AN APPLICATION OF PROHOROV'S THEOREM TO PROBABILISTIC NUMBER THEORY

BY

PATRICK BILLINGSLEY[1]

The purpose of this paper is to give a probabilistic proof of a theorem which contains as special cases certain results about the natural densities of arithmetically interesting sets of integers, results usually conjectured on probabilistic grounds but proved by nonprobabilistic methods. The principal tool is Prohorov's theorem on the weak compactness of probability measures [8]. Results similar to those of this paper, but for logarithmic density rather than natural density, have been proved, by very different methods, by Paul [7] (see the end of the paper for a comparison).

1. Introduction

Let μ_N be the probability measure on the space of positive integers that places mass $1/N$ at each of the points 1, 2, $\cdots$, N. To ask if a set A of integers has a natural density is to ask if the limit $\lim_N \mu_N(A)$ exists. Thus we are lead to ask whether the measures μ_N converge in some sense. In order to obtain a satisfactory answer, we must first complete the space of integers in some way. The following completion is useful for problems of multiplicative number theory.

Let X be the space of sequences $x = (x_1, x_2, \cdots)$ of nonnegative integers. For each $n \geq 1$ let $\alpha(n) = (\alpha_1(n), \alpha_2(n), \cdots)$, where $\alpha_i(n)$ is the exponent of the i^{th} prime p_i in the factorization of $n = \prod_i p_i^{\alpha_i(n)}$. The mapping α provides a one-to-one correspondence between the set of positive integers and the subset X_0 of X consisting of those x that have only finitely many nonzero coordinates. The completion that we will use is a space X_λ, between X_0 and X $(X_0 \subset X_\lambda \subset X)$, defined as follows. For each component x_i of x, let x_i' be x_i or 0 according as $x_i \leq 1$ or $x_i > 1$, and let x_i'' be x_i or 0 according as $x_i > 1$ or $x_i \leq 1$. For a fixed sequence $\lambda = (\lambda_1, \lambda_2, \cdots)$ of positive constants, let X_λ consist of those points x of X for which each of the sums $\sum_i \lambda_i x_i'$ and $\sum_i x_i''$ is finite. (Of course the second sum is finite if and only if at most finitely many of the x_i exceed 1.) Under the metric

$$d(x, y) = \sum_i \lambda_i \,|\, x_i' - y_i' \,|\, + \sum_i \,|\, x_i'' - y_i'' \,|,$$

X_λ is a complete, separable metrix space. (To see this, identify x with (x', x''), where x' and x'' have coordinates x_i' and x_i'', respectively. With this identification, X_λ is the topological product of two sets, the first [second]

Received August 24, 1963.

[1] This research was carried out in the Department of Statistics, University of Chicago, under partial sponsorship of the Statistics Branch, Office of Naval Research.

being a closed subset of the Banach space of real sequences x' $[x'']$ with $\| x' \| = \sum_i \lambda_i | x_i' |$ $[\| x'' \| = \sum_i | x_i'' |]$ finite.)

Let $\mathfrak{X}_\lambda$ be the σ-field of Borel sets in X_λ, i.e., the σ-field generated by the open sets. We now redefine μ_N to be that probability measure on $\mathfrak{X}_\lambda$ corresponding to a mass of $1/N$ at each of the points $\alpha(1), \alpha(2), \cdots, \alpha(N)$, and we ask whether μ_N converges weakly to some probability measure μ on $\mathfrak{X}_\lambda$. Weak convergence (denoted $\mu_N \Rightarrow \mu$) here means that $\int g\, d\mu_N \to \int g\, d\mu$ for all bounded, continuous functions g on X_λ or, equivalently, that $\mu_N(A) \to \mu(A)$ for all μ-continuity sets A, a μ-continuity set being an element A of $\mathfrak{X}_\lambda$ whose boundary $A^\sim$ satisfies $\mu(A^\sim) = 0$ (see [1] or [8]). If $\mu_N \Rightarrow \mu$ and if A is a μ-continuity set then clearly

$$D\{n : \alpha(n) \in A\} = \mu(A), \tag{1.1}$$

where D denotes natural density. (For an account of density, see [4], for example.) The main result will be that μ_N does converge weakly to an appropriate limit if

$$\textstyle\sum_i \lambda_i'/p_i < \infty, \tag{1.2}$$

where

$$\begin{aligned} \lambda_i' &= \lambda_i \quad \text{if} \quad \lambda_i \leq 1 \\ &= 1 \quad \text{if} \quad \lambda_i > 1. \end{aligned} \tag{1.3}$$

As we will see, this result enables us to compute the natural densities of various number-theoretically interesting sets.

2. Weak convergence in X_λ

In this section we consider arbitrary probability measures ν_N and ν on X_λ. Prohorov [8] has generalized the classical Helly theorem by showing that $\{\nu_N\}$ contains a subsequence that converges weakly (to some ν) if it is *tight* in the sense that for every positive ε there is a compact set K such that $\nu_N(K) > 1 - \varepsilon$ for all N. (Prohorov's theorem is valid in any metric space; the converse also holds if the space is separable and complete.) A routine application of this theorem yields a useful convergence criterion for measures on $\mathfrak{X}_\lambda$.

For each k, the equation $\pi_k(x) = (x_1, \cdots, x_k)$ defines a mapping π_k from X_λ to the space I^k of k-tuples of nonnegative integers; for any k and ν, $\nu\pi_k^{-1}$ is a purely atomic probability measure defined for all subsets of I^k. It is easy to show that if $\nu_N \Rightarrow \nu$ then $\nu_N \pi_k^{-1} \Rightarrow \nu\pi_k^{-1}$ for each k, where here the weak convergence in I^k refers to the discrete topology (i.e., there is convergence for every subset of I^k). Although the reverse implication does not hold in general, it does if one adds the hypothesis that $\{\nu_N\}$ is tight, which is the content of the following result.

LEMMA 2.1. *If $\nu_N \pi_k^{-1} \Rightarrow \nu\pi_k^{-1}$ for each k and if $\{\nu_N\}$ is tight then $\nu_N \Rightarrow \nu$.*

Proof. It is enough to show that any subsequence $\{\nu_{N'}\}$ contains a further subsequence $\{\nu_{N''}\}$ converging weakly to ν. But, since $\{\nu_{N'}\}$ is tight, it follows by Prohorov's theorem that it contains a subsequence $\{\nu_{N''}\}$ converging weakly to some ν_0. To see that ν_0 must coincide with ν, note first that, since

$$\nu_{N''}\pi_k^{-1} \Rightarrow \nu\pi_k^{-1} \quad \text{and} \quad \nu_{N''}\pi_k^{-1} \Rightarrow \nu_0\pi_k^{-1},$$

ν and ν_0 must agree on the field $\mathfrak{X}_\lambda^0$ of sets of the form $\pi_k^{-1}M$ ($k \geq 1$, $M \subset I^k$). Hence it suffices to show that $\mathfrak{X}_\lambda^0$ generates $\mathfrak{X}_\lambda$, which follows from the separability of X_λ, together with the fact that the closed sphere of radius δ about y is the limit of the sets

$$\{x : \textstyle\sum_{i=1}^k \lambda_i | x_i' - y_i' | + \sum_{i=1}^k | x_i'' - y_i'' | \leq \delta\},$$

which all lie in $\mathfrak{X}_\lambda^0$.

To use Lemma 2.1, we need a condition that implies $\{\nu_N\}$ is tight.

LEMMA 2.2. *The sequence $\{\nu_N\}$ is tight if* (i) *for each k the sequence $\{\nu_N \pi_k^{-1}\}$ is tight and* (ii) *for any positive ε and δ there exists an integer k such that*

$$\nu_N\{x : \textstyle\sum_{i=k}^\infty \lambda_i x_i' + \sum_{i=k}^\infty x_i'' > \delta\} < \varepsilon$$

for all N.

Proof. Given a positive ε, we must produce a compact set K such that $\nu_N(K) > 1 - \varepsilon$ for all N. It will be convenient to write

$$r_k(x) = \textstyle\sum_{i=k}^\infty \lambda_i x' + \sum_{i=k}^\infty x_i''. \tag{2.1}$$

There is by (ii) an increasing sequence $(k_1, k_2, \cdots)$ of integers such that

$$\nu_N\{x : r_{k_j}(x) > 1/j\} < \varepsilon/2^{j+1}$$

for all N and j. And there is by (i) a number c such that

$$\nu_N\{x : \textstyle\sum_{i=1}^{k_1} \lambda_i x_i' + \sum_{i=1}^{k_1} x_i'' > c\} < \varepsilon/2$$

for all N. Let A be the set of x for which

$$r_{k_j}(x) \leq 1/j, \qquad j = 1, 2, \cdots, \tag{2.2}$$

and

$$\textstyle\sum_{i=1}^{k_1} \lambda_i x_i' + \sum_{i=1}^{k_1} x_i'' \leq c;$$

take K to be the closure of A. Clearly $\nu_N(K) > 1 - \varepsilon$ for all N. Since only finitely many distinct nonnegative integers can appear in any given coordinate of the members of A, it is possible, by the diagonal method, to select from any sequence in A a subsequence that is in each coordinate eventually constant; (2.2) now implies there is convergence in the sense of the topology of X_λ. Hence K is compact.

3. Weak convergence of the μ_N

Consider again the particular measures μ_N defined in Section 1.

THEOREM 3.1. *If* (1.2) *holds then* $\mu_N \Rightarrow \mu$, *where the measure* μ *on* $\mathfrak{X}_\lambda$ *is uniquely determined by the relation*

$$(3.1)\qquad \mu\{x: x_i = v_i, i = 1, \cdots, k\} = \prod_{i=1}^{k} \left(\frac{1}{p_i}\right)^{v_i} \left(1 - \frac{1}{p_i}\right),$$

valid for any finite sequence $(v_i, \cdots, v_k)$ *of nonnegative integers.*

Proof. The main thing is to show that $\{\mu_N\}$ is tight by verifying that it satisfies conditions (i) and (ii) of Lemma 2.2. We must estimate $\mu_N\{x : r_k(x) > \delta\}$, with $r_k(x)$ defined by (2.1.). Let $\sum [\sum'][\sum'']$ denote summation over those indices i for which $i \geq k$ $[i \geq k$ and $\lambda_i \leq 1]$ $[i \geq k$ and $\lambda_i > 1]$; then

$$(3.2)\qquad \begin{aligned} \mu_N\{x : r_k(x) > \delta\} &\leq \mu_N\{x : \textstyle\sum' \lambda_i x_i' > \delta\} \\ &\quad + \mu_N\{x : \textstyle\sum'' \lambda_i x_i' > 0\} + \mu_N\{x : \sum x_i'' > 0\}. \end{aligned}$$

Using Chebyshev's inequality, we obtain

$$\mu_N\{x : \textstyle\sum' \lambda_i x_i' > \delta\} \leq \delta^{-1} \sum' \lambda_i \displaystyle\int x_i' \mu_N(dx);$$

since

$$\int x_i' \mu_N(dx) = \mu_N\{x : x_i' = 1\} \leq \mu_N\{x : x_i \geq 1\} = \frac{1}{N}\left[\frac{N}{p_i}\right] \leq \frac{1}{p_i},$$

we have

$$\mu_N\{x : \textstyle\sum' \lambda_i x_i' > 0\} \leq \delta^{-1} \sum' \lambda_i/p_i .$$

Moreover

$$\mu_N\{x : \textstyle\sum'' \lambda_i x_i' > 0\} \leq \sum'' \mu_N\{x : x_i' > 0\} \leq \sum'' 1/p_i$$

and

$$\mu_N\{x : \textstyle\sum x_i'' > 0\} \leq \sum \mu_N\{x : x_i'' > 0\} = \displaystyle\sum \frac{1}{N}\left[\frac{N}{p_i^2}\right] \leq \sum \frac{1}{p_i^2} .$$

Applying the last three inequalities to (3.2) and using the definition (1.3) of λ_i', we arrive at

$$\mu_N\{x : r_k(x) > \delta\} \leq (\delta^{-1} + 1) \textstyle\sum_{i=k}^{\infty} \lambda_i'/p_i + \sum_{i=k}^{\infty} 1/p_i^2 .$$

As $k \to \infty$, the right-hand member of this inequality goes to 0, because of the assumption (1.2); hence condition (ii) of Lemma 2.1 is satisfied. Condition (i) is easily verified.

By Prohorov's theorem, there is a subsequence $\{\mu_{N'}\}$ converging weakly to some μ. Since $\mu_{N'} \pi_k^{-1} \Rightarrow \mu\pi_k^{-1}$, it is not hard to show that μ satisfies (3.1). (That μ is then uniquely determined by (3.1) follows from the fact, noted in Section 2, that $\mathfrak{X}_\lambda$ is generated by the field $\mathfrak{X}_\lambda^0$.) Having established the existence of μ, one can show without difficulty that $\mu_N\pi_k^{-1} \Rightarrow \mu\pi_k^{-1}$. Since $\{\mu_N\}$ is tight, $\mu_N \Rightarrow \mu$.

Remark. Theorem 3.1 remains true if $p_1, p_2, \cdots$ is some subsequence of the primes, rather than the entire sequence. Suppose the subsequence satisfies $\sum_i 1/p_i < \infty$; if λ_i is identically 1, then Theorem 3.1 applies. In this case, X_0 is a closed subset of X_λ, the distance between any two elements of X_0 is at least 1, and $\mu(X_0) = 1$. If A is any set in $\mathfrak{X}_\lambda$, and if B is the set of points within distance $\frac{1}{2}$ of $A \cap X_0$, then

$$\{n : \alpha(n) \in A\} = \{n : \alpha(n) \in B\}, \quad \mu(B^\sim) = 0, \quad \text{and} \quad \mu(A) = \mu(B).$$

Thus $D\{n : \alpha(n) \in A\} = \mu(A)$ for any A in $\mathfrak{X}_\lambda$ in this special case. This fact can be restated as a known result: D is a (completely additive) probability measure on the σ-field generated by the class of sets $\{n : p^u \mid n\}$ with $p \in C$ and $u \geq 1$, provided $\sum_{p \in C} 1/p < \infty$.

The relation

$$D\{n : \alpha(n) \in A\} = \mu(A) \qquad (\mu(A^\sim) = 0)$$

implies a number of corollary results. Let h be a continuous function from X_λ to the line, say; if the linear Borel set M is a μh^{-1}-continuity set then $h^{-1}M$ is a μ-continuity set and hence

$$D\{n : h(\alpha(n)) \in M\} = \mu h^{-1}(M). \tag{3.3}$$

(This result also follows if we only assume that $h(x)$ is continuous on a set of μ-measure 1.) Now an additive arithmetic function $f(n)$ has the form

$$f(n) = \sum_i f(p_i^{\alpha_i(n)}),$$

where the numbers $f(p_i^v)$ are arbitrary, except that $f(p_i^0) = f(1) = 0$. If

$$h(x) = \sum_i f(p_i^{x_i}) \tag{3.4}$$

then $f(n) = h(\alpha(n))$. If we impose on the function $f(n)$ conditions that ensure $h(x)$ is well defined and continuous on X_λ then it will follow by (3.3) that

$$D\{n : f(n) \in M\} = \mu\{x : \sum_i f(p_i^{x_i}) \in M\} \tag{3.5}$$

holds for any linear Borel set M such that

$$\mu\{x : \sum_i f(p_i^{x_i}) \in M^\sim\} = 0. \tag{3.6}$$

Put

$$\begin{aligned} f'(p_i) &= f(p_i) \quad \text{if} \quad |f(p_i)| \leq 1 \\ &= 1 \qquad\;\; \text{if} \quad |f(p_i)| > 1. \end{aligned}$$

If $\sum_i |f'(p_i)|/p_i < \infty$ then there is a sequence $\lambda = (\lambda_1, \lambda_2, \cdots)$ of positive constants such that $|f(p_i)| < \lambda_i$ and $\sum_i \lambda_i'/p_i < \infty$. Since (3.4) is then continuous in the topology of X_λ, we may state the following result.

THEOREM 3.2. *If $f(n)$ is an additive arithmetic function such that*

$$\sum_i |f'(p_i)|/p_i < \infty, \tag{3.7}$$

then (3.5) *holds for any linear Borel set M satisfying* (3.6).

This theorem was proved by Erdös in Part II of [2]. If w is a continuity point of the distribution function

$$F(w) = \mu\{x : \textstyle\sum_i f(p_i^{x_i}) \leq w\}$$

then it follows upon taking $M = (-\infty, w]$ in (3.5) that

$$D\{n : f(n) \leq w\} = F(w). \tag{3.8}$$

Erdös stated his result in this form, but he proved it for *all* w. If the series

$$\textstyle\sum_{f(p_i)\neq 0} 1/p_i \tag{3.9}$$

diverges then it follows from a probability theorem of Lévy's that $F(w)$ is everywhere continuous (see [3]); thus (3.8) holds for all w in this case. On the other hand, if (3.9) converges then it follows by the remark after the proof of Theorem 3.1 that (3.5) holds for *all* linear Borel sets M and hence that (3.8) holds for all w.

The above proofs depend strongly on Prohorov's theorem. It is perhaps interesting to note that the use of Prohorov's theorem (presumably to be regarded as nonelementary) enables one to bypass all the sieve arguments (usually regarded as elementary) used in [2]. Another nonelementary approach to these problems is via Fourier analysis (see [4]).

Theorem 3.1 can be used to analyse further the fluctuations of arithmetic functions. For example, take $f(n)$ to be additive, as before, and put

$$h(x) = \sup_k \textstyle\sum_{i=1}^k (f(p_i^{x_i}) - m_i),$$

where

$$m_i = \sum_{v=0}^{\infty} \frac{1}{p_i^v}\left(1 - \frac{1}{p_i}\right) f(p_i^v)$$

is assumed finite. Then (under the hypothesis of Theorem 3.2) $h(x)$ is continuous on X_λ, from which it follows that

$$D\{n : \sup_k \textstyle\sum_{i=1}^k (f(p_i^{\alpha_i(n)}) - m_i) \in M\} = \mu\{x : h(x) \in M\}$$

for μh^{-1}-continuity sets M. The idea is to see how far the partial sums of the components $f(p_i^{\alpha_i(n)})$ of $f(n)$ deviate from their "average" values. To specialize further, if $f(p_i^v) = \log(1 - 1/p_i)$ for $v \geq 1$, then $f(n) = \log \varphi(n)/n$, where $\varphi(n)$ is Euler's function. In this case $\sup_k \sum_{i=1}^k (f(p_i^{\alpha_i(n)}) - m_i)$ reduces to the logarithm of

$$\max_\theta \left\{ \prod_{p \mid n, p \leq \theta} \left(1 - \frac{1}{p}\right) \Big/ \prod_{p \leq \theta} \left(1 - \frac{1}{p}\right)^{1/p} \right\}, \tag{3.10}$$

which thus has a distribution. The numerator in (3.10) is the fraction of integers less than n having in common with n no prime factor not exceeding θ.

Moreover, one could in principle find the joint distributions of arithmetic functions by taking h to be an appropriate mapping into R^2. Thus the present method gives information beyond that contained it Theorem 3.2. On the other hand, in Part II of [2] Erdös proved (3.8) under a hypothesis weaker than (3.7), namely, the hypothesis that the two series $\sum_i f'(p_i)/p_i$ and $\sum_i (f'(p_i))^2/p_i$ converge. Whether a result analogous to Theorem 3.1 holds in this case is unknown to me. The difficulty is that X_λ must be replaced by a space of conditionally convergent series and the Chebyshev inequality used in the proof of Theorem 3.1 must be replaced by an inequality of the Kolmogorov type.

4. Comparison with Paul's results

Theorem 3.1 has an interesting connection with the results of Paul [7]. We first reformulate his results; the reformulation will only be sketched, since nothing really new is involved.

With each subset A of the space X introduced in Section 1 associate the set A^- consisting of the elements of A together with those x in X for which $(x_1, \cdots, x_k, 0, 0, \cdots)$ lies in A for infinitely many values of k. This defines a closure operator which determines a topology $\mathfrak{I}$; if $V_n(x)$ consists of x together with the points $(x_1, \cdots, x_k, 0, 0, \cdots)$ for $k \geq n$, then

$$\{V_n(x) : n \geq 1, x \in X\}$$

is a base for $\mathfrak{I}$. With this topology, X is a locally compact, completely disconnected, completely regular Hausdorff space; it is first countable but not second countable (or even Lindelöf); any compact subset of X is countable; the countable set X_0 is dense in X and is discrete in the relative topology. The σ-field $\mathcal{B}$ of Baire sets (the σ-field generated by the continuous functions) coincides with that generated by the cylinders, or sets of the form

$$\{x : x_i = v_i, i = 1, \cdots, k\}$$

with $(v_1, \cdots, v_k)$ a finite sequence of nonnegative integers. The class $\mathcal{B}$ is properly contained in the σ-field $\mathcal{S}$ of Borel sets; $\mathcal{S}$ contain all subsets of X (in fact, any subset of X is a G_δ).

Let μ_N denote the measure (on $\mathcal{B}$ this time) corresponding to a mass of $1/N$ at the point $\alpha(n)$, $n = 1, 2, \cdots, N$; let μ_N^* denote the measure on $\mathcal{B}$ corresponding to a mass of $n^{-1}/\sum_{k=1}^N k^{-1}$ at the point $\alpha(n)$, $n = 1, 2, \cdots, N$. Since $\mathcal{B}$ coincides with the σ-field generated by the cylinders, it follows by Kolmogorov's existence theorem [5] that there exists a unique probability measure μ on $\mathcal{B}$ satisfying (3.1). (It is interesting to note that μ is not tight, or even τ-smooth; see [6] or [9].) Paul has shown in effect that if $A \subset X_0$ then the upper and lower logarithmic densities of $\{n : \alpha(n) \in A\}$ lie between the μ-measure of the set A^- and that of its interior $(A^-)^0$; this fact can be restated:

$$(4.1) \quad \mu((A^-)^0) \leq \liminf_N \mu_N^*(A) \leq \limsup_N \mu_N^*(A) \leq \mu(A^-) \quad (A \subset X_0).$$

Paul's result is analogous to Theorem 3.1 because it can be shown that (4.1) is equivalent to $\mu_N^* \Rightarrow \mu$. (We are here dealing with weak convergence in a general topological space; see [6] or [9].) Paul's result is in a sense stronger than Theorem 3.1, since the topology $\mathfrak{J}$, relativized to X_λ, is much finer than the topology introduced into X_λ in Section 1. On the other hand, his result fails if logarithmic density is replaced by natural density: it can be shown that μ_N does not converge weakly to μ in $\mathfrak{B}$.

References

1. Patrick Billingsley, *The invariance principle for dependent random variables*, Trans. Amer. Math. Soc., vol. 83 (1956), pp. 250–268.
2. Paul Erdös, *On the density of some sequences of numbers*, J. London Math. Soc., *Part I*, vol. 10 (1935), pp. 120–125; *Part II*, vol. 12 (1937), pp. 7–11; *Part III*, vol. 13 (1938), pp. 119–127.
3. Paul Erdös and Aurel Wintner, *Additive arithmetical functions and statistical independence*, Amer. J. Math., vol. 61 (1939), pp. 713–721.
4. Marc Kac, *Statistical independence in probability, analysis and number theory*, Carus Mathematical Monographs, no. 12, New York, Wiley, 1959.
5. A. N. Kolmogorov, *Grundbegriffe der Wahrscheinlichkeitsrechnung*, Erg. Mat., vol. 2, No. 3 (1933).
6. L. LeCam, *Convergence in distribution of stochastic processes*, Univ. California Publ. Statist., vol. 2 (1957), pp. 207–236.
7. E. M. Paul, *Density in the light of probability theory*, Sankhyā, vol. 24 (1962), pp. 103–114.
8. Yu. V. Prohorov, *Convergence of random processes and limit theorems in probability theory*, Teor. Veroyatnost. i Primenen, vol. 1 (1956), pp. 177–238.
9. V. S. Varadarajan, *Measures in topological spaces*, Mat. Sb., vol. 55 (97) (1961), pp. 35–100.

University of Chicago
Chicago, Illinois