The Annals of Mathematical Statistics
1969, Vol. 40, No. 1, 40–50

THE SPEED OF MEAN GLIVENKO-CANTELLI CONVERGENCE

By R. M. Dudley

Massachusetts Institute of Technology

Let (S, d) be a separable metric space. Let $\mathcal{P}(S)$ be the set of all Borel probability measures on S, and $\mu \varepsilon \mathcal{P}(S)$. Let $X_1, X_2, \cdots$, be independent S-valued random variables with distribution μ. For each $x \varepsilon S$ let δ_x be the unit mass at x. Let μ_n be the "empirical measure"

$$(\delta_{X_1} + \cdots + \delta_{X_n})/n.$$

Then the Glivenko-Cantelli theorem states that with probability 1, $\mu_n \to \mu$ weak-star as $n \to \infty$, i.e. for every bounded continuous real-valued function f on S,

$$\int f\, d\mu_n \to \int f\, d\mu.$$

(For a fixed f, this is the strong law of large numbers.) In this generality, the Glivenko-Cantelli theorem apparently is due to Varadarajan [16].

Weak-star convergence in $\mathcal{P}(S)$ is metrizable, by various metrics. In this paper we consider two such metrics: that of Prokhorov [11], which we call $\boldsymbol{\rho}$, and the "BL^* norm" metric β (see details in Section 2 below). β was apparently first used by Fortet and Mourier [9], who proved $\beta(\mu_n, \mu) \to 0$ almost surely.

If for some $K < \infty$ and $k > 2$, S can be covered by at most $K\epsilon^{-k}$ sets with diameter $\leqq 2\epsilon$ whenever $0 < \epsilon < 1$, we prove in Section 3 below that for some $M < \infty$, $E\beta(\mu_n, \mu) < Mn^{-1/k}$ for all n. In Section 4 we prove $E\rho(\mu_n, \mu) < Mn^{-1/(k+2)}$. Moreover the covering may omit a set of μ-measure ϵ (for ρ) or $\epsilon^{k/(k-2)}$ (for β). These results are shown to be best possible by certain examples; for β, by Lebesgue measure on the unit cube in R^d, $k = d \geqq 3$; and for ρ, by the d-fold product of Cantor measure spaces, $k = (d \log 2)/\log 3$. However, for other measures the convergence may be faster, especially for ρ.

It is worth noting that the above results are consistent with, but not related by, the best possible general inequalities between small values of β and ρ, which are of the form

$$c\rho(\mu, \nu)^2 \leqq \beta(\mu, \nu) \leqq C\rho(\mu, \nu)$$

for some constants $c, C > 0$ and all $\mu, \nu \varepsilon \mathcal{P}(S)$ ([6], latter part of Section 2).

In Section 6 we briefly discuss the "classical" case in which μ is Lebesgue measure on the unit interval [0, 1]. Here both $E\beta(\mu_n, \mu)$ and $E\rho(\mu_n, \mu)$ approach 0 at the rate of $n^{-\frac{1}{2}}$, and this is connected to the central limit theorem. In higher dimensions, there seems to be no such connection and the convergence is slower. If $\alpha < \frac{1}{2}$, then for any fixed $f \varepsilon L^2(S, \mu)$, $\int fn^\alpha\, d(\mu_n - \mu) \to 0$ in probability as $n \to \infty$, so our speed of convergence theorems do not seem to be connected to convergence of a renormalized $\mu_n - \mu$ in law to any non-zero limit. We shall not

Received 6 February 1968.

find here the rate of probability 1 convergence (analogous to the law of the iterated logarithm).

A fundamental problem in statistics is, given $X_1, \cdots, X_n$ or an empirical measure ν_n, to test the hypothesis that they arise from a given μ, i.e. $\mu = \nu$. In principle, the results of this paper provide such tests. However, the metrics ρ and β are not easily computed in practice (by any method known to me). Another method is to decompose S into subsets S_j, say m of them, of equal μ-measure $1/m$, and compute $S(n, m)$, defined as

$$\sum_j |(\mu_n - \mu)(S_j)|.$$

It is easily shown (Proposition 3.1) that this has expectation less than $(m/n)^{\frac{1}{2}}$, and we prove in Section 5 that it is of this order of magnitude with probability bounded away from 0 for $n \geqq m$. These results are used to prove some of those mentioned previously. The distribution of $S(n, m)$ is independent of μ.

If m is fixed and small and n sufficiently large, one can apply a χ^2 test, i.e. consider

$$\sum_j (\mu_n - \mu)(S_j)^2$$

(cf. [13]). In other cases, perhaps the results of this paper will suggest other, more suitable tests.

Given μ_m and ν_n one may also wish to test the hypothesis that $\mu = \nu$. If S has finite dimension k in our sense (e.g. if it is a compact set with interior in R^k) then one may apply our results to obtain tests involving $\beta(\mu_m, \nu_n)$ and $\rho(\mu_m, \nu_n)$, but *a priori* we do not know what sets S_j have equal μ- or ν-measure. In such cases one may apply the Fisher permutation principle ([2], [14], [17]). We shall further discuss this "two-sample" problem in this paper.

2. Definitions and preliminaries. Given $\epsilon > 0$ let $N(S, \epsilon)$ be the minimal number of sets (possibly $+\infty$) in a covering of S by sets of diameter at most 2ϵ. Then $H(S, \epsilon)$, the ϵ-entropy of S, is defined as $\log N(S, \epsilon)$ (Kolmogorov). We define the *entropic dimension* of S by

$$k(S) = \limsup_{\epsilon \downarrow 0} H(S, \epsilon)/\log(1/\epsilon).$$

Suppose $\mu \in \mathcal{P}(S)$ and $\epsilon, \delta > 0$. Let $N(\mu, \epsilon, \delta)$ be the minimal number of sets of diameter $\leqq 2\epsilon$ which cover S except for a set A with $\mu(A) \leqq \delta$ (cf. Posner et al. [10]). Clearly $N(\mu, \epsilon, \delta) \leqq N(S, \epsilon)$. Let

$$N(\mu, \epsilon) = N(\mu, \epsilon, \epsilon), \qquad H(\mu, \epsilon) = \log N(\mu, \epsilon),$$
$$k(\mu) = \limsup_{\epsilon \downarrow 0} H(\mu, \epsilon)/\log(1/\epsilon).$$

Let $BL(S, d)$ be the Banach space of all bounded Lipschitzian real-valued functions f on S with the norm

$$\|f\|_{BL} \equiv \|f\|_\infty + \|f\|_L$$
$$\equiv \sup_x |f(x)| + \sup_{y \neq z} |f(y) - f(z)|/d(y, z).$$

Let $\|\alpha\|_{BL}^* = \sup\{|\int f\,d\alpha| : \|f\|_{BL} \leqq 1\}$, and $\beta(\mu, \nu) = \|\mu - \nu\|_{BL}^*$. Then β metrizes the weak-star topology on $\mathcal{P}(S)$ [4].

Given $F \subset S$ and $\epsilon > 0$ let

$$F^\epsilon = \{x \in S : d(x, y) < \epsilon \quad \text{for some} \quad y \in F\}.$$

Then Prokhorov's metric ρ is defined by

$$\rho(\mu, \nu) = \inf\{\epsilon > 0 : \mu(F) \leqq \nu(F^\epsilon) + \epsilon \quad \text{for all closed} \quad F \subset S\}$$

$$= \inf\{\epsilon > 0 : \nu(F) \leqq \mu(F^\epsilon) + \epsilon \quad \text{for all closed} \quad F \subset S\}$$

where $\mu, \nu \in \mathcal{P}(S)$ ([11], [15], and Proposition 1 of [6]). Here is a first simple result.

2.1. PROPOSITION. *Let $\mu \in \mathcal{P}(S)$ and suppose for some $c > 0$, $N(\mu, \epsilon, \frac{1}{2}) \geqq c\epsilon^{-k}$ for all small enough $\epsilon > 0$. Then there is a $\gamma > 0$ and an n_0 such that if $\nu \in \mathcal{P}(S)$ is concentrated in n points* (e.g. *ν is a value of μ_n*), *$n \geqq n_0$, then*

$$\beta(\mu, \nu) \geqq \gamma n^{-1/k}.$$

PROOF. Let F have n points, $\nu(F) = 1$. Let $f(x) = \min(1, d(x, y) : y \in F)$. Then $\| f \|_{BL} \leqq 2$ and if we let $n + 1 = c\epsilon^{-k}$, then $\int f d(\mu - \nu) > \epsilon/2$ for n large enough, so

$$\beta(\mu, \nu) \geqq \tfrac{1}{4}[(n + 1)/c]^{-1/k},$$

hence the result.

It is not hard to show that the hypothesis of 2.1 holds if μ is any absolutely continuous probability on k-dimensional Euclidean space R^k.

3. β-convergence. If T is a measurable subset of S, then $E\mu_n(T) = \mu(T)$ and $\sigma^2(\mu_n(T)) = (\mu(T) - \mu^2(T))/n$. Let S_j, $j = 1, \cdots, m$, be disjoint measurable sets with union T. Summing over j and using the Schwartz inequality we get

3.1. PROPOSITION. $E\sum_j (\mu_n - \mu)(S_j)^2 = (\mu(T) - \sum_j \mu^2(S_j))/n < \mu(T)/n$, $E\sum_j |(\mu_n - \mu)(S_j)| \leqq (m\mu(T)/n)^{\frac{1}{2}}$.

Note that if μ is nonatomic one can make $\mu(S_j)/\mu(T)$ small for each j (m large, $\mu(T) > 0$) and then $\sum \mu^2(S_j)/\mu(T)$ is small.

3.2. THEOREM. *Suppose that for some real number $k > 2$, there is a $K < \infty$ such that*

$$N(\mu, \epsilon, \epsilon^{k/(k-2)}) \leqq K\epsilon^{-k}$$

whenever $0 < \epsilon \leqq 1$. Then there is an $M = M(k, K) < \infty$ such that $E\beta(\mu_n, \mu) \leqq Mn^{-1/k}$ for all n.

PROOF. For each positive integer r, S is the disjoint union of measurable sets S_{rj}, $j = 0, \cdots, m_r$, where $m_r \leqq K\cdot 3^{k(r+2)}$, for $j \geqq 1$ the diameter of S_{rj} is at most 3^{-r-1}, and $\mu(S_{r0}) \leqq 3^{-k(r+2)/(k-2)}$.

Given a positive integer n let $\epsilon = n^{-1/k}$ and let t be the smallest integer such that $3^{-t} < \epsilon$. Then $3^t \leqq 3/\epsilon$. Let s be the smallest integer such that $3^{-s} < \epsilon^{(k-2)/k}$. Then $3^s \leqq 3\epsilon^{(2-k)/k}$ and $s \leqq t$.

We define sets $A_{t-u,j}$, $j = 1, \cdots, m_{t-u}$, inductively on $u = 0, \cdots, t - s$, as follows. Let $A_{tj} = S_{tj}$. Given the $A_{t-u,j}$, each one which is not included in $S_{t-u-1,0}$ intersects some $S_{t-u-1,q}$, $q \geqq 1$, and we choose such a $q = q(t-u, j)$. Then we let

$$A_{t-u-1,z} = \bigcup \{A_{t-u,j}\colon q(t-u, j) = z\}.$$

Then for each z, we have for diameters

$$\operatorname{diam}(A_{t-u-1,z}) \leqq 2 \max_j \operatorname{diam}(A_{t-u,j}) + 3^{u-t}.$$

Thus by induction on u, the diameter of each A_{rj} is at most 3^{-r} and each $A_{r-1,q}$ is the disjoint union of those A_{rj} which it intersects, $r = s+1, \cdots, t$. Also

$$\bigcup_j A_{rj} \subset S_{r-1,0} \cup \bigcup_q A_{r-1,q}.$$

Let $M_r = \sum_{j=1}^{m_r} |(\mu_n - \mu)(A_{rj})|$. Let $f \in BL(S, d)$, $\|f\|_{BL} \leqq 1$. For each $r = s, \cdots, t$ and $j = 1, \cdots, m_r$ we choose if possible $x_{rj} \in A_{rj}$ and let $f(x_{rj}) = f_{rj}$. Then for $r = s+1, \cdots, t$,

$$|f_{rj} - f_{r-1,q(r,j)}| \leqq 3^{1-r}$$

whenever the left side is defined (i.e. $A_{rj} \not\subset S_{r-1,0}$). Now

$$\begin{aligned} |\textstyle\int f\, d(\mu_n - \mu)| &\leqq (\mu_n + \mu)(S_{t0}) + |\textstyle\sum_{j=1}^{m_t} \int_{S_{tj}} f(x) - f_{tj} + f_{tj}\, d(\mu_n - \mu)(x)| \\ &\leqq (\mu_n + \mu)(S_{t0}) + 3^{-t} + |\textstyle\sum_{j=1}^{m_t} f_{tj}(\mu_n - \mu)(A_{tj})| \\ &\leqq (\mu_n + \mu)(S_{t0} \cup S_{t-1,0}) + \epsilon \\ &\quad + |\textstyle\sum_{q=1}^{m_{t-1}} \sum_{j:q(t,j)=q} (f_{tj} - f_{t-1,q} + f_{t-1,q})(\mu_n - \mu)(A_{tj})| \\ &\leqq (\mu_n + \mu)(S_{t0} \cup S_{t-1,0}) + \epsilon + 3^{1-t} M_t \\ &\quad + |\textstyle\sum_{q=1}^{m_{t-1}} f_{t-1,q}(\mu_n - \mu)(A_{t-1,q})|. \end{aligned}$$

Continuing inductively in this fashion from $r = t$ down to $r = s$ we obtain

$$\beta(\mu_n, \mu) \leqq \epsilon + M_s + \textstyle\sum_{r=s}^{t} \{(\mu_n + \mu)(S_{r0}) + 3^{1-r} M_r\}.$$

Thus by Proposition 3.1

$$\begin{aligned} E\beta(\mu_n, \mu) &\leqq \epsilon + (m_s/n)^{\frac{1}{2}} + \textstyle\sum_{r=s}^{t} \{2 \cdot 3^{-k(r+2)/(k-2)} - 3^{1-r}(3^{k(r+2)} K/n)^{\frac{1}{2}}\} \\ &\leqq \epsilon + (K/n)^{\frac{1}{2}} \{3^{k(s+2)/2} + 27[3^{(k-2)(t+3)/2} - 1]/(3^{(k-2)/2} - 1)\} \\ &\quad + 2 \cdot 3^{-k(s+2)/(k-2)}/(1 - 3^{-k/(k-2)}). \end{aligned}$$

Facts in the second paragraph of this proof and calculation yield $E\beta(\mu_n, \mu) \leqq M\epsilon$ for some $M < \infty$ depending only on k and K, q.e.d.

3.3 COROLLARY. *Let (S, d) be compact. Suppose for some $k > 2$ and $K < \infty$, $N(S, \epsilon) \leqq K\epsilon^{-k}$ whenever $0 < \epsilon \leqq 1$. Then for any $\mu \in \mathcal{P}(S)$, $E\beta(\mu_n, \mu) \leqq M(k, K)n^{-1/k}$ for all n.*

3.4. PROPOSITION. *Suppose S is d-dimensional Euclidean space R^d and $\int |x|^\alpha\, d\mu(x) < \infty$ where $\mu \in \mathcal{P}(S)$ and $\alpha = dk/(k-d)(k-2) > 0$, $d < k$. Then the hypothesis of Theorem 3.2 holds for μ and k.*

PROOF. Let $N = \int |x|^\alpha \, d\mu(x)$, $0 < \epsilon \leqq 1$, and $B_r = \{x: |x| \leqq r\}$, $|x| = (x_1^2 + \cdots + x_d^2)^{\frac{1}{2}}$. Choose r so that

$$\mu(B_{r/2}) \leqq 1 - \epsilon^{k/(k-2)} < \mu(B_r)$$

(we may choose the origin so that $\mu(B_0) = 0$). Then $\epsilon^{k/(k-2)}(r/2)^\alpha \leqq N$, and for some $c < \infty$, $r \leqq c\epsilon^{-(k-d)/d}$ for all ϵ. Let c_d be the (Lebesgue) volume of B_1 in R^d. We choose a maximal set Q of q points of B_r with $|x - y| \geqq \epsilon$ for $x \neq y$ in Q. Then

$$qc_d(\epsilon/2)^d \leqq c_d(r + \epsilon)^d, \qquad q \leqq [2(r + \epsilon)/\epsilon]^d.$$

$B_r \subset \bigcup_{x \varepsilon Q} (x + B_\epsilon)$, so

$$N(\mu, \epsilon, \epsilon^{k/(k-2)}) \leqq q \leqq 2^d(1 + r\epsilon^{-1})^d \leqq 2^d(c\epsilon^{-k/d} + 1)^d \leqq K\epsilon^{-k}$$

for some $K = K(\alpha, N, d) < \infty$, q.e.d.

It is easy to show that if μ is a Gaussian probability on R^d, then the hypothesis of 3.2 holds for any $k > \max(d, 2)$.

Suppose for some $K < \infty$, $N(S, \epsilon) \leqq K\epsilon^{-2}$ whenever $0 < \epsilon \leqq 1$, e.g. S is a bounded subset of R^2. Then we can apply the method of proof of Theorem 3.2, letting each S_{r0} be empty, eliminating the part of the proof concerning s, and inducting from $r = t$ down to $r = 1$. We obtain

$$\beta(\mu_n, \mu) \leqq \epsilon + M_1 + \textstyle\sum_{r=1}^t 3^{1-r} M_r,$$

$$E\beta(\mu_n, \mu) \leqq \epsilon + n^{-\frac{1}{2}}(m_1^{\frac{1}{2}} + 27K^{\frac{1}{2}}t)$$

$$\leqq cn^{-\frac{1}{2}}(1 + \log n) \quad \text{for all} \quad n,$$

where $c = c(m_1, K) < \infty$. I do not know whether the logarithmic factor can be improved or removed.

4. ρ-convergence. We bound the size of $\rho(\mu_n, \mu)$ by relating it to sums $\sum_j |(\mu_n - \mu)(S_j)|$ for suitable sets S_j. We first prove a positive result, then give examples where it is best possible depending on Theorem 5.1 below.

4.1. THEOREM. *For any* $\epsilon > 0$,

$$E\rho(\mu_n, \mu) \leqq n^{-1/(k+2+\epsilon)}$$

for n large enough, where $k = k(\mu)$ as defined in Section 2.

PROOF. Given n let $\delta = n^{-1/(k+2+\epsilon)}$. Then for n large enough, S is the union of at most $\delta^{-k-\epsilon}$ sets A_j of diameter at most 2δ and a set A_0 with $\mu(A_0) < \delta$. Now if F is any measurable set, then

$$\mu_n(F) \leqq \mu(\textstyle\bigcup_{j \geqq 1}\{A_j: A_j \cap F \neq \emptyset\}) + \delta + \sum_{j \geqq 0} |(\mu_n - \mu)(A_j)|$$

$$\leqq \mu(F^{2\delta}) + \delta + \textstyle\sum |(\mu_n - \mu)(A_j)|,$$

so $\rho(\mu_n, \mu) \leqq 2\delta + \sum |(\mu_n - \mu)(A_j)|$. Hence by Proposition 3.1, for n large enough

$$E\rho(\mu_n, \mu) \leqq 2\delta + \delta^{-(k+\epsilon)/2} n^{-\frac{1}{2}} \leqq 3\delta.$$

The factor of 3 is irrelevant as $\epsilon \downarrow 0$, so the proof is complete.

A more precise dimension for μ or S yields a more precise result by about the same proof:

4.2. COROLLARY. *If for some finite k and K, $N(\mu, \delta) \leqq K\delta^{-k}$ for $0 < \delta \leqq 1$, then for some $M = M(k, K) < \infty$,*

$$E\rho(\mu_n, \mu) \leqq Mn^{-1/(k+2)}$$

for all n.

Next, we show, conversely, that the above result cannot be improved (except for finding the least possible constant M) under its hypotheses. It should be noted, however, that if μ is Lebesgue measure on the unit cube in R^k, then the hypothesis of 4.2 holds for the given k but the result is not best possible for $k = 1$ (see Section 6 below), and I don't know whether it is best for $k > 1$ in these cases.

Let $S^d \subset R^d$ be the Cartesian product of d Cantor sets

$$S^1 = \{\textstyle\sum_{j=1}^{\infty} a_j/3^j \colon a_j = 0 \quad \text{or} \quad 2\}.$$

On S^1 we put the Cantor measure μ^1, i.e. the a_j are independent and $\mu^1(a_j = 0) \equiv \frac{1}{2}$. Then on S^d we have a Cartesian product measure μ^d. For $m = 1, 2, \cdots$, let A_{mr}, $r = 1, \cdots, 2^{md}$, be the subsets of S^d where a_j have given values for each co-ordinate and $j = 1, \cdots, m$. Then $\mu^d(A_{mr}) \equiv 2^{-md}$ and the distance from A_{mr} to A_{ms} is at least 3^{-m} for $r \neq s$.

4.3. PROPOSITION. *We have ϵ-entropic dimensions*

$$k(S^d) = k(\mu^d) = d \log 2/\log 3.$$

PROOF. Let $0 < \epsilon < 1/6$ and let m be the positive integer such that $3^{-m-1} \leqq 2\epsilon < 3^{-m}$. If $A \subset S^d$ and $\mu^d(A) > \frac{1}{2}$, then a covering of A by sets of diameter $\leqq 2\epsilon$ must contain at least 2^{md-1} sets, so $N(\mu^d, \epsilon) \geqq 2^{md-1}$, $H(\mu^d, \epsilon) \geqq (md - 1) \log 2$, and $(m + 1) \log 3 \geqq \log (1/2\epsilon)$, so

$$H(\mu^d, \epsilon) \geqq (\log 2)(d[\log 3]^{-1} \log (1/2\epsilon) - d - 1),$$

$$d(\log 2)/(\log 3)^{-1} \leqq k(\mu^d).$$

Conversely $N(S^d, d^{\frac{1}{2}}\epsilon) \leqq 2^{(m+1)d}$,

$$H(S^d, d^{\frac{1}{2}}\epsilon) \leqq (m + 1)\, d \log 2 \leqq d(\log 2)(1 + \log (1/2\epsilon)/\log 3),$$

so letting $\delta = d^{\frac{1}{2}}\epsilon$,

$$H(S^d, \delta) \leqq d(\log 2)(1 + [\log (1/2\delta) + \tfrac{1}{2} \log d]/\log 3),$$

$$k(\mu^d) \leqq k(S^d) \leqq d(\log 2)/(\log 3),$$

and the proof is complete.

4.4. LEMMA. $\sum_{r=1}^{2^{md}} |(\mu - \nu)(A_{mr})| \geqq \gamma$ *implies*

$$\rho(\mu, \nu) \geqq \min(\gamma/2, 3^{-m}) \quad \text{for} \quad \mu, \nu \in \mathcal{P}(S^d).$$

PROOF. Given m let F be a maximal union of A_{mr} such that $(\mu - \nu)(A_{mr})$ have the same sign, choosing the sign so that $|(\mu - \mu_n)(F)| \geqq \gamma/2$. Then letting $\epsilon = 3^{-m}$, either $\mu(F) \geqq \mu_n(F^\epsilon) + \gamma/2$ or $\mu_n(F) \geqq \mu(F^\epsilon) + \gamma/2$, and the Lemma is proved.

4.5. PROPOSITION. *If $\mu = \mu^d$ as above, then for some $\alpha > 0$ we have have*

$$\Pr(\rho(\mu_n, \mu) \geqq \alpha n^{-1/(2+k(\mu))}) \geqq \alpha$$

for all large enough n.

PROOF. By 4.4 and Theorem 5.1 below, there is a $c > 0$ such that if $2^{md} \leqq n$,

$$\Pr(\rho(\mu_n, \mu) \geqq \min(3^{-m}, 2^{md/2}c/2n^{\frac{1}{2}}) > c.$$

Given n, let m be the smallest integer such that $n^{\frac{1}{2}} \leqq 2^{md/2}3^m c/2$. Then for n large enough, $2^{md} \leqq n$, $\Pr(\rho(\mu_n, \mu) \geqq 3^{-m}) > c$, and

$$n^{\frac{1}{2}} > 2^{md/2}3^m c/6 \cdot 2^{d/2},$$

so for some constant $\beta > 0$

$$n^{\frac{1}{2}} > (3^m\beta)^{1+(k/2)} \quad \text{where} \quad k = d \log 2/\log 3,$$

$$n^{1/(2+k)} > 3^m\beta, \qquad 3^{-m} > \beta n^{-1/(2+k)},$$

$$\Pr(\rho(\mu_n, \mu) > \beta n^{-1/(2+k)}) > c.$$

Letting $\alpha = \min(c, \beta)$ the proof is finished.

Note that Proposition 4.5 is more special than the corresponding result for the metric β (Proposition 2.1). Thus it appears that more remains to be done for ρ than for β.

5. Sums over small sets. Here we shall see that the last estimate in Proposition 3.1 is best possible up to a constant factor. Let (S, μ) be any nonatomic probability space and let S be decomposed into m sets A_j with $\mu(A_j) = 1/m$, $j = 1, \cdots, m$. Let

$$S(m, n) = \sum_{j=1}^{m} |(\mu_n - \mu)(A_j)|.$$

Then $S(m, n)$ is a random variable whose distribution does not depend on μ.

5.1. THEOREM. *For some $c > 0$,*

$$\Pr(S(m, n) \geqq c(m/n)^{\frac{1}{2}}) \geqq c$$

for all integers m and n such that $2 \leqq m \leqq n$.

PROOF. We can choose $c > 0$ for any given finite set of values of n and $m \geqq 2$. For fixed m, c can be bounded away from 0 for $n \geqq m$ by the central limit theorem. Thus if the Theorem were false, we could choose $n_i \geqq m_i \to \infty$ such that for $S(m_i, n_i)$ the largest possible values c_i of c approach 0 as $i \to \infty$.

If $n_i/m_i \leqq N < \infty$ for infinitely many i, we may assume it holds for all i. Then it is enough to show that $ES(m, n)$ is bounded away from 0 for $m = m_i$, $n = n_i$ since $S(m, n) \leqq 2$. Equivalently, we show that for some $\alpha > 0$, $E|(\mu_n - \mu)(A_1)| \geqq \alpha(mn)^{-\frac{1}{2}}$ for all i. We may assume that as $i \to \infty$, $n_i/m_i \to \lambda < \infty$. Then $n\mu_n(A_1)$ converges in law to a Poisson random variable φ with $E\varphi = \lambda$ ([8], VI.5). Then $E|\varphi - \lambda| > 0$ since $\lambda \geqq 1$, and $n\mu(A_1) \to \lambda$ as $i \to \infty$. Hence

$$\liminf_{i\to\infty} E|n(\mu_n - \mu)(A_1)| > 0.$$

Thus there is a $\kappa > 0$ such that for all i,

$$E|(\mu_n - \mu)(A_1)| \geqq \kappa/n \geqq \kappa/(Nmn)^{\frac{1}{2}}.$$

This yields the desired conclusion.

Thus we may assume $n_i/m_i \to \infty$ as $i \to \infty$. Given $n = n_i$ and $m = m_i$ large enough, let q be an integer such that $m/4 < q < m/3$. Let

$$B_t = \bigcup_{j=1}^{t-1} A_j, \qquad b_t = \mu_n(B_t), \qquad t = 1, \cdots, q.$$

Let $\mathcal{A}_t$ be the event $b_t \leqq \frac{1}{2}$. Let $\Pr_t$ denote conditional probability given $\mathcal{A}_t$ and the values of $\mu_n(A_j)$, $j = 1, \cdots, t-1$ (a function of these values). For any such values, the distribution of $\mu_n(A_t)$ for $\Pr_t$ is exactly that of $r\nu_r(A_t)/n$ where $r = n(1 - b_t)$ and for any measurable set C,

$$\nu(C) = \mu(C \sim B_t)/\mu(S \sim B_t).$$

On $\mathcal{A}_t$, $r \geqq n/2$.

Now let $r^{\frac{1}{2}}(\nu_r(A_t) - \nu(A_t)) = G_{tr}$. G_{tr} is approximated in law by a Gaussian random variable G with mean 0 and variance

$$\sigma^2 = \nu(A_t) - \nu^2(A_t) \geqq \nu(A_t)/2 \geqq 1/2m.$$

Specifically, by the Berry-Esséen theorem ([1], [7]) there is an absolute constant $K < \infty$ such that for any real number b,

$$\begin{aligned} |\Pr(G_{tr} \leqq b) - \Pr(G \leqq b)| &\leqq KE|G_{t1}|^3/\sigma^3 r^{\frac{1}{2}} \\ &\leqq 24K\nu(A_t)/\nu(A_t)^{3/2} r^{\frac{1}{2}} \\ &\leqq 24K(m/r)^{\frac{1}{2}}. \end{aligned}$$

Thus for any real number κ and $\zeta \geqq 0$,

$$\Pr(|\nu_r(A_t) - \kappa| \leqq \zeta) \leqq \Pr(\beta - 2\zeta r^{\frac{1}{2}} \leqq G \leqq \beta) + \delta_i$$

where $\beta = r^{\frac{1}{2}}(\zeta + \kappa - \nu(A_t))$ and $\delta_i \to 0$ as $i \to \infty$ for $r \geqq n/2$. Now let $\zeta = (mn)^{-\frac{1}{2}}$. Then

$$\Pr(\beta - 2\zeta r^{\frac{1}{2}} \leqq G \leqq \beta) = \Pr(m^{\frac{1}{2}}\beta - 2(r/n)^{\frac{1}{2}} \leqq m^{\frac{1}{2}}G \leqq m^{\frac{1}{2}}\beta).$$

This is the measure of an interval of length $\leqq 2$ for a Gaussian measure of variance $\geq \frac{1}{2}$. Hence it is less than $1 - 2\eta$ for some absolute constant $\eta > 0$. Now

$$\Pr_t(|(\mu_n - \mu)(A_t)| \leqq \zeta/2) = \Pr_t(|\nu_r(A_t) - n\mu(A_t)/r| \leqq n\zeta/2r)$$
$$\leqq \Pr_t(|\nu_r(A_t) - \kappa| \leqq \zeta) \leqq 1 - 2\eta + \delta_i$$

where $\kappa = n\mu(A_t)/r$, for any $\mu_n(A_j)$, $j < t$. Thus for i large enough, we have for each $t = 1, \cdots, q$

$$\Pr_t(|(\mu_n - \mu)(A_t)| > \zeta/2) > \eta$$

for any $\mu_n(A_j)$, $j < t$, on $\mathfrak{A}_t$.

We say we have a "success at the tth trial" if $|(\mu_n - \mu)(A_t)| > \zeta/2$ or if $\mu_n(B_t) > \frac{1}{2}$. Then the conditional probability of such a success, given *any* values of $\mu_n(A_j)$, $j < t$, is at least η. Hence the probability of at least $\eta q/2$ successes in the first q trials is at least what it would be for independent binomial trials with probability η of success in each trial. By the central limit theorem, this probability is $> \frac{1}{2}$ for m and hence q large enough. Then, since $\mu_n(B_q) \geqq \mu_n(B_t)$, $t \leqq q$,

$$\Pr\left(\sum_{j=1}^{q} |(\mu_n - \mu)(A_j)| \geqq \eta q\zeta/4 \quad \text{or} \quad \mu_n(B_q) > \tfrac{1}{2}\right) > \tfrac{1}{2}.$$

Now $\eta q\zeta/4 = \eta q/4(mn)^{\frac{1}{2}} \geqq \eta m^{\frac{1}{2}}/16n^{\frac{1}{2}} \cdot \mu_n(B_q) > \frac{1}{2}$ implies

$$S(m, n) > \tfrac{1}{6} \geqq m^{\frac{1}{2}}/6n^{\frac{1}{2}}.$$

Thus for i large enough

$$c_i \geqq \min(\tfrac{1}{6}, \eta/16),$$

a contradiction, and the proof is complete.

6. The classical case. In this section S is the unit interval [0, 1] and μ is Lebesgue measure. We shall see that $E\beta(\mu_n, \mu)$ and $E\rho(\mu_n, \mu)$ both approach 0 as $n^{-\frac{1}{2}}$ for $n \to \infty$, while $k(\mu) = k(S) = 1$. Thus the rates of convergence $n^{-1/k}$ for β and $n^{-1/(k+2)}$ for ρ do not apply here.

Defining the distribution functions

$$F_n(x) = \mu_n([0, x]),\ F(x) = x,$$

we have $(F_n - F)(0) = (F_n - F)(1) = 0$. Let $p(f) = \|f\|_L + |f(0)|$. Then

$$\sup\{|\textstyle\int_0^1 f\, d(\mu_n - \mu)|: p(f) \leqq 1\} = \sup\{|\textstyle\int_0^1 f'(x)(F_n - F)(x)\, dx|: \sup |f'| \leqq 1\}$$
$$= \textstyle\int_0^1 |(F_n - F)(x)|\, dx.$$

Now $|f(0)| \leqq \|f\|_\infty \leqq p(f)$ on S so $p(f) \leqq \|f\|_{BL} \leqq 2p(f)$, and

$$\beta(\mu_n, \mu) \leqq \textstyle\int_0^1 |F_n - F| \leqq 2\beta(\mu_n, \mu).$$

The functional $\Phi(G) = \int_0^1 |G|$ is defined and continuous for $\|\cdot\|_\infty$ on the space of functions G on S continuous except for at most finitely many jumps. Let $G_n(t) = n^{\frac{1}{2}}(F_n(t) - F(t))$. Then $\Phi(G_n)$ is a well-defined random variable for each n. By

Donsker's theorem [3] (cf. also [5]), $\Phi(G_n)$ converges in law as $n \to \infty$ to $\Phi(x_t)$ where $\{x_t\}$ is a certain Gaussian stochastic process with continuous sample functions, and

$$\Pr(\Phi(x_t) > 0) = \Pr(\Phi(G_n) > 0) = 1$$

for all n. Thus for some $c > 0$,

$$\Pr(\Phi(G_n) > c) > c \quad \text{for all} \quad n.$$

In the converse direction we have the following result, which follows from results of N.V. Smirnov and specifically from [6a], Lemma 2 p. 646.

6.1. PROPOSITION. $\sup_n E \, \|G_n\|_\infty < \infty$.

We infer that for some $M < \infty$,

$$M^{-1}n^{-\frac{1}{2}} \leqq E\beta(\mu_n, \mu) \leqq Mn^{-\frac{1}{2}}$$

for all n.

Now for ρ, we also connect $\rho(\mu_n, \mu)$ to $\|F_n - F\|_\infty$ by the following result.

6.2. PROPOSITION. *For any $\nu \in \mathcal{P}(S)$ with $\nu([0, x]) = G(x)$,*

$$\|G - F\|_\infty/2 \leqq \rho(\mu, \nu) \leqq 2\|G - F\|_\infty .$$

PROOF. If for some $x \in S$,

$$|(G - F)(x)| \geqq 2\epsilon > 0,$$

then either $\nu([0, x]) \geqq \mu([0, x]^\epsilon) + \epsilon$ or

$$\nu([x, 1]) \geqq \mu([x, 1]^\epsilon) + \epsilon.$$

Hence $\rho(\mu, \nu) \geqq \|F - G\|_\infty/2$.

Conversely, suppose $0 < \epsilon < \rho(\mu, \nu)$. We choose a closed set K such that

$$\nu(K) > \mu(K^\epsilon) + \epsilon.$$

We may assume that whenever $x, y \in K$ and $|x - y| < 2\epsilon$, we have $[x, y] \subset K$. Then K is a finite union of disjoint closed intervals $I_j = [b_j, c_j]$, $j = 1, \cdots, m$, where possibly $b_j = c_j$ for some j's. Now

$$\mu(K^\epsilon) = \mu(K) + \lambda\epsilon \quad \text{where} \quad 2m - 2 \leqq \lambda \leqq 2m,$$

so

$$\textstyle\sum_{j=1}^m \nu(I_j) > (\lambda + 1)\epsilon + \sum_{j=1}^m \mu(I_j).$$

Hence for some j,

$$\nu(I_j) \geqq \mu(I_j) + (2m - 1)\epsilon/m, \qquad \epsilon \leqq (\nu - \mu)(I_j) \leqq 2\|G - F\|_\infty .$$

Letting $\epsilon \uparrow \rho(\mu, \nu)$ the proof is complete.

We infer that $n^{\frac{1}{2}}E\rho(\mu_n, \mu)$, like $n^{\frac{1}{2}}E\|F_n - F\|_\infty$, is bounded and bounded away from 0 as $n \to \infty$.

REFERENCES

[1] BERRY, A. C. (1941). The accuracy of the Gaussian approximation to the sum of independent variates. *Trans. Amer. Math. Soc.* **49** 122–136.

[2] BICKEL, P. J. A distribution free version of the Smirnov two sample test in the p-variate case (preprint).

[3] DONSKER, M. D. (1952). Justification and extension of Doob's heuristic approach to the Kolmogorov-Smirnov theorems. *Ann. Math. Statist.* **23** 277–281.

[4] DUDLEY, R. M. (1966). Convergence of Baire measures. *Studia Math.* **27** 251–268.

[5] DUDLEY, R. M. (1966). Weak convergence of probabilities on non-separable metric spaces and empirical measures on Euclidean spaces. *Illinois J. Math.* **10** 109–126.

[5a] DUDLEY, R. M. (1967). Measures on non-separable metric spaces. *Illinois J. Math.* **11** 449–453.

[6] DUDLEY, R. M. (1968). Distances of probability measures and random variables. *Ann. Math. Statist.* **38** 1563–1572.

[6a] DVORETZKY, A., KIEFER, J., and WOLFOWITZ, J. (1956). Asymptotic minimax character of the sample distribution function and of the classical multinomial estimator. *Ann. Math. Statist.* **27** 642–669.

[7] ESSÉEN, C. G. (1945). Fourier analysis of distribution functions. A mathematical study of the Laplace-Gaussian law. *Acta. Math.* **77** 1–125.

[8] FELLER, W. (1950, 1957, 1968). *An Introduction to Probability Theory and its Applications* Wiley, New York.

[9] FORTET, R., and MOURIER, E. (1953). Convergence de la répartition empirique vers la répartition théorique. *Ann. Scient. École Norm. Sup.* **70** 266–285.

[10] POSNER, E. C., RODEMICH, E. R., and RUMSEY, H. (1967). Epsilon entropy of stochastic processes *Ann. Math. Statist.* **38** 1000–1020.

[11] PROKHOROV, YU. V. (1956). Convergence of random processes and limit theorems in probability theory. *Theor. Probability Appl.* **1** 157–214.

[12] RANGA RAO, R. (1962). Relations between weak and uniform convergence of measures and applications. *Ann. Math. Statist.* **33** 659–680.

[13] ROSENBLATT, MURRAY (1952). Remarks on a multivariate transformation. *Ann. Math. Statist.* **23** 470–472.

[14] SEN, P. K., and CHATTERJEE, S. K. (1964). Nonparametric tests for the bivariate two sample location problem. *Calcutta Statist. Assoc. Bull.* **13** 18–58.

[15] STRASSEN, V. (1965). The existence of probability measures with given marginals. *Ann. Math. Statist.* **36** 423–439.

[16] VARADARAJAN, V. S. (1958). On the convergence of sample probability distributions. *Sankhya* **19** 23–26.

[17] WALD, A., and WOLFOWITZ, J. (1944). Statistical tests based on permutations of the observations. *Ann. Math. Statist.* **15** 358–372.