous functions F with $\|F\| \leq 1$ (such F_n exist since $\mathfrak{C}(K_N)$ is a separable Banach space for each N and we can use the Tietze extension theorem).

Given N and n , let $\delta > 0$ be such that $\rho(x, y) < \delta$ and $x \in K_N$ imply

$$|F_n(x) - F_n(y)| < 1/N.$$

Let x_1, \dots, x_r be points of K_N such that for each $x \in K_N$,

$$\rho(x_j, x) < \delta$$

for some j . For $j = 1, \dots, r$ let A_j be the set of all $x \in S$ such that $\rho(x_i, x) \geq \delta$ for $i < j$ and $\rho(x_j, x) < \delta$. Then the sets A_j are disjoint and belong to \mathfrak{S} ; if A is their union,

$$K_N \subset A \subset K_N^\delta.$$

Since A is open, $K_N^\gamma \subset A$ for some $\gamma > 0$, so that $\beta(A) > 1 - 1/N$ for all but finitely many $\beta \in \mathfrak{K}$. Let $\varepsilon = 1/N$,

$$\begin{aligned} G(x) &= F_n(x_j) - \varepsilon, & x \in A_j, j = 1, \dots, r \\ &= -1, & x \notin A \\ H(x) &= F_n(x_j) + \varepsilon, & x \in A_j, j = 1, \dots, r \\ &= 1, & x \notin A. \end{aligned}$$

Then $G \leq F_n \leq H$, G and H are \mathfrak{S} -measurable, and

$$\int (H - G) d\beta \leq 2\varepsilon + 2\varepsilon = 4\varepsilon$$

for any $\beta \in \mathfrak{K}$ with $\beta(A) > 1 - \varepsilon$.

If $\{\beta_m\}$ is any sequence of distinct elements of \mathfrak{K} , we can find a subsequence $\{\beta_{m_r}\}$ such that

$$\int_* F_n d\beta_{m_r}$$

is convergent for a given n , so that

$$\limsup \int_* F_n d\beta_{m_r} - \lim \int_* F_n d\beta_{m_r} \leq 4\varepsilon.$$

Taking further subsequences and diagonalizing, we can assume this holds for all n . Letting ε tend to zero through some sequence and diagonalizing again, we get a subsequence $\{\gamma_q\}$ of $\{\beta_m\}$ such that

$$\liminf_{q \rightarrow \infty} \int_* F_n d\gamma_q = \limsup_{q \rightarrow \infty} \int_* F_n d\gamma_q$$

for all n , so that \liminf and \limsup can be replaced by \lim .

Now let F be a bounded continuous function on S with $\|F\| \leq 1$. Given N , choose n so that

$$|(F - F_n)(x)| < 1/N$$

for all $x \in K_N$; then this will hold for all $x \in K_N^\delta$ for some $\delta > 0$. Then except for finitely many $\beta \in \mathfrak{K}$,

$$\begin{aligned} \int_* F d\beta &\geq \int_* F_n d\beta - 3/N, \\ \int_*^* F d\beta &\leq \int_*^* F_n d\beta + 3/N. \end{aligned}$$

Thus

$$\limsup_{q \rightarrow \infty} \int^* F d\gamma_q - \liminf_{q \rightarrow \infty} \int_* F d\gamma_q \leq 6/N.$$

Since this holds for all N , the limit $M(F)$ defined by

$$M(F) = \lim_{q \rightarrow \infty} \int^* F d\gamma_q = \lim_{q \rightarrow \infty} \int_* F d\gamma_q$$

exists. Then this clearly holds for all $F \in \mathcal{C}$ without the restriction $\|F\| \leq 1$. Clearly M is linear on \mathcal{C} , $M(1) = 1$, and $M(F) \geq 0$ if $F \geq 0$. If $F_n \in \mathcal{C}$ and $F_n \downarrow 0$ pointwise, then $F_n \downarrow 0$ uniformly on compact sets. Given $F \in \mathcal{C}$ with $\|F\| \leq 1$ and

$$\sup_{x \in K_N} |F(x)| \leq 1/N,$$

there is a $\delta > 0$ such that

$$\sup_{x \in K_N^\delta} |F(x)| \leq 2/N$$

so that

$$M(F) \leq 3/N.$$

Thus $M(F_n) \downarrow 0$.

Hence there is a nonnegative, countably additive probability measure P on S such that each $F \in \mathcal{C}$ is P -measurable with

$$M(F) = \int F dP.$$

Since S is a metric space, P is a Borel measure on S . Now $\gamma_q \rightarrow P$ (weak*), and the proof is complete.

PROPOSITION 1. *Suppose (S, ρ) is a complete metric space, α is a finite Borel measure on S , concentrated in a separable subspace, the α_n are defined on S , and $\alpha_n \rightarrow \alpha$ (weak*). Then for any bounded real function F on S which is continuous almost everywhere with respect to α ,*

$$\lim_{n \rightarrow \infty} \int^* F d\alpha_n = \lim_{n \rightarrow \infty} \int_* F d\alpha_n = \int F d\alpha.$$

Proof. We can assume that the α_n and α are all probability measures and $\|F\| = 1$. Given $\varepsilon > 0$, there is a compact set K on which F is continuous with $\alpha(K) \geq 1 - \varepsilon$. Take $\delta > 0$ such that whenever $\rho(x, y) < \delta$ and $x \in K$, $|F(x) - F(y)| < \varepsilon$.

Let G be F restricted to K . Then by the Tietze extension theorem G can be extended to a continuous function on all of S with $\|G\| \leq 1$. There is a $\gamma > 0$ such that $\rho(x, y) < \gamma$ and $x \in K$ imply

$$|G(x) - G(y)| < \varepsilon;$$

we can assume $\gamma < \delta$. For $x \in K^\gamma$,

$$|F(x) - G(x)| \leq 2\varepsilon.$$

Let f be a continuous function on S with $0 \leq f(x) \leq 1$ for all x , $f(x) = 1$ for $x \in K$, and $f(x) = 0$ for $x \notin K^\gamma$. Then

$$\int_* f d\alpha_n \rightarrow \int f d\alpha \geq 1 - \varepsilon$$

so that $\alpha_n(K^\gamma) \geq 1 - 2\varepsilon$ for n large enough, say $n \geq n_0$. Now

$$\lim_{n \rightarrow \infty} \int_* G d\alpha_n = \lim_{n \rightarrow \infty} \int^* G d\alpha_n = \int G d\alpha,$$

$$\left| \int F d\alpha - \int G d\alpha \right| \leq \int |F - G| d\alpha \leq 2\varepsilon.$$

For $n \geq n_0$,

$$\int_* F d\alpha_n \geq \int_* (G - 2\varepsilon) d\alpha_n - \alpha_n(S \sim K^\gamma)$$

$$\geq \int_* G d\alpha_n - 4\varepsilon,$$

$$\int^* F d\alpha_n \leq \int^* G d\alpha_n + 4\varepsilon.$$

Thus

$$\lim_{n \rightarrow \infty} \int^* F d\alpha_n = \lim_{n \rightarrow \infty} \int_* F d\alpha_n = \int F d\alpha, \quad \text{Q.E.D.}$$

In proving weak*-precompactness using Theorem 1, the following is useful: suppose (C, ρ) is a metric space, let $(B, \| \cdot \|)$ be the Banach space of all bounded real-valued functions on C , and for ε and $\delta > 0$ let B_ε^δ be the set of all $f \in B$ such that for some $x, y \in C$, $\rho(x, y) < \delta$ and $|f(x) - f(y)| \geq \varepsilon$.

PROPOSITION 2. *Suppose (C, ρ) is compact and $\mu_n, n = 1, 2, \dots$, are probability measures on the σ -field S in B such that for any $\varepsilon > 0$ there are $\delta > 0, n_0$, and M such that*

$$(\mu_n)^*(B_\varepsilon^\delta) < \varepsilon \quad \text{for } n \geq n_0,$$

and

$$\mu_n\{f : \|f\| \geq M\} < \varepsilon \quad \text{for all } n.$$

Then for any $\varepsilon > 0$ there is a compact set K in B , consisting entirely of continuous functions, such that for any $\gamma > 0$,

$$\mu_n(K^\gamma) \geq 1 - \varepsilon$$

for n sufficiently large.

Proof. We may assume $0 < \varepsilon < 1$. Choose $M > 1$ such that

$$\mu_n\{f : \|f\| \geq M\} < \varepsilon/2$$

for all n , and for $m = 1, 2, \dots, \beta > 0$, let

$$A_m^\beta = B_{\varepsilon/2^m}^\beta.$$

We choose a decreasing sequence $\{\beta_m\}$ of positive numbers satisfying the following two conditions:

- (I) If $\delta_m = \beta_m \varepsilon / 2^{m+1} M$, then $\delta_{m+1} < \delta_m / 4$ for all m .
- (II) For some sequence $\{n_0(m)\}$, $(\mu_n)^*(A_m^{\beta_m}) < \varepsilon / 2^m$ for $n \geq n_0(m)$.

Then let $A_m = A_m^{\beta_m}$. Now if $\|f\| \leq M$, $f \notin A_j$, and $\rho(x, y) \geq \beta_j$, then

$$|f(x) - f(y)| \leq 2M \leq \frac{\varepsilon \beta_j / 2^j}{\varepsilon \beta_j / 2^{j+1} M} \leq \frac{\varepsilon \rho(x, y)}{2^j \delta_j},$$

while if $\rho(x, y) < \beta_j$, $|f(x) - f(y)| \leq \varepsilon / 2^j$. Thus for any $x, y \in C$,

$$(*) \quad |f(x) - f(y)| \leq \frac{\varepsilon}{2^j} \max \left(1, \frac{\rho(x, y)}{\delta_j} \right).$$

Let B_m be the set of $f \in B$ such that $\|f\| \leq M$ and $(*)$ holds for $j = 2, \dots, m$ and all $x, y \in C$. Then

$$(\mu_n)_*(B_m) \geq 1 - \varepsilon \quad \text{for } n \geq N = N(m).$$

Now let K be the set of all $g \in B$ such that $\|g\| \leq M$ and for all j ,

$$\rho(x, y) < \delta_j / 2 \quad \text{implies} \quad |g(x) - g(y)| < 3\varepsilon / 2^j.$$

K is a set of continuous functions, compact by Ascoli's theorem.

Given a $\gamma > 0$, choose an integer $m > 1$ such that $\varepsilon / 2^m < \gamma / 2$. Let us show that $B_m \subset K^\gamma$. Choose a finite set C_m of points of C such that $\rho(x, y) \geq \delta_m$ for any distinct $x, y \in C_m$, and such that for any $z \in C$, $\rho(x, z) < \delta_m$ for some $x \in C_m$ (choosing points one by one to satisfy the first condition, we end with a finite set satisfying the second condition).

If $f \in B_m$ and $x, y \in C_m$,

$$|f(x) - f(y)| \leq \frac{\varepsilon \rho(x, y)}{2^m \delta_m}.$$

Let f_m be f restricted to C_m . Then f_m can be extended to a function g on C satisfying

$$|g(x) - g(y)| \leq \varepsilon \rho(x, y) / 2^m \delta_m$$

for all $x, y \in C$ (Czipszer and Geher [2]). We can assume $\|g\| \leq M$. Let us show that $g \in K$. For $j \geq m$, since $\varepsilon / 2^m \delta_m \leq \varepsilon / 2^j \delta_j$,

$$|g(x) - g(y)| \leq \varepsilon / 2^j \quad \text{for } \rho(x, y) < \delta_j.$$

For $j < m$, given $x, y \in C$ with $\rho(x, y) < \delta_j / 2$, choose $x_m, y_m \in C_m$ with $\rho(x, x_m) < \delta_m$ and $\rho(y, y_m) < \delta_m$. Then

$$\rho(x_m, y_m) < \delta_j / 2 + 2\delta_m < \delta_j,$$

$$\begin{aligned} |g(x) - g(y)| &\leq |g(x) - g(x_m)| + |f(x_m) - f(y_m)| + |g(y_m) - g(y)| \\ &\leq \varepsilon / 2^m + \varepsilon / 2^j + \varepsilon / 2^m \leq 3\varepsilon / 2^j. \end{aligned}$$

Thus $g \in K$. Finally $\|f - g\| < \gamma$ since for any $x \in C$ and $x_m \in C_m$ with

$$\rho(x, x_m) < \delta_m,$$

$$|f(x) - g(x)| \leq |f(x) - f(x_m)| + |g(x_m) - g(x)| \leq 2\varepsilon/2^m < \gamma.$$

Thus indeed $B_m \subset K^\gamma$, so that $\mu_n(K^\gamma) \geq 1 - \varepsilon$ for n sufficiently large, Q.E.D.

3. Empirical measures on Euclidean spaces

Let R^k be the Cartesian space of ordered k -tuples $t = \langle t_1, \dots, t_k \rangle$ of real numbers. Suppose μ is a Borel probability measure on R^k . Let X_1, X_2, \dots , be independent R^k -valued random variables with distribution μ ; specifically, let Ω be a countably infinite product of probability spaces isomorphic to (R^k, μ) , with X_i as coordinate functions, and call the product measure Pr . For any $t \in R^k$, let δ_t be the unit measure at t , and let μ_n be the measure

$$(\delta_{x_1} + \dots + \delta_{x_n})/n, \quad n = 1, 2, \dots$$

Then the μ_n will be called ‘‘empirical measures’’ for μ . They may be thought of as approximations to μ given by a series of independent trials. Let

$$\mu^n = \sqrt{n}(\mu_n - \mu).$$

Let H be the Hilbert space $L^2(R^k, \mu)$. For any $f_1, \dots, f_m \in H$ the multi-dimensional central limit theorem implies that the joint distribution of

$$\int f_1 d\mu^n, \dots, \int f_m d\mu^n$$

converges as $n \rightarrow \infty$ to that of

$$L(f_1), \dots, L(f_m)$$

where L is a linear map of H into a space of Gaussian random variables with mean zero and

$$E(L(f)L(g)) = \int \left(f - \int f d\mu \right) \left(g - \int g d\mu \right) d\mu,$$

$f, g \in H$. L is the ‘‘centered noise’’ r.l.f. with spectral measure μ as defined in [5, §7].

Given $t = \langle t_1, \dots, t_k \rangle \in R^k$ let A_t be the indicator function of the set B_t of all $s \in R^k$ such that for each j ,

$$t_j \leq s_j < 0 \quad \text{or} \quad 0 \leq s_j < t_j.$$

(B_t is empty if any t_j is zero.)

For $k = 1$, if μ is a nonatomic measure, the celebrated Kolmogorov-Smirnov theorems give information on the limiting behavior as $n \rightarrow \infty$ of the distribution of

$$\sup_t |\mu^n(B_t)|$$

and related quantities (see e.g. Fortet [6, Chapter 5]). It is not hard to see that, still for $k = 1$, the distributions are the same for any nonatomic μ on the half-line $t > 0$. One approach to the Kolmogorov-Smirnov results is to show that the limiting distribution is that of

$$\sup_t |L(A_t)|,$$

and that this is true not only for the supremum but for a large class of other functionals (Donsker [3], [4]). In this paper I shall extend this last result to $k > 1$ although the "invariance principle" no longer holds, i.e. the distributions depend on what continuous measure μ is chosen.

Let J be the space of all bounded real-valued functions f on R^k such that for any $t \in R^k$, $f(s) \rightarrow f(t)$ if $s_j \uparrow t_j$ for $j = 1, \dots, k$. Then it is easy to verify that J is a Banach space with the supremum norm

$$\|f\| = \sup_{t \in R^k} |f(t)|,$$

and that if Q is any countable dense set in R^k and $f \in J$,

$$\|f\| = \sup_{t \in Q} |f(t)|.$$

It is clear that the functions V_n^μ or V_n ,

$$V_n(t) = \mu^n(B_t),$$

belong to J , and that the set of such functions includes no countable dense subset for the supremum norm, even if a set of zero probability is removed, unless μ is purely atomic.

I want to show that for any real-valued function F on J which is continuous for the given norm, the limiting distribution of $F(\mu^n(B_t))$ is that of $F(L(A_t))$. The formulation requires special attention since the distribution of $\mu^n(B_t)$ will in general *not* be a Borel measure on J , and F may not be measurable for this distribution, so that $F(\mu^n(B_t))$ will not have a well-defined distribution. Nevertheless its distribution is defined with increasing precision as $n \rightarrow \infty$ and does approach that of $F(L(A_t))$ in a sense to be explained below.

The functions $\mu^n(B_t)$ for $k = 1$ have only jump discontinuities, and by introducing the "Skorokhod topology" (Skorokhod [10] and [11], Kolmogorov [7]) on the space of such functions and considering only functionals F continuous for this topology, one can avoid the imprecision in the definition of the distribution of $F(\mu^n(B_t))$ (Prokhorov [9, Theorem 2.4]). The method used here yields a larger class of functionals and easily implies the results using the Skorokhod topology. Also, no useful generalization of the Skorokhod topology to functions on R^k seems to be known.

Let Q_μ be as defined in [5, §4], i.e. the set of real-valued functions on R^k , continuous except on hyperplanes $t_j = \text{constant}$ having positive μ -measure, and there being continuous from below with limits from above, and with limits at $\pm \infty$. Then $Q_\mu \subset J$, and Q_μ is a separable Banach space.

Let L or L^μ be the centered noise r.l.f. with spectral measure μ . Then by [5, Theorem 4.2] the functions

$$t \rightarrow L^\mu(A_t)$$

may be taken in Q_μ with probability 1. Since $(Q_\mu, \|\cdot\|)$ is a separable Banach space, and a function in Q_μ is determined, for purposes of membership in an open ball and hence any set in \mathcal{S} , by its values on any countable dense set in R^k , there is a Borel measure α or α^μ on Q_μ such that for any $t(1), \dots, t(n) \in R^k$ the joint probability law of $f(t(1)), \dots, f(t(n))$, where f has distribution α , is the same as that of

$$L(A_{t(1)}), \dots, L(A_{t(n)}).$$

α can also be regarded as a Borel measure on J since Q_μ is a closed subspace.

Now let $\alpha_n = \alpha_n^\mu$ be the probability distribution of the function $V_n^\mu = \mu^n(B_{(\cdot)})$ in J for $n = 1, 2, \dots$. Then the α_n are, under certain mild assumptions, *not* definable as (countably additive) Borel measures on J (or any other space containing all the V_n^μ , with supremum norm). For example, if μ is not purely atomic, α_1 gives positive measure to an uncountable set \mathfrak{X} such that $\|f - g\| = 1$ for any $f, g \in \mathfrak{X}$ with $f \neq g$. If α_1 had a Borel extension, we would have a countably additive probability measure on all subsets of \mathfrak{X} (since every subset is closed), giving points measure zero, which is impossible assuming the continuum hypothesis (Banach and Kuratowski [1]).

What we have, then, is the following: each $\alpha_n, n = 1, 2, \dots$, is defined by mapping W_n of an n -fold product of R^k 's into J , and α_n is defined exactly on those sets A such that $W_n^{-1}(A)$ is a measurable set in the product.

Here is my main theorem on empirical measures, whose proof will be completed in §5:

THEOREM 2. *If F is a bounded real-valued function on J , continuous almost everywhere with respect to α , then*

$$\lim_{n \rightarrow \infty} \int_* F d\alpha_n = \lim_{n \rightarrow \infty} \int^* F d\alpha_n = \int F d\alpha.$$

In particular, $\alpha_n \rightarrow \alpha$ (weak).*

For $k = 1$, Donsker [4] asserts that under the same hypotheses

$$\int F d\alpha_n \rightarrow \int F d\alpha.$$

There is not even a measurability assumption on F away from the support of α , and an examination of his proof indicates that one has only convergence of upper and lower integrals.

In order to treat unbounded functionals such as the supremum, we have

THEOREM 3. *If F is a real-valued function on J , continuous for the supremum*

norm almost everywhere with respect to α , and b is a real number such that $\alpha(f : F(f) = b) = 0$, then

$$\lim_{n \rightarrow \infty} \alpha_n^*(f : F(f) < b) = \lim_{n \rightarrow \infty} (\alpha_n)_*(f : F(f) < b) = \alpha(f : F(f) < b).$$

Proof. We use Theorem 2. Given $\varepsilon > 0$ take continuous functions g and h from the real line to the unit interval with

$$\begin{aligned} g(x) &= 1, & x &\leq b - \varepsilon \\ g(x) &= 0, & x &\geq b; \\ h(x) &= 1, & x &\leq b \\ h(x) &= 0, & x &\geq b + \varepsilon. \end{aligned}$$

Then we apply Theorem 2 to $g \circ F$ and $h \circ F$. Noting that

$$\begin{aligned} \int_* g \circ F \, d\alpha_n &\leq (\alpha_n)_*(F < b) \leq \alpha_n^*(F < b) \\ &\leq \int^* h \circ F \, d\alpha_n, \end{aligned}$$

and

$$\int (h - g) \circ F \, d\alpha \rightarrow 0 \quad \text{as } \varepsilon \downarrow 0,$$

the proof is complete.

Theorem 2 will be proved first for “continuous” measures on the unit cube. In this case, we have an “equicontinuity” result (Theorem 4, §4) which implies, using the abstract results in §2, that the α_n form a “weak*-precompact” set. The final details and the passage from a general probability measure on R^k to a continuous one on the unit cube will be given in §5.

4. “Equicontinuity”

A measure μ on R^k will be called *continuous* if each hyperplane $t_j = \text{constant}$ has measure zero. Let C be the unit cube

$$\{t : 0 \leq t_j < 1, j = 1, \dots, k\}.$$

We use the notation of Proposition 2 for this choice of C .

THEOREM 4. *Suppose μ is a continuous Borel probability measure on C . Then for any $\varepsilon > 0$ there is a $\delta > 0$ such that*

$$\Pr(V_n \in B_\varepsilon^\delta) \leq \varepsilon$$

for all large enough n .

For the proof, we first mention a sort of Markov property of the random measures μ^n . Note that μ^n and μ_n are functions of each other and of μ , and for a given measurable set E , $\mu^n(E)$ has only finitely many possible values, so that the definition of conditional probabilities is elementary.

PROPOSITION 3. Given a probability space (S, μ) and measurable sets

$$E_i \subset F \subset G \subset S$$

for finitely many values of i , $\mu^n(G)$ is conditionally independent of the $\mu^n(E_i)$ given $\mu^n(F)$. The conditional distribution of $\mu_n(G \sim F)$ given that $\mu_n(F) = r/n$, and hence also given any consistent values of $\mu_n(E_i)$, is exactly that of

$$\frac{n-r}{n} \nu_{n-r}(G \sim F)$$

where ν is μ restricted to $C \sim F$ and then normalized to total mass 1 (or, if $\mu(C \sim F) = 0$, $\mu_n(G \sim F)$ is almost surely zero).

Proof. It suffices to show that the conditional distribution of $\mu_n(G \sim F)$, given $n\mu_n(F) = r$ and given the $\mu_n(E_i)$, is as indicated. Taking independent random variables X_1, \dots, X_n with distribution μ defining μ_n , the conditional distribution clearly does not depend on which set of r of the X_i is included in F . For a given set, since the $\mu_n(E_i)$ are independent of the complementary set, the conclusion follows by an elementary calculation.

Now to prove Theorem 4 we use induction on k . For $k = 1$, given μ , let F be its distribution function:

$$F(x) = \mu([0, x]).$$

Let λ be Lebesgue measure on $[0, 1) = I$,

$$W_n(t) = \lambda^n(B_t), \quad V_n(t) = \mu^n(B_t).$$

Then we can put

$$V_n(t) = W_n(F(t)).$$

Since F is uniformly continuous, it suffices to prove our assertion where μ is Lebesgue measure. Here, Prokhorov [9, Lemmas 2.8 and 2.9, pp. 209–210 (original), pp. 187–188 (translation)] has proved the following: given $\beta > 0$, $\varepsilon > 0$, $s, t \in I$, let $A(\beta, \varepsilon, s, t)$ be the set of all $\omega \in \Omega$ such that

$$\min(|V_n(s - \beta) - V_n(t)|, |V_n(s + \beta) - V_n(t)|) \geq \varepsilon.$$

Then there is a $\delta > 0$ such that

$$\Pr(\omega \in A(\beta, \varepsilon, s, t) \text{ for some } \beta, s, t \text{ with } |s - t| \leq \beta \leq \delta) \leq \varepsilon.$$

Now if $\omega \notin A(\beta, \varepsilon, s, t)$ whenever $|s - t| \leq \beta \leq \delta$, and there are $s \in I$ and $\beta, 0 < \beta \leq \delta$, such that

$$|V_n(s + \beta) - V_n(s - \beta)| \geq 3\varepsilon,$$

then there is at least one jump of height ε or more in the graph of V_n between $s - \beta$ and $s + \beta$. However, for $n > 1/\varepsilon^2$, since μ_n is concentrated in n distinct points with probability one, the probability of such a jump is zero. Thus

$$Pr(V_n \in B_{3\epsilon}^{2\delta}) \leq \epsilon$$

for $n > 1/\epsilon^2$, and the theorem is proved for $k = 1$.

Now suppose the conclusion is true for $1, \dots, k - 1$. Let L_m be the cubical lattice of all points

$$\langle r_1/2^m, \dots, r_k/2^m \rangle,$$

$r_j = 0, 1, \dots, 2^m, j = 1, \dots, k$. We put the lexical ordering on L_m :

$$\langle a_1, \dots, a_k \rangle < \langle b_1, \dots, b_k \rangle$$

if and only if there is a j such that $a_i = b_i$ for $i < j$ and $a_j < b_j$.

It suffices to prove the theorem inserting the condition that $s_j = t_j$ for all $j \neq i$ for some i , say $i = k$. We can also assume that s and t both belong to L_m for some m (of course, δ must not depend on m). This shows in particular that the probability which appears in the theorem is well-defined. Let

$$\phi_n(x) = V_n(x, 1, \dots, 1), \quad 0 \leq x \leq 1.$$

Let $\epsilon > 0$ be given. By the result for $k = 1$, there is a positive integer M such that

$$Pr(\phi_n \in B_{\epsilon/4}^{1/2M}) \leq \epsilon/4$$

for $n > N_1 = 144/\epsilon^2$.

For $j = 0, 1, \dots, 2^M - 1$ let S_j be the slab

$$j/2^M < t_1 \leq (j + 1)/2^M.$$

We assume M is large enough so that $\mu(S_j) < 1$ for all j .

By the result for $k - 1$ there are a $\delta_2 > 0$ and an $N_2 > 0$ such that

$$\begin{aligned} Pr(|V_n(s) - V_n(t)| \geq \epsilon/2^{M+2} \text{ for some } s, t \\ \text{with } |s - t| < \delta_2 \text{ and } s_1 = t_1 = 1) \\ \leq \epsilon/2^{M+2} \end{aligned}$$

if $n \geq N_2$. Also, there is a $\delta_3, 0 < \delta_3 < \frac{1}{2}^M$, such that if λ is an arbitrary probability measure on a measurable space $(S, \mathfrak{J}), A \in \mathfrak{J}$, and $\lambda(A) < \delta_3$, then for all r ,

$$Pr(|\lambda^r(A)| > \epsilon/2^{M+2}) < \epsilon/2^{M+2}$$

(this follows e.g. from Chebyshev's inequality). Let

$$b = \min_j \mu(C \sim S_j);$$

then $b > 0$ by assumption. Choose $\delta_4 > 0$ so that

$$\sup_a \mu\{s : a \leq s_k \leq a + \delta_4\} \leq b\delta_3.$$

Let $\delta = \min(\delta_2, \delta_4)$. We shall show that this δ satisfies the desired condition. For the rest of the proof, not only ϵ and δ but m will be fixed, $m \geq M$.

Let $\Gamma(j, n)$ be the set of all ω such that for some $s, t \in S_j \cap L_m$,

$$|s - t| < \delta, \quad s_i = t_j \quad \text{for } i < k,$$

and

$$|\mu^n(B_s \cap S_j) - \mu^n(B_t \cap S_j)| \geq \varepsilon/2^M.$$

For $\omega \in \Gamma(j, n)$ let $s(\omega) = s(j, n, \omega)$ be the least s (lexically) in $S_j \cap L_m$ for which a t satisfying the above three conditions exists, and then let $t(\omega)$ be the least such t . Let

$$E = E(j, n, \omega) = \{u \in C : j/2^M < u_1 < s(\omega)_1\},$$

and for any $s \in C$ let

$$P(s) = (1, s_2, \dots, s_k).$$

Then, since $\delta \leq \delta_2$,

$$\Pr [|V_n(P(s(\omega))) - V_n(P(t(\omega)))| \geq \varepsilon/2^{M+2}] \leq \varepsilon/2^{M+2}$$

for any j and n (a set defined by a condition on $s(\omega)$ and $t(\omega)$ will be regarded as a subset of $\Gamma(j, n)$, their domain of definition).

Let $F = F(j, n, \omega)$ be the set

$$(B_{P(t(\omega))} \sim B_{P(s(\omega))}) \sim (B_{t(\omega)} \cap S_j).$$

Then F is disjoint from E , and E is determined by F , $E = E(F)$.

The conditional distribution of $\mu^n(F)$ given a value F_0 of F and given $\mu^n(E)$ is the same as the conditional distribution of $\mu^n(F_0)$ given $\mu^n(E(F_0))$, by Proposition 3, since knowing that $F = F_0$ in addition yields information only on values of μ_n on subsets of $E(F_0)$.

Let $\Omega_1 = \Omega_1(n)$ be the set of ω with ($\omega \in \Gamma(j, n)$ and)

$$|\mu^n(E(j, n, \omega))| > \varepsilon/4$$

for some j . Then for $n \geq N_1$, $\Pr(\Omega_1(n)) \leq \varepsilon/4$. Let $G = C \sim E$ and $r = n\mu_n(G)$.

For given values of j, m, n, F and r , the conditional distribution of $n\mu_n(F)$ is exactly that of $r\lambda_r(F)$, where λ is μ restricted to G and normalized to mass 1, by Proposition 3. Since $\mu(F)/\mu(G) \leq \delta_3$,

$$\Pr \left\{ \left| \frac{n\mu_n(F)}{\sqrt{r}} - \sqrt{r} \frac{\mu(F)}{\mu(G)} \right| \geq \varepsilon/2^{M+2} \mid F, r \right\} \leq \varepsilon/2^{M+2}.$$

Thus if $\Omega_2(j, n)$ is the set of $\omega \in \Gamma(j, n)$ with

$$\left| \frac{n\mu_n(F)}{\sqrt{r}} - \frac{\sqrt{r} \mu(F)}{\mu(G)} \right| \geq \varepsilon/2^{M+2},$$

then $\Pr(\Omega_2(j, n)) \leq \varepsilon/2^{M+2}$.

Let $N = \max(N_1, N_2)$. We fix $n \geq N$ and suppress “ n ” in some notations.

Now note that

$$\begin{aligned} \mu^n(F) &= \sqrt{n}(\mu_n(F) - \mu(F)) \\ &= \frac{\sqrt{r}}{\sqrt{n}} \left(\frac{n\mu_n(F)}{\sqrt{r}} - \frac{\sqrt{r}\mu(F)}{\mu(G)} \right) + \left(\frac{r\mu(F)}{\sqrt{n}\mu(G)} - \sqrt{n}\mu(F) \right), \end{aligned}$$

and

$$\frac{r\mu(F)}{\sqrt{n}\mu(G)} - \sqrt{n}\mu(F) = \frac{\mu(F)}{\mu(G)}\mu^n(G) = -\frac{\mu(F)}{\mu(G)}\mu^n(E).$$

For $\omega \in \Gamma(j)$ but neither in Ω_1 nor in $\Omega_2(j)$, $r \leq n$ and $\mu(F)/\mu(G) \leq \frac{1}{2}^M$ imply

$$|\mu^n(F)| \leq 2\varepsilon/2^{M+2} = \varepsilon/2^{M+1},$$

and hence

$$|V_n(P(t(\omega))) - V_n(P(s(\omega)))| \geq \varepsilon/2^{M+1}.$$

If $\Omega_3(j)$ is the set of $\omega \in \Gamma(j)$ for which the latter inequality holds, then since $\delta \leq \delta_2$,

$$\Pr(\Omega_3(j)) < \varepsilon/2^{M+2}.$$

Thus for each j ,

$$\Pr(\Gamma(j) \sim \Omega_1) \leq \Pr(\Omega_2(j)) + \Pr(\Omega_3(j)) \leq \varepsilon/2^{M+1}.$$

Now if $s, t \in C$, $|s - t| < \delta$, and $s_j = t_j$ for $j < k$, then $V_n(s) - V_n(t)$ is the sum of at most 2^M terms

$$\mu^n(B_{f(s,q)} \cap S_q) - \mu^n(B_{f(t,q)} \cap S_q)$$

where for any $u \in C$, $f(u, q) \in S_q$,

$$\bar{B}_u \cap S_q = \bar{B}_{f(u,q)} \cap S_q,$$

and for $u \in L_m$ and $m \geq M$, $f(u, q) \in L_m$. Thus

$$|f(s, q) - f(t, q)| < \delta$$

for each q , and the probability that at least one of the 2^M terms exceeds $\varepsilon/2^{M+1}$ in absolute value is less than

$$\Pr(\Omega_1) + \varepsilon/2 < \varepsilon.$$

Thus for $n \geq N$,

$$\Pr(|V_n(s) - V_n(t)| \geq \varepsilon \text{ for some } s, t \in C \text{ with } s_j = t_j$$

$$\text{for } j < k \text{ and } |s - t| < \delta)$$

$$< \varepsilon.$$

For any $\delta > 0$ and $s, t \in R^k$ with $|s - t| < \delta$, there are $u, v \in C$ with $|u - v| < \delta$ and $V_n(s) \equiv V_n(u)$, $V_n(t) \equiv V_n(v)$. Thus the condition “ $s, t \in C$ ” can be removed and the proof of Theorem 4 is complete.

5. Proof of Theorem 1

We first complete the proof assuming μ is continuous and concentrated in C . In this case, Theorems 1 and 4 and Proposition 2 show that the measures α_n

on J form a weak*-precompact set. Now suppose

$$\alpha_{n_m} \rightarrow \beta \quad (\text{weak}^*)$$

for some subsequence $\{\alpha_{n_m}\}$ of $\{\alpha_n\}$. Then for any $t^{(1)}, \dots, t^{(n)} \in C$ the joint distribution of $f(t^{(1)}), \dots, f(t^{(n)})$ for f distributed according to β is the same as for f distributed according to α , by the central limit theorem convergence mentioned in §3. Thus $\beta = \alpha$.

Suppose there is a bounded continuous function F on $(J, \|\cdot\|)$ and an $\varepsilon > 0$ such that

$$\max \left(\left| \int^* F d\alpha_n - F d\alpha \right|, \left| \int F d\alpha - \int_* F d\alpha_n \right| \right) \geq \varepsilon$$

for an infinite sequence of values of n . Then taking a subsequence n_m such that α_{n_m} is weak* convergent, we have a contradiction. Thus

$$\alpha_n \rightarrow \alpha \quad (\text{weak}^*).$$

Applying Proposition 1, §2 the proof for μ continuous on C is complete.

The general case of a probability measure on R^k is easily reduced to that of a measure on the cube

$$-1 < t_j < 1, \quad j = 1, \dots, k,$$

by the transformation

$$t_j \rightarrow (2/\pi) \arctan t_j, \quad j = 1, \dots, k,$$

which preserves all the structure we need.

Let E_t be the indicator function of the set

$$\{s : -1 \leq s_j < t_j, j = 1, \dots, k\}$$

Then for $t_j \geq -1, j = 1, \dots, k,$

$$A_t = \sum_F S(F) E_{P(t, F)}$$

where F runs over all finite subsets of $(1, \dots, k), S(F) = (-1)^r$ where $r = k - \#(F)$ and $\#(F)$ is the number of elements in F , and $P(t, F) = s$ with $s_j = \max(t_j, 0)$ for $j \in F$ and $s_j = \min(t_j, 0)$ otherwise. Since the functions $t \rightarrow P(t, F)$ are uniformly continuous for each F , results for the E_t , obtained by an obvious linear transformation, yield corresponding results for the A_t . Thus we have reduced to the case of a measure on C .

In [5, §4] the following was proved: if μ is any probability measure on C , there is a continuous measure ν on some rectangular solid

$$C_1 : 0 \leq t_j < a_j, \quad j = 1, \dots, k,$$

and a continuous mapping G of the closure \bar{C}_1 of C_1 onto \bar{C} , of the form

$$G(t_1, \dots, t_k) = \langle G_1(t_1), \dots, G_k(t_k) \rangle,$$

with each G_j (weakly) monotone increasing, such that $\nu \circ G^{-1} = \mu$. Clearly we can assume $a_j \equiv 1$, $C_1 = C$, and extend each G_j to a monotone function from $[0, \infty)$ onto itself.

If X_n , $n = 1, 2, \dots$, are independent with distribution ν then $G(X_n)$ are independent with distribution μ , so we can put

$$\mu^n = \nu^n \circ G^{-1}.$$

For $j = 1, \dots, k$, $0 \leq t_j \leq 1$, let

$$F_j(t_j) = \inf (s : G_j(s) = t_j),$$

and let $F(t) = \langle F_1(t_1), \dots, F_k(t_k) \rangle$.

Now for any j , s_j , t_j we have $G_j(s_j) < t_j$ if and only if $s_j < F_j(t_j)$. Thus $A_t = A_s \circ G$ if and only if $F(s) = t$, and for any $s \in \bar{C}$,

$$\begin{aligned} V_n^\mu(s) &= \int A_s(v) d\mu^n(v) = \int A_s(G(u)) d\nu^n(u) \\ &= \int A_{F(s)}(\mu) d\nu^n(u) = V_n^\nu(F(s)). \end{aligned}$$

Now for any continuous function f on \bar{C} , $F^*(f) = f \circ F \in Q_\mu$ since $F_j \in Q_\mu$ for each j . The map $F^* : f \rightarrow f \circ F$ is linear with norm one. Since

$$\alpha_n^\mu = \alpha_n^\nu \circ (F^*)^{-1} \quad \text{and} \quad \alpha_n^\nu \rightarrow \alpha^\nu \text{ (weak*)},$$

we have $\alpha_n^\mu \rightarrow \alpha^\mu$ (weak*). Then applying Proposition 1 of §2, the proof of Theorem 2 is complete.

Note. The supremum norm $N = \| \|$ is an \mathcal{S} -measurable function. Letting $s \in C_t$ if and only if $s_j \leq t_j, j = 1, \dots, k$, and letting $f_n(t) = \mu^n(C_t)$, it is clear that $N(f_n)$ has a well-defined distribution for each n . For $k = 1$ and μ continuous, it was shown by Kolmogorov (see e.g. [6, Chapter 5]) that the limiting distribution of $N(f_n)$ as $n \rightarrow \infty$ is

$$\Pr (N < z) = K(z) = 1 + 2 \sum_{r=1}^\infty (-1)^r \exp (-2r^2 z^2).$$

On a k -fold product of real lines for $k > 1$,

$$R^k = \prod_{j=1}^k R_j,$$

the limiting distribution of N will not be the same for all continuous measures μ . However, it will be the same for all product measures

$$\mu = \prod_{j=1}^k \mu_j,$$

where μ_j is continuous on R_j for each j . In this case,

$$\mu^n(C_t) = \prod_{j=1}^k \mu_j^n(C_{t_j}),$$

the μ_j^n are independent for different j , and if f_j is a function of t_j only,

$$N(\prod_{j=1}^k f_j) = \prod_{j=1}^k N(f_j).$$

Thus we have a product of n independent random variables with the same distribution function K . I have not found the distribution of this product.

Although the function(al) S defined by

$$S(f) = \sup_t f(t)$$

has a simpler limiting distribution for a continuous measure in one dimension,

$$\Pr(S < z) = 1 - \exp(-2z^2)$$

[6, Chapter 5], the situation is more complicated in several dimensions because

$$S(\prod_{j=1}^k f_j(t_j))$$

depends on the infima as well as the suprema of the f_j .

REFERENCES

1. S. BANACH AND C. KURATOWSKI, *Sur une généralisation du problème de la mesure*, Fund. Math., vol. 14 (1929), pp. 127–131.
2. J. CZIPSZER AND L. GEHER, *Extension of functions satisfying a Lipschitz condition*, Acta Math. Acad. Sci. Hungar., vol. 6 (1955), pp. 213–220.
3. M. D. DONSKER, *An invariance principle for certain probability limit theorems*, Mem. Amer. Math. Soc., no. 6 (1951).
4. ———, *Justification and extension of Doob's heuristic approach to the Kolmogorov-Smirnov theorems*, Ann. Math. Statist., vol. 23 (1952), pp. 277–281.
5. R. M. DUDLEY, *Gaussian processes on several parameters*, Ann. Math. Statist., vol. 36 (1965), pp. 771–788.
6. R. FORTET, *Recent advances in probability theory*, Some aspects of analysis and probability, New York, Wiley, 1958, pp. 171–240.
7. A. N. KOLMOGOROV, *On the Skorokhod convergence*, Theory of probability and its applications, vol. 1 (1956), pp. 239–247 (Russian), pp. 215–222 (English translation).
8. E. MARCZEWSKI AND P. SIKORSKI, *Measures in nonseparable metric spaces*, Colloq. Math., vol. 1 (1948), pp. 133–139.
9. YU. V. PROKHOROV, *Convergence of random processes and limit theorems in probability*, Theory of probability and its applications, vol. 1, (1956), pp. 177–238 (Russian), pp. 151–214 (English translation).
10. A. V. SKOROKHOD, *On passage to the limit from sequences of sums independent random variables to a homogeneous random process with independent increments*, Dokl. Akad. Nauk SSR, vol. 104 (1955), pp. 364–367 (in Russian).
11. ———, *On a class of limit theorems for Markov chains*, Dokl. Akad. Nauk SSR, vol. 106 (1956), pp. 781–784 (in Russian).

UNIVERSITY OF CALIFORNIA
BERKELEY, CALIFORNIA